Reassessment of coding-sequence constraint across the volvocine multicellularity gradient: correction of versions 1 and 2

Masato Tanigawa

Department of Biophysics, Faculty of Medicine, Oita University, Oita 879-5593, Japan. tanigawa@oita-u.ac.jp. ORCID 0000-0001-5877-2271

This is version 3 of EcoEvoRxiv preprint 10.32942/X2DD40. It replaces versions 1 and 2, whose central conclusions are not supported by the corrected comparisons reported here. No new data were generated. The result summaries in this text are rendered from a machine-readable manifest built from the deposited result files; input data, values quoted from the earlier versions and static metadata are documented separately (Data availability).

Abstract

Versions 1 and 2 of this preprint interpreted lower dN/dS and shorter amino-acid terminal branches as evidence that coding-sequence constraint strengthens across the volvocine multicellularity gradient. Re-examination of the models and comparisons does not support that interpretation. The original rooted branch-model fits often reached boundary estimates. An unrooted specification, which is a different model for the original branch labels, gives lower foreground than background dN/dS in 71.5 % of 1,374 quality-controlled orthogroups, compared with the previously reported 88.6 % of 845; model specification and orthogroup selection both change. Separating the two lineages inferred in previous studies to represent independent origins of multicellularity yields different patterns: on a common set of 1,336 orthogroups, the fraction is 48.4 % for Tetrabaenaceae and 73.5 % for Goniaceae + Volvocaceae. At the latter origin, lower fitted dN/dS accompanies greater synonymous divergence relative to the *Chlamydomonas reinhardtii* branch (median path ratio 1.96), without a lower median nonsynonymous path (ratio 1.25). This arithmetic decomposition does not identify a cause or exclude stronger selection. On a separate public 263-gene dataset, the primary somatic-differentiation analysis retains only 2 usable contrasts. Numerically validated refits give a wide interval for the mean log path-length ratio (-0.33 to 0.45), including slower and faster evolution; applying both previously omitted filters is reported as a post hoc sensitivity check. Terminal branch length also confounds rate with time since the last sampled split, and no significant cell-number association was detected on the denser 57-OTU tree. The earlier comparisons therefore do not establish genome-wide strengthening of constraint or test disposable-soma theory. This corrected version identifies the unsupported interpretations, documents specification changes and numerical diagnostics, and provides the materials needed to reproduce its summaries.

Keywords: dN/dS; branch models; volvocine algae; multicellularity; negative result; correction

1. Introduction

Versions 1 (3 July 2026) and 2 (16 July 2026) of this preprint analysed ten volvocine green algae spanning the unicellular *Chlamydomonas reinhardtii*, the colonial *Tetrabaena*, *Gonium*, *Astrephomene*, *Yamagishiella*, *Eudorina* and *Pleodorina*, and three *Volvox* species, using PAML branch models (Yang 2007) on 1,755 single-copy nuclear orthogroups. They reported four headline results: foreground (multicellular) ω below background (*Chlamydomonas*) ω in 88.6 % of 845

quality-controlled orthogroups (median ω_bg 0.117, median ω_fg 0.046); the same direction in all 20 COG functional categories; shorter aggregated multicellular amino-acid branches in 60.6 % of orthogroups (882/1,456); and a negative phylogenetic regression of log median terminal branch length on log2 cell number (β = -0.123, P = 0.002, n = 10), with a negative per-orthogroup correlation in 96.1 % of orthogroups (permutation P = 0.0107). The pattern was interpreted as consistent with the disposable-soma prediction that somatic-maintenance investment strengthens with multicellular complexity.

The manuscript was submitted to a journal and rejected. The reviewer identified the species-tree topology, which had been imposed on every analysis rather than inferred, as a fundamental problem, on two grounds: the analyses treated all colonial species as descendants of a single origin, whereas the phylogenomic studies of Lindsey et al. (2021, 2024) support two independent origins of multicellularity (in Tetrabaenaceae and in Goniaceae + Volvocaceae); and a single extant species, *C. reinhardtii*, stood for the unicellular condition. The review prompted three re-analyses, each with decision criteria written down and hash-recorded before its outcome analyses were run (Methods 2.2). While the first was being specified, a problem the reviewer had not raised was found in the original outputs: the branch models had been fitted to a rooted tree.

This version reports those re-analyses. Its purpose is to record, under the same DOI, that the claims of versions 1 and 2 are not supported by the corrected comparisons, to identify what in the original material remains usable, and to state the errors precisely enough that they can be avoided. It is a negative result about the comparisons of versions 1 and 2, not a test of the disposable-soma theory, which concerns allocation between reproduction and somatic maintenance and does not by itself predict the direction of genome-wide nonsynonymous substitution.

2. Materials and Methods

2.1 Data (unchanged from versions 1 and 2)

The ten genome assemblies, gene annotations, OrthoFinder v2.5.5 orthogroup assignment (1,755 single-copy orthogroups) and MAFFT-based codon alignments are those of versions 1 and 2 (Supplementary Table S1 lists accessions and annotation sources; Featherston et al. 2018; Hanschen et al. 2016; Hamaji et al. 2018; Yamashita et al. 2021; Yamamoto et al. 2021; Nozaki et al. 2022). These inputs are retained as reusable material; their quality (annotation, orthology, alignment) was not independently reassessed in this re-analysis, and only the trees passed to the codon models and the comparisons drawn from them differ.

2.2 Pre-specification, amendments and departures

Each re-analysis was specified in a document whose SHA-256 hash was recorded before the corresponding statistics were computed: gate 1 (unrooted branch models, two origins, topology; hash prefix 49ee0f9e), gate 2 (sister-clade contrasts for somatic differentiation on public data; 5f0f106f) and gate 3 (decomposition of the residual ω difference; 20db750f). The primary decision criteria in those documents were fixed before the outcome analyses were run and were not edited afterwards. What each document already contained differs: the gate-1 specification records the unconstrained species-tree estimate and the rooted-tree finding, both obtained while it was being written, and fixes the criteria for the branch-model runs that followed; the gate-2 specification was written before any of its analyses and amended as described below; the gate-3 specification was written after gates 1 and 2 had been evaluated and before the decomposition was computed. Three qualifications apply. First, the questions were formulated after the original

results were known, so the analyses are confirmatory only with respect to their own pre-specified criteria. Second, the gate-2 specification was amended on 11 September 2026 after intermediate outputs had been seen (amendments A1–A4, hash-recorded). Amendments A1–A3 followed the first ancestral-state reconstruction, which had shown the number of contrasts and their tip sets: a second background rule (v1.1) was added as a sensitivity analysis, the parsimony routine was replaced by an implementation with an explicit cost[from, to] convention and a small-tree oracle check, and the parsimony root was fixed at the undifferentiated state. Amendment A4 added a relaxed gene filter as a sensitivity analysis after a trial run on 98 of the 263 genes had left one contrast without any gene under the strict filter; the addendum records this as preceding the aggregation of r, and the available record does not establish that every amendment preceded every calculation of a contrast statistic. The primary criteria were evaluated under the original rules. Third, three items of the gate-2 specification were omitted from the recorded analysis: a stochastic-mapping sensitivity with rate matrices sampled by MCMC, the descriptive regression on the coalescent topology, and a 200-replicate orthogroup-cluster bootstrap. These were completed post hoc for this correction on 15 September 2026, under a separately recorded execution protocol, with implementation changes documented in Supplement S7. The additional planned reconstruction on a chronogram was also found to be absent and was completed post hoc (Table S7f). Two omitted filters were also applied post hoc (Section 2.6). A hash records the content of a specification; it is not an independent time stamp. The specifications, amendments, hashes and automatically generated decision tables are in the code and data archive, and Supplementary Table S6 lists every departure.

2.3 Unrooted branch models and the two origins

PAML codeml v4.10.10 was re-run on each orthogroup with model = 2, NSsites = 0, CodonFreq = 2 and the control file of versions 1 and 2, changing only the tree. Trees were supplied unrooted (basal trifurcation *C. reinhardtii* | *T. socialis* | Goniaceae + Volvocaceae). The output parser recorded the number of branch-length parameters and the presence of codeml's "rooted tree" warning for each orthogroup in every run. The decision script evaluated the H0u modal count against the specified criterion of 17 parameters for an unrooted ten-taxon tree. Modal counts and warning totals for all runs are reported in Table S2; the nine-species runs have a modal count of 15, and warning totals are 0 across the quality-controlled orthogroups in each run. Three branch-class settings were run on two topologies (Table M1): T_dat, the unconstrained maximum-likelihood species tree of the present data, in which *Astrephomene* is sister to Volvocaceae (as in Lindsey et al. 2021); and T_imp, the topology imposed in versions 1 and 2, in which *Astrephomene* is sister to *Gonium* (Goniaceae monophyletic, as in Lindsey et al. 2024). Quality control was that of versions 1 and 2 (all class ω estimates present and between 0.0001 and 10). Results are reported on the orthogroups passing quality control in all four ten-species runs (1,336, "common set"), with each run's own set in Supplementary Table S2; the nine-species runs share 1,436 orthogroups. As a sensitivity analysis, orthogroups whose *Chlamydomonas*-branch dS estimate was 10 or more were removed; this threshold removes large dS estimates and is not a criterion for saturation.

Table M1. Branch-model runs.

Branch classes	T_imp (*Gonium* + *Astrephomene*)	T_dat (*Astrephomene* + Volvocaceae)
two ratios, all colonial branches one foreground class	H0u	H3
three ratios, origin A (*Tetrabaena* terminal branch, ω1) and origin B (Goniaceae + Volvocaceae clade including its stem, ω2) against the *Chlamydomonas* branch (ω0)	H1	H1p
two ratios, *Tetrabaena* removed (its sequence deleted from the existing alignment; columns kept, not re-aligned)	H2	H2p

Origin A denotes Tetrabaenaceae and origin B Goniaceae + Volvocaceae throughout, following the broader phylogenetic and ancestral-state analyses of Lindsey et al. (2021, 2024). To characterise the original fits, the original rooted model-2 outputs were scanned for the lengths assigned to the two root-adjacent branches (the *Chlamydomonas* terminal branch and the stem leading to all colonial species), and the original rooted free-ratio (model = 1) outputs for the *Chlamydomonas*-branch ω and the nonsynonymous distances of the same two branches. Both scans are repeated by the manifest script: 1,431 model-2 outputs were rescanned, of which 1,430 are in the stored diagnostic table and agree with it to within printing precision (maximum difference 5×10^{-5}). The unrooted pipeline was compared with the original on the subset of orthogroups whose original stem branch had length ≤ 0.001 (four of them positive), where the two specifications are numerically close but not identical.

2.4 Species tree, gene trees and concordance

Per-orthogroup amino-acid trees were inferred without topological constraint (IQ-TREE v3.0.1, LG+G4, 1,000 ultrafast bootstraps, seed 20260911), replacing the constrained trees of versions 1 and 2. The species tree was inferred from the unconstrained amino-acid supermatrix (LG+G4; unpartitioned with 1,000 ultrafast bootstraps and 1,000 SH-aLRT replicates, and partitioned by orthogroup with 1,000 ultrafast bootstraps) and by ASTRAL-III v5.7.8 from the gene trees (support as local posterior probability). Gene-concordance factors from the gene trees and site-concordance factors from 100 quartets per branch were computed in IQ-TREE, and bipartition frequencies across the 1,456 gene trees with all ten species were counted with ape::prop.part. Support values are the support of each analysis for its own tree; no formal test between topologies was performed.

2.5 Decomposition of the residual ω difference (gate 3)

In a branch model every branch of a class shares one fitted ω, and codeml reports each branch's dN and dS with dN/dS equal to the class ω. For any set of branches within one class, therefore, ΣdN/ΣdS = ω, and for a foreground class B and background class 0 the identity ln(ω_B/ω_0) =

ln(ΣdN_B/ΣdN_0) − ln(ΣdS_B/ΣdS_0) holds exactly within an orthogroup (proved in Lean 4 with Mathlib; Supplementary Note S5). For each orthogroup of the H1p run, I took the basal trifurcation node C = {*Chlamydomonas*} | {*Tetrabaena*} | {Goniaceae + Volvocaceae stem} and computed, from the dN and dS trees printed by codeml, the path length from C to each of the eight Goniaceae + Volvocaceae tips and to the *C. reinhardtii* tip. With dN_B and dS_B the medians over the eight tips, a = ln(dN_B/dN_chl) and b = ln(dS_B/dS_chl) are log ratios of path lengths, and a − b = ln(ω2/ω0) within each orthogroup; this identity was checked in every orthogroup (maximum deviation 0.0161, from six-decimal printing). Across orthogroups, medians of a, of b and of a − b are reported separately, because the identity does not carry over to medians. Under the maximum-likelihood rooting of Lindsey et al. (2024) restricted to these ten species, the root lies on the Tetrabaenaceae lineage and C is the most recent common ancestor of *C. reinhardtii* and the Goniaceae + Volvocaceae clade, so that the paths compared span the same time; under their coalescent-based arrangement the root lies on the Goniaceae + Volvocaceae stem and they do not. Any reading of a or b as a log rate ratio is conditional on the former rooting, which the ten-genome data cannot determine. Sensitivity analyses used the mean over tips instead of the median, excluded *V. reticuliferus* (which has the largest dS path ratios), excluded orthogroups with *Chlamydomonas*-branch dS ≥ 10, and repeated the calculation on the T_imp run (H1). The same quantities were computed for each of the nine colonial species separately; within an orthogroup all origin-B paths share the fitted ω2/ω0, so per-species differences lie in the path lengths, not in ω.

2.6 Sister-clade contrasts for somatic differentiation (gate 2)

To test whether somatic differentiation is accompanied by slower protein evolution, I used the public amino-acid alignments of Lindsey et al. (2024; Dryad 10.5061/dryad.mcvdnck6b): 263 single-copy genes conserved across the three major Archaeplastida clades, for 164 taxa, pruned to the Reinhardtinia clade plus *Chlamydomonas moewusii* and *Chlorella variabilis* as outgroups (68 tips) and, for the contrasts, to one representative strain per species (the strain with the most genes), except for the polyphyletic *Eudorina elegans*, all 7 strains of which were kept as separate OTUs (57 tips: 48 species representatives, the 7 *E. elegans* strains and the 2 outgroups). Differentiation was coded in three states from Lindsey et al. (2021, Table S2): 0, none; 1, partial (some somatic cells; *Astrephomene*, *Pleodorina*, most *Volvox*); 2, complete germ–soma differentiation (*V. carteri*, *V. africanus*, *V. gigas*, *V. obversus*). Cell numbers for this dataset are a single value per genus, the geometric midpoint sqrt(low x high) of the genus-level range given by Lindsey et al. (2021). The cell-number regressions on this 57-OTU dataset inherit that coarse genus-level coding. The ten-species regressions in Section 2.7 retain the separate per-species values used in versions 1 and 2. Ancestral states were reconstructed under seven scenarios on the maximum-likelihood topology of Lindsey et al. (2024) with branch lengths re-estimated from the concatenated alignment (LG+G4, partitioned by gene): equal-rates (S1) and all-rates-different (S2) Mk models and an ordered 0 ↔ 1 ↔ 2 model (S7), each summarised over 1,000 stochastic maps simulated on the fitted rate matrix (phytools); and Sankoff parsimony with gain:loss costs of 1:1 (S3), 1:2 (S4), 2:1 (S5) and 100:1 (S6), with the root fixed at state 0. For the Mk scenarios a branch's score is the mean fraction of its length spent in states 1 or 2 across the maps; for the parsimony scenarios it is the mean of the scores at its two ends. Branches with score ≥ 0.9 were classed foreground and ≤ 0.1 background. Each connected component of foreground branches whose stem is not foreground counts as a raw origin; the design contrast for a component pairs it with the nearest sister subtree that contains no foreground-classed branch (rule v1; rule v1.1 takes the largest such subtree nearest the sister's root). Raw components whose recorded foreground or background branch sets overlapped were merged into one contrast (tip sets united; the larger

node distance kept), so a foreground tip set need not comprise every descendant of its most recent common ancestor, and a background set defined this way can contain tips coded as differentiated whose branches fell below the foreground threshold and can lie several nodes above the foreground component (Supplementary Table S3e lists the tips of every contrast). Contrasts share no branches. Three counts are reported: raw origins, design contrasts (K), and contrasts with at least one gene after the gene filter below.

For gene g and contrast k, r_gk = ln(mean path length from the contrast's common ancestor to the foreground tips / the corresponding mean for the background tips). The shared ancestor makes the elapsed time equal, but the paths do not isolate the period after differentiation evolved. Path lengths came from per-gene LG+G4 trees on the fixed topology. The strict filter required each side to be monophyletic in the unconstrained gene tree; the relaxed sensitivity filter required reciprocal monophyly after pruning to the two sides. Two pre-specified filters were omitted from the recorded implementation: at least two sampled tips on each side per gene, and at least 150 aligned amino-acid columns. Although every design contrast has at least two tips per side, missing sequences leave 35 of the primary observations (32 genes) with a single-tip side. Across the input set, 45 alignments are shorter than 150 columns; 12 such genes contribute 14 primary observations. A second statistic, Δrad_gk, is the between-side difference in the fraction of inferred substitutions with Grantham distance > 100; each side must have at least 10 inferred substitutions.

Both statistics were fitted with $y_{gk} = \mu + a_g + b_k + e_{gk}$ in brms/rstan, with independent gene and contrast random intercepts, $\mu \sim N(0, 1)$, $\sigma_g \sim$ half-N(0, 1), $\sigma_k \sim$ half-N(0, 0.5), and residual scale $\sigma \sim$ half-Student-t(3, 0, 2.5). For r, each Gaussian log-likelihood contribution was weighted by min(ΣPL_F, ΣPL_B) × alignment length, with weights scaled to mean 1; Δrad was unweighted. The original four-chain run used 4,000 iterations per chain; archived sensitivity runs and departures are detailed in Table S6. The pre-specified directional criterion was $P(\mu < 0) \geq 0.9$ with at least $K - 1$ negative contrasts under S2. Archived fits, including leave-one-contrast-out and prior-sensitivity fits, remain in Tables S3b–S3f, with their diagnostic limitations; they are not treated as numerically validated merely because they share the same estimated direction.

Post hoc numerical validation retained the data, model, priors and seed (20260911). Two longer brms attempts on the primary model retained divergent transitions. I therefore evaluated the same model by exact Gaussian marginalization of the population mean and gene/contrast intercepts, then reconstructed their conditional posterior draws (Table S3g). The normalized weighted likelihood is equivalent, up to a parameter-independent constant, to Gaussian errors with variance σ^2/w_i. Its integrated likelihood was checked against the direct dense covariance calculation. Four chains initially used 8,000 iterations per chain, half discarded as warm-up, adapt_delta = 0.99 and maximum tree depth 15; a fixed fallback used 16,000 iterations and adapt_delta = 0.999 if diagnostics failed. Acceptance required zero post-warm-up divergences, no maximum-tree-depth hits, $\hat{R} < 1.01$, bulk and tail effective sample sizes of at least 400 for monitored model parameters, energy BFMI ≥ 0.3 in each chain, and Monte Carlo standard error for $\mu \leq 0.01$. Acceptance did not depend on the sign or size of the effect. The primary r model was fitted on all recorded observations and with each omitted filter separately and both together; the primary Δrad model was also checked. Attempts, accepted settings, chain summaries and sampler diagnostics are retained in the code and data archive and Table S3g. The original decision tables are unchanged; these validated refits supply the quantitative estimates reported below. The complete archived scenario grid was then audited and refitted with the same marginal likelihood, including the primary prior-scale and leave-one-contrast-out sensitivities (Table S7a). Identical datasets and priors share a fit through an explicit registry. All prior, weighting and acceptance

settings were retained; failed attempts remain in the archive.

The previously omitted ARD rate-uncertainty sensitivity used four independent phytools rate chains, independent Gamma(shape 1, rate 1) priors on the six transition rates, and an equal root-state prior on the original OTU substitution-length tree. These prior details were fixed in the completion protocol because the earlier specification did not define them. Fixed retry settings and rate-chain/branch-occupancy diagnostics are given in Table S7b. Accepted branch-occupancy means were passed through the unchanged thresholds and contrast rules, then into the gene statistics and hierarchical fits where observations were available. This integrates rate uncertainty into the thresholded reconstruction; it does not average the final effect posterior over every possible contrast design.

The 200-replicate bootstrap resampled contributing genes with replacement, keeping every observation of a selected gene, giving each selected copy a distinct gene identifier and renormalizing the weights. Contrasts remained fixed. Each replicate estimated the posterior mean of μ, using its conditional Gaussian mean to reduce Monte Carlo noise. All replicates had to pass the sampling criteria; an additional MCSE threshold of 0.002 applied to this estimator. The percentile interval measures gene-resampling uncertainty conditional on the two contrasts, separately from the hierarchical credible interval across origins (Table S7d).

A descriptive phylogenetic regression of log median terminal branch length on log2 cell number and differentiation state (caper::pgls, λ by maximum likelihood, on a chronogram from ape::chronos) was pre-specified as a joint model; the single-predictor regressions reported here are additional descriptive fits, and the joint fit is in Supplementary Table S3d. When the figure was drawn it emerged that six representative strains had near-zero terminal branches because a strain of the same species remained in the gene trees. Three post hoc variants were therefore added: B, removing representatives whose median terminal branch was below 0.001; C, measuring terminal branches after pruning each gene tree to the representatives; and D, variant C without the two representatives whose median terminal branch remained below 0.001. The 57 OTUs include the outgroups; cell numbers for *Volvox* are genus-level values. All four variants are reported (Supplementary Table S3d). The ASTRAL covariance sensitivity was completed using branch lengths estimated from the same partitioned concatenation with topology fixed to ASTRAL. Both ML and ASTRAL trees were rooted at the midpoint of the outgroup edge before time transformation, avoiding the zero-duration branch produced by the initial rooting convention. An objective–gradient mismatch was identified in the installed correlated chronos implementation; a separately labelled local sensitivity used a consistent squared-rate-difference penalty and a numerically verified gradient. The archived ML chronogram and both implementations on the matched root convention are compared in Table S7c. The gene-level responses were held fixed. These remain descriptive regressions on relative-time proxies. The previously omitted chronogram reconstructions were also completed under ER, ARD and ordered Mk models on the locally consistent ML chronogram, using 1,000 maps at each fitted Q and both contrast rules. Gene-level statistics and fits were propagated as for the rate-MCMC sensitivity (Table S7f).

2.7 The terminal-branch gradient

The ten-species gradient of versions 1 and 2 retained the per-species cell-number values recorded in `per_species_iqtree_summary.csv` and was recomputed from the unconstrained gene trees (median terminal amino-acid branch length per species; PGLS with caper on the unconstrained species tree, λ by maximum likelihood, on log2 cell number with and without *C. reinhardtii*; 10,000 species-label permutations of the per-orthogroup Spearman correlation, seed 42). The

same regression was run on per-species median free-ratio ω taken from the original rooted model-1 run of versions 1 and 2 and restricted to the nine colonial terminal branches; these ω values were not re-estimated on unrooted trees, so their robustness to the rooting is not established, and the result is reported as a diagnostic only. The interpretation of the gradient rests on the observation that a terminal branch length is the product of a rate and the time since the lineage's last sampled split; in an untransformed ordinary least-squares illustration, the slope of branch length on a trait X equals r·Cov(T, X)/Var(X) when all lineages share the rate r (Supplementary Note S5), and the time component does not disappear under a log transform or a phylogenetic correction.

2.8 Functional categories

Versions 1 and 2 reported the direction of the ω comparison separately for each COG functional category. That tally was recomputed here, post hoc and descriptively, on the corrected fits. The orthogroup-to-category assignment of versions 1 and 2 was retained unchanged: each orthogroup inherits the COG letters of its eggNOG annotation, every letter of an available annotation votes, and ties are resolved by first occurrence in the fixed species and annotation order. The reproduction was verified by regenerating the historical category table from this mapping and comparing it cell by cell with the archived table before the mapping was applied to any corrected run. Tallies were computed for the original rooted run, for the unrooted pooled run (H0u) and for each foreground class of the three-ratio run (H1p), each on its own quality-controlled set and on the intersection of the quality-controlled sets of the ten-species runs; a category enters a tally only if at least five orthogroups carry it in that subset. The archived mapping table allows the tallies to be reproduced without the raw annotation files. These are re-counts of the same per-orthogroup ω values grouped by annotation, not independent tests, and no significance test is attached to them.

2.9 Software and reproducibility

IQ-TREE 3.0.1, PAML 4.10.10, ASTRAL-III 5.7.8, R 4.3.3 (ape 5.8.1, phytools 2.4.4, caper 1.0.3, brms 2.23.0, rstan 2.32.7), Python 3.11.0 for the analysis scripts (conda environment) and 3.12.3 for the manifest, figures (matplotlib 3.6.3) and machine checks (numpy 1.26.4, pandas 3.0.1, SymPy 1.14.0, z3 5.1.0), and Lean 4.32.2 with Mathlib; versions were read from the executables that ran each script and are recorded in the accompanying archive. The result summaries in this text are rendered from a manifest that regenerates them from the accompanying result files; values quoted from versions 1 and 2, software versions, identifiers and method constants are typed and documented separately. Result placeholders must resolve to non-null manifest values. The renderer also checks decimals and large numeric literals against a documented allowlist, excluding references, identifiers, software versions and cross-references; small integers are not covered by this check. Validation must pass before the generated manuscript is replaced. A manifest guarantees that a printed value matches its file; it does not guarantee that the file was computed on the right set, or that the value means what the sentence says (an earlier draft of this version printed a tip count of 134 for a 68-tip tree because the manifest counted commas), which is why Supplementary Table S6 reconciles each Methods statement with the code.

3. Results

3.1 The original rooted fits were unstable, and an unrooted specification gives a smaller fraction

The original runs supplied codeml with a rooted tree in which the *Chlamydomonas* terminal branch (background class) and the stem of all colonial species (foreground class) meet at the root. When the two root-adjacent branches carry the same substitution process, only their sum is constrained by the data; when they carry different ω classes, as here, the model is formally identifiable (Álvarez-Carretero et al. 2023), but in these data the fits frequently placed one of the two branches at a boundary. The stored diagnostic table of the original fits contains 844 of the 845 orthogroups that passed the original quality control (the remaining one, OG0003521, has readable root-adjacent branch lengths, with a zero-length stem, but non-finite dS entries; the rescan described in Section 2.3 includes it); in 266 of these the stem branch was longer than 0.001. The 111 orthogroups with background ω above 1 (the upper mode of the original Figure 1A) had a median *Chlamydomonas* branch length of 0.0, and in 93 of them that branch was at most 0.001 long: their background ω was estimated from a branch of essentially zero length. The 46 orthogroups with background ω below 0.01 had a median *Chlamydomonas* branch length of 50.0, the upper bound of codeml, whereas the main mode had 0.508 (Figure 1B). In the original free-ratio run, $\omega \leq 0.0001$ was assigned to the *Chlamydomonas* branch in 930 of 1,364 orthogroups, and in 928 of those the nonsynonymous distance on the stem exceeded that on the *Chlamydomonas* branch. Versions 1 and 2 attributed the free-ratio result to synonymous saturation without examining the synonymous divergence of that branch; in the unrooted re-run its median dS across quality-controlled orthogroups is 0.83.

Re-run unrooted (H0u, T_imp), 1,374 orthogroups pass the unchanged quality control (against 845 originally), the fraction with foreground ω below background ω is 71.5 % (72.0 % on T_dat, run H3; median ω0 0.074, median ω1 0.05), and the upper mode is absent (6 orthogroups with background ω above 1; Figure 1A). The lower fitted ω therefore persists in a majority of orthogroups, at a smaller fraction. Where the change comes from can be described but not fully attributed (Figure 1C). Of the original 845 orthogroups, 834 pass the unrooted quality control. Among the 575 of these whose rooted stem was at most 0.001 long, the two specifications give closely similar results (90.6 % rooted, 91.0 % unrooted; median background ω the same), which shows that the pipeline reproduces the original where the two specifications are numerically close. Among the 259 with a longer stem, the fraction falls from 83.8 % to 59.1 %, and the 540 orthogroups that failed the original quality control and pass the unrooted one show 56.9 %. Because the unrooted specification is a different model for this branch labelling and the two quality-controlled sets differ, the separate contributions of the model, of optimisation behaviour and of orthogroup selection to the difference between 88.6 % and 71.5 % were not isolated. Removing orthogroups with *Chlamydomonas*-branch $dS \geq 10$ (97 orthogroups, plus 1 without a branch record; 1,276 remain) raises the fraction to 77.0 %, so large dS estimates on the reference branch work against, not for, the original claim.

3.2 The two origins of multicellularity do not share the pattern

With the Tetrabaenaceae and Goniaceae + Volvocaceae origins as separate foreground classes (H1p, T_dat; common set of 1,336 orthogroups), the *Tetrabaena* branch ω is below the *Chlamydomonas* branch ω in 48.4 % of orthogroups (own-set medians ω1 0.082, ω0 0.077), whereas the Goniaceae + Volvocaceae class is below it in 73.5 % (own-set median ω2 0.049; Figure 2A). The pooled result of versions 1 and 2 was therefore carried by the branches of one origin, that of

the eight Goniaceae + Volvocaceae species. The imposed topology gives similar fractions (49.2 % and 73.6 %; Figure 2B), and removing *Tetrabaena* (H2p) gives 75.0 % of 1,437. Removing orthogroups with *Chlamydomonas*-branch dS ≥ 10 gives 51.6 % and 79.5 %.

The unconstrained species tree of the present data places *Astrephomene* as sister to Volvocaceae, with aLRT and ultrafast-bootstrap support of 100/100 and an ASTRAL local posterior probability of 1.00 for the Goniaceae + Volvocaceae clade, whereas the topology imposed in versions 1 and 2 followed Lindsey et al. (2024), who recovered monophyletic Goniaceae; no test between the two arrangements was performed. Gene-level support for the basal arrangement is weak: the bipartition separating {*Chlamydomonas*, *Tetrabaena*} from the eight Goniaceae + Volvocaceae species occurs in 33.0 % of the 1,456 gene trees (gene concordance factor 33.3 %, site concordance factor 39.0 %, internal branch 0.0479 substitutions per site), Goniaceae monophyly in 17.3 % and *Astrephomene* + Volvocaceae in 21.4 % (gene concordance factor for that branch 21.5 %; Supplementary Table S2). Without an outgroup, the placements "*C. reinhardtii* nested within the colonial species" and "*Tetrabaena* nested" are the same unrooted tree, so whether the *C. reinhardtii* lineage is secondarily unicellular cannot be tested with these ten genomes; the branch labels used here do not depend on the root. Interpreting the two foreground groups as independent gains follows the broader sampling and ancestral-state analyses of Lindsey et al. (2021, 2024); these ten genomes do not independently establish two gains of multicellularity. The fitted branch classes remain defined by their taxon sets without resolving that root.

The functional-category tally was recomputed on the corrected fits using the annotation assignments of versions 1 and 2 (Section 2.8; Table S7e). A majority of orthogroups still had lower foreground ω in all 20 eligible categories under H0u and in all 20 for H1p origin B. For *Tetrabaena*, the corresponding majority occurred in 8 of 20 categories under its own QC set (9 on the common set). Category-wide directionality in the pooled comparison therefore persists, but is not shared by both origins and does not establish a mechanism of stronger selection.

3.3 Lower ω accompanies greater synonymous divergence without a lower median nonsynonymous path length

For the 1,354 orthogroups of the H1p run with finite path lengths, the median of a = ln(dN_B/dN_chl) is +0.221 and the median of b = ln(dS_B/dS_chl) is +0.675: the median ratio of nonsynonymous path length between the Goniaceae + Volvocaceae paths and the *C. reinhardtii* branch is 1.25, and the median ratio of synonymous path length is 1.96 (Figure 3A). The median of the per-orthogroup ω ratio, exp(a − b), is 0.63 (the ratio of the two separate medians is 0.635). The nonsynonymous path was shorter than the *C. reinhardtii* branch in 30.2 % of orthogroups, and the synonymous path longer in 83.1 %. The lower fitted ω of the colonial branches therefore does not correspond to a lower median nonsynonymous path length; it corresponds to a larger synonymous denominator. Whether these path lengths are rates depends on the rooting (Section 2.5). The pattern is similar with the mean over tips (dN ratio 1.49, dS ratio 2.44), without *V. reticuliferus* (1.18, 1.86), without orthogroups with *Chlamydomonas*-branch dS ≥ 10 (1.25, 2.07; n = 1,255) and on the imposed topology (median a +0.158, median b +0.594). For the *Tetrabaena* branch the corresponding median ratios are close to one (dN 1.08, dS 1.05), consistent with Section 3.2.

Within each orthogroup all origin-B paths share the fitted ω2/ω0, so the per-species ω ratio is not a separate estimate; what differs between species is the divergence along the path (Figure 3B; Supplementary Table S4). Among the colonial species without somatic differentiation, the median synonymous path ratio exceeds one in all three (*Gonium* 1.39, *Yamagishiella* 1.53, *Eudorina* 1.86),

whereas the median nonsynonymous path ratio is below one in *Gonium* (0.89) and *Yamagishiella* (0.97) and above one in *Eudorina* (1.19). The species with somatic differentiation have larger ratios of both kinds (*V. carteri* 1.45 and 2.34; *V. africanus* 1.52 and 2.44), and *V. reticuliferus* has the largest (1.84 and 3.12), for which sequence or annotation quality and estimation effects remain possible explanations. These descriptive per-species summaries share the class ω and the basal path and do not test an effect of differentiation.

3.4 Sister-clade contrasts give no consistent support for slower amino-acid evolution with somatic differentiation

On the 57-OTU tree, the number of origins of somatic differentiation depends on the reconstruction model. The ordered Mk model fits best (AIC 72.9; all-rates-different 76.4; equal-rates 79.6) and, like the all-rates-different and 1:1 parsimony reconstructions, places one gain at the base of the *Eudorina–Volvox–Pleodorina* clade followed by losses within *Eudorina*, giving 6 raw origins and K = 3 design contrasts, whereas the equal-rates and loss-penalised parsimony reconstructions give 8 to 10 raw origins and K = 4 to 5 (Figure 4B). Raw origins are connected foreground components under the threshold rule, not counts of trait gains, and are not directly comparable with the four to six origins of Lindsey et al. (2021) or the five of Lindsey et al. (2024). The primary scenario met its own pre-specified design-count threshold (K ≥ 3); the combined origin-count criterion failed because the 1:1 parsimony scenario had three design contrasts rather than the required four. After the strict gene filter the primary scenario retains 2 usable contrasts, *Astrephomene* against *Gonium* (102 genes) and *Volvox* section *Volvox* against the *Chlamydomonas* group (104 genes), the latter being a comparison three nodes above the foreground component rather than with a direct sister. Thus, the fitted primary model uses fewer contrasts than the design-count threshold, despite S2 meeting that threshold before gene filtering. These two somatic-differentiation contrasts are distinct from the two multicellularity foreground groups in Section 3.2.

The numerically validated primary refit on all 206 recorded observations gives μ = +0.070 (95 % equal-tailed credible interval -0.333 to 0.454, corresponding to rate ratios 0.72 to 1.57; P(μ < 0) = 0.21). Applying the length and per-gene tip-count filters together leaves 160 observations and gives μ = +0.068 (-0.331 to 0.449; P(μ < 0) = 0.21). Separate-filter results are reported in Table S3g. All five accepted primary and filter/Δrad refits have zero post-warm-up divergences; their maximum $\hat{R}$ is 1.001, and the minimum bulk and tail effective sample sizes across monitored parameters are 3,580 and 2,255. These diagnostics address numerical sampling, not the biological validity of the comparison. The two usable contrasts provide limited information about variation among independent origins.

For descriptive comparison, the equal-rates reconstruction adds two contrasts to the primary pair: the *V. aureus–P. japonica–V. carteri* group against *Eudorina* A (33 genes; weighted median r +0.327), and *P. starrii* + *P. indica* against a background including differentiated *V. powersii* and *V. gigas* (79 genes; weighted median r -0.189). Figure 4 distinguishes these descriptive contrast summaries from the validated S2 posterior estimates. The audit found divergent transitions in all 54 archived scenario fits (7,761 total). Their replacements and the prior/leave-one-contrast-out sensitivities give 58 accepted entries from 34 distinct model inputs, with zero divergences, maximum $\hat{R}$ 1.0023 and minimum bulk/tail ESS 2,865/1,567. All corresponding hierarchical intervals include zero (Table S7a); the historical chains are retained only as an audit trail. The validated primary Δrad refit gives μ = -0.008 (-0.350 to 0.340; P(μ < 0) = 0.60), also inconclusive. These conserved loci and the small number of contrasts provide no precise genome-wide test. The analyses do not provide clear support for a slowdown, and do not establish its absence.

The completed rate-MCMC sensitivity used 16,000 maps and yielded 5 design contrasts under rule v1 and 4 under v1.1. The largest branch-occupancy MCSE was 0.0051; 0 branches had Monte Carlo uncertainty intervals crossing a classification threshold. The rule-v1 strict r fit on 405 observations from 5 retained contrasts gives μ = +0.089 (-0.135 to 0.305; P(μ < 0) = 0.160), conditional on the rate prior and thresholded design. The resulting contrasts, gene statistics and other posterior fits are reported in Table S7b. Fixed-Q reconstruction on the relative-time chronogram produces 3 design contrasts for ARD under rule v1; ER, ARD and ordered results and their downstream fits are reported in Table S7f. With a narrower contrast-scale prior, μ was +0.073 (-0.174 to 0.318); the wider prior gave +0.069 (-0.541 to 0.651). Removing either contrast also left wide intervals. With only two retained contrasts, information about between-contrast variation is limited, and the interval for μ is sensitive to the prior on σ_k. The interval is conditional on that prior and provides limited precision for generalization across origins.

All 200 gene-cluster bootstrap replicates passed the diagnostic criteria. Their mean estimate was +0.069, with a percentile interval of 0.012 to 0.134 (Figure S3). This interval lies above zero, indicating stable positive estimates when genes are resampled within these two fixed contrasts. It does not quantify uncertainty across origins and does not override the wide hierarchical credible interval. Gene-resampling stability therefore does not establish a general effect of differentiation.

3.5 The cell-number gradient cannot be separated from sampling depth

Recomputed from unconstrained gene trees, the ten-species regression of log median terminal branch length on log2 cell number remains negative (β = -0.116, $P = 0.002$; without *C. reinhardtii*, β = -0.128, $P = 0.006$; per-orthogroup correlation negative in 93.9 % of 1,456 orthogroups for the ten species, permutation $P = 0.0155$, and in 90.2 % without *C. reinhardtii*, permutation $P = 0.0299$), so the association is not itself an artefact of the constrained trees; the diagnostic regression on the original free-ratio ω is flat (β = 0.002, $P = 0.94$). The ten-species panel contains one unicellular species and single representatives of the early-diverging *Tetrabaena* and *Gonium*, whose terminal branches span the whole time since those lineages last split from a sampled relative, while the *Volvox* species are sampled within a recent radiation and have short terminal branches. Because cell number increases along the same axis as sampling depth, a regression of terminal branch length on cell number cannot separate the two (Figure S1C); the time component was not quantified, and the observed slope is not thereby explained in full. On the 57-OTU tree, where sampling is denser across the gradient and the gene set, tree and cell-number coding differ, no significant cell-number association was detected in any of the four variants: pre-specified variant A, β = -0.04 ($P = 0.69$, $n = 57$); post hoc variant C, with terminal branches measured on gene trees pruned to the selected OTUs, β = +0.08 ($P = 0.32$, $n = 57$, $\hat{\lambda} = 0.86$); variant D, β = +0.05 ($P = 0.15$, $n = 55$) (Figure S1B; Supplementary Table S3d). For differentiation state as the predictor, variant D gives β = +0.29 with $P = 0.050$, the other variants larger P; variant A had been affected by the sampling problem it was meant to address, six representatives having near-zero terminal branches because a conspecific strain remained in the gene trees.

The completed matched-root ASTRAL sensitivity likewise gave no significant cell-number coefficient in the joint representative-pruned model (variant C): β = -0.090, $P = 0.481$. The corresponding ML result was β = -0.087, $P = 0.499$. Both installed and locally consistent chronos implementations, the archived ML reference and all variants are reported in Table S7c; these checks do not turn terminal divergence into a direct rate measurement.

4. How the errors arose and how they were found

This section is written for readers of versions 1 and 2. It records, for each error, where it was, what contributed to it as far as the record shows, what signs of it were visible at the time, and how it was found. The errors fall into three groups: the inputs to the analysis programs, the design of the comparisons, and reporting. The genome accessions, orthogroups and alignments were not changed by the re-analysis, though their quality was not reassessed either.

4.1 Timeline

Version 1 was posted on 3 July 2026 and the manuscript was submitted to a journal on 11 July; version 2, which corrected one figure and one abstract value, was posted on 16 July. The journal rejected the manuscript on 9 September on the reviewer's finding that the species-tree topology had been imposed rather than tested and that a single extant species stood in for the unicellular condition. On 11 September, while writing the specification of the first re-analysis (gate 1), I read the original codeml output files and found codeml's rooted-tree warning and a zero-length root-adjacent branch; the specification therefore required unrooted input and the recording of branch counts and warnings for every run, and the finding is recorded in the specification. Gate 1 and the differentiation contrasts (gate 2) ran the same day; the decomposition of the residual ω difference (gate 3) ran on 12 September, when the Zenodo record was also annotated. This version was prepared in the following week.

4.2 Errors in the inputs to the analysis programs

A rooted tree was passed to codeml. The script that prepared the branch-model runs wrote the species tree as a rooted binary tree with *C. reinhardtii* as the outgroup. codeml accepts such a tree and prints "This is a rooted tree. Please check!" in every output file; the original parser extracted the ω estimates and the log-likelihood but did not extract or flag the warning. With different ω classes on the two root-adjacent branches the model is formally identifiable, so a rooted input is not wrong in general; here the fits frequently reached boundary estimates (Section 3.1). Two signs were visible in versions 1 and 2 and were not pursued: the bimodal distribution of the *Chlamydomonas*-branch ω in Figure 1A, which I described but did not explain, and the collapse of the *Chlamydomonas* ω to zero in 930 orthogroups of the free-ratio analysis, which I attributed in the text to synonymous saturation without checking the synonymous divergence of that branch. The reviewer's comment on the bimodality was correct, and the explanation offered in versions 1 and 2 was unsupported.

Gene trees were constrained to the species tree. The amino-acid branch-length analysis ran IQ-TREE with a topology constraint so that every gene tree matched the imposed species tree. Constrained trees have legitimate uses (this version uses them for branch-length estimation on a published topology in gate 2), but they cannot test the topology they were constrained to, and the species tree itself was never inferred from the ten-genome data. An unconstrained run of the concatenated alignment supports a different arrangement (Section 3.2).

4.3 Errors in the design of the comparisons

One reference lineage. The ten-genome panel contained one unicellular species, so every foreground class was compared with the *C. reinhardtii* branch. Public genomes of three unicellular relatives (*C. incerta*, *C. schloesseri*, *Edaphochlamys debaryana*) existed and were not used. With one reference branch, any peculiarity of that branch, in synonymous divergence, codon usage

or saturation, becomes the "unicellular" value. Section 3.3 shows that the residual ω difference accompanies a synonymous-divergence difference; a second reference lineage would test the sensitivity of the comparison to the reference.

Two origins pooled. Versions 1 and 2 cited Lindsey et al. (2021) for two independent origins of multicellularity and then treated all colonial branches as one foreground class. Pooling is not itself illegitimate; the error was to generalise from the pooled result without testing whether each origin showed the pattern. Separated, one origin does not (Section 3.2).

Terminal branch lengths regressed on cell number. The gradient analysis was inherited from earlier seven- and eight-species versions of the pipeline and was read as a rate gradient. A terminal branch length is rate × time since the lineage's last sampled split, and in a panel with one unicellular species and single representatives of the early-diverging genera that time increases toward the unicellular end of the gradient. The confound is visible in the branch lengths themselves (Figure S1A) but was not recognised until the same regression was run on a densely sampled tree (Section 3.5). The pre-specified version of that regression then had a problem of the same kind, six representative strains with near-zero terminal branches, which was found only when the figure was drawn; the post hoc variants and the original are all reported.

4.4 Errors in reporting

Supplementary Tables S5b and S5c of versions 1 and 2 were posted with "To be computed" placeholders; the record does not establish why. The free-ratio result was explained by saturation without evidence. Version 2 corrected the abstract value 73.1 % to 60.6 % in the PDF, but the abstract shown on the record page and registered with Crossref kept the earlier value for two months; the available record does not establish why the metadata were not updated.

4.5 Why the pre-submission checks did not catch them

Before submission the numbers in the manuscript were compared with the pipeline output (the change log of versions 1 and 2 records a staged pre-manuscript check; no complete number-by-number log is preserved). Whatever those checks caught, they compared text with output; the output was faithful to a problematic input. What was missing was a second kind of check, on the inputs and on the diagnostics that the analysis programs emit: the number of branch parameters implied by the tree, the warnings printed by codeml, whether a topology constraint was in force, and whether an imposed topology agreed with an unconstrained estimate. None of these is a universal invariant, since a rooted model or a constrained tree can be the intended specification; each is a diagnostic whose disagreement with the intended model should be recorded and resolved. The three re-analyses recorded these diagnostics and evaluated them against their specified criteria.

4.6 What would have stopped each error

Table 2 summarises the errors with the sign that was visible at the time, the step that found it, and the check adopted in this version. Two general practices follow. First, an analysis pipeline should surface the diagnostics its programs emit, record how each disagreement with the intended model was resolved, and check that the root and branch labels supplied correspond to the model intended. Second, a comparison that rests on one reference lineage, one imposed topology, or terminal branch lengths should be tested against the alternative (a second reference, an unconstrained estimate, a sister-clade contrast) before it is interpreted.

5. Discussion

5.1 What is not supported

None of the four headline interpretations of versions 1 and 2 is supported by the corrected comparisons. The pooled branch-model fraction falls from 88.6 % to 71.5 % under an unrooted specification, by an amount this analysis does not partition between model specification, optimisation behaviour and orthogroup selection (Section 3.1), and the remaining majority is carried by one of the two origins; the lower ω at that origin accompanies greater synonymous divergence without a lower median nonsynonymous path length; the functional-category direction survives recomputation on the corrected fits, but it re-counts the same per-orthogroup ω values rather than providing twenty independent tests, and it is not shared by the Tetrabaenaceae origin (8 of 20 categories); and the terminal-branch gradient, although numerically reproducible, is confounded by sampling depth. The disposable-soma interpretation therefore has no support from these data. Because the theory concerns resource allocation rather than substitution rates, and because dN/dS against a single reference lineage does not measure maintenance investment, the data do not refute it either; they show that the comparisons used in versions 1 and 2 could not have tested it.

5.2 What the corrected data show, and what they do not

Two observations remain. First, along the paths from the basal node to the colonial tips, synonymous divergence is greater than on the *C. reinhardtii* branch in most orthogroups, and the median nonsynonymous path length is not smaller; the synonymous excess is present in the colonial species without somatic cells, whereas their nonsynonymous path ratios lie on either side of one. Second, lineages with somatic differentiation are not consistently slower than the background tip sets defined by the threshold rule on the public dataset, with the resolution limits stated in Section 3.4. Neither observation identifies a mechanism. The decomposition in Section 3.3 is an exact rearrangement of one model fit, not an independent measurement: it shows that the lower ω is not accompanied by a lower median nonsynonymous path, but it does not show why the synonymous divergence differs. Candidate explanations, all untested here, are a longer elapsed time on the colonial paths under a different rooting (under the coalescent-based arrangement of Lindsey et al. 2024 the two paths are not of equal duration); a higher per-year mutation supply in the colonial lineages; a lower synonymous rate on the *C. reinhardtii* branch from selection on codon usage, which is strong in this species (Naya et al. 2001; Popescu et al. 2006); residual saturation or model misspecification on the long reference branch; and gene-model, orthology or alignment errors, which were not independently reassessed. Neither the presence nor the direction of bias from these input errors was established. Distinguishing these possibilities would require validation of gene models, orthology and codon alignments, several unicellular reference lineages, explicit models of synonymous-site selection and codon-composition change, and direct mutation-rate estimates for colonial species. “Lower dN/dS” and “stronger purifying selection” are not synonyms when the denominator differs between lineages, and a higher dN path does not establish relaxed selection: a lineage with a higher mutation supply and stronger selection on nonsynonymous sites could show both a higher dN and a lower ω.

5.3 Design errors and how to avoid them

Three problems undermine the earlier interpretations. (i) Boundary estimates in a rooted branch model were not investigated. The model is formally identifiable, so a rooted input is not always wrong; in these data the fits were unstable, and supplying unrooted trees, recording the number of branch parameters and surfacing codeml’s warnings makes the intended model explicit and

its diagnostics visible. (ii) Terminal branch lengths were regressed on a trait. Because a terminal branch is rate × time since the last sampled split, its length depends on taxon sampling, and a trait that co-varies with sampling depth produces a slope without any rate difference. Comparing sister clades from a shared ancestor, or estimating rates on a time-calibrated tree, avoids this. (iii) Constraint was inferred from ω against a single reference lineage. With one unicellular species, any peculiarity of its branch becomes the "unicellular" value; decomposing ω into its numerator and denominator, and using several reference lineages, is required before ω differences are read as differences in selection.

5.4 Relation to earlier work

Hu et al. (2019) compared chloroplast substitution patterns (55 protein-coding genes) between unicellular and colonial taxa, using *Chloromonas radiata* as the reference for pairwise distances and comparing dN separately. Hu et al. (2020) compared nuclear-gene evolution (105 single-copy orthologues) among colonial taxa, including a branch model with *Tetrabaena* as the foreground, and used *C. reinhardtii* as the reference for pairwise distances. Neither study is a test of the comparison examined here, and the present result neither confirms nor contradicts them; it shows that, for the nuclear genome and the *C. reinhardtii* reference, the ω difference between colonial and reference branches accompanies a synonymous-divergence difference, which is a reason to decompose any comparison built on a single reference lineage before interpreting it. Lindsey et al. (2021) inferred four to six origins of somatic differentiation and Lindsey et al. (2024) five; the reconstructions here yield 2 to 10 connected foreground components depending on the model and threshold, which is a different quantity and not a reproduction of those counts, though it makes the same point that the number of independent origins available for a comparative test depends on the model of trait evolution.

5.5 Limitations

The primary sister-clade analysis has only 2 usable contrasts, prior-sensitive posterior intervals and a conserved gene set. Some comparisons are not between direct sister clades, and background tip sets can contain differentiated taxa under the threshold rule. Cell numbers and some differentiation assignments use genus-level information; the chronogram is a penalised-likelihood substitute. Numerical validation resolved the sampling diagnostics for the primary and filter/Δrad refits, but does not validate the Gaussian model, likelihood weights or biological design; no simulation calibration was performed. The full scenario and primary sensitivity refits now pass the sampling diagnostics, while the failed historical chains remain documented. The three initially omitted analyses and two omitted filters were completed post hoc. Rate-MCMC sensitivity still thresholds an averaged reconstruction, and the bootstrap conditions on the two retained contrasts. The chronos objective–gradient mismatch required a separately labelled local sensitivity; convergence and gradient checks do not independently calibrate the relative-time proxy. The codon-path decomposition can be interpreted as a rate comparison only under the stated rooting, and the ten-genome data have weak basal gene-tree support and cannot determine that root. These limitations restrict inference about biological effects. The correction withdraws the earlier interpretation because its comparisons do not establish it; it does not establish an absence of selection differences.

6. Version history

Version 1 (posted 3 July 2026) reported that 73.1 % of orthogroups had shorter aggregated multicellular amino-acid branches. Version 2 (posted 16 July 2026) corrected this figure to 60.6 % (882/1,456) in the manuscript PDF, but the abstract displayed on the preprint record and the abstract registered with Crossref retained the version-1 value until this version was prepared; the corrected value therefore did not reach readers who consulted the record page or the DOI metadata rather than the PDF. Both figures are in any case superseded by this version, because the comparison on which they rest is confounded by the time since each lineage's last sampled split (Section 3.5). Version 3 (this version) replaces the title, abstract, results and conclusions; Table 1 lists every headline value of versions 1 and 2 with its status. The earlier Zenodo data version (10.5281/zenodo.21344531) carries a file-level validity manifest for the original deposit; the exact materials for this correction are collected in an independent code and data archive (`volvocine_v3_reproduction.zip`), identified below.

Data and code availability

Original data and earlier reanalyses are at Zenodo version DOI 10.5281/zenodo.21344531. The independent correction archive `volvocine_v3_reproduction.zip` (version DOI `10.5281/zenodo.22782522`) is prepared as a new version under the same concept DOI `10.5281/zenodo.21094822`. It contains derived inputs, results, specifications, diagnostics, figure data, `numbers.json`, scripts, machine-checked proofs and toolchain information. The manuscript and finished figures are supplied separately for the EcoEvoRxiv preprint. `REPRODUCE.md` describes regenerating summaries and figures locally from saved results, separately from rerunning analyses; the manifest identifies the contents. Lindsey et al. (2024) source alignments are at Dryad 10.5061/dryad.mcvdnck6b. Individual alignments are not duplicated; derived lengths and the concatenated ASTRAL subset are included. Original per-orthogroup codeml files remain in the earlier Zenodo archive; derived diagnostic scans are included here.

Acknowledgements

I thank the anonymous reviewer whose comment on the imposed topology prompted this reanalysis. The errors in versions 1 and 2 are mine.

AI-assistance disclosure

AI tools assisted with drafting analysis scripts, mathematical checks, figures and manuscript text for all three versions. The analyses were executed under the author's direction; scripts, results and numerical checks are documented in the archive. AI-assisted code in versions 1 and 2 contained the input choices discussed in Section 4; the available evidence does not isolate AI assistance as their cause, and the author's pre-submission checks did not examine those inputs. For version 3, Claude (Anthropic) assisted with drafting and OpenAI Codex with review, literature checks, document revision, implementation and validation of the exact marginal computation, and reproduction testing; the Methods distinguish pre-specified analyses, amendments and post hoc checks. The AI-assisted reviews and the machine-checked identities verify arithmetic and internal consistency; they are not an independent verification of the biological conclusions. The author is responsible for the interpretations and final submitted content.

Table 1. Headline statements of versions 1 and 2 and their status

Statement (versions 1–2)	Corrected diagnostic (this version)	Status of the interpretation
ω_fg < ω_bg in 88.6 % of 845 orthogroups; median ω_bg 0.117, ω_fg 0.046	unrooted: 71.5 % of 1,374; median ω0 0.074, ω1 0.05; by origin (common set) 48.4 % (Tetrabaenaceae) and 73.5 % (Goniaceae + Volvocaceae)	model-dependent: not reproduced under an unrooted specification; a lower ω persists for one origin
Lower ω interpreted as stronger constraint	median per-orthogroup ω ratio 0.63; median dN path ratio 1.25; median dS path ratio 1.96	interpretation not supported: the ω difference accompanies greater synonymous divergence, which the comparison cannot attribute to selection or to mutation supply
Signal across all 20 COG categories	majority lower ω remains in all 20 H0u/origin-B categories; in 8 of 20 *Tetrabaena* categories (own QC)	pooled direction retained descriptively; not shared by both origins and not evidence for the proposed mechanism (Table S7e)
Bimodal *Chlamydomonas* ω (Figure 1A of v1–2) attributed to dS saturation	111 orthogroups with ω_bg > 1 in the rooted fits, 93 with a *Chlamydomonas* branch ≤ 0.001; 6 with ω_bg > 1 unrooted	attribution not supported: bimodality associated with boundary estimates of the rooted fits
60.6 % of orthogroups with shorter multicellular amino-acid branches	terminal-branch comparison; no consistent slowdown in sister-clade contrasts	interpretation not supported: a reduced rate cannot be inferred from terminal branches (sampling-depth confound)
PGLS β = -0.123, P = 0.002 (n = 10); 96.1 % negative per-orthogroup correlations	reproduced from unconstrained trees (β = -0.116, P = 0.002); on 57 OTUs no significant association (β = +0.08, P = 0.32)	numerically reproduced; interpretation as a rate gradient not supported
Species tree with Goniaceae monophyletic	unconstrained trees support *Astrephomene* + Volvocaceae (aLRT/ultrafast bootstrap 100/100); gene concordance 33.3 % at the basal branch, 21.5 % at the *Astrephomene* + Volvocaceae branch	model-dependent: imposed topology not supported by the ten-genome data; gene-level support weak either way
Free-ratio *Chlamydomonas* ω estimable in 353 orthogroups “because of dS saturation”	ω → 0 in 930 orthogroups of the rooted fits	attribution not supported

Table 2. Problems in versions 1 and 2 and checks adopted

Error	Where	Documented contributing factor	Sign visible in versions 1–2	How it was found	Check adopted in this version
Boundary estimates in a rooted branch model were interpreted without diagnostic follow-up	branch-model script	tree written with *C. reinhardtii* as outgroup; warning not surfaced by the parser	"rooted tree" warning in every output; bimodal ω_bg (Fig. 1A of v1–2); *Chlamydomonas* ω = 0 in 930 free-ratio fits	reading the original outputs while specifying gate 1	tree supplied in the form the intended model requires (here unrooted); branch counts and warnings recorded and evaluated
Species tree imposed, gene trees constrained	IQ-TREE scripts	literature topology adopted without an unconstrained estimate; gene trees forced to match	none in the constrained output by construction	reviewer; unconstrained ML and ASTRAL in gate 1	unconstrained trees for topology; gCF/sCF; constrained trees only for branch-length estimation on a stated topology
Two origins pooled	branch labels	design did not follow the cited phylogeny	—	three-ratio model in gate 1	pooled and origin-specific models compared
Terminal branch length read as a rate	gradient analysis	statistic inherited from earlier versions; confound not recognised	long terminal branches of the sparsely sampled lineages (Fig. S1A)	sister-clade contrasts and a denser tree (gate 2)	path lengths from a shared ancestor; confound stated
Single reference lineage	panel design	one unicellular genome in the panel; public relatives unused	—	decomposition of ω into dN and dS paths (gate 3)	decompose ω; sensitivity to the reference to be tested with additional lineages
"dS saturation" explanation of the free-ratio result	text of v1–2	explanation supplied without checking dS	—	gate 1 diagnostics	check the quantity before naming the cause
Placeholder supplementary tables	Tables S5b–c	undetermined	"To be computed"	reviewer	supplementary tables generated from the manifest; missing keys stop the build
Version-2 abstract not propagated	record page and Crossref	undetermined	73.1 % on the record page after the update	this version	verify PDF, record page and DOI metadata after every update

Figures and legends

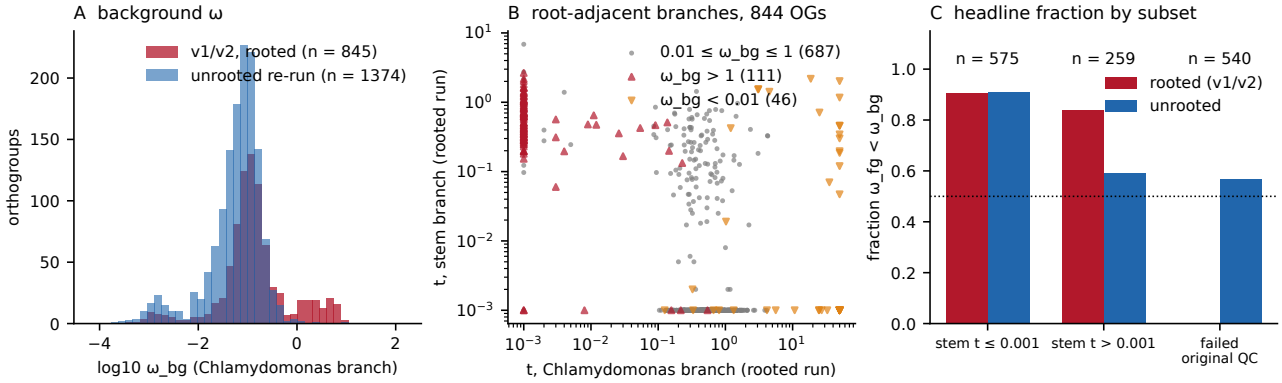

Figure 1. The original rooted fits and the unrooted re-run. (A) Distribution of background (*Chlamydomonas*-branch) ω in the original rooted run (845 orthogroups passing quality control) and in the unrooted re-run (1,374). (B) Lengths assigned by the original rooted run to the two root-adjacent branches, for the 844 retained orthogroups in the stored diagnostic table, coloured by background ω; 93 of the 111 orthogroups with ω_bg > 1 have a *Chlamydomonas* branch of at most 0.001. (C) Fraction of orthogroups with ω_fg < ω_bg in the rooted and unrooted runs, for retained orthogroups whose rooted stem branch was at most 0.001, retained orthogroups with a longer stem, and orthogroups that failed the original quality control and pass the unrooted one.

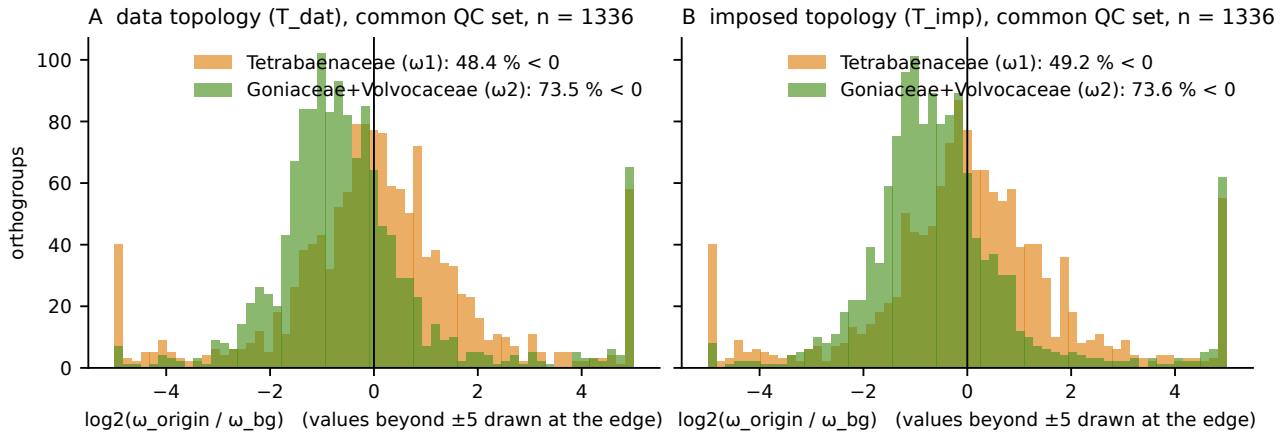

Figure 2. The two origins of multicellularity as separate foreground classes. Distribution of log2(ω_origin/ω0) for the *Tetrabaena* branch (ω1) and the Goniaceae + Volvocaceae clade (ω2) on (A) the data topology and (B) the imposed topology, on the common set of 1,336 orthogroups; values beyond ±5 are drawn at the edge.

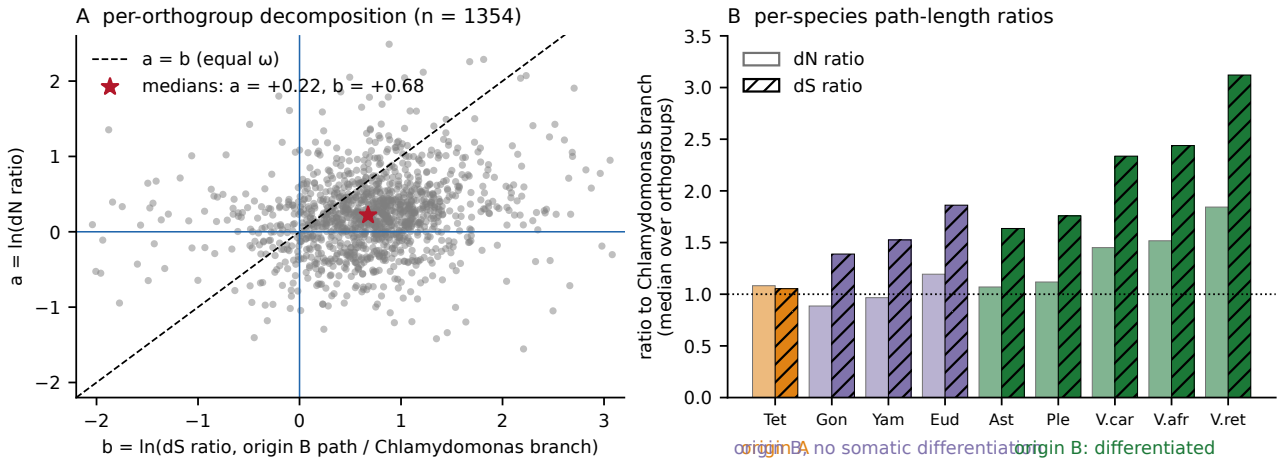

Figure 3. Decomposition of the residual ω difference. (A) For each orthogroup, a = ln(dN_B/dN_chl) against b = ln(dS_B/dS_chl); the dashed line marks equal ω, and a − b = ln(ω2/ω0) within each orthogroup. (B) Per-species median ratios of nonsynonymous and synonymous path length to the *C. reinhardtii* branch; within an orthogroup all Goniaceae + Volvocaceae paths share the fitted ω2/ω0.

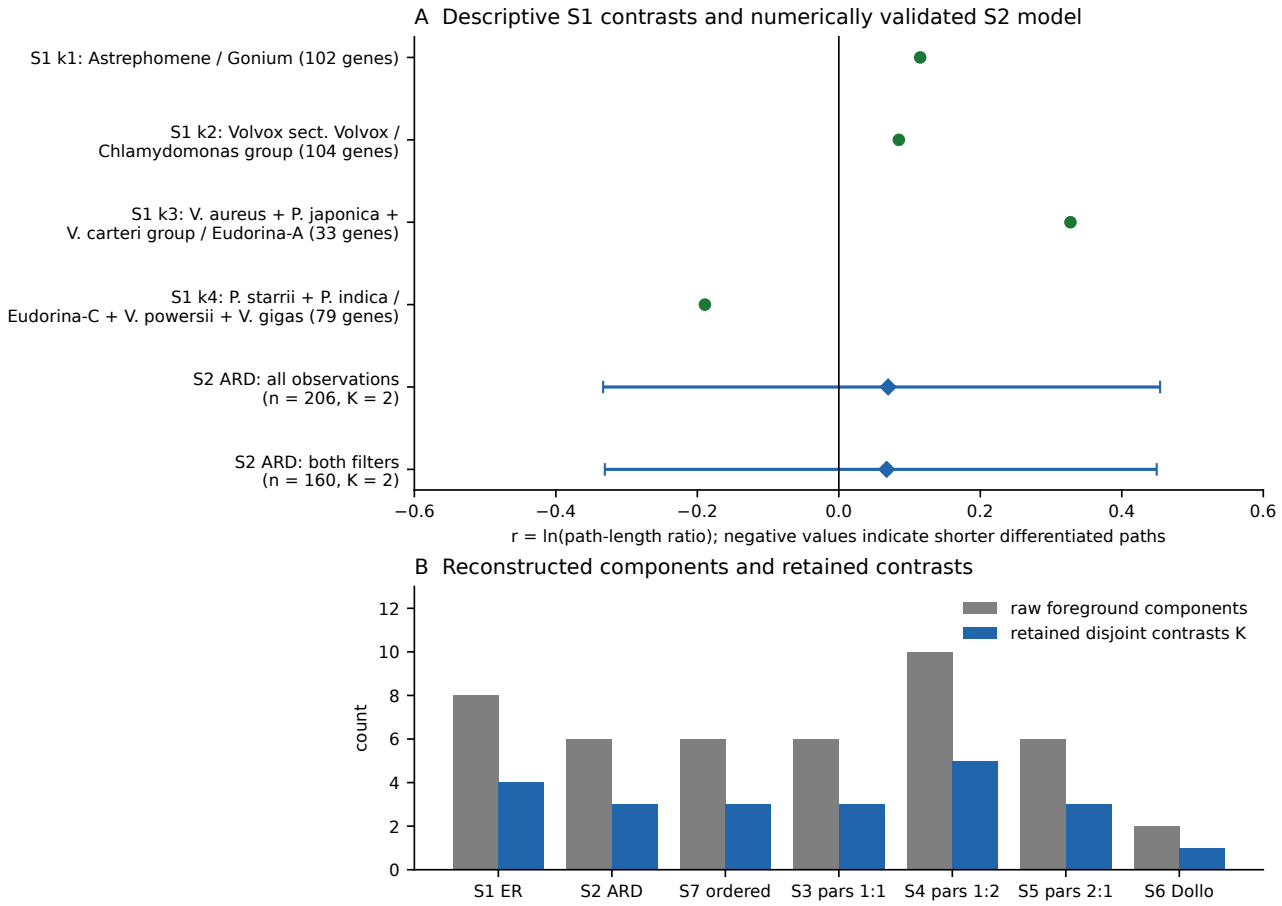

Figure 4. Somatic-differentiation comparisons. (A) Descriptive weighted medians of r for the four contrasts under the equal-rates reconstruction (circles), and validated S2 posterior means of μ with 95 % credible intervals for all recorded observations and for both filters applied post hoc (diamonds). The primary S2 analysis uses only the first two contrasts. Negative r indicates a shorter foreground path, without isolating the period after differentiation evolved. (B) Connected foreground components and design contrasts under the seven reconstruction scenarios; component counts are not counts of inferred trait gains.

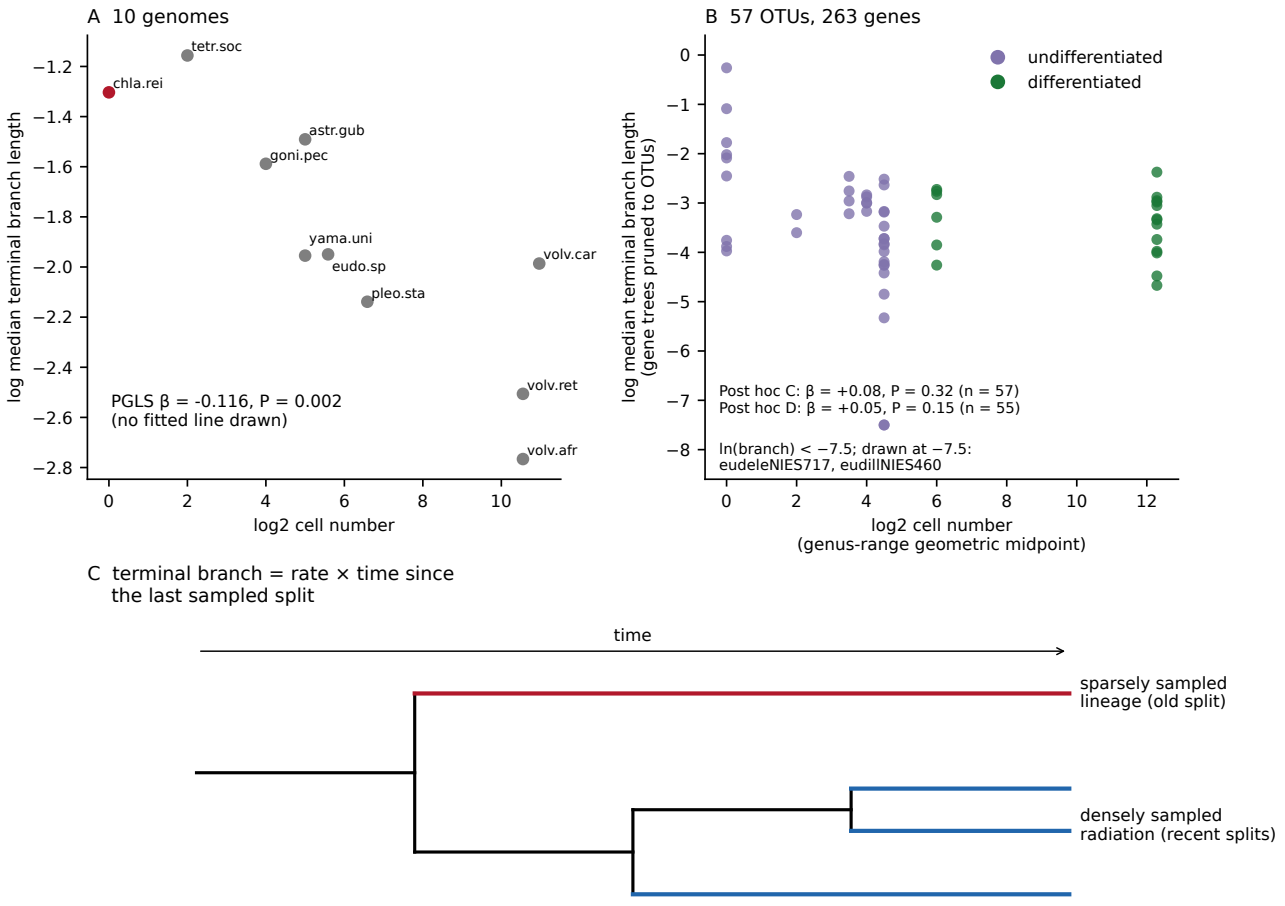

Figure S1. Terminal branch length and sampling depth. (A) Ten genomes: log median terminal branch length against log2 cell number (no fitted line; the PGLS estimate is given in the panel). (B) 57 OTUs on the 263-gene dataset, using genus-range geometric midpoints for cell number; terminal branches measured after pruning gene trees to the selected OTUs, including all seven *E. elegans* strains (variant C); two representatives whose log median terminal branch lengths are below the plotting limit are drawn at the floor indicated in the panel; PGLS estimates for variants C and D are given in the panel. (C) Schematic: a sparsely sampled lineage has a long terminal branch because its last sampled split is old, not because its rate is high.

Supplementary material

Table S1, species and assemblies (unchanged from versions 1 and 2). Table S2, branch-model metrics for every run and set, topology support and bipartition frequencies. Table S3, Stage 0 origins, hierarchical-model estimates under every scenario and rule, the tips of every contrast, convergence and sensitivity checks, and the terminal-branch regressions. Table S4, per-species dN and dS path ratios. Note S5, the identities used in Sections 2.5 and 2.7 with their machine checks (SymPy symbolic residuals, z3 counterexample search) and the Lean proof of the path-sum ratio identity. Table S6, a reconciliation of Methods statements with their implementation and departures from the specifications. Supplement S7, completed numerical audits and scenario refits, rate-MCMC sensitivity, ASTRAL/chronogram comparisons, gene-cluster bootstrap, corrected COG tallies and diagnostic Figures S2–S4.

References

Álvarez-Carretero S, Kapli P, Yang Z. 2023. Beginner's guide on the use of PAML to detect positive selection. Mol Biol Evol 40:msad041. doi:10.1093/molbev/msad041 Featherston J, Arakaki Y, Hanschen ER, Ferris PJ, Michod RE, Olson BJSC, Nozaki H, Durand PM. 2018. The 4-celled *Tetrabaena socialis* nuclear genome reveals the essential components for genetic control of cell number at the origin of multicellularity in the volvocine lineage. Mol Biol Evol 35:855–870. doi:10.1093/molbev/msx332 Hamaji T, Kawai-Toyooka H, Uchimura H, Suzuki M, Noguchi H, Minakuchi Y, Toyoda A, Fujiyama A, Miyagishima S, Umen JG, Nozaki H. 2018. Anisogamy evolved with a reduced sex-determining region in volvocine green algae. Commun Biol 1:17. doi:10.1038/s42003-018-0019-5 Hanschen ER, Marriage TN, Ferris PJ, Hamaji T, Toyoda A, Fujiyama A, Neme R, Noguchi H, Minakuchi Y, Suzuki M, Kawai-Toyooka H, Smith DR, Sparks H, Anderson J, Bakarić R, Luria V, Karger A, Kirschner MW, Durand PM, Michod RE, Nozaki H, Olson BJSC. 2016. The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. Nat Commun 7:11370. doi:10.1038/ncomms11370 Hu Y, Xing W, Song H, Zhu H, Liu G, Hu Z. 2019. Evolutionary analysis of unicellular species in Chlamydomonadales through chloroplast genome comparison with the colonial volvocine algae. Front Microbiol 10:1351. doi:10.3389/fmicb.2019.01351 Hu Y, Xing W, Song H, Hu Z, Liu G. 2020. Comparison of colonial volvocine algae based on phylotranscriptomic analysis of gene family evolution and natural selection. Eur J Phycol 55:100–112. doi:10.1080/09670262.2019.1663269 Lindsey CR, Rosenzweig F, Herron MD. 2021. Phylotranscriptomics points to multiple independent origins of multicellularity and cellular differentiation in the volvocine algae. BMC Biol 19:182. doi:10.1186/s12915-021-01087-0 Lindsey CR, Knoll AH, Herron MD, Rosenzweig F. 2024. Fossil-calibrated molecular clock data enable reconstruction of steps leading to differentiated multicellularity and anisogamy in the volvocine algae. BMC Biol 22:79. doi:10.1186/s12915-024-01878-1 Naya H, Romero H, Carels N, Zavala A, Musto H. 2001. Translational selection shapes codon usage in the GC-rich genome of *Chlamydomonas reinhardtii*. FEBS Lett 501:127–130. doi:10.1016/S0014-5793(01)02644-8 Nozaki H, Mahakham W, Heman W, Matsuzaki R, Kawachi M. 2022. Morphology, mating system and taxonomy of *Volvox africanus* (Volvocaceae, Chlorophyceae) from Thailand. Bot Stud 63:1. doi:10.1186/s40529-022-00332-1 Popescu CE, Borza T, Bielawski JP, Lee RW. 2006. Evolutionary rates and expression level in *Chlamydomonas*. Genetics 172:1567–1576. doi:10.1534/genetics.105.047399 Yamamoto K, Hamaji T, Kawai-Toyooka H, Matsuzaki R, Takahashi F, Nishimura Y, Kawachi M, Noguchi H, Minakuchi Y, Umen JG, Toyoda A, Nozaki H. 2021. Three genomes in the algal genus *Volvox* reveal the fate of a haploid sex-determining region after a transition to homothallism. Proc Natl Acad Sci USA 118:e2100712118. doi:10.1073/pnas.2100712118 Yamashita S, Yamamoto K, Matsuzaki R, Suzuki S, Yamaguchi H, Hirooka S, Minakuchi Y, Miyagishima S, Kawachi M, Toyoda A, Nozaki H. 2021. Genome sequencing of the multicellular alga *Astrephomene* provides insights into convergent evolution of germ-soma differentiation. Sci Rep 11:22231. doi:10.1038/s41598-021-01521-x Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591. doi:10.1093/molbev/msm088

Supplementary material for version 3

Numerical result summaries are rendered from `numbers.json`, which `build_numbers_manifest.py` regenerates from the accompanying result files and which records the source file of every value; species metadata (Table S1), values quoted from versions 1 and 2 and the contrast tip lists (Table S3e) have separately identified sources.

Table S1. Species, assemblies and gene-model sources (unchanged from versions 1 and 2)

Species	Strain	Cells	Family	Assembly	Gene models
Chlamydomonas reinhardtii	CC-503 cw92 mt+	1	Chlamydomonadaceae	GCF_000002595.2	Deposited RefSeq
Tetrabaena socialis	NIES-571	4	Tetrabaenaceae	GCA_002891735.1	Featherston et al. 2018
Gonium pectorale	—	16	Goniaceae	GCA_001584585.1	Hanschen et al. 2016
Astrephomene gubernaculifera	NIES-4017	32	Goniaceae	GCA_021605115.1	Deposited GFF
Yamagishiella unicocca	2012-1026-YU-F2-6 *MT+*	32	Volvocaceae	GCA_003116995.1	Augustus *de novo* (trained on *C. reinhardtii*)
Eudorina sp.	2010-623-F1-E4 (female, NIES-3984)	~48	Volvocaceae	GCA_003117195.1	Augustus *de novo* (trained on *C. reinhardtii*)
Pleodorina starrii	NIES-4479	~96	Volvocaceae	GCA_030267565.1	Deposited
Volvox carteri f. *nagariensis*	Eve	~2,000	Volvocaceae	GCF_000143455.1	Deposited RefSeq
Volvox africanus	NIES-4468 (1101-NZ-11, Thai)	~1,500	Volvocaceae	GCA_030268105.1	Deposited GFF (Nozaki et al. 2022)
Volvox reticuliferus	NIES-3785 (female)	~1,500	Volvocaceae	GCA_019650235.1	Deposited GFF (Yamamoto et al. 2021)

Table S2. Unrooted branch-model runs (gate 1)

Fraction of orthogroups with foreground ω below background (*Chlamydomonas*-branch) ω and median ω by class. ‘Own set’ = orthogroups passing quality control in that run; ‘common set’ = passing in all ten-species runs (1,336) or both nine-species runs (1,436). dS < 10 = *Chlamydomonas*-branch dS below 10.

run	topology	foreground classes	n (own)	frac ω1 < ω0	frac ω2 < ω0	median ω0 / ω1 / ω2	frac ω1 < ω0 (common)	frac ω2 < ω0 (common)	n (dS < 10)	frac ω1 < ω0 (dS < 10)	frac ω2 < ω0 (dS < 10)	branch parameters	rooted warnings
H0u	T_imp	all colonial (1 class)	1,374	0.715		0.074 / 0.05	0.72		1276	0.77		17	0
H3	T_dat	all colonial (1 class)	1,370	0.72		0.077 / 0.051	0.724		1275	0.773		17	0
H1	T_imp	origin A (*Tetrabaena*) / origin B (Goniaceae + Volvocaceae)	1,359	0.493	0.734	0.076 / 0.08 / 0.049	0.492	0.736	1260	0.523	0.792	17	0
H1p	T_dat	origin A (*Tetrabaena*) / origin B (Goniaceae + Volvocaceae)	1,357	0.486	0.736	0.077 / 0.082 / 0.049	0.484	0.735	1257	0.516	0.795	17	0
H2	T_imp	all colonial, *Tetrabaena* removed	1,444	0.746		0.076 / 0.05	0.745		1347	0.799		15	0
H2p	T_dat	all colonial, *Tetrabaena* removed	1,437	0.75		0.078 / 0.051	0.75		1346	0.8		15	0

Topology support for the Goniaceae + Volvocaceae clade (T_dat): ML LG+G4 (aLRT/UFBoot): 100/100; ML partitioned (UFBoot): 100.00; ASTRAL-III (local PP): 1.00. Goniaceae monophyletic in none of the three trees.

Bipartition frequencies across the 1,456 unconstrained gene trees with all ten species:

bipartition	frequency
GonVolv8	0.33
Goniaceae	0.173
Astre Volvocaceae7	0.214
Volvocaceae6	0.392
Chlamy Gon	0.215
Tet Gon	0.121
Tet Volvocaceae6	0.125
afr ret	0.787
Volvox3	0.488
EudPleVolvox	0.598

Concordance factors at the basal split (branch to Gon+Volv): gCF 33.3 %, sCF 39.0 %; at the *Astrephomene* + Volvocaceae branch: gCF 21.5 %, sCF 36.0 %.

Table S3. Somatic-differentiation contrasts (gate 2 / Stage 0)

S3a. Origins per reconstruction scenario. Raw origins is the historical column name for connected foreground components, not a count of independently established evolutionary gains; K = retained disjoint contrasts (rule v1 / rule v1.1).

scenario	raw origins	K (v1)	K (v1.1)
S1 equal rates	8	4	4
S2 all rates different (primary)	6	3	4
S7 ordered	6	3	4
S3 parsimony 1:1	6	3	4
S4 parsimony gain 1 : loss 2	10	5	4
S5 parsimony gain 2 : loss 1	6	3	4
S6 Dollo-like	2	1	2

Mk model fit (AIC): equal rates 79.6, all rates different 76.4, ordered 72.9.

S3b. Archived fits of the hierarchical model r_gk = μ + a_g + b_k + e_gk (and the same for Δrad). Rule v1 strict is the primary analysis under the pre-specified rules, with the departures listed in Table S6; rule v1.1 and the relaxed gene filter are sensitivity analyses added by the dated amendments (Methods 2.2). The model for r carries likelihood weights; the model for Δrad is unweighted. These are the recorded fits, including fits with unresolved sampling diagnostics. A small $\hat{R}$ alone does not establish their accuracy. They are retained to document the original decision tables; the numerical results used in the revised main text are the validated refits in S3g and S7a.

rule	scenario	statistic	K	n	μ	95 % equal-tailed credible interval	P(μ < 0)	contrasts with negative median	max $\hat{R}$
v1 strict	S1_ER	r	4	318	+0.071	-0.264 to 0.429	0.30	1 / 4	1.003
v1 strict	S1_ER	rad	4	204	-0.006	-0.027 to 0.016	0.79	2 / 4	1.002
v1 strict	S2_ARD	r	2	206	+0.077	-0.201 to 0.378	0.19	0 / 2	1.020
v1 strict	S2_ARD	rad	2	143	-0.009	-0.133 to 0.109	0.62	1 / 2	1.011
v1 strict	S7_ORD	r	2	206	+0.077	-0.201 to 0.378	0.19	0 / 2	1.020
v1 strict	S7_ORD	rad	2	143	-0.009	-0.133 to 0.109	0.62	1 / 2	1.011
v1 strict	S3_pars11	r	2	206	+0.077	-0.201 to 0.378	0.19	0 / 2	1.020
v1 strict	S3_pars11	rad	2	143	-0.009	-0.133 to 0.109	0.62	1 / 2	1.011
v1 strict	S4_pars_g1_l2	r	5	405	+0.090	-0.130 to 0.303	0.17	2 / 5	1.005
v1 strict	S4_pars_g1_l2	rad	5	204	-0.008	-0.028 to 0.013	0.81	3 / 5	1.004

rule	scenario	statistic	K	n	μ	95 % equal-tailed credible interval	P(μ < 0)	contrasts with negative median	max R̂
v1 strict	S5_pars_g2_l1	r	2	206	+0.077	-0.201 to 0.378	0.19	0 / 2	1.020
v1 strict	S5_pars_g2_l1	rad	2	143	-0.009	-0.133 to 0.109	0.62	1 / 2	1.011
v1 strict	S6_dollo	r	1	1	+0.350	-1.486 to 2.039	0.34	0 / 1	1.007
v1.1 strict	S1_ER	r	4	236	+0.037	-0.250 to 0.309	0.36	1 / 4	1.006
v1.1 strict	S1_ER	rad	3	115	+0.009	-0.031 to 0.100	0.42	1 / 3	1.106
v1.1 strict	S2_ARD	r	4	133	+0.113	-0.066 to 0.295	0.06	0 / 4	1.014
v1.1 strict	S2_ARD	rad	4	93	+0.002	-0.041 to 0.035	0.40	2 / 4	1.007
v1.1 strict	S7_ORD	r	4	133	+0.113	-0.066 to 0.295	0.06	0 / 4	1.014
v1.1 strict	S7_ORD	rad	4	93	+0.002	-0.041 to 0.035	0.40	2 / 4	1.007
v1.1 strict	S3_pars11	r	4	133	+0.113	-0.066 to 0.295	0.06	0 / 4	1.014
v1.1 strict	S3_pars11	rad	4	93	+0.002	-0.041 to 0.035	0.40	2 / 4	1.007
v1.1 strict	S4_pars_g1_l2	r	4	177	+0.090	-0.068 to 0.304	0.10	1 / 4	1.014
v1.1 strict	S4_pars_g1_l2	rad	3	92	+0.011	-0.095 to 0.091	0.25	1 / 3	1.059
v1.1 strict	S5_pars_g2_l1	r	4	133	+0.113	-0.066 to 0.295	0.06	0 / 4	1.014
v1.1 strict	S5_pars_g2_l1	rad	4	93	+0.002	-0.041 to 0.035	0.40	2 / 4	1.007
v1.1 strict	S6_dollo	r	2	21	-0.000	-0.502 to 0.703	0.59	0 / 2	1.149
v1 relaxed	S1_ER	r	4	482	+0.056	-0.207 to 0.318	0.28	1 / 4	1.007
v1 relaxed	S1_ER	rad	4	296	-0.003	-0.023 to 0.021	0.67	1 / 4	1.008
v1 relaxed	S2_ARD	r	3	320	+0.083	-0.046 to 0.211	0.08	0 / 3	1.011

rule	scenario	statistic	K	n	μ	95 % equal-tailed credible interval	P(μ < 0)	contrasts with negative median	max R̂
v1 relaxed	S2_ARD	rad	3	223	-0.002	-0.051 to 0.046	0.57	1 / 3	1.013
v1 relaxed	S7_ORD	r	3	320	+0.083	-0.046 to 0.211	0.08	0 / 3	1.011
v1 relaxed	S7_ORD	rad	3	223	-0.002	-0.051 to 0.046	0.57	1 / 3	1.013
v1 relaxed	S3_pars11	r	3	320	+0.083	-0.046 to 0.211	0.08	0 / 3	1.011
v1 relaxed	S3_pars11	rad	3	223	-0.002	-0.051 to 0.046	0.57	1 / 3	1.013
v1 relaxed	S4_pars_g1_l2	r	5	586	+0.086	-0.075 to 0.232	0.10	1 / 5	1.008
v1 relaxed	S4_pars_g1_l2	rad	5	291	-0.008	-0.035 to 0.012	0.81	2 / 5	1.009
v1 relaxed	S5_pars_g2_l1	r	3	320	+0.083	-0.046 to 0.211	0.08	0 / 3	1.011
v1 relaxed	S5_pars_g2_l1	rad	3	223	-0.002	-0.051 to 0.046	0.57	1 / 3	1.013
v1 relaxed	S6_dollo	r	1	45	+0.043	-0.766 to 0.748	0.39	0 / 1	1.012
v1 relaxed	S6_dollo	rad	1	40	+0.000	-0.564 to 0.553	0.49	0 / 1	1.033
v1.1 relaxed	S1_ER	r	4	418	+0.053	-0.216 to 0.381	0.30	1 / 4	1.035
v1.1 relaxed	S1_ER	rad	3	194	-0.002	-0.032 to 0.022	0.60	1 / 3	1.008
v1.1 relaxed	S2_ARD	r	4	338	+0.124	-0.061 to 0.332	0.07	1 / 4	1.008
v1.1 relaxed	S2_ARD	rad	4	264	-0.003	-0.036 to 0.029	0.59	2 / 4	1.006
v1.1 relaxed	S7_ORD	r	4	338	+0.124	-0.061 to 0.332	0.07	1 / 4	1.008
v1.1 relaxed	S7_ORD	rad	4	264	-0.003	-0.036 to 0.029	0.59	2 / 4	1.006
v1.1 relaxed	S3_pars11	r	4	338	+0.124	-0.061 to 0.332	0.07	1 / 4	1.008
v1.1 relaxed	S3_pars11	rad	4	264	-0.003	-0.036 to 0.029	0.59	2 / 4	1.006

rule	scenario	statistic	K	n	μ	95 % equal-tailed credible interval	P(μ < 0)	contrasts with negative median	max R̂
v1.1 relaxed	S4_pars_g1_l2	r	4	337	+0.053	-0.176 to 0.281	0.28	1 / 4	1.006
v1.1 relaxed	S4_pars_g1_l2	rad	3	166	-0.002	-0.040 to 0.034	0.56	2 / 3	1.009
v1.1 relaxed	S5_pars_g2_l1	r	4	338	+0.124	-0.061 to 0.332	0.07	1 / 4	1.008
v1.1 relaxed	S5_pars_g2_l1	rad	4	264	-0.003	-0.036 to 0.029	0.59	2 / 4	1.006
v1.1 relaxed	S6_dollo	r	2	146	+0.138	-0.154 to 0.457	0.11	0 / 2	1.006
v1.1 relaxed	S6_dollo	rad	2	120	-0.029	-0.317 to 0.174	0.50	1 / 2	1.315

S3c. Individual contrasts (rule v1 strict).

scenario	k	genes	weighted median r	fraction r < 0	median Δrad
S1_ER	1	102	+0.115	0.47	0.0
S1_ER	2	104	+0.085	0.49	-0.0168
S1_ER	3	33	+0.327	0.09	-0.0018
S1_ER	4	79	-0.189	0.7	0.0101
S2_ARD	1	102	+0.115	0.47	0.0
S2_ARD	2	104	+0.085	0.49	-0.0168
S4_pars_g1_l2	1	102	+0.115	0.47	0.0
S4_pars_g1_l2	2	104	+0.085	0.49	-0.0168
S4_pars_g1_l2	3	33	+0.327	0.09	-0.0018
S4_pars_g1_l2	4	150	-0.114	0.59	-0.0107
S4_pars_g1_l2	5	16	-0.054	0.62	0.0033

S3d. Descriptive PGLS of log median terminal branch length (caper, λ by ML, chronogram). A = pre-specified version (68-tip gene trees, all representatives); B = A without six representatives whose median terminal branch is < 0.001 because a conspecific strain remained in the trees (post hoc); C = terminal branches measured after pruning each gene tree to the selected OTUs, including all seven E. elegans strains (post hoc); D = C without the two representatives whose median terminal branch remained below 0.001 (post hoc). n = 57 includes the two outgroups.

variant	predictor	β	P	n	λ̂ (ML)
A	log2 cell number	-0.041	0.691	57	0.000
A	differentiation state	-0.320	0.634	57	0.000
B	log2 cell number	+0.044	0.211	51	0.990
B	differentiation state	+0.296	0.090	51	0.987
C	log2 cell number	+0.078	0.316	57	0.859
C	differentiation state	+0.744	0.085	57	0.866
D	log2 cell number	+0.046	0.152	55	0.996
D	differentiation state	+0.293	0.050	55	0.995
A, joint model (pre-specified)	log2 cell number	-0.001	0.994	57	0.000
A, joint model (pre-specified)	differentiation state	-0.312	0.794	57	0.000

S3e. Tips of every design contrast (rule v1). Foreground and background denote the tip sets used in each design contrast: foreground = connected set of branches classed as differentiated, background = nearest sister subtree containing no foreground-classed branch; raw components with overlapping recorded foreground/background branch sets were merged, and a foreground tip set can exclude descendants reconstructed as losses and need not comprise every descendant of its most recent common ancestor; ‘nodes above’ = number of nodes between the foreground component and the node at which the background attaches (1 = direct sister). Strain labels as in the Lindsey et al. (2024) alignments.

scenario	k	nodes above	foreground tips	background tips
S1_ER	1	1	astperNIES564, astgubNIES418 (2)	gonpec_NCBI, gonoctNIES851, gonquaNIES653, gonvirNIES654, gonmulNIES737 (5)
S1_ER	2	3	volbarNIES730, volkirNIES543, volferNIES3986, volgloSAG199_80 (4)	chlschSAG2486, chlgloSAG81_72, chlrei_Phytozome, chlamydebSAG70_81 (4)
S1_ER	3	2	volobvNIES868, volcar_Phytozome, volafrNIES863, voldisNIES4128, volovaNIES2569, volterNIES544, volaurNIES541, pleojapUTEX2523 (8)	eudeleNIES456, eudminNIES856, eudeleNIES719, eudeleNIES720 (4)
S1_ER	4	1	pleostarNIES1362, pleoindNIES736 (2)	eudeleNIES568, eudeleFACHB2321, eudeleNIES458, volpowNIES4127, volgigNIES867 (5)
S2_ARD	1	1	astperNIES564, astgubNIES418 (2)	gonpec_NCBI, gonoctNIES851, gonquaNIES653, gonvirNIES654, gonmulNIES737 (5)
S2_ARD	2	3	volbarNIES730, volkirNIES543, volferNIES3986, volgloSAG199_80 (4)	chlschSAG2486, chlgloSAG81_72, chlrei_Phytozome, chlamydebSAG70_81 (4)
S2_ARD	3	1	volaurNIES541, pleojapUTEX2523, volobvNIES868, volcar_Phytozome, volafrNIES863, voldisNIES4128, volovaNIES2569, volterNIES544, pleothompNIES4126, pleostarNIES1362, pleoindNIES736, volpowNIES4127, volgigNIES867 (13)	eudeleNIES456, eudminNIES856, eudeleNIES719, eudeleNIES720 (4)

scenario	k	nodes above	foreground tips	background tips
S7_ORD	1	1	astperNIES564, astgubNIES418 (2)	gonpec_NCBI, gonoctNIES851, gonquaNIES653, gonvirNIES654, gonmulNIES737 (5)
S7_ORD	2	3	volbarNIES730, volkirNIES543, volferNIES3986, volgloSAG199_80 (4)	chlschSAG2486, chlgloSAG81_72, chlrei_Phytozome, chlamydebSAG70_81 (4)
S7_ORD	3	1	volaurNIES541, pleojapUTEX2523, volobvNIES868, volcar_Phytozome, volafrNIES863, voldisNIES4128, volovaNIES2569, volterNIES544, pleothompNIES4126, pleostarNIES1362, pleoindNIES736, volpowNIES4127, volgigNIES867 (13)	eudeleNIES456, eudminNIES856, eudeleNIES719, eudeleNIES720 (4)
S3_pars1	1	1	astperNIES564, astgubNIES418 (2)	gonpec_NCBI, gonoctNIES851, gonquaNIES653, gonvirNIES654, gonmulNIES737 (5)
S3_pars1	2	3	volbarNIES730, volkirNIES543, volferNIES3986, volgloSAG199_80 (4)	chlschSAG2486, chlgloSAG81_72, chlrei_Phytozome, chlamydebSAG70_81 (4)
S3_pars1	3	1	volaurNIES541, pleojapUTEX2523, volobvNIES868, volcar_Phytozome, volafrNIES863, voldisNIES4128, volovaNIES2569, volterNIES544, pleothompNIES4126, pleostarNIES1362, pleoindNIES736, volpowNIES4127, volgigNIES867 (13)	eudeleNIES456, eudminNIES856, eudeleNIES719, eudeleNIES720 (4)
S4_pars_g1_l2	1	1	astperNIES564, astgubNIES418 (2)	gonpec_NCBI, gonoctNIES851, gonquaNIES653, gonvirNIES654, gonmulNIES737 (5)
S4_pars_g1_l2	2	3	volbarNIES730, volkirNIES543, volferNIES3986, volgloSAG199_80 (4)	chlschSAG2486, chlgloSAG81_72, chlrei_Phytozome, chlamydebSAG70_81 (4)
S4_pars_g1_l2	3	2	volobvNIES868, volcar_Phytozome, volafrNIES863, voldisNIES4128, volovaNIES2569, volterNIES544, volaurNIES541, pleojapUTEX2523 (8)	eudeleNIES456, eudminNIES856, eudeleNIES719, eudeleNIES720 (4)
S4_pars_g1_l2	4	1	volpowNIES4127, volgigNIES867 (2)	eudeleNIES568, eudeleFACHB2321, eudeleNIES458 (3)
S4_pars_g1_l2	5	2	pleostarNIES1362, pleoindNIES736 (2)	eudperNIES725, euduniSAG24_1c, eudillNIES460, eudeleNIES717, eudcylNIES722 (5)
S5_pars_g2_l1	1	1	astperNIES564, astgubNIES418 (2)	gonpec_NCBI, gonoctNIES851, gonquaNIES653, gonvirNIES654, gonmulNIES737 (5)
S5_pars_g2_l1	2	3	volbarNIES730, volkirNIES543, volferNIES3986, volgloSAG199_80 (4)	chlschSAG2486, chlgloSAG81_72, chlrei_Phytozome, chlamydebSAG70_81 (4)
S5_pars_g2_l1	3	1	volaurNIES541, pleojapUTEX2523, volobvNIES868, volcar_Phytozome, volafrNIES863, voldisNIES4128, volovaNIES2569, volterNIES544, pleothompNIES4126, pleostarNIES1362, pleoindNIES736, volpowNIES4127, volgigNIES867 (13)	eudeleNIES456, eudminNIES856, eudeleNIES719, eudeleNIES720 (4)

scenario	k	nodes above	foreground tips	background tips
S6_dollo	1	1	astperNIES564, astgubNIES418, volgloSAG199_80, volbarNIES730, volkirNIES543, volferNIES3986, pleothompNIES4126, volaurNIES541, pleojapUTEX2523, volobvNIES868, volcar_Phytozome, volafrNIES863, voldisNIES4128, pleostarNIES1362, pleoindNIES736, volovaNIES2569, volterNIES544, volpowNIES4127, volgigNIES867 (19)	chlschSAG2486, chlgloSAG81_72, chlrei_Phytozome, chlamydebSAG70_81 (4)

S3f. Historical refits and sensitivity checks of the primary model (rule v1 strict, S2, statistic r). The entries below are retained as an audit trail; incomplete diagnostics or divergent transitions prevent treating them as validated sensitivity results. Blank diagnostic cells mean not recorded. Leave-one-contrast-out refits and prior-sensitivity refits for the between-contrast scale σ_k; the primary fit uses σ_k ~ half-N(0, 0.5).

fit	μ	95 % equal-tailed credible interval	P(μ < 0)	max $\hat{R}$	divergent transitions
recorded primary fit (K = 2; 4 × 4,000 iterations, adapt_delta 0.8)	+0.077	-0.201 to 0.378	0.19	1.020	421
without contrast 1 (*Astrephomene* vs *Gonium*)	-0.026	-1.000 to 0.812	0.44		
without contrast 2 (sect. *Volvox* vs *Chlamydomonas* group)	+0.051	-0.671 to 0.764	0.40		
σ_k ~ half-N(0, 0.25)	+0.067	not recorded	0.18		
σ_k ~ half-N(0, 1)	+0.076	not recorded	0.22		
earlier refit (residual divergences): 4 × 20,000 iterations, adapt_delta 0.99 (same data, priors, weights, seed); n = 206, 153 genes	+0.068	-0.359 to 0.457 (rate ratio 0.70 to 1.58)	0.21	1.0006	75

fit	μ	95 % equal-tailed credible interval	P(μ < 0)	max R̂	divergent transitions
without alignments < 150 columns (pre-specified filter applied post hoc); recorded settings; n = 192, 141 genes	+0.070	-0.225 to 0.372 (rate ratio 0.80 to 1.45)	0.21	1.0038	103
without alignments < 150 columns; 4 × 20,000 iterations, adapt_delta 0.99; n = 192, 141 genes	+0.067	-0.331 to 0.474 (rate ratio 0.72 to 1.61)	0.23	1.0013	31
without observations with a single-tip side (pre-specified filter applied post hoc); recorded settings; n = 171, 136 genes	+0.060	-0.414 to 0.462 (rate ratio 0.66 to 1.59)	0.23	1.1385	945
without single-tip observations; 4 × 20,000 iterations, adapt_delta 0.99; n = 171, 136 genes	+0.074	-0.308 to 0.457 (rate ratio 0.74 to 1.58)	0.20	1.0006	75
both filters; 4 × 20,000 iterations, adapt_delta 0.99; n = 160, 126 genes	+0.069	-0.306 to 0.453 (rate ratio 0.74 to 1.57)	0.21	1.0006	42

Maximum R̂ over all fitted models (all rules, scenarios and statistics): 1.315; values above 1.05 occur only in sensitivity fits with 2,000 iterations (Table S3b, last column).

S3g. Post hoc numerical validation of S2 ARD, rule v1 strict. Two longer brms attempts retained divergences. Exact Gaussian marginalization subsequently preserved the model, priors, normalized weights and seed. Four chains used 8,000 iterations (4,000 warm-up), adapt_delta 0.99 and maximum tree depth 15; a fixed fallback allowed 16,000 iterations and adapt_delta 0.999 if diagnostics failed. Selection depended only on diagnostics. All attempts, diagnostics and chain-labelled draws are in the code and data archive. The length filter (≥ 150 columns) and tip-count filter (≥ 2 tips per side per gene) repair earlier implementation omissions and were applied post hoc. n counts gene–contrast observations.

fit	n	genes	K	μ	95 % credible interval	P(μ < 0)
r: all	206	153	2	+0.070	-0.333 to 0.454	0.210
r: length filter	192	141	2	+0.064	-0.357 to 0.454	0.227
r: tip-count filter	171	136	2	+0.072	-0.335 to 0.463	0.203
r: both filters	160	126	2	+0.068	-0.331 to 0.449	0.209
Δrad: all, unweighted	143	113	2	-0.008	-0.350 to 0.340	0.598

fit	accepted attempt	iterations / warm-up	adapt_delta	max R̂	min bulk / tail ESS	MCSE(μ)	min E-BFMI	divergences / tree-depth hits
r: all	1	8,000 / 4,000	0.99	1.0008	4125 / 2547	0.0014	0.822	0 / 0
r: length filter	1	8,000 / 4,000	0.99	1.0011	3927 / 2255	0.0014	0.753	0 / 0
r: tip-count filter	1	8,000 / 4,000	0.99	1.0013	3761 / 2903	0.0014	0.837	0 / 0
r: both filters	1	8,000 / 4,000	0.99	1.0008	3580 / 2689	0.0014	0.825	0 / 0
Δrad: all, unweighted	1	8,000 / 4,000	0.99	1.0013	3777 / 3472	0.0013	0.854	0 / 0

Acceptance required zero divergent transitions and maximum-tree-depth hits, rank-normalized $\hat{R} < 1.01$, bulk and tail effective sample size (ESS) ≥ 400 for every monitored model parameter, E-BFMI ≥ 0.3 in each chain and Monte Carlo standard error (MCSE) of $\mu \leq 0.01$. These checks assess numerical sampling only. They do not remove the uncertainty arising from two retained contrasts, reconstructed trait states or prior assumptions.

Exact marginalization used in S3g. The original r model has Gaussian likelihood contributions raised to weights w_i normalized so that Σw_i = n. Up to a parameter-independent constant, this equals a Gaussian likelihood with residual variances σ^2/w_i. Let Z_g and Z_k be the gene and contrast incidence matrices. Integrating the normal priors on μ, a_g and b_k gives $y \sim N(0, V)$, where $V = \mathrm{diag}(\sigma^2/w_i) + \sigma_g^2 Z_g Z_g' + \sigma_k^2 Z_k Z_k' + 11'$; the last term uses the prior variance 1 of μ. The original half-normal priors on σ_g and σ_k and half-Student-t prior on σ are retained. The unweighted Δrad model is the case w_i = 1. Low-rank matrix identities evaluate this likelihood without forming a large dense matrix. The implementation checks this expression against direct dense covariance evaluation on a fixed grid for each dataset; original-effect draws are reconstructed from their conditional Gaussian distribution. Scripts: stage0_marginal.stan and stage0_marginal_refits.R. Numerical equivalence is a computational check, not a test of model adequacy.

Table S4. Per-species decomposition (gate 3)

Median over orthogroups of the ratio of the path length from the basal trifurcation to the species' tip to the *C. reinhardtii* branch, for dN and dS separately (run H1p). Within each orthogroup the origin-B paths share the fitted ω2/ω0; its across-orthogroup median is 0.63.

species	origin	differentiation	n	dN ratio	dS ratio	fraction dN higher	fraction dS higher
Tetrabaena socialis	A	none	1,353	1.08	1.05	0.55	0.54
Gonium pectorale	B	none	1,350	0.89	1.39	0.4	0.68
Yamagishiella unicocca	B	none	1,354	0.97	1.53	0.48	0.74
Eudorina sp.	B	none	1,354	1.19	1.86	0.65	0.81
Astrephomene gubernaculif-era	B	partial	1,354	1.07	1.64	0.55	0.75
Pleodorina starrii	B	partial	1,354	1.12	1.76	0.59	0.78
Volvox reticuliferus	B	partial	1,354	1.84	3.12	0.85	0.88
Volvox africanus	B	complete	1,354	1.52	2.44	0.82	0.86
Volvox carteri	B	complete	1,354	1.45	2.34	0.79	0.85
origin B, colonial, without somatic differentiation (median over 3 species)			1,354	1.02	1.6	0.52	0.76
origin B, with somatic differentiation (median over 5 species)			1,354	1.41	2.26	0.78	0.85

Sensitivity of the pooled decomposition: main (median over tips) dN 1.25 / dS 1.96; mean over tips 1.49 / 2.44; without *V. reticuliferus* 1.18 / 1.86; *Chlamydomonas*-branch dS < 10 only 1.25 / 2.07 (n = 1,255); imposed topology (H1) median a +0.158, median b +0.594 (n = 1,355); *Tetrabaena* vs *Chlamydomonas* 1.08 / 1.05.

Note S5. Identities used in the analysis and their machine checks

P1 (path-sum ω identity; Sections 2.5 and 3.3). In a codeml branch class with shared ω, dN_b = ω·dS_b for every branch b, so for any set of branches in the class ΣdN_b/ΣdS_b = ω, and for two classes ln(ω_B/ω_0) = ln(ΣdN_B/ΣdN_0) − ln(ΣdS_B/ΣdS_0) exactly. Checked symbolically (SymPy 1.14.0; residual 0), by counterexample search (z3 5.1.0: negation unsatisfiable) and proved in Lean 4.32.2 with Mathlib (`LeanVerify.omega_class_path_ratio`, `LeanVerify.log_omega_ratio_decomposition`; axioms propext, Classical.choice, Quot.sound only). In the data, |(a − b) − ln(ω2/ω0)| ≤ 0.0161 in every orthogroup (six-decimal printing).

P2 (root placement; Section 3.1). For a reversible substitution process with the same generator on both root-adjacent branches, the likelihood depends on their lengths only through the sum (Felsenstein's pulley principle); with different generators on the two branches, as when they carry different ω classes, it depends on the split. Checked numerically on a four-state reversible model: maximum likelihood difference across splits of a fixed total 1.4e-16 with one generator and 0.009 with two; the two-state closed form was verified in SymPy. This example establishes that different generators can make the likelihood depend on the root-adjacent split (Álvarez-Carretero et al. 2023, supplementary Fig. S1D); it does not by itself establish weak identifiability in the empirical codon fits, whose boundary estimates are reported in Section 3.1.

P3 (terminal branch length; Section 3.5). With BL_i = r·T_i, where T_i is the time since lineage i's last sampled split, the ordinary least-squares slope of BL on a trait X equals r·Cov(T, X)/Var(X): non-zero whenever T co-varies with X even when all lineages share the rate r (SymPy identity).

P4 (sister-clade contrast; Section 2.6). Two sister clades share the time T since their common ancestor, so the ratio of mean path lengths (r_F·T)/(r_B·T) = r_F/r_B is free of T (SymPy identity).

Scripts: `scripts/11_gate3/g3_math_checks.py` (results in `results/gate3/math_checks.json`) and `lean_verify/LeanVerify/RoukaOmegaClass.lean`.

Table S6. Methods ↔ code reconciliation (two-way)

Methods statements and their implementations are reconciled below; departures and their resolution are recorded at the end. Analysis scripts are under `scripts/08_gate/`, `scripts/09_stage0/`, `scripts/11_gate3/`, `scripts/12_preprint_v3/`, `scripts/13_completion/` and `lean_verify/LeanVerify/` in the independent code and data archive (Data and code availability). Manuscript-production and release-check scripts cited below use local publication sources and are excluded from that archive.

Methods statement	Implementation (script: element)	Status
2.1 Genomes, OrthoFinder v2.5.5 orthogroups, MAFFT codon alignments unchanged	inputs `results/paml_dnds/volvocine_10species/<OG>/cds_aligned.phy` and `codeml_m2.ctl` reused by `gate_g2_paml_runs.py`; original scripts `02_ortholog/run_orthofinder_volvocine_10species.sh`, `04_paml/run_paml_volvocine.py`: `run_mafft_codon_align`	stated = implemented
2.2 Three locally hashed specifications; primary decision criteria recorded before their gate evaluations, with timing qualifications in Methods 2.2	`gate_spec_v1.md`, `stage0_spec_v1.md`, `gate3_spec_v1.md`; hashes in `results/*/…sha256`; decision tables `gate_evaluate.py`, `s0_06_evaluate.py`, `gate3_decision.md`	stated = implemented

Methods statement	Implementation (script: element)	Status
2.2 Gate-2 amendments A1–A4 (11 September 2026, after the first ancestral-state run and a 98-gene trial; the record does not establish that all amendments preceded all trial statistics)	`stage0_spec_v1_addendum_20260911.md` (hash-recorded); `s0_03_scenarios.R: mpr_prob` (A1, A3), `contrasts_from_p(rule="v11")` (A2), `s0_04_contrasts.py: relaxed` (A4)	**amendments previously described as pre-specified → dated and separated from the original rules**
2.2 Initially omitted rate-MCMC sensitivity, ASTRAL PGLS and 200-replicate orthogroup-cluster bootstrap	absent from the historical `s0_03_scenarios.R` / `s0_05_model.R`; completed by `scripts/13_completion/` under the dated completion protocol	**initial omission retained in the history; completed post hoc in Supplement S7**
2.3 Gate-1 diagnostic C0-b (near-zero-stem subset comparison)	`gate_evaluate.py` C0-b; not in `gate_spec_v1.md`	**present-but-unstated as a post-specification addition → stated as a diagnostic added after the specification**
2.6 Contrast construction merges raw components with overlapping recorded foreground/background branch sets; foreground tip sets need not be complete clades	`s0_03_scenarios.R: contrasts_from_p` (merge loop, lines 113–118)	**present-but-unstated → added (Methods 2.6, Table S3e)**
2.6 Pre-specified joint PGLS (cell number + differentiation state)	`s0_05_model.R` (formula `log_bl ~ log2_cells + diff`, `model/pgls.csv`); manifest `stage0.pgls.joint.*`	**present-but-unstated → added (Table S3d); single-predictor fits stated as additional descriptive fits**
2.3 codeml model = 2, NSsites = 0, CodonFreq = 2, original control file, only tree changed	`gate_g2_paml_runs.py`: ctl reused, `seqfile/treefile/outfile` lines substituted	stated = implemented
2.3 Unrooted trees, basal trifurcation	`scripts/08_gate/trees/*.nwk`	stated = implemented
2.3 Per-orthogroup branch counts and warnings recorded in every run; H0u modal count tested against 17; all-run modes (ten species 17, nine species 15) and warning totals reported	`gate_g2_paml_runs.py: parse_out` extracts `ntime` and the warning; `gate_evaluate.py:metrics` summarizes every run; `main` applies the branch-count criterion to H0u	**run-stopping assertion and per-orthogroup validation were overstated; scope of recording, summaries and H0u criterion now explicit**
2.3 Runs H0u, H3, H1, H1p, H2, H2p on T_imp / T_dat	`gate_g2_paml_runs.py: RUNS`	**present-but-unstated → runs listed in text**
2.3 Tetrabaena removal	`gate_g2_paml_runs.py: write_phylip_dropped` deletes the sequence block; columns kept	**claimed-but-absent ("re-aligned") → text corrected**
2.3 QC 0.0001 < ω < 10 (and < 100), all classes present	`gate_evaluate.py: LO, HI, HUGE; load()`	stated = implemented
2.3 Common sets (1,336 / 1,436)	`gate_evaluate.py: main` (intersections); `build_numbers_manifest.py`	stated = implemented
2.3 dS ≥ 10 sensitivity	`gate_diag_chlamy_branch_dS.py` (Chlamydomonas branch 11..2 of each run); `gate_evaluate.py` diag lines; manifest `gate1.*.dS10.*`	stated = implemented

Methods statement	Implementation (script: element)	Status
2.3 Scan of original rooted model-2 and model-1 outputs	`build_numbers_manifest.py` (regenerates `diag_m2_root_branch_split.csv`: 1,431 outputs rescanned, 1,430 in the stored table, maximum difference 5×10^{-5}; free-ratio scan 11..2 vs 11..12)	**provenance added → text states the regeneration**
2.3 Pipeline check on zero-stem subset	`gate_evaluate.py` C0-b; manifest `artefact.h0u.clean.*`	stated = implemented
2.4 Gene trees LG+G4, 1,000 UFBoot, seed 20260911, no constraint	`gate_g1_iqtree_unconstrained_per_og.py: cmd`	stated = implemented (seed added to text)
2.4 Species tree unpartitioned (UFBoot 1,000 + SH-aLRT 1,000) and partitioned (UFBoot 1,000)	`results/gate/species_tree_unconstrained/*.log` command lines (`-B 1000 -alrt 1000`; `-p partitions.nex -B 1000`)	**present-but-unstated (SH-aLRT) → added**
2.4 ASTRAL-III 5.7.8, local posterior support	`gate_g1_concordance.sh` (`astral.5.7.8.jar -t 2`)	**present-but-unstated (`-t 2`) → added**
2.4 gCF from gene trees, sCF from 100 quartets	`gate_g1_concordance.sh` (`--gcf … --scf 100`)	**present-but-unstated (100 quartets) → added**
2.4 Bipartition frequencies (ape::prop.part)	`manifest_r_parts.R` (recomputed; equals the 2026-09-11 table)	stated = implemented
2.4 Support check for the basal clade	`gate_g4_topology_check.R` → `c4_topology_support.csv`	stated = implemented
2.5 Decomposition: trifurcation node, median over 8 tips, identity tolerance 0.02, sensitivities	`g3_part0_dn_ds_decomposition.py` (`root_to_tip`, `metrics`, TOL)	stated = implemented
2.5 Per-species decomposition	`g3_part0b_per_species.py`	stated = implemented
2.5 Lean identities	`RoukaOmegaClass.lean` (two theorems; `lake build`; standard axioms)	stated = implemented
2.6 Pruning to Reinhardtinia + *C. moewusii* + *Chlorella*, 68 tips; CR line endings normalised	`s0_00_prepare_lindsey_data.py` (`keep_labels`; byte replacement)	**present-but-unstated (outgroups) → text corrected**
2.6 OTU representatives: one strain per species with the most genes, except all seven *E. elegans* strains retained separately	`s0_01_build_trait_table.py` (`otu_representative`; every *E. elegans* row receives 1); manifest `stage0.n_species_reps`, `n_eudorina_elegans_otus`, `n_outgroup_otus` (48 + 7 + 2 = 57)	**earlier three-clade wording corrected to the seven retained strains; historical `otu_note` is not the selection rule**
2.6 Differentiation coding 0/1/2 from Lindsey 2021 Table S2 and footnote	`s0_01_build_trait_table.py: VOLVOX_GERM_SOMA`, genus table	stated = implemented
2.6 Branch lengths on the Lindsey 2024 ML topology, LG+G4 partitioned	`s0_02_branch_lengths.sh` (`-te … -p charpartition.nex -m LG+G4`)	**present-but-unstated (partitioned) → added**
2.6 Scenarios S1/S2/S7 (fitMk; 1,000 stochastic maps on the fitted Q)	`s0_03_scenarios.R` (`fitMk`, `simmap(Q, nsim=1000)`)	stated = implemented
2.6 Scenarios S3–S6 Sankoff costs, root fixed at 0, oracle test	`s0_03_scenarios.R: mpr_prob(root0=TRUE)`, oracle `stopifnot`	stated = implemented

Methods statement	Implementation (script: element)	Status
2.6 Thresholds 0.9 / 0.1; contrast rules v1 and v1.1; disjoint contrasts	`s0_03_scenarios.R` (lines 88–130)	stated = implemented
2.6 r_gk from fixed-topology per-gene trees; monophyly in unconstrained gene tree (strict / relaxed)	`s0_02b_per_gene.sh` (`-te` + `--ancestral`; `-B 1000`); `s0_04_contrasts.py: is_monophyletic, relaxed`	stated = implemented
2.6 ≥ 2 tips per side; alignment ≥ 150 aa (pre-specified filters)	not in `s0_04_contrasts.py` / `s0_03_scenarios.R`; every design contrast has ≥ 2 tips per side (`scenarios/contrasts.csv`), but per gene the sides are formed from the sequences present, so 35 of the 206 primary observations (32 genes) have a single-tip side (`contrasts_r.csv`, `nF`/`nB`; manifest `stage0.refit.n_obs_singleton_primary`); 45 of 263 alignments < 150 columns (`stage0_alignment_lengths.csv`), 12 genes of the primary set	**spec-claimed-but-absent → disclosed with the counts; both filters applied post hoc in numerically validated refits (Table S3g)**
2.6 Δrad: Grantham > 100; ≥ 10 substitutions per side	`s0_04_contrasts.py` (`GR[(x, y)] > 100`; `tF >= 10 and tB >= 10`)	**present-but-unstated (≥ 10) → added**
2.6 Hierarchical model: priors μ ~ N(0, 1), σ_k ~ half-N(0, 0.5), σ_g ~ half-N(0, 1), brms default on the residual scale; chains and iterations; seed	`s0_05_model.R: fit_one` (`prior(normal(0,1), class="Intercept")`, `sd` priors; `chains=4, iter=ITER, warmup=ITER/2, seed=20260911`; ITER = 4000 for the primary run, 2000 for the sensitivity runs in `s0_run_models.sh`)	**present-but-unstated → added**
2.6 Likelihood weights for r only (w = min(ΣPL_F, ΣPL_B) × alignment length, scaled to mean 1); Δrad unweighted	`s0_04_contrasts.py` (`w = min(sum(plF), sum(plB)) * nsites`); `s0_05_model.R` (`r | weights(wn)` with `wn = w/mean(w)`; `fit_one(dd, "d_rad", FALSE, s)`)	**text previously described both statistics as weighted with an inverse-variance model → corrected to the brms observation-weight specification, r only**
2.6 $\hat{R}$, leave-one-contrast-out and prior sensitivity (σ_k ~ half-N(0, 0.25) and half-N(0, 1)) for the primary scenario	`s0_05_model.R` (`rhat(m)`; LOO loop over contrasts of `PRIMARY`; `sdk in c(0.25, 1.0)`); results `model/loo_contrast_S2.csv`, `s0_05_v1.log`; Table S3f	**present-but-unstated → added (Table S3f)**
2.6 Criteria P(μ < 0) ≥ 0.9 and ≥ K − 1 negative	`s0_06_evaluate.py`	stated = implemented
2.6 Post hoc numerical validation of S2 ARD (all observations, each omitted filter, both filters, and unweighted Δrad); unchanged model and priors, acceptance based only on sampling diagnostics	`stage0_release_refits.R` (failed original parameterization); `stage0_marginal_refits.R` and `stage0_marginal.stan` (same Gaussian model integrated analytically) → `stage0_release_refits.csv`; full diagnostics and all attempts in `results/preprint_v3/release_refits/`; manifest `stage0.release.*`; the recorded fits and decision tables remain unchanged	**historical fits retain unresolved diagnostics; Section 3.4 and Figure 4 use the accepted refits in Table S3g**

Methods statement	Implementation (script: element)	Status
2.6 Descriptive PGLS (caper, λ ML, chronos λ = 1 correlated): variant A pre-specified, B–D post hoc	`s0_05_model.R` (chronos in `s0_03`); `stage0_pgls_check.R` (variants B–D added after the six near-zero terminal branches were seen when the figure was drawn)	stated = implemented (post hoc status of B–D stated)
2.6 Contrast tip lists (Table S3e)	`results/stage0/scenarios/contrasts.csv` (`F_tips`, `B_tips`, `steps_to_background`) rendered by `build_supplementary.py`	stated = implemented
2.7 Ten-species gradient from unconstrained gene trees; PGLS on T_dat; 10,000 permutations, seed 42	`gate_g3_chlamy_free_gradient.R` (`N_PERM`, `set.seed(42)`, `pgls(lambda="ML")` with fallback λ = 0 on error, not triggered)	**present-but-unstated (permutation count, seed) → added**
2.7 Free-ratio ω regression	`gate_g3_chlamy_free_gradient.R` reads `model1_per_species_omega.csv` from the original rooted model-1 run (T_imp)	**claimed-inaccurately ("same regression") → source and caveat stated**
2.7 Identity BL = r·T and slope r·Cov/Var	`g3_math_checks.py` (P3)	stated = implemented
2.9 Software versions from the executing binaries	`versions_s0_02*.txt`, log headers (IQ-TREE 3.0.1, codeml 4.10.10), `tool_check.sh` output, R `packageVersion` lines in `s0_05` logs	stated = implemented (interpreters and matplotlib added)
Functional-category tally: assignment of versions 1 and 2 retained, every annotation letter votes, ties by first occurrence, historical table reproduced cell by cell before reuse, categories with < 5 orthogroups dropped, no significance test	`scripts/13_completion/cog_reassessment.py` (imports `06_cog_breakdown/cog_category_breakdown_volvocine_10species.py`; `pd.testing.assert_frame_equal` against the archived table; `if len(g)<5: continue`) → `completion/cog/{og_to_cog.csv, category_reassessment.csv, category_summary.csv, methods.txt}`	**present-but-unstated → Methods 2.8 added (2026-09-15 verification pass)**
2.6 Cell number on the 57-OTU dataset: geometric midpoint sqrt(low x high) of the Lindsey 2021 genus range; 2.7 retains separate ten-species cell values	`s0_01_build_trait_table.py` (`modal = math.sqrt(lo * hi)`); ten-species `per_species_iqtree_summary.csv`; `make_figures.py` uses the separate inputs for Figure S1A and S1B	**modal-value wording corrected in Methods and Figure S1B; scope restricted to the 57-OTU regressions**
Permutation test of the per-orthogroup correlation, ten species and nine species reported separately	`gate_g3_chlamy_free_gradient.R:perm_test` → `permutation_results.csv` rows "ref BL, 10 spp" and "G-c BL, 9 multicellular"	**stated inaccurately (nine-species value quoted in a ten-species sentence) → both now reported (2026-09-15)**
Machine checks of the identities: SymPy symbolic residual, z3 counterexample search, Lean 4 + Mathlib proofs	`g3_math_checks.py` (imports sympy, numpy, z3, scipy; `math_checks.json` records those three versions); `RoukaOmegaClass.lean`	**claimed-but-absent (Note S5 also named Maxima and Wolfram Engine, for which the archive holds no input or output) → those two engines removed from Note S5 (2026-09-15)**

Methods statement	Implementation (script: element)	Status
Text rendering from the manifest; typed literals restricted to a documented allowlist; build fails otherwise; unknown format specs fail; `int` rounds	`render_manuscript.py` (exit 1 on a missing key, exit 2 on a literal absent from `allowed_literals.tsv`, exit on an unrecognised format spec; `int` uses `round` so that a value printed here and by the supplement generator's `:.0f` cannot differ by one); `numbers.json` (1,479 entries); submission texts by `build_submission_texts.py`	stated = implemented
Data availability: independent code and data archive for Zenodo, separate from the earlier data version	`make_deposit_bundle.py` (curated layout, reproduction instructions, SHA256SUMS); `test_reproduction.py` (reproduce numbers and plots from the archive alone; check manuscript and PDF separately with local publication inputs); raw codeml scans and alignment-length summaries are supplied when raw data are not redistributed	code/data archive prepared for deposition; full manuscript, supplement and finished figures excluded; earlier Zenodo DOI kept distinct
Completed audit of every saved scenario fit; exact-marginal refits of the scenario grid, priors and leave-one-contrast-out	`scripts/13_completion/audit_saved_models.R`, `refit_scenarios.R`, `gaussian_marginal.stan`, `marginal_common.R`; accepted fit registry and full parameter diagnostics	post hoc completion; Table S7a and Figure S2; unchanged likelihood and priors
S2 MCMC rate uncertainty and propagation to gene statistics	`stochastic_mapping.R`, `mapping_contrasts.py`, `refit_mapping.R`; chain-aware rate/branch diagnostics and F/B reuse provenance	initially omitted, now completed post hoc; Gamma(1,1) rate priors and equal root prior explicit; averaged thresholded design, not full averaging of effect posteriors over designs
ASTRAL branch lengths and descriptive PGLS	`run_astral_lengths.sh`, `chronos_diagnostics.R`, `astral_pgls.R`; both topologies use the same midpoint root convention; installed and locally consistent penalties retained	initially omitted, now completed post hoc; local algorithm sensitivity and objective–gradient mismatch explicit (Table S7c)
200 gene-cluster bootstrap replicates	`gene_cluster_bootstrap.R`; fixed gene selections, distinct copy identifiers, renormalized weights, posterior-mean estimator and diagnostics	initially omitted, now completed post hoc; conditional on the two contrasts (Table S7d)
Original and corrected COG category tallies	`cog_reassessment.py`; archived mapping, annotation input hashes and exact check against historical table	recomputed post hoc, distinguishing both H1p foreground classes; descriptive interpretation (Table S7e)
Reconstruction on a chronogram in addition to the substitution-length tree	`chronogram_mapping.R`, followed by `mapping_contrasts.py` and `refit_mapping.R` with the chronogram directory	additional original-specification omission, completed post hoc under the locally consistent relative-time proxy; fixed Q, ER/ARD/ordered models (Table S7f)

Departures from the pre-specified analyses. (a) Gate 2 was amended on 11 September 2026 (A1–A4). A1–A3 followed the first ancestral-state reconstruction (A1 parsimony implementation with an explicit cost[from, to] convention and an oracle check; A2 background rule v1.1 as a sensitivity analysis; A3 parsimony root fixed at state 0); A4 added a relaxed gene filter as a sensitivity analysis after a 98-gene trial run had revealed an empty strict-filter contrast, and the addendum describes this as preceding the aggregation of r; the available record (`s0_05_dryrun.log` contains a trial fit) does not establish that every amendment preceded every calculation of a contrast statistic. The primary criteria were evaluated under the original rules. (b) Three items of the gate-2 specification were initially omitted and completed post hoc on 15 September 2026: the stochastic-mapping sensitivity with rate matrices sampled by MCMC, the descriptive PGLS on the ASTRAL topology, and the 200-replicate orthogroup-cluster bootstrap (Tables S7b–d); gate 1's near-zero-stem comparison (C0-b) was added after its specification. (c) Two pre-specified filters (≥ 2 tips per side; alignment ≥ 150 amino acids) were not implemented in the recorded analysis; per gene, 35 of the 206 primary observations have a single-tip side, and 45 short alignments (12 genes of the primary set) were included; both filters were applied post hoc in the validated refits reported in Table S3g. (d) PGLS variants B–D were added post hoc. (e) The recorded primary fit did not meet the pre-specified convergence diagnostic ($\hat{R}$ 1.02 > 1.01); its early longer-chain refits still had divergent transitions. The accepted numerical-validation fits are reported separately in Table S3g; the recorded fit remains the analysis evaluated against the historical criteria. All 54 archived scenario fits were subsequently audited and found to contain divergences. Their replacements and four primary prior/leave-one-contrast-out checks now pass the numerical criteria (Table S7a); the old chains remain diagnostic records. The locally recorded gate criteria were not changed for this revision; their timing and the departures from their specifications are described in Methods 2.2. They were not a prospective public preregistration.

Discrepancies found by this check and their resolution. (1) "Re-aligned" nine-species runs: the sequence was removed from the existing alignment; corrected. (2) The free-ratio ω gradient used the original rooted model-1 estimates for the nine colonial terminal branches; now stated with its caveat. (3) The pre-specified filters of at least two tips per side and at least 150 amino acids were not implemented; an earlier draft of this table stated that the first had no effect because every design contrast has two or more tips per side, but per gene the sides are formed from the sequences present and 35 of 206 primary observations have a single-tip side; both filters are now disclosed with their counts and applied post hoc as refits. (3b) The text had described the branch-parameter and warning checks as run-stopping assertions; the code records them and the decision script evaluates them; corrected. (3c) The text had described both contrast statistics as fitted with inverse-variance weights; only r carries brms observation weights and Δrad is unweighted; corrected. (4) Run H3 (T_dat, one class), SH-aLRT support, the sCF quartet count, the ASTRAL support option, the partitioned branch-length model, the *E. elegans* OTU split, the outgroup composition, the Δrad substitution minimum, the model weights, MCMC settings, the LOO and prior-sensitivity checks and the permutation settings were implemented but not stated; added. (5) The two scans of the original outputs had no script of record; they are now regenerated by the manifest script.

Verification pass of 15 September 2026. A further two-way reconciliation, an independent audit of every numeric literal in the manuscript and supplement, and a rebuild of the release on a second machine reported corrections to text, figures and the renderer: two statements in Section 5.1 that contradicted Section 3.1 and Section 3.2; a claim in Note S5 of Maxima and Wolfram Engine checks for which the archive holds no input or output; "genus-level modal values" for a geometric midpoint; "the three clades" of *E. elegans* for seven retained strains; a nine-species permutation quoted in a ten-species sentence; a functional-category analysis with no Methods section; a branch-parameter verification described more strongly than the code performs; a figure count with no saved derivation; and a figure group label that disagreed with the text. The renderer silently ignored unrecognised format specifications (printing two values at full float precision) and truncated floats under its `int` specification (disagreeing by one with the supplement's rounding); it now fails on an unknown specification and rounds. No analysis was re-run: every change is to the text, the rendering code or the figure code.

Follow-up of 16 September 2026. Figure S1B and the cell-number, OTU and branch-check descriptions were reconciled with the implementation. The text clarifies prior sensitivity, design versus retained contrasts, the published basis for the multicellularity origins, and unassessed input errors. Release checks now cover numeric formats and the Figure S1B axis. Trait coding, observations, model fits and accepted numerical results were retained.

Supplement S7. Analyses completed after the initial correction

All analyses here were completed post hoc on 15 September 2026. The execution protocol was hash-recorded before these runs; the implementation notes record subsequent numerical repairs. The original specifications and gate decisions are unchanged. Source tables, every retained attempt, chain diagnostics and scripts are in the independent code and data archive under `results/preprint_v3/completion/` and `scripts/13_completion/`. The archive is identified in Data and code availability. Supplement S7 contains the following tables and figures; scenario labels refer to reconstruction models.

Table S7a. Complete audit and validated scenario refits

All 54 historical scenario/rule/statistic fits had divergent transitions (7,761 total). These archived chains are retained for diagnosis; their $\hat{R}$ values alone do not validate their estimates. Exact marginalization of the same Gaussian model produced 58 accepted scenario/prior/leave-one-contrast-out entries from 34 distinct inputs. Identical inputs explicitly share a fit. All accepted fits have zero divergences and tree-depth hits. Maximum $\hat{R}$ = 1.0023; minimum bulk/tail ESS = 2865/1567. All equal-tailed 95% intervals include zero. Acceptance criteria are those of Table S3g, including the direct dense-covariance implementation check. Parameter-level diagnostics, MCSE, E-BFMI and attempted settings are in the CSV files. This resolves computational warnings, not model adequacy or the biological limitations of the contrasts.

Rule labels: S = strict gene monophyly filter; R = relaxed gene filter. v1 and v1.1 identify the background rule. Scenario codes follow Table S3a: S1 = equal rates; S2 = all rates different; S3 = parsimony 1:1; S4 = gain 1, loss 2; S5 = gain 2, loss 1; S6 = Dollo-like; S7 = ordered.

Rule	Scenario	Statistic	n / genes / K	μ	95% interval	P(μ < 0)	Old diver-gences	New $\hat{R}$	Min bulk / tail ESS
v1/S	S1	r	318/181/4	0.067	-0.283 to 0.401	0.299	22	1.0005	7133 / 4981
v1/S	S1	rad	204/125/4	-0.007	-0.033 to 0.020	0.781	14	1.0012	4306 / 3334
v1/S	S2	r	206/153/2	0.070	-0.333 to 0.454	0.210	421	1.0008	4125 / 2547
v1/S	S2	rad	143/113/2	-0.008	-0.350 to 0.340	0.598	389	1.0013	3777 / 3472
v1/S	S3	r	206/153/2	0.070	-0.333 to 0.454	0.210	421	1.0008	4125 / 2547
v1/S	S3	rad	143/113/2	-0.008	-0.350 to 0.340	0.598	389	1.0013	3777 / 3472
v1/S	S4	r	405/204/5	0.089	-0.135 to 0.305	0.160	6	1.0006	8737 / 8216
v1/S	S4	rad	204/125/5	-0.007	-0.030 to 0.016	0.806	12	1.0010	3196 / 2735
v1/S	S5	r	206/153/2	0.070	-0.333 to 0.454	0.210	421	1.0008	4125 / 2547
v1/S	S5	rad	143/113/2	-0.008	-0.350 to 0.340	0.598	389	1.0013	3777 / 3472
v1/S	S6	r	1/1/1	0.382	-1.438 to 2.004	0.320	452	1.0005	4967 / 3839

Rule	Scenario	Statistic	n / genes / K	μ	95% interval	P(μ < 0)	Old diver-gences	New R̂	Min bulk / tail ESS
v1/S	S7	r	206/153/2	0.070	-0.333 to 0.454	0.210	421	1.0008	4125 / 2547
v1/S	S7	rad	143/113/2	-0.008	-0.350 to 0.340	0.598	389	1.0013	3777 / 3472
v1.1/S	S1	r	236/151/4	0.041	-0.231 to 0.311	0.338	13	1.0007	7526 / 6062
v1.1/S	S1	rad	115/88/3	0.002	-0.065 to 0.066	0.452	245	1.0009	4250 / 3878
v1.1/S	S2	r	133/109/4	0.119	-0.056 to 0.334	0.060	15	1.0011	3553 / 2382
v1.1/S	S2	rad	93/77/4	0.002	-0.044 to 0.045	0.414	44	1.0013	3902 / 3487
v1.1/S	S3	r	133/109/4	0.119	-0.056 to 0.334	0.060	15	1.0011	3553 / 2382
v1.1/S	S3	rad	93/77/4	0.002	-0.044 to 0.045	0.414	44	1.0013	3902 / 3487
v1.1/S	S4	r	177/127/4	0.081	-0.098 to 0.252	0.116	73	1.0011	4660 / 2582
v1.1/S	S4	rad	92/78/3	0.017	-0.107 to 0.151	0.251	276	1.0013	3952 / 3460
v1.1/S	S5	r	133/109/4	0.119	-0.056 to 0.334	0.060	15	1.0011	3553 / 2382
v1.1/S	S5	rad	93/77/4	0.002	-0.044 to 0.045	0.414	44	1.0013	3902 / 3487
v1.1/S	S6	r	21/21/2	0.043	-0.668 to 0.854	0.475	899	1.0018	2865 / 1855
v1.1/S	S7	r	133/109/4	0.119	-0.056 to 0.334	0.060	15	1.0011	3553 / 2382
v1.1/S	S7	rad	93/77/4	0.002	-0.044 to 0.045	0.414	44	1.0013	3902 / 3487
v1/R	S1	r	482/228/4	0.061	-0.236 to 0.342	0.275	15	1.0009	8949 / 7660
v1/R	S1	rad	296/164/4	-0.003	-0.026 to 0.023	0.649	11	1.0013	3448 / 2708
v1/R	S2	r	320/210/3	0.081	-0.129 to 0.272	0.112	91	1.0012	5852 / 4899
v1/R	S2	rad	223/157/3	-0.001	-0.089 to 0.094	0.559	41	1.0012	3392 / 2807
v1/R	S3	r	320/210/3	0.081	-0.129 to 0.272	0.112	91	1.0012	5852 / 4899

Rule	Scenario	Statistic	n / genes / K	μ	95% interval	P(μ < 0)	Old diver-gences	New R̂	Min bulk / tail ESS
v1/R	S3	rad	223/157/3	-0.001	-0.089 to 0.094	0.559	41	1.0012	3392 / 2807
v1/R	S4	r	586/241/5	0.086	-0.073 to 0.239	0.102	12	1.0005	9745 / 8346
v1/R	S4	rad	291/166/5	-0.007	-0.031 to 0.014	0.805	10	1.0011	3319 / 2520
v1/R	S5	r	320/210/3	0.081	-0.129 to 0.272	0.112	91	1.0012	5852 / 4899
v1/R	S5	rad	223/157/3	-0.001	-0.089 to 0.094	0.559	41	1.0012	3392 / 2807
v1/R	S6	r	45/45/1	0.044	-0.817 to 0.894	0.412	117	1.0013	2882 / 2511
v1/R	S6	rad	40/40/1	0.001	-0.836 to 0.823	0.497	479	1.0014	3070 / 2445
v1/R	S7	r	320/210/3	0.081	-0.129 to 0.272	0.112	91	1.0012	5852 / 4899
v1/R	S7	rad	223/157/3	-0.001	-0.089 to 0.094	0.559	41	1.0012	3392 / 2807
v1.1/R	S1	r	418/215/4	0.048	-0.221 to 0.301	0.300	126	1.0005	7096 / 4442
v1.1/R	S1	rad	194/133/3	-0.002	-0.051 to 0.051	0.584	25	1.0010	4831 / 3583
v1.1/R	S2	r	338/190/4	0.119	-0.082 to 0.309	0.074	19	1.0010	4135 / 3725
v1.1/R	S2	rad	264/159/4	-0.003	-0.045 to 0.037	0.595	13	1.0017	4232 / 3756
v1.1/R	S3	r	338/190/4	0.119	-0.082 to 0.309	0.074	19	1.0010	4135 / 3725
v1.1/R	S3	rad	264/159/4	-0.003	-0.045 to 0.037	0.595	13	1.0017	4232 / 3756
v1.1/R	S4	r	337/202/4	0.049	-0.228 to 0.302	0.289	12	1.0023	3535 / 1567
v1.1/R	S4	rad	166/129/3	-0.001	-0.063 to 0.062	0.551	62	1.0012	3696 / 3518
v1.1/R	S5	r	338/190/4	0.119	-0.082 to 0.309	0.074	19	1.0010	4135 / 3725
v1.1/R	S5	rad	264/159/4	-0.003	-0.045 to 0.037	0.595	13	1.0017	4232 / 3756
v1.1/R	S6	r	146/119/2	0.130	-0.295 to 0.527	0.134	41	1.0014	3567 / 2953

Rule	Scenario	Statistic	n / genes / K	μ	95% interval	P(μ < 0)	Old diver-gences	New R̂	Min bulk / tail ESS
v1.1/R	S6	rad	120/102/2	0.005	-0.406 to 0.385	0.449	707	1.0009	3974 / 4511
v1.1/R	S7	r	338/190/4	0.119	-0.082 to 0.309	0.074	19	1.0010	4135 / 3725
v1.1/R	S7	rad	264/159/4	-0.003	-0.045 to 0.037	0.595	13	1.0017	4232 / 3756

Primary prior and leave-one-contrast-out checks. σ_k prior scale is 0.5 unless stated otherwise. The two-contrast interval for μ is sensitive to this scale because information about between-contrast variation is limited. Interpret it conditional on the prior when generalizing across origins. Removing one contrast leaves K = 1 and still less information about variation among origins.

Check	n / K	μ	95% interval	P(μ < 0)	R̂	Min bulk / tail ESS
σ_k prior scale 0.25	206 / 2	0.073	-0.174 to 0.318	0.158	1.0008	4267 / 2733
σ_k prior scale 1	206 / 2	0.069	-0.541 to 0.651	0.240	1.0007	3784 / 2284
Remove contrast 1	104 / 1	0.046	-0.779 to 0.839	0.391	1.0008	3727 / 2636
Remove contrast 2	102 / 1	0.047	-0.810 to 0.878	0.397	1.0009	3811 / 2585

Table S7b. MCMC transition-rate uncertainty

Four independent ARD rate chains use Gamma(shape 1, rate 1) priors on the six off-diagonal rates and an equal root-state prior on the original substitution-length OTU tree. The accepted attempt used 16,000 conditional maps in total, 10,000 burn-in proposals per chain and one map every 50 proposals (proposal variance 0.1). Maximum rate $\hat{R}$ = 1.0007; minimum bulk/tail ESS = 5327/3656; largest branch-occupancy MCSE = 0.0051. Branch occupancy means the fraction of a branch spent in either differentiated state. 0 branches have a mean ± 1.96 MCSE interval crossing a foreground/background threshold (0.9/0.1); this quantifies Monte Carlo uncertainty, not a posterior interval. Thresholds were not tuned.

Rule	Raw foreground components	Design contrasts K	Foreground / background / intermediate branches
v1	10	5	26 / 71 / 15
v11	10	4	26 / 71 / 15

The 4,000-map first attempt was retained after its maximum branch-occupancy MCSE exceeded 0.01. The 16,000-map fallback used a compiled proposal-likelihood kernel to reduce execution time. Priors, proposals, seeds, burn-in, thinning and phytools conditional-history generation were unchanged. On 200 test matrices, the largest log-likelihood discrepancy from the original implementation was 3.5e-11; a same-seed short chain produced exactly identical rate matrices and maps. Ill-conditioned numerical cases use the original likelihood. Tests and source are deposited.

The accepted branch means define a thresholded contrast design; conditional histories are averaged for that design. This sensitivity integrates transition-rate uncertainty for reconstruction, but does not average the downstream trait-effect posterior over all possible contrast designs. Identical F/B tip sets reuse the original per-gene statistics; newly encountered sets are calculated with the unchanged historical algorithm. The provenance table records which route was used.

Table S7b, continued. Downstream model fits.

Rule / gene filter / statistic / observation filter	n / genes / K	μ	95% interval	P(μ < 0)	$\hat{R}$	Min bulk / tail ESS
v1 / strict / r / all	405 / 204 / 5	0.089	-0.135 to 0.305	0.160	1.0006	8737 / 8216
v1 / strict / r / both_filters	286 / 159 / 5	0.092	-0.121 to 0.293	0.141	1.0013	6367 / 3299
v1 / strict / rad / all	204 / 125 / 5	-0.008	-0.030 to 0.016	0.810	1.0006	3086 / 2669
v1 / relaxed / r / all	586 / 241 / 5	0.086	-0.073 to 0.239	0.102	1.0005	9745 / 8346
v1 / relaxed / rad / all	291 / 166 / 5	-0.007	-0.031 to 0.014	0.805	1.0011	3319 / 2520
v11 / strict / r / all	177 / 127 / 4	0.081	-0.098 to 0.252	0.116	1.0011	4660 / 2582
v11 / strict / rad / all	92 / 78 / 3	0.017	-0.103 to 0.151	0.254	1.0014	4263 / 3848
v11 / relaxed / r / all	337 / 202 / 4	0.049	-0.228 to 0.302	0.289	1.0023	3535 / 1567
v11 / relaxed / rad / all	166 / 129 / 3	-0.001	-0.065 to 0.060	0.553	1.0009	3647 / 3356

Table S7c. ASTRAL and chronogram sensitivity

The supplied ASTRAL tree had no branch lengths. IQ-TREE fitted lengths on that fixed topology from the same partitioned 68-taxon concatenation (LG+G4; topology RF distance 0). Both ML and ASTRAL trees were pruned to the same OTUs and rooted at the midpoint of the outgroup edge, preserving unrooted substitution lengths. This avoids the zero root-to-ingroup edge generated by the initial rooting convention. The original ML chronogram is retained as a separate reference. Per-gene responses are held fixed; this tests the PGLS covariance topology, not re-estimated ASTRAL gene responses.

The installed ape 5.8.1 correlated chronos code has a demonstrated objective–gradient mismatch: basal-rate variance is added to the objective but subtracted in its gradient, and the gradient degree can differ from the number of adjacent rates in the smoothing penalty. A local sensitivity subtracts basal variance and computes this degree directly. Independent numerical derivatives at three interior vectors on each tree agree with the local implementation (maximum relative discrepancy below 1e-5). The installed implementation remains labelled as such; an optimizer convergence flag does not resolve its mismatch. The local version is an explicit post hoc algorithm change. Both use lambda = 1, a root age of 1 and tips at 0; dates are relative penalized-likelihood proxies. Source, diagnostic scripts and GPL attribution are included in the code and data archive.

Topology	Implementation	Convergence flag	Positive / ultrametric	Minimum duration	Maximum scaled analytic gradient
ML	installed	True	True / True	0.001702	0.00035
ML	consistent_penalty	True	True / True	0.001656	1.3e-05
ASTRAL	installed	True	True / True	0.001850	7.6e-05
ASTRAL	consistent_penalty	True	True / True	0.001792	6.1e-06

Joint model: log median terminal branch length ~ log2 cell number + differentiation state. Variants A–D are defined in Table S3d. λ is estimated by ML; an asterisk marks a retry with bounds 0.001–0.999 after the default-bound fit failed. All single-predictor fits are supplied in `astral_pgls.csv`. The installed and local chronograms used iter.max 50,000, eval.max 100,000, dual.iter.max 100 and epsilon 1e-9.

Topology labels: ML0 = archived ML chronogram; ML-A / AS-A = midpoint-root ML / ASTRAL, installed ape implementation; ML-C / AS-C = midpoint-root ML / ASTRAL, locally consistent implementation.

Topology	Variant	Predictor	n	β	SE	P	λ
ML0	A	log2 cells	57	-0.001	0.181	0.994	0.000
ML0	A	differentiation	57	-0.312	1.190	0.794	0.000
ML0	B	log2 cells	51	-0.006	0.052	0.901	0.986
ML0	B	differentiation	51	0.321	0.265	0.231	0.986
ML0	C	log2 cells	57	-0.105	0.125	0.404	0.764
ML0	C	differentiation	57	1.165	0.725	0.114	0.764
ML0	D	log2 cells	55	-0.001	0.045	0.988	0.995
ML0	D	differentiation	55	0.296	0.214	0.174	0.995
ML-A	A	log2 cells	57	0.124	0.201	0.541	0.595
ML-A	A	differentiation	57	-0.700	1.200	0.562	0.595
ML-A	B	log2 cells	51	-0.017	0.053	0.754	0.985
ML-A	B	differentiation	51	0.320	0.275	0.251	0.985

Topology	Variant	Predictor	n	β	SE	P	λ
ML-A	C	log2 cells	57	-0.087	0.127	0.498	0.849
ML-A	C	differentiation	57	1.052	0.705	0.142	0.849
ML-A	D	log2 cells	55	-0.018	0.047	0.708	0.996
ML-A	D	differentiation	55	0.312	0.230	0.180	0.996
ML-C	A	log2 cells	57	0.123	0.201	0.543	0.594
ML-C	A	differentiation	57	-0.697	1.200	0.564	0.594
ML-C	B	log2 cells	51	-0.017	0.053	0.753	0.985
ML-C	B	differentiation	51	0.320	0.275	0.250	0.985
ML-C	C	log2 cells	57	-0.087	0.127	0.499	0.852
ML-C	C	differentiation	57	1.051	0.705	0.142	0.852
ML-C	D	log2 cells	55	-0.018	0.047	0.708	0.996
ML-C	D	differentiation	55	0.312	0.230	0.180	0.996
AS-A	A	log2 cells	57	0.139	0.205	0.502	0.663
AS-A	A	differentiation	57	-0.764	1.201	0.527	0.663
AS-A	B	log2 cells	51	-0.016	0.053	0.758	0.987
AS-A	B	differentiation	51	0.317	0.271	0.248	0.987
AS-A	C	log2 cells	57	-0.090	0.127	0.481	0.854
AS-A	C	differentiation	57	1.055	0.698	0.136	0.854
AS-A	D	log2 cells	55	-0.018	0.046	0.690	0.997
AS-A	D	differentiation	55	0.314	0.226	0.171	0.997
AS-C	A	log2 cells	57	0.138	0.205	0.503	0.663
AS-C	A	differentiation	57	-0.761	1.200	0.529	0.663
AS-C	B	log2 cells	51	-0.016	0.053	0.757	0.987
AS-C	B	differentiation	51	0.317	0.271	0.248	0.987
AS-C	C	log2 cells	57	-0.090	0.127	0.481	0.857
AS-C	C	differentiation	57	1.056	0.698	0.136	0.857
AS-C	D	log2 cells	55	-0.018	0.046	0.691	0.997*
AS-C	D	differentiation	55	0.314	0.226	0.170	0.997*

Table S7d. Gene-cluster bootstrap

The primary r model was refitted in 200 replicates, each resampling the 153 contributing genes with replacement, keeping all observations of each selected gene and assigning each selected copy a new gene identifier. Weights were renormalized within each replicate. The two contrasts remain fixed. The statistic is the posterior mean of μ, calculated by averaging the conditional Gaussian mean to reduce Monte Carlo noise. Sampling follows the prespecified completion fallback schedule and diagnostic criteria; no replicate is dropped or selected by its effect. The sampling seed is 20260915 and the Stan seed is 20260911 + replicate.

Replicates	Mean estimate	SD across replicates	95% percentile interval	Fraction < 0	Maximum estimator MCSE	Maximum $\hat{R}$	Minimum bulk / tail ESS
200	0.069	0.032	0.012 to 0.134	0.010	0.0004	1.0096	591 / 418

All accepted bootstrap fits have zero divergences and tree-depth hits. The percentile interval describes gene-resampling uncertainty conditional on these two contrasts and the chosen model/prior. It is not a credible interval across evolutionary origins and does not replace the much wider hierarchical interval in Table S3g.

Table S7e. Corrected functional-category tallies

The original five-species eggNOG-to-orthogroup mapping and voting order are retained. Every category letter votes; ties use first occurrence in the fixed input order. The historical table was reproduced before applying the mapping to the corrected models. Rows below use each run's own QC set and require at least five annotated orthogroups per category. Each cell reports n and the fraction with foreground ω below background ω. H1p separates Tetrabaena (ω1) from Goniaceae + Volvocaceae (ω2). Common-QC counts and fractions, median ω values and input hashes are in the accompanying CSV files. These are descriptive category summaries, not independent evidence for stronger selection.

COG	Original n; fraction	H0u n; fraction	H1p Tetrabaena n; fraction	H1p Gon+Volv n; fraction
A	51; 0.843	73; 0.740	73; 0.466	73; 0.781
B	6; 1.000	16; 0.688	16; 0.500	16; 0.625
C	38; 0.947	60; 0.867	61; 0.590	61; 0.852
D	13; 1.000	19; 0.842	20; 0.700	20; 0.900
E	43; 0.930	61; 0.803	61; 0.492	61; 0.770
F	11; 0.909	17; 0.941	17; 0.471	17; 0.941
G	30; 0.933	44; 0.841	44; 0.545	44; 0.886
H	23; 0.870	33; 0.727	33; 0.485	33; 0.727
I	22; 0.864	36; 0.722	35; 0.571	35; 0.714
J	54; 0.926	97; 0.732	97; 0.402	97; 0.753
K	31; 0.806	51; 0.686	48; 0.521	48; 0.708
L	17; 0.824	31; 0.677	32; 0.281	32; 0.688
M	6; 1.000	9; 0.778	9; 0.556	9; 0.778
O	59; 0.915	94; 0.755	91; 0.505	91; 0.791
P	25; 1.000	35; 0.857	34; 0.588	34; 0.882
Q	18; 0.778	27; 0.519	27; 0.407	27; 0.519
S	182; 0.929	288; 0.736	288; 0.500	288; 0.774
T	42; 0.857	77; 0.675	75; 0.453	75; 0.667
U	39; 0.897	60; 0.683	56; 0.446	56; 0.732
Z	24; 1.000	29; 0.897	29; 0.448	29; 0.897

Table S7f. Reconstruction on the relative-time chronogram

The original specification also called for reconstruction on a chronogram; the recorded implementation used only the substitution-length tree. This additional omission was identified and completed post hoc. ER, ARD and ordered Mk models were fitted to the locally consistent ML chronogram, with 1,000 conditional maps at each fitted Q, equal root-state prior and seed 20260911 + scenario index. These are fixed-Q sensitivities; they are separate from the rate-MCMC chains on the original substitution-length tree. Parsimony reconstruction is unaffected because it does not use branch lengths. Both contrast rules and the original strict/relaxed gene filters are propagated; the v1 strict r results also apply both omitted observation filters. All fits and diagnostic criteria follow Table S7a.

Scenario	Rule	Design K	Foreground / background / intermediate	Max occupancy MCSE	Threshold-uncertain branches
S1_ER	v1	4	24 / 69 / 19	0.0151	4
S1_ER	v11	4	24 / 69 / 19	0.0151	4
S2_ARD	v1	3	34 / 51 / 27	0.0151	1
S2_ARD	v11	4	34 / 51 / 27	0.0151	1
S7_ORD	v1	3	34 / 52 / 26	0.0151	3
S7_ORD	v11	4	34 / 52 / 26	0.0151	3

The fitted-Q optimizers returned convergence code zero, and independent fixed-Q likelihood evaluations reproduced their log likelihoods within 1e-8. The ARD maximum-likelihood estimate has 2 rates at the numerical boundary; its fixed-Q reconstruction therefore does not express rate uncertainty.

The occupancy MCSE uses independent conditional maps at a fitted Q. Its interval crossing a classification threshold is reported as Monte Carlo uncertainty, not posterior certainty about the contrast design. These chronogram results inherit the local algorithm and relative-time limitations described in Table S7c.

Scenario codes S1, S2 and S7 denote ER, ARD and ordered Mk reconstruction on the chronogram. Gene filters and observation filters follow the labels in Table S7b.

Scenario	Rule / gene filter / statistic / observation filter	n / genes / K	μ	95% interval	P(μ < 0)	R̂	Min bulk / tail ESS
S1	v1 / strict / r / all	318 / 181 / 4	0.067	-0.283 to 0.401	0.299	1.0005	7133 / 4981
S1	v1 / strict / r / both_filters	242 / 144 / 4	0.066	-0.275 to 0.398	0.297	1.0015	4728 / 2445
S2	v1 / strict / r / all	206 / 153 / 2	0.070	-0.333 to 0.454	0.210	1.0008	4125 / 2547
S7	v1 / strict / r / all	206 / 153 / 2	0.070	-0.333 to 0.454	0.210	1.0008	4125 / 2547

Scenario	Rule / gene filter / statistic / observation filter	n / genes / K	μ	95% interval	P(μ < 0)	R̂	Min bulk / tail ESS
S2	v1 / strict / r / both_filters	160 / 126 / 2	0.068	-0.331 to 0.449	0.209	1.0008	3580 / 2689
S7	v1 / strict / r / both_filters	160 / 126 / 2	0.068	-0.331 to 0.449	0.209	1.0008	3580 / 2689
S1	v1 / strict / rad / all	204 / 125 / 4	-0.007	-0.033 to 0.023	0.776	1.0018	3822 / 3312
S2	v1 / strict / rad / all	143 / 113 / 2	-0.007	-0.358 to 0.348	0.595	1.0010	3810 / 3156
S7	v1 / strict / rad / all	143 / 113 / 2	-0.007	-0.358 to 0.348	0.595	1.0010	3810 / 3156
S1	v1 / relaxed / r / all	482 / 228 / 4	0.061	-0.236 to 0.342	0.275	1.0009	8949 / 7660
S2	v1 / relaxed / r / all	320 / 210 / 3	0.081	-0.129 to 0.272	0.112	1.0012	5852 / 4899
S7	v1 / relaxed / r / all	320 / 210 / 3	0.081	-0.129 to 0.272	0.112	1.0012	5852 / 4899
S1	v1 / relaxed / rad / all	296 / 164 / 4	-0.003	-0.026 to 0.023	0.649	1.0013	3448 / 2708
S2	v1 / relaxed / rad / all	223 / 157 / 3	-0.001	-0.089 to 0.094	0.559	1.0012	3392 / 2807
S7	v1 / relaxed / rad / all	223 / 157 / 3	-0.001	-0.089 to 0.094	0.559	1.0012	3392 / 2807
S1	v11 / strict / r / all	236 / 151 / 4	0.041	-0.231 to 0.311	0.338	1.0007	7526 / 6062
S2	v11 / strict / r / all	133 / 109 / 4	0.119	-0.056 to 0.334	0.060	1.0011	3553 / 2382
S7	v11 / strict / r / all	133 / 109 / 4	0.119	-0.056 to 0.334	0.060	1.0011	3553 / 2382
S1	v11 / strict / rad / all	115 / 88 / 3	0.002	-0.065 to 0.066	0.452	1.0009	4250 / 3878
S2	v11 / strict / rad / all	93 / 77 / 4	0.002	-0.044 to 0.044	0.413	1.0015	3580 / 3562
S7	v11 / strict / rad / all	93 / 77 / 4	0.002	-0.044 to 0.044	0.413	1.0015	3580 / 3562
S1	v11 / relaxed / r / all	418 / 215 / 4	0.048	-0.221 to 0.301	0.300	1.0005	7096 / 4442
S2	v11 / relaxed / r / all	338 / 190 / 4	0.119	-0.082 to 0.309	0.074	1.0010	4135 / 3725

Scenario	Rule / gene filter / statistic / observation filter	n / genes / K	μ	95% interval	P(μ < 0)	R̂	Min bulk / tail ESS
S7	v11 / relaxed / r / all	338 / 190 / 4	0.119	-0.082 to 0.309	0.074	1.0010	4135 / 3725
S1	v11 / relaxed / rad / all	194 / 133 / 3	-0.002	-0.051 to 0.051	0.584	1.0010	4831 / 3583
S2	v11 / relaxed / rad / all	264 / 159 / 4	-0.003	-0.045 to 0.037	0.595	1.0017	4232 / 3756
S7	v11 / relaxed / rad / all	264 / 159 / 4	-0.003	-0.045 to 0.037	0.595	1.0017	4232 / 3756

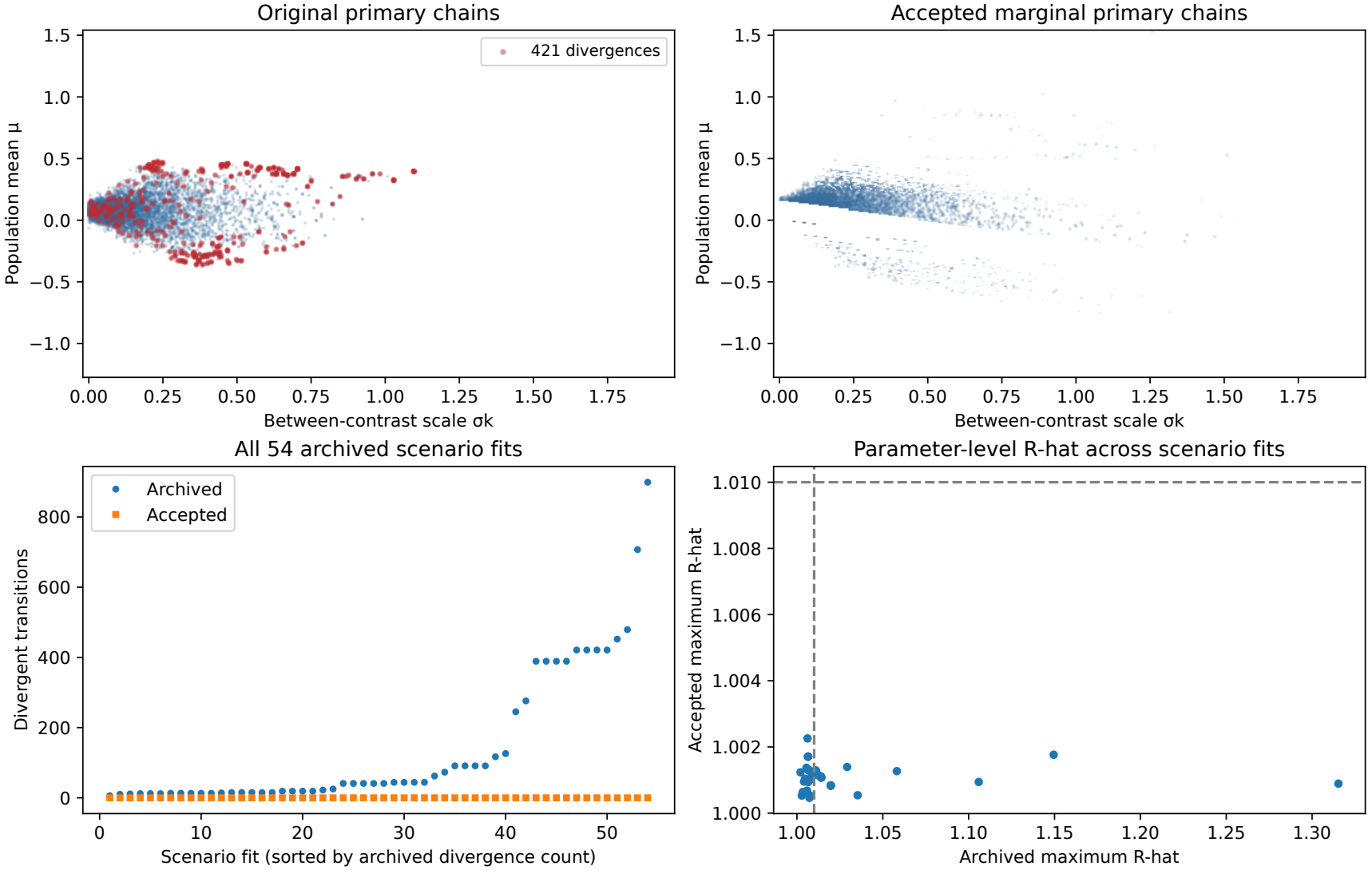

Figure S2. Sampling diagnostics. The first panels use shared axes to show all post-warm-up primary draws in the original and marginal parameterizations; divergent draws are red. The lower panels compare every archived scenario fit with its accepted refit. Accepted fits have no divergences. Full parameter and chain diagnostics are provided for all fits.

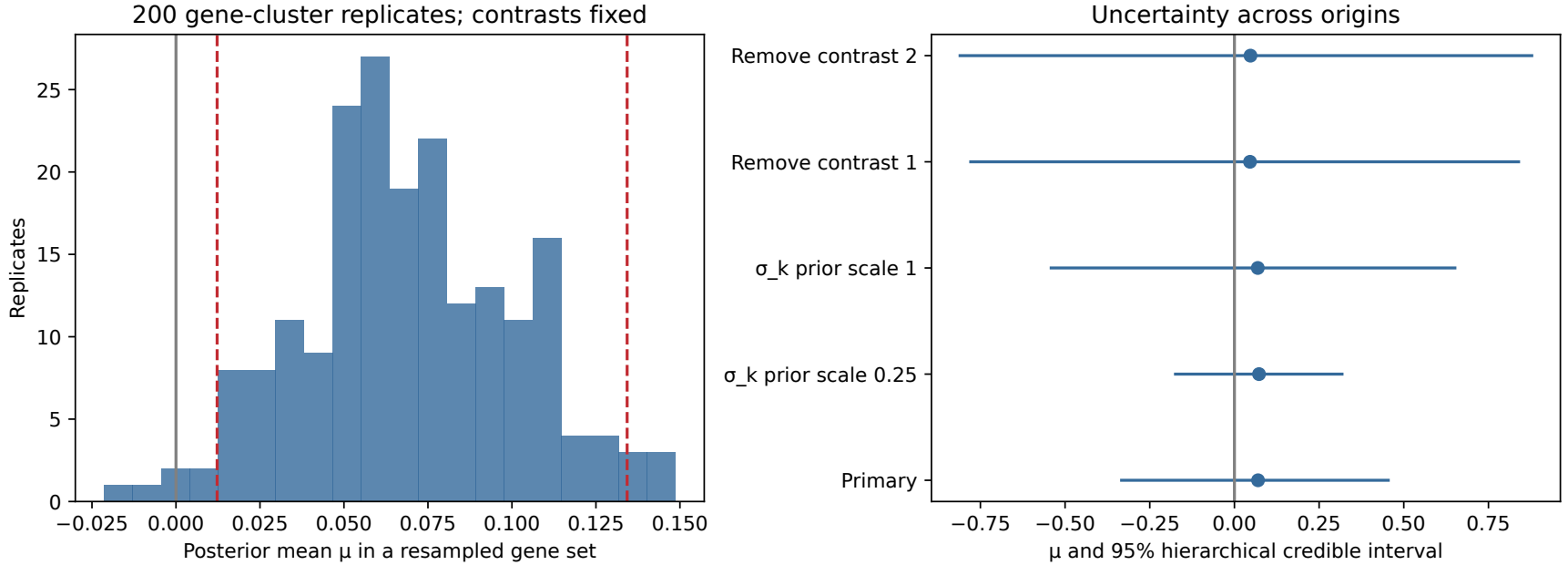

Figure S3. Conditional gene-cluster bootstrap and origin-level uncertainty. Left: all 200 gene-resampling estimates; dashed bounds are their percentile interval. Right: hierarchical credible intervals for the primary model, prior-scale sensitivities and leave-one-contrast-out fits. These intervals describe different sources of uncertainty.

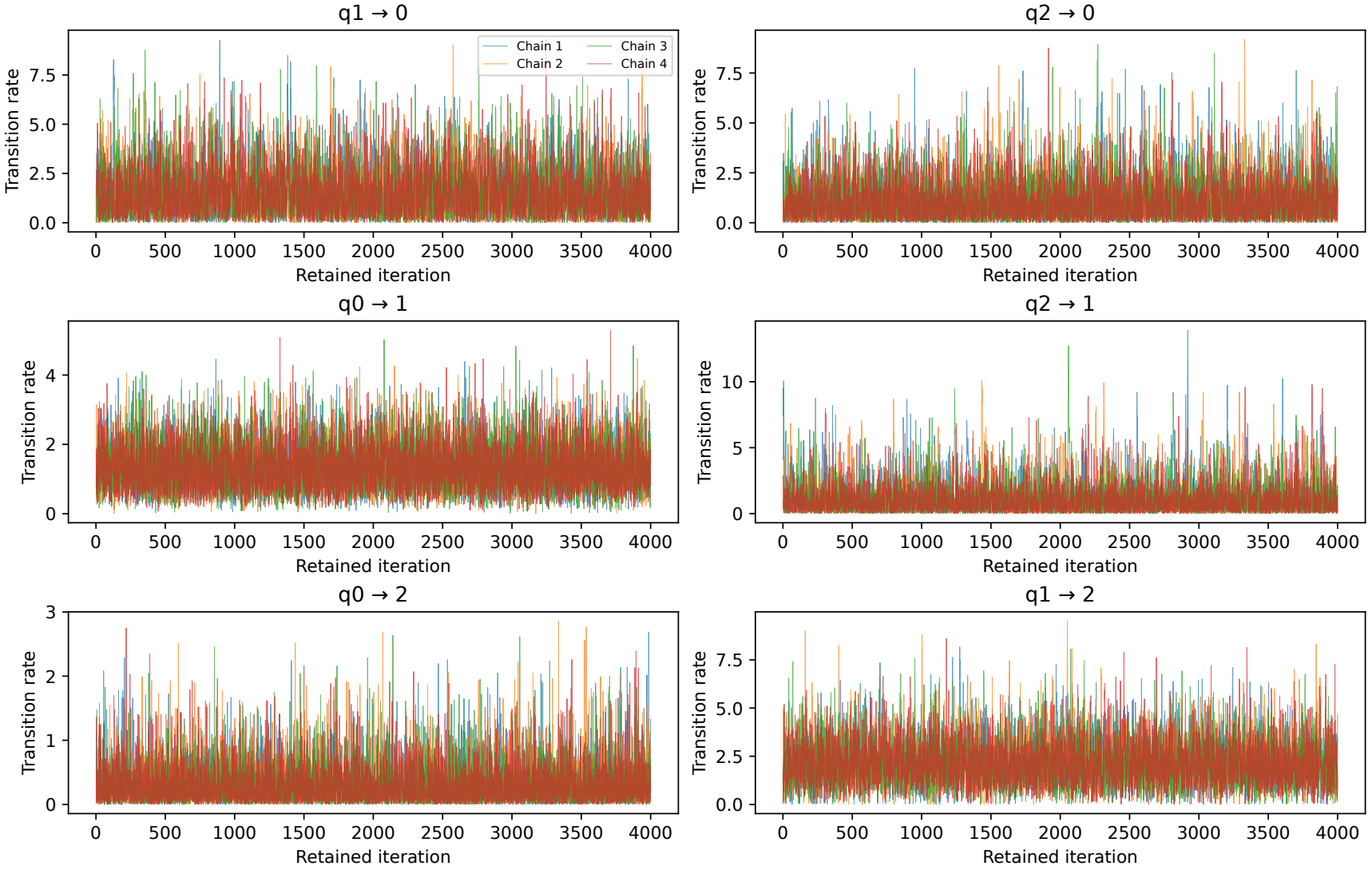

Figure S4. Retained transition-rate draws in the four independent mapping chains. All six ARD rates are shown; chain diagnostics are in Table S7b and the archive.