

VESTIGIAL ANATOMY AND THE BURDEN OF PROOF

A Synthesis Review of Four Independent Critiques of “Vestigial Anatomy in a Technologically Modified Selective Environment”

Document ID	VABP-2026-001
Version	1.1 (amended house-style reissue)
Date	2026-09-11
Status	Review — demo specimen for review-site demonstration
Author of reviewed paper	Kevin Alexander
Review basis	Four independent LLM critiques (Deepseek, Claude Incognito, Qwen, fourth reviewer) + synthesis
Mode	Thinking-out-loud; deliverable at author’s request
Classification	Deliberately non-viable paper under review; review functions as the demonstration

Version Notes

- v1.0 (2026-09-11) — Initial synthesis, delivered in chat; structure ported from the Megalithic Archaeology and the Burden of Proof document.
- v1.1 (2026-09-11) — House-style reissue per Thread Continuity Brief §2: metadata table, version notes, standardized tier tags, new Section 10 (Honest Assessment of This Review), correction log, versioned footer. Content unchanged.
- v1.1 amended (2026-09-11) — Removed megalithic-content carryover at author flag: vase-replication disconfirmation bullet, inundation-season analogy, Scheuring counterweight analogy, Houdin-correction note, and one ‘megalithic context’ framing phrase. All replaced with domain-internal equivalents. See Correction Log.

Contents

A Note on Sources

What This Document Does and Does Not Claim

The Standing Methodological Principle

The Plasticity Corollary

The Logical Structure

The Three-Tier Critique Structure

Summary of Hard Constraints — Tier 1 Checklist

A Note on Case Selection

Section 1 — The Logical Structure Constraint (1.1–1.5)

Section 2 — The Extrapolation Constraint (2.1–2.3)

Section 3 — The Combined Overreach Pattern (Overreaches 1–8, 3.9)

Section 4 — What the Paper Actually Shows

Section 5 — What This Review Establishes

Section 6 — Conclusion

Section 7 — What Would Disconfirm This Review

Section 8 — Investigative Methodology: How the Impasse Might Actually Be Resolved (8.1–8.8)

Section 9 — Next Phase: Application to Other Evolutionary Claims

Section 10 — Honest Assessment of This Review

Correction Log

A Note on Sources

Four independent critiques of the paper were solicited and analyzed: Deepseek, Claude Incognito, Qwen, and a fourth reviewer. No two shared a drafting process. All four arrived, independently, at the same central complaint — that the paper’s central hypothesis is structured to accommodate every possible observation, and therefore predicts none. Three of the four independently identified the lactase-persistence overgeneralization. Three of the four independently identified that the candidate structures do not form a coherent category.

When four independent analyses converge on the same structural complaint, the complaint is structural. This document synthesizes that convergence, adds four points not raised by any of the four, and organizes the whole into a single framework.

Continued argument without adequate scientific testing serves neither side. The evidence should decide.

What This Document Does and Does Not Claim

This document presents a two-part review. Part one establishes a set of logical and evidentiary standards — derived from the philosophy of scientific testing, population genetics, and evolutionary biology — that any evolutionary hypothesis must meet to be considered testable rather than merely interpretive. These standards apply equally to the paper under review and to this review itself. Measured against these standards, the paper’s central hypothesis does not currently account for its own predictions at several tested points: no disconfirming observation is specified, no risky prediction is committed to, and the strongest available evidence (plasticity) is not the evidence the thesis requires (genetic evolution). Whether that gap can be closed by revision or reflects a genuine limitation of the hypothesis is left open.

This document does not claim the transitional-structure hypothesis is false. It does not claim the conventional retrospective interpretation of vestigial anatomy is complete. It does not claim that technology has not altered human evolution — the lactase-persistence case demonstrates that it has, in at least one well-documented instance. It claims only that, like any account, the transitional-structure hypothesis must meet the burden of proof its own framing implies — and that at several specific, tested points, the predictions required for that proof have not yet been stated, let alone met. The same standard is applied throughout to the retrospective interpretation this review defends as the null.

This review is not neutral in the sense of withholding criticism from the paper. It is neutral in the methodological sense that the same evidentiary requirements are applied to the hypothesis under review and to the competing account. Identifying an unresolved constraint in the transitional model does not establish the retrospective model; it establishes only that the transitional model remains incompletely demonstrated.

The Standing Methodological Principle

Any evolutionary hypothesis must specify at least one observation that would count against it. Whatever outcomes a hypothesis permits, it must render some class of observations less likely under it than under its rivals. A hypothesis compatible with decline, persistence, variability, stability, repurposing, disappearance, and no-change-at-all renders no observation less likely — and a hypothesis that forbids nothing predicts nothing. This test applies equally to the transitional-structure hypothesis, to the retrospective interpretation it proposes to supplement, and to any future alternative. The paper under review nowhere identifies a single observation that would disconfirm its central claim.

The Plasticity Corollary

The Standing Methodological Principle applies equally to mechanisms. Any claimed anatomical response to technology must distinguish genetic evolution (heritable allele-frequency change) from developmental plasticity (environmentally induced, non-heritable change within an organism's lifetime). Plastic responses to technologically modified environments — craniofacial and mandibular reduction under soft, processed diets being the best-documented example — are real, fast, well-evidenced, and leave no allele-frequency trace. A plastic response therefore cannot support a claim about evolutionary trajectories, however impressive the anatomical change. Until a claimed response has been separated from plasticity by design, plasticity should be treated as the competing hypothesis to be defeated, not as one confound in a list. A negative plasticity control is the equivalent of a blank sample in an assay: without it, no positive reading is interpretable.

The Logical Structure

The argument throughout this document follows a single logical structure: IF the transitional-structure hypothesis is correct, THEN specific observable consequences must follow — for example, prevalence variability across structures should correlate

with the degree to which technology has substituted each structure’s ancestral function, and structures under recent relaxed constraint should show genomic signatures of constraint decay. Currently, those consequences have not been stated as predictions, let alone observed. THEREFORE the hypothesis, as currently formulated, has not met the burden of proof its own logic would generate. This is not an argument from ignorance. It is an argument from the requirements of scientific testing, applied consistently to what the hypothesis proposes — and applicable with the same force to the retrospective account it seeks to supplement.

The Three-Tier Critique Structure

The critiques in this document operate at three distinct levels of strength. Each section states its tier at the outset.

Tier	Type	Examples
Tier 1	Structural — hard constraints from the logic of testing.	Unfalsifiability of the transitional claim; stasis counted as transitional; compatibility with all rival explanations; missing stop criterion; asymmetric evidential standard; temporal misalignment
Tier 2	Evidentially constrained but conditional on an interpretive premise.	Lactase-to-anatomy extrapolation; variability treated as diagnostic; category error across candidate structures; contested single-paper dependency
Tier 3	Disproportionality and omission arguments — explicitly OPEN.	Missing risky prediction; missing longitudinal engagement; missing adjacent literatures; citation hygiene; detection-asymmetry

		feasibility
--	--	-------------

Summary of Hard Constraints — Tier 1 Checklist

Constraint	Type	What It Contradicts
Stasis counted as a transitional state	Logical — definitional	Any possible falsification of the ‘transitional’ claim
Outcome-list immunization (decline, persistence, variability, repurposing, stability all compatible)	Logical — probabilistic	The hypothesis having predictive content
Compatibility with every rival mechanism (Sec. 9 of the paper)	Logical — identifiability	The hypothesis being distinguishable from drift, relaxed selection, exaptation
No stop criterion for ‘transitional’	Logical — operational	Any test period after which ‘still transitional’ could be rejected
Temporal misalignment (deep-time variation vs. recent technology)	Chronological	Recent technological niche construction as the trigger of observed variation
Category error across candidate structures	Biological	Treating appendix, palmaris longus, plantaris, and third molars as instances of one phenomenon
Asymmetric evidential standard (paper’s Sec. 8)	Methodological	Absence of a soft-tissue record counting against stability but not against transition
Plasticity not separated	Mechanistic	Claimed evolutionary

from genetic evolution		trajectories being interpretable as evidence
------------------------	--	--

A Note on Case Selection

This review takes the paper’s four candidate structures — the vermiform appendix, the palmaris longus, the plantaris, and the third molars — as its primary worked examples throughout Sections 1 and 2, together with the lactase-persistence case the paper itself relies upon. This is a choice of documentation, not a verdict. These are the most thoroughly discussed ‘vestigial’ structures in the literature, which makes them the strongest available test cases for the standards proposed here — not because the conventional retrospective interpretation of them is considered secure. The retrospective interpretation may ultimately prove correct in every particular, and this document does not claim otherwise. Its claim is narrower: that specific, documented logical and evidential constraints, outlined in the sections that follow, have not yet been met by the transitional hypothesis at several identified points. Whether future revision resolves those gaps in favor of the hypothesis, a narrowed hypothesis, or its abandonment is an open question this document does not attempt to answer.

Section 1 — The Logical Structure Constraint

Tier structure: 1.1 Tier 1 (logic). 1.2 Tier 1 directional (inferential mass balance). 1.3 Tier 2 (plasticity dependency chain). 1.4 Tier 3 (open — pending the comparative test proposed in Section 8.1). 1.5 Tier 1 (methodological principle applies to the novel framework itself).

1.1 — The Unfalsifiability Constraint [Tier 1]

[TIER 1 — STRUCTURAL]

A falsifiable hypothesis must forbid some observation. The paper’s central claim — that some vestigial structures ‘may be interpreted’ as transitional — is hedged with ‘may,’ ‘need not,’ and ‘not exclusively,’ and its Section 6 lists every possible fate of a structure (decline, variability, persistence, secondary function, integration, repurposing, stability) as compatible with the model.

When stasis itself is counted as a transitional state, the definition is tautological: change confirms the hypothesis, and no-change confirms it. When every rival

mechanism — relaxed selection, drift, developmental constraint, retained secondary function, exaptation — is explicitly declared compatible with the model (the paper's Section 9), no dataset can favor the hypothesis over those mechanisms, because it does not compete with them; it relabels their outputs. A hypothesis that cannot be distinguished from its alternatives by any evidence is not making an independent empirical claim.

Note: this argument applies specifically to the hypothesis as formulated. A narrower formulation — one structure, one dated selection change, one committed prediction — would not face this problem. That narrower formulation is what Section 8 proposes to extract from the paper's own logic.

1.2 — The Analogy Mass Balance: One Demonstrated Case, Many Extrapolated Ones [Tier 1 — directional]

[TIER 1 — DIRECTIONAL]

The paper's only case with direct population-genetic evidence — lactase persistence — is a single-regulatory-variant trait under positive selection with a well-documented sweep, a dated cultural trigger (dairying), and a geographic correlation. From this one case the paper licenses the inference 'behavior → modified environment → altered selection → biological adaptation' (its Section 3) and applies that license to four anatomical structures — none of which has comparable evidence tying its variation to technological selection specifically.

The hedging defense, at its strongest, holds that the paper only proposes the transitional reading as a possibility, not a finding. This does not resolve the mass-balance problem, because hedging extends how long a claim remains defensible; it does not recover the inferential content consumed in the extrapolation. The fraction of the argument actually supported by demonstrated cases — one enzymatic trait — is permanently lost into a thesis covering all variable anatomy, independent of how carefully the surviving claims are phrased. No version of the 'we only say may' defense restores this: the unsupported share of the thesis is the mass-balance problem, independent of how the supported share is managed.

Converting this directional observation into a demonstrated one requires the paper to state, per candidate structure, the specific genetic or archaeological evidence that would tie its variation to a dated technological transition. None is currently stated.

1.3 — The Two-Stage Explanation Problem and Plasticity Dependency [Tier 2]

Any claim that a structure is evolutionarily transitional requires two separable stages: first, establishing that the observed anatomical pattern is genetic rather than plastic; second, establishing that the genetic component reflects selection (relaxed or directional) associated with a technological change rather than drift, population history, or ancient constraint. The paper's best anatomical evidence — craniofacial and dental change associated with food processing — sits almost entirely in stage one, where it is well documented but non-heritable. Its thesis, however, lives in stage two.

The plasticity dependency chain: you cannot claim an evolutionary trajectory without first excluding plasticity, and you cannot exclude plasticity with the paper's current evidence inventory. The conventional tool inventory of the paper — cross-sectional prevalence variation — has no solution to this bootstrapping problem at all. The dependency applies specifically to heritable change, not to anatomical change generally: soft-food-driven jaw reduction within individuals' lifetimes is real, well-documented change that requires no allele-frequency change whatsoever and therefore does no work for the thesis.

1.4 — The Missing-Risky-Prediction Problem [Tier 3 — open]

[TIER 3 — OPEN]

The paper's own logic generates a comparative, ranked, testable prediction that the paper never states: candidate structures should rank in prevalence variability and agenesis frequency according to the degree to which technology has substituted their ancestral function (grip mechanics for palmaris and plantaris; mastication for third molars; no clear substitution for the appendix). Under the transitional model, variability should track degree of substitution; under drift, it should track mutational target size and developmental-genetic architecture. This is treated here as an open investigative question rather than a settled constraint: Section 8.1 proposes a formal comparative study to establish what the hypothesis does and does not predict.

1.5 — The Framework Novelty Problem [Tier 1 — methodological]

The hypothesis is sometimes defended as a new conceptual framework for vestigial anatomy. The framework is logically available. Its novelty implications are not — it shifts rather than resolves the explanatory burden.

IF relaxed selection, drift, developmental constraint, and exaptation already explain the observed phenomena (as the paper's own Section 9 concedes), THEN a new label for

their outputs must add either new predictions or new evidence. The transitional framework adds a technological attribution but no new prediction and no new evidence. Relaxed selection is relaxed selection whether the environment was modified by a Pleistocene glacier or a modern tractor. The framework does not apply to the hard cases it is asked to carry — it provides no criterion for identifying which structures are transitional, no stop rule, and no disconfirming observation.

Element	Conventional account (relaxed selection / drift / exaptation)	Transitional-structure account
Mechanism	Standard population genetics, established	Same mechanisms, relabeled as responses to technology
Trigger	Unspecified environmental change	Technological niche construction — undated per structure
Predictions	Variable prevalence under weak constraint; no ranked expectation	None stated beyond ‘anything possible’
Evidence required	Already consistent with observed variation	Same evidence, plus a dated technological correlation per structure — not supplied
Evidence present	Consistent with cross-sectional variation	No novel prediction or dated causal link demonstrated

Section 2 — The Extrapolation Constraint

Tier structure: 2.1 Tier 1 directional (inferential arithmetic). 2.2 Tier 1 (category coordination). 2.3 Tier 1 (chronological alignment).

2.1 — The Case-Carrying Rate [Tier 1 — directional]

[TIER 1 — DIRECTIONAL]

One demonstrated case (lactase persistence) is asked to carry four candidate structures, plus an open-ended class of future candidates. There is no production quota to divide — but the inferential arithmetic has the same shape as any supply problem: the breadth of the thesis grows while the demonstrated base stays fixed at one. The more structures the ‘transitional’ label covers, the thinner the evidential coverage per structure.

Chronology-compounding analogy: the paper’s own Section 12 concedes that modern technology has existed for too few generations to observe outcomes — the way a shortened working window sharpens a throughput problem rather than solving it. The concession does not soften the constraint; it compounds it. A hypothesis whose central mechanism (recent relaxation) cannot yet have produced observable outcomes, invoked to explain variation that is demonstrably ancient, must carry the whole inferential load on chronology alone.

2.2 — The Multi-Structure Category Problem [Tier 1]

[TIER 1 — STRUCTURAL]

IF four structures of different tissue type (lymphoid-gut, skeletal muscle, skeletal muscle, dental), different developmental origin, different genetic architecture, and different evolutionary history are treated as instances of one phenomenon, THEN interpretive error compounds across categories the way geometric error compounds across uncoordinated work faces. The paper acknowledges this in its Section 12 — and then continues to reason as if the category held.

The rhetorical-accumulation fallacy: a transport system can accelerate a load’s lift without accelerating its final placement — twenty fast lifts are not twenty completed placements. The same fallacy operates here at the level of evidence: stacking four individually hedged examples creates the impression of a building evidential base, when the examples are not independent confirmations of one mechanism but four different unresolved causal stories. A count of inconclusive cases is not cumulative evidence, in exactly the way that twenty fast lifts are not twenty completed placements.

2.3 — The Temporal Misalignment Constraint [Tier 1]

[TIER 1 — STRUCTURAL]

The variation the hypothesis invokes — palmaris longus absence rates, third-molar agenesis frequencies, plantaris variability — is deep-time in origin; the technological changes the hypothesis proposes as its trigger are, in the paper’s own words, ‘only an instant on an evolutionary timescale.’ A trigger cannot explain variation that substantially predates it. Correction note: this document takes no position on when the variation arose, and does not rule out that some component is recent — only that the trigger must be dated before it can be invoked, and the paper dates none of its proposed triggers to the origins of the variation they are asked to explain.

Section 3 — The Combined Overreach Pattern [All Tier 3 — explicitly OPEN]

All eight items in this section are Tier 3: explicitly OPEN. Tier 1 arguments in Sections 1 and 2 stand independently of these. The combined pattern across eight independent omission categories is analytically significant — each individual omission might have an innocent explanation, but all eight simultaneously, in a paper otherwise careful and self-aware, defines the shape of the gap between the hypothesis and a testable one.

Overreach 1 — The Missing Risky Prediction

[TIER 3 — OPEN]

The paper never states what would be surprising under its model. Heterogeneity per se is not the flaw — evolutionary theory routinely predicts heterogeneous outcomes — the flaw is the absence of any stated probability ordering over outcomes, without which the model and the null (drift + relaxed selection + plasticity) are observationally indistinguishable.

Overreach 2 — The Missing Stop Criterion

[TIER 3 — OPEN]

The paper never specifies what a ‘stable evolutionary condition’ would look like or how long observation without change would need to continue before a structure is ruled out as transitional. Since the paper elsewhere notes that cultural change outpaces genetic response, there is no operational test period after which ‘still transitional’ could be rejected in favor of ‘actually just stable.’

Overreach 3 — The Missing Longitudinal Engagement

[TIER 3 — OPEN]

Published — and contested — claims of rising palmaris longus absence and rising third-molar agenesis frequency in younger cohorts exist in the literature and go uncited. These are the strongest available evidence for the paper’s own thesis, however disputed; a paper claiming ongoing transition should engage its only directly relevant data stream, not rely on cross-sectional variation that all parties agree is non-diagnostic.

Overreach 4 — The Missing Adjacent Literatures

[TIER 3 — OPEN]

The paper does not position itself against the human gracilization literature, the self-domestication hypothesis, or the ‘ongoing evolution of upper extremity muscles’ hypothesis it cites but does not critically weigh. Gracilization substantially predates agriculture, which complicates attributing anatomical reduction to recent technology — a complication the framework must address rather than inherit silently.

Overreach 5 — The Asymmetric Evidential Standard

[TIER 3 — OPEN]

The paper’s Section 8 argues that the absence of a recoverable soft-tissue record ‘cannot necessarily establish anatomical stability’ — true, but applied one-way. Missing data is used to keep the transitional interpretation alive (absence of evidence $\neq$ evidence of stability), while the corresponding standard is never applied to the hypothesis itself (absence of evidence $\neq$ evidence of transition). Combined with the Section 12 concession that modern technology is too recent for outcomes to be observable, Sections 8 and 12 form a closed loop in which every evidential gap resolves in favor of the hypothesis.

Overreach 6 — The Terminological Drift

[TIER 3 — OPEN]

‘Transitional’ already denotes transitional forms between taxa and transient polymorphisms in population genetics. The paper coins a third usage without flagging it, inviting equivocation and reviewer resistance. A less loaded coinage (‘unresolved-trajectory structure’) would carry the same content without the collision.

Overreach 7 — The Single-Contested-Paper Dependency

[TIER 3 — OPEN]

The plantaris argument leans on one 2025 paper arguing against the muscle’s vestigial status — itself a proposal, not a consensus — and is cited without listed authors. A framework built partly on a single contested source overweights it; and an anonymous-format citation in an otherwise complete reference list looks like an incomplete citation rather than an authorless source.

Overreach 8 — The Detection-Asymmetry Feasibility Problem

[TIER 3 — OPEN]

The canonical gene–culture cases work as evidence because they involve positive selection — sweeps that are detectable. The paper’s proposed mechanism is almost entirely relaxed selection, which leaves no sweep and is notoriously close to undetectable at the genomic level. The paper therefore proposes a research program whose central mechanism is the one least amenable to the very methods it recommends (ancient DNA, allele-frequency reconstruction). This is a feasibility constraint on the program, not merely a rhetorical one.

3.9 — The Citation Hygiene Constraint

[TIER 3 — OPEN]

Minor but telling in a document whose authority rests on evidentiary standards: the plantaris reference appears as an anonymous title. Flag it as authorless if genuinely so; as printed, it reads as an omission.

Section 4 — What the Paper Actually Shows

Honest engagement with the paper requires acknowledging what it has documented.

[CONCEDED]

- The retrospective-only interpretation of vestigial anatomy may be incomplete for humans specifically — a defensible, though weak, claim.
- Gene–culture coevolution is real: cultural behavior can precede and generate biological adaptation (lactase persistence, amylase copy number, malaria resistance associated with agriculture).
- Human anatomy remains variable, and some structures can be reduced or absent without substantial functional consequences under current conditions.
- Technological substitution plausibly alters selective environments — directionally reasonable, even where undemonstrated per structure.

- The paper is unusually self-aware: its Section 12 concedes correlation≠causation, the recency of modern technology, and the non-unity of ‘vestigial’ as a category — concessions most programmatic papers do not make.
- The paper correctly frames itself as a framework for investigation rather than a demonstrated result — which is precisely why it should be held to the standard of frameworks: generating testable predictions.

The paper’s best section is the lactase section — the only section that does what the whole paper proposes: a dated cultural trigger, a specific genetic mechanism, and a testable causal chain. The review’s constructive claim is that a version of the paper consisting of four such sections — one per candidate structure, each with a committed prediction and a stated disconfirming observation — would be a substantially stronger paper at roughly the same length.

Section 5 — What This Review Establishes

What This Review Can State

- The central hypothesis, as formulated, is unfalsifiable: it accommodates every possible observation, including stasis. [Tier 1]
- One strong case (lactase persistence) is carrying the inferential weight of many weak ones, across a category (anatomical structures) that differs from the demonstrated case (single-variant enzymatic trait) in exactly the properties that made the case demonstrable. [Tier 1 — directional]
- Plasticity — the best-documented technology-driven anatomical change — is not heritable evolution and cannot support the thesis; it is a competing hypothesis, not a confound. [Tier 2]
- The candidate structures do not form a coherent category, and the paper’s rhetorical accumulation of hedged examples creates an impression of cumulative evidence that case-by-case analysis does not supply. [Tier 1]
- The observed variation substantially predates the technological triggers proposed to explain it; no trigger is dated to the origin of the variation it is asked to explain. [Tier 1]
- Eight independent categories of omission — no risky prediction, no stop criterion, no longitudinal engagement, no adjacent-literature positioning, asymmetric evidential standards, terminological drift, single-contested-source dependency, and a detection-asymmetry feasibility problem — together define the gap between the hypothesis and a testable one. [Tier 3 — explicitly OPEN]

What This Review Cannot State

- That the transitional reading of any particular structure is false. It may prove correct for some structures; this review identifies the absence of the predictions that would demonstrate it.
- What mechanism actually produced the observed variation. Genuinely unknown — drift, relaxed selection, plasticity, and genuinely recent technological selection all remain live candidates for different structures.
- That technology has not altered selection on human anatomy. It plausibly has; the claim under review is that this has not yet been demonstrated per structure, per trigger, per prediction.
- That the hypothesis is permanently untestable. Section 8 proposes seven investigations, several executable immediately, that would convert the framework into a test.

Section 6 — Conclusion

The paper's hypothesis has not yet demonstrated the predictions, disconfirming conditions, and mechanism-separation that its own framing implies — not because the question is uninteresting, but because, at several specific and tested points, the testable content has not yet been stated. This does not establish that the hypothesis is false; it establishes that it has not yet met the same evidentiary burden this review applies to every account, including the retrospective one. The honest position is that the question remains open, the gap is well-defined by the logic of testing and by evolutionary chronology, and the standing challenge to the hypothesis — state one risky prediction and one disconfirming observation — has not yet been met. The next step is not more framing. It is targeted empirical commitment.

Section 7 — What Would Disconfirm This Review

These are genuine commitments. The framework is falsifiable — and applies to itself the same standard it applies to the paper.

Findings That Would Require Significant Revision

- The paper (or a successor) commits to the comparative ranked prediction of Section 8.1 — variability scales with degree of technological substitution across structures — and the correlation is found in either direction.

- A per-structure causal case of the lactase type: a dated technological transition, a specific developmental-pathway allele, and a frequency shift in ancient DNA correlated with the transition, for any candidate structure.
- A well-designed plasticity-controlled study of third-molar agenesis (or palmaris longus prevalence) showing heritable frequency change beyond what plasticity and population structure predict.
- A replication study independently confirming the plantaris sensorimotor function claim or the palmaris dispensability claim under standardized protocols, closing the functional-biology question the review treats as open.

Partial Disconfirmation

Tier 1 claims are independent. Closing one does not close the others: a confirmed genomic signature for one structure would revise, not refute, the claims about the other three. Tier 2 claims are conditional — if the interpretive premise is refuted, that specific argument does not apply. Tier 3 claims are explicitly OPEN. Evaluate on the combined pattern.

Findings That Would Require Review Abandonment

This review is abandoned, not merely revised, if the paper is revised into a form that commits to specific risky predictions with stated disconfirming observations — because the review’s central claim is that no such commitments currently exist. A revised paper stating one ranked comparative prediction (Section 8.1), one dated per-structure causal case, and one disconfirming observation per candidate structure would satisfy the review’s entire Tier 1 case regardless of what the subsequent evidence shows.

If any of the above findings emerge in the literature, they will be folded into this document with the same honesty that has characterized all corrections to date.

Section 8 — Investigative Methodology: How the Impasse Might Actually Be Resolved

These investigations are not designed to confirm the review’s verdict. They are designed to answer specific physical questions that have remained untested. Under the transitional hypothesis, specific results are predicted. Under the null (drift + relaxed selection + plasticity), different results are predicted. The evidence will determine which

set is correct, and this review is committed to accepting whichever result the investigations produce.

8.1 — The Comparative Substitution-Rank Study [Tier A — Immediate Priority]

The hypothesis's own logic generates a ranked, cross-structure prediction it has never stated.

PROPOSED INVESTIGATION: Rank the candidate structures (plus controls) by independently assessed degree of technological substitution of their ancestral function — grip mechanics (palmaris, plantaris), mastication (third molars), none identified (appendix; controls such as the pyramidalis or auricular muscles). Compile prevalence variability and agenesis-frequency data across populations from the anatomical literature. Test whether variability correlates with substitution rank, controlling for mutational target size, developmental-genetic architecture, and population history. Predicted outcome under the transitional hypothesis: significant positive correlation with substitution rank. Predicted outcome under the null: correlation with genetic-architecture variables only, no residual correlation with substitution. No excavation-equivalent permission required — literature synthesis and biostatistics only. Estimated effort: 6–12 months. Can be initiated immediately.

8.2 — Plasticity-Controlled Cohort Analysis of Third-Molar Agenesis [Tier A — Immediate Priority]

PROPOSED INVESTIGATION: Assemble cohort data on third-molar agenesis and impaction with documented dietary/masticatory-load histories, separating agenesis (genetic; PAX9/MSX1/WNT10A/AXIN2 pathways) from impaction/crowding (largely plastic consequence of jaw dimensions). Control for jaw-size reduction measured independently of the molar outcome. Predicted outcome under the transitional hypothesis: agenesis frequency shifts not explained by plasticity, population structure, or known developmental genes. Predicted outcome under the null: impaction tracks plastic jaw reduction; agenesis tracks known genes and population history with no residual technological signal. Can be initiated immediately with existing clinical and archaeological dental datasets.

8.3 — Longitudinal Prevalence Meta-Analysis [Tier A — Immediate Priority]

PROPOSED INVESTIGATION: Systematic meta-analysis of the published (and contested) claims of rising palmaris longus absence and rising third-molar agenesis in younger cohorts, rating each study's methodology (ascertainment, cohort comparability,

population stratification). Where claims survive quality screening, identify the datasets needed to test them prospectively. Predicted outcome under the transitional hypothesis: at least some rising-frequency claims survive methodological screening. Predicted outcome under the null: rising-frequency claims dissolve on ascertainment correction, consistent with stable deep-time frequencies. This converts the paper's most-cited-but-uncited evidence stream into a citable verdict either way.

8.4 — Ancient DNA Relaxed-Constraint Scan [Tier B]

PROPOSED INVESTIGATION: Genome-wide scan of ancient and modern DNA for allele-frequency trajectories in developmental-pathway genes affecting the candidate structures, testing for (a) constraint decay (increasing derived-allele accumulation) and (b) dated shifts coincident with documented technological transitions. Honest feasibility note: relaxed selection is close to undetectable at current sample sizes; the scan is worthwhile precisely because it quantifies how close. Predicted outcome under the transitional hypothesis: dated frequency shifts at some pathway loci. Predicted outcome under the null: trajectories consistent with drift and demographic history.

8.5 — Functional Replication Study (Plantaris and Palmaris Claims) [Tier B]

PROPOSED INVESTIGATION: Formal functional-anatomy replication of the plantaris sensorimotor claim and the palmaris-dispensability claim, using standardized biomechanical and proprioceptive protocols across presence/absence phenotypes. Document what each structure does and does not contribute, independent of the evolutionary interpretation placed on the result. Predicted outcome is deliberately uncommitted: whichever result emerges constrains both the hypothesis and the review.

8.6 — Anatomical Collection Survey [Tier C]

PROPOSED INVESTIGATION: Systematic re-examination of documented anatomical-variation collections (historical dissection records, museum osteological/dental collections) with modern imaging, to establish baseline prevalence distributions and their population structure before any evolutionary interpretation is placed on them.

8.7 — Adjacent-Literature Positioning Review [Tier C]

PROPOSED INVESTIGATION: A structured review positioning the transitional framework against the self-domestication hypothesis, the gracilization literature, and the ongoing-evolution-of-upper-extremity-muscles literature — establishing which observed patterns each framework claims, and where their predictions diverge. This is

the cheapest investigation on the list and the one most likely to force the hypothesis to commit to something falsifiable.

8.8 — Priority Research Agenda

#	Proposal	Priority	Effort	Permission	Who Can Execute
8.1	Comparative substitution-rank study	TIER A	6–12 mo	None	Biostatistics + evolutionary biology
8.2	Plasticity-controlled agenesis cohort analysis	TIER A	6–18 mo	Clinical/archival data access	Dental anthropology + epidemiology
8.3	Longitudinal prevalence meta-analysis	TIER A	6–12 mo	None	Systematic-review team
8.4	Ancient DNA relaxed-constraint scan	TIER B	12–24 mo	Dataset access	Population-genetics lab
8.5	Functional replication study	TIER B	12–24 mo	Ethics approval	Functional-anatomy lab
8.6	Anatomical collection survey	TIER C	12–24 mo	Museum access	Anatomy + imaging partnership
8.7	Adjacent-literature positioning	TIER C	3–6 mo	None	Any of the four reviewing

	review				parties
--	--------	--	--	--	---------

IMMEDIATE ACTION: Proposals 8.1, 8.2, 8.3, and 8.7 require no permissions beyond ordinary data access, use standard methods, and have high diagnostic value at modest cost. These should be the first phase of any resolution program. Results from Tier A investigations will inform the design and targeting of subsequent investigations in Tier B and C categories.

Section 9 — Next Phase: Application to Other Evolutionary Claims

This review was developed for one paper, but the analytical framework — the Standing Methodological Principle, the Plasticity Corollary, the three-tier critique structure, the Tier 1 checklist, the Combined Overreach Pattern, and the Section 8 investigative methodology — is a reusable template applicable to any claim of ongoing or technology-driven human evolution. One caution applies specifically to Sections 1 and 2: the particular constraints (unfalsifiability via outcome-listing, lactase overgeneralization, palmaris/plantar third-molar specificity) do not transfer directly to other claims. Each claim requires its own analysis using its own proposed mechanisms, timelines, and evidential base. The reusable element is the general principle: that any claim of ongoing evolution must commit to at least one risky prediction, one disconfirming observation, one plasticity control, and one dated trigger — and that a framework failing all four is an interpretive lens, not a hypothesis.

- The human self-domestication hypothesis — a strong candidate for the same treatment: selection mechanism contested, timeline deep, plasticity confounds prominent, risky predictions scarce.
- ‘Ongoing evolution’ claims in general — the four-commitment test (risky prediction, disconfirming observation, plasticity control, dated trigger) discriminates hypotheses from framings across the entire literature.
- Comparative claims across species — the comparative ranked-prediction method of Section 8.1 applies wherever a hypothesized selective force varies in degree across comparable cases.

The analytical template used here is not specific to vestigial anatomy and is available as a standalone method for application to any of the claims listed above.

Section 10 — Honest Assessment of This Review

[SELF-APPLICATION]

- Source-bias flag (standing rule): all four constituent critiques and this synthesis were produced by large language models. Convergence of four independent models is evidence that the paper's flaws are structurally visible to multiple reviewers — it is not evidence that the critique is correct. On any contested point, a human anatomist or philosopher of science outranks this document.
- Scope limit: this review evaluates logical structure and evidential architecture. It performed no independent verification of the underlying anatomical, embryological, or genomic claims made by either the paper or its critics (agenesis frequencies, pathway genetics, prevalence figures). Those claims were accepted as stated. The review's Tier 1 findings survive this limit because they concern the paper's structure, not its facts; the Tier 3 proposed investigations inherit the unverified factual base and are only as good as it.
- Intent-dependence: the Tier 1 findings rest on interpreting the document as a hypothesis. The document's metadata records its status as a deliberately non-viable specimen prepared for review-site demonstration. The review's content was not modified in light of that status — a review that went easier on a known specimen would be demonstrating preference, not analysis — but readers should note that 'non-viable as designed' and 'failed attempt' are different verdicts, and this review renders the first.
- Feasibility uncertainty: Section 8's investigations are proposals. Their feasibility — data access, ascertainment quality of historical dissection records, sample sizes adequate for relaxed-selection detection — has not been validated. Estimated efforts are order-of-magnitude.
- Recommended next check: this review should itself be cold-burned — one adversarial pass by a model not among the four, assigned to defend the paper and attack the review. A review that cannot survive its own method should not be published on a site that sells the method.

Correction Log

Version	Date	Section	Change	Reason
1.0	2026-09-11	All	Initial synthesis	Demo scaffold at author's

			delivered in chat; structure ported from the Megalithic Archaeology document	request
1.1	2026-09-11	Front matter; tags; new §10; end matter	House-style reissue: metadata table, version notes, standardized tier tags, honest-assessment section, correction log, versioned footer	Product-family consistency for site demo; per Thread Continuity Brief §2
1.1 (amended)	2026-09-11	§2.1; §2.2; §2.3; §7	Removed megalithic-content carryover: vase-replication disconfirmation bullet; inundation-season, Scheuring counterweight, and Houdin analogies — replaced with domain-	Author flag (2026-09-11): structure-only port must not import content

			internal equivalents	
--	--	--	-------------------------	--

References

Vestigial Anatomy in a Technologically Modified Selective Environment: A Reassessment of Transitional Structures in Human Evolution. Manuscript under review; the paper analyzed throughout this document.

Deepseek. Independent critique of the above manuscript (synthesized, Sections 1–3 sources).

Claude Incognito. Independent critique of the above manuscript (synthesized, Sections 1–3 sources).

Qwen. Independent critique of the above manuscript (synthesized, Sections 1–3 sources).

Gerbault, P., et al. (2011). Evolution of lactase persistence: An example of human niche construction. *Philosophical Transactions of the Royal Society B*, 366, 863–877.

Laland, K. N., Odling-Smee, J., & Myles, S. (2010). How culture shaped the human genome. *Nature Reviews Genetics*, 11, 137–148.

Capdarest-Arest, N., Gonzalez, J. P., & Türker, T. (2014). Hypotheses for ongoing evolution of muscles of the upper extremity. *Medical Hypotheses*, 82(4), 452–456.

Kooij, I. A., et al. (2016). The immunology of the vermiform appendix. *Clinical & Experimental Immunology*, 186(1), 1–9.

Smith, H. F. (2023). A review of the function and evolution of the cecal appendix. *The Anatomical Record*, 306(5), 972–982.

The Plantaris Muscle Is Not Vestigial: Developmental, Comparative, and Functional Evidence for Its Sensorimotor Role. (2025). *Biology*, 14(6), 696.

Angelakopoulos, N., et al. (2026). Third molar agenesis: An updated systematic review and meta-analysis. *Archives of Oral Biology*, 189, 106662.

Vestigial Anatomy and the Burden of Proof — v1.1 — September 2026. This document applies to itself the Standing Methodological Principle stated in its front matter: if the paper under review produces the risky predictions and dated causal cases this review calls for, this review's Tier 1 case collapses, and this document will say so.