Historical warming consistently decreased size, dispersal and speciation rate of fish

Article

Accepted Version

Avaria-Llautureo, J. ORCID: https://orcid.org/0000-0002-8610-7428, Venditti, C. ORCID: https://orcid.org/0000-0002-6776-2355, Rivadeneira, M. M. ORCID: https://orcid.org/0000-0002-1681-416X, Inostroza-Michael, O., Rivera, R. J. ORCID: https://orcid.org/0000-0001-7903-0314, Hernández, C. E. ORCID: https://orcid.org/0000-0002-9811-2881 and Canales-Aguirre, C. B. ORCID: https://orcid.org/0000-0002-8468-6139
(2021) Historical warming consistently decreased size, dispersal and speciation rate of fish. Nature Climate Change, 11. pp. 787-793. ISSN 1758-678X doi: 10.1038/s41558-021-01123-5 Available at https://centaur.reading.ac.uk/99750/

Title: Warmer temperatures decrease size, dispersal ability and speciation rate in Clupeiform fish

There is an ongoing debate as to whether fish body size will decrease with global warming and how changes in body may impact dispersal ability and speciation rate. Theory predicts that, when fish face warmer temperatures, they grow to smaller adult sizes, undergo a reduction in their ability to move, and increase their probability of speciation. However, evaluations of such predictions are hampered owing to the lack of empirical data spanning both wide temporal and geographical scales. Here, using phylogenetic methods, temperature, and 21,795 globally distributed occurrences for 158 Clupeiform fish species, we show that smaller fish have occurred in warmer waters for over 150 million years, across marine and freshwater realms. Smaller fish have historically moved the shortest distances and at low speeds. In addition, small fish display the lowest probability of giving rise to new species. Further, we found that fish species that displayed high speeds of geographical movement and rates of size evolution experienced higher rates of temperature change in their lineage. These results together with global warming predicts a future where smaller Clupeiform fish that have reduced ability to move over aquatic systems will be more prevalent. In turn, this will result in fewer species contributing to global biodiversity.

A great deal of scientific research seeks to anticipate the impact of human-induced global warming on Earth's biodiversity[1–5]. Compelling evidence suggests that global warming will increase species extinction risk[6–8], but there are hints in the literature pointing to the idea that species have several alternative strategies which might enable them to survive such adversity[2,3,9,10]. Local adaptive changes to decrease body size or tracking of suitable environmental conditions over geographic space have emerged as common responses allowing species survival, especially in fish[8,11–19]. However, it is unknown to what extent fish get smaller with warming[20] and how these climate-induced changes in size will impact the ability of species to track optimal environmental conditions over aquatic systems, i.e., species dispersal ability[4,5,10]. Furthermore, the consequences that the interaction between temperature, size, and dispersal ability may have on speciation is less explored, even though speciation is the principal buffer preventing biodiversity loss in the face of species extinction[21].

Based on previous knowledge, we expect a positive association between fish size and dispersal ability given that bigger species are more efficient in terms of consuming energy for long-distance dispersals[22], and their correlated life history strategies promote resilience under unpredictable environments[23]. Moreover, population genetics theory postulates that organisms with a high capacity to move can increase the gene flow within species; predicting a low probability of population divergence and speciation[24]. When these predictions are taken together, it is expected that the evolution of smaller fish under global warming (Fig. 1a) will decrease their dispersal ability (Fig 1b) but increase the rate at which they contribute with new species to biodiversity by local genetic differentiation (Fig. 1c and d). Nevertheless, there is a big gap between theoretical expectations and evidence owing to the lack of combined data on size evolution, temperature change, species dispersal ability and speciation rates. This patchy evidence comes from the fact that, first, the relationships between size, dispersal, and temperature change have only been evaluated across small temporal scales (i.e. decades)[12,13,17–20,25,26], where the process of speciation cannot be observed. Second, species movement is notoriously difficult to quantify[27–29] so that most studies use data from extremely few individuals within species, measured in recent decades[19].

Here, for the first time, we test these predictions (and potential alternatives; Fig. 1) in Clupeiformes - a highly diverse order of fishes with a worldwide distribution, inhabiting the marine and freshwater realms[30] (Supplementary Figure 1). Clupeiformes include some of the most important species for fisheries[31], such as the anchovy (*Engraulis ringens*), Atlantic herring (*Clupea*

harengus), Japanese pilchard (*Sardinops melanostictus*), Pacific herring (*Clupea pallasi*), and the South American pilchard (*Sardinops sagax*). We evaluated the relationship between water temperature tolerance (WTT) and standard length (SL) across the nodes of the Clupeiformes phylogenetic tree spanning ~150 Myr of evolutionary history (Supplementary Figure 2), and across their full global distribution in the present (Supplementary Figure 1). We estimated the *posterior* distribution of WTT values at phylogenetic nodes, which represent the subset of temperatures to which each species was adapted at the given node age. This does not represent a climate model-based or proxy-based measurement of paleotemperature *per se*. To evaluate the relationship between WTT, SL and the species ability to move over aquatic systems we inferred the historical distance and speed of fish historical movement in a three-dimensional space, using the Geo (Geographical) model[32] (Methods). This phylogenetic model estimates the *posterior* distribution of the estimated ancestral geographical locations for all nodes in a time-calibrated phylogenetic tree – allowing us to have a measure of the distance each species moved per-time unit (speed). Then, we evaluated the effect of SL and dispersal ability on Clupeiformes tip speciation rates.

Our approach also provides information on the rate at which WTT has changed over lineages evolutionary history (phylogenetic branch). Thus, we can uniquely seek to know the range of rates at which the thermal environment of fish has changed (how fast) which, in turn, can reveal how quickly a species adapts. Studying species responses to the rate at which their thermal environmental change is now more pertinent than ever given the alarming accelerating-rates of heating of the oceans[33] and because species and populations respond differently when faced with a fast or slow change in their environment[34,35].

If higher temperatures select smaller fish, we expect to observe a negative relationship between SL and WTT over both evolutionary history and across extant species (Fig. 1a). If size reductions under global warming decrease the ability to move and increase the probability of speciation, we expect to observe a positive relationship between dispersal ability and SL (Fig. 1b) and a negative effect of SL and dispersal ability on speciation rate (Fig. 1c, d; Scenario 1). We evaluated an alternative scenario in which SL reductions and low dispersal ability decrease the probabilities of speciation so that we expect a positive effect of SL and dispersal ability on speciation rate (Fig. 1f, g, Scenario 2). This alternative scenario can have support if dispersal promotes geographical expansion which increase the probabilities of vicariant speciation (range fragmentation by a physical barrier)[36]. Finally, if the rate of climate change can additionally modulate species dispersal and adaptation, the rate of WTT change should has a significant effect of on both the speed of movement and the rate of SL evolution (Fig. 1h-m). The slope of the relationships between rates will differ depending on how species respond when climate changes faster. Both slopes should be positive if species move faster and evolve rapidly (Fig. 1h, i; Scenario 3) – indicating rapid evolution away from its original location. The slopes can be positive and negative if species move faster and evolve slowly (Fig. 1j, k; Scenario 4) – indicating geographic tracking of optimal environmental conditions; and the slopes can be negative and positive if species move slowly and evolve rapidly (Fig. 1l, m; Scenario 5) – indicating rapid local adaptation.

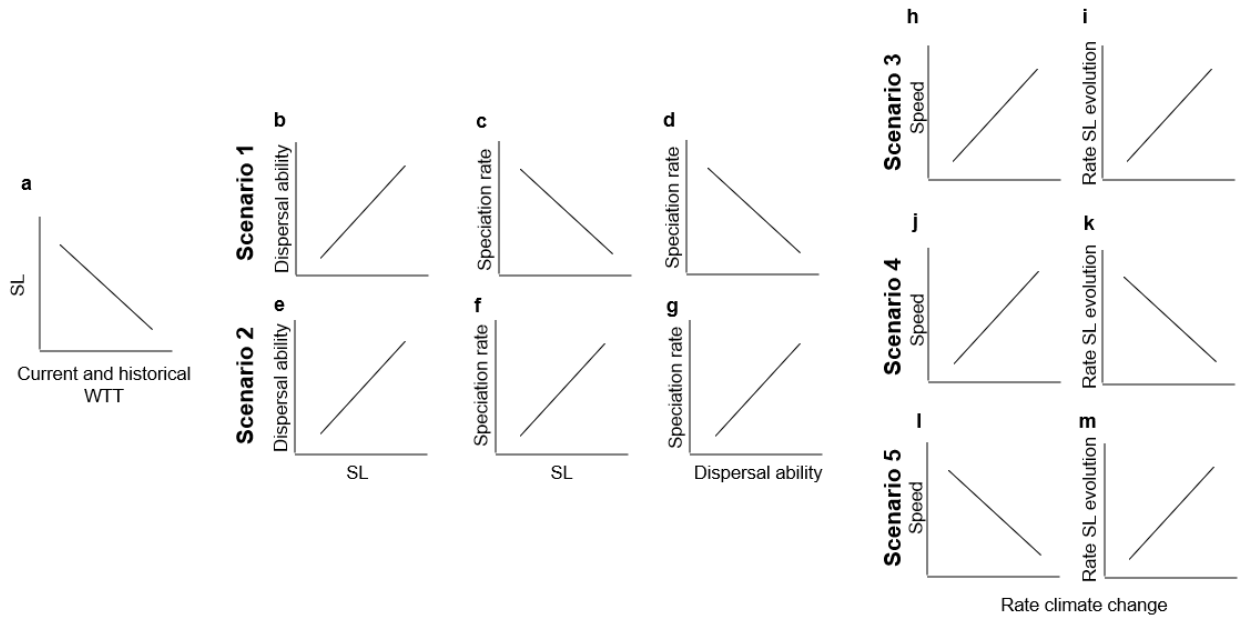


Figure 1. Global warming can impact fish species in multiple ways. a, a negative relationship between standard length (SL) and water temperature tolerance (WTT), across the phylogeny and the extant global distribution of fish, support the idea that warmer temperatures have selected small fish over million years and at wide geographical scales. **b - d,** if small fish are less likely to disperse but more prone to speciate we should observe a positive relationship between dispersal ability and SL (**b**) and a negative effect of dispersal ability and SL on speciation rate (**c**, **d**). **e – g,** if small fish with lower dispersal ability are less prone to speciate we should observe a positive effect of dispersal ability and SL on speciation rates (**h**, **j**). **h – m**, additionally, species can respond differently to the rate at which temperature changes. When temperature changes faster species can move faster and adapt rapidly (**h**, **i**); move faster and adapt slowly (**j**, **k**); or move slowly and adapt rapidly (**i**, **m**).

SL and WTT over current and historical time

We studied the relationship between fish SL and WTT over their extant geographic distribution using the phylogenetic variable rates regression model[37] (Methods). This approach enables the simultaneous estimation of both an overall relationship between SL as a function of WTT across extant species, and any significant shifts in the rate of SL evolution that apply to the phylogenetically structured residual variance in the relationship. We also included the type of migration (diadromous and non-diadromous) as an additional binary variable in the regression, as previous studies show that diadromous fish are larger on average[31]. We used a Bayesian approach, that allows the estimation of regression coefficients while sampling the WTT data within each species. With this approach we can effectively evaluate the effect of WTT on SL while considering the temperature variability over the entire native distributional range of each species (Methods).

Results show that WTT has a significant negative effect on SL across the current geographic distribution of Clupeiformes (Fig. 2a; P_{MCMC} = 0.001). This reveals that smaller Clupeiformes are found in warmer WTT, supporting the "*temperature-size rule*"[38]. Diadromous species were significantly larger than non-diadromous species on average (Supplementary Table 1; P_{MCMC} = 0). Additionally, the variable rate regression did not detect any significant shifts in the rate of SL evolution, and fish SL was better explained by Brownian motion on the scaled phylogeny according to the Pagel's Lambda (λ) parameter (Fig. 2a; Supplementary Table 1).

To study the relationship between fish size and temperature in the deep past, we evaluated the relationship between the *posterior* sample of SL and WTT reconstructed at phylogenetic nodes, which comprises a temporal window of ~150 Myr. To conduct this analysis, we, firstly, inferred the *posterior* distribution of ancestral states of SL across nodes of the λ-scaled phylogeny (Fig. 2b; Methods). Secondly, we inferred the *posterior* distribution of ancestral WTT across nodes of the rate-scaled phylogeny (Fig. 2c) obtained from the variable rate regression between WTT and absolute latitude across the 21,795 occurrence records (Methods; Supplementary Table 2). We found a significant negative association between the *posterior* sample of ancestral SL and WTT (Fig. 2d; $P_{MCMC} = 0$), which support that Clupeiformes evolved smaller sizes under warmer WTT for over 150 Myr (Fig. 2d). Our results agree with the theoretical expectations (Fig. 1a), supporting that warmer temperatures select for smaller Clupeiform fish across large temporal and spatial scales.

Finally, the variable rate regression for WTT indicates that the lower rate of temperature change at which Clupeiformes have adapted is 0.0014 ºC Myr^{-1}, while the upper rate is 0.79 ºC Myr^{-1} (0.000000014 and 0.0000079 ºC per decade, respectively). These historical rates of change of WTT, given our data and approach, are far lower than the average rates of global warming that the planet is experiencing in the last decades; 0.07 ºC per decade since 1880 to 1981, and 0.18 ºC per decade since 1981 (according to the NOAA 2019 Global Climate Summary). These results are comparable to that observed in terrestrial vertebrates[39]. The difference in rates of thermal change we observe might be because of the difference in time scale – millions of years vs decades. However, what is relevant in our results is that the estimated rates of WTT change per phylogenetic branch represent the rates of thermal change to which all species have adapted during their entire life. The thermal environments where species live are highly heritable at phylogenetic scales[35,39], so if some species kept pace with rates of thermal change equal or faster than actual rates, over their entire life, then our methodology is highly likely to detect it. Together, if species are not able to track optimal environmental conditions, then a great part of biodiversity will not be able to adapt to the actual rates of local temperature change.

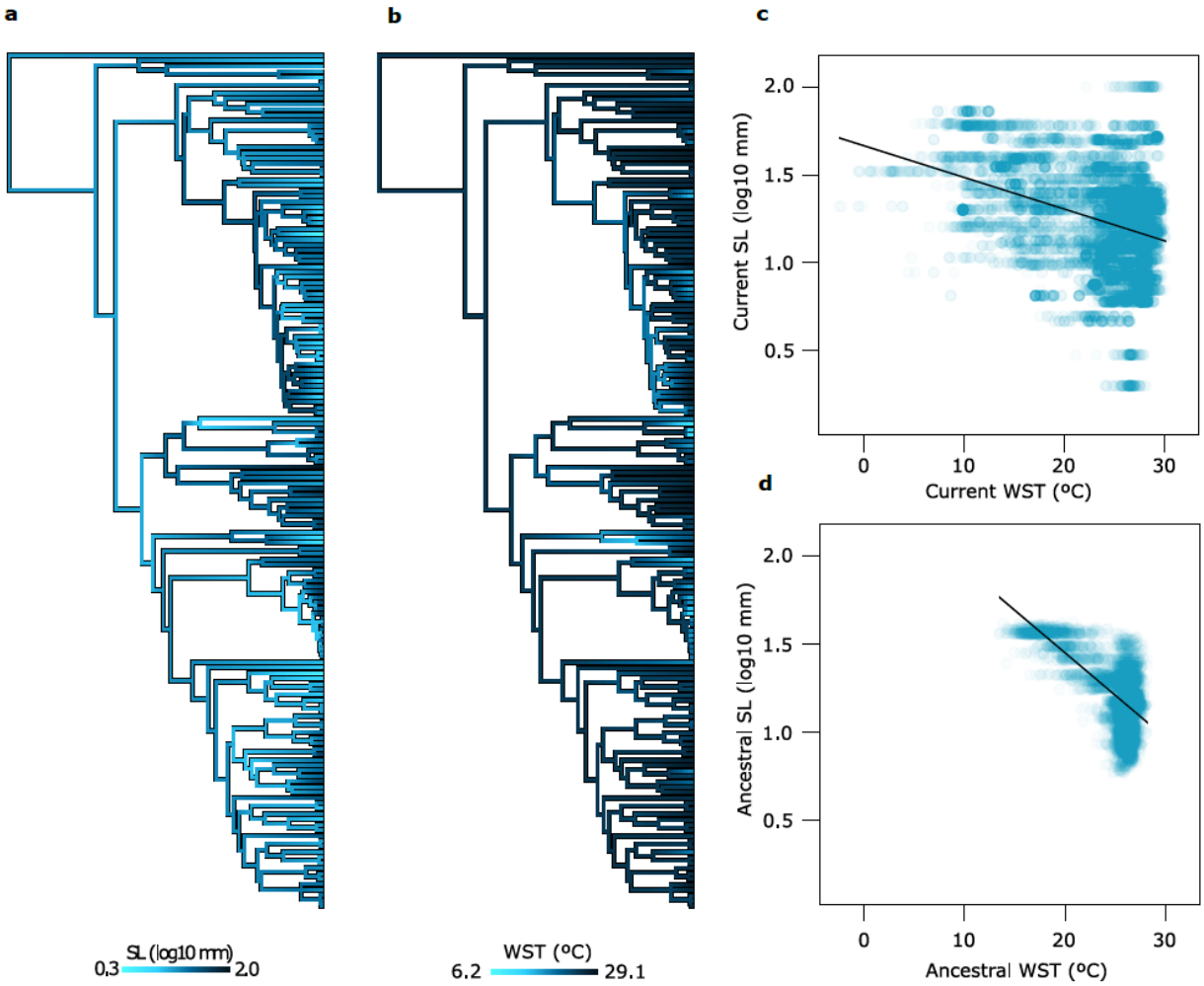


Figure 2. Clupeiformes evolved smaller size in warmer temperatures for million years and in recent times. a, b. Clupeiformes phylogenetic time tree with branches coloured according to the ancestral states for SL (**a**) and WTT (**b**). Ancestral states were estimated using the λ-model and the variable rate regression model for SL and WTT, respectively. **c.** Bayesian phylogenetic generalized least squares sustain that SL and WTT are negatively correlated across extant species (P_{MCMC} = 0.001; *n* = 158,000 observations sampled from extant species data). The black line represents the posterior mean slope of the phylogenetic regression, which was estimated while sampling within species WTT data. **d.** Bayesian generalized least squares shows a significant negative correlation between the ancestral SL and WTT values across nodes (P_{MCMC} = 0; *n* = 157,000 observations sampled from the *posterior* distribution of the estimated ancestral states across phylogenetic nodes). These results support the prediction in Fig. 1a. Line equation in **c**: y = 1.3 + 0.15(Diadromus) - 0.0077(WTT). Line equation in **d**: y = 2.44 - 0.047(WTT).

SL and dispersal ability

The geographic analyses support a model with significant variation in the speed of fish movement across phylogenetic branches (Supplementary Table 3). This implies that the current spatial diversity of Clupeiformes has been assembled by species dispersal at variable speed from the location of the most recent common ancestor (MRCA) of the group (Clupeoidei, after excluding *Denticeps clupeoides*; see Methods). The highest *posterior* density for the geographic distribution of the MRCA indicates that this ancestral species was distributed between the western Tethys Ocean and eastern of Proto Atlantic Ocean, mainly between Eurasia and Africa, around 111 Mya (Fig. 3a). The ancestral reconstruction of the habitat type for the MRCA supports that it was more

likely a marine species (Fig. 3b). However, there are some *posterior* coordinates that fall on continents, and there is also some *posterior* probability indicating that the MRCA inhabited freshwater realms (0.29; Fig. 3b) or both freshwater and marine realms (0.29; Fig. 3b). This degree of uncertainty in the estimation of location and the habitat type (see Supplementary Figure 3 for more node locations) suggests that the MRCA was a marine species with ability to occupy the freshwater space. The biology of the living species of Clupeoidei support this notion because there are living species adapted to live in both freshwater (rivers, lakes, swamps) and saline waters (estuaries, bays, sea).

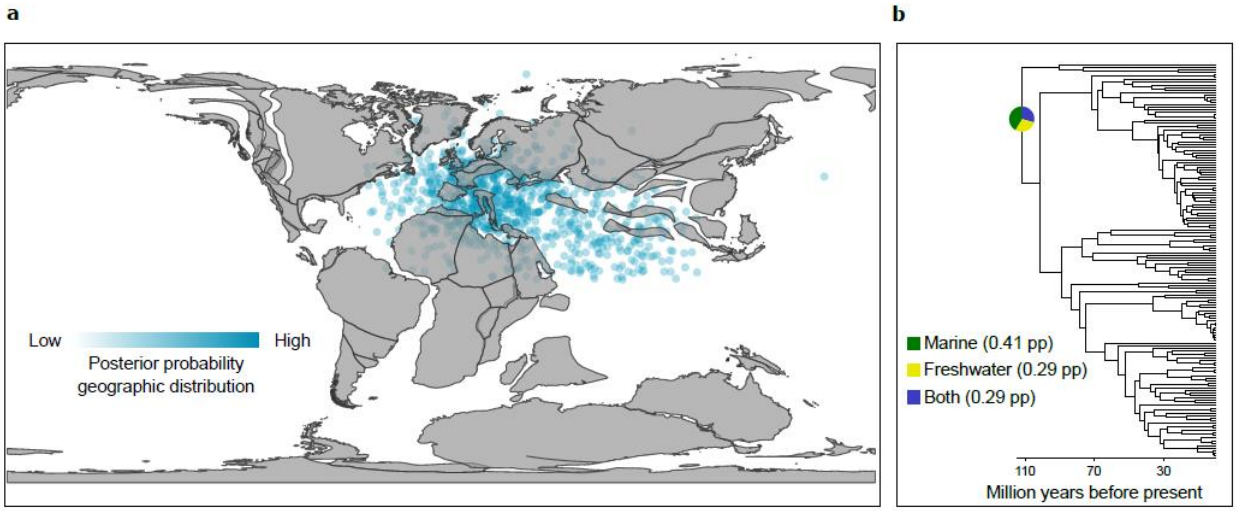


Figure 3. The ancestor of Clupeoidei was distributed across the western Tethys Ocean and the eastern of the Proto Atlantic Ocean 111 million years ago. a. *posterior* geographic distribution of the phylogenetic node representing the ancestor of Clupeoidei. **b.** habitat type for the ancestor of Clupeoidei. pp: *posterior* probability.

When we calculate the total distance that each species dispersed - along the lineage leading from the MRCA to the living species (Supplementary Figure 4) - we observe that the shortest distance was taken by the lineage of *Chirocentrus dorab* (9, 608 km) while the largest distance by the lineage of *Engraulis australis* (53,885 km). Note that this total distance was calculated across the geographic centroids of the *posterior* locations at each phylogenetic node. However, the distances dispersed vary due to the uncertainty in the estimation of species at each phylogenetic node (Figure 3; Supplementary Figure 3). Thus, this uncertainty should be considered when studying the correlates of species movement. We evaluated the effect of SL on the total distance moved for each species from the MRCA (pathwise distance; Methods), and the median of the branch-specific speed of movement along the path that links the MRCA with extant species (pathwise speed; Methods) considering the uncertainty in the estimated ancestral locations. These relationships were evaluated using Bayesian phylogenetic regression models that include the *posterior* sample of 1,000 pathwise distances and speeds for each species in the estimation of regression coefficients (Methods). Results show that SL correlates positively with both the pathwise distances and the pathwise speed of movement (Fig. 4b and c; Supplementary Table 4 and 5, respectively). There were no significant differences in either the mean pathwise distances or the pathwise speed of movement travelled by diadromous and non-diadromous species (Supplementary Table 4 and 5). These results agree with theoretical expectations (Fig. 1b and e). Smaller fish have had a reduced ability to disperse through water bodies over their evolutionary history. They may find it hard to track suitable temperatures over geological time, thus making them more prone to extinction if they cannot keep pace with the actual rates of local heating of the oceans.

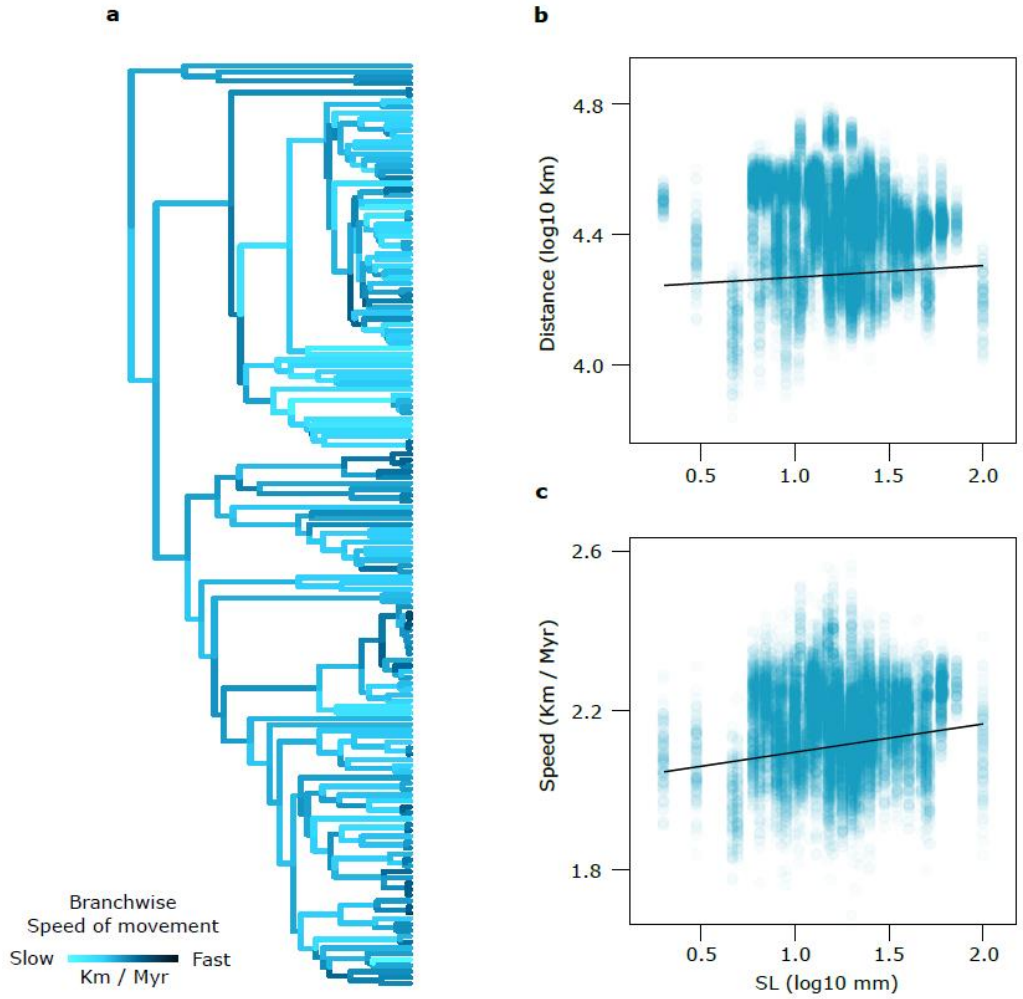


Figure 4. Fish dispersal ability depend on body size. a. Clupeiformes phylogenetic tree with branches coloured according to the speed of movement. **b.** Bayesian phylogenetic generalized least squares show that pathwise distance correlates positively with SL (*BF* > 5; *n* = 157,000 observations sampled from species data). **c.** SL has also a significant positive effect on pathwise speed of movement (*BF* > 10; *n* = 157,000 observations sampled from species data). These results support the prediction in Fig. 1b and e. Black lines (b and c) represent the mean slope estimated from the *posterior* distribution of phylogenetic slopes. Line equation in **b**: y = 4.21 + 0.026(SL). Line equation in **c**: y = 2.02 + 0.071(SL).

Fish response to the historical rate of WTT change

We evaluated the effect that the rates of WTT change may have on both the rates of SL evolution and the speed of movement across all branches of the Clupeiformes phylogeny, using Bayesian GLS regressions that use samples of the data. We included the *posterior* sample of 1,000 branchwise rates estimated at each phylogenetic branch as sample data in regression analyses (Methods). All branchwise rates were estimated by dividing the scaled branches (with the λ-model for SL, the variable rate regression model for WTT, and the variable rate Geo model for speed) with original branch lengths measured in time. The rate of WTT change had a positive effect on both the rate of SL evolution and the speed of fish movement (P_{MCMC} = 0, Fig. 5a, b), meaning that the SL of Clupeiformes have evolved rapidly, and they have dispersed faster when the temperature of their aquatic environments changed at higher rates. These results agree with theoretical expectation in Fig. 1h and b (Scenario 3), indicating that clupeiforms have evolved

rapidly, away from its original location, when climate changed faster. Under accelerated rates of warming clupeiforms will evolve rapidly towards smaller sizes, concomitantly losing their ability to move as size and temperature correlates negatively.

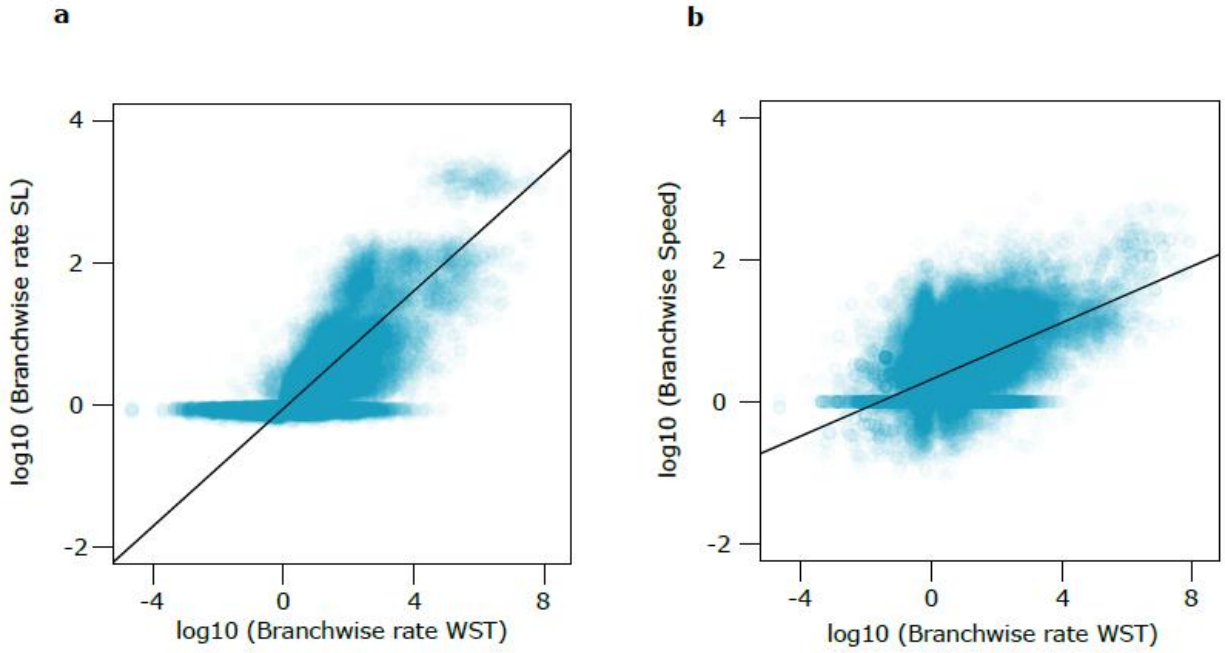


Figure 5. Clupeiformes have evolved rapidly and moved faster when temperature changed at higher rates. a. Bayesian generalized least squares support that the branchwise rates of SL evolution are positively correlated with the branchwise rates of WTT change (P_{MCMC} = 0; n = 314,000 phylogenetic branches). **b.** The branchwise speed of fish movement are also positively correlated with the branchwise rates of WTT change (P_{MCMC} = 0; n = 312,000 phylogenetic branches). These results support the predictions in Fig. 1h and i. Black lines represent the mean slope estimated from the *posterior* distribution of phylogenetic slopes. Line equation in **a**: y = -0.05 + 0.41(Branchwise rate WTT). Line equation in **b**: y = 0.31 + 0.2(Branchwise rate WTT).

Effect of SL and dispersal ability on speciation rates

We evaluated the relationship between speciation with dispersal ability and SL of Clupeiformes. We used Bayesian phylogenetic regression models that include the uncertainty in parameter estimation and samples of dispersal ability within species (Methods). Our results show that the independent additive effect of pathwise distance and pathwise speed were significant (P_{MCMC} = 0.04 and 0 respectively; Supplementary Table 6) – species that move longer distances and faster were more likely to originate new species – supporting theoretical predictions in Fig. 1d. SL did not have a significant effect on speciation when its independent additive effect or their interaction with dispersal ability was evaluated (Supplementary Table 6), rejecting theoretical predictions in Fig. 1c and f. These results suggest that fish SL, by its positive association with dispersal ability, has an indirect effect on speciation rates. We speculate that SL by itself does not related to speciation rate in clupeiform fish because natural selection on SL has not split populations in two or more isolated groups (i.e., selection was not disruptive). Considering that warmer temperatures have selected for smaller fish (and colder temperatures for bigger one) we also speculate that changes in temperature generated a process of directional selection on fish SL – moving the population mean towards smaller values when the temperature increased and vice versa. Directional selection is unlikely to split populations, which can explain why we do not observe a significant relationship between SL and speciation rates. The fact that higher dispersal ability does

correlate positively with speciation rates point to the idea that speciation in clupeiforms was determined principally by geographic process - species moving further and faster could be more likely to experience geographic isolation. Taking together, the speciation rates of smaller fish that move slowly are lower than the speciation rates of their larger counterparts that moved faster and larger distances. A scenario of smaller fish under global warming may cause a decrease of speciation rate in fish, a phenomenon that can impoverish future biodiversity[21].

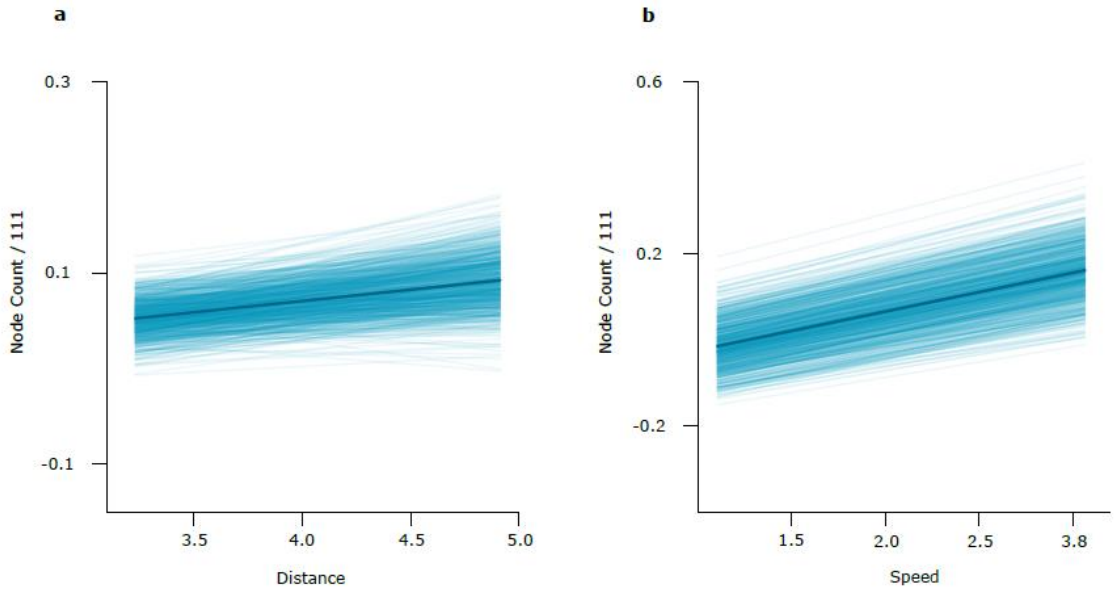


Figure 6. Clupeiformes with lower dispersal abilities have lower probabilities of originate new species. a - b. The Bayesian phylogenetic generalized least squares show that the pathwise distance of movement and the pathwise speed of movement has a positive effect on speciation (P_{MCMC} = 0.04 and 0, respectively; *n* = 157,000 observations sampled from species data). These results support the prediction in Fig. 1g. Lighter lines show the *posterior* distribution of slopes and dark lines shows the *posterior* mean slopes. These slopes were estimated while sampling the pathwise distance and speed within species (Methods). Line equation for **a** and **b**: y = -0.21 + 0.023(Distance) + 0.09(Speed). The node count values were divided by the tree length after excluding *D. clupeoides* (111 Myr).

Conclusion

Global change poses double jeopardy for fish body size, as both overfishing[40] and climate drive populations towards smaller sizes. The phenomena of fish shrinking when facing hotter waters is general in the evolutionary history of Clupeiformes and over their entire worldwide geographic distribution. Provided that smaller Clupeiform fish adapted to warmer conditions are less capable of disperse and in turn less able to originate new species, the scenario of global warming could limit both their ability to find optimal environments to live and their capacity to buffer their increasing extinction risk by the process of speciation. Furthermore, Clupeiform fish living in the present are the survivors of a long evolutionary history under variable rates of temperature change. They have responded to such historical changes by SL adaptation and dispersal at variable rate and speed, respectively. However, such evolutionary processes have never involved the current accelerating rates of heating of the water bodies. Clupeiformes will probably face an increasing risk of extinction. This conclusion can be generalized to other fish if body size, dispersal abilities, and speciation rates relates to each other as in Clupeiformes.

References

1. Parmesan, C. Ecological and Evolutionary Responses to Recent Climate Change. *Annu. Rev. Ecol. Evol. Syst.* **37**, 637–669 (2006).
2. Sheridan, J. A. & Bickford, D. Shrinking body size as an ecological response to climate change. *Nat. Clim. Chang.* **1**, 401–406 (2011).
3. Gardner, J. L., Peters, A., Kearney, M. R., Joseph, L. & Heinsohn, R. Declining body size: A third universal response to warming? *Trends Ecol. Evol.* **26**, 285–291 (2011).
4. McCauley, S. J. & Mabry, K. E. Climate change, body size, and phenotype dependent dispersal. *Trends Ecol. Evol.* **26**, 554–555 (2011).
5. Norberg, J., Urban, M. C., Vellend, M., Klausmeier, C. A. & Loeuille, N. Eco-evolutionary responses of biodiversity to climate change. *Nat. Clim. Chang.* **2**, 747–751 (2012).
6. Amigo, I. The Amazon's fragile future. **578**, 505–507 (2020).
7. Reddin, C. J., Nätscher, P. S., Kocsis, Á. T., Pörtner, H. O. & Kiessling, W. Marine clade sensitivities to climate change conform across timescales. *Nat. Clim. Chang.* **10**, (2020).
8. Comte, L. & Olden, J. D. Climatic vulnerability of the world's freshwater and marine fishes. *Nat. Clim. Chang.* **7**, 718–722 (2017).
9. Skelly, D. K. *et al.* Evolutionary responses to climate change. *Conserv. Biol.* **21**, 1353–1355 (2007).
10. Chen, I., Hill, J. K., Ohlemûller, R., Roy, D. B. & Thomas, C. D. Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science (80-.).* **1024**, 17–20 (2012).
11. Cheung, W. W. L. *et al.* Projecting global marine biodiversity impacts under climate change scenarios. *Fish Fish.* **10**, 235–251 (2009).
12. Cheung, W. W. L. *et al.* Shrinking of fishes exacerbates impacts of global ocean changes on marine ecosystems. *Nat. Clim. Chang.* **3**, 254–258 (2013).
13. Crozier, L. G. & Hutchings, J. A. Plastic and evolutionary responses to climate change in fish. *Evol. Appl.* **7**, 68–87 (2014).
14. Travis, J. M. J. *et al.* Dispersal and species' responses to climate change. *Oikos* **122**, 1532–1540 (2013).
15. Pauly, D. & Cheung, W. W. L. Sound physiological knowledge and principles in modeling shrinking of fishes under climate change. *Glob. Chang. Biol.* **24**, e15–e26 (2018).
16. Tamario, C., Sunde, J., Petersson, E., Tibblin, P. & Forsman, A. Ecological and Evolutionary Consequences of Environmental Change and Management Actions for Migrating Fish. *Front. Ecol. Evol.* **7**, 1–24 (2019).
17. Ljungström, G., Claireaux, M., Fiksen, Ø. & Jørgensen, C. Body size adaptions under climate change: zooplankton community more important than temperature or food abundance in model of a zooplanktivorous fish. *Mar. Ecol. Prog. Ser.* **636**, 1–18 (2020).
18. Daufresne, M., Lengfellner, K. & Sommer, U. Global warming benefits the small in aquatic ecosystems. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12788–12793 (2009).
19. Lenoir, J. *et al.* Species better track climate warming in the oceans than on land. *Nat. Ecol. Evol.* (2020) doi:10.1038/s41559-020-1198-2.
20. Audzijonyte, A. *et al.* Fish body sizes change with temperature but not all species shrink with warming. *Nat. Ecol. Evol.* 1–6 (2020) doi:10.1038/s41559-020-1171-0.
21. Rosenzweig, M. L. Loss of speciation rate will impoverish future diversity. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 5404–5410 (2001).
22. Burns, M. D. & Bloom, D. D. Migratory lineages rapidly evolve larger body sizes than non-migratory relatives in ray-finned fishes. *Proceedings. Biol. Sci.* **287**, 20192615 (2020).
23. Comte, L. & Olden, J. D. Evidence for dispersal syndromes in freshwater fishes. *Proc. R. Soc. B Biol. Sci.* **285**, (2018).
24. Bohonak, A. J. Dispersal, gene flow, and population structure. *Q. Rev. Biol.* **74**, 21–45

(1999).
25. Perry, A. L., Low, P. J., Ellis, J. R. & Reynolds, J. D. Climate change and distribution shifts in marine fishes. *Science (80-.).* **308**, 1912–1915 (2005).
26. Pinsky, M. L., Worm, B., Fogarty, M. J., Sarmiento, J. L. & Levin, S. A. Marine taxa track local climate velocities. *Science (80-.).* **341**, 1239–1242 (2013).
27. Stevens, V. M. *et al.* A comparative analysis of dispersal syndromes in terrestrial and semi-terrestrial animals. *Ecol. Lett.* **17**, 1039–1052 (2014).
28. Dieckmann, U., O'Hara, B. & Weisser, W. The evolutionary ecology of dispersal. *Trends Ecol. Evol.* **14**, 88–90 (1999).
29. Kokko, H. & López-Sepulcre, A. From individual dispersal to species ranges: Perspectives for a changing world. *Science (80-.).* **313**, 789–791 (2006).
30. Lavoué, S., Miya, M., Musikasinthorn, P., Chen, W. J. & Nishida, M. Mitogenomic Evidence for an Indo-West Pacific Origin of the Clupeoidei (Teleostei: Clupeiformes). *PLoS One* **8**, (2013).
31. Bloom, D. D., Burns, M. D. & Schriever, T. A. Evolution of body size and trophic position in migratory fishes: A phylogenetic comparative analysis of Clupeiformes (anchovies, herring, shad and allies). *Biol. J. Linn. Soc.* **125**, 302–314 (2018).
32. O'Donovan, C., Meade, A. & Venditti, C. Dinosaurs reveal the geographical signature of an evolutionary radiation. *Nat. Ecol. Evol.* **2**, 452–458 (2018).
33. Cheng, L. *et al.* Record-Setting Ocean Warmth Continued in 2019. *Adv. Atmos. Sci.* **37**, 137–142 (2020).
34. Pinek, L., Mansour, I., Lakovic, M., Ryo, M. & Rillig, M. C. Rate of environmental change across scales in ecology. *Biol. Rev.* **1**, (2020).
35. Avaria-Llautureo, J., Hernández, C. E., Rodríguez-Serrano, E. & Venditti, C. The decoupled nature of basal metabolic rate and body temperature in endotherm evolution. *Nature* **572**, 651–654 (2019).
36. Gaston, K. J. Species-range size distributions: products of speciation, extinction and transformation. *Philos. Trans. R. Soc. B Biol. Sci.* **353**, 219–230 (1998).
37. Baker, J., Meade, A., Pagel, M. & Venditti, C. Positive phenotypic selection inferred from phylogenies. *Biol. J. Linn. Soc.* **118**, 95–115 (2016).
38. Angilletta, M. J. & Dunham, A. E. The Temperature-Size Rule in Ectotherms: Simple Evolutionary Explanations May Not Be General. *Am. Nat.* **162**, 332–342 (2003).
39. Quintero, I. & Wiens, J. J. Rates of projected climate change dramatically exceed past rates of climatic niche evolution among vertebrate species. *Ecol. Lett.* **16**, 1095–1103 (2013).
40. Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres Jr, F. Fishing Down Marine Food Webs. *Science (80-.).* **279**, 860–863 (1998).
41. Rabosky, D. L. *et al.* An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**, 392–395 (2018).
42. Whitehead, P. J. P. FAO Species Catalogue: Vol. 7 Clupeoid Fishes of the World. *FAO Fish. synopsis* **7**, 303 (1985).
43. Charnov, E. L. & Berrigan, D. Evolution of life history parameters in animals with indeterminate growth, particularly fish. *Evol. Ecol.* **5**, 63–68 (1991).
44. Önsoy, B., Tarkan, A. S., Filiz, H. & Bilge, G. Determination of the best length measurement of fish. *North. West. J. Zool.* **7**, 178–180 (2011).
45. Mohseni, O. & Stefan, H. G. Stream temperature/air temperature relationship: A physical interpretation. *J. Hydrol.* **218**, 128–141 (1999).
46. Morrill, J. C., Bales, R. C. & Conklin, M. H. Estimating stream temperature from air temperature: Implications for future water quality. *J. Environ. Eng.* **131**, 139–146 (2005).
47. Sharma, S., Jackson, D. A., Minns, C. K. & Shuter, B. J. Will northern fish populations be in hot water because of climate change? *Glob. Chang. Biol.* **13**, 2052–2064 (2007).

48. Pagel, M., Meade, A. & Barker, D. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**, 673–684 (2004).
49. Venditti, C., Meade, A. & Pagel, M. Multiple routes to mammalian diversity. *Nature* **479**, 393–396 (2011).
50. Kocsis, Á. T. & Raja, N. B. chronosphere: Earth system history variables. (2020) doi:10.5281/zenodo.3530703.
51. Raftery, A. E. Hypothesis testing and model selection. in *Markov Chain Monte Carlo in Practice* (eds. Gilks, W., Richardson, S. & Spiegelhalter, D.) 163–187 (Chapman & Hall, 1996).
52. Hijmans, R. J. geosphere: Spherical Trigonometry. R package version 1.5-10. https://CRAN.R-project.org/package=geosphere. (2019).
53. Harvey, M. G. & Rabosky, D. L. Continuous traits and speciation rates: Alternatives to state-dependent diversification models. *Methods Ecol. Evol.* **9**, 984–993 (2018).
54. Title, P. O. & Rabosky, D. L. Tip rates, phylogenies and diversification: What are we estimating, and how good are the estimates? *Methods Ecol. Evol.* **10**, 821–834 (2019).
55. Louca, S. & Pennell, M. W. Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505 (2020).
56. Shafir, A., Azouri, D., Goldberg, E. E. & Mayrose, I. Heterogeneity in the rate of molecular sequence evolution substantially impacts the accuracy of detecting shifts in diversification rates. *Evolution (N. Y).* (2020) doi:https://doi.org/10.1111/evo.14036.
57. Ganzach, Y. Misleading Interaction and Curvilinear Terms. *Psychol. Methods* **2**, 235–247 (1997).
58. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
59. Lunt, D. J. *et al.* Palaeogeographic controls on climate and proxy interpretation. *Clim. Past* **12**, 1181–1198 (2016).

Methods

Data. Analyses were performed on the most recent time-calibrated phylogeny of 158 Clupeiformes species (Supplementary Figure 2). This phylogeny was obtained from The Fish Tree of Life[41]. We used the maximum Standard Length (SL) in mm for these 158 species (Supplementary Table 7). We obtained the SL from FishBase and the FAO Species Catalogue for clupeoid fishes[42]. *Sensu* the FishBase System Glossary, the fish SL is the measurement from the most anterior tip of the body to the mid lateral posterior edge of the hypural plate (in fish with a hypural plate) or to the posterior end of the vertebral column (in fish lacking hypural plates). The maximum SL was used because of three reasons. First, maximum SL is preferred over mean SL because fishes have indeterminate growth[43]. Second, it is a more stable measure of size in teleosts to compare museum and collection samples[44]. Third, and most important, individuals that are commonly larger than the population average, and are outside the central distribution of size, are likely the individuals that allow the species to shift their geographic ranges[4]. 21,795 georeferenced occurrences (Supplementary Figure 1; Supplementary Table 7) were obtained from marine and freshwater bodies (i.e., rivers and lakes) from Aquamaps (https://www.aquamaps.org/) and the IUCN (https://www.iucnredlist.org/) respectively. We obtained the geographic locations (within the native range) of 116 species available in Aquamaps, and locations within the polygon of distribution for 42 additional species available in the IUCN. To obtain the geographic locations from the IUCN, we sampled 100 random locations within each species polygon. All georeferenced occurrences were matched with information of water temperature, which represent water temperature tolerances for species (WTT; Supplementary

Table 7). For marine species, we used the mean annual sea surface temperature estimated from the Aquamaps database. For freshwater species, the mean annual air temperatures estimated from the WordClim database (https://worldclim.org/) were used as a first-order proxy of the water surface temperature of the freshwater bodies[45–47]. By maximizing the number of locations and temperature records per species, instead of using single estimates (e.g., mean temperature at the geographic centroid of species distributional range) we can produce more precise estimates of both the ancestral locations and the ancestral thermal environments where Clupeiformes inhabited. Finally, information about the type of migration for each species (diadromous, non-diadromous) was obtained from Bloom et al[31] (Supplementary Table 7).

Inferring ancestral locations. From the geographic locations within each species in the Clupeiformes phylogeny, we inferred the ancestral geo-distribution in a continuous, three-dimensional space. Ancestral locations were estimated for each phylogenetic node using the Geo model[32] in the computer program BayesTraits 3.0[48]. This model estimates the *posterior* distribution of ancestral locations measured in longitude and latitude, while sampling across all location-data within species, and considering the spherical nature of Earth. This natural assumption of the Earth as a spherical object avoids the erroneous calculation of distances between the inferred ancestral locations due to the non-continuity of the longitude scale. When based on a time-calibrated phylogeny, the Geo model simultaneously estimates the speed of species movement across each branch that links pairs of phylogenetic nodes (branchwise speed of movement). Additionally, the ancestral locations across phylogenetic nodes are estimated while considering the continuous variation in dispersal ability of each ancestral species – ranging from species quiescence (no movement), through constant movement in direct proportion of the passage of time, to fast species movement. Estimation of the branchwise speed of species movement are based on the variable rates model[49], which detects shifts away from a background rate of evolution in continuous traits (expected under Brownian motion) in whole clades or individual branches. We also include data of the geographic locations of two Clupeiform fossils, one for the crown group of Engraulidae and another for the crown group of Dorosoma (Supplementary Figure 2). They were included as branches linked to the nodes where the two fossil belongs. We assigned ~zero branch-length (0.000001) to each fossil. The aim of assigning zero branch-length to each fossil is to ensure that the Geo model will not modify the branch so that the estimated location of the node will be at the fossil location with high accuracy and precision. Some variation will be present in the inference given the data of the remainder species in clade. The fossil data we used are those whose phylogenetic position at phylogenetic nodes are well known in The Fish Tree of Life[41]. The use of well-known node-fossils allowed us a more secure placing of paleo coordinates given that our methodological approach place fossil data at phylogenetic nodes. The geographic locations were extracted from the original papers describing the fossils. Then we reconstructed the paleo coordinates for the two fossils using the function *reconstruct* in the chronosphere R-package[50]. We used the age of the nodes for each fossil and the PALEOMAP model for the paleo coordinate reconstruction. Finally, we used the paleo coordinates as input in the Geo model analyses.

We ran four MCMC chains for 250,000,000 iterations, sampling every 50,000 iterations, and discarding 200,000,000 as burn in. These procedures were conducted based on the Brownian motion (BM) model and the Variable Rates (VR) model (Supplementary Table 3). We checked for chain convergence using the Effective Sample Size (ESS) in Tracer v1.6, ensuring outputs with ESS > 200. The final sample includes 1,000 posterior locations for each phylogenetic node. We selected the model that fit the data better by means of Bayes factors (*BF*), using the marginal likelihoods estimated by stepping stone sampling. *BF* is calculated as the double of the difference between the log marginal likelihood of the complex model and the simple model. By convention, *BF* > 2 indicates positive support for the complex model, *BF* = 5–10 indicates strong support and

BF > 10 is considered very strong support[51]. We excluded the species *Denticeps clupeoides* from the Geo model analyses because its pathwise distance and speed of movement obtained from previous analyses were extreme outliers (Supplementary Figure 5), which can bias the inferences made from further regression analyses to evaluate the correlates of dispersal abilities.

Pathwise distances and speed of species movement. We first define species dispersal as the movement of the species, considering its entire geographic range. We additionally define speed of species movement as the distance a species moves in an interval of time – kilometres per million year (see Supplementary Figure 4). In order to obtain the total distance that each species have historically dispersed through the oceans and rivers – starting from the location of the root of Clupeoidei (Clupeiformes without *D. clupeoides*) phylogenetic tree - we calculated the distances dispersed across each phylogenetic branch (branchwise distances) and then we summed these distances along the path that links the root with extant species (pathwise distances; Supplementary Figure 4). The branchwise distances were calculated using the distCosine function in the geosphere R package[52]. The distCosine function brings the shortest distance between two points, assuming a spherical earth. The distance is calculated according to the law of the cosines, and the method works at both large and small scales[32]. We calculated the branchwise distances for every location in the posterior sample, meaning that we have 1,000 distances for every branch in the tree, and therefore, 1,000 pathwise distances for each species in the tree (Supplementary Table 7). With this approach we have the historical distance dispersed for each species, considering the uncertainty in ancestral locations estimates (Fig. 3 and Supplementary Figure 3). In order to have a measure of the speed at which each species in phylogeny have dispersed over historical time, we calculated the branchwise speed of movement in km per Myr - diving the branchwise distances by the branch length of the time-calibrated tree. We also calculate the speed of movement for all the *posterior* sample of branchwise distances, and then we calculated the median speed of movement in the path that links the MRCA with extant species. Finally, we have 1,000 measures of the historical speed of movement for each species (Supplementary Table 7), which include the uncertainty in ancestral location estimates (Fig. 3 and Supplementary Figure 3).

Phylogenetic regressions. To evaluate the expected relationships between SL, WTT, pathwise distance, pathwise speed of movement, and speciation rates, we performed Phylogenetic Generalized Least Squares regression models (PGLS) with Bayesian inference which allowed us to consider the uncertainty in both, parameters estimation and within species data. We consider the uncertainty within species by using the samples of data for WTT, georeferences, pathwise distances, and speed of movement. We also considered the uncertainty in ancestral states and branchwise rates of SL and WSL, and the Speed across phylogenetic branches. Under this approach, the MCMC samples the regression parameters and the sample data simultaneously, integrating the uncertainty of both factors in the results. All Bayesian regressions were done in the computer program BayesTraits 3.0.

First, we conducted a multiple phylogenetic regression to evaluate the relationship between SL, WTT and type of migration, including the sample of WTT within species. We compared the BM, Lambda model (LA), and Ornstein-Uhlenbeck model (OU) for these regressions. We also evaluated the variation in the SL evolution rate using the variable rates (VR) regression model[37], and model that integrate both the VR and LA model (VRLA). The VR regression model enable the simultaneous estimation of both an overall relationship between SL as a function of WTT and type of migration, and any shift in the rate that applies to the phylogenetically structured residual variance in the relationship. The VR regression model identifies heterogeneity in the rate of evolution along phylogenetic branches (branchwise rates) by dividing the rate into two parameters: a background rate parameter (σ^2_b), which assumes that changes in the trait of interest

are drawn from an underlying BM process, and a second parameter, *r*, which identifies a branch-specific rate shift. A full set of branchwise rates are estimated by adjusting the lengths of each branch in a time-calibrated tree (stretching or compressing a branch is equivalent to increasing or decreasing the phenotypic rate of change relative to the underlying Brownian rate of evolution). Branchwise rates are defined by a set of branch-specific scalars *r* ($0 < r < \infty$) that scale each branch to optimize the phenotypic rate of change to a BM process ($\sigma^2_b \times r$). If phenotypic change occurred at rates faster than the background rate, along a specific branch of the tree, then $r > 1$ and the branch is stretched. Rates slower than the background rate are detected by $r < 1$ and the branch is compressed. If the trait evolves at a constant rate along a branch, then the branch will not be modified (that is, $r = 1$). There is no limit or prior expectation in the number of the *r* branch scalars, *r* numbers vary from zero (no branch is scaled) to *n*, in which *n* is the number of branches in the phylogenetic tree. Regarding the values of each *r* parameter, we used a gamma prior, with $\alpha = 1.1$ and a β parameter that is rescaled such that the median of the distribution is equal to 1[37,49]. With this setting, the numbers of the rate increases and decreases that are proposed are balanced[49]. We ran four MCMC chains for 151,000,000 iterations, sampling every 50,000 iterations, and discarding 101,000,000 as burn-in. We checked for chain convergence using the ESS in Tracer v1.6, ensuring of using outputs with ESS > 200.

Second, in order to estimate the rates of WTT change through the Clupeiformes phylogeny, we conducted a Bayesian VRLA regression between WTT and latitude (comparing it with the BM, LA, OU, and VR regression models; Supplementary Table 2). We included the sample of WTT and latitude within each species in regression analyses. We ran four MCMC chains for 300,000,000 iterations, sampling every 250,000 iterations, and discarding 150,000,000 as burn-in. We checked for chain convergence using the ESS in Tracer v1.6, ensuring of using outputs with ESS > 200.

Third, we evaluated the relationship between the pathwise distance with SL and the type of migration, and between the pathwise speed with SL and type of migration. We included in the phylogenetic regressions the sample of species data for the pathwise distance and speed of movement, comparing regressions fitted with the BM, LA, OU, VR, and VRLA model (Supplementary Table 4 and 5). We ran MCMC chains with different number of iterations, sampling, and burn-in, in order to ensure of using outputs with ESS > 100. Regressions for pathwise distance had all ESS > 100 (Supplementary Table 8). Regressions for Speed had ESS > 100 for the BM and LA model, and ESS <100 for the OU, VR, and VRLA model (Supplementary Table 9). However, regressions for pathwise speed based on all models (including those with ESS <100) give the same result: SL had a positive effect on pathwise speed.

Fourth, we evaluated the relationship between speciation rates with pathwise speed, SL, pathwise distance, and WTT - including the sample of data for pathwise distances and speed of movement. We used tip-specific estimates of speciation rates to evaluate the regression between speciation rates and the multiple explanatory variables. Among the recommended non-model-based tip-rate metrics to study the correlates of speciation rates (i.e. inverse of equal splits [ES], node density [ND] and the inverse of terminal branch length [TB])[53] we based our interpretations on the node density along the phylogenetic paths, divided by the age of the phylogeny (111 Myr after excluding *D. clupeoides*). Our choice is based on the fact that ND is the least influenced metric by potential biases and sources of uncertainty associated with branch length estimation from empirical data[54] – ND capture the average speciation rate over the entire phylogenetic path and weight equally all branch lengths along the paths. We did not use the tip-rate speciation metric estimated from time-varying birth-death diversification models owing to the striking uncertainty in the speciation rates values when they are estimated from phylogenies with extant species only[55], and due to the

erroneous inference of the general diversification patterns when the variation in rates of sequence evolution are not properly considered in time-tree inference[56].

Additionally, we used PGLS regression models to evaluate regression-coefficients-significance because PGLS-ND has the highest statistical power when compared with PGLS-ES and PGLS-TB[53]. Furthermore, PGLS allow us to evaluate the simultaneous effect of multiple explanatory variables whose effect on speciation rates can be modelled as a linear or non-linear function. This last point is of utmost importance for our objective because there are expected interactions between the main explanatory variables (e.g. pathwise speed and SL, WTT and SL) and also because there are statistical complications associated with estimating interactions without including quadratic terms (i.e. non-linear functions between the independent and explanatory variables)[57]. Our full PGLS-ND regression model is described by the following equation: ND ~ Speed + SL + Distance + WTT + $Speed^2$ + SL^2 + $Distance^2$ + WTT^2 + (Speed * SL) + (Distance * SL) + (WTT * SL). Then, we reached the simpler reduced PGLS-ND regression model based on strict criteria: we removed the single most non-significant regression-coefficient from the full regression model, then we reiterated this procedure across every simpler regression until we get the regression with significant covariates only. We conducted these regression analyses comparing the BM and LA model. The final regression is in Supplementary Table 6. We ran 51,000,000 iterations, sampling every 50,000 iterations, and discarding the first 10,000,000 iterations as burn in. Regression coefficients were judged to be significant according to a calculated P_{MCMC} value for each posterior of regression coefficients. For cases in which <5% of samples in the posterior distribution crossed zero, this indicates that the coefficient is significantly different from zero.

Nonphylogenetic regressions. We applied Bayesian GLS regressions to evaluate the relationship between the branchwise rates of SL evolution, the branchwise speed of movement and the branchwise rates of WTT change. We obtained these branchwise rates and speed of movement using the rate-scaled branches as dividend and the original branch lengths (measured in time) as the divisor. Specifically, we divided the branches from the LA-scaled posterior sample of trees for SL, the VRLA-scaled posterior sample of trees for WTT, and the VR-scaled posterior sample of trees for geographic occurrences. We used 1,000 scaled trees so that we had 1,000 observations of rates per phylogenetic branch. The use of the posterior sample of rate-scaled branches allows us to include the uncertainty of rates estimation in regression analyses (Fig. 4; Supplementary Table 9).

Additionally, we regressed the *posterior* sample of ancestral SL on the *posterior* sample of ancestral WTT, inferred at each node of the Clupeiformes phylogeny. Ancestral states were inferred with the fastAnc function of the phytools R-package[58], which assumes a constant-rate Brownian motion model for the evolution of continuous traits. We used the *posterior* sample of scaled trees, obtained from the model outputs that fit the data better, i.e., LA for SL and the VRLA model for WTT. The use of rate scaled trees allow us to include the variation in the rate of evolution when estimating the ancestral states at each phylogenetic node. We used a sample of 1,000 scaled trees, which also allow us to include the uncertainty of ancestral states estimation in regression analyses (Fig. 2d; Supplementary Table 10). We validated the WTT inferred at nodes with phylogenetic models by comparing them to model-based temperature reconstructions. We randomly selected eight nodes (plus the MRCA of Clupeoidei) and matched the median temperature estimated from the phylogenetic approach with the environmental temperatures reconstructed from the output of the HadCM3L Earth-System-Model encompassing from the Jurassic to the Eocene[59]. We matched the age of each node with the respective geologic stage. We extracted both air (mainland) temperature and sea surface temperature, available at a 3.75 x 2.5º longitude-latitude resolution, based on the 95% of the posterior density of coordinates at each

node. It is important to note that the HadCM3L model is based on the Getech model as boundary conditions, while we used a different model (the PALEOMAP model) to reconstruct the paleo coordinates of the two fossils. As such, the PALEOMAP reconstruction for the fossils will be different to what their position in the Getech palaeogeography would be. However, as the difference between Getech and PALEOMAP from the Cretaceous onwards is small and given that we used a 3.75x2.5 resolution, the points will probably still fall in the same grid cell under either reconstruction. On the other hand, the difference will also be negligible as the vast majority of the *posterior* density of coordinates came from the Geo model.

The results show that the phylogenetic estimation of temperatures is positively correlated with both the estimation of air temperature ($t = 5.4$, $p = 0.0009$) and the estimation of sea temperature ($t = 3.17$, $p = 0.01$; Supplementary Figure 6) which are based on the HadCM3L Earth-System-Model. The phylogenetic estimations fall also within the range of air temperature or sea temperature, depending on whether the phylogenetic node was more likely a freshwater or marine species. There were three phylogenetic nodes (node 4, node 6, and node 7; Supplementary Figure 3), with the highest *posterior* distribution for marine habitat, in which the phylogenetic estimations fall outside the range of sea temperature. However, these three nodes are reconstructed with high precision around islands (Supplementary Figure 3) which suggest that those species occupied the inland waters. When considering the range of air temperature, the phylogenetic estimations of these three nodes fall withing the range of the HadCM3L Earth-System-Model.

Finally, we conducted the Bayesian nonphylogenetic GLS regressions in BayesTraits by setting the Pagel's Lambda parameter to zero, which discard the phylogenetic covariance of the data values. To sample from the posterior distribution of rates per phylogenetic branch, and from the posterior distribution of ancestral states at each phylogenetic node, we ran Bayesian regressions that sample within tips data. We ran 51,000,000 iterations, sampling every 50,000 iterations, and discarding the first 1,000,000 iterations as burn in. All chains had ESS > 200. Regression coefficients were judged to be significant according to a calculated P_{MCMC} value for each posterior of regression coefficients. For cases in which <5% of samples in the posterior distribution crossed zero, this indicates that the coefficient is significantly different from zero. We used a uniform prior for regression coefficients (slopes) as we do not know what the relationship between the response and predictor variable is. The prior ranged from −100 to 100, to ensure that all possible slope values are sampled.

Code availability

All analyses in this study were done using BayesTraits version 3 available at http://www.evolution.rdg.ac.uk/BayesTraitsV3/ BayesTraitsV3.html

Competing interests

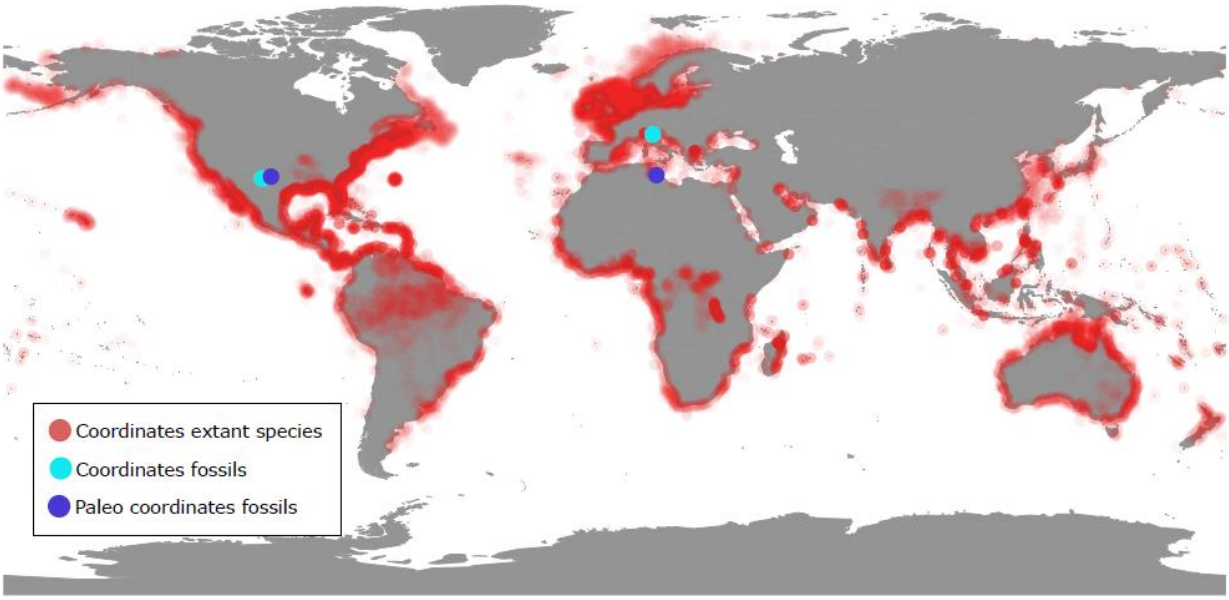
The authors declare no competing interests.

Data availability statement

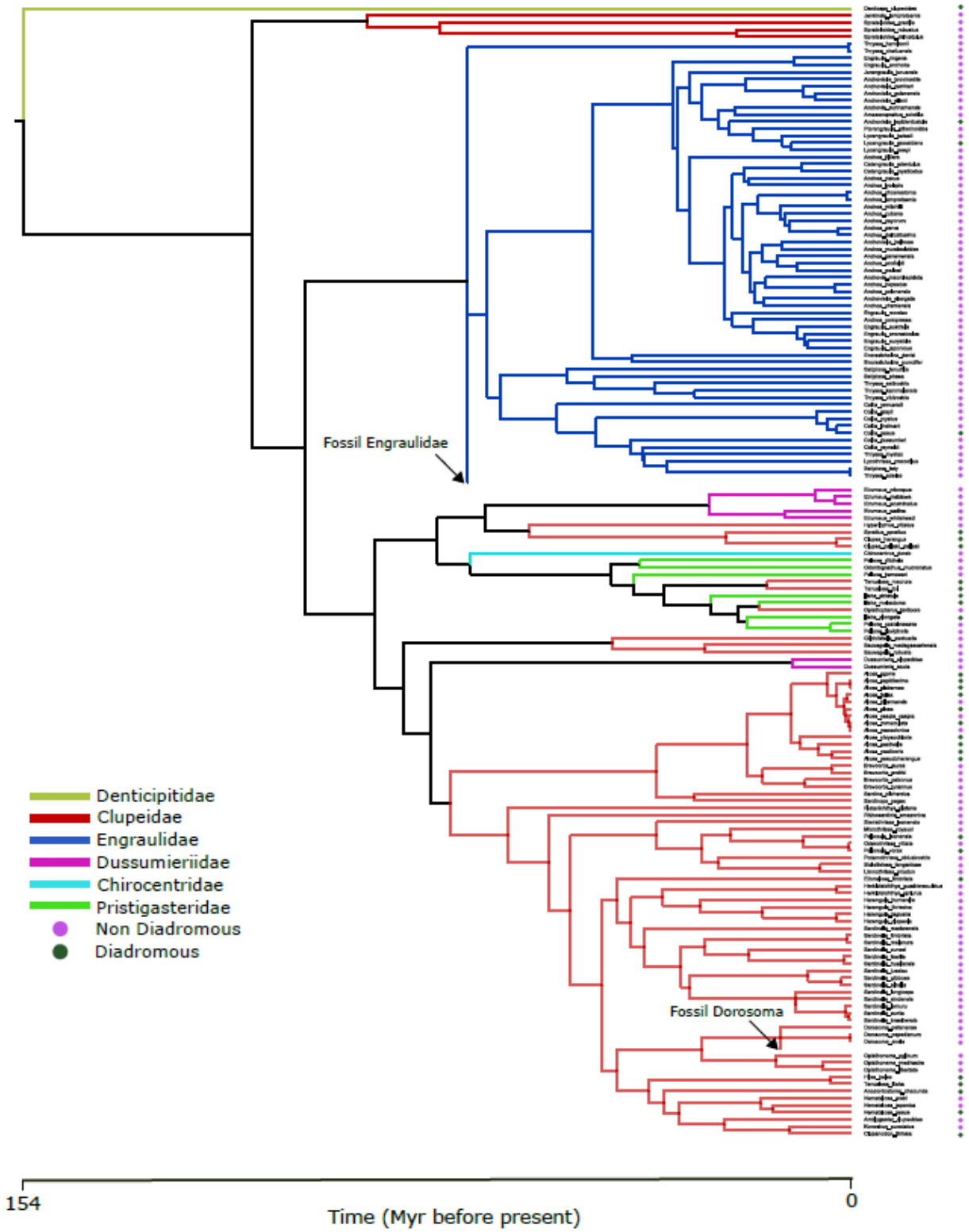
No new data were generated for this study. The data used for this paper are available from the original sources cited in the Methods and Supplementary Information.

Correspondence and request for materials should be addressed to the corresponding author.

Supplementary Figures

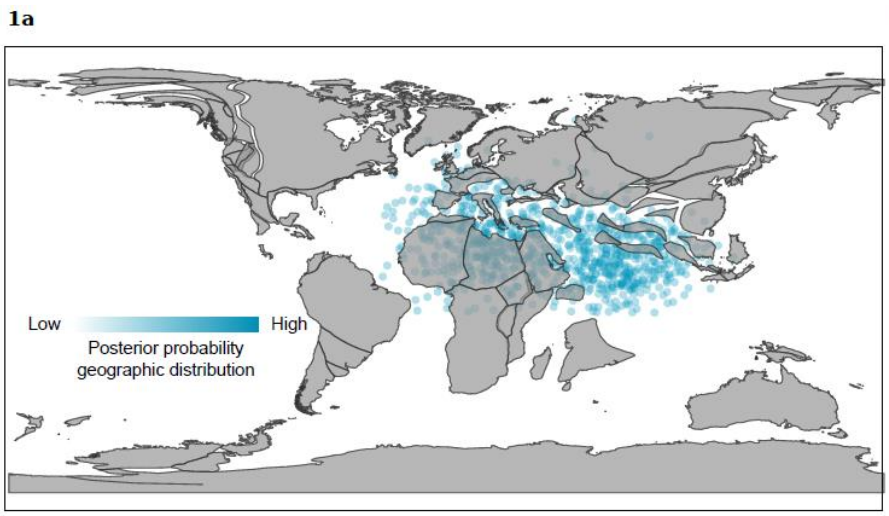


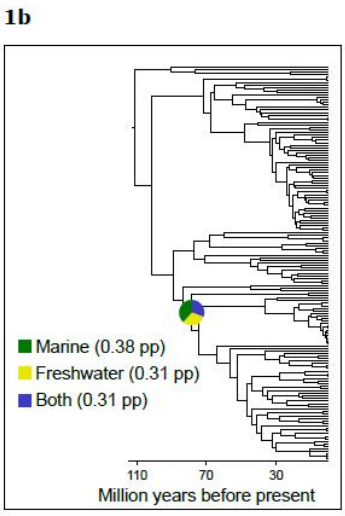
Supplementary Figure 1. Geographic distribution of Clupeiformes species used in this study. Red dots represent the geographic occurrences obtained from Aquamaps and the random sample within IUCN polygons, which comprises 21,795 datapoints for 158 species. The paleo coordinates for the fossils of Dorosoma (America) and Engraulidae (Europe) were estimated using the PALEOMAP model in the chronosphere R package. The coordinates of extant species plus the two paleo coordinates were used as input data to reconstruct ancestral locations across phylogenetic nodes.



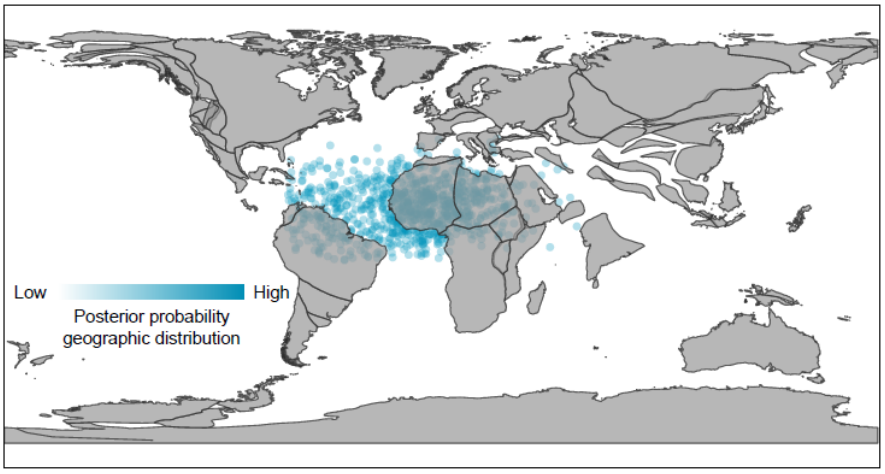
Supplementary Figure 2. Clupeiformes phylogenetic tree used in this study. The phylogenetic tree was obtained from the Fish Tree of Life and represent the most updated topology and divergence times of the group. Note that branch colours represent the taxonomic arrangement of the group and are used for reference only. Fossils added, and type of migration are indicated. For the Geo model analyses we excluded *Denticeps clupeoides* (Methods). Nevertheless, we included *D. cupleoides* in all other analyses.

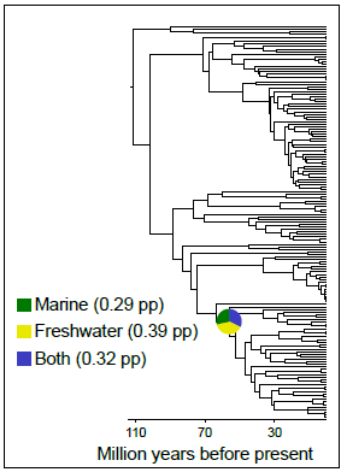
726


1a
Low
High
Posterior probability
geographic distribution

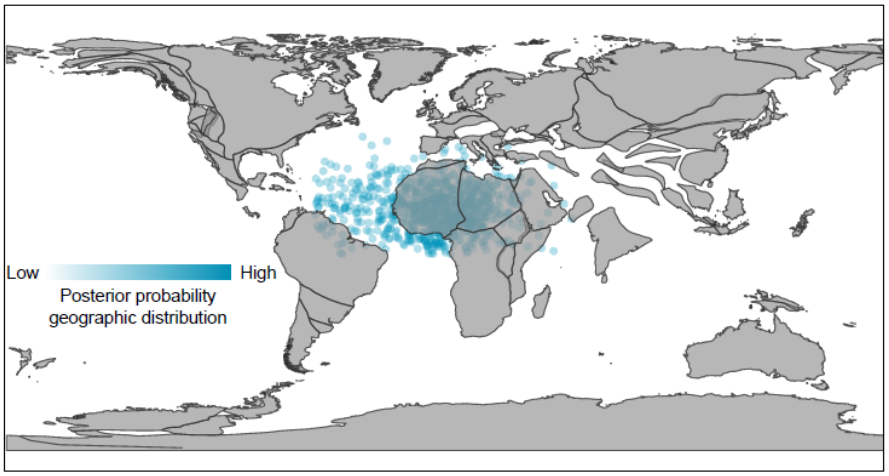

1b
Marine (0.38 pp)
Freshwater (0.31 pp)
Both (0.31 pp)
110
70
30
Million years before present

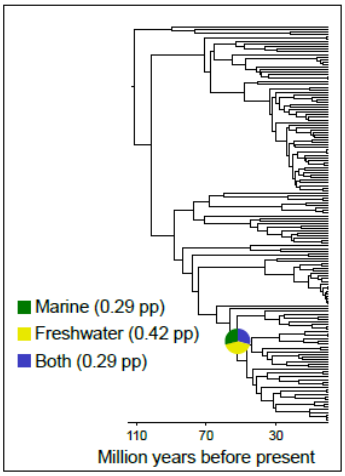
727


2a
Low
High
Posterior probability
geographic distribution


2b
Marine (0.29 pp)
Freshwater (0.39 pp)
Both (0.32 pp)
110
70
30
Million years before present

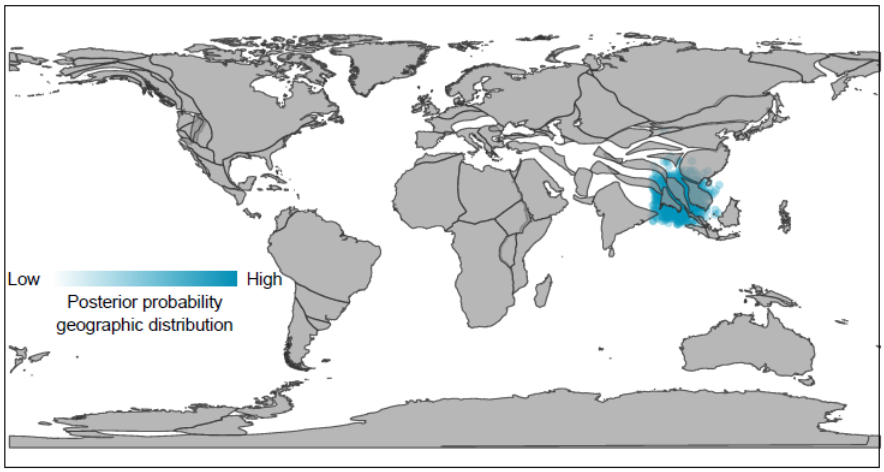
728


3a
Low
High
Posterior probability
geographic distribution

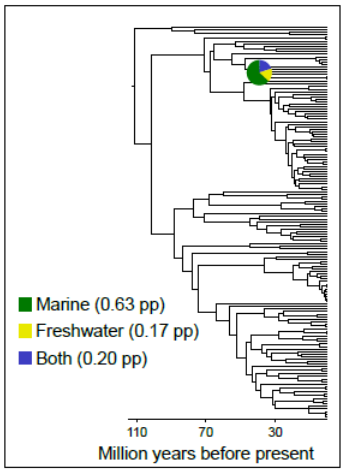

3b
Marine (0.29 pp)
Freshwater (0.42 pp)
Both (0.29 pp)
110
70
30
Million years before present

729

4a

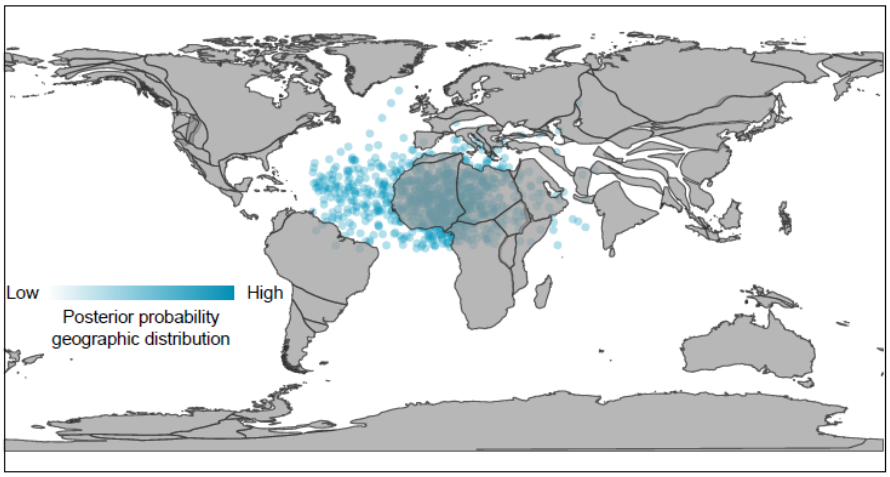


4b

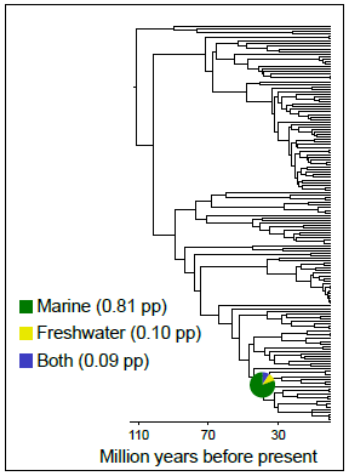


730

5a

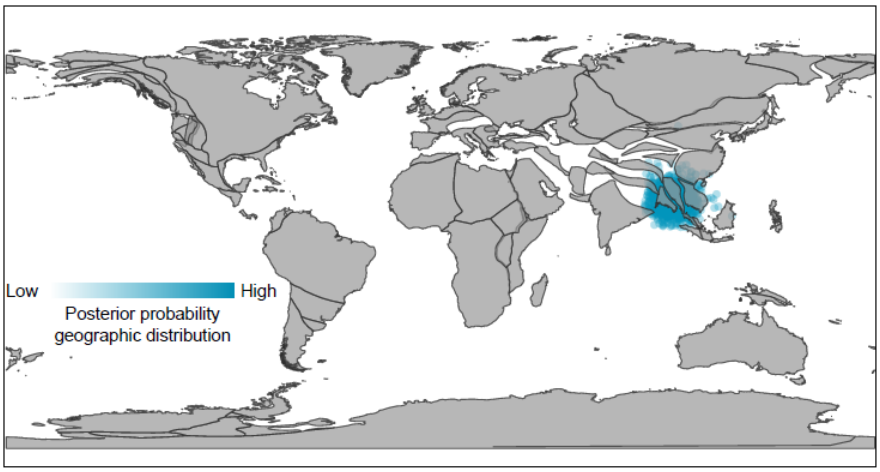


5b

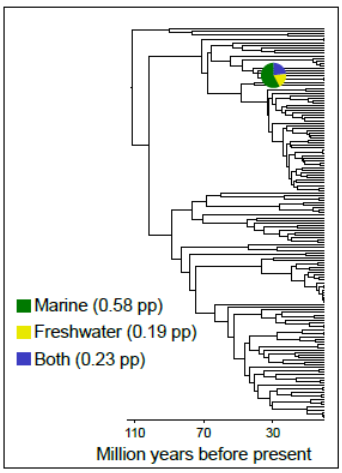


731

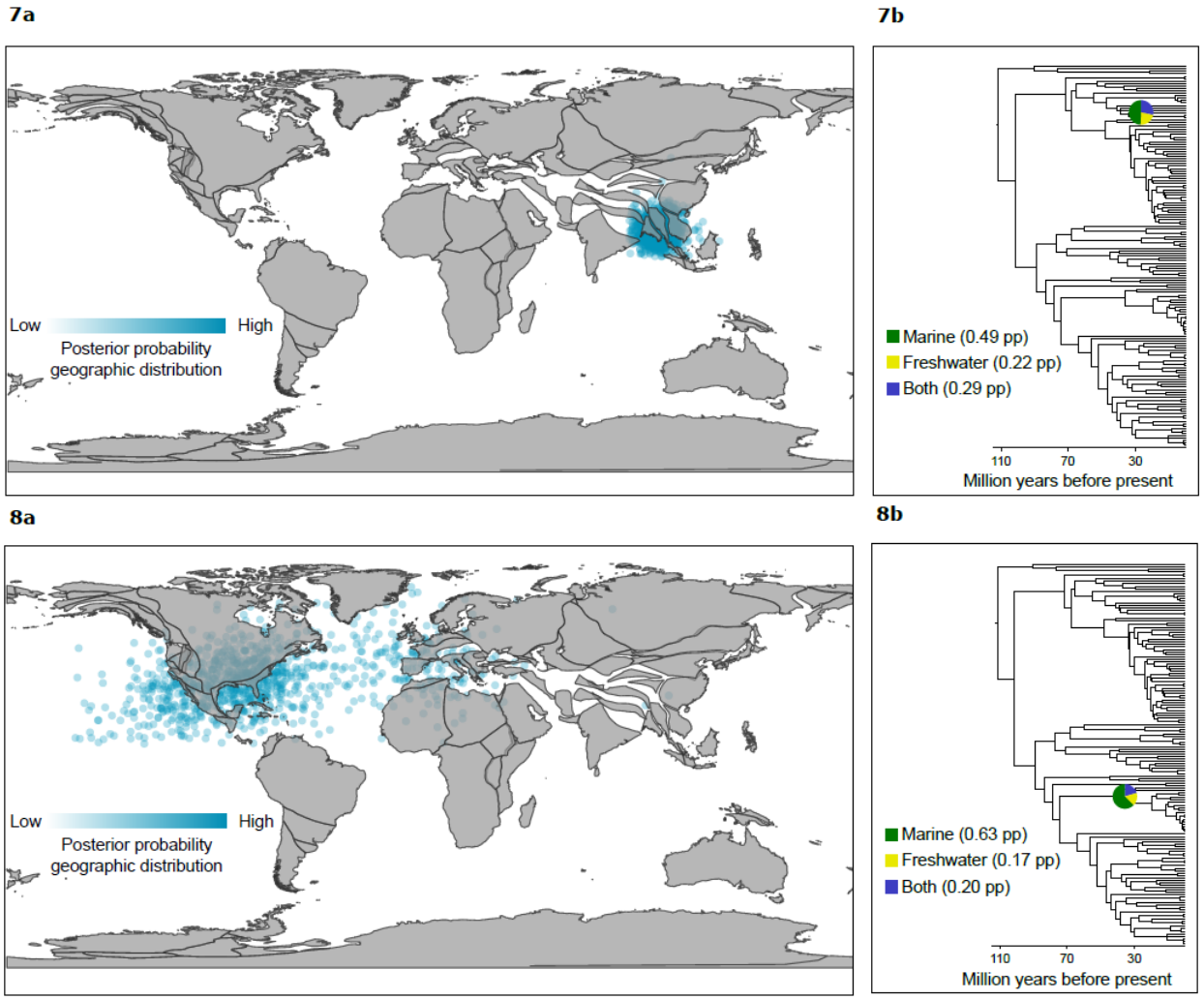
6a



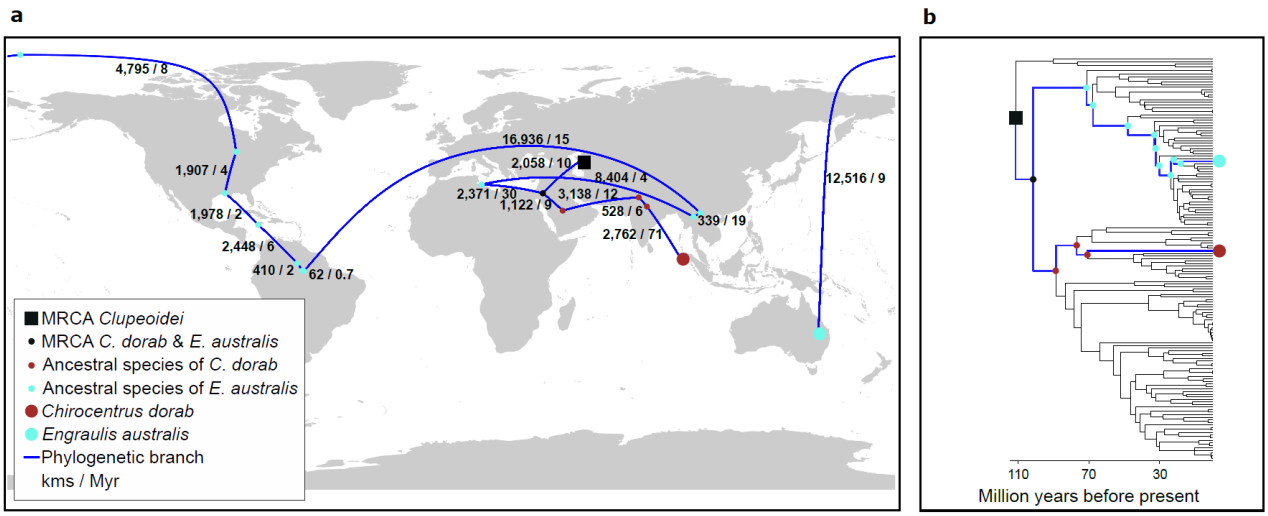
6b



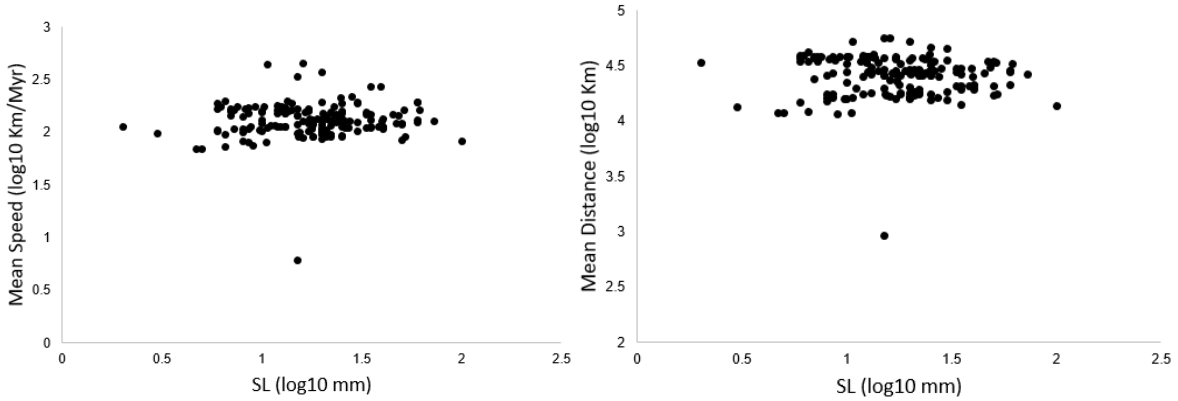
732



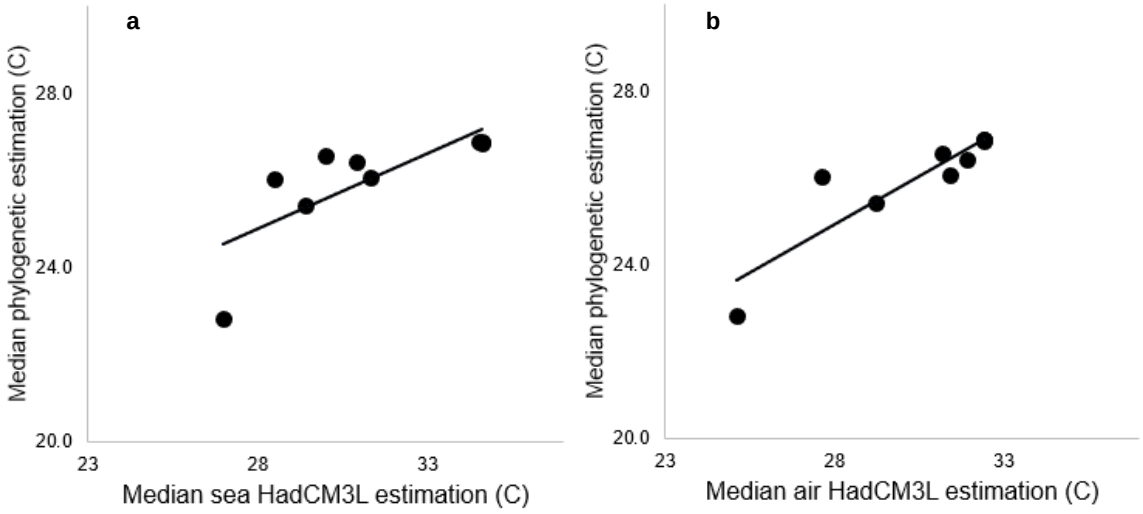
Supplementary Figure 3. *Posterior* geographic distribution and *posterior* probability of habitat type for eight phylogenetic nodes. We selected eight random nodes ranging from 111 to 33 Mya. 1 = 78 Mya; 2 = 56 Mya; 3 = 52 Mya; 4 = 41 Mya; 5 = 40 Mya; 6 = 38.8 Mya; 7 = 38.4 Mya; 8 = 36 Mya. **a.** The *posterior* coordinates were estimated with Geo model. **b.** The ancestral habitat type for these eight random nodes was estimated using phylogenetic models for discrete trait evolution.



Supplementary Figure 4. Two continuous geographic routes for the lines of descend leading to *Chirocentrus dorab* and *Engraulis australis*. The Geo model estimate the *posterior* probability of ancestral species locations (phylogenetic nodes) from geo-referenced occurrences of individuals within extinct and extant species. Ancestral locations are estimated while allowing the speed of species movement to vary across phylogenetic branches. The circles and squares are the geographic centroid estimated from the posterior distribution of coordinates (phylogenetic nodes) and the sample of coordinates from extant species. Note that the geographic centroids are used to obtain an example of the average route travelled for each species. However, we used 1,000 values of total distance and speed (using the full posterior distribution of estimated locations) for each species in all the analyses of this study. Note also that the map represents the actual location of continents – which is included as reference only.



Supplementary Figure 5. The estimated location for *Denticeps clupeiodes* made their speed and distance of movement to be an outlier in regression analyses. We removed *D. clupeoides* from the Geo analyses because that species descends directly from the MRCA of Clupeiformes and its location is estimated near to the location of the MRCA. This means that species has dispersed a short distance in an exceptionally long time period of 150 million years. This causes the speed and distance of movement for that species to be extremely low and far away from the rest of data when evaluating the correlates of speed and distance. We plot the mean speed and distance for all species.



Supplementary Figure 6. Comparison between median temperatures inferred independently from the phylogenetic approach and the HadCM3L Earth-System-Model. We selected eight random nodes plus the MRCA of Clupeoidei for comparison. Line equation in **a**: y = 15 + 0.35(x). Line equation in **b**: y = 12 + 0.45(x).

Supplementary Tables

Table 1. Evolutionary model fitting for the regression that evaluate the effect of type of migration and water surface temperature (WTT) on fish standard length (SL). Data analysed includes the maximum SL and samples of WTT, within the native range, for each species. The log Marginal Likelihood (Marginal Lh) estimated by stepping stone sampling, provides the models support given the data and priors. More positive values support a given model, where differences >1 indicates positive evidence; differences between 2.5 - 5 indicates strong support; and differences > 5 indicates very strong support for a model over the other. BM = Brownian Motion, LA = Lambda, OU = Ornstein-Uhlenbeck, VR = Variable Rate, VRLA = Variable Rate and Lambda.

SL Phylogenetic Regression Model	Marginal Lh. BM	Marginal Lh. LA	Marginal Lh. OU	Marginal Lh. VR	Marginal Lh. VRLA
$SL \sim \alpha + \beta_1(Diadromous) + \beta_2(WTT)$	-59.11	**8.09**	-19.84	-16.29	8.13

Table 2. Evolutionary model fitting for the regression that evaluates the effect of absolute latitude on WTT. Data analysed includes a sample of WTT and absolute latitude (AbsLat) within the native range of each species. The log Marginal Likelihood (Marginal Lh) estimated by stepping stone sampling, provides the models support given the data and priors. More positive values support a given model, where differences >1 indicates positive evidence; differences between 2.5 - 5 indicates strong support; and differences > 5 indicates very strong support for a model over the other. BM = Brownian Motion, LA = Lambda, OU = Ornstein-Uhlenbeck, VR = Variable Rate, VRLA = Variable Rate and Lambda.

WTT Phylogenetic Regression Model	Marginal Lh. BM	Marginal Lh. LA	Marginal Lh. OU	Marginal Lh. VR	Marginal Lh. VRLA
$WTT \sim \alpha + \beta_1(AbsLat) + \beta_2(AbsLat)^2$	-421.8	-338.9	-340.1	-318.2	**-304.3**

Table 3. Geographical model (Geo model) fitting for Clupeiformes georeferenced data. The Geo model estimate the longitudes and latitudes across the nodes of the phylogenetic tree by means of Bayesian inference. These coordinates are estimated onto a three-dimensional cartesian coordinates system which were modelled using Brownian motion (BM) – the rate of location change across the tree is constant. We also allowed the rate of location-change to vary across phylogenetic branches by fitting the Variable Rate model (VR). The log Marginal Likelihood (Marginal Lh) estimated by stepping stone sampling, provides the models support given the data and priors. More positive values support a given model, where differences >1 indicates positive evidence (Bayes Factor > 2); differences between 2.5 - 5 indicates strong support (Bayes Factor 5 – 10); and differences > 5 indicates very strong support for a model over the other (Bayes Factor > 10).

Chain	Marginal Lh. Geographical model BM	Marginal Lh. Geographical model VR	Bayes Factor BM vs VR
1	-8545.36	-8008.62	1073.48
2	-8546.16	-8008.82	1074.68
3	-8546.74	-8002.93	1087.62
4	-8545.16	-8005.31	1079.70

Table 4. Evolutionary model fitting for the regression that evaluate the effect of SL and type of migration on the speed of fish movement. The log Marginal Likelihood (Marginal Lh) estimated by stepping stone sampling, provides the models support given the data and priors. More positive values support a given model, where differences >1 indicates positive evidence (Bayes Factor > 2); differences between 2.5 - 5 indicates strong support (Bayes Factor 5 – 10); and differences > 5 indicates very strong support for a model over the other (Bayes Factor > 10). BM = Brownian Motion, LA = Lambda, OU = Ornstein-Uhlenbeck, VR = Variable Rate, VRLA = Variable Rate and Lambda.

	Marginal Lh. BM	Marginal Lh. LA	Marginal Lh. OU	Marginal Lh. VR	Marginal Lh. VRLA
Distance ~ $\alpha + \beta_1(SL)$	137.83	131.89	133,33	**145.63**	142.58
Distance ~ $\alpha + \beta_1(SL) + \beta_2(Diadromous)$	128.75	126.14	124.12	140.50	134.74
Distance	121.68	115.44	114.99	128.99	127.81

Table 5. Evolutionary model fitting for the regression that evaluate the effect of SL and type of migration on the distance of fish movement. The log Marginal Likelihood (Marginal Lh) estimated by stepping stone sampling, provides the models support given the data and priors. More positive values support a given model, where differences >1 indicates positive evidence (Bayes Factor > 2); differences between 2.5 - 5 indicates strong support (Bayes Factor 5 – 10); and differences > 5 indicates very strong support for a model over the other (Bayes Factor > 10). BM = Brownian Motion, LA = Lambda, OU = Ornstein-Uhlenbeck, VR = Variable Rate, VRLA = Variable Rate and Lambda.

	Marginal Lh. BM	Marginal Lh. LA	Marginal Lh. OU	Marginal Lh. VR	Marginal Lh. VRLA
Speed ~ $\alpha + \beta_1(SL)$	106.41	102.70	99.71	**135.16**	109.77
Speed ~ $\alpha + \beta_1(SL) + \beta_2(Diadromous)$	97.90	99.33	89.61	113.02	101.47
Speed	93.24	85.39	89.13	94.13	92.18

Table 6. Phylogenetic regression model for Node Density (ND) obtained after reducing the full model ND ~ Speed + SL + Distance + WTT + Speed^2 + SL^2 + Distance^2 + WTT^2 + (Speed * SL) + (Distance * SL) + (WTT * SL). The log Marginal Likelihood (Marginal Lh) estimated by stepping stone sampling, provides the models support given the data and priors. More positive values support a given model, where differences >1 indicates positive evidence (Bayes Factor > 2); differences between 2.5 - 5 indicates strong support (Bayes Factor 5 – 10); and differences > 5 indicates very strong support for a model over the other (Bayes Factor > 10).

	Marginal Lh. BM	Marginal Lh. LA
ND ~ $\alpha + \beta_1(Speed) + \beta_2(Distance)$	**481.54**	473.75

Table 7. Data sample size (n) for each species used in this study. SL: Maximum standard length, Lon-Lat: Longitude and latitude, WTT: Water temperature tolerance. Note that there is 1 standard length for every species as we used one data point – the maximum standard length. The number of coordinates and temperature are the same as temperature data was obtained for every coordinate.

Species	SL	Migration type	Lon-Lat	WTT	Pathwise Distance	Pathwise Speed	Node Density
Alosa aestivalis	1	1	219	219	1,000	1,000	1
Alosa agone	1	1	100	100	1,000	1,000	1
Alosa alabamae	1	1	100	100	1,000	1,000	1
Alosa alosa	1	1	243	243	1,000	1,000	1
Alosa caspia caspia	1	1	6	6	1,000	1,000	1
Alosa chrysochloris	1	1	42	42	1,000	1,000	1
Alosa fallax	1	1	466	466	1,000	1,000	1
Alosa immaculata	1	1	11	11	1,000	1,000	1
Alosa killarnensis	1	1	1	1	1,000	1,000	1
Alosa macedonica	1	1	100	100	1,000	1,000	1
Alosa mediocris	1	1	77	77	1,000	1,000	1
Alosa pseudoharengus	1	1	344	344	1,000	1,000	1
Alosa sapidissima	1	1	304	304	1,000	1,000	1
Amazonsprattus scintilla	1	1	100	100	1,000	1,000	1
Amblygaster clupeoides	1	1	26	26	1,000	1,000	1
Anchoa cayorum	1	1	69	69	1,000	1,000	1
Anchoa chamensis	1	1	4	4	1,000	1,000	1
Anchoa choerostoma	1	1	100	100	1,000	1,000	1
Anchoa colonensis	1	1	45	45	1,000	1,000	1
Anchoa compressa	1	1	18	18	1,000	1,000	1
Anchoa cubana	1	1	100	100	1,000	1,000	1
Anchoa delicatissima	1	1	8	8	1,000	1,000	1
Anchoa filifera	1	1	47	47	1,000	1,000	1
Anchoa hepsetus	1	1	299	299	1,000	1,000	1
Anchoa lamprotaenia	1	1	80	80	1,000	1,000	1
Anchoa lyolepis	1	1	186	186	1,000	1,000	1
Anchoa mitchilli	1	1	206	206	1,000	1,000	1
Anchoa mundeoloides	1	1	100	100	1,000	1,000	1
Anchoa nasus	1	1	97	97	1,000	1,000	1
Anchoa panamensis	1	1	5	5	1,000	1,000	1
Anchoa parva	1	1	38	38	1,000	1,000	1
Anchoa scofieldi	1	1	8	8	1,000	1,000	1
Anchoa walkeri	1	1	43	43	1,000	1,000	1
Anchovia macrolepidota	1	1	90	90	1,000	1,000	1
Anchovia surinamensis	1	1	100	100	1,000	1,000	1
Anchoviella alleni	1	1	100	100	1,000	1,000	1

Anchoviella balboae	1	1	100	100	1,000	1,000	1
Anchoviella brevirostris	1	1	15	15	1,000	1,000	1
Anchoviella carrikeri	1	1	100	100	1,000	1,000	1
Anchoviella elongata	1	1	100	100	1,000	1,000	1
Anchoviella guianensis	1	1	100	100	1,000	1,000	1
Anchoviella lepidentostole	1	1	44	44	1,000	1,000	1
Anodontostoma chacunda	1	1	163	163	1,000	1,000	1
Brevoortia aurea	1	1	20	20	1,000	1,000	1
Brevoortia patronus	1	1	66	66	1,000	1,000	1
Brevoortia smithi	1	1	49	49	1,000	1,000	1
Brevoortia tyrannus	1	1	153	153	1,000	1,000	1
Cetengraulis edentulus	1	1	115	115	1,000	1,000	1
Cetengraulis mysticetus	1	1	81	81	1,000	1,000	1
Chirocentrus dorab	1	1	294	294	1,000	1,000	1
Clupanodon thrissa	1	1	13	13	1,000	1,000	1
Clupea harengus	1	1	2,140	2,140	1,000	1,000	1
Clupea pallasii pallasii	1	1	647	647	1,000	1,000	1
Coilia dussumieri	1	1	37	37	1,000	1,000	1
Coilia grayii	1	1	13	13	1,000	1,000	1
Coilia lindmani	1	1	100	100	1,000	1,000	1
Coilia mystus	1	1	100	100	1,000	1,000	1
Coilia nasus	1	1	20	20	1,000	1,000	1
Coilia ramcarati	1	1	100	100	1,000	1,000	1
Coilia reynaldi	1	1	7	7	1,000	1,000	1
Denticeps clupeoides	1	1	100	100	-	-	1
Dorosoma anale	1	1	100	100	1,000	1,000	1
Dorosoma cepedianum	1	1	121	121	1,000	1,000	1
Dorosoma petenense	1	1	96	96	1,000	1,000	1
Dussumieria acuta	1	1	65	65	1,000	1,000	1
Dussumieria elopsoides	1	1	271	271	1,000	1,000	1
Encrasicholina devisi	1	1	60	60	1,000	1,000	1
Encrasicholina punctifer	1	1	120	120	1,000	1,000	1
Engraulis anchoita	1	1	93	93	1,000	1,000	1
Engraulis australis	1	1	251	251	1,000	1,000	1
Engraulis encrasicolus	1	1	863	863	1,000	1,000	1
Engraulis eurystole	1	1	187	187	1,000	1,000	1
Engraulis japonicus	1	1	125	125	1,000	1,000	1
Engraulis mordax	1	1	292	292	1,000	1,000	1
Engraulis ringens	1	1	63	63	1,000	1,000	1
Ethmalosa fimbriata	1	1	105	105	1,000	1,000	1
Etrumeus acuminatus	1	1	100	100	1,000	1,000	1
Etrumeus makiawa	1	1	100	100	1,000	1,000	1
Etrumeus micropus	1	1	100	100	1,000	1,000	1

Etrumeus sadina	1	1	258	258	1,000	1,000	1
Etrumeus whiteheadi	1	1	100	100	1,000	1,000	1
Gilchristella aestuaria	1	1	14	14	1,000	1,000	1
Harengula clupeola	1	1	129	129	1,000	1,000	1
Harengula humeralis	1	1	124	124	1,000	1,000	1
Harengula jaguana	1	1	296	296	1,000	1,000	1
Harengula thrissina	1	1	105	105	1,000	1,000	1
Herklotsichthys quadrimaculatus	1	1	207	207	1,000	1,000	1
Herklotsichthys spilurus	1	1	9	9	1,000	1,000	1
Hilsa kelee	1	1	59	59	1,000	1,000	1
Hyperlophus vittatus	1	1	88	88	1,000	1,000	1
Ilisha elongata	1	1	40	40	1,000	1,000	1
Ilisha melastoma	1	1	51	51	1,000	1,000	1
Ilisha striatula	1	1	4	4	1,000	1,000	1
Jenkinsia lamprotaenia	1	1	118	118	1,000	1,000	1
Jurengraulis juruensis	1	1	100	100	1,000	1,000	1
Konosirus punctatus	1	1	42	42	1,000	1,000	1
Limnothrissa miodon	1	1	100	100	1,000	1,000	1
Lycengraulis batesii	1	1	100	100	1,000	1,000	1
Lycengraulis grossidens	1	1	89	89	1,000	1,000	1
Lycengraulis poeyi	1	1	20	20	1,000	1,000	1
Lycothrissa crocodilus	1	1	100	100	1,000	1,000	1
Microthrissa royauxi	1	1	100	100	1,000	1,000	1
Nematalosa erebi	1	1	100	100	1,000	1,000	1
Nematalosa japonica	1	1	14	14	1,000	1,000	1
Nematalosa nasus	1	1	69	69	1,000	1,000	1
Odaxothrissa vittata	1	1	100	100	1,000	1,000	1
Odontognathus mucronatus	1	1	41	41	1,000	1,000	1
Opisthonema libertate	1	1	125	125	1,000	1,000	1
Opisthonema medirastre	1	1	59	59	1,000	1,000	1
Opisthonema oglinum	1	1	374	374	1,000	1,000	1
Opisthopterus tardoore	1	1	32	32	1,000	1,000	1
Pellona castelnaeana	1	1	100	100	1,000	1,000	1
Pellona ditchela	1	1	290	290	1,000	1,000	1
Pellona flavipinnis	1	1	100	100	1,000	1,000	1
Pellona harroweri	1	1	66	66	1,000	1,000	1
Pellonula leonensis	1	1	38	38	1,000	1,000	1
Pellonula vorax	1	1	100	100	1,000	1,000	1
Platanichthys platana	1	1	100	100	1,000	1,000	1
Potamothrissa obtusirostris	1	1	100	100	1,000	1,000	1
Pterengraulis atherinoides	1	1	100	100	1,000	1,000	1
Rhinosardinia amazonica	1	1	100	100	1,000	1,000	1
Sardina pilchardus	1	1	714	714	1,000	1,000	1

Sardinella albella	1	1	190	190	1,000	1,000	1
Sardinella aurita	1	1	674	674	1,000	1,000	1
Sardinella brasiliensis	1	1	75	75	1,000	1,000	1
Sardinella fimbriata	1	1	56	56	1,000	1,000	1
Sardinella gibbosa	1	1	226	226	1,000	1,000	1
Sardinella hualiensis	1	1	11	11	1,000	1,000	1
Sardinella jussieu	1	1	7	7	1,000	1,000	1
Sardinella lemuru	1	1	61	61	1,000	1,000	1
Sardinella longiceps	1	1	38	38	1,000	1,000	1
Sardinella maderensis	1	1	168	168	1,000	1,000	1
Sardinella melanura	1	1	76	76	1,000	1,000	1
Sardinella sindensis	1	1	24	24	1,000	1,000	1
Sardinella tawilis	1	1	100	100	1,000	1,000	1
Sardinella zunasi	1	1	33	33	1,000	1,000	1
Sardinops sagax	1	1	745	745	1,000	1,000	1
Sauvagella madagascariensis	1	1	100	100	1,000	1,000	1
Sauvagella robusta	1	1	100	100	1,000	1,000	1
Setipinna phasa	1	1	100	100	1,000	1,000	1
Setipinna taty	1	1	47	47	1,000	1,000	1
Setipinna tenuifilis	1	1	86	86	1,000	1,000	1
Sierrathrissa leonensis	1	1	100	100	1,000	1,000	1
Spratelloides delicatulus	1	1	311	311	1,000	1,000	1
Spratelloides gracilis	1	1	169	169	1,000	1,000	1
Spratelloides robustus	1	1	69	69	1,000	1,000	1
Sprattus sprattus	1	1	825	825	1,000	1,000	1
Stolothrissa tanganicae	1	1	100	100	1,000	1,000	1
Tenualosa ilisha	1	1	47	47	1,000	1,000	1
Tenualosa macrura	1	1	7	7	1,000	1,000	1
Tenualosa toli	1	1	19	19	1,000	1,000	1
Thryssa adelae	1	1	3	3	1,000	1,000	1
Thryssa chefuensis	1	1	100	100	1,000	1,000	1
Thryssa hamiltonii	1	1	166	166	1,000	1,000	1
Thryssa kammalensis	1	1	8	8	1,000	1,000	1
Thryssa mystax	1	1	33	33	1,000	1,000	1
Thryssa setirostris	1	1	211	211	1,000	1,000	1
Thryssa vitrirostris	1	1	65	65	1,000	1,000	1

Table 8. Chain settings for regression analysis of pathwise distance. Number of iterations / burn-in / sampling frequency / and effective sample size (ESS). BM = Brownian Motion, LA = Lambda, OU = Ornstein-Uhlenbeck, VR = Variable Rate, VRLA = Variable Rate and Lambda. SL = Maximum standard length.

	BM	LA	OU	VR	VRLA
Distance ~ α + β_1(Diadromous) + β_2(SL)	150E6/50E6/ 10E4/176	150E6/50E6/ 10E4/424	300E6/200E6/ 10E4/414	600E6/500E6/ 10E4/150	150E6/50E6/ 10E4/229
Distance ~ α + β_1(SL)	200E6/100E6/ 10E4/332	51E6/1E6/ 5E4/537	51E6/1E6/ 5E4/271	600E6/500E6/ 10E4/129	200E6/100E6/ 10E4/296
Distance	51E6/1E6/ 5E4/262	51E6/1E6/ 5E4/605	100E6/50E6/ 5E4/276	300E6/200E6/ 10E4/153	51E6/1E6/5 E4/212

Table 9. Chain settings for regression analysis of pathwise speed. Number of iterations / burn-in / sampling frequency / and effective sample size (ESS). BM = Brownian Motion, LA = Lambda, OU = Ornstein-Uhlenbeck, VR = Variable Rate, VRLA = Variable Rate and Lambda. SL = Maximum standard length.

	BM	LA	OU	VR	VRLA
Speed ~ α + β_1(diadromous) + β_2(SL)	100E6/50E6/ 5E4/126	110E6/10E6/ 10E4/382	-	-	-
Speed ~ α + β_1(SL)	150E6/50E6/ 4E4/133	100E6/50E6/ 5E4/194	-	-	-
Speed	100E6/50E6/ 5E4/168	51E6/1E6/ 5E4/461	51E6/1E6/ 5E4/198	-	300E6/250E6/ 5E4/171

Table 10. Data sample size (n) of the estimated ancestral states, and the estimated branchwise rates. SL: Maximum standard length, WTT: Water temperature tolerance.

Variable	**n**
Ancestral SL	157,000
Ancestral WTT	157,000
Branchwise Rate SL	314,000
Branchwise Rate WTT	314,000
Branchwise Speed	312,000