Online Resource 1

Decoding assumptions and inference about a robustness preference in codon reassignments along sampled histories

Jonathan Reich
Independent Researcher, Melbourne, Victoria, Australia
Corresponding author: jonathanreich100@gmail.com
ORCID: 0009-0000-0108-0717

Supplementary material for the main article. Tables, sections and propositions numbered without S refer to the main article.

S1 Codon capture

The disappearance proxy of a codon set X of family s is $P_{\text{dis}} = (1-f)^N$, with f the composition-expected share of X in the family and N the expected count of the family's other codons, scored at the reconstructed stem of each origin from fourfold-site base frequencies and codon frequencies reconstructed by Brownian motion on the log scale and renormalised. Sengupta et al. (2007, p. 667) define the same quantity with N the total number of stop codons, and apply it in post-change proxy genomes with the count of the other stop codons (p. 669); Swire et al. (2005, p. 129) used sister-taxon usage as the proxy for the ancestral genome. The statistic sums each moved set's mid-rank in disappearance-proneness within its pool, and its null, exact under the stated model, draws pool members uniformly (Table S1). At the stem the moved sets rank slightly above random sets of the same type and kind (25.88 against 23.5, $P = 0.093$, $n = 47$), and the bootstrap over reconstructed traits spans $P = 0.027$ to 0.45. The controls do not support it: in the same-family pool, which compares the moved set only with codon sets of its own synonymous family, where disappearance has to happen, $P = 0.88$; at the crown $P = 0.32$, and in the crown same-family pool the moved sets rank lower than chance (lower-tail $P = 0.018$); with at least two ancestral-side tips $P = 0.24$. The mitochondrial subset, stated in advance as a sensitivity, gives $P = 0.035$. The counting convention matters: with the expected total count $N/(1-f)$ in the exponent, which is what a zero count among all family positions implies when only the other codons are counted, the stem result becomes $P = 0.010$ (mitochondria 0.0027), mainly through two-member stop pools in which each origin can only score 0.25 or 0.75, while the stem same-family comparison still does not detect enrichment ($P = 0.48$). Scored in the extant genomes, the same origins give the opposite tail ($P = 0.011$ for depletion), largely because a genome that already reads `UGA` as Trp leaves its comparator `UAA`+`UAG` with no competitor; stem scoring removes that artefact.

The quantities validate differently. On the stop-dominated validation set the pre-stated proxy flags 4 of 10 capture-supported cases, with class confounded with kind. On moved codons that are not stop codons (4 capture-supported against 7 not; Table S1, panel b, with the variants of the validation set), the ancestral relative synonymous codon usage of the moved codons is lower in the supported cases (one-sided Mann–Whitney $P = 0.0061$), the pre-stated proxy flags 2 of 4 supported cases and 7 of 7 not-capture cases correctly, but the disappearance rule flags 0 of 4, and the per-origin mid-rank that the test sums does not separate the classes ($P = 0.42$). Low ancestral usage of the moved codons thus separates the small literature classes, whereas the disappearance rank that the historical test sums does not. Codon capture is therefore a weak,

(a) Historical tests

		$(1-f)^N$		$(1-f)^{N/(1-f)}$
Analysis	n	P upper	P lower	P upper
stem (primary)	47	0.093	0.91	0.010
same-family pool, stem	47	0.88	0.14	0.48
crown	47	0.32	0.68	0.072
crown, same-family pool	47	0.99	0.018	0.86
at least two ancestral-side tips	37	0.24	0.76	0.14
mitochondria only	30	0.035	0.96	0.0027
stem, taxonomy tree	47	0.080	0.92	0.0088
extant scores, same origins	47	0.99	0.011	0.99

(b) Sense-codon validation

			Usage	Mid-rank P		Usage proxy		Rule C
Validation set	Supp.	Not	P	$(1-f)^N$	$(1-f)^{N/(1-f)}$	sensitivity	specificity	sensitivity
sense to sense (primary)	4	7	0.0061	0.42	0.16	2 of 4	7 of 7	0 of 4
M23 dropped	4	6	0.0095	0.48	0.24	2 of 4	6 of 6	0 of 4
with sense to stop M37, M42	4	9	0.0028	0.24	0.097	2 of 4	9 of 9	0 of 4
M23 dropped, with M37, M42	4	8	0.0040	0.28	0.14	2 of 4	8 of 8	0 of 4

(c) Capture weighting of the target-choice null, six merge histories

Model	λ	Observed	Expected	P ($2\bar{P}$)	Largest P
`main`	0	38.1	40.1	1.00	0.77
`main`	1	38.1	39.1	1.00	0.69
`comp`	0	38.1	35.5	0.58	0.34
`comp`	1	38.1	36.2	0.67	0.40

Table S1: Codon capture. (a) Historical capture tests of the disappearance proxy, with the controls and the alternative counting convention. Upper tail: moved sets more disappearance-prone than their pool. Bootstrap of the primary over reconstructed traits: upper-tail P from 0.027 to 0.45 under $(1-f)^N$ and from 0.0017 to 0.066 under $(1-f)^{N/(1-f)}$. (b) The validation on moved codons that are not stop codons: capture-supported and not-capture cases; one-sided Mann–Whitney P for lower ancestral relative synonymous codon usage of the moved codons in the supported cases; one-sided Mann–Whitney P for a higher per-origin mid-rank (the score the historical test sums) under each counting convention; the pre-stated usage proxy's sensitivity and specificity; and the sensitivity of the disappearance rule (rule C). Usage, the usage proxy and rule C do not depend on the counting convention. On the stop-dominated validation set the proxy flags 4 of 10 capture-supported cases. (c) Observed and expected optimum-attaining transitions under the two-stage target-choice null, in which the moved codon set is drawn from the sets of the same type and kind with probability proportional to P_{dis}^{λ} ($\lambda = 0$: uniformly) and the target with the model's accessibility weights, weighted over the six merge histories, with the merged $2\bar{P}$ and the largest per-history P.

test-dependent signal in a composition proxy, sensitive to the comparator and to the counting convention. In the six-merge-history weighting sensitivity, weighting codon sets by P_{dis} in the target-choice null moves the expected number of optimum-attaining transitions under `main` from 40.1 to 39.1 (observed 38.1; Table S1, panel c), so capture weighting changes the expectation little. This sensitivity was run on the six merge histories only, not on the sampled histories of the primary analysis.

S2 The dynamic programme for the packing maximum

Fix a partition of the codons into B blocks of size q ($q = 4$ for boxes) and a code with fixed stop codons. Let block β hold c_β sense codons, and let the m sense classes have sizes $n_1 \geq n_2 \geq \cdots \geq n_m$, so that $\sum_i n_i = \sum_\beta c_\beta$. A placement assigns $x_{i\beta} \geq 0$ codons of class i to block β, with $\sum_\beta x_{i\beta} = n_i$ and $\sum_i x_{i\beta} = c_\beta$; the sense contribution to the within-block silent-edge count is $\sum_{i,\beta} \binom{x_{i\beta}}{2}$, and its maximum over placements, plus the fixed stop–stop term, is the fixed-stop maximum. Blocks with equal remaining capacity are interchangeable, so a state is a vector $a = (a_0, \ldots, a_q)$ in which a_r counts the blocks with r sense slots still unfilled. Let $V(i, a)$ be the maximum of $\sum_{l \geq i} \sum_\beta \binom{x_{l\beta}}{2}$ over placements of classes $i, \ldots, m$ that fill the capacities a exactly, with $V(i, a) = -\infty$ when no such placement exists. The terminal values are $V(m + 1, a) = 0$ if $a_1 = \cdots = a_q = 0$ and $V(m + 1, a) = -\infty$ otherwise, and for $i \leq m$

$$V(i, a) = \max_{\pi} \Big\{ \sum_{(r,j,k) \in \pi} k \tbinom{j}{2} + V(i + 1, a^\pi) \Big\},$$

where π ranges over multisets of triples (r, j, k) (a level r, a part $1 \leq j \leq r$ and a number of blocks $k \geq 1$) with $\sum_{(r,j,k) \in \pi} jk = n_i$ and, for each level r, $\sum k \leq a_r$ over the triples at that level; a^π is obtained from a by moving, for each triple, k blocks from level r to level $r - j$; and a maximum over no admissible π is $-\infty$. The programme starts from $a^{(0)}$, where $a^{(0)}_r$ is the number of blocks with $c_\beta = r$, and the sense maximum is $V(1, a^{(0)})$. Each step removes n_i units of capacity, so every state reached from $a^{(0)}$ at stage i has total capacity $\sum_r r a_r = \sum_{l \geq i} n_l$; every such state admits a placement (each remaining class can be split over the blocks that still have capacity), so all values met in the computation are finite. With B blocks there are at most $\binom{B+q}{q}$ states per stage (4,845 for 16 boxes and $q = 4$). Storing a maximising π at each state and replaying the choices from $a^{(0)}$ yields a placement, and hence a witness code, attaining the maximum.

S3 Supplementary tables

References

Barcytė D, Žihala D, Eliáš M (2025) Expanded diversity of pedinophytes provides a window into the evolution of the genetic code in organelles. PLoS Genet 21:e1011901. `https://doi.org/10.1371/journal.pgen.1011901`

Błażej P, Wnętrzak M, Mackiewicz D, Gagat P, Mackiewicz P (2019) Many alternative and theoretical genetic codes are more robust to amino acid replacements than the standard genetic code. J Theor Biol 464:21–32. `https://doi.org/10.1016/j.jtbi.2018.12.030`

Fučíková K, Lewis PO, Lewis LA (2016) Chloroplast phylogenomic data from the green algal order Sphaeropleales (Chlorophyceae, Chlorophyta) reveal complex patterns of sequence evolution. Mol Phylogenet Evol 98:176–183. `https://doi.org/10.1016/j.ympev.2016.01.022`

Model	β, 20 histories	Pairs	Largest shift over pairs			Pairs changing	
			median	lower end	upper end	zero inclusion	decision
`comp fixed`	0.15 (−1.01, 1.37)	190	0.14	0.15	0.25	0	0
`compi fixed`	0.36 (−0.97, 1.70)	190	0.16	0.14	0.18	0	0
`compr fixed`	0.67 (−0.71, 1.94)	190	0.17	0.12	0.22	0	0
`compr mix`	−0.01 (−1.44, 1.37)	190	0.27	0.25	0.31	0	0
`compri fixed`	0.97 (−0.61, 2.37)	190	0.18	0.14	0.28	0	0
`compri mix`	−0.23 (−1.69, 1.20)	190	0.12	0.11	0.15	0	0
`lineage fixed`	0.12 (−1.04, 1.34)	190	0.17	0.14	0.22	0	0
`main fixed`	−0.43 (−1.94, 1.11)	190	0.14	0.16	0.21	0	0
`pooli fixed`	−0.40 (−1.99, 1.22)	190	0.15	0.18	0.24	0	0

Table S2: The two-history approximation on the observed data: every two-history mixture of the 20 fitted primary histories (specification `reach`) against the 20-history mixture, for the median and 95% interval of β, whether the interval contains zero, and the decision $P(\beta > 0) > 0.975$.

Grafen A (1989) The phylogenetic regression. Philos Trans R Soc Lond B 326:119–157. https://doi.org/10.1098/rstb.1989.0106

Haag S, Sloan KE, Ranjan N, Warda AS, Kretschmer J, Blessing C, Hübner B et al. (2016) NSUN3 and ABH1 modify the wobble position of mt-tRNA^{Met} to expand codon recognition in mitochondrial translation. EMBO J 35:2104–2119. https://doi.org/10.15252/embj.201694885

Knight RD, Freeland SJ, Landweber LF (2001a) Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet 2:49–58. https://doi.org/10.1038/35047500

Knight RD, Landweber LF, Yarus M (2001b) How mitochondria redefine the code. J Mol Evol 53:299–313. https://doi.org/10.1007/s002390010220

Kollmar M, Mühlhausen S (2017) Nuclear codon reassignments in the genomics era and mechanisms behind their evolution. BioEssays 39:1600221. https://doi.org/10.1002/bies.201600221

Krassowski T, Coughlan AY, Shen XX, Zhou X, Kominek J, Opulente DA, Riley R et al. (2018) Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. Nat Commun 9:1887. https://doi.org/10.1038/s41467-018-04374-7

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB (2022) TimeTree 5: an expanded resource for species divergence times. Mol Biol Evol 39:msac174. https://doi.org/10.1093/molbev/msac174

Marshall AW, Olkin I, Arnold BC (2011) Inequalities: theory of majorization and its applications, 2nd edn. Springer, New York. https://doi.org/10.1007/978-0-387-68276-1

Matsuyama S, Ueda T, Crain PF, McCloskey JA, Watanabe K (1998) A novel wobble rule found in starfish mitochondria. J Biol Chem 273:3363–3368. https://doi.org/10.1074/jbc.273.6.3363

Nakano S, Suzuki T, Kawarada L, Iwata H, Asano K, Suzuki T (2016) NSUN3 methylase initiates 5-formylcytidine biogenesis in human mitochondrial tRNA^{Met}. Nat Chem Biol 12:546–551. https://doi.org/10.1038/nchembio.2099

Noutahi E, Calderon V, Blanchette M, El-Mabrouk N, Lang BF (2019) Rapid genetic code evolution in green algal mitochondrial genomes. Mol Biol Evol 36:766–783. https://doi.org/10.1093/molbev/msz016

(a) Upper-tail ranks under the sequential generator, beside the paired frozen skeleton

			Sequential				Frozen skeleton	
Generator	Test	Odds	$(k+1)/(n+1)$	k/n	95% CI	Over seeds	$(k+1)/(n+1)$	k/n
restrictive	`compri cell`	1.5	0.099	9/100	0.042–0.164	0.079–0.119	0.109	10/100
		2	0.025	4/200	0.005–0.050	0.025–0.035	0.0498	9/200
		2.5	0.025	4/200	0.005–0.050	0.025–0.025	0.020	3/200
		3	0.010	1/200	0.000–0.028	0.010–0.010	0.015	2/200
	`main`	1.5	0.139	13/100	0.071–0.212	0.119–0.168	0.168	16/100
		2	0.045	8/200	0.017–0.077	0.040–0.060	0.065	12/200
		2.5	0.025	4/200	0.005–0.050	0.025–0.030	0.025	4/200
		3	0.010	1/200	0.000–0.028	0.010–0.010	0.015	2/200
pooled	`compri cell`	2	0.040	3/100	0.006–0.085	0.040–0.040	0.0495	4/100
		2.5	0.025	4/200	0.005–0.050	0.025–0.025	0.030	5/200
		3	0.005	0/200	0.000–0.018	0.005–0.005	0.005	0/200
	`main`	2	0.059	5/100	0.016–0.113	0.059–0.069	0.089	8/100
		2.5	0.025	4/200	0.005–0.050	0.025–0.035	0.045	8/200
		3	0.005	0/200	0.000–0.018	0.005–0.005	0.010	1/200

(b) Rejections under the sequential no-preference generators

		Calibrated detection rule		Merged $2\bar{P}$ rule	
Generator	Test	k/n	rate (95% CI)	k/n	rate (95% CI)
restrictive	`uniform`	7/140	0.050 (0.020–0.100)	38/140	0.271 (0.200–0.353)
	fixed-state `compri`	8/140	0.057 (0.025–0.109)	13/140	0.093 (0.050–0.154)
	`compri cell`	9/140	0.064 (0.030–0.119)	2/140	0.014 (0.002–0.051)
	`main`	8/140	0.057 (0.025–0.109)	2/140	0.014 (0.002–0.051)
pooled	`uniform`	5/140	0.036 (0.012–0.081)	140/140	1.000 (0.974–1.000)
	fixed-state `compri`	8/140	0.057 (0.025–0.109)	140/140	1.000 (0.974–1.000)
	`compri cell`	6/140	0.043 (0.016–0.091)	5/140	0.036 (0.012–0.081)
	`main`	6/140	0.043 (0.016–0.091)	5/140	0.036 (0.012–0.081)

(c) Generation audit and checks, all sequential cells

Generation	Count	Check	Result
replicates (9 cells)	1,480	predecessor code	0 of 128,549 parts
units	112,897	inadmissible	0 of 128,549 parts
on a changed predecessor code	34,035	catalogue-codon coverage	0
target differs from frozen draw	10,740	tips	0 of 2,949,640
parts (members of units)	128,549	carrier rows	0 of 341,880
parts with a changed source	8,439	rows not compared	205
parts split	0	node constraints	0 of 1,480
later no-ops, full	37	coupling	0 of 78,862 units
later no-ops, partial	3	failed replicates	0

Table S3: The sequential generator: each simulated unit chooses its target or target tuple on its simulated predecessor code. Accessibility coefficients are refitted at each tested preference on the reference (first truth) history and then used in the sequential generator. (a) Upper-tail ranks of the observed statistic, with Clopper–Pearson intervals for k/n and the range over the 20 sampler seeds of the observed statistic, beside the paired frozen-skeleton rank. These are pointwise ranks under fitted conditional generators; repeated-sampling exclusion coverage for a procedure that re-estimates histories and accessibility has not been established. No replicate ties the observed statistic (0 ties); a rank that three decimals would round to 0.050 has four. (b) The calibrated detection rule (critical value from the frozen-skeleton no-preference cell of the same generator) and the merged $2\bar{P}$ rule; the calibrated cell test rejects in 9/140 replicates under the restrictive generator and the pooled test in 6/140 under the pooled one. These replicates reuse the seeds of the frozen no-preference replicates, of the original cells that set the critical values and of the validation cells (Table S14), so they check how the frozen critical values transport to the sequential generator, not a fresh, independent validation. (c) Every check is a mismatch count. Coupling: a unit on the true predecessor code must receive the paired frozen target; rows not compared: carriers whose codons received different targets.

Model	Route hit, directly documented (17 groups)	Route hit minus co-occupancy-only	Route rank minus neighbour-only	Target hit minus co-occupancy-only
`main`	0.68	0.394 (0.251 to 0.541)	0.024 (−0.062 to 0.112)	0.047 (0.005 to 0.100)
`comp`	0.62	0.351 (0.196 to 0.500)	0.045 (−0.047 to 0.140)	0.080 (0.008 to 0.152)
`compr`	0.56	0.396 (0.249 to 0.543)	0.106 (0.004 to 0.205)	−0.013 (−0.105 to 0.077)
`pooli`	0.49	0.370 (0.243 to 0.495)	−0.037 (−0.112 to 0.035)	0.044 (0.001 to 0.095)
`compi`	0.37	0.302 (0.178 to 0.424)	−0.010 (−0.099 to 0.078)	0.045 (−0.014 to 0.098)
`compri`	0.30	0.343 (0.224 to 0.465)	0.074 (−0.035 to 0.175)	−0.098 (−0.191 to −0.004)
`lineage`	0.62	0.372 (0.215 to 0.520)	0.034 (−0.058 to 0.127)	0.086 (0.013 to 0.157)

Table S4: The benchmark's paired and direct-evidence results: the route hit rate on the groups whose route is directly documented, and paired differences over groups (bootstrap 95% intervals) of the route hit rate from the co-occupancy-only baseline (positive: better than the baseline), of the mean normalised route rank from the neighbour-only baseline (positive: worse) and of the target hit rate from the co-occupancy-only baseline (positive: better).

Model	Odds	$d < \infty$	$d \leq 1$	$d = 0$	Range over the first ten truth histories		
`compri`	1	1.989	2.565	2.091	1.88–2.05	2.47–2.74	1.82–2.35
`compri`	1.5	1.905	2.561	1.842	1.80–1.96	2.47–2.74	1.57–2.13
`compri`	2	1.853	2.561	1.688	1.74–1.91	2.47–2.74	1.42–1.99
`compri`	2.5	1.815	2.562	1.581	1.71–1.87	2.48–2.74	1.30–1.90
`compri`	e	1.802	2.563	1.544	1.69–1.85	2.48–2.74	1.26–1.86
`compri`	3	1.787	2.565	1.503	1.68–1.84	2.48–2.74	1.22–1.83
`main`	1	0.000	1.618	2.200	0.00–0.00	1.60–1.79	2.09–2.28
`main`	1.5	0.000	1.627	1.941	0.00–0.00	1.61–1.80	1.84–2.02
`main`	2	0.000	1.635	1.773	0.00–0.00	1.61–1.81	1.67–1.86
`main`	2.5	0.000	1.644	1.653	0.00–0.00	1.62–1.82	1.55–1.74
`main`	e	0.000	1.648	1.610	0.00–0.00	1.62–1.82	1.51–1.70
`main`	3	0.000	1.652	1.560	0.00–0.00	1.62–1.82	1.46–1.65

Table S5: Accessibility coefficients of the generators: MAP fits of the reach design under the model's priors, with the robustness coefficient fixed at $\log \psi$ (the constrained fit; $\psi = 1$ is the no-preference fit), on the first truth history, and their range when refitted on each of the first ten.

Model	Units	Class inf.	Non-const.	β-inf.	Inacc.	Fitted	Constant units
`main`	66–68	11	64–66	47–49	0	66–68	single candidate (sense->stop pool of size 1): M37, M42
`comp`	66–68	23	64–66	47–49	0	66–68	single candidate (sense->stop pool of size 1): M37, M42
`compr`	66–68	26	64–66	47–49	0	66–68	single candidate (sense->stop pool of size 1): M37, M42
`pooli`	66–68	12	62–64	47–49	0	66–68	every candidate shares the reach classes and at_max: B10, M32; single candidate (sense->stop pool of size 1): M37, M42
`compi`	66–68	23	56–58	47–49	0	66–68	every candidate shares the reach classes and at_max: B07, B08, B09, B10, M32, N26, N27, N28; single candidate (sense->stop pool of size 1): M37, M42
`compri`	66–68	40–42	56–58	47–49	0	66–68	every candidate shares the reach classes and at_max: B07, B08, B09, B10, M32, N26, N27, N28; single candidate (sense->stop pool of size 1): M37, M42
`lineage`	66–68	23	64–66	47–49	0	66–68	single candidate (sense->stop pool of size 1): M37, M42

Table S6: Observation flow on the six merge histories (ranges over histories). Units: units per history, the same under every model. Class inf.: units whose class stratum holds both outcomes. Non-const.: units with a non-constant likelihood. β-inf.: units whose pool varies in the optimum. Inacc.: observed targets inaccessible under the model. Fitted: units in the hierarchical arrays (all of them). Constant units: those with a constant likelihood in the modal history, with the reason; keeping them leaves the likelihood unchanged.

Model	β, all units	β, per model	β, intersection	P, all	P, intersection
`main`	−0.75 (−2.36, 0.91)	−0.74 (−2.35, 0.88)	−0.75 (−2.35, 0.91)	1.00	1.00
`comp`	−0.03 (−1.22, 1.19)	−0.02 (−1.15, 1.18)	−0.02 (−1.20, 1.19)	1.00	1.00
`compr`	0.69 (−0.62, 1.89)	0.68 (−0.61, 1.90)	0.65 (−0.60, 1.87)	1.00	1.00
`pooli`	−0.71 (−2.35, 0.98)	−0.70 (−2.36, 1.06)	−0.71 (−2.33, 0.96)	1.00	1.00
`compi`	0.11 (−1.14, 1.43)	0.11 (−1.17, 1.48)	0.12 (−1.16, 1.42)	1.00	1.00
`compri`	0.95 (−0.58, 2.32)	0.94 (−0.57, 2.30)	0.96 (−0.61, 2.32)	0.0022	0.0022
`lineage`	−0.05 (−1.20, 1.21)	−0.05 (−1.21, 1.18)	−0.05 (−1.21, 1.16)	1.00	1.00

Table S7: The three unit sets on the six merge histories, at a fixed decoding state: hierarchical β (median, 95% interval) and merged class-conditional P.

Model	Obs.	Unif. E	Unif. P	Class E	Class P	$\bar{P}$ (range)	Inform.
`main`	30.6	16.9	4.8×10^{-5}	31.2	1.00	0.75 (0.75–0.75)	11
`comp`	30.6	16.9	4.8×10^{-5}	32.2	1.00	0.84 (0.84–0.85)	23
`compr`	30.6	16.9	4.8×10^{-5}	30.7	1.00	0.59 (0.58–0.60)	40
`pooli`	30.6	16.9	4.8×10^{-5}	31.4	1.00	0.78 (0.78–0.78)	12
`compi`	30.6	16.9	4.8×10^{-5}	32.0	1.00	0.82 (0.81–0.82)	23
`compri`	30.6	16.9	4.8×10^{-5}	24.3	0.046	0.023 (0.014–0.028)	41
`lineage`	30.6	16.9	4.8×10^{-5}	32.2	1.00	0.84 (0.83–0.84)	23

Table S8: As Table 4, on the six merge histories, weighted by their reconstruction weights (`compr` and `compri` with the default priors, by the earlier prior mixture, which is not exact). Inform.: units of the modal history.

Model	β median	95% interval	$P(\beta > 0)$	e^{median}	Table-1 analysis
`main`	−0.75	(−2.36, 0.91)	0.18	0.47	−0.34 (−1.84, 1.24)
`comp`	−0.03	(−1.22, 1.19)	0.48	0.98	0.26 (−0.89, 1.44)
`compr`	−0.12	(−1.52, 1.25)	0.43	0.89	0.74 (−0.61, 1.96)
`pooli`	−0.71	(−2.35, 0.98)	0.20	0.49	−0.36 (−1.99, 1.27)
`compi`	0.11	(−1.14, 1.43)	0.58	1.12	0.34 (−0.98, 1.70)
`compri`	−0.29	(−1.71, 1.10)	0.34	0.74	0.98 (−0.46, 2.27)
`lineage`	−0.05	(−1.20, 1.21)	0.46	0.95	0.27 (−0.90, 1.45)

Table S9: As Table 6, on the six merge histories.

Model	Decoding state	Class E	Class P	Largest P	β	$P(\beta > 0)$	Mito. slope
`compr`	fixed state	31.2	1.00	0.69	0.69 (−0.62, 1.89)	0.87	1.18 (0.19, 2.56)
`compr`	E3c, default priors	30.7	1.00	0.60	−0.12 (−1.52, 1.25)	0.43	0.21 (−1.17, 1.75)
`compr`	E3c, low priors	31.3	1.00	0.70	0.66 (−0.64, 1.89)	0.87	1.17 (0.15, 2.54)
`compr`	E3c, high priors	30.6	1.00	0.59	−0.55 (−2.01, 0.90)	0.21	−0.16 (−1.61, 1.42)
`compri`	fixed state	21.2	0.0022	0.0016	0.95 (−0.58, 2.32)	0.90	2.09 (1.00, 3.51)
`compri`	E3c, default priors	24.3	0.046	0.028	−0.29 (−1.71, 1.10)	0.34	0.16 (−1.26, 1.79)
`compri`	E3c, low priors	21.5	0.0036	0.0026	0.89 (−0.64, 2.20)	0.89	1.93 (0.78, 3.34)
`compri`	E3c, high priors	27.7	0.36	0.20	−0.51 (−2.04, 0.95)	0.24	0.03 (−1.44, 1.66)

Table S10: The restrictive models at a fixed decoding state and with the three settings of the modification-availability priors (the earlier prior mixture, which is not exact), on the six merge histories.

Design or sampler	Histories	Units	P `main`	P `compri`	β `main`	β `compri`
primary_catalogue	200	64–84	1.00	0.0041	−0.43 (−1.94, 1.11)	0.97 (−0.61, 2.37)
primary_all	200	83–100	1.00	0.0029	−0.50 (−1.97, 1.02)	0.96 (−0.56, 2.34)
clade_error_catalogue	50	62–80	1.00	0.023	−0.43 (−1.92, 1.16)	0.83 (−0.70, 2.25)
classes_20_catalogue	50	67–80	1.00	0.0053	−0.40 (−1.85, 1.16)	0.95 (−0.65, 2.37)
classes_min_catalogue	50	70–83	1.00	0.0029	−0.39 (−1.92, 1.18)	1.02 (−0.61, 2.46)
computational_carrier_error_catalogue	50	61–74	1.00	0.0043	−0.42 (−1.99, 1.14)	0.96 (−0.58, 2.36)
error_x2_catalogue	50	65–83	1.00	0.0066	−0.39 (−1.97, 1.24)	1.02 (−0.58, 2.50)
exact_tips_catalogue	50	70–83	1.00	0.0011	−0.49 (−2.04, 1.12)	0.96 (−0.64, 2.37)
fixed_rates_catalogue	50	68–82	1.00	0.0047	−0.34 (−1.88, 1.25)	0.98 (−0.65, 2.41)
flat_root_catalogue	50	56–80	1.00	0.094	−0.43 (−2.02, 1.21)	0.75 (−0.90, 2.21)
grafen_lengths_catalogue	50	62–72	1.00	0.0084	−0.46 (−1.97, 1.07)	0.84 (−0.65, 2.20)
hard_readers_catalogue	50	68–85	1.00	0.0037	−0.37 (−1.89, 1.25)	0.97 (−0.62, 2.43)
hydro_ala_catalogue	50	67–84	1.00	0.0036	−0.44 (−1.95, 1.10)	0.98 (−0.56, 2.36)
ncbi_trees_catalogue	6	76–84	1.00	0.0012	–	–
no_literature_catalogue	50	68–84	1.00	0.014	−0.35 (−1.90, 1.28)	0.91 (−0.65, 2.40)
no_readers_catalogue	50	72–85	1.00	0.0018	−0.43 (−1.94, 1.11)	0.97 (−0.60, 2.31)
ott_bases_catalogue	50	72–83	1.00	0.0034	−0.37 (−1.91, 1.30)	0.94 (−0.66, 2.40)
primary_ordered_catalogue	400	74–94	1.00	0.0022	−0.23 (−1.71, 1.32)	1.09 (−0.43, 2.46)
s1_tips_catalogue	50	67–82	1.00	0.0022	−0.45 (−1.97, 1.11)	1.00 (−0.57, 2.38)
unit_lengths_catalogue	50	65–77	1.00	0.0021	−0.40 (−1.92, 1.13)	0.92 (−0.67, 2.35)
unlinked_catalogue	50	72–89	1.00	0.0019	−0.26 (−1.81, 1.32)	1.05 (−0.61, 2.53)

Table S11: Every sampled-history design: the primary catalogue design, every sampled change, and the sensitivity samplers (catalogue design; `compri` at a fixed decoding state). Class P: merged; β: median (95% interval). `primary_ordered`: the order replicates; `unlinked`: every codon its own character; `hard_readers`: no reader replacement; `ncbi_trees`: six draws only, not fitted.

Generator	Odds	n	Restrictive, cell null		Pooled	
			$2\bar{P}$	calibrated	$2\bar{P}$	calibrated
independent states	1.5	160	0.125	0.250 (0.185–0.324)	0.131	0.225
	2	160	0.275	0.450 (0.371–0.531)	0.269	0.419
	e	40	0.500	0.725 (0.561–0.854)	0.500	0.650
	3	60	0.550	0.683 (0.550–0.797)	0.550	0.667
pooled	1.5	140	0.086	0.086	0.079	0.093 (0.050–0.154)
	2	140	0.193	0.193	0.157	0.164 (0.107–0.236)
	e	40	0.325	0.325	0.200	0.275 (0.146–0.439)
	3	40	0.425	0.450	0.350	0.400 (0.249–0.567)

Table S12: Power of the merged $2\bar{P}$ rule and of the calibrated detection rule (the mean per-history P against its distribution over the generator's no-preference replicates) at four odds of preference, with Clopper–Pearson intervals for the calibrated rates; n: replicates of the cell. These are the frozen-skeleton power experiments, with accessibility at the no-preference fit; they are separate from the constrained-fit ranks of Tables 8 and S3.

Reich J (2026) Decoding accessibility and constrained optimality in natural codon reassignments. Manuscript.

Sengupta S, Higgs PG (2015) Pathways of genetic code evolution in ancient and modern organisms. J Mol Evol 80:229–243. https://doi.org/10.1007/s00239-015-9686-8

Sengupta S, Yang X, Higgs PG (2007) The mechanisms of codon reassignments in mitochondrial genetic codes. J Mol Evol 64:662–688. https://doi.org/10.1007/s00239-006-0284-7

Shulgina Y, Eddy SR (2021) A computational screen for alternative genetic codes in over 250,000 genomes. eLife 10:e71402. https://doi.org/10.7554/eLife.71402

Swire J, Judson OP, Burt A (2005) Mitochondrial genetic codes evolve to match amino acid requirements of proteins. J Mol Evol 60:128–139. https://doi.org/10.1007/s00239-004-0077-9

Turmel M, Lopes dos Santos A, Otis C, Sergerie R, Lemieux C (2019) Tracing the evolution of the plastome and mitogenome in the Chloropicophyceae uncovered convergent tRNA gene losses and a variant plastid genetic code. Genome Biol Evol 11:1275–1292. https://doi.org/10.1093/gbe/evz074

Yokobori S, Suzuki T, Watanabe K (2001) Genetic code variations in mitochondria: tRNA as a major determinant of genetic code plasticity. J Mol Evol 53:314–326. https://doi.org/10.1007/s002390010221

Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol 39:306–314. https://doi.org/10.1007/BF00160154

Žihala D, Eliáš M (2019) Evolution and unprecedented variants of the mitochondrial genetic code in a lineage of green algae. Genome Biol Evol 11:2992–3007. https://doi.org/10.1093/gbe/evz210

exclusion	restrictive	null fit	compri cell	2	5	160	0	0.037	0.010–0.071	0.037–0.037
exclusion	restrictive	null fit	compri cell	2.5	2	100	0	0.030	0.002–0.070	0.030–0.040
exclusion	restrictive	null fit	compri cell	e	0	40	0	0.024	0.000–0.088	0.024–0.024
exclusion	restrictive	null fit	compri cell	3	0	60	0	0.016	0.000–0.060	0.016–0.016
exclusion	restrictive	null fit	main	1	53	120	0	0.446	0.351–0.535	0.405–0.496
exclusion	restrictive	null fit	main	1.5	20	160	0	0.130	0.078–0.186	0.130–0.161
exclusion	restrictive	null fit	main	2	8	160	0	0.056	0.022–0.096	0.056–0.062
exclusion	restrictive	null fit	main	2.5	5	100	0	0.059	0.016–0.113	0.059–0.059
exclusion	restrictive	null fit	main	e	0	40	0	0.024	0.000–0.088	0.024–0.024
exclusion	restrictive	null fit	main	3	0	60	0	0.016	0.000–0.060	0.016–0.016
exclusion	restrictive	constrained fit	compri cell	1.5	10	100	0	0.109	0.049–0.176	0.099–0.139
exclusion	restrictive	constrained fit	compri cell	2	9	200	0	0.0498	0.021–0.084	0.045–0.0498
exclusion	restrictive	constrained fit	compri cell	2.5	3	200	0	0.020	0.003–0.043	0.020–0.025
exclusion	restrictive	constrained fit	compri cell	e	0	200	0	0.005	0.000–0.018	0.005–0.005
exclusion	restrictive	constrained fit	compri cell	3	2	200	0	0.015	0.001–0.036	0.010–0.015
exclusion	restrictive	constrained fit	main	1.5	16	100	0	0.168	0.094–0.247	0.158–0.198
exclusion	restrictive	constrained fit	main	2	12	200	0	0.065	0.031–0.102	0.055–0.095
exclusion	restrictive	constrained fit	main	2.5	4	200	0	0.025	0.005–0.050	0.025–0.035
exclusion	restrictive	constrained fit	main	e	1	200	0	0.010	0.000–0.028	0.010–0.010
exclusion	restrictive	constrained fit	main	3	2	200	0	0.015	0.001–0.036	0.015–0.015
exclusion	shared	null fit	compri cell	1	27	100	0	0.277	0.186–0.368	0.277–0.307
exclusion	shared	null fit	compri cell	2	1	60	0	0.033	0.000–0.089	0.033–0.033
exclusion	shared	constrained fit	compri cell	2	3	150	0	0.026	0.004–0.057	0.020–0.026
exclusion	shared	constrained fit	compri cell	3	0	150	0	0.007	0.000–0.024	0.007–0.007
exclusion	pooled	null fit	main	1	19	60	0	0.328	0.203–0.450	0.328–0.344
exclusion	pooled	null fit	main	1.5	35	140	0	0.255	0.181–0.330	0.255–0.270
exclusion	pooled	null fit	main	2	12	140	0	0.092	0.045–0.145	0.085–0.113
exclusion	pooled	null fit	main	2.5	4	100	0	0.0495	0.011–0.099	0.0495–0.059
exclusion	pooled	null fit	main	e	0	40	0	0.024	0.000–0.088	0.024–0.049
exclusion	pooled	null fit	main	3	1	40	0	0.049	0.001–0.132	0.049–0.049
exclusion	pooled	null fit	compri cell	1	15	60	0	0.262	0.147–0.379	0.262–0.262
exclusion	pooled	null fit	compri cell	1.5	22	140	0	0.163	0.101–0.228	0.142–0.170
exclusion	pooled	null fit	compri cell	2	5	140	0	0.043	0.012–0.081	0.035–0.043
exclusion	pooled	null fit	compri cell	2.5	3	100	0	0.040	0.006–0.085	0.040–0.040
exclusion	pooled	null fit	compri cell	e	0	40	0	0.024	0.000–0.088	0.024–0.024
exclusion	pooled	null fit	compri cell	3	0	40	0	0.024	0.000–0.088	0.024–0.049
exclusion	pooled	constrained fit	main	2	8	100	0	0.089	0.035–0.152	0.089–0.109
exclusion	pooled	constrained fit	main	2.5	8	200	0	0.045	0.017–0.077	0.040–0.0498
exclusion	pooled	constrained fit	main	e	4	200	0	0.025	0.005–0.050	0.025–0.025
exclusion	pooled	constrained fit	main	3	1	200	0	0.010	0.000–0.028	0.010–0.010
exclusion	pooled	constrained fit	compri cell	2	4	100	0	0.0495	0.011–0.099	0.0495–0.069
exclusion	pooled	constrained fit	compri cell	2.5	5	200	0	0.030	0.008–0.057	0.030–0.030
exclusion	pooled	constrained fit	compri cell	e	3	200	0	0.020	0.003–0.043	0.020–0.020
exclusion	pooled	constrained fit	compri cell	3	0	200	0	0.005	0.000–0.018	0.005–0.010
detection	restrictive	null fit	uniform	1	5	120	0	0.0496	0.014–0.095	0.017–0.0496
detection	restrictive	null fit	main	1	67	120	0	0.562	0.465–0.649	0.512–0.603
detection	restrictive	null fit	compri	1	3	120	0	0.033	0.005–0.071	0.017–0.041
detection	restrictive	null fit	compri cell	1	76	120	0	0.636	0.541–0.719	0.620–0.636
detection	shared	null fit	uniform	1	4	100	0	0.0495	0.011–0.099	0.040–0.0495
detection	shared	null fit	main	1	68	100	0	0.683	0.579–0.770	0.653–0.683
detection	shared	null fit	compri	1	5	100	0	0.059	0.016–0.113	0.020–0.079
detection	shared	null fit	compri cell	1	73	100	0	0.733	0.632–0.814	0.703–0.733
detection	pooled	null fit	uniform	1	57	60	0	0.951	0.861–0.990	0.820–0.984
detection	pooled	null fit	main	1	41	60	0	0.689	0.550–0.797	0.672–0.689
detection	pooled	null fit	compri	1	57	60	0	0.951	0.861–0.990	0.738–0.984
detection	pooled	null fit	compri cell	1	45	60	0	0.754	0.621–0.853	0.754–0.754

Table S13: Every frozen-skeleton simulation comparison with the observed statistic: the exclusion tail (share of replicates under a preference generator whose mean per-history P is at least the observed one) and the detection tail (share under the no-preference generator at most the observed one), with counts, exact ties with the observed statistic (0 in all), Clopper–Pearson intervals for k/n and the range of the reported value over the 20 sampler seeds of the observed statistic; a rank that three decimals would round to 0.050 is given to four. Accessibility: at the no-preference fit or refitted

Null cell	Test	Critical value	In sample	k	n	Fresh rejection rate
compri_state b0	`uniform`	1.9×10^{-4}	6 of 120	1	100	0.010 [0.000–0.054]
compri_state b0	`main`	0.070	6 of 120	5	100	0.050 [0.016–0.113]
compri_state b0	`compri`	0.014	6 of 120	4	100	0.040 [0.011–0.099]
compri_state b0	`compri cell`	0.077	6 of 120	6	100	0.060 [0.022–0.126]
main b0	`uniform`	1.7×10^{-12}	3 of 60	6	100	0.060 [0.022–0.126]
main b0	`main`	0.031	3 of 60	3	100	0.030 [0.006–0.085]
main b0	`compri`	9.8×10^{-10}	3 of 60	4	100	0.040 [0.011–0.099]
main b0	`compri cell`	0.026	3 of 60	3	100	0.030 [0.006–0.085]

Table S14: Independent validation of the calibrated detection rule: the critical value (empirical 5% quantile of the mean per-history P over the original no-preference replicates, with its in-sample rejections) applied to fresh no-preference replicates with disjoint seeds. The validation concerns the no-preference generators only; it does not establish repeated-sampling coverage for the exclusion ranks at positive preferences.

Test	Reported	Over 20 seeds	Median	All 200 histories
`uniform`	7.0×10^{-5}	5.8×10^{-6}–1.8×10^{-4}	3.5×10^{-5}	5.7×10^{-5}
`main`	0.68	0.64–0.69	0.67	0.67
`compri cell`	0.74	0.71–0.76	0.74	0.74

Table S15: The observed statistic (mean per-history P) of the calibrated comparisons: the reported value, its range and median over sampler seeds of the 10-history procedure, and the mean over all sampled histories.

Cell	Replicates	Retargeted	With later change	No-ops (catalogue)	Stale catalogue changes	Stale units	At-max changed	Inadmissible	Mismatches
compri b0	40	2675	416	26 (24)	1089	985 of 3044	512	24	0
compri_shared b0	100	6537	938	38 (37)	2638	2365 of 7612	1176	37	0
compri_shared or2	60	3836	570	26 (25)	1547	1381 of 4542	691	25	0
compri_shared or2 cf	150	9731	1443	82 (78)	3986	3564 of 11408	1831	78	0
compri_shared or3 cf	150	9822	1442	83 (81)	3998	3597 of 11373	1851	81	0
compri_state b0	120	7813	1147	47 (47)	3153	2836 of 9115	1479	47	0
compri_state b0 val	100	6590	1006	44 (44)	2723	2434 of 7626	1271	44	0
compri_state b0_tau1	40	2615	411	18 (17)	1102	992 of 3039	503	17	0
compri_state b1	40	2553	341	14 (13)	1010	917 of 3083	464	13	0
compri_state b1 cf	200	13093	1927	73 (69)	5286	4757 of 15221	2488	69	0
compri_state or15	160	10355	1521	62 (61)	4250	3812 of 12159	1950	61	0
compri_state or15 cf	100	6530	973	47 (46)	2683	2419 of 7594	1235	46	0
compri_state or2	160	10346	1483	71 (69)	4187	3775 of 12233	1922	69	0
compri_state or2 cf	200	13156	1951	76 (72)	5346	4812 of 15251	2479	72	0
compri_state or25	100	6408	952	38 (37)	2594	2349 of 7595	1174	37	0
compri_state or25 cf	200	13051	1890	78 (74)	5206	4678 of 15255	2450	74	0
compri_state or3	60	3804	563	26 (25)	1538	1390 of 4567	691	25	0
compri_state or3 cf	200	13150	1900	80 (79)	5281	4727 of 15295	2491	79	0
main b0	60	3613	516	48 (46)	1419	1309 of 4554	593	46	0
main b0 val	100	6019	809	90 (87)	2297	2132 of 7617	1012	87	0
main b1	40	2357	316	41 (38)	927	849 of 3028	403	38	0
main b1 cf	200	12394	1696	172 (166)	4777	4393 of 15255	2203	166	0
main or15	140	8381	1169	147 (144)	3207	2962 of 10707	1453	144	0
main or2	140	8369	1159	153 (149)	3162	2933 of 10717	1423	149	0
main or2 cf	100	6193	850	93 (89)	2339	2159 of 7659	1088	89	0
main or25	100	5842	808	113 (111)	2278	2103 of 7575	1007	111	0
main or25 cf	200	12421	1762	194 (189)	4833	4449 of 15278	2300	189	0
main or3	40	2360	312	48 (48)	885	819 of 3054	409	48	0
main or3 cf	200	12400	1760	198 (186)	4804	4431 of 15260	2279	186	0

Table S16: Retargeting audit of the calibration replicates: changes retargeted, those with a later change of the same codons below them, later changes that become no-ops, catalogue changes and units drawn on a predecessor code that the retargeting has changed, units in which a candidate's at-maximum outcome changes on that code, drawn targets inadmissible on it, and tip, row and constraint mismatches with the retargeted true history.

Generator	Fit	Coverage (95% CI)	$P(\beta > 0) > 0.975$ (95% CI)	Fits accepted
restrictive, no preference	mixture	0.967 (0.885–0.996)	0.017 (0.000–0.089)	113 of 120
	fixed state	0.950 (0.861–0.990)	0.050 (0.010–0.139)	113 of 120
pooled, no preference	pooled	1.000 (0.940–1.000)	0.000 (0.000–0.060)	115 of 120
	restrictive mixture	0.850 (0.734–0.929)	0.150 (0.071–0.266)	117 of 120
restrictive, random slopes	mixture	0.900 (0.763–0.972)	0.050 (0.006–0.169)	78 of 80
restrictive, odds e	mixture	0.975 (0.868–0.999)	0.175 (0.073–0.328)	73 of 80
pooled, odds e	pooled	1.000 (0.912–1.000)	0.100 (0.028–0.237)	75 of 80

Table S17: Calibration of the hierarchical posterior on sampled histories: coverage of the 95% interval for β, the rate of $P(\beta > 0) > 0.975$, and the number of component fits that passed the diagnostics out of those attempted (2 per replicate).

Generator	Uniform	Fixed-state `compri`	Cell null (`compri`) / pooled (`main`)
restrictive, no preference	0.050	0.033	0.64
pooled, no preference	0.95	0.95	0.69

Table S18: Lower-tail ranks of the observed statistics under the no-preference generators: the share of replicates whose mean per-history P is at most the observed one (the detection tail; small values would indicate a preference).

Generator	Test	True history	$2\bar{P}$ [95%]	Largest P	$\bar{P} \leq 0.05$
compri b0	`uniform`	0.275	0.200 [0.091–0.356]	0.200	0.350
compri b0	`main`	0.025	0.000 [0.000–0.088]	0.000	0.000
compri b0	`compri`	0.050	0.050 [0.006–0.169]	0.050	0.050
compri b0	`compri cell`	0.025	0.000 [0.000–0.088]	0.000	0.000
compri b0	`compri worst`	0.000	0.000 [0.000–0.088]	0.000	0.000
compri b0	`compri priormix`	0.050	0.025 [0.001–0.132]	0.050	0.050
compri b0	`compr cell`	0.025	0.000 [0.000–0.088]	0.000	0.000
compri_shared b0	`uniform`	0.610	0.410 [0.313–0.513]	0.430	0.520
compri_shared b0	`main`	0.020	0.000 [0.000–0.036]	0.010	0.060
compri_shared b0	`compri`	0.240	0.130 [0.071–0.212]	0.130	0.180
compri_shared b0	`compri cell`	0.030	0.010 [0.000–0.054]	0.020	0.040
compri_shared b0	`compri worst`	0.000	0.000 [0.000–0.036]	0.000	0.000
compri_shared b0	`compri priormix`	0.190	0.120 [0.064–0.200]	0.130	0.160
compri_shared b0	`compr cell`	0.040	0.000 [0.000–0.036]	0.010	0.060
compri_shared or2	`uniform`	1.000	0.883 [0.774–0.952]	0.917	0.933
compri_shared or2	`main`	0.550	0.317 [0.203–0.450]	0.333	0.433
compri_shared or2	`compri`	0.767	0.583 [0.449–0.709]	0.600	0.650
compri_shared or2	`compri cell`	0.550	0.317 [0.203–0.450]	0.333	0.417
compri_shared or2	`compri worst`	0.117	0.050 [0.010–0.139]	0.067	0.083
compri_shared or2	`compri priormix`	0.733	0.533 [0.400–0.663]	0.550	0.633
compri_shared or2	`compr cell`	0.533	0.267 [0.161–0.397]	0.367	0.433
compri_state b0	`uniform`	0.517	0.283 [0.205–0.373]	0.292	0.383
compri_state b0	`main`	0.033	0.017 [0.002–0.059]	0.017	0.017
compri_state b0	`compri`	0.108	0.100 [0.053–0.168]	0.100	0.117
compri_state b0	`compri cell`	0.033	0.017 [0.002–0.059]	0.017	0.017
compri_state b0	`compri worst`	0.000	0.000 [0.000–0.030]	0.000	0.000
compri_state b0	`compri priormix`	0.075	0.050 [0.019–0.106]	0.058	0.108
compri_state b0	`compr cell`	0.033	0.017 [0.002–0.059]	0.017	0.017
compri_state b0_tau1	`uniform`	0.475	0.450 [0.293–0.615]	0.450	0.475
compri_state b0_tau1	`main`	0.125	0.100 [0.028–0.237]	0.150	0.175
compri_state b0_tau1	`compri`	0.325	0.250 [0.127–0.412]	0.250	0.300
compri_state b0_tau1	`compri cell`	0.150	0.125 [0.042–0.268]	0.150	0.150
compri_state b0_tau1	`compri worst`	0.075	0.025 [0.001–0.132]	0.050	0.050
compri_state b0_tau1	`compri priormix`	0.300	0.200 [0.091–0.356]	0.225	0.300
compri_state b0_tau1	`compr cell`	0.125	0.150 [0.057–0.298]	0.150	0.150
compri_state b1	`uniform`	1.000	1.000 [0.912–1.000]	1.000	1.000
compri_state b1	`main`	0.800	0.500 [0.338–0.662]	0.525	0.625
compri_state b1	`compri`	0.975	0.875 [0.732–0.958]	0.875	0.875
compri_state b1	`compri cell`	0.800	0.500 [0.338–0.662]	0.575	0.650
compri_state b1	`compri worst`	0.275	0.075 [0.016–0.204]	0.100	0.150
compri_state b1	`compri priormix`	0.950	0.850 [0.702–0.943]	0.875	0.875
compri_state b1	`compr cell`	0.800	0.500 [0.338–0.662]	0.525	0.600
compri_state or15	`uniform`	0.931	0.812 [0.743–0.870]	0.806	0.869
compri_state or15	`main`	0.237	0.131 [0.083–0.194]	0.138	0.200
compri_state or15	`compri`	0.537	0.388 [0.312–0.468]	0.419	0.531
compri_state or15	`compri cell`	0.219	0.125 [0.078–0.186]	0.131	0.206
compri_state or15	`compri worst`	0.006	0.000 [0.000–0.023]	0.000	0.000
compri_state or15	`compri priormix`	0.463	0.312 [0.242–0.390]	0.362	0.456
compri_state or15	`compr cell`	0.237	0.119 [0.073–0.179]	0.138	0.212
compri_state or2	`uniform`	0.988	0.925 [0.873–0.961]	0.925	0.956
compri_state or2	`main`	0.456	0.269 [0.202–0.344]	0.306	0.356
compri_state or2	`compri`	0.819	0.644 [0.564–0.718]	0.669	0.800
compri_state or2	`compri cell`	0.469	0.275 [0.207–0.351]	0.275	0.369
compri_state or2	`compri worst`	0.113	0.069 [0.035–0.120]	0.069	0.075
compri_state or2	`compri priormix`	0.800	0.575 [0.494–0.653]	0.613	0.713
compri_state or2	`compr cell`	0.463	0.256 [0.191–0.331]	0.269	0.350
compri_state or25	`uniform`	1.000	0.990 [0.946–1.000]	0.990	1.000
compri_state or25	`main`	0.630	0.400 [0.303–0.503]	0.410	0.480
compri_state or25	`compri`	0.960	0.900 [0.824–0.951]	0.880	0.930
compri_state or25	`compri cell`	0.620	0.420 [0.322–0.523]	0.420	0.510
compri_state or25	`compri worst`	0.260	0.090 [0.042–0.164]	0.080	0.180
compri_state or25	`compri priormix`	0.960	0.800 [0.708–0.873]	0.840	0.920
compri_state or25	`compr cell`	0.590	0.400 [0.303–0.503]	0.410	0.510
compri_state or3	`uniform`	1.000	1.000 [0.940–1.000]	1.000	1.000
compri_state or3	`main`	0.733	0.550 [0.416–0.679]	0.533	0.617
compri_state or3	`compri`	0.983	0.917 [0.816–0.972]	0.900	0.917
compri_state or3	`compri cell`	0.783	0.550 [0.416–0.679]	0.567	0.633
compri_state or3	`compri worst`	0.450	0.183 [0.095–0.304]	0.183	0.267
compri_state or3	`compri priormix`	0.967	0.883 [0.774–0.952]	0.900	0.917

Truth	Scenario	Analysis	True-history P	$\bar{P}$	$2\bar{P}$	Largest P	Tilt covers	Hier. coverage	Hier. $P(\beta > 0) > 0.975$	TOST equiv.
`compr`	$\beta = 0$	`compr`	0.035	0.035	0.017	0.032	0.970	0.975	0.000	0.000
`compr`	$\beta = 0$, $\tau_c = 1$	`compr`	0.172	0.169	0.135	0.163	0.788	0.950	0.025	0.000
`compr`	$\beta = 1$	`compr`	0.794	0.793	0.701	0.785	0.962	1.000	0.375	0.000
`compri`	$\beta = 0$	`compri`	0.038	0.038	0.020	0.033	0.966	0.975	0.025	0.000
`compri`	$\beta = 0$	`main`	0.037	0.038	0.017	0.030	0.970	–	–	0.000
`compri`	$\beta = 0$, fixed compartment pattern	`compri`	0.089	0.095	0.046	0.080	0.956	1.000	0.000	0.000
`compri`	$\beta = 0$, fixed compartment pattern	`main`	0.177	0.172	0.104	0.155	0.902	–	–	0.000
`compri`	$\beta = 0$, $\tau_c = 1$	`compri`	0.201	0.199	0.153	0.186	0.751	0.975	0.025	0.000
`compri`	$\beta = 0$, $\tau_c = 1$	`main`	0.169	0.167	0.120	0.153	0.797	–	–	0.000
`compri`	$\beta = 1$	`compri`	0.875	0.874	0.797	0.861	0.968	1.000	0.375	0.000
`compri`	$\beta = 1$	`main`	0.748	0.747	0.635	0.716	0.981	–	–	0.000
`compri`	$\beta = 1$, $\tau_c = 1$	`compri`	0.683	0.684	0.626	0.671	0.728	1.000	0.325	0.000
`compri`	$\beta = 1$, $\tau_c = 1$	`main`	0.608	0.610	0.540	0.588	0.798	–	–	0.000
`compri`	$\beta = \log 1.5$	`compri`	0.254	0.256	0.168	0.234	0.967	1.000	0.050	0.000
`compri`	$\beta = \log 1.5$	`main`	0.211	0.206	0.127	0.189	0.974	–	–	0.000
`main`	$\beta = 0$	`main`	0.034	0.035	0.014	0.033	0.973	1.000	0.000	0.000
`main`	$\beta = 0$	`compri`	0.989	0.989	0.974	0.986	0.026	–	–	0.000
`main`	$\beta = 0$, fixed compartment pattern	`main`	0.069	0.072	0.034	0.068	0.960	1.000	0.000	0.000
`main`	$\beta = 0$, fixed compartment pattern	`compri`	0.981	0.982	0.964	0.979	0.034	–	–	0.000
`main`	$\beta = 0$, $\tau_c = 1$	`main`	0.089	0.088	0.063	0.083	0.871	1.000	0.000	0.000
`main`	$\beta = 0$, $\tau_c = 1$	`compri`	0.858	0.860	0.816	0.844	0.182	–	–	0.000
`main`	$\beta = 1$	`main`	0.509	0.513	0.383	0.498	0.977	1.000	0.325	0.000
`main`	$\beta = 1$	`compri`	1.000	1.000	1.000	1.000	0.044	–	–	0.000
`main`	$\beta = 1$, $\tau_c = 1$	`main`	0.444	0.443	0.343	0.439	0.880	0.975	0.250	0.000
`main`	$\beta = 1$, $\tau_c = 1$	`compri`	0.994	0.993	0.991	0.993	0.239	–	–	0.000
`main`	$\beta = \log 1.5$	`main`	0.110	0.110	0.071	0.107	0.978	1.000	0.000	0.000
`main`	$\beta = \log 1.5$	`compri`	1.000	1.000	1.000	1.000	0.025	–	–	0.000

Table S20: Restricted calibration on the six merge histories, which reruns only their merge step: rejection rates at the 5% level of the class-conditional test using the true history's P, the mean per-history P, the merged $2\bar{P}$ and the largest P; coverage of the tilt interval and of the hierarchical interval; the rate of $P(\beta > 0) > 0.975$; and the rate at which the two one-sided tests declare equivalence at odds 1.5. Rows with $\beta = 0$ and the analysis model equal to the generating one measure size; the others measure rejection of a false null.

Model	TOST P	Margin (class)	P(inside)	Margin (Bayes)	Shift	Calibration
`main`	0.91	5.19	0.35	5.40	3.3	0.000
`comp`	1.00	3.76	0.52	3.13	3.0	–
`compr`	0.96	5.46	0.45	3.29	3.9	0.000
`pooli`	1.00	5.37	0.35	5.60	3.4	–
`compi`	1.00	3.70	0.42	4.28	3.2	–
`compri`	1.00	5.57	0.42	4.19	4.1	0.000
`lineage`	1.00	3.75	0.52	3.04	3.0	–

Table S21: Equivalence at odds 1.5 over the sampled histories: merged two one-sided tests on the class-conditional tilt; the narrowest symmetric margin (odds) those tests and the posterior 90% interval would support; the posterior probability inside the margin; the expected extra optimum-attaining units at the margin (modal merge history); and the rate of declared equivalence when $\beta = 0$ in the calibration.

Model	β	Mitochondrial	Plastid	Bacterial	Nuclear	$P(\beta > 0)$
`main`	−0.45 (−1.98, 1.14)	0.40 (−1.07, 2.06)	−0.74 (−3.06, 1.31)	−0.02 (−1.80, 2.25)	**−1.75 (−3.46, −0.09)**	0.28
`compr`	−0.27 (−1.74, 1.23)	0.51 (−0.86, 2.13)	−0.85 (−3.49, 1.10)	0.32 (−1.27, 2.60)	−1.23 (−2.80, 0.26)	0.35
`compri`	−0.26 (−1.82, 1.32)	0.92 (−0.47, 2.55)	−1.03 (−3.80, 1.11)	0.38 (−1.33, 2.68)	−1.55 (−3.15, 0.05)	0.37
`compr (E3c)`	−0.32 (−1.75, 1.13)	0.26 (−1.27, 1.95)	−0.64 (−3.07, 1.19)	−0.01 (−1.72, 2.23)	−0.99 (−2.65, 0.45)	0.32
`compri (E3c)`	−0.33 (−1.80, 1.14)	0.19 (−1.40, 2.01)	−0.55 (−2.96, 1.31)	−0.15 (−1.92, 2.04)	−0.92 (−2.75, 0.59)	0.32

Table S22: The prespecified adjustment on the sampled histories: reach, co-occupancy, neighbour status and polar-requirement cost, then the at-maximum term; global β and compartment slopes (median, 95% interval), intervals excluding zero in bold. Models without a tag at a fixed decoding state.

Model	$d \leq 1$	$d = 0$	Co-occupancy	Neighbour	Cost
`main`	0.51 (−0.53, 1.53)	**3.64 (1.60, 6.23)**	−0.82 (−3.34, 1.15)	**1.53 (0.56, 2.53)**	−0.41 (−1.51, 0.73)
`compr`	**1.45 (0.57, 2.41)**	−0.31 (−1.35, 1.05)	**1.55 (0.65, 2.52)**	**1.08 (0.13, 2.00)**	−0.55 (−1.70, 0.66)
`compri`	**1.22 (0.23, 2.27)**	−0.27 (−1.33, 1.13)	**1.96 (1.05, 2.90)**	**1.58 (0.85, 2.38)**	−0.39 (−1.53, 0.81)
`compr (E3c)`	**1.22 (0.33, 2.18)**	1.38 (−0.59, 6.33)	1.01 (−1.35, 2.35)	**1.16 (0.21, 2.09)**	−0.58 (−1.78, 0.68)
`compri (E3c)`	**1.48 (0.29, 2.68)**	4.36 (−0.17, 7.69)	0.04 (−2.35, 2.35)	**1.44 (0.62, 2.27)**	−0.75 (−2.02, 0.61)

Table S23: Accessibility and alternative-predictor coefficients of the adjustment fit (median, 95% interval); intervals excluding zero in bold.

Prior	Model	β	Mitochondrial	Plastid	Bacterial	Nuclear	$P(\beta > 0)$
half-normal(0, 1) (primary)	`main`	−0.43 (−1.94, 1.11)	0.49 (−0.91, 2.07)	−0.78 (−3.06, 1.19)	0.08 (−1.71, 2.40)	**−1.84 (−3.49, −0.24)**	0.28
half-normal(0, 1) (primary)	`compr`	0.67 (−0.71, 1.94)	**1.38 (0.24, 2.89)**	0.26 (−2.27, 1.96)	1.31 (−0.04, 3.44)	0.02 (−1.31, 1.21)	0.86
half-normal(0, 1) (primary)	`compri`	0.97 (−0.61, 2.37)	**2.35 (1.02, 3.86)**	0.39 (−2.42, 2.47)	**1.79 (0.25, 4.00)**	−0.03 (−1.36, 1.30)	0.90
half-normal(0, 1) (primary)	`compr (E3c)`	−0.01 (−1.44, 1.37)	0.46 (−1.02, 2.14)	−0.21 (−2.45, 1.50)	0.23 (−1.51, 2.42)	−0.51 (−2.23, 0.76)	0.49
half-normal(0, 1) (primary)	`compri (E3c)`	−0.23 (−1.69, 1.20)	0.32 (−1.19, 2.08)	−0.39 (−2.78, 1.49)	−0.07 (−1.86, 2.09)	−0.85 (−2.82, 0.53)	0.37
τ scale 0.25	`main`	−0.59 (−1.71, 0.46)	−0.15 (−1.32, 1.08)	−0.65 (−2.03, 0.61)	−0.44 (−1.71, 0.90)	−1.18 (−2.58, 0.00)	0.14
τ scale 0.25	`compr`	0.65 (−0.11, 1.42)	**0.82 (0.04, 1.81)**	0.59 (−0.45, 1.51)	0.77 (−0.07, 1.81)	0.44 (−0.49, 1.22)	0.95
τ scale 0.25	`compri`	**1.05 (0.20, 1.90)**	**1.58 (0.58, 2.71)**	0.94 (−0.35, 2.05)	**1.23 (0.27, 2.46)**	0.54 (−0.51, 1.47)	0.99
τ scale 0.25	`compr (E3c)`	−0.10 (−1.23, 0.90)	0.01 (−1.17, 1.14)	−0.13 (−1.38, 0.97)	−0.07 (−1.28, 1.09)	−0.22 (−1.41, 0.75)	0.43
τ scale 0.25	`compri (E3c)`	−0.32 (−1.46, 0.68)	−0.21 (−1.39, 0.93)	−0.34 (−1.59, 0.77)	−0.31 (−1.54, 0.83)	−0.44 (−1.65, 0.54)	0.27
τ scale 2.5	`main`	−0.34 (−2.11, 1.54)	0.70 (−0.81, 2.64)	−0.91 (−3.74, 1.45)	0.34 (−1.70, 3.36)	**−2.01 (−3.84, −0.24)**	0.35
τ scale 2.5	`compr`	0.63 (−1.19, 2.23)	**1.63 (0.34, 3.54)**	−0.06 (−3.49, 2.12)	**1.65 (0.01, 4.56)**	−0.14 (−1.66, 1.24)	0.78
τ scale 2.5	`compri`	0.85 (−1.22, 2.62)	**2.63 (1.17, 4.72)**	−0.01 (−3.70, 2.64)	**2.11 (0.31, 5.21)**	−0.18 (−1.80, 1.35)	0.81
τ scale 2.5	`compr (E3c)`	0.01 (−1.66, 1.67)	0.69 (−0.95, 2.65)	−0.34 (−3.43, 1.74)	0.40 (−1.62, 3.27)	−0.66 (−2.62, 0.78)	0.50
τ scale 2.5	`compri (E3c)`	−0.21 (−1.94, 1.54)	0.59 (−1.11, 2.64)	−0.52 (−3.96, 1.79)	0.08 (−2.01, 2.93)	−1.11 (−3.42, 0.51)	0.40
fixed effect	`main`	−0.63 (−1.63, 0.27)	–	–	–	–	0.09
fixed effect	`compr`	0.60 (−0.05, 1.26)	–	–	–	–	0.97
fixed effect	`compri`	**1.02 (0.35, 1.68)**	–	–	–	–	1.00
fixed effect	`compr (E3c)`	−0.15 (−1.24, 0.76)	–	–	–	–	0.38
fixed effect	`compri (E3c)`	−0.37 (−1.43, 0.55)	–	–	–	–	0.23

Table S24: Heterogeneity-prior sensitivity on the sampled histories: the primary half-normal(0, 1) prior on the compartment and lineage scales, both scales set to 0.25 and to 2.5 together, and the fixed-effect fit (no random slopes).

Left out	Model	β	Mitochondrial	Plastid	Bacterial	Nuclear	$P(\beta > 0)$
without mitochondrial	`main`	−0.61 (−2.20, 1.09)	–	−0.74 (−2.79, 1.19)	−0.08 (−1.83, 2.35)	−1.24 (−2.95, 0.25)	0.23
without mitochondrial	`compr`	−0.17 (−1.53, 1.23)	–	−0.31 (−2.36, 1.31)	0.04 (−1.46, 2.11)	−0.27 (−1.50, 0.85)	0.40
without mitochondrial	`compri`	−0.07 (−1.45, 1.36)	–	−0.20 (−2.32, 1.50)	0.20 (−1.32, 2.36)	−0.21 (−1.45, 0.94)	0.46
without mitochondrial	`compr (E3c)`	−0.67 (−2.31, 0.93)	–	−0.79 (−2.96, 1.04)	−0.45 (−2.26, 1.77)	−0.94 (−2.67, 0.42)	0.19
without mitochondrial	`compri (E3c)`	−0.60 (−2.24, 1.04)	–	−0.74 (−3.06, 1.22)	−0.28 (−2.08, 2.09)	−0.95 (−2.78, 0.45)	0.22
without plastid	`main`	−0.22 (−1.85, 1.49)	0.62 (−0.80, 2.24)	–	0.25 (−1.58, 2.64)	**−1.75 (−3.43, −0.10)**	0.39
without plastid	`compr`	0.88 (−0.54, 2.27)	**1.46 (0.30, 2.95)**	–	**1.43 (0.04, 3.56)**	0.08 (−1.26, 1.34)	0.90
without plastid	`compri`	1.19 (−0.46, 2.68)	**2.39 (1.09, 3.94)**	–	**1.88 (0.33, 4.11)**	0.02 (−1.35, 1.42)	0.93
without plastid	`compr (E3c)`	0.17 (−1.35, 1.72)	0.62 (−0.93, 2.32)	–	0.40 (−1.40, 2.66)	−0.43 (−2.17, 0.91)	0.59
without plastid	`compri (E3c)`	−0.12 (−1.65, 1.46)	0.43 (−1.12, 2.23)	–	0.03 (−1.83, 2.30)	−0.83 (−2.86, 0.60)	0.44
without bacterial	`main`	−0.68 (−2.32, 1.04)	0.43 (−1.06, 2.09)	−0.95 (−3.29, 1.11)	–	**−1.97 (−3.68, −0.33)**	0.21
without bacterial	`compr`	0.41 (−1.19, 1.81)	**1.50 (0.27, 3.03)**	0.02 (−2.62, 1.86)	–	−0.11 (−1.46, 1.10)	0.72
without bacterial	`compri`	0.66 (−1.14, 2.23)	**2.50 (1.16, 4.01)**	0.10 (−2.77, 2.35)	–	−0.14 (−1.50, 1.16)	0.78
without bacterial	`compr (E3c)`	−0.15 (−1.81, 1.40)	0.52 (−1.25, 2.37)	−0.36 (−2.78, 1.52)	–	−0.63 (−2.61, 0.76)	0.42
without bacterial	`compri (E3c)`	−0.43 (−2.06, 1.11)	0.18 (−1.47, 2.07)	−0.58 (−3.18, 1.41)	–	−1.04 (−3.19, 0.41)	0.28
without nuclear	`main`	0.12 (−1.48, 1.64)	0.35 (−1.10, 1.87)	−0.15 (−2.46, 1.53)	0.21 (−1.53, 2.10)	–	0.56
without nuclear	`compr`	1.23 (−0.43, 2.60)	**1.56 (0.45, 2.96)**	0.90 (−1.89, 2.52)	**1.56 (0.22, 3.49)**	–	0.94
without nuclear	`compri`	1.74 (−0.11, 3.12)	**2.35 (1.20, 3.78)**	1.39 (−1.61, 3.11)	**2.04 (0.56, 3.89)**	–	0.97
without nuclear	`compr (E3c)`	0.85 (−0.79, 2.32)	1.13 (−0.32, 2.62)	0.62 (−1.87, 2.27)	1.04 (−0.71, 2.95)	–	0.86
without nuclear	`compri (E3c)`	0.62 (−1.04, 2.12)	0.88 (−0.60, 2.43)	0.46 (−1.94, 2.19)	0.67 (−1.16, 2.52)	–	0.78

Table S25: Leaving out one compartment at a time, specification `reach`, on the sampled histories.

Model	Histories 1–20	Histories 21–40	All	Difference (s.e.)	Stable
`main`	−0.43 (−1.94, 1.11)	−0.41 (−1.92, 1.18)	−0.42 (−1.93, 1.14)	0.03 (0.021)	yes
`compr`	0.67 (−0.71, 1.94)	0.69 (−0.68, 1.98)	0.68 (−0.70, 1.96)	0.02 (0.024)	yes
`compri`	0.97 (−0.61, 2.37)	0.98 (−0.60, 2.40)	0.97 (−0.60, 2.39)	0.01 (0.025)	yes
`compr (E3c)`	−0.01 (−1.44, 1.37)	−0.01 (−1.45, 1.41)	−0.01 (−1.45, 1.39)	0.00 (0.046)	yes
`compri (E3c)`	−0.23 (−1.69, 1.20)	−0.21 (−1.70, 1.24)	−0.22 (−1.69, 1.22)	0.02 (0.017)	yes

Table S26: Two disjoint sets of sampled histories for the hierarchical fits: β (median, 95% interval) and the difference of medians with its Monte Carlo standard error.

Design	Model	Fits	Dropped	Retried	Largest $\hat{R}$	Bulk ESS	Tail ESS	Div.
primary_catalogue	`comp`	20	0	0	1.0084	1229	1353	0
primary_catalogue	`compi`	20	0	1	1.0093	1540	1080	0
primary_catalogue	`compr`	20	0	0	1.0090	1409	1242	0
primary_catalogue	`compr mix`	20	0	1	1.0096	1368	1393	0
primary_catalogue	`compri`	20	0	0	1.0098	1537	1029	0
primary_catalogue	`compri mix`	20	0	2	1.0090	1645	1292	0
primary_catalogue	`lineage`	20	0	1	1.0089	1464	1296	0
primary_catalogue	`main`	20	0	0	1.0094	1552	905	0
primary_catalogue	`pooli`	20	0	0	1.0087	1576	1354	0
primary_catalogue (evidence)	`compri`	10	0	0	1.0078	1938	1453	0
primary_catalogue (evidence)	`compri mix`	10	0	0	1.0078	1724	1499	0
primary_catalogue (evidence)	`main`	10	0	0	1.0088	1679	1158	0
primary_all	`compr`	10	0	2	1.0091	1268	1380	0
primary_all	`compri`	10	0	1	1.0099	1689	1304	0
primary_all	`main`	10	0	0	1.0082	1704	1388	0
clade_error_catalogue	`compri`	5	0	0	1.0077	1521	1357	0
clade_error_catalogue	`main`	5	0	0	1.0087	1679	1233	0
classes_20_catalogue	`compri`	5	0	1	1.0088	1721	1354	0
classes_20_catalogue	`main`	5	0	0	1.0064	1971	1337	0
classes_min_catalogue	`compri`	5	0	0	1.0081	1822	1432	0
classes_min_catalogue	`main`	5	0	0	1.0064	1801	1436	0
computational_carrier_error_catalogue	`compri`	5	0	0	1.0065	1685	1255	0
computational_carrier_error_catalogue	`main`	5	0	0	1.0084	2085	1386	0
error_x2_catalogue	`compri`	5	0	0	1.0059	1921	1446	0
error_x2_catalogue	`main`	5	0	0	1.0091	1801	1301	0
exact_tips_catalogue	`compri`	5	0	0	1.0070	2054	1298	0
exact_tips_catalogue	`main`	5	0	0	1.0090	2088	1413	0
fixed_rates_catalogue	`compri`	5	0	0	1.0088	1957	1399	0
fixed_rates_catalogue	`main`	5	0	0	1.0077	1637	1442	0
flat_root_catalogue	`compri`	5	0	0	1.0086	1808	1398	0
flat_root_catalogue	`main`	5	0	0	1.0068	1743	1294	0
grafen_lengths_catalogue	`compri`	5	0	0	1.0080	1715	1375	0
grafen_lengths_catalogue	`main`	5	0	0	1.0061	1774	1504	0
hard_readers_catalogue	`compri`	5	0	0	1.0059	2047	1469	0
hard_readers_catalogue	`main`	5	0	0	1.0067	1641	1303	0
hydro_ala_catalogue	`compri`	5	0	0	1.0078	1817	1402	0
hydro_ala_catalogue	`main`	5	0	0	1.0071	1876	1320	0
no_literature_catalogue	`compri`	5	0	0	1.0078	1743	1274	0
no_literature_catalogue	`main`	5	0	0	1.0092	1881	1431	0
no_readers_catalogue	`compri`	5	0	0	1.0083	1984	1465	0
no_readers_catalogue	`main`	5	0	0	1.0093	1579	1345	0
ott_bases_catalogue	`compri`	5	0	0	1.0079	1949	1413	0
ott_bases_catalogue	`main`	5	0	0	1.0088	1661	1105	0
primary_ordered_catalogue	`compri`	5	0	0	1.0075	1656	1404	0
primary_ordered_catalogue	`main`	5	0	0	1.0089	1707	1234	0
s1_tips_catalogue	`compri`	5	0	0	1.0082	1924	1422	0
s1_tips_catalogue	`main`	5	0	0	1.0076	1708	1220	0
unit_lengths_catalogue	`compri`	5	0	0	1.0095	1678	1315	0
unit_lengths_catalogue	`main`	5	0	0	1.0088	1657	1183	0
unlinked_catalogue	`compri`	5	0	0	1.0079	1639	1359	0
unlinked_catalogue	`main`	5	0	0	1.0079	1968	1554	0

Table S27: Posterior diagnostics of every hierarchical fit on the sampled histories: 4 chains of 1000 draws after 1000 of warm-up; a fit failing $\hat{R} < 1.01$, bulk or tail effective sample size of 400, or with a divergence is rerun once with a smaller step, and dropped if it still fails. `mix`: the latent-state mixture likelihood.

Test	ρ^* ($2\bar{P}$)	ρ^* (largest P)	$\hat{\rho}$
`uniform`	1.85	1.45	3.55
`main`	1 (no rejection)	1 (no rejection)	0.89
`comp`	1 (no rejection)	1 (no rejection)	0.72
`compr`	1 (no rejection)	1 (no rejection)	0.94
`pooli`	1 (no rejection)	1 (no rejection)	0.83
`compi`	1 (no rejection)	1 (no rejection)	0.73
`compri`	1.41	1.06	2.97
`lineage`	1 (no rejection)	1 (no rejection)	0.72
`compr cell`	1 (no rejection)	1 (no rejection)	0.84
`compri cell`	1 (no rejection)	1 (no rejection)	0.80
`compr priormix`	1 (no rejection)	1 (no rejection)	1.03
`compri priormix`	1 (no rejection)	1 (no rejection)	2.12

Table S28: Outcome-dependent ascertainment: the smallest ratio ρ^* of recording probabilities for optimum-attaining against other changes at which each merged test stops rejecting at the 5% level (1: the test does not reject), and the ratio $\hat{\rho}$ that matches the mean observed count under the tilt (the tilt estimate).

Test	Obs.	Class E	P ($2\bar{P}$)	Largest P	β (95%)
`comp`	23.9	24.1	1.00	0.88	–
`compi`	23.9	23.9	1.00	0.83	–
`compr`	23.9	23.7	1.00	0.77	–
`compr cell`	23.9	22.5	0.34	0.41	–
`compri`	23.9	14.9	0.0037	0.057	–
`compri cell`	23.9	22.4	0.30	0.41	–
`lineage`	23.9	24.1	1.00	0.88	–
`uniform`	23.9	11.9	4.1×10^{-4}	0.015	–
`main`	23.9	22.5	0.34	0.41	–
`pooli`	23.9	22.4	0.31	0.41	–
`compri` (hierarchical)	–	–	–	–	1.16 (−0.70, 2.79)
`compri mix` (hierarchical)	–	–	–	–	0.25 (−1.44, 1.92)
`main` (hierarchical)	–	–	–	–	−0.11 (−2.01, 1.79)

Table S29: The catalogue restricted to units whose rows are all confirmed or supported (32–50 units per history): class-conditional tests (`cell`: the exact cell null) and hierarchical fits on 10 histories.

Fit	Hist.	Largest $\hat{R}$	Bulk ESS	Tail ESS	Div.	First att.	Retried
`main`	6	1.0059	1296	1269	0	0	0
`comp`	6	1.0091	1457	1449	0	0	0
`compr` (E3c, default priors)	6	1.0082	1441	1202	0	0	0
`compr` (E3c, low priors)	6	1.0075	1628	1351	0	0	0
`compr` (E3c, high priors)	6	1.0092	1668	1324	0	0	1
`compr` (fixed state)	6	1.0086	1277	1204	0	0	0
`pooli`	6	1.0088	2048	1340	0	0	0
`compi`	6	1.0082	1680	1225	0	0	0
`compri` (E3c, default priors)	6	1.0070	1599	1314	0	0	0
`compri` (E3c, low priors)	6	1.0076	1762	1410	0	0	0
`compri` (E3c, high priors)	6	1.0088	1778	1299	0	0	0
`compri` (fixed state)	6	1.0077	1503	1289	0	0	0
`lineage`	6	1.0097	1478	1258	0	0	0
`main` (+cooc)	6	1.0081	1988	1274	0	0	0
`compr` (+cooc)	6	1.0083	1642	1100	0	0	0
`compri` (+cooc)	6	1.0090	1652	1565	0	0	0
`compr cost`	6	1.0076	1279	1199	0	5	2
`compr dU_all`	6	1.0077	860	733	0	13	3
`compr deficit`	6	1.0068	1364	1365	0	0	0
`compri cost`	6	1.0096	1024	1355	0	4	3
`compri dU_all`	6	1.0062	1212	1231	0	6	5
`compri deficit`	6	1.0091	1448	1347	0	0	0
`main cost`	6	1.0069	1266	1354	0	2	1
`main dU_all`	6	1.0068	1080	1031	0	2	2
`main deficit`	6	1.0059	1540	1570	0	0	0
units common to every model							
`main` (intersection)	6	1.0082	1763	1413	0	0	0
`comp` (intersection)	6	1.0074	1720	1459	0	0	1
`compr` (intersection)	6	1.0095	1499	1414	0	1	1
`pooli` (intersection)	6	1.0076	1760	1044	0	0	0
`compi` (intersection)	6	1.0071	1524	1366	0	0	0
`compri` (intersection)	6	1.0066	1423	1233	0	0	0
`lineage` (intersection)	6	1.0077	1599	1308	0	0	1
units informative under each model							
`main` (per model)	6	1.0091	1738	1290	0	0	0
`comp` (per model)	6	1.0057	1588	1359	0	0	0
`compr` (per model)	6	1.0067	1605	1354	0	0	0
`pooli` (per model)	6	1.0077	2026	1125	0	0	0
`compi` (per model)	6	1.0082	1618	1389	0	0	0
`compri` (per model)	6	1.0092	1631	1334	0	0	0
`lineage` (per model)	6	1.0070	1792	1410	0	0	0
prior settings							
`main` (β prior s.d. 0.5)	6	1.0078	1843	1399	0	0	1
`main` (β prior s.d. 3)	6	1.0069	1435	874	0	0	0
`main` (τ scale 0.25)	6	1.0088	1824	1480	0	0	0
`main` (τ scale 2.5)	6	1.0062	1419	1601	0	0	0
`compr` (β prior s.d. 0.5)	6	1.0074	1580	1361	0	0	0
`compr` (β prior s.d. 3)	6	1.0072	1410	1474	0	0	0
`compr` (τ scale 0.25)	6	1.0085	1421	1337	0	0	0
`compr` (τ scale 2.5)	6	1.0075	1281	1322	0	1	1
`compri` (β prior s.d. 0.5)	6	1.0071	1771	1497	0	0	0
`compri` (β prior s.d. 3)	6	1.0085	1594	1408	0	0	0
`compri` (τ scale 0.25)	6	1.0078	1533	1193	0	0	0
`compri` (τ scale 2.5)	6	1.0088	1191	1592	0	0	0

Table S30: Posterior diagnostics of every hierarchical fit on the six merge histories, 306 in all (36 with the E3c priors and 270 without): one fit per history and specification, so each row summarises six fits (largest $\hat{R}$, truncated to four decimals; smallest effective sample sizes; divergences in the accepted fits and on first attempts; fits retried with a smaller step). The rows above the first subheading are the reported models, the co-occupancy specification and the graded outcomes; the rows below are the unit-set fits of Table S7 and the prior settings. Every fit was accepted.

Design	Model	Global β	Mitochondrial	Plastid	Bacterial	Nuclear	$P(\tau_c > 0.5)$
catalogue	`comp`	0.15 (−1.01, 1.37)	0.42 (−0.69, 1.87)	−0.02 (−2.07, 1.50)	0.47 (−0.79, 2.48)	−0.25 (−1.55, 0.84)	0.62
catalogue	`compi`	0.37 (−0.97, 1.70)	0.97 (−0.23, 2.49)	0.14 (−2.20, 1.94)	0.84 (−0.55, 2.96)	−0.33 (−1.77, 0.90)	0.79
catalogue	`compr`	0.67 (−0.71, 1.94)	**1.38 (0.24, 2.89)**	0.26 (−2.27, 1.96)	1.31 (−0.04, 3.44)	0.02 (−1.31, 1.21)	0.82
catalogue	`compr (E3c)`	−0.01 (−1.44, 1.37)	0.46 (−1.02, 2.14)	−0.21 (−2.45, 1.50)	0.23 (−1.51, 2.42)	−0.51 (−2.23, 0.76)	0.68
catalogue	`compri`	0.97 (−0.61, 2.37)	**2.35 (1.02, 3.86)**	0.39 (−2.42, 2.47)	**1.79 (0.25, 4.00)**	−0.03 (−1.36, 1.30)	0.95
catalogue	`compri (E3c)`	−0.23 (−1.69, 1.20)	0.32 (−1.19, 2.08)	−0.39 (−2.78, 1.49)	−0.07 (−1.86, 2.09)	−0.85 (−2.82, 0.53)	0.72
catalogue	`lineage`	0.12 (−1.04, 1.34)	0.37 (−0.74, 1.84)	−0.04 (−2.04, 1.49)	0.44 (−0.82, 2.46)	−0.26 (−1.56, 0.81)	0.61
catalogue	`main`	−0.43 (−1.94, 1.11)	0.49 (−0.91, 2.07)	−0.78 (−3.06, 1.19)	0.08 (−1.71, 2.40)	**−1.84 (−3.49, −0.24)**	0.94
catalogue	`pooli`	−0.40 (−1.99, 1.22)	0.71 (−0.70, 2.32)	−0.76 (−3.13, 1.33)	0.27 (−1.59, 2.67)	**−2.21 (−3.88, −0.56)**	0.98
all changes	`compr`	0.65 (−0.69, 1.88)	**1.28 (0.23, 2.61)**	0.32 (−2.06, 1.93)	1.20 (−0.08, 3.13)	0.05 (−1.24, 1.20)	0.79
all changes	`compri`	0.96 (−0.56, 2.34)	**2.22 (0.95, 3.63)**	0.50 (−2.19, 2.49)	**1.69 (0.23, 3.71)**	0.00 (−1.30, 1.31)	0.94
all changes	`main`	−0.50 (−1.97, 1.02)	0.32 (−0.99, 1.77)	−0.78 (−2.98, 1.18)	−0.04 (−1.80, 2.13)	**−1.85 (−3.51, −0.25)**	0.94

Table S31: Compartment slopes of the sampled-history fits (median, 95% interval); intervals excluding zero in bold. Models without a tag are at a fixed decoding state; (E3c): the latent-state mixture likelihood under the default modification-availability priors.

Fit	Global β	Mitochondrial	Plastid	Bacterial	Nuclear	$P(\tau_c > 0.5)$
`main`	−0.75 (−2.36, 0.91)	−0.06 (−1.54, 1.50)	−0.75 (−3.09, 1.66)	−0.36 (−2.34, 1.96)	**−2.35 (−4.24, −0.57)**	0.94
`comp`	−0.03 (−1.22, 1.19)	0.15 (−0.97, 1.51)	−0.09 (−1.97, 1.61)	0.20 (−1.11, 2.04)	−0.37 (−1.73, 0.74)	0.56
`compr` (E3c, default priors)	−0.12 (−1.52, 1.25)	0.21 (−1.17, 1.75)	−0.22 (−2.34, 1.59)	0.03 (−1.66, 2.00)	−0.54 (−2.25, 0.72)	0.61
`compr` (E3c, low priors)	0.66 (−0.64, 1.89)	**1.17 (0.15, 2.54)**	0.46 (−1.83, 2.27)	1.12 (−0.15, 3.08)	0.11 (−1.24, 1.27)	0.74
`compr` (E3c, high priors)	−0.55 (−2.01, 0.90)	−0.16 (−1.61, 1.42)	−0.63 (−2.89, 1.41)	−0.44 (−2.25, 1.58)	−1.15 (−3.20, 0.26)	0.68
`compr` (fixed state)	0.69 (−0.62, 1.89)	**1.18 (0.19, 2.56)**	0.48 (−1.88, 2.25)	1.16 (−0.11, 3.15)	0.12 (−1.22, 1.27)	0.74
`pooli`	−0.71 (−2.35, 0.98)	0.12 (−1.34, 1.68)	−0.71 (−3.21, 1.86)	−0.20 (−2.28, 2.23)	**−2.70 (−4.58, −0.87)**	0.98
`compi`	0.11 (−1.14, 1.43)	0.54 (−0.65, 1.95)	0.00 (−2.10, 1.87)	0.45 (−0.95, 2.50)	−0.50 (−2.03, 0.75)	0.71
`compri` (E3c, default priors)	−0.29 (−1.71, 1.10)	0.16 (−1.26, 1.79)	−0.38 (−2.71, 1.54)	−0.18 (−1.93, 1.89)	−0.84 (−2.84, 0.49)	0.68
`compri` (E3c, low priors)	0.89 (−0.64, 2.20)	**1.93 (0.78, 3.34)**	0.55 (−2.12, 2.61)	1.48 (0.00, 3.47)	0.00 (−1.38, 1.34)	0.90
`compri` (E3c, high priors)	−0.51 (−2.04, 0.95)	0.03 (−1.44, 1.66)	−0.64 (−3.13, 1.48)	−0.35 (−2.22, 1.76)	−1.30 (−3.61, 0.26)	0.75
`compri` (fixed state)	0.95 (−0.58, 2.32)	**2.09 (1.00, 3.51)**	0.57 (−2.17, 2.68)	**1.62 (0.12, 3.65)**	0.00 (−1.40, 1.36)	0.93
`lineage`	−0.05 (−1.20, 1.21)	0.11 (−1.02, 1.46)	−0.12 (−1.91, 1.57)	0.17 (−1.11, 2.07)	−0.38 (−1.75, 0.73)	0.55
`main` (+cooc)	−0.65 (−2.26, 0.95)	0.02 (−1.47, 1.52)	−0.65 (−3.06, 1.75)	−0.27 (−2.20, 2.06)	**−2.27 (−4.16, −0.44)**	0.94
`compr` (+cooc)	−0.20 (−1.64, 1.28)	0.51 (−0.76, 1.96)	−0.53 (−3.13, 1.57)	0.30 (−1.25, 2.37)	−1.20 (−2.81, 0.30)	0.87
`compri` (+cooc)	−0.28 (−1.88, 1.32)	1.04 (−0.23, 2.54)	−0.92 (−3.82, 1.60)	0.45 (−1.33, 2.65)	**−1.94 (−3.52, −0.29)**	0.98
`compr cost`	0.21 (−0.40, 0.71)	0.21 (−0.19, 0.63)	0.19 (−0.82, 0.94)	0.15 (−0.45, 0.62)	0.28 (−0.13, 0.74)	0.19
`compr dU_all`	−0.19 (−0.83, 0.48)	−0.26 (−0.82, 0.24)	−0.18 (−1.09, 0.86)	−0.18 (−0.85, 0.52)	−0.16 (−0.72, 0.35)	0.22
`compr deficit`	0.35 (−0.44, 1.24)	**0.60 (0.12, 1.27)**	0.38 (−0.85, 1.91)	0.59 (−0.09, 1.50)	−0.09 (−0.65, 0.45)	0.57
`compri cost`	0.22 (−0.36, 0.72)	0.24 (−0.12, 0.60)	0.20 (−0.77, 0.96)	0.16 (−0.41, 0.59)	0.30 (−0.08, 0.74)	0.19
`compri dU_all`	0.45 (−0.08, 1.03)	**0.39 (0.03, 0.73)**	0.45 (−0.32, 1.40)	**0.57 (0.13, 1.28)**	**0.40 (0.00, 0.77)**	0.21
`compri deficit`	0.52 (−0.41, 1.46)	**0.95 (0.44, 1.66)**	0.57 (−0.83, 2.14)	**0.73 (0.03, 1.58)**	−0.05 (−0.60, 0.55)	0.71
`main cost`	0.07 (−0.48, 0.61)	0.08 (−0.35, 0.51)	0.06 (−0.80, 0.86)	0.02 (−0.61, 0.54)	0.13 (−0.34, 0.66)	0.19
`main dU_all`	−0.44 (−1.01, 0.22)	**−0.56 (−1.09, −0.11)**	−0.41 (−1.21, 0.68)	−0.40 (−0.99, 0.32)	−0.43 (−0.88, 0.05)	0.22
`main deficit`	−0.06 (−0.98, 0.99)	0.24 (−0.34, 0.90)	0.14 (−1.07, 1.88)	0.17 (−0.64, 1.12)	**−0.84 (−1.47, −0.15)**	0.77

Table S32: Compartment slopes of every hierarchical fit on the six merge histories, including the co-occupancy specification (+cooc), the graded outcomes and the modification-prior settings (median, 95% interval); intervals excluding zero in bold.

Model	Mitochondrial	Plastid	Bacterial	Nuclear	Homogeneity P
`main`	24.0/23.4, 0.90, 1.00	1.0/1.5, 1.00, 0.99	5.9/5.4, 0.81, 1.00	4.6/5.4, 1.00, 0.79	0.58
`comp`	24.0/23.4, 0.90, 1.00	1.0/1.5, 1.00, 0.99	5.9/5.5, 1.00, 1.00	4.6/6.4, 1.00, 0.46	0.80
`compr`	24.0/23.0, 0.61, 1.00	1.0/1.5, 1.00, 0.99	5.9/5.5, 1.00, 1.00	4.6/5.8, 1.00, 0.72	0.67
`compr cell`	24.0/23.4, 0.90, 1.00	1.0/1.5, 1.00, 0.99	5.9/5.4, 0.81, 1.00	4.6/5.5, 1.00, 0.67	0.47
`pooli`	24.0/23.3, 0.80, 1.00	1.0/1.5, 1.00, 0.99	5.9/5.4, 0.85, 1.00	4.6/5.7, 1.00, 0.52	0.41
`compi`	24.0/23.3, 0.80, 1.00	1.0/1.6, 1.00, 0.94	5.9/5.4, 0.91, 1.00	4.6/6.5, 1.00, 0.43	0.42
`compri`	24.0/12.1, 4.2×10^{-6}, 1.00	1.0/1.6, 1.00, 0.94	5.9/5.4, 0.91, 1.00	4.6/5.8, 1.00, 0.72	2.0×10^{-4}
`compri cell`	24.0/23.3, 0.80, 1.00	1.0/1.5, 1.00, 0.99	5.9/5.4, 0.85, 1.00	4.6/5.8, 1.00, 0.47	0.35
`lineage`	24.0/23.4, 0.89, 1.00	1.0/1.5, 1.00, 0.99	5.9/5.5, 1.00, 1.00	4.6/6.4, 1.00, 0.46	0.79
Six merge histories (largest per-history P)					
`main`	20.6/20.4, 0.58, 0.77	1.0/1.5, 1.00, 0.50	5.0/4.5, 0.44, 0.94	4.0/4.8, 1.00, 0.39	0.43
`comp`	20.6/20.4, 0.58, 0.77	1.0/1.5, 1.00, 0.50	5.0/4.7, 0.55, 0.90	4.0/5.7, 0.93, 0.25	0.55
`compr`	20.6/20.0, 0.43, 0.84	1.0/1.5, 1.00, 0.50	5.0/4.7, 0.55, 0.90	4.0/5.1, 0.83, 0.38	0.56
`pooli`	20.6/20.2, 0.51, 0.81	1.0/1.5, 1.00, 0.50	5.0/4.5, 0.46, 0.93	4.0/5.1, 1.00, 0.27	0.28
`compi`	20.6/20.2, 0.51, 0.81	1.0/1.5, 1.00, 0.50	5.0/4.6, 0.49, 0.92	4.0/5.8, 0.93, 0.23	0.38
`compri`	20.6/9.9, 1.3×10^{-6}, 1.00	1.0/1.5, 1.00, 0.50	5.0/4.6, 0.49, 0.92	4.0/5.2, 0.84, 0.37	1.8×10^{-4}
`lineage`	20.6/20.4, 0.57, 0.77	1.0/1.5, 1.00, 0.50	5.0/4.7, 0.55, 0.90	4.0/5.7, 0.93, 0.25	0.55

Table S33: Class-conditional tests within each compartment: observed/expected optimum-attaining units, upper-tail P, lower-tail P, and the homogeneity test across compartments. Upper block: sampled histories, merged P (each per-history law exact under the stated model; the merged values are not); models without a tag at a fixed decoding state; (E3c): the earlier prior mixture over decoding states, which is not exact (Proposition 10).

History or catalogue variant	Units	P `main`	P `compr`	P `compri`	β `compr` (95%)
Primary (literature-resolved reconstruction)	66–68	1.00	1.00	0.0022	0.69 (−0.62, 1.89)
Codon-state reconstruction without the literature resolution	65–67	1.00	1.00	0.0047	0.64 (−0.65, 1.87)
Every transition from table 1 (same transitions)	66–68	0.73	0.81	4.6×10^{-4}	0.81 (−0.48, 2.00)
Ancestral codes by the path rule	66–68	1.00	1.00	0.0015	0.68 (−0.57, 1.86)
Fitch ties resolved by the Mk argmax	66–68	1.00	1.00	0.0035	0.61 (−0.78, 1.91)
Catalogue source for every moved codon	66–68	1.00	1.00	0.0022	0.68 (−0.64, 1.89)
M43 and M45 one gain (codon states)	66–68	1.00	1.00	0.0020	0.74 (−0.55, 2.00)
M43: Hydrodictyaceae read AGG as Arg	66–68	1.00	1.00	0.0022	0.70 (−0.60, 1.91)
AUA→Met gained 3 times (Bilateria) and twice (yeasts)	68–70	1.00	1.00	4.1×10^{-4}	0.71 (−0.63, 1.92)
N23 as three gains	68–70	1.00	1.00	0.0020	0.69 (−0.60, 1.91)
B06 and B08 one episode (order RFN20)	65–67	1.00	1.00	0.0015	0.80 (−0.62, 2.19)
B06 and B08 one lineage cluster	66–68	1.00	1.00	0.0022	0.69 (−0.63, 1.88)
With the ten phage rows	76–78	1.00	1.00	0.019	0.60 (−0.68, 1.82)
With phage rows, G09 and G10 removed	74–76	1.00	1.00	0.019	0.56 (−0.71, 1.80)
f^5C available before bilaterian AUA→Met (ALKBH1)	66–68	1.00	1.00	0.0030	0.68 (−0.61, 1.93)

Table S34: Alternative merge histories and catalogue variants, at a fixed decoding state: merged class-conditional P and the hierarchical β under `compr`. The catalogue variants come from the literature curation: Hydrodictyaceae reading `AGG` as Arg; two bacterial rows in one GTDB order (Shulgina and Eddy, 2021, PDF p. 12); phage rows resting on an unread supplementary table; and the yeast `CUN`→Thr tRNA derived from tRNA-His, which changes no reach class because reach is a minimum over routes.