Journal of Archaeological Method and Theory

Supplementary Information for "Characterizing core morphology and scar patterning within a unified spherical harmonic framework"

Peiyuan Xiao[a,b,c], Hao Li[a,d,✉], Ben Marwick[c,✉]

[a] Alpine Paleoecology and Human Adaptation Group (ALPHA Group), State Key Laboratory of Tibetan Plateau Earth System, Environment and Resources, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China
[b] University of Chinese Academy of Sciences, Beijing, China
[c] Department of Anthropology, University of Washington, Seattle, WA, USA
[d] MOE Key Laboratory of Western China's Environmental System, Center for Excellence in Archaeological Science and Cultural Heritage, Lanzhou University, Lanzhou, China

✉ Correspondence: Hao Li <lihao@itpcas.ac.cn>, Ben Marwick <bmarwick@uw.edu>

Sensitivity analysis

To evaluate the stability of the spherical harmonic framework, we conducted a parameter sensitivity robustness analysis. The purpose of this analysis was to determine whether the main computational results depend on a single selected parameter setting, or whether they remain stable across a reasonable range of values for key control parameters. This is an important step in assessing algorithmic performance, because the outputs of a computational method may be shaped not only by the empirical structure of the data, but also by analyst-defined parameters (Hamby, 1994; Saltelli, 2002).

In computational modelling, sensitivity analysis is commonly used to examine how changes in model inputs or assumptions affect model outputs and the resulting inferences (Saltelli, 2002). In our study, the tested parameters may influence how the spherical harmonic descriptors represent core morphology and scar patterning. Therefore, if the main patterns of variation, group separation, effect size ranking, and covariation results remain stable when these parameters are varied within reasonable ranges, the method and the selected main analysis settings can be considered robust and justified. Conversely, if the analytical conclusions appear only under a specific parameter setting, those conclusions would be parameter dependent, which would also limit the broader generalizability of the analytical method.

We therefore assessed whether the main results depend on the principal free parameters of the two descriptors. For the scar pattern descriptor (SP-SPHARM), we varied the von Mises–Fisher kernel bandwidth (h) and the scar minimum-size inclusion threshold. For the core morphology descriptor (M-SPHARM), we varied the mesh decimation target and the number of Laplacian smoothing iterations. In each parameter sweep, all settings other than the focal parameter were held at the values used in the main analysis. At the main analysis setting, the recomputed

descriptors reproduced the previously saved descriptor values, confirming that each sensitivity sweep was anchored to the results reported in the main analysis. As in the main text, we interpret the results primarily in terms of effect magnitudes rather than significance thresholds. The analyses below show repeatedly that parameter perturbation affects the significance of small-to-modest effects near the 0.05 threshold, but not the magnitude of the effects themselves.

Bandwidth of the von Mises–Fisher kernel

The spherical kernel density bandwidth was fixed at $h = 0.35$ (concentration $\kappa = 1/h^2 = 8.16$) in the main analysis. We recomputed the SP-SPHARM power spectra over $h \in \{0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50\}$ ($\kappa = 4.0$–25.0), holding everything else fixed (alignment, the spherical evaluation grid and Driscoll–Healy transform, $l_{\max} = 20$, normalisation, and seeds). Morphology (M-SPHARM) is independent of h and was reused unchanged. At $h = 0.35$ the recomputed spectra reproduce the cached values.

Across all h, SP-SPHARM power saturates early: cumulative power through $l = 6$ is 96.71–99.99%, and the cross specimen coefficient of variation (CV) first exceeds 100% at degree $l = 7$–14. The $l = 1$–6 truncation is robust to h; it is most complete at moderate-to-large h (at the smallest h a little reliable signal persists beyond $l = 6$).

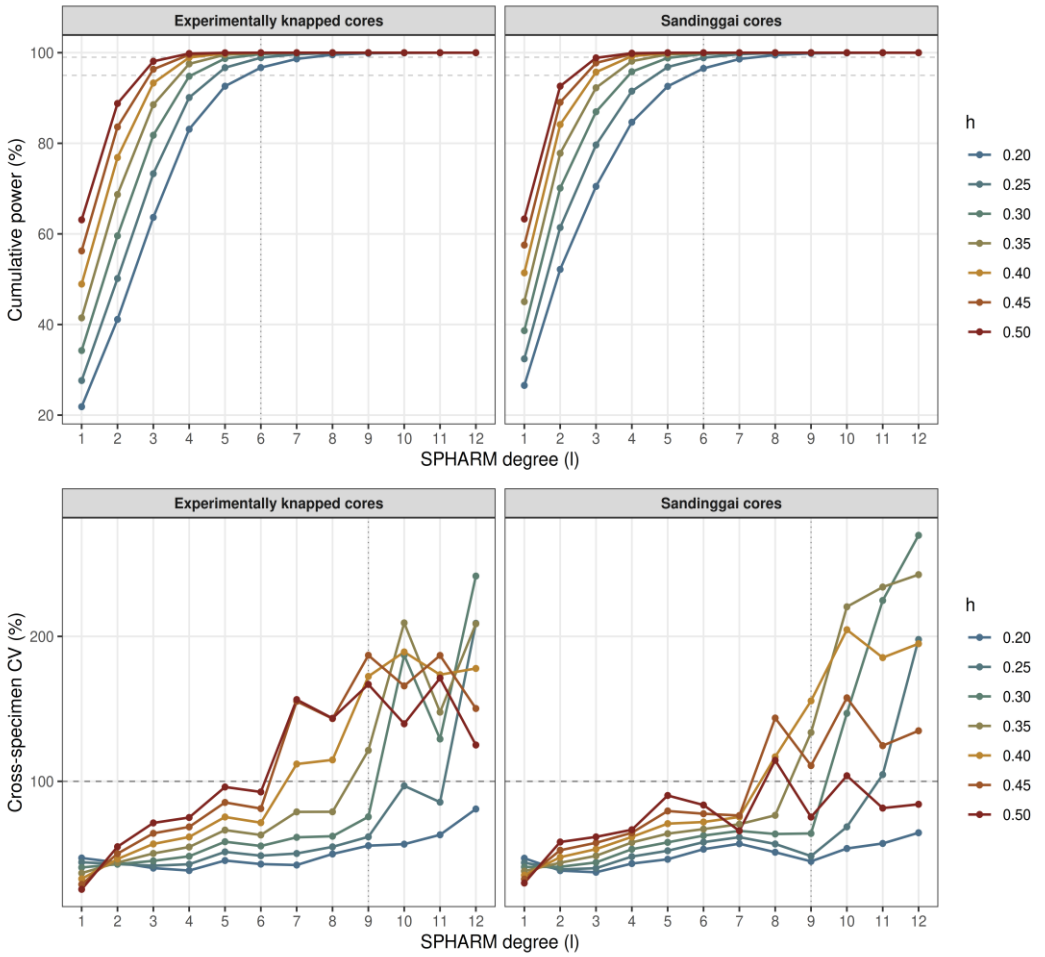

Figure S1 SP-SPHARM cumulative power (top) and cross specimen CV (bottom) by spherical harmonic degree l, overlaid across bandwidths h, for the experimental (EXP) and archaeological (SDG) assemblages. Dashed lines mark the 95%/99% power and 100% CV references.

The pairwise standardised Euclidean distance structure is highly stable: each h's distance matrix correlates with the $h = 0.35$ matrix at Pearson $r \geq 0.961$, and key separations persist at every h (e.g. discoidal vs Levallois).

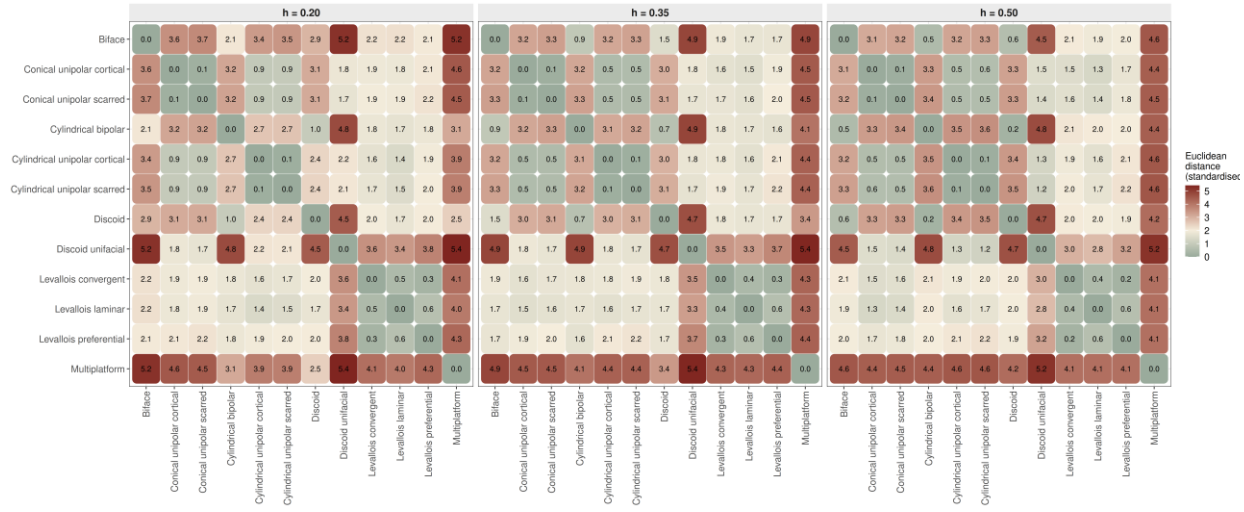

Figure S2 Idealized core standardised Euclidean distance matrices at $h = 0.20$, 0.35 and 0.50.

The scar pattern PERMANOVA by experimental core type stays significant across h ($R^2 = 0.298$–0.329; $p < 0.05$ throughout). The pairwise resolution profile is stable: 7 of 10 typology pairs are resolved at the three smallest bandwidths ($h = 0.20$–0.30) and 8 of 10 from $h = 0.35$ to 0.50, including the chosen $h = 0.35$), dropping to 6/10 at the two smallest bandwidths and 7/10 at the largest. The association between morphology and scar patterning is negligible in magnitude at every bandwidth: Mantel $r \approx 0$ throughout (EXP -0.062 to -0.008; SDG 0.001 to 0.017) and RV ≈ 0.10–0.12 (EXP) and ≈ 0.09 (SDG). For SDG both measures are non-significant across the entire range. For the experimental assemblage the Mantel test is non-significant at every h, and the RV coefficient is non-significant from $h = 0.20$ to 0.40, becomes marginal at $h = 0.45$ ($p = 0.078$), and crosses $p < 0.05$ only at the most over smoothed setting ($h = 0.50$: RV = 0.117, $p = 0.044$), where the association is still negligible and is not corroborated by the Mantel test ($r = -0.008$, $p = 0.54$). Read on effect size, the decoupling is robust across the full range, and the chosen $h = 0.35$ sits well inside the non-significant regime. The SDG scar–core type PERMANOVA remains significant throughout ($R^2 = 0.166$–0.177; $p < 0.05$).

Table S1 Per bandwidth sensitivity metrics. The asterisk marks the main analysis setting (h = 0.35). Reported metrics include cumulative power, coefficient of variation (CV), idealized-model (IM) correlation, R^2, and Mantel/RV estimates.

h	κ	EXP cum%	CV>100%	IM corr vs .35	EXP R^2	EXP p	EXP sig pairs	SDG type R^2	SDG p	EXP Mantel r (p)	EXP RV (p)	SDG Mantel r (p)	SDG RV (p)
0.20	25.00	96.71	14	0.961	0.298	0.001	7/10	0.166	0.039	-0.058 (0.789)	0.100 (0.128)	0.017 (0.396)	0.093 (0.290)
0.25	16.00	98.90	12	0.984	0.299	0.001	7/10	0.167	0.038	-0.061 (0.800)	0.100 (0.128)	0.016 (0.401)	0.093 (0.297)
0.30	11.11	99.69	10	0.996	0.300	0.001	7/10	0.167	0.038	-0.062 (0.805)	0.100 (0.126)	0.015 (0.410)	0.092 (0.305)
0.35*	8.16	99.92	9	1.000	0.302	0.001	8/10	0.168	0.038	-0.061 (0.801)	0.101 (0.120)	0.012 (0.425)	0.091 (0.313)
0.40	6.25	99.98	7	0.997	0.305	0.001	8/10	0.169	0.037	-0.055 (0.777)	0.103 (0.106)	0.008 (0.444)	0.090 (0.316)
0.45	4.94	99.99	7	0.990	0.313	0.001	8/10	0.171	0.035	-0.039 (0.704)	0.108 (0.078)	0.005 (0.464)	0.090 (0.307)
0.50	4.00	99.99	7	0.980	0.329	0.001	8/10	0.177	0.029	-0.008 (0.539)	0.117 (0.044)	0.001 (0.485)	0.091 (0.304)

Flaking scar size threshold

To test whether the results depend on flaking scar size, and in particular whether they are driven by small, less reliably measured scars, we recomputed the SP-SPHARM power spectra after dropping scars at or below increasing cutoffs $T \in \{2, 5, 10\}$ mm (retaining length > T), holding everything else fixed (alignment, $h = 0.35$, the evaluation and transform grids, $l_{\max} = 20$, normalisation, seeds). The cutoff $T = 2$ was used as the main analysis, at which the recomputed descriptors reproduced the saved descriptor values used in the main analysis. The >5 mm rule retains 95.9% (EXP) and 99.2% (SDG) of scars and >10 mm retains 80.7% and 88.5%. The threshold is not applied to the idealized cores, whose scar lengths are fixed values, so their discriminability is unchanged by construction.

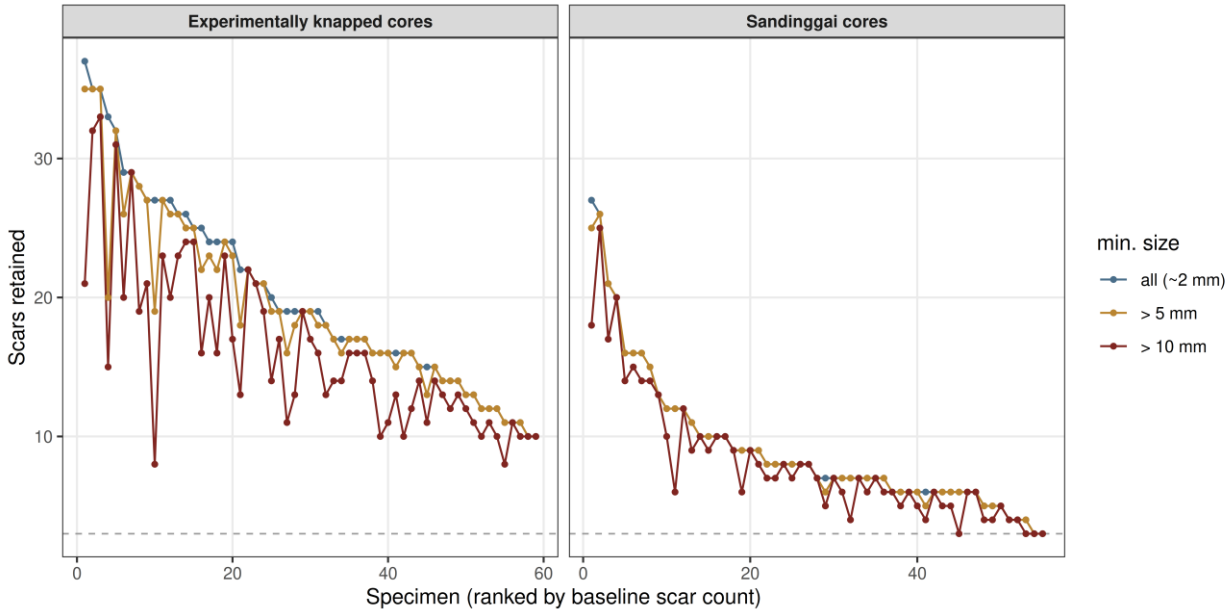

Figure S3 Per specimen scar retention by size threshold T; the dashed line is the three-scar advisory floor, below which no specimen falls.

SP-SPHARM power saturates early at every threshold: cumulative power through $l = 6$ is 99.91–99.92%, and the cross-specimen CV first exceeds 100% at degree $l = 9$ for all cutoff value. The $l = 1$–6 truncation is robust to the inclusion rule.

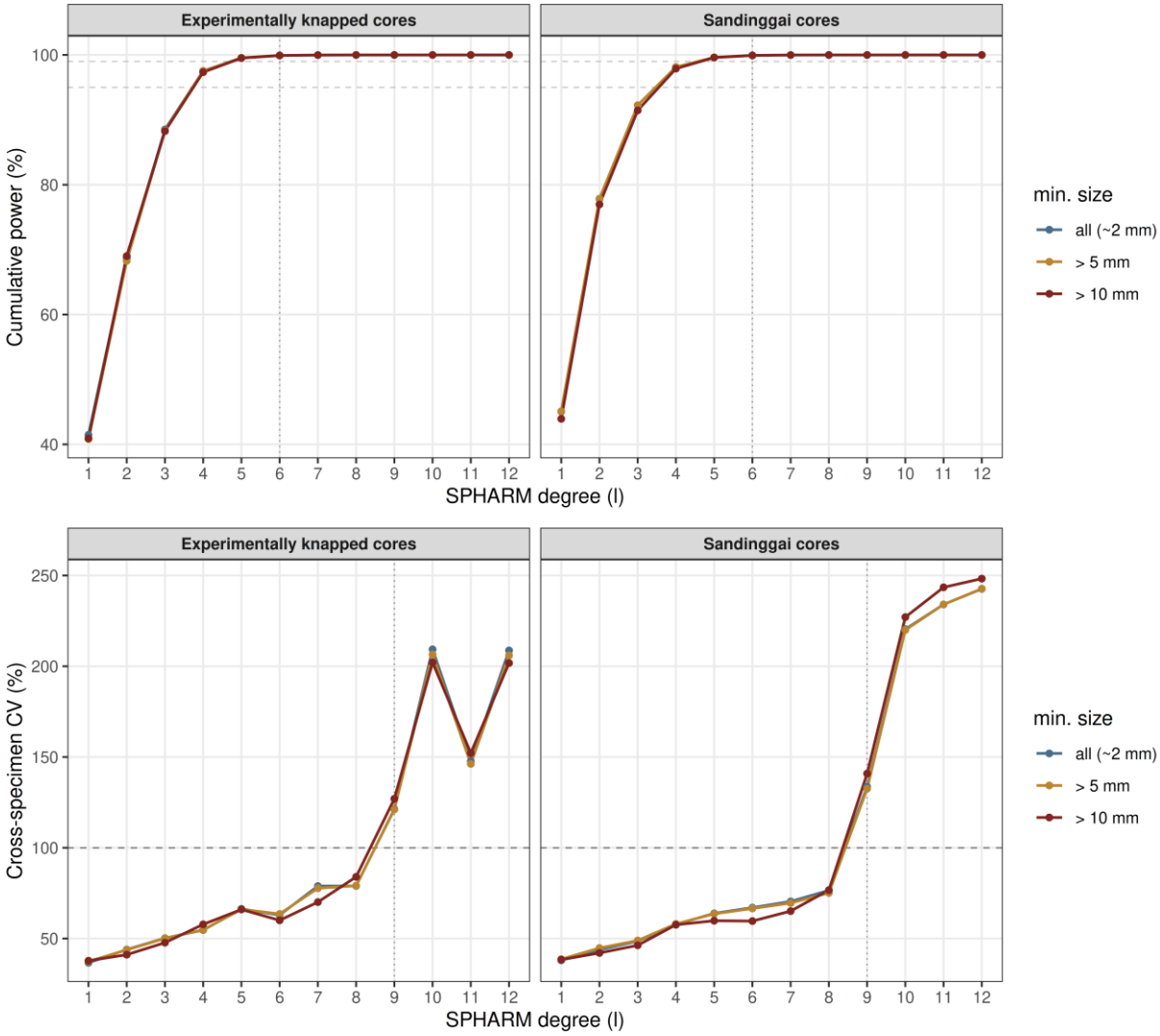

Figure S4 SP-SPHARM power saturation and CV by degree across scar size thresholds T, for experimentally knapped cores and SDG cores.

The scar pattern PERMANOVA by core type is significant at every threshold (R^2 = 0.282–0.322; $p < 0.05$ throughout). The pairwise resolution profile is 8/10 typology pairs at the main analysis baseline, 7/10 at >5 mm, and 8/10 at >10 mm. Restricting the analysis to larger, more reliably measured scars therefore does not weaken type discrimination; the experimental core-type signal is carried by the larger scars and is not an artifact of small scars.

Morphology and scar patterning remain negligibly associated and non-significant at every threshold: Mantel $r \approx 0$ (EXP -0.061 to 0.004; SDG -0.047 to 0.012) and RV = 0.101–0.109 (EXP) and 0.057–0.091 (SDG), all $p > 0.05$. The SDG scar–core type PERMANOVA is significant throughout (R^2 = 0.168–0.171).

Table S2 Per threshold sensitivity metrics. The asterisk marks the main analysis baseline, in which > 2 mm scars are retained. The EXP and SDG scar columns report the total number of retained scars.

T	EXP scars	SDG scars	EXP cum%	CV>100%	EXP R^2	EXP p	EXP sig pairs	SDG type R^2	SDG p	EXP Mantel r (p)	EXP RV (p)	SDG Mantel r (p)	SDG RV (p)
2*	1185	503	99.92	9	0.302	0.001	8/10	0.168	0.038	-0.061 (0.801)	0.101 (0.120)	0.012 (0.425)	0.091 (0.313)
5	1136	499	99.92	9	0.282	0.001	7/10	0.170	0.036	-0.049 (0.751)	0.104 (0.098)	0.008 (0.446)	0.087 (0.358)
10	956	445	99.91	9	0.322	0.001	9/10	0.171	0.037	0.004 (0.483)	0.109 (0.082)	-0.047 (0.756)	0.057 (0.827)

Mesh decimation and Laplacian smoothing

Core morphology is decimated by quadric edge collapse to a fixed target face count (main analysis = 20,000) and Laplacian-smoothed (main analysis = 3 iterations) before the spherical harmonic expansion. We recomputed the M-SPHARM power spectra over two one-dimensional sweeps through the main analysis setting: decimation target ∈ {10000, 20000, 50000} faces (at production smoothing) and smoothing ∈ {0, 3, 6} iterations (at the production target), holding everything else fixed (the quadric-decimation algorithm, mesh normalisation, deterministic area-weighted PCA alignment, the spherical sampling grid, l_{max} = 20, normalisation, seeds).

M-SPHARM power concentrates by l = 8 (cumulative power 97.98–98.58% EXP, 97.75–98.59% SDG) and the cross-specimen CV stays below the 90% reliability line through l = 8 (maximum CV 82.5–88.6% EXP, 54.6–69.6% SDG). The l = 1–8 truncation is robust to decimation and smoothing.

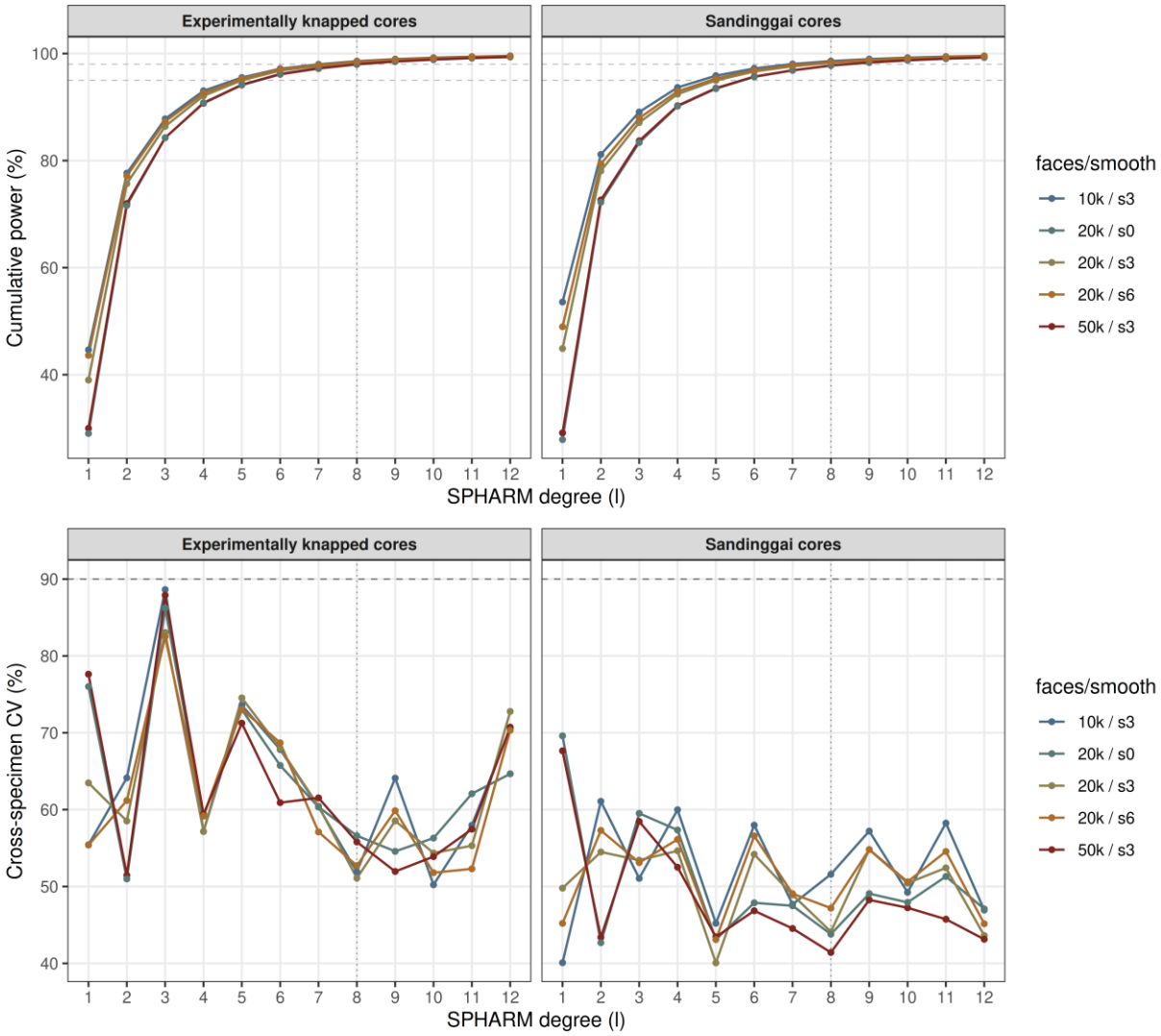

Figure S5 M-SPHARM power concentration and CV by degree across mesh-preprocessing settings, for EXP and SDG; dashed line = 90% reliability.

The experimental morphology PERMANOVA by core type is weak but consistently significant (R^2 = 0.121–0.149; 0 of 10 typology pairs resolved). In the experimental assemblage, type information is therefore carried far more strongly by the scar pattern descriptor (which resolves 8/10 pairs; Table S 2) than by morphology, evidence that the two descriptors are complementary rather than redundant. Core type is the dominant, significant morphology term across all settings (R^2 = 0.178–0.216; $p < 0.05$ throughout), while raw material is small (R^2 = 0.013–0.051; significant only at the heaviest smoothing) and layer is negligible (R^2 = 0.007–0.030; non-significant throughout). The "core type dominant, technological continuity across layers and raw materials" picture holds across all settings.

With the perturbed morphology and the fixed scar descriptor, the association between morphology and scar patterning is negligible in magnitude at every setting: Mantel $r \approx 0$ (EXP -0.106 to -0.048; SDG -0.058 to 0.102) and RV ≈ 0.10 (EXP 0.095–0.116; SDG 0.072–0.093). The Mantel test is non-significant throughout (EXP and SDG), and the RV is non-significant at every setting except the unsmoothed mesh (0 smoothing iterations: EXP RV = 0.116, p = 0.043) — where, as for the most over smoothed bandwidth in Table S1, the association is still negligible and is not corroborated by the Mantel test (r = -0.069, p = 0.83). Read on effect size, the decoupling holds under perturbation of the morphology descriptor as well as the scar descriptor (Table S1, Table S2).

Table S3 Per setting sensitivity metrics. The asterisk marks the main analysis setting, defined as 20,000 faces and three Laplacian smoothing iterations.

faces	smooth	EXP cum%	EXP CV_{max}	SDG cum%	EXP morph R^2 (p)	SDG type R^2 (p)	SDG rawmat R^2 (p)	SDG layer R^2 (p)	EXP Mantel r (p)	EXP RV (p)	SDG Mantel r (p)	SDG RV (p)
10000	3	98.58	88.6	98.59	0.149 (0.005)	0.185 (0.013)	0.031 (0.164)	0.030 (0.201)	-0.064 (0.800)	0.099 (0.137)	-0.058 (0.776)	0.072 (0.613)
20000*	3	98.37	83.0	98.29	0.123 (0.025)	0.188 (0.011)	0.041 (0.083)	0.018 (0.484)	-0.061 (0.801)	0.101 (0.120)	0.012 (0.425)	0.091 (0.313)
50000	3	97.98	87.9	97.76	0.149 (0.004)	0.193 (0.009)	0.048 (0.056)	0.014 (0.619)	-0.106 (0.933)	0.100 (0.118)	0.102 (0.098)	0.093 (0.266)
20000	0	98.06	86.2	97.75	0.138 (0.004)	0.216 (0.003)	0.013 (0.624)	0.007 (0.874)	-0.069 (0.828)	0.116 (0.043)	0.071 (0.167)	0.090 (0.301)
20000	6	98.49	82.5	98.41	0.121 (0.025)	0.178 (0.018)	0.051 (0.041)	0.023 (0.334)	-0.048 (0.735)	0.095 (0.182)	-0.002 (0.496)	0.085 (0.393)

Summary

Across both descriptors, every conclusion the analysis relies on is robust to its principal free parameters: early truncation (l = 1–6 for the scar descriptor, l = 1–8 for morphology); idealized core discrimination; the dominance of core type over raw material and layer in the archaeological morphology and scar patterning; and the decoupling of morphology and scar patterning, which holds under perturbation of either descriptor.

Statistical power of the morphology and scar patterning covariation test

The tests of covariation between core morphology and scar patterning returned null results in both assemblages (main text; Table S1, Table S2, Table S3). Because a non-significant result does not by itself distinguish a genuine absence of association from insufficient power to detect one, we assessed the informativeness of these nulls directly. Rather than computing observed power from the realized effect, which is uninformative (Hoenig & Heisey, 2001), we followed the recommendation to characterize the range of effect sizes a fixed design can resolve (Lakens, 2022). Three quantities were computed for each assemblage, reusing the same ILR transformed power spectra as the main analysis. At the main analysis setting the recomputed coefficients reproduced the reported raw RV and Mantel statistics, confirming that the analysis was anchored to the values in the main text.

We evaluated the null results in three complementary ways. First, because raw RV coefficients can be upwardly biased when the number of variables is large relative to sample size (Adams, 2016; Smilde et al., 2009), we recalculated the association using the bias-corrected modified RV (Smilde et al., 2009). Second, we used a percentile bootstrap over specimens (2,000 resamples) to estimate how large the association might plausibly be, while keeping each specimen's morphology and scar pattern coordinates paired. Third, we used Monte Carlo simulation to estimate how sensitive the Mantel and RV permutation tests were at the realized sample sizes (Manly, 2007; Rohlf & Corti, 2000). In these simulations, data were generated with a known degree of association between the two blocks, following established practice in morphometric studies of modularity and integration (Adams, 2016). Variation within each block was drawn from the empirical mean and covariance of that block, and a shared component was added to create a controlled level of between-block association. The strength of this component was calibrated to a target population RV, by which we mean the underlying between-block association imposed in the simulated population, rather than the RV estimated from any single finite sample. Because the simulated association was aligned with the leading axis of each block, the resulting minimum detectable effects should be understood as best-case values: a less favorably structured association would require a larger effect to be detected (Lakens, 2022). Simulation settings were kept modest, and all procedures used fixed seeds.

These checks suggest that the observed decoupling between morphology and scar patterning is unlikely to be explained by sample size alone. After bias correction, the modified RV fell from 0.101 to 0.038 in the EXP assemblage and from 0.091 to 0.011 in the SDG assemblage. This indicates that much of the already small raw association was attributable to small-sample inflation rather than to a strong underlying signal. The bootstrap results also placed only modest upper bounds on the possible association: approximately RV = 0.19 for EXP and 0.20 for SDG, and Mantel $r = 0.13$ for EXP and 0.22 for SDG. The simulation results showed that the RV permutation test had sufficient power to detect a moderate association at the realized sample sizes. It reached at least 80% power at a population RV of 0.072 in EXP and 0.079 in SDG, and exceeded 96% power by a population RV of about 0.13. The Mantel test was less sensitive, reaching 80% power only at a population RV of 0.148 in EXP and 0.191 in SDG, consistent with the lower power of Mantel tests relative to covariance-based statistics. Power curves are shown in Figure S6, and all values are summarized in Table S4.

Taken together, these results support the interpretation that core morphology and scar patterning are largely decoupled at the assemblage level. If a moderate or stronger association had been present, it would probably have been detected by the RV test. Instead, the bias-corrected estimates were close to zero in both assemblages. A very weak residual association, below a population RV of roughly 0.07–0.08, cannot be excluded. The small per-type subsamples also remain underpowered, which is why the within-type analyses are treated as exploratory in the main text.

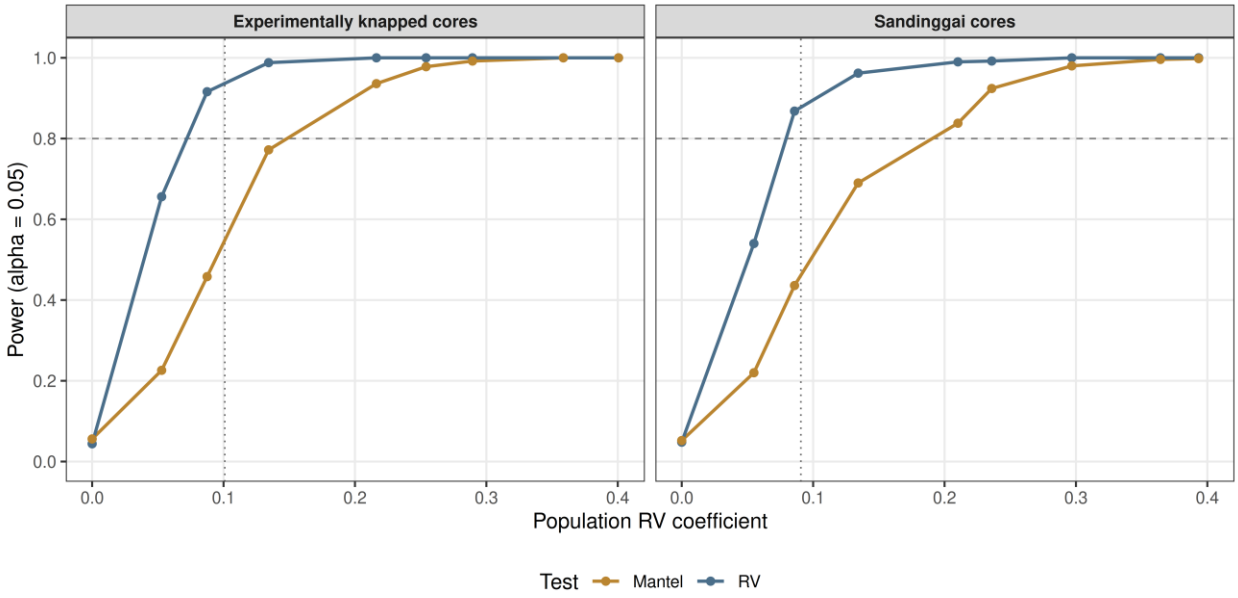

Figure S6 Monte Carlo power of the RV permutation test and the Mantel test as a function of the population RV coefficient, at the realized sample sizes of the experimental and Sandinggai assemblages. The dashed horizontal line marks 80% power and the dotted vertical line marks the observed raw RV.

Table S4 Power and sensitivity metrics for the morphology and scar patterning covariation test, by assemblage. Raw RV and Mantel r are the values reported in the main text (anchor). RV2 is the bias-corrected modified RV coefficient. The 95% confidence intervals are percentile bootstrap intervals (2,000 resamples) and are interpreted as upper bounds on the association (see text). MDES is the minimum population effect detectable at 80% power, estimated by Monte Carlo simulation, and is a lower bound. Values in parentheses are permutation p values.

Assemblage	n	Raw RV (p)	RV2	RV2 95% CI	Mantel r (p)	Mantel r 95% CI	MDES RV	MDES Mantel
EXP	58	0.101 (0.120)	0.038	[0.030, 0.190]	-0.061 (0.806)	[-0.123, 0.127]	0.072	0.148
SDG	51	0.091 (0.313)	0.011	[0.011, 0.198]	0.012 (0.425)	[-0.062, 0.220]	0.079	0.191

References

Adams, D. C. (2016). Evaluating modularity in morphometric data: Challenges with the RV coefficient and a new test measure. Methods in Ecology and Evolution, 7(5), 565–572. https://doi.org/10.1111/2041-210X.12511

Hamby, D. M. (1994). A review of techniques for parameter sensitivity analysis of environmental models. Environmental Monitoring and Assessment, 32(2), 135–154. https://doi.org/10.1007/BF00547132

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. The American Statistician, 55(1), 19–24. https://doi.org/10.1198/000313001300339897

Lakens, D. (2022). Sample size justification. Collabra: Psychology, 8(1), 33267. https://doi.org/10.1525/collabra.33267

Manly, B. F. J. (2007). Randomization, bootstrap and monte carlo methods in biology (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Rohlf, F. J., & Corti, M. (2000). Use of two-block partial least-squares to study covariation in shape. Systematic Biology, 49(4), 740–753. https://doi.org/10.1080/106351500750049806

Saltelli, A. (2002). Sensitivity Analysis for Importance Assessment. Risk Analysis, 22(3), 579–590. https://doi.org/10.1111/0272-4332.00040

Smilde, A. K., Kiers, H. A. L., Bijlsma, S., Rubingh, C. M., & van Erk, M. J. (2009). Matrix correlations for high-dimensional data: The modified RV-coefficient. Bioinformatics, 25(3), 401–405. https://doi.org/10.1093/bioinformatics/btn634