



# Evidence from sperm whale clans of symbolic marking in non-human cultures

Taylor A. Hersh<sup>a,b,1</sup>, Shane Gero<sup>a,c,d</sup>, Luke Rendell<sup>e,f</sup>, Maurício Cantor<sup>g,h,i,j,k</sup>, Lindy Weilgart<sup>a</sup>, Masao Amano<sup>l</sup>, Stephen M. Dawson<sup>m</sup>, Elisabeth Slooten<sup>n</sup>, Christopher M. Johnson<sup>o,p,q</sup>, Iain Kerr<sup>q</sup>, Roger Payne<sup>q</sup>, Andy Rogan<sup>q</sup>, Ricardo Antunes<sup>r,s</sup>, Olive Andrews<sup>r,s</sup>, Elizabeth L. Ferguson<sup>t,3</sup>, Cory Ann Hom-Weaver<sup>t</sup>, Thomas F. Norris<sup>t</sup>, Yvonne M. Barkley<sup>u,v</sup>, Karlina P. Merkens<sup>v,w</sup>, Erin M. Oleson<sup>y</sup>, Thomas Doniol-Valcroze<sup>x</sup>, James F. Pilkington<sup>x</sup>, Jonathan Gordon<sup>f</sup>, Manuel Fernandes<sup>y</sup>, Marta Guerra<sup>m</sup>, Leigh Hickmott<sup>f,z</sup>, and Hal Whitehead<sup>a</sup>

Edited by Frans de Waal, Emory University, Atlanta, GA; received January 31, 2022; accepted July 19, 2022

Culture, a pillar of the remarkable ecological success of humans, is increasingly recognized as a powerful force structuring nonhuman animal populations. A key gap between these two types of culture is quantitative evidence of symbolic markers—seemingly arbitrary traits that function as reliable indicators of cultural group membership to conspecifics. Using acoustic data collected from 23 Pacific Ocean locations, we provide quantitative evidence that certain sperm whale acoustic signals exhibit spatial patterns consistent with a symbolic marker function. Culture segments sperm whale populations into behaviorally distinct clans, which are defined based on dialects of stereotyped click patterns (codas). We classified 23,429 codas into types using contaminated mixture models and hierarchically clustered coda repertoires into seven clans based on similarities in coda usage; then we evaluated whether coda usage varied with geographic distance within clans or with spatial overlap between clans. Similarities in within-clan usage of both “identity codas” (coda types diagnostic of clan identity) and “nonidentity codas” (coda types used by multiple clans) decrease as space between repertoire recording locations increases. However, between-clan similarity in identity, but not nonidentity, coda usage decreases as clan spatial overlap increases. This matches expectations if sympatry is related to a measurable pressure to diversify to make cultural divisions sharper, thereby providing evidence that identity codas function as symbolic markers of clan identity. Our study provides quantitative evidence of arbitrary traits, resembling human ethnic markers, conveying cultural identity outside of humans, and highlights remarkable similarities in the distributions of human ethnolinguistic groups and sperm whale clans.

culture | symbolic markers | bioacoustics | distribution | sperm whales

Culture—information or behavior shared within a community and acquired from conspecifics through social learning—has propelled the rapid ecological success of humans across the globe, profoundly shaping our individual and collective lives (1, 2). Despite the mounting evidence for culture as a powerful structuring force of nonhuman animal lives and societies (3), a key, unresolved gap between human and nonhuman cultures is the use and recognition of symbolic, arbitrary markers that broadcast cultural group membership and modulate conspecific interactions (4–8).

In humans, symbolic markers—sometimes referred to as ethnic markers, social markers, tags, or identity signals—are universal across cultures and can include accent, dialect, music, personal ornaments, tattoos, and/or clothing styles (9–11). Symbolic markers go beyond identity cues in that they not only denote cultural group identity but are also perceived as doing so by others and are then used as a means of categorizing individuals (7). As such, models predict (and empirical data supports) that different cultural groups will be most marked at boundary regions, cultural norm and symbolic marker distributions will correspond, and potential marker traits with the greatest initial differences will become marked first (4). While there are abundant examples of animals associating in ways that map onto differences in acoustic signals (12, 13), comparatively fewer studies have explicitly quantified whether spatial or temporal trends in those acoustic signals meet any of the aforementioned predictions for symbolic markers (5). Given that culture is a pervasive aspect of the lives of sperm whales (*Physeter macrocephalus*) (14–16), here we ask whether there is evidence of acoustic cultural traits being used to erect social barriers among groupings of whales.

Sperm whales live in multilevel societies composed of long-lasting social units (15). In social situations, whales communicate using socially learned, stereotyped patterns of clicks, called codas (listen to an example at <https://osf.io/ae6pd/>; *SI Appendix, Fig. S1*) (17). Variation in the number of clicks, rhythm, and tempo results in different coda types (17,

## Significance

Symbolic marking is a hallmark of human cultures, but quantitative evidence for nonhuman animal cultures is comparatively limited. We show evidence that certain acoustic communication signals—“identity codas”—function as symbolic markers of cultural identity among Pacific Ocean sperm whale clans. Within clans, identity and nonidentity coda usage shows similar patterns of stability over geographic space, but between-clan similarity in identity coda usage decreases with increasing spatial overlap. The distribution of sperm whale clans—with multiple cultural groups that vary widely in ranging, are unevenly distributed, and frequently overlap along a latitudinal gradient—resembles that of human ethnolinguistic groups. These findings demonstrate how culture can structure and fortify segmentation of animal populations and reveal similarities between human and nonhuman cultures.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: [taylor.a.hersh@gmail.com](mailto:taylor.a.hersh@gmail.com).

<sup>2</sup>Present address: independent researcher, Largs KA30 8EW, United Kingdom.

<sup>3</sup>Present address: Ocean Science Analytics, San Diego, CA 92129.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2201692119/-/DCSupplemental>.

Published September 8, 2022.

18). Even in sympatry, whales from one social unit only associate with whales from other units that use similar coda types (i.e., have a similar dialect), thereby delineating a higher cultural level—the vocal clan—in their populations (16). This preferential assortment around shared coda dialects suggests that sperm whales can discriminate cultural in-group vs. out-group members, resembling a discrimination that in humans is greatly facilitated by the use of symbolic markers (4, 6, 10). Given evidence that sperm whale clan dialects are the product of cultural transmission via biased social learning (19), it has been posited that certain coda types act as symbolic markers of clan identity (16, 20–22), but quantitative evidence is lacking. However, recent work showed that sperm whale clans in both the Atlantic and Pacific Oceans could be computationally distinguished by largely clan-specific “identity codas”, as opposed to “nonidentity codas” made by multiple clans (20). If these identity codas can be used by researchers to tell different clans apart, can they serve a similar function for the whales themselves?

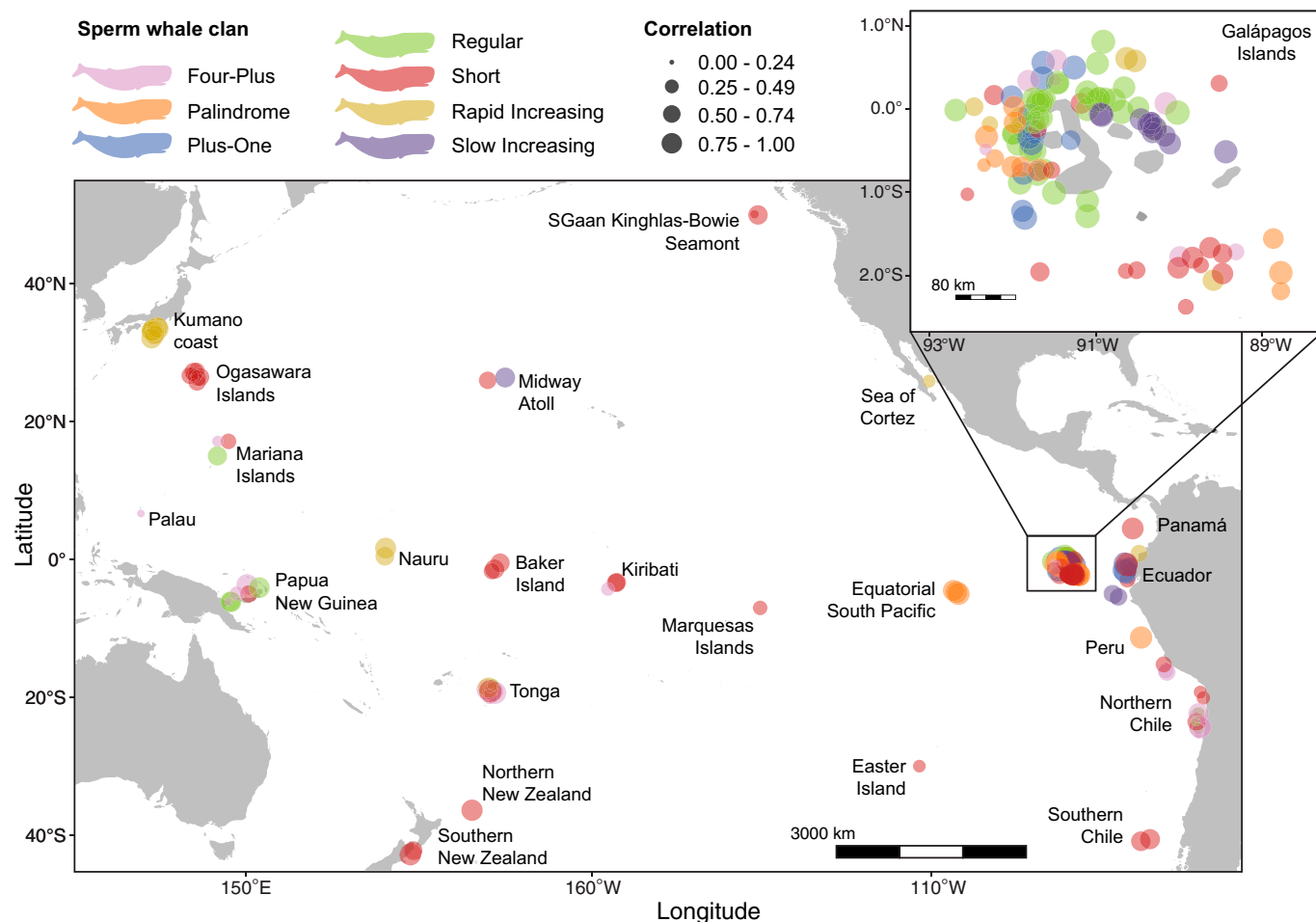
We address this question on an evolutionarily meaningful scale by combining coda recordings from an entire ocean basin. We determine how the coda usage similarity of dialects varies (A) spatially within clans and (B) with spatial overlap between clans across the Pacific Ocean. These objectives aim to discern possible evolutionary processes at play in the development and usage of identity and nonidentity codas comprising dialects and to investigate whether identity codas show hallmarks of being

symbolic markers of clan identity. We do so by quantifying whether and how identity and nonidentity coda usage is modulated by geographic clan overlap, a proxy for degree of sympatry. If identity codas are used as symbolic markers of clan identity, we hypothesize that their usage within clans will be more stable over geographic distance (due to selective pressures incurred from maintaining an identity signal within clans) than nonidentity coda usage. Informed by research on symbolic marking in humans (4, 6, 10, 23), we also hypothesize that identity coda usage will become more distinct as clan overlap increases, with no change predicted for nonidentity coda usage.

## Results

**Coda Dataset.** In total, 23,429 codas with 3–10 clicks were extracted from acoustic recordings of sperm whales and classified into types using contaminated mixture models via the “identity call” (IDcall) method (*SI Appendix, Table S1*) (20). Only recording days with at least 25 codas were included in subsequent clan analyses (16), with each day comprising a single repertoire (22,829 codas and 191 repertoires).

**Pacific Clans.** Hierarchical clustering of repertoires (via the IDcall method) suggests that seven clans inhabit the Pacific Ocean (Fig. 1 and *SI Appendix, Discussion S1*). Four are well known (“Four-Plus”, “Plus-One”, “Regular”, and “Short”



**Fig. 1.** Composite map of sperm whale clan distribution across 23 labeled regions in the Pacific Ocean using the seven-clan tree clan designations (*SI Appendix, Table S5* for regional recording metadata). Each point is a single repertoire, colored by clan. Point size indicates the within-clan correlation of that repertoire, with smaller points having lower correlations. The scale bar is approximate and most accurate along the Equator. *SI Appendix, Fig. S3* includes clan-specific maps.

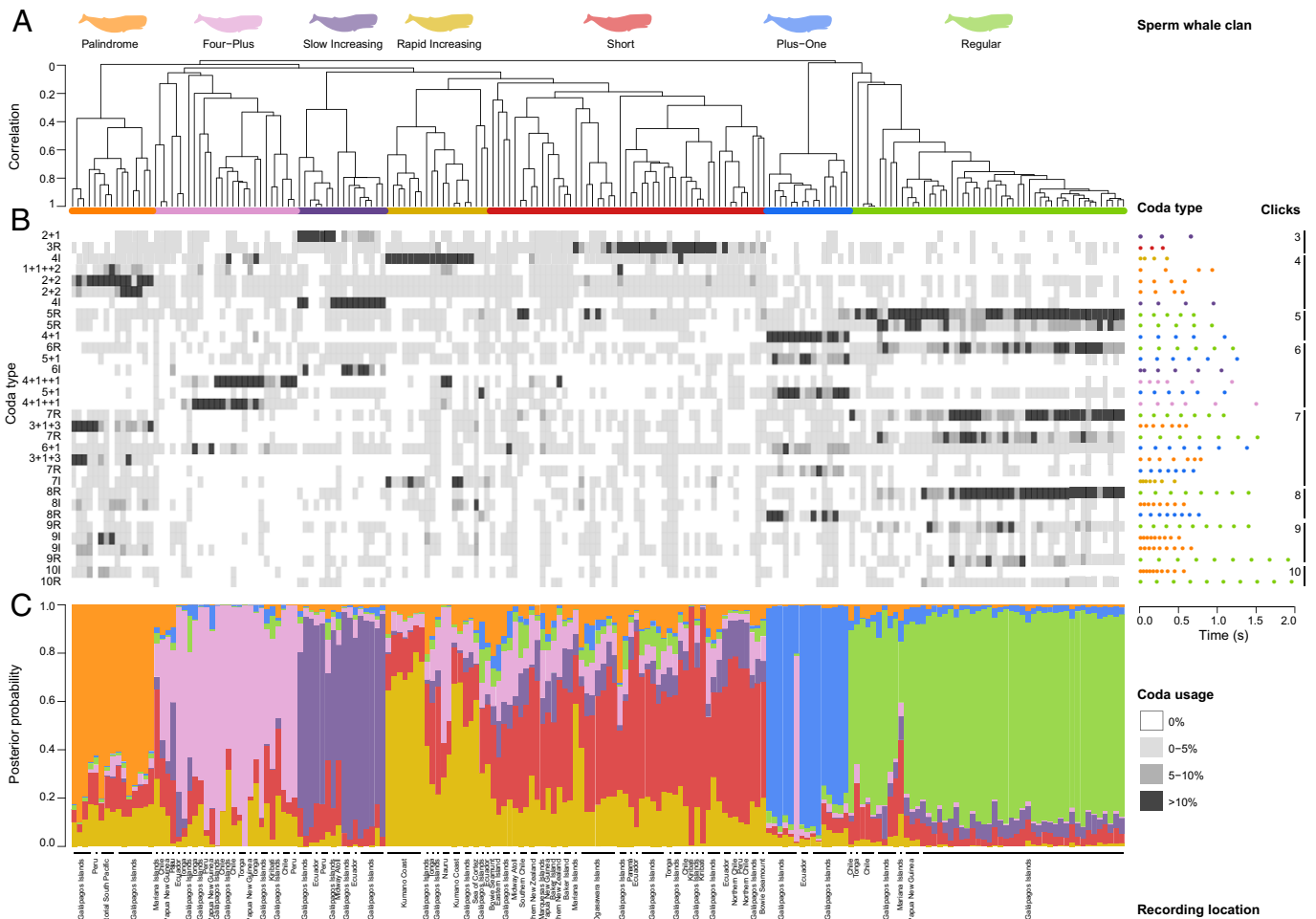
clans) (16, 24); two were previously documented, but not named (the “Palindrome” (20) and “Rapid Increasing” (25) clans); and one is new (“Slow Increasing” clan) (*SI Appendix, Discussion S2* includes clan naming conventions).

Within-clan repertoire correlation varies across clans, with some clans showing high consistency in coda type usage across repertoires (e.g., Plus-One and Regular) and others showing much lower uniformity (e.g., Four-Plus and Short) (*SI Appendix, Table S2*). The number of identity codas per clan also varies (*SI Appendix, Table S2*), and most clans display rhythmic “motifs” (16) in their identity codas (Fig. 2 and *SI Appendix, Discussion S2*). For example, the Regular clan’s nine identity codas all have equally spaced (i.e., isochronous) clicks, while the Four-Plus clan’s two identity codas both start with four isochronous clicks but subsequent click spacing increases. Codas with three to six clicks dominated, with preferences varying by clan (*SI Appendix, Fig. S2*).

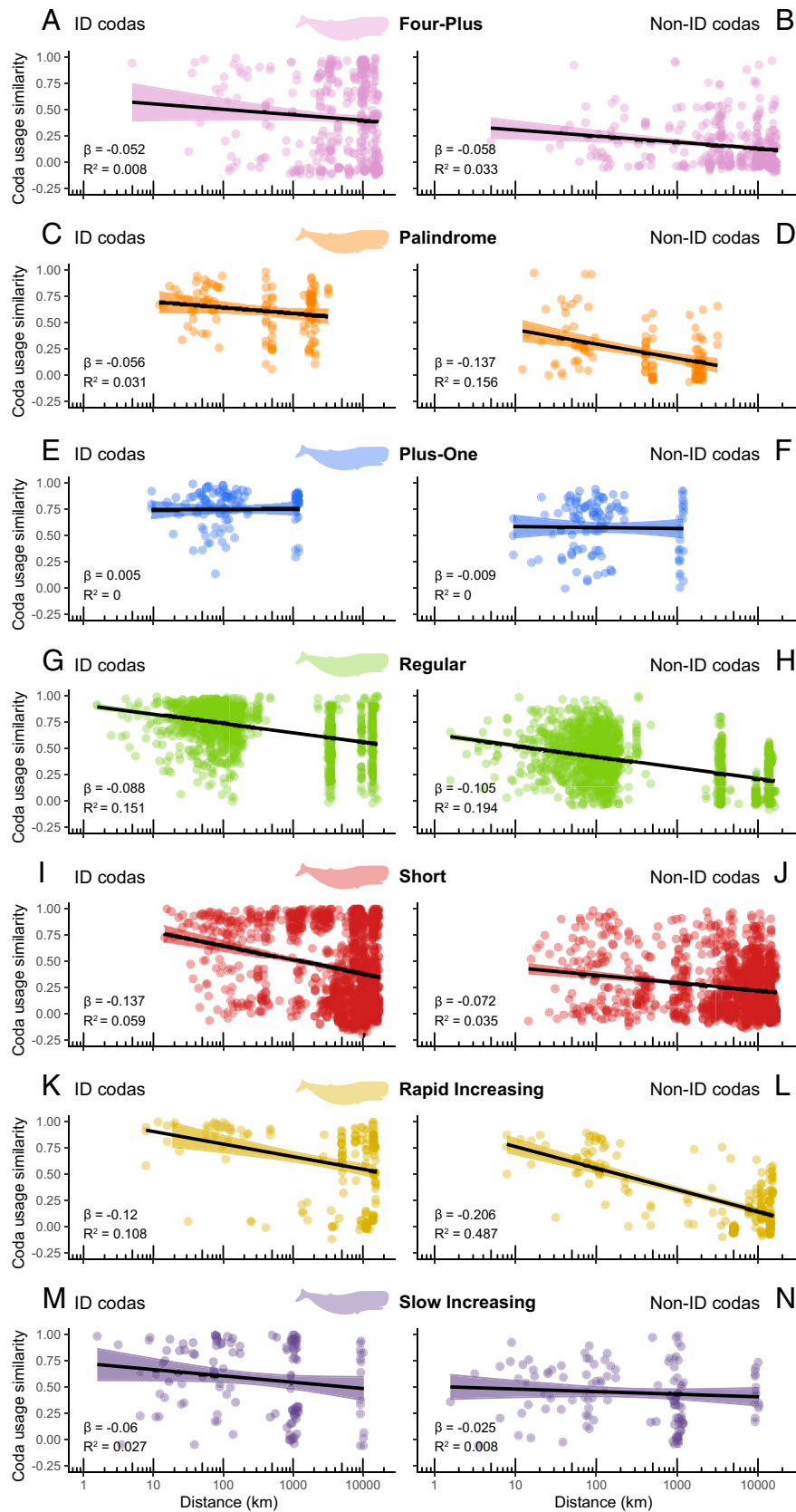
The distribution of clans across the Pacific Ocean varied, with some clans showing small ranges (~1,000 km, e.g., Plus-One), while others spanned the ocean basin (~10,000 km, e.g., Short) (Fig. 1 and *SI Appendix, Fig. S3*). Given unequal sampling across regions, our distribution maps provide evidence of clan presence, but not necessarily clan absence, especially in

regions with few recorded repertoires. Geographically sympatric clans were detected in 9/23 Pacific regions, with the highest diversity off the Galápagos Islands (seven clans), Ecuador (five clans), Tonga (four clans), and Northern Chile (four clans) (Fig. 1). There is a positive relationship between regional sampling effort (in terms of number of extracted codas and, especially, number of repertoires) and number of detected clans (*SI Appendix, Table S3*).

**Within-Clan Trends.** Similarity in identity coda usage between pairs of repertoires within a clan ( $sim_{IDwi}$ , where “sim” stands for coda usage similarity hereafter) decreased with logged geographic distance for all clans except Plus-One (Fig. 3). In general, a decrease in  $sim_{IDwi}$  is apparent at ranges greater than 1,000 km (the approximate span of the Plus-One clan). Similarity in nonidentity coda usage between pairs of repertoires within clans ( $sim_{nonIDwi}$ ) also decreased with logged geographic distance for all clans. For five out of seven clans, the slope of the linear regression was steeper for  $sim_{nonIDwi}$  than for  $sim_{IDwi}$  (Fig. 3). Although this provides support for our alternative hypothesis that within-clan identity coda usage is more stable over geographic space than nonidentity coda usage, the overall observed trend across clans was not significant ( $\beta = -0.019$ ,



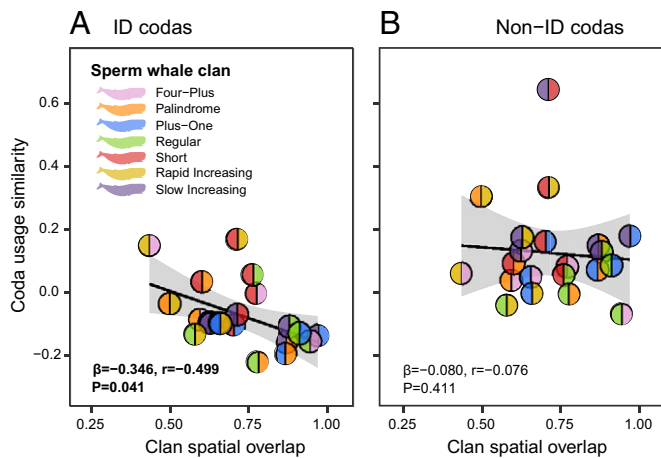
**Fig. 2.** Seven-clan tree with identity coda types (constructed in IDcall using  $critfact = 5$  and  $minrep = 15$ ). (A) Average linkage hierarchical clustering dendrogram depicts similarity among sperm whale coda repertoires recorded across the Pacific Ocean. Repertoires that are more similar, in terms of coda type usage, cluster closer together and have a higher correlation value (y axis). (B) Heat map depicts identity coda type usage (rows) for each repertoire (columns) in shades of gray. Usage is calculated based on probabilistic assignment of codas to types. Identity coda type names are on the left and type centroid rhythm plots (colored by clan) are on the right of the heat map (each dot represents a click). (C) Stacked bar plot shows the posterior probabilities of repertoire assignments to clans. Recording locations are listed along the bottom.



**Fig. 3.** Logged geographic distance vs. coda usage similarity for each clan, calculated using identity codas ( $sim_{IDw}$ ; left panels, *A/C/E/G/I/K/M*) or nonidentity codas ( $sim_{nonIDw}$ ; right panels, *B/D/F/H/J/L/N*) within a clan. Each circle represents a pair of coda repertoires. For each panel, the regression line slope ( $\beta$ ) and  $R^2$  value are provided. The 95% CIs are shaded.

$P = 0.267$ ). As the generally low  $R^2$  values indicate (Fig. 3), there is underlying variation in clan repertoire coda usage similarity that is not explained by geographic distance alone, and

the patchy sampling of clans across the Pacific Ocean means that linear regressions are not complete descriptions of the data but do indicate broad patterns.



**Fig. 4.** Mean clan spatial overlap vs. coda usage similarity, calculated using identity codas ( $sim_{IDbt}$ ; A) or nonidentity codas ( $sim_{nonIDbt}$ ; B). Each circle represents a pair of clans. For each panel, the regression line slope ( $\beta$ ), Mantel test matrix correlation coefficient ( $r$ ), and Mantel test  $P$  value ( $P$ ) are provided. Significant results are bolded, and 95% CIs are in gray. The simple Mantel tests are one tailed against the alternative that  $sim_{IDbt}/sim_{nonIDbt}$  decreases as clan spatial overlap increases.

**Between-Clan Trends.** Similarity in identity coda usage calculated for pairs of clans ( $sim_{IDbt}$ ) significantly ( $P = 0.043$ ) decreased as the mean spatial overlap (*SI Appendix, Table S4*) of the clans increased (Fig. 4). In contrast, similarity in nonidentity coda usage calculated for pairs of clans ( $sim_{nonIDbt}$ ) did not significantly change with increasing mean spatial overlap (Fig. 4). The Mantel test matrix correlation values were always more negative for identity codas compared with nonidentity codas, regardless of how clan spatial overlap was calculated (Fig. 4 and *SI Appendix, Fig. S4*). Thus, clans with overlapping distributions much more rarely use each others' identity codas than clans whose members rarely encounter one another.

## Discussion

We found that seven sperm whale cultural clans inhabit the Pacific Ocean and (i) have variable and overlapping geographic distributions, (ii) are generally well distinguished by identity codas, and (iii) exhibit some vocal characteristics that are modulated by clan spatial overlap.

**Distribution of Clans in the Pacific Ocean.** Each clan has a different distribution, with order of magnitude differences in the spatial extent of occupancy for some clans. Many regions have sympatric clans, and all clans spatially overlap with at least one other clan in some portion of their detected range. The general increase in the number of detected clans in regions with greater sampling effort suggests that sympatric clans may be the rule rather than the exception across the Pacific and that there may be additional clans in un(der)sampled regions.

The distributions of sperm whale clans across the Pacific Ocean echo patterns of human ethnolinguistic diversity. Resemblances include multiple cultural groups (that vary widely in ranging) being unevenly distributed in areas of high and low diversity along a latitudinal gradient from the equator to the poles, with frequent regions of overlap (26). The global distribution of human ethnolinguistic groups has been driven by political complexity, environmental productivity and heterogeneity, and subsistence strategies (26–28). Some of these have homologs or proxies in sperm whales (e.g., social complexity,

ocean productivity and heterogeneity, and foraging strategies) and should be the focus of future research.

**Clan Identity Codas.** All repertoires recorded across the Pacific Ocean were assigned to clans based on identity codas made frequently within, and rarely outside of, the clans. Within clans, coda usage similarity generally decreased as distance between repertoire recording locations increased. An exception was the most geographically restricted clan, the Plus-One clan, which spanned  $\sim 1,000$  km (the approximate annual home range of eastern tropical Pacific sperm whales) (29, 30). Isolation by distance has been documented in culturally transmitted attributes of other species, such as human material culture (e.g., the beadwork patterns of High Plains ethnolinguistic tribes) (31) and red-faced cisticola (*Cisticola erythropis*) song (32). The observed decrease in coda usage similarity over space was steeper for nonidentity codas compared with identity codas, but this trend was not significant across all clans. Thus, while most clans exhibit patterns in line with our first hypothesis—that identity coda usage is more stable than nonidentity coda usage over geographic distance, potentially due to selective pressure for symbolic marker stability—the effect across clans is not statistically significant.

Pacific Ocean sperm whale clans vary, sometimes substantially, in number of identity codas, coda click length preferences, and within-clan repertoire correlations. Some commonalities also emerge, such as rhythmic motifs in the identity codas of most clans (exact motifs vary by clan). A more in-depth characterization of the full coda repertoires of each clan will help determine how far these motifs extend beyond identity codas. One caveat is our choice to restrict analyses to 3- to 10-click codas, which could obscure some clan boundaries. For example, two-click codas are made off the Ogasawara Islands of Japan and the Mariana Islands but were not analyzed here because they have been inconsistently marked across datasets; these very short codas may be culturally relevant to certain clans, and their omission could explain why some of the Mariana Islands repertoires have low within-clan correlations.

**Modulation of Clan Vocal Behavior by Clan Overlap.** Identity coda usage similarity significantly decreased as clan spatial overlap increased, in contrast to no significant trend in nonidentity coda usage. This finding supports the hypothesis that identity codas act as symbolic markers of clan identity, aligning with human empirical and simulation work showing that ethnic group differences are strongest at boundaries for symbolically marked groups (4, 23). Here, more spatially overlapped clans appear to be “more marked” than less spatially overlapped clans, which suggests that between-clan interactions have increased selection on usage of identity markers (in this case, identity codas) (4). This pattern fits with cross-species evidence showing that dialects of groups or species in sympatry are typically more distinct than in allopatry (33–37). As sperm whale acoustic sampling efforts continue in regions with just one or a few repertoires, our understanding of possible interactions between clan spatial overlap and coda usage similarity will continue to improve.

No single coda type is unique to a single clan, which means that clans occasionally make each other's identity codas. However, the between-clan analyses suggest that clans make other clans' identity codas less frequently in areas of higher spatial overlap. Theoretically, this adjustment would improve a whale's ability to quickly discern whether nearby individuals are from their own clan or not, which could guide decisions about social assortment or avoidance (21, 38). Although sperm whales

associate at the surface, over 70% of their time is spent independently foraging (39). Identity coda symbolic markers could help whales reunite and reaffirm their social ties in between long foraging dives, even in the absence of other clans. To borrow from Savage et al. (40), identity codas likely “serve multiple evolutionary functions within the umbrella of social bonding”, similar to human music.

When quantifying selection pressure, understanding the characteristics of the basal or source population can be informative and is a requisite for drawing conclusions about whether and how between-group interactions have impacted selection on group identity markers (4). For example, phylogenetic models informed by cognate evolution rates can be used to infer ancestral human dialects, which can help track historic migration patterns of speakers (41, 42). Results from our analyses provide preliminary support for the Short clan dialect as basal in the Pacific Ocean (*SI Appendix, Discussion S3*), but a detailed understanding of how codas evolve over time is necessary before we can determine the “ancestral” sperm whale dialect.

**Implications and Conclusions.** Symbolic marking is considered a defining trait of human cultures (5, 43). Our study provides quantitative evidence that identity codas act as symbolic markers of sperm whale cultural (i.e., clan) identity in the Pacific Ocean, with whales adjusting their vocal behavior in ways that could reaffirm their cultural identity in areas of high spatial overlap. The distributional scales and trends seen in Pacific sperm whales resemble those found in human ethnic groups (4, 6). Although we cannot know for certain how whales perceive codas or label the identity of conspecifics, the observed patterns could arise from the selection for identity signals that enable social assortment and cooperation in sperm whales, as has been shown for humans (6, 44) and suggested for other taxa (5, 37, 45, 46). Playback experiments comparing behavioral responses of whales exposed to same-clan, different-sympatric clan, or different-allopatric clan identity codas are essential for validating the patterns described here (47), but such experimental studies with wild marine mammals remain logistically challenging (48). In the interim, studies like ours that examine whether specific acoustic signals meet symbolic marker predictions can be informative for cultural species.

Collectively, these discoveries shed light on how different clans maintain behavioral distinctiveness and how cultural boundaries persist in areas of sympatry. In doing so, this work takes a key step toward understanding the evolutionary mystery of why clans exist by documenting evidence of how they may persist. A preference for interacting with individuals whose symbolic markers match your own has been favored by natural selection in humans (44). Our results suggest that the conditions requisite for the evolution of these markers exist among nonhuman animals as well.

## Materials and Methods

**Data Collection and Coda Extraction.** Acoustic recordings of sperm whales were collected between 1978 and 2017 in 23 Pacific Ocean locations (Fig. 1; for regional methods, *SI Appendix, Table S5*). For 12 regions (starred in *SI Appendix, Table S5*), coda metrics (number of clicks and interclick intervals) had previously been extracted (16, 18, 20, 24, 25), and additional codas were extracted for our analyses. Codas recorded off Tonga were extracted but unpublished. For the remaining regions, audio recordings of sperm whales were audited for codas in Audacity (version 2.3.0). Coda metrics were extracted using “Coda Sorter”—a custom software implemented in LabView and run in MATLAB (version R 2020a). Extracted codas with 3–10 clicks were divided into repertoires by recording day, and repertoire locations were inferred from satellite navigation

data with varying levels of precision (*SI Appendix, Table S5*). When multiple positional fixes were available for a given day of recordings, we took the average. For repertoires with identical recording locations, we jittered the locations using the “geoR” R package (49) prior to analyses. The full extracted coda dataset can be accessed via Open Science Framework (<https://osf.io/ae6pd/>).

**Coda Typing and Clan Assignment.** We used the IDcall method (20) to classify codas into types and assign repertoires to clans. Codas with the same number of interclick intervals are classified into types using contaminated mixture models, with each coda assigned a probability of belonging to each type. Repertoires are hierarchically clustered based on similarities in coda type usage. An “identity clade” of repertoires (i.e., a putative sperm whale clan) is denoted if it has at least one identity coda (i.e., a coda type made frequently in that clade of repertoires but rarely in any other). Two corresponding parameters must be set to delineate identity clades and codas: *minrep* (i.e., minimum number of repertoires required for an identity clade to form) and *critfact* (i.e., factor by which a coda type must be made more frequently in a given clade compared with other clades for it to become an identity coda and for the clade to become an identity clade). We used the default parameter settings from (20) with a few deviations (*SI Appendix, Method S1*). With *minrep* set to 15, we tested a range of values for the *critfact* parameter (from 3 to 20) during hierarchical clustering and compared the resultant dendrograms to determine the most probable number of sperm whale clans in the Pacific Ocean (*SI Appendix, Discussion S1*). We also calculated the Bayesian posterior probabilities (using all coda types) that each repertoire is a member of each identity clade using the IDcall posterior probability (IDcallPP) methodology (50). In IDcallPP, a jackknife procedure ensures that information from a repertoire is not used to define the clans with which it is then compared (50). Maps showing clan distributions were created using the “rgeos” (51), “sp” (52), and “ggspatial” R packages.

**Within-Clan Analyses.** Within clans, our goal was to determine whether coda usage similarity of repertoires (calculated as correlations between usage of coda types) varied with geographic distance between repertoires. For each pair of repertoires, we plotted the logged geographic distance separating them (*SI Appendix, Method S2*) against their coda usage similarity and fit a linear regression using the “stats” R package. The regression slope indicated how coda usage similarity varied with increasing distance. The coda usage similarity of each pair of repertoires within clans was calculated based on correlations in usage of identity codas ( $\text{sim}_{\text{IDwi}}$ ) or nonidentity codas ( $\text{sim}_{\text{nonIDwi}}$ ).

Having found a general pattern of decreasing coda usage similarity with increasing geographic distance, we tested the null hypothesis that coda usage similarity of repertoires within a clan, calculated as  $\text{sim}_{\text{IDwi}}$  or  $\text{sim}_{\text{nonIDwi}}$ , drops off similarly with increasing distance. If identity codas are used by sperm whales as symbolic markers of clan identity, we predicted that within-clan identity coda usage would be more stable over geographic space than nonidentity coda usage, with nonidentity codas more susceptible to drift. We explicitly tested these hypotheses across clans using a weighted linear mixed effects model, where the response variable is the slope parameter estimates of the individual clan regressions ( $\beta$ ). The predictors include a clan random effect and an identity/nonidentity fixed effect (i.e., whether repertoire coda usage similarity was calculated from just identity codas or just nonidentity codas). We then weight this model by the inverse squared SD of each clan parameter estimate, such that more precisely estimated slopes have higher influence.

**Between-Clan Analyses.** Between clans, our goal was to determine whether coda usage similarity significantly varied with spatial overlap. For each pair of clans, spatial overlap was calculated as the proportion of the first clan’s repertoires that were recorded within 1,000 km (the approximate annual home range span of eastern tropical Pacific sperm whales) of at least one of the second clan’s repertoires (29, 30). This produces an asymmetric matrix, as a clan found in only one region might overlap completely with a clan that spans the ocean, while the inverse is not true (*SI Appendix, Table S4*). Mean spatial overlap was calculated as the average of the two values for each pair of clans. Given the patchy sampling of codas in different regions in different years (with 87% of regions represented by just 1 or 2 years; *SI Appendix, Table S1*), our analyses do not include a temporal component. Theoretically, two clans could thus have a nonzero spatial overlap value without having been recorded in the same year.

Between-clan coda usage similarity was calculated by first averaging the identity or nonidentity coda type usages across all the repertoires in each clan into a single repertoire per clan and then calculating the overall correlation in coda type (identity or nonidentity) usage for all clan pairs. We then plotted the minimum (SI Appendix, Fig. S4), mean (Fig. 4), and maximum (SI Appendix, Fig. S4) spatial overlap against  $sim_{IDbt}$  or  $sim_{nonIDbt}$  for all clan pairs. We used one-tailed simple Mantel tests with 10,000 permutations to determine the direction and significance of any correlation, testing the null hypothesis that coda usage similarity between clans does not decrease with increasing spatial overlap.

Having found a general pattern of decreasing  $sim_{IDbt}$  (but not  $sim_{nonIDbt}$ ) with increasing geographic distance, we tested the null hypothesis that between-clan coda usage similarity, whether calculated as  $sim_{IDbt}$  or  $sim_{nonIDbt}$ , falls off similarly with increasing clan overlap. If identity codas are used by sperm whales as symbolic markers of clan identity, we predicted that  $sim_{IDbt}$  would decrease with increasing clan overlap, because whales would modulate identity coda production in some way that enhances clan distinctiveness in areas of greater overlap. In contrast, we did not expect a trend in  $sim_{nonIDbt}$  if, as we posit, nonidentity codas are generally not the primary markers of clan identity.

**Data, Materials, and Software Availability.** [csv file of coda interclick-intervals for 3- to 10-click codas included in analyses. R file of analytical code] data have been deposited in [Open Science Framework] (<https://osf.io/a66pd/>) (53). Previously published data were used for this work (T. A. Hersh, S. Gero, L. Rendell, H. Whitehead, Using identity calls to detect structure in acoustic datasets. *Methods Ecol. Evol.* 2021, 1–11 (2021). M. Amano, A. Kourogi, K. Aoki, M. Yoshioka, K. Mori, Differences in sperm whale codas between two waters off Japan: possible geographic separation of vocal clans. *J. Mammal.* 95, 169–175 (2014)).

**ACKNOWLEDGMENTS.** This manuscript emanates from the Global Coda Dialect Project, a consortium of scientists conducting sperm whale acoustics research worldwide. Portions of this work were developed from the lead author's doctoral thesis (54). We are grateful to the individuals that helped collect acoustic data and the organizations that funded regional research efforts; Noor Youssef and Scott McCain for conversations that helped orient this work; Stephanie King, Andrew Horn, Aaron Newman, and three anonymous referees for providing helpful feedback on manuscript drafts; Kristian Beedholm for creating and sharing

Coda Sorter; Robin Baird and Katie Kowarski for connecting us with individuals who had sperm whale acoustic data; and H eloise Frouin-Mouy, Lynn Lee, Jasper Kanes, Ocean Networks Canada, and Richard Dewey for sharing sperm whale recordings that were ultimately not included in this publication. SI Appendix, Table S5 includes sperm whale fieldwork funding information.

Author affiliations: <sup>a</sup>Department of Biology, Dalhousie University, Halifax, NS B3H 4R2, Canada; <sup>b</sup>Comparative Bioacoustics Group, Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, Netherlands; <sup>c</sup>Department of Biology, Carleton University, Ottawa, ON K1S 5B6, Canada; <sup>d</sup>Zoophysiology, Department of Bioscience, Aarhus University, Aarhus 8000, Denmark; <sup>e</sup>Centre for Social Learning and Cognitive Evolution, School of Biology, University of St. Andrews, St. Andrews KY16 9ST, United Kingdom; <sup>f</sup>Sea Mammal Research Unit, Scottish Oceans Institute, University of St. Andrews, St. Andrews KY16 8LB, United Kingdom; <sup>g</sup>Department of Fisheries, Wildlife, and Conservation Sciences, Marine Mammal Institute, Oregon State University, Newport, OR 97365; <sup>h</sup>Departamento de Ecologia e Zoologia, Universidade Federal de Santa Catarina, Florian opolis 88040-970, Brazil; <sup>i</sup>Department for the Ecology of Animal Societies, Max Planck Institute of Animal Behavior, Radolfzell 78315, Germany; <sup>j</sup>Centro de Estudos do Mar, Universidade Federal do Paran , Pontal do Paran  83255-000, Brazil; <sup>k</sup>School of Plant, Animal and Earth Sciences, University of the Witwatersrand, Johannesburg 2000, South Africa; <sup>l</sup>Graduate School of Fisheries and Environmental Sciences, Nagasaki University, Nagasaki 852-8521, Japan; <sup>m</sup>Department of Marine Science, University of Otago, Dunedin 9016, New Zealand; <sup>n</sup>Department of Zoology, University of Otago, Dunedin 9016, New Zealand; <sup>o</sup>World Wide Fund for Nature, Melbourne, VIC 3053, Australia; <sup>p</sup>Centre for Marine Science and Technology, Curtin University, Perth, WA 6845, Australia; <sup>q</sup>Ocean Alliance, Gloucester, MA 01930; <sup>r</sup>The South Pacific Whale Research Consortium, Avarua, Cook Islands; <sup>s</sup>Institute of Marine Science, University of Auckland, Auckland 1142, New Zealand; <sup>t</sup>Bio-Waves, Inc., Encinitas, CA 92024; <sup>u</sup>Hawai  Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawai  at Manoa, Kaneohe, HI 96744; <sup>v</sup>National Oceanic and Atmospheric Administration Pacific Islands Fisheries Science Center, Honolulu, HI 96818; <sup>w</sup>Lynker Technologies, Limited Liability Company, Portland, OR 97219; <sup>x</sup>Fisheries and Oceans Canada, Cetacean Research Program, Pacific Biological Station, Nanaimo, BC V9T 6N7, Canada; <sup>y</sup>Biological Sciences Department, University of Canterbury, Christchurch 8140, New Zealand; and <sup>z</sup>Open Ocean Consulting, Hampshire GU32 3LF, United Kingdom

Author contributions: T.A.H., S.G., and H.W. designed research; T.A.H., S.G., L.R., M.C., L.W., M.A., S.M.D., E.S., C.M.J., I.K., R.P., A.R., R.A., O.A., E.L.F., C.A.H.-W., T.F.N., Y.M.B., K.P.M., E.M.O., T.D.-V., J.F.P., J.G., M.F., M.G., L.H., and H.W. performed research; T.A.H., S.G., L.R., M.C., and H.W. analyzed data; T.A.H., S.G., L.R., M.C., and H.W. wrote the paper; T.A.H., L.R., M.C., L.W., M.A., S.M.D., E.S., C.M.J., I.K., R.P., A.R., R.A., O.A., E.L.F., C.A.H.-W., T.F.N., Y.M.B., K.P.M., E.M.O., T.D.-V., J.F.P., J.G., M.F., M.G., L.H., and H.W. contributed data; L.W., M.A., S.M.D., E.S., C.M.J., I.K., R.P., A.R., R.A., O.A., E.L.F., C.A.H.-W., T.F.N., Y.M.B., K.P.M., E.M.O., T.D.-V., J.F.P., J.G., M.F., M.G., and L.H. provided feedback; L.W., M.A., S.M.D., E.S., C.M.J., I.K., R.P., A.R., R.A., O.A., E.L.F., C.A.H.-W., Y.M.B., K.P.M., E.M.O., T.D.-V., J.F.P., J.G., M.F., M.G., and L.H. read and approved the manuscript.

- R. Boyd, P. J. Richerson, *The Origin and Evolution of Cultures* (Oxford University Press, 2005).
- A. V. Bell, P. J. Richerson, R. McElreath, Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17671–17674 (2009).
- A. Whiten, The burgeoning reach of animal culture. *Science* **372**, eabe6514 (2021).
- R. McElreath, R. Boyd, P. J. Richerson, Shared norms and the evolution of ethnic markers. *Curr. Anthropol.* **44**, 122–130 (2003).
- S. Perry, "Are nonhuman primates likely to exhibit cultural capacities like those of humans?" in *The Question of Animal Culture*, K. N. Laland, B. G. Galef, Eds. (Harvard University Press, Cambridge, MA, 2009), pp. 247–268.
- R. Boyd, P. J. Richerson, The evolution of ethnic markers. *Cult. Anthropol.* **21**, 65–79 (1987).
- F. Barth, "Ethnic groups and boundaries: The social organization of cultural difference" in *Introduction* (Waveland Press, 1969), pp. 9–39.
- C. Efferson, R. Lalive, E. Fehr, The coevolution of cultural groups and ingroup favoritism. *Science* **321**, 1844–1849 (2008).
- A. V. Bell, A. Paegle, Ethnic markers and how to find them: An ethnographic investigation of marker presence, recognition, and social information. *Hum. Nat.* **32**, 470–481 (2021).
- E. Cohen, The evolution of tag-based cooperation in humans: The case for accent. *Curr. Anthropol.* **53**, 588–616 (2012).
- D. Nettle, R. I. M. Dunbar, Social markers and the evolution of reciprocal exchange. *Curr. Anthropol.* **38**, 93–99 (1997).
- E. Balaban, Cultural and genetic variation in swamp sparrows (*Melospiza georgiana*): II. Behavioral salience of geographic song variants. *Behaviour* **105**, 292–322 (1988).
- R. Riesch, L. G. Barre t-Lennard, G. M. Ellis, J. K. B. Ford, V. B. Deecke, Cultural traditions and the evolution of reproductive isolation: Ecological speciation in killer whales? *Biol. J. Linn. Soc. Lond.* **106**, 1–17 (2012).
- H. Whitehead, L. Rendell, *The Cultural Lives of Whales and Dolphins* (University of Chicago Press, 2014).
- H. Whitehead, *Sperm Whales: Social Evolution in the Ocean* (University of Chicago Press, 2003).
- L. E. Rendell, H. Whitehead, Vocal clans in sperm whales (*Physeter macrocephalus*). *Proc. R. Soc. London Ser. B Biol. Sci.* **270**, 225–231 (2003).
- W. A. Watkins, W. E. Schevill, Sperm whale codas. *J. Acoust. Soc. Am.* **62**, 1485–1490 (1977).
- L. Weilgart, H. Whitehead, Group-specific dialects and geographic variation in coda repertoire in South Pacific sperm whales. *Behav. Ecol. Sociobiol.* **40**, 277–285 (1997).
- M. Cantor *et al.*, Multilevel animal societies can emerge from cultural transmission. *Nat. Commun.* **6**, 8091 (2015).
- T. A. Hersh, S. Gero, L. Rendell, H. Whitehead, Using identity calls to detect structure in acoustic datasets. *Methods Ecol. Evol.* **2021**, 1–11 (2021).
- M. Cantor, H. Whitehead, The interplay between social networks and culture: Theoretically and among whales and dolphins. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120340 (2013).
- S. Gero, H. Whitehead, L. Rendell, Individual, unit and vocal clan level identity cues in sperm whale codas. *R. Soc. Open Sci.* **3**, 150372 (2016).
- T. Lund n, "Religious symbols as boundary markers in physical landscapes: An aspect of human geography" in *De-Bordering, Re-Bordering, and Symbols on the European Boundaries*, J. Ja ncaz, Ed. (Logos Verlag Berlin, 2011), pp. 9–19.
- M. Cantor, H. Whitehead, S. Gero, L. Rendell, Cultural turnover among Gal pagos sperm whales. *R. Soc. Open Sci.* **3**, 160615 (2016).
- M. Amano, A. Kourogi, K. Aoki, M. Yoshioka, K. Mori, Differences in sperm whale codas between two waters off Japan: Possible geographic separation of vocal clans. *J. Mammal.* **95**, 169–175 (2014).
- T. E. Currie, R. Mace, The evolution of ethnolinguistic diversity. *Adv. Complex Syst.* **15**, 1–20 (2012).
- T. E. Currie, R. Mace, Political complexity predicts the spread of ethnolinguistic groups. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7339–7344 (2009).
- T. Honkola *et al.*, Evolution within a language: Environmental differences contribute to divergence of dialect groups. *BMC Evol. Biol.* **18**, 132 (2018).
- H. Whitehead, A. Coakes, N. Jaquet, S. Lusseau, Movements of sperm whales in the tropical Pacific. *Mar. Ecol. Prog. Ser.* **361**, 291–300 (2008).
- H. Whitehead, Analysis of animal movement using opportunistic individual identifications: Application to sperm whales. *Ecology* **82**, 1417–1432 (2001).
- S. J. Lycett, Confirmation of the role of geographic isolation by distance in among-tribe variations in beadwork designs and manufacture on the High Plains. *Archaeol. Anthropol. Sci.* **11**, 2837–2847 (2019).
- L. Benedict, R. C. K. Bowie, Macrogeographical variation in the song of a widely distributed African warbler. *Biol. Lett.* **5**, 484–487 (2009).
- K. S. Pfennig, D. W. Pfennig, Character displacement: Ecological and reproductive responses to a common evolutionary problem. *Q. Rev. Biol.* **84**, 253–276 (2009).
- Y. Jang, H. C. Gerhardt, Divergence in female calling song discrimination between sympatric and allopatric populations of the southern wood cricket *Gryllus fultoni* (Orthoptera: Gryllidae). *Behav. Ecol. Sociobiol.* **60**, 150–158 (2006).
- J. Haavie *et al.*, Flycatcher song in allopatry and sympatry—convergence, divergence and reinforcement. *J. Evol. Biol.* **17**, 227–237 (2004).
- G. H obel, H. C. Gerhardt, Reproductive character displacement in the acoustic communication system of green tree frogs (*Hyla cinerea*). *Evolution* **57**, 894–904 (2003).
- C. Crockford, I. Herbinger, L. Vigilant, C. Boesch, Wild chimpanzees produce group-specific calls: A case for vocal learning? *Ethology* **110**, 221–243 (2004).
- M. Cantor, H. Whitehead, How does social behavior differ among sperm whale clans? *Mar. Mamm. Sci.* **31**, 1275–1290 (2015).

39. S. L. Watwood, P. J. O. Miller, M. Johnson, P. T. Madsen, P. L. Tyack, Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*). *J. Anim. Ecol.* **75**, 814–825 (2006).
40. P. E. Savage *et al.*, Music as a coevolved system for social bonding. *Behav. Brain Sci.* **44**, e59 (2020).
41. T. E. Currie, A. Meade, M. Guillon, R. Mace, Cultural phylogeography of the Bantu Languages of sub-Saharan Africa. *Proc. Biol. Sci.* **280**, 20130695 (2013).
42. R. Bouckaert, *et al.*, Mapping the origins and expansion of the Indo-European language family. *Science* **337**, 957–960 (2012).
43. E. Jablonka, M. J. Lamb, *Evolution in Four Dimensions, Revised Edition: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life* (MIT Press, 2014).
44. M. W. Moffett, Human identity and the evolution of societies. *Hum. Nat.* **24**, 219–267 (2013).
45. A. M. Van Cise, S. D. Mahaffy, R. W. Baird, T. A. Mooney, J. Barlow, Song of my people: Dialect differences among sympatric social groups of short-finned pilot whales in Hawai'i. *Behav. Ecol. Sociobiol.* **72**, 193 (2018).
46. O. A. Filatova *et al.*, Call diversity in the North Pacific killer whale populations: Implications for dialect evolution and population history. *Anim. Behav.* **83**, 595–603 (2012).
47. J. Andreas *et al.*, Toward understanding the communication in sperm whales. *iScience* **25**, 104393 (2022).
48. V. B. Deecke, Studying marine mammal cognition in the wild: A review of four decades of playback experiments. *Aquat. Mamm.* **32**, 461–482 (2006).
49. P. J. Ribeiro, Jr, P. J. Diggle, The geoR package. *R News* **1**, 14–18 (2001).
50. H. Whitehead, T. A. Hersh, Posterior probabilities of membership of repertoires in acoustic clades. *PLoS One* **17**, e0267501 (2022).
51. R. Bivand *et al.*, *Package 'rgeos* (Compr. R Arch. Netw) 2017).
52. E. J. Pebesma, Simple features for R: Standardized support for spatial vector data. *R J.* **10**, 439 (2018).
53. T. Hersh *et al.*, Evidence from sperm whale clans of symbolic marking in non-human cultures. OSF. <https://osf.io/ae6pd/>. Deposited 31 January 2022.
54. T. A. Hersh, *Dialects Over Space and Time: Cultural Identity and Evolution in Sperm Whale Codas* (Dalhousie University, Halifax, Nova Scotia, Canada, 2021).