Systematic inference of mutation rates and spectra across the tree of life via a scalable read-based framework

Supplementary Information

Supplementary Information contents:

- **Supplementary Notes 1-4**
  - **Note 1: Performance and parameter choices**
  - **Note 2: External validation and filtering criteria**
  - **Note 3: Additional analyses of rate variation with divergence time**
  - **Note 4: Coverage and positional analyses, plus extended applications**
- **Supplementary Figures S1-S14**
  - **Figures supporting benchmarking, validation, robustness checks, additional exploratory analyses, and extended applications**
- **Supplementary Table legends 1-8**
  - **Legends describing the contents of datasets, summary statistics, metadata, and analysis outputs provided in the supplementary tables**
- **Supplementary References**
  - **References cited only in the Supplementary Information**

Supplementary Note 1

CORAL performance in different settings

Genomic fragments from different species align with their outgroups in varying proportions due to genomic divergence and variation in reference genome completeness. Although some mutational features can be reliably inferred from a relatively small portion of the genome, others may be biased under low coverage. In particular, when genome coverage is low, retained pseudo-reads tend to originate primarily from conserved regions, which can lead to systematic underestimation of mutation rates. This issue affects all alignment-based genome comparison methods but warrants special attention here, as aligners differ in sensitivity.

We therefore evaluated the performance of several short-read aligners on a set of species triplets spanning diverse phylogenetic groups. Ten triplets with diverse genome sizes were selected, and CORAL was run using BWA-MEM[1], Minimap2[2], and BBMap[3], each tested across multiple fragment lengths (Methods). BBMap exhibited unstable CPU usage and high memory demands, leading to unexpected pipeline failures, and was therefore excluded from further analysis.

BWA-MEM was more sensitive than Minimap2 in pseudo-read mapping, retaining a higher proportion of pseudo-reads aligned to the outgroup across nearly all tested triplets (Figure S1a). This advantage was most pronounced in species with a low overall fraction of retained pseudo-reads. In contrast, Minimap2 was substantially faster than BWA-MEM across fragment lengths and genome sizes (Figure S1b). Prioritizing sensitivity over runtime, and given that both aligners scale approximately linearly with genome size and can align large genomes within reasonable timeframes (<72 hours on a single CPU), we selected BWA-MEM for database generation and downstream analyses. The more recent BWA-MEM2 further improves runtime without altering alignment behavior, supporting this choice [4].

As shown in Figure S1c,d, both the fraction of retained pseudo-reads and alignment time increase with fragment length. However, for BWA-MEM, while the alignment fraction rises steeply between 50 bp and 150 bp, it plateaus between 150 bp and 200 bp. On the contrary, alignment time - especially for large genomes - increases sharply over this range. Based on this trade-off, we selected 150 bp as the optimal fragment length for downstream analyses.

CORAL implementation natively supports all three aligners tested here, and users may also supply a custom aligner command, provided that standard input (FASTA/FASTQ) and output (BAM/SAM)[5] formats are maintained. This ensures flexibility for future improvements and adaptation of alignment strategies to the evolutionary context.

Across all species, the fraction of retained pseudo-reads passing quality control (Methods) varied substantially (Figure S1e). This alignment fraction was strongly influenced by divergence time from the outgroup (Figure S1f) and by phylogenetic group (Figure S1g; Figure S3b), with additional contributions likely arising from genome assembly completeness and sequencing quality. Alignment fractions were generally higher in well-characterized clades such as Mammalia, Aves, Reptilia, and Actinopteri compared with most insects, plants, and fungi. Nonetheless, alignment-related biases are likely smaller than those associated with reference-based multiple sequence alignments, which typically retain only a limited subset of highly conserved regions relative to the reference genome.

Two notable deviations from these trends were observed. First, a subset of bird species overrepresented by parrots (*Psittaciformes*) showed relatively low fractions of retained pseudo-reads, despite having closely related outgroups (Figure S3c). Parrots are known to have unusually frequent events of chromosome fusions and fissions, as well as a high proportion of transposable elements compared to other birds [6]. Second, reptiles showed the opposite pattern, maintaining high alignment fractions despite having the longest outgroup divergence times among all clades (Figure S3b). Specifically, Lacertidae (wall lizards) exhibited exceptionally high alignment (>55%), even though most diverged from their outgroups over 40 MYA, suggesting long-term genomic stability.

Importantly, alignment runtime scaled well with the target genome sizes (Figure S1h). Doubling of genome size increases runtime by roughly 2.5-fold, yielding a high Pearson correlation (r=0.78) between the total sister taxa genome sizes and alignment time. This scaling enables CORAL to align even the largest genomes within hours, supporting its broad applicability (Table S2).

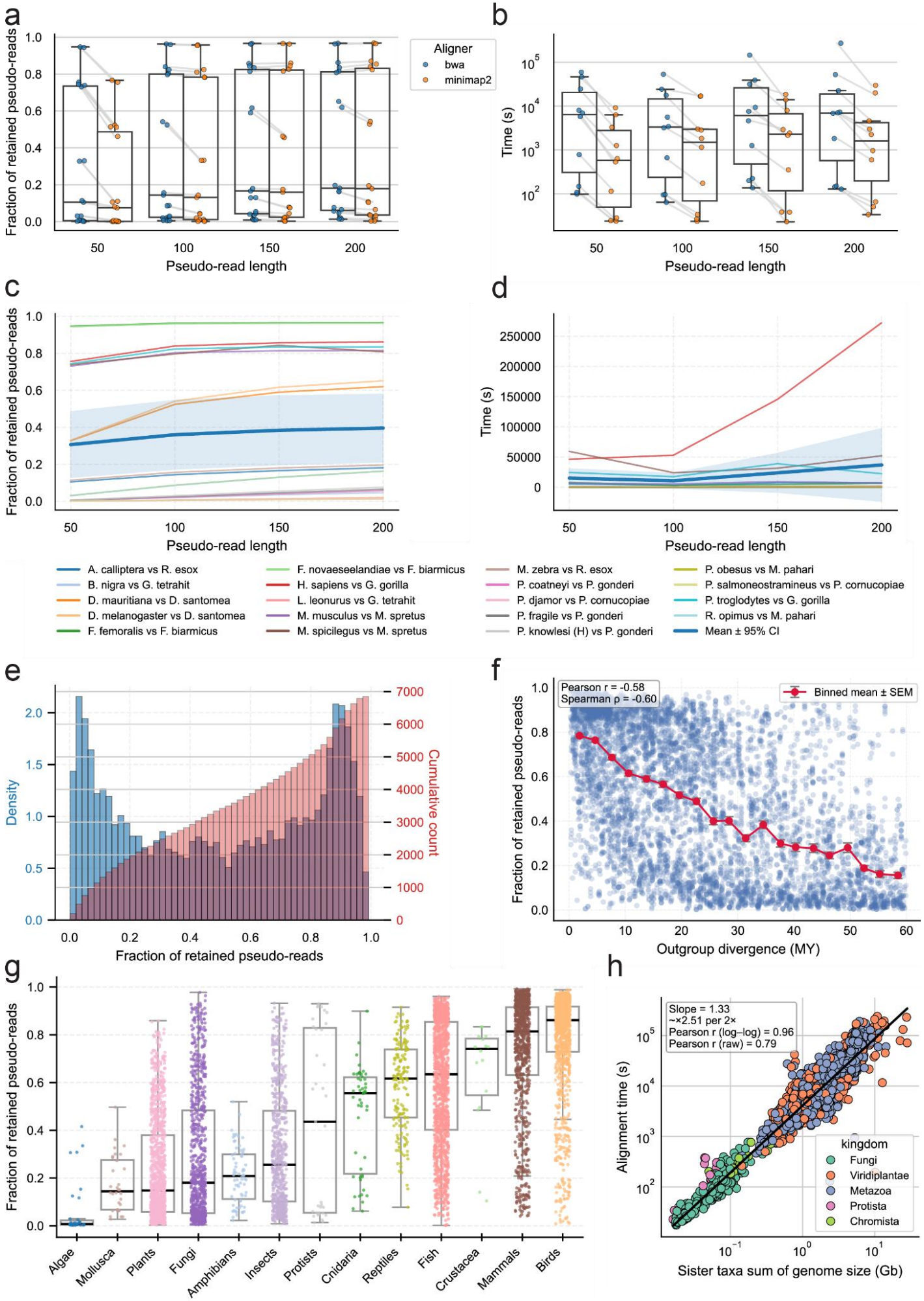

a
Fraction of retained pseudo-reads
Pseudo-read length
Aligner
bwa
minimap2
b
Time (s)
Pseudo-read length
c
Fraction of retained pseudo-reads
Pseudo-read length
d
Time (s)
Pseudo-read length
A. calliptera vs R. esox
B. nigra vs G. tetrahit
D. mauritiana vs D. santomea
D. melanogaster vs D. santomea
F. femoralis vs F. biarmicus
F. novaeseelandiae vs F. biarmicus
H. sapiens vs G. gorilla
L. leonurus vs G. tetrahit
M. musculus vs M. spretus
M. spicilegus vs M. spretus
M. zebra vs R. esox
P. coatneyi vs P. gonderi
P. djamor vs P. cornucopiae
P. fragile vs P. gonderi
P. knowlesi (H) vs P. gonderi
P. obesus vs M. pahari
P. salmoneostramineus vs P. cornucopiae
P. troglodytes vs G. gorilla
R. opimus vs M. pahari
Mean ± 95% CI
e
Density
Cumulative count
Fraction of retained pseudo-reads
f
Pearson r = -0.58
Spearman ρ = -0.60
Binned mean ± SEM
Fraction of retained pseudo-reads
Outgroup divergence (MY)
g
Fraction of retained pseudo-reads
Algae
Mollusca
Plants
Fungi
Amphibians
Insects
Protists
Cnidaria
Reptiles
Fish
Crustacea
Mammals
Birds
h
Slope = 1.33
~×2.51 per 2×
Pearson r (log–log) = 0.96
Pearson r (raw) = 0.79
kingdom
Fungi
Viridiplantae
Metazoa
Protista
Chromista
Alignment time (s)
Sister taxa sum of genome size (Gb)

Figure S1 Alignment performance and coverage across species and aligners.

a, Fraction of pseudo-reads retained after filtering for BWA-MEM (blue) and Minimap2 (orange) across pseudo-read lengths. Boxplots summarize selected species triplets; points represent species-branches, with lines connecting the same branches across aligners. **b,** Alignment runtime for BWA-MEM and Minimap2 across pseudo-read lengths, plotted as in **a**. **c,** Fraction of retained pseudo-reads versus pseudo-read length for selected species triplets using BWA-MEM. Thin lines indicate individual triplets; the thick line shows the mean ±95% confidence interval. **d,** Alignment runtime versus pseudo-read length for selected species triplets using BWA-MEM, plotted as in **c**. **e,** Distribution of the fraction of retained pseudo-reads across all analyzed species-branches, shown as a density histogram (blue) with cumulative count (red). **f,** Fraction of retained pseudo-reads versus outgroup divergence time (MY). Points represent species-branches; the red line shows the binned mean ± SEM. **g,** Fraction of retained pseudo-reads across clades. Boxplots summarize per-clade distributions with species-branch points overlaid. **h,** Alignment runtime versus the sum of sister-taxa genome sizes (Gb). Points represent species triplets, colored by kingdom; the fitted line indicates the scaling trend.

Supplementary Note 2

External validation of CORAL spectra and species-branch filtration

To further validate our spectra, we performed three comparisons: to spectra extracted from the Multiz100 alignment using the same three-way approach (Methods), to spectra from *msad213* [7], and to spectra derived from three-way Multiz alignments in fungi we generated. For each species, we computed the cosine similarity between its CORAL-derived spectrum and the corresponding external spectrum.

In the Multiz100 comparison (Figure S2a), human and chimpanzee (*P. troglodytes*) showed nearly identical spectra (cosine = 0.9995 and 0.9991, respectively). Across species, cosine similarity between matching spectra showed a significant negative correlation with evolutionary distance from humans (Spearman = -0.564, p = 0.00267) and no positive correlation with alignment coverage (Spearman = -0.316, p = 0.115). These findings highlight the value of the non-single-reference alignment strategy and indicate that differences in alignment coverage do not systematically bias the inferred spectra.

We next compared CORAL spectra with polymorphism-based mutation spectra reported in *msad213* for 13 species [7] (Figure S2b). For matching species, nearly all pairs showed cosine similarity greater than 0.98, and matching pairs were significantly more similar than non-matching pairs (mean 0.988 vs. 0.954; one-sided Wilcoxon rank-sum test, $p = 1.69\times10^{-6}$). In addition, species' spectra were consistently more similar to those of closely related taxa than to distantly related ones, motivating further exploration of the phylogenetic structure embedded in mutation spectra.

To further validate the accuracy of CORAL spectra, we compared fungal mutation spectra inferred by CORAL with spectra derived from three-way Multiz [8] alignments across 65 fungal triplets, yielding 130 fungal species-branches. Importantly, fungi exhibited low pseudo-read alignment fractions, making them a particularly challenging clade. For each species-branch, spectra were compared using cosine similarity, yielding a CORAL-by-Multiz similarity matrix (Figure S2c). Matched similarities were extremely high (diagonal mean = 0.9944; median = 0.9952) and markedly exceeded non-matched pairs (off-diagonal mean = 0.9178; median = 0.9466). This diagonal enrichment was highly significant (one-sided Mann–Whitney U test, $p = 2.47\times10^{-82}$), indicating that CORAL recapitulates branch-specific fungal spectra with near-identical composition to Multiz in alignable regions.

However, mutation rates showed a different relationship with alignment coverage. Comparison with Zoonomia [9] mutation rates showed that low alignment coverage biases mutation-rate estimates (Figure 2a), whereas mutation spectra remained stable across coverage levels (Figure S2; Figure S7c). To avoid including potentially unreliable branches, analyses that depend directly on mutation counts, such as those in comparison of traits with mutation rates, and the distribution of mutation rate across clades (Figure 2c,d), were restricted to species-branches with more than 25% aligned reads. We applied the same threshold to de novo signature extraction as a precaution, ensuring that only well-supported branches were used for calculation.

For all other analyses based on mutation spectra and their derived features, all species-branches were included. Notably, all triplets were retained when computing in-triplet features (e.g., sister-taxa mutation ratios), as mutations are called symmetrically from outgroup-aligned regions and are therefore less sensitive to absolute rate biases.

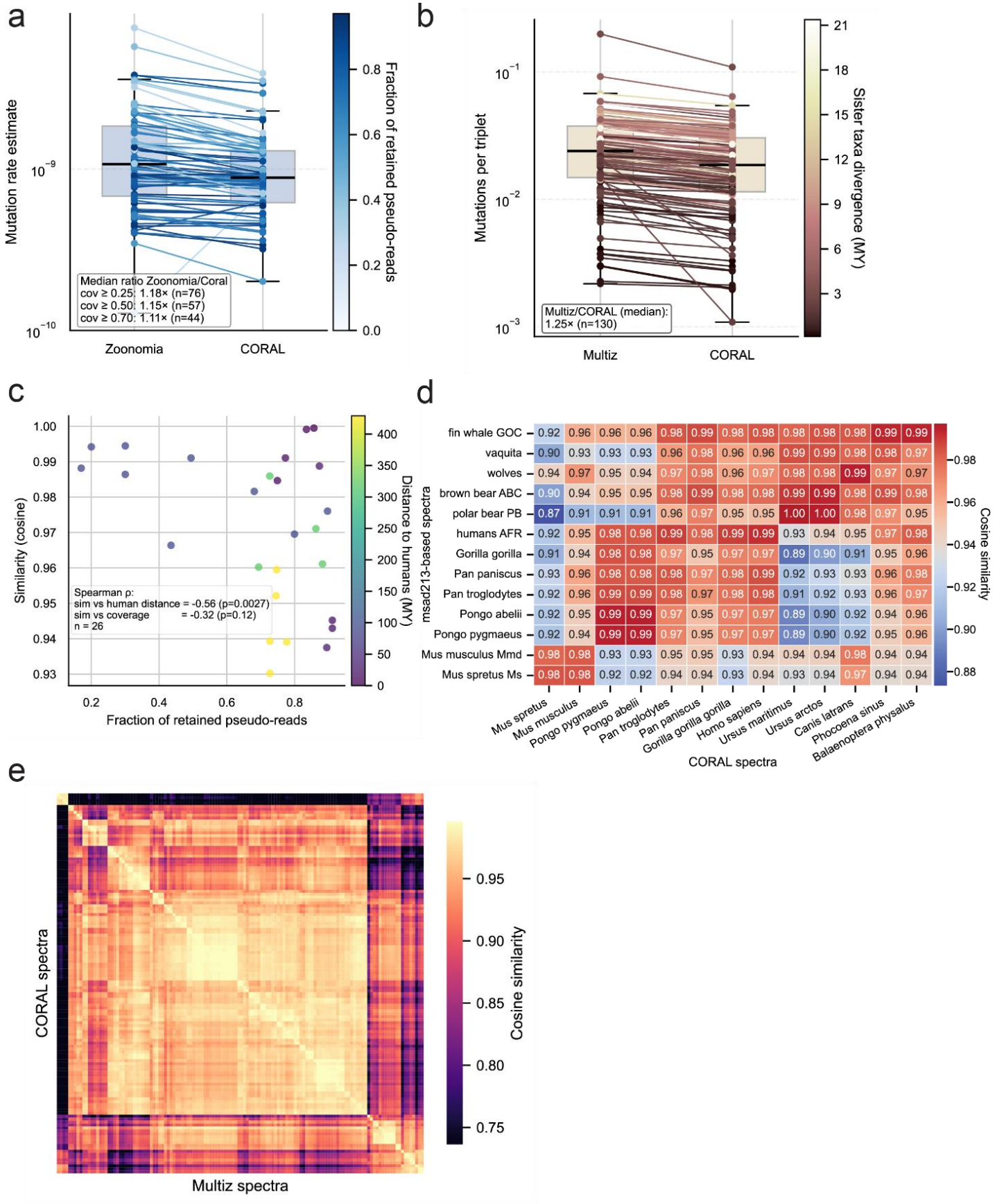

a
Mutation rate estimate
Zoonomia
CORAL
Median ratio Zoonomia/Coral
cov ≥ 0.25: 1.18× (n=76)
cov ≥ 0.50: 1.15× (n=57)
cov ≥ 0.70: 1.11× (n=44)
Fraction of retained pseudo-reads
b
Mutations per triplet
Multiz
CORAL
Multiz/CORAL (median):
1.25× (n=130)
Sister taxa divergence (MY)
c
Similarity (cosine)
Fraction of retained pseudo-reads
Spearman ρ:
sim vs human distance = -0.56 (p=0.0027)
sim vs coverage = -0.32 (p=0.12)
n = 26
Distance to humans (MY)
d
msad213-based spectra
fin whale GOC
vaquita
wolves
brown bear ABC
polar bear PB
humans AFR
Gorilla gorilla
Pan paniscus
Pan troglodytes
Pongo abelii
Pongo pygmaeus
Mus musculus Mmd
Mus spretus Ms
Mus spretus
Mus musculus
Pongo pygmaeus
Pongo abelii
Pan troglodytes
Pan paniscus
Gorilla gorilla gorilla
Homo sapiens
Ursus maritimus
Ursus arctos
Canis latrans
Phocoena sinus
Balaenoptera physalus
CORAL spectra
Cosine similarity
e
CORAL spectra
Multiz spectra
Cosine similarity

Figure S2 External validation of CORAL-derived mutation rates and spectra.

a, Paired boxplot comparing CORAL- and Zoonomia-derived mutation-rate estimates (right and left boxplots, respectively). Each dot represents a species-branch estimate, with lines connecting the same branches. Points are colored by the fraction of aligned pseudo-reads retained for that branch in CORAL. The median Zoonomia/CORAL ratio is reported for coverage thresholds of ≥0.25, ≥0.5, and ≥0.7. **b,** Paired boxplot comparing Multiz- and CORAL-derived mutations per callable site for fungal species (left and right boxplots, respectively). Each dot represents a species-branch estimate, with lines connecting the same branches. Points are colored by sister-taxon divergence time, and the annotated value indicates the median Multiz/CORAL ratio. **c,** Cosine similarity between CORAL-derived mutation spectra and spectra extracted from Multiz100 with analogous three-way comparisons, plotted against the fraction of retained pseudo-reads. Each point represents a species and is colored by evolutionary distance from humans (MY). Spearman correlation coefficients and p-values are shown for similarity versus distance from humans and similarity versus alignment coverage. **d,** Cosine similarity heatmap comparing CORAL-derived mutation spectra (columns) with msad213 spectra (rows). Cell values indicate cosine similarity. **e,** Cosine similarity heatmap comparing CORAL-derived mutation spectra (rows) with Multiz-based spectra (columns) for the chosen fungi species. Cell values indicate cosine similarity.

Supplementary Note 3

Time dependence of mutation rates with divergence time

We observed a strong and consistent negative correlation between mutation rate estimates and both sister taxa and outgroup divergence times (Figure 2d; Figure S3b; Figure S5). One possible explanation is coverage bias, arising from the preferential alignment of conserved regions. Indeed, divergence time negatively correlates with alignment coverage (Figure S1f; Figure S3b), which in turn correlates with mutation-rate estimates in some groups. However, even when comparing species with similar alignment coverage, we observed the same time-dependent decline in mutation rate, suggesting that coverage bias is not the primary cause of this pattern (Figure S5b).

Alternative explanations include: mutation saturation, potentially exacerbated by regional clustering of mutations and leading to masking of multiple hits or reduced apparent mutation probability; purifying selection, which gradually removes non-fixed variants over time; and systematic underestimation of divergence times, which would bias mutation-rate estimates upward [10,11]. Overall, our results reveal a clear hyperbolic time-dependence of mutation rates across phyla (Figure S3c). Reptiles represent a notable exception, showing little or no such trend. This anomaly may reflect their generally low mutation rates, potentially indicating saturation effects or other factors, as reptiles were also outliers in read-alignment behavior (Supplementary Note 1).

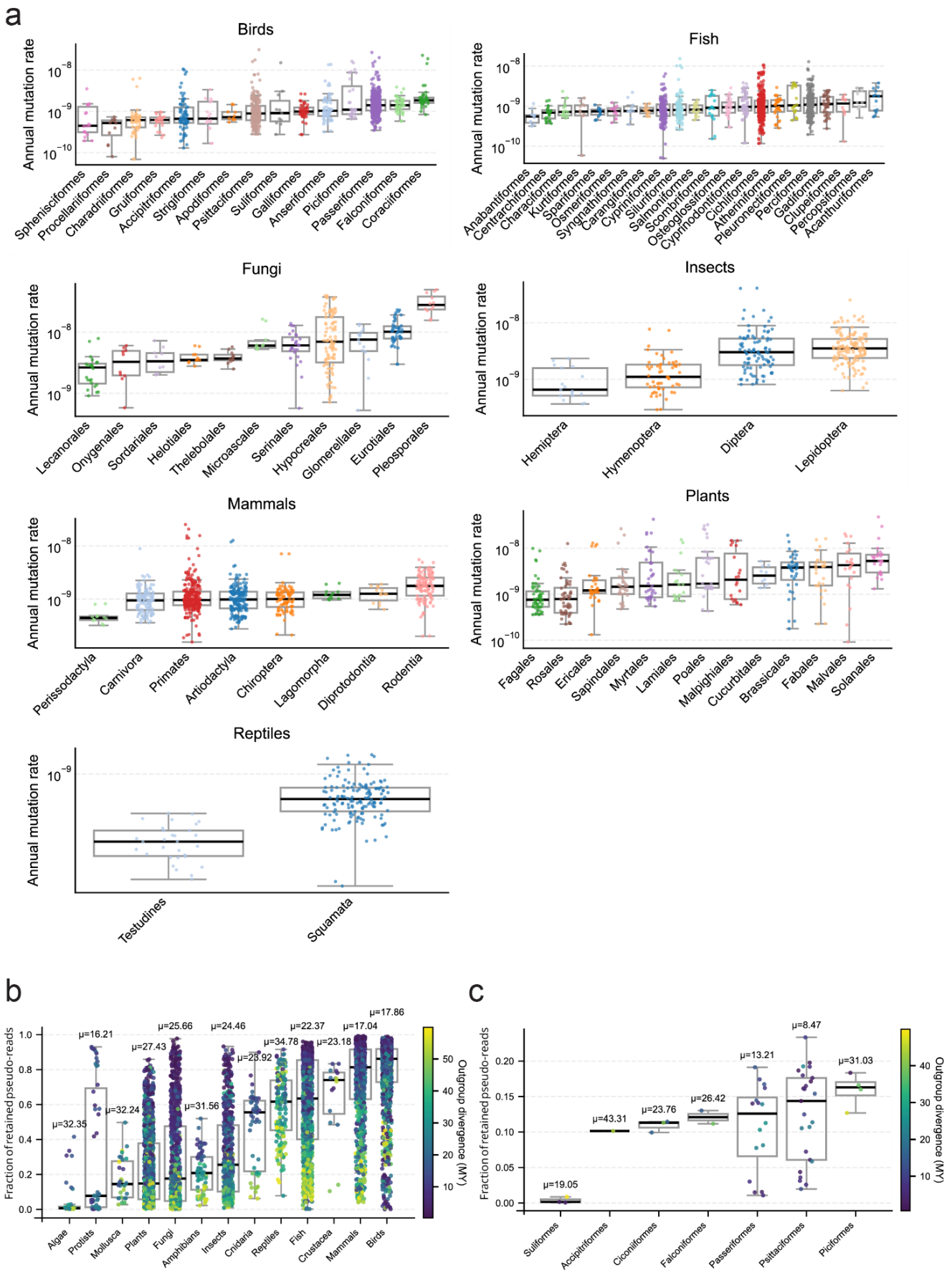
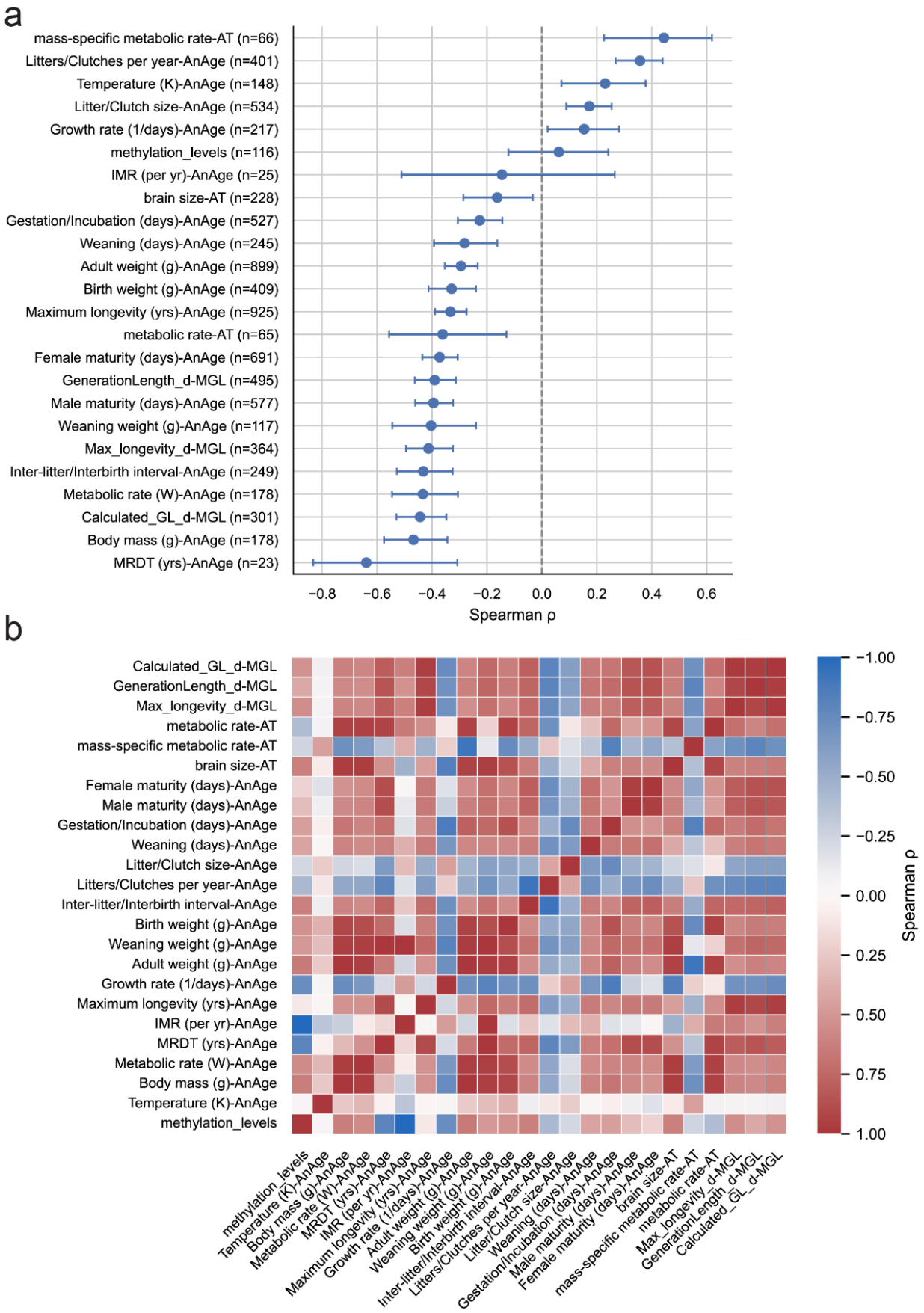


Figure S3 Variation in mutation rates and alignment across taxonomic groups. a, Annual mutation rate estimates across orders within each clade (panels). Boxplots summarize per-order distributions; points represent individual species-branches. Only orders with more than 10 species-branches are shown. **b,** Fraction of retained pseudo-reads

across clades. Boxplots summarize per-clade distributions; points represent species-branches and are colored by outgroup divergence time (MY). Mean outgroup divergence time (μ) is shown above each box. **c,** Fraction of retained pseudo-reads across bird orders with <0.25 fraction of kept pseudo-reads, plotted as in panel **b**.



161

162 **Figure S4 Association of mutation rates with life-history and ecological traits. a,** Spearman correlation coefficients

163 between mean annual mutation rate per species and life-history traits from AnAge, AnimalTraits (AT), and Pacifici *et*

164 *al.* Points indicate correlation estimates with 95% confidence intervals; the dashed line denotes zero correlation. P-

values are shown in Table S4. **b,** Pairwise Spearman correlation matrix among life-history traits used in the analysis. Cell color indicates correlation strength and direction.

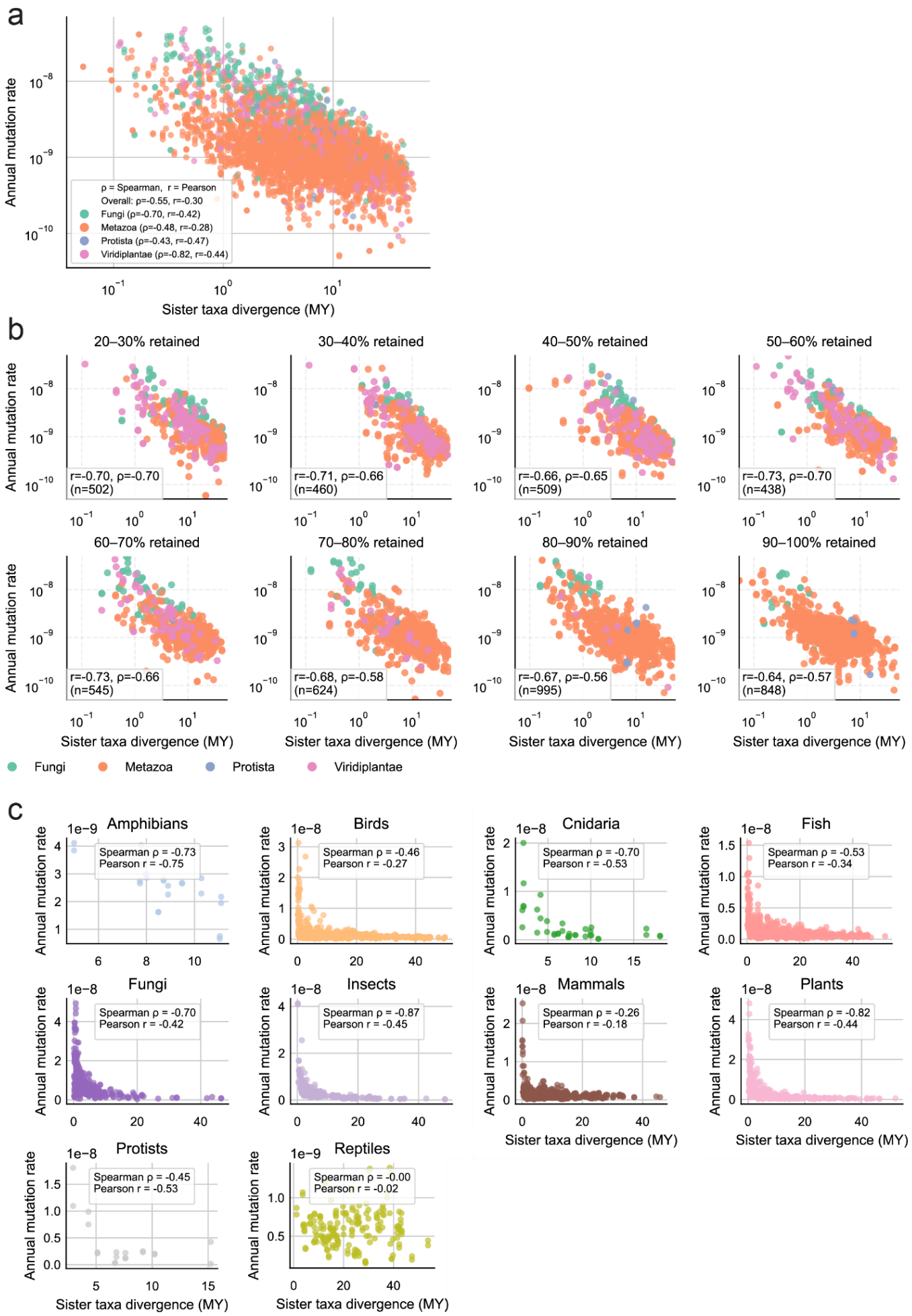

a
Annual mutation rate
Sister taxa divergence (MY)
ρ = Spearman, r = Pearson
Overall: ρ=-0.55, r=-0.30
Fungi (ρ=-0.70, r=-0.42)
Metazoa (ρ=-0.48, r=-0.28)
Protista (ρ=-0.43, r=-0.47)
Viridiplantae (ρ=-0.82, r=-0.44)
b
20–30% retained
r=-0.70, ρ=-0.70 (n=502)
30–40% retained
r=-0.71, ρ=-0.66 (n=460)
40–50% retained
r=-0.66, ρ=-0.65 (n=509)
50–60% retained
r=-0.73, ρ=-0.70 (n=438)
60–70% retained
r=-0.73, ρ=-0.66 (n=545)
70–80% retained
r=-0.68, ρ=-0.58 (n=624)
80–90% retained
r=-0.67, ρ=-0.56 (n=995)
90–100% retained
r=-0.64, ρ=-0.57 (n=848)
Fungi
Metazoa
Protista
Viridiplantae
c
Amphibians
Spearman ρ = -0.73
Pearson r = -0.75
Birds
Spearman ρ = -0.46
Pearson r = -0.27
Cnidaria
Spearman ρ = -0.70
Pearson r = -0.53
Fish
Spearman ρ = -0.53
Pearson r = -0.34
Fungi
Spearman ρ = -0.70
Pearson r = -0.42
Insects
Spearman ρ = -0.87
Pearson r = -0.45
Mammals
Spearman ρ = -0.26
Pearson r = -0.18
Plants
Spearman ρ = -0.82
Pearson r = -0.44
Protists
Spearman ρ = -0.45
Pearson r = -0.53
Reptiles
Spearman ρ = -0.00
Pearson r = -0.02

Figure S5 Dependence of mutation rate estimates on divergence time. a, Annual mutation rate estimates versus sister-taxa divergence time (MY) across species. Points are colored by kingdom; Spearman and Pearson correlation coefficients are shown overall and per kingdom. b, Annual mutation rate versus sister-taxa divergence time stratified by alignment coverage. Panels show species-branches as points; Spearman correlation coefficients are reported within each coverage bin. **c,** Annual mutation rate versus sister-taxa divergence time shown separately by clade. Each panel corresponds to one clade; points represent species, with Spearman and Pearson correlation coefficients reported.

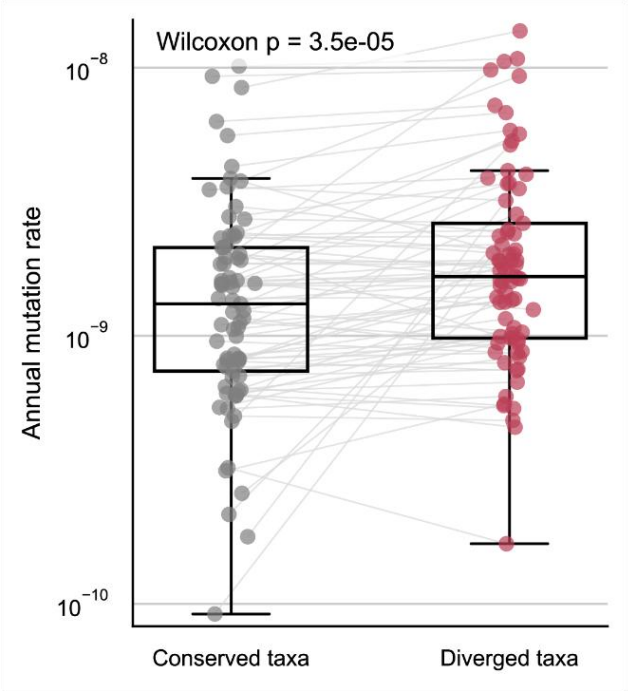
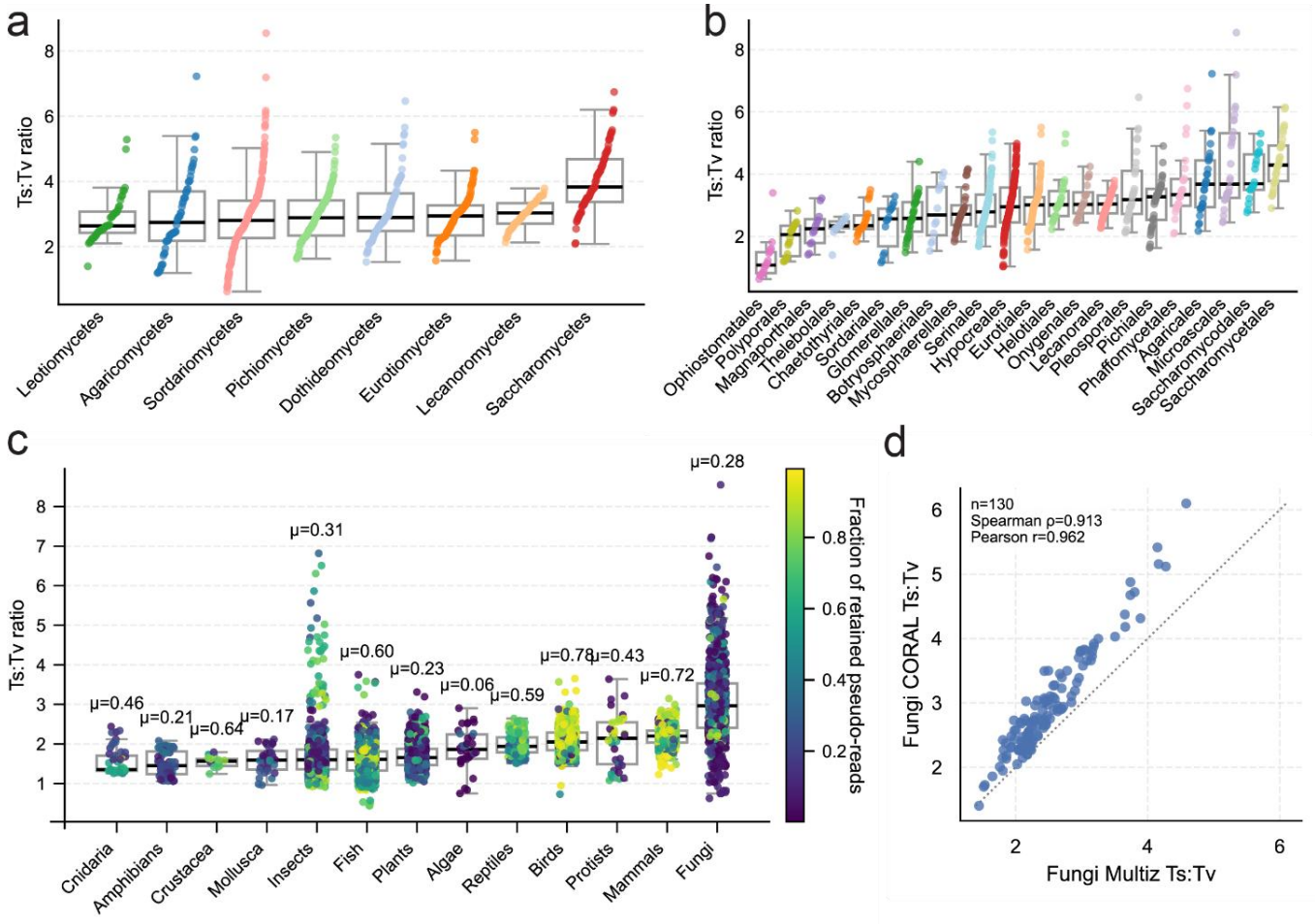


Figure S6 Genus divergence analysis. Annual mutation rates for same-genus and genus-diverged sister taxa within triplets. Sister-taxon estimates are connected by lines; boxplots summarize distributions. The Wilcoxon signed-rank test p-value is shown.



180

181 **Figure S7 Transition:transversion ratios across taxa. a,** Transition–transversion (Ts:Tv) ratios across fungal classes.
182 Boxplots summarize class-level distributions; points represent individual species-branches. **b,** Ts:Tv ratios across
183 fungal orders, plotted as in **a**. **c,** Ts:Tv ratios across clades. Boxplots summarize per-clade distributions; points
184 represent species-branches and are colored by the fraction of retained pseudo-reads. The mean fraction of retained
185 pseudo-reads (μ) is indicated above each clade. **d,** Comparison of Ts:Tv ratios between CORAL (y-axis) and Multiz (x-
186 axis) for selected fungal species-branches. Spearman and Pearson correlation coefficients are shown.

187

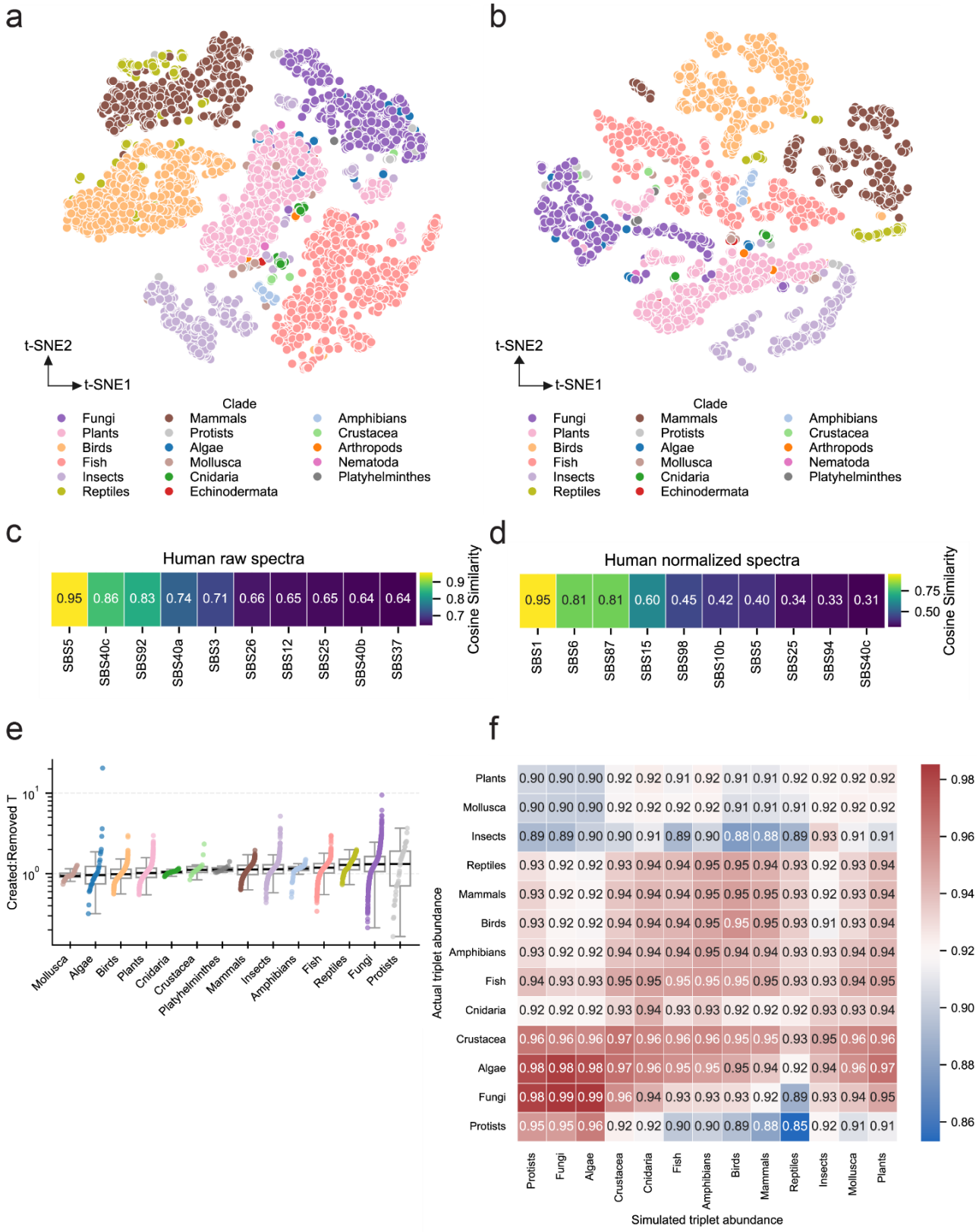


Figure S8 Normalization of mutation spectra and relationships to trinucleotide composition. a, t-SNE embedding of 96-category pre-normalized mutation spectra across species-branches, with points colored by clade. **b,** t-SNE embedding of 32-category 3-mer abundance across species, plotted as in **a**. **c,** Cosine similarity between the raw human mutation spectrum and top-most similar COSMIC SBS signatures; values indicate cosine similarity. **d,** Cosine similarity between the normalized human mutation spectrum and top-most similar COSMIC SBS signatures, plotted

as in **c**. **e,** Ratios of created to removed AT bases across clades (Methods). Boxplots summarize per-clade distributions; points represent species-branches. **f,** Cosine similarity matrix comparing observed trinucleotide abundances (rows) with steady-state compositions simulated from clade-mean normalized mutation spectra (columns). Values indicate cosine similarity.

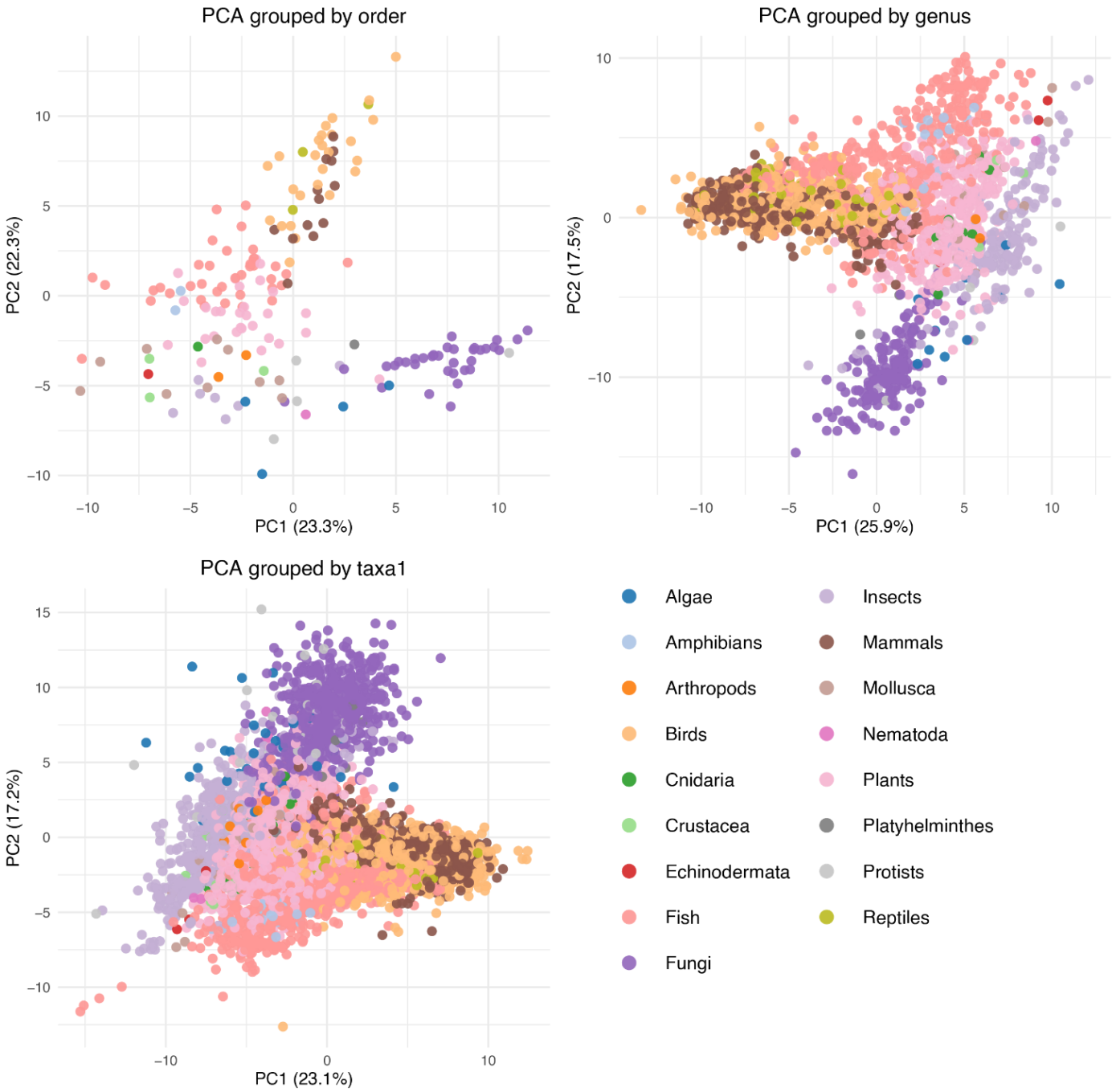


Figure S9 Principal component analysis of normalized mutation spectra. PCA plots showing the first two principal components of mean normalized mutation spectra at three taxonomic resolutions: orders (top left), genera (top

right), and species (bottom left). Points represent groups of species-branches at the corresponding level and are colored by clade. Axes indicate the variance explained by each principal component.

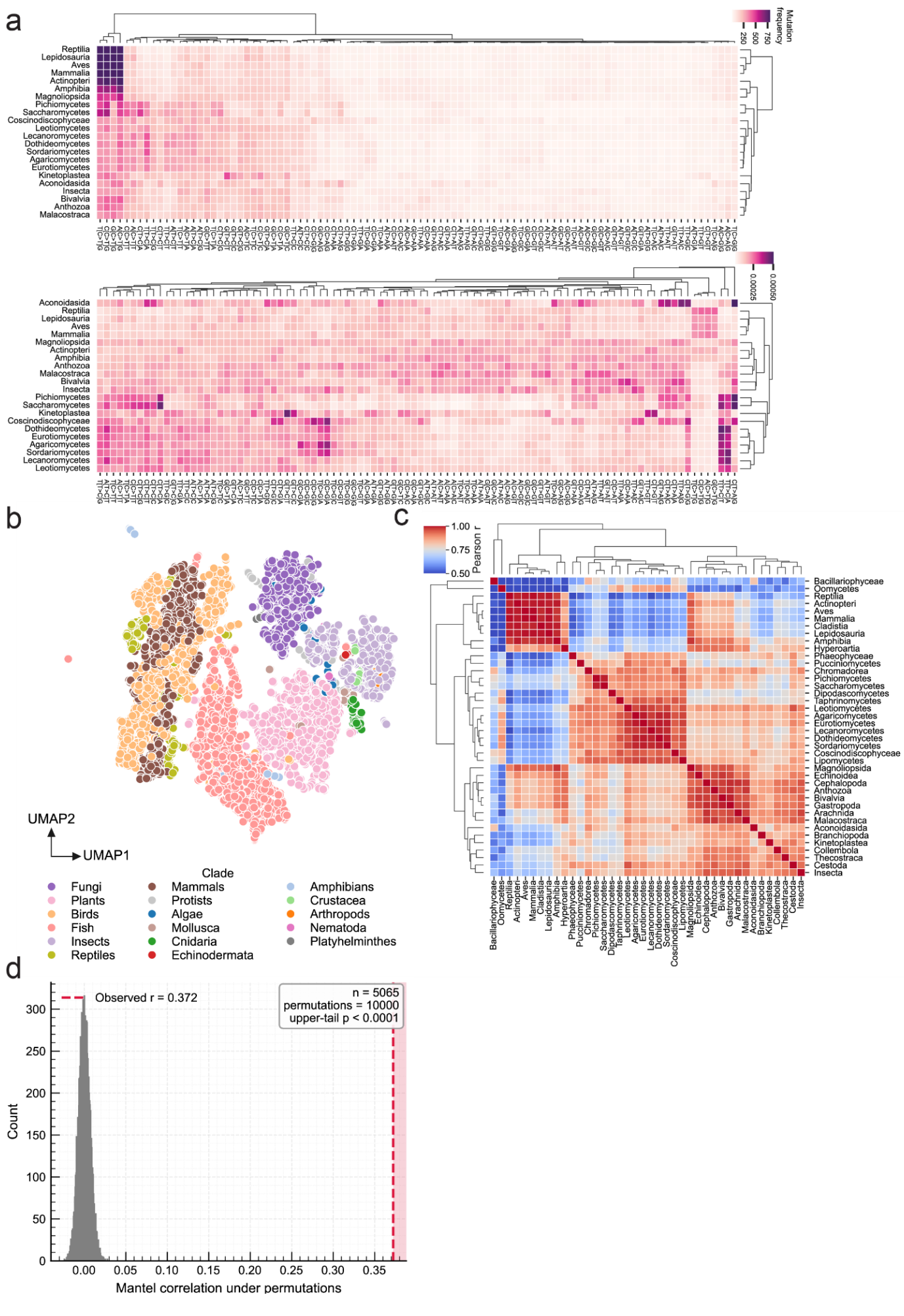
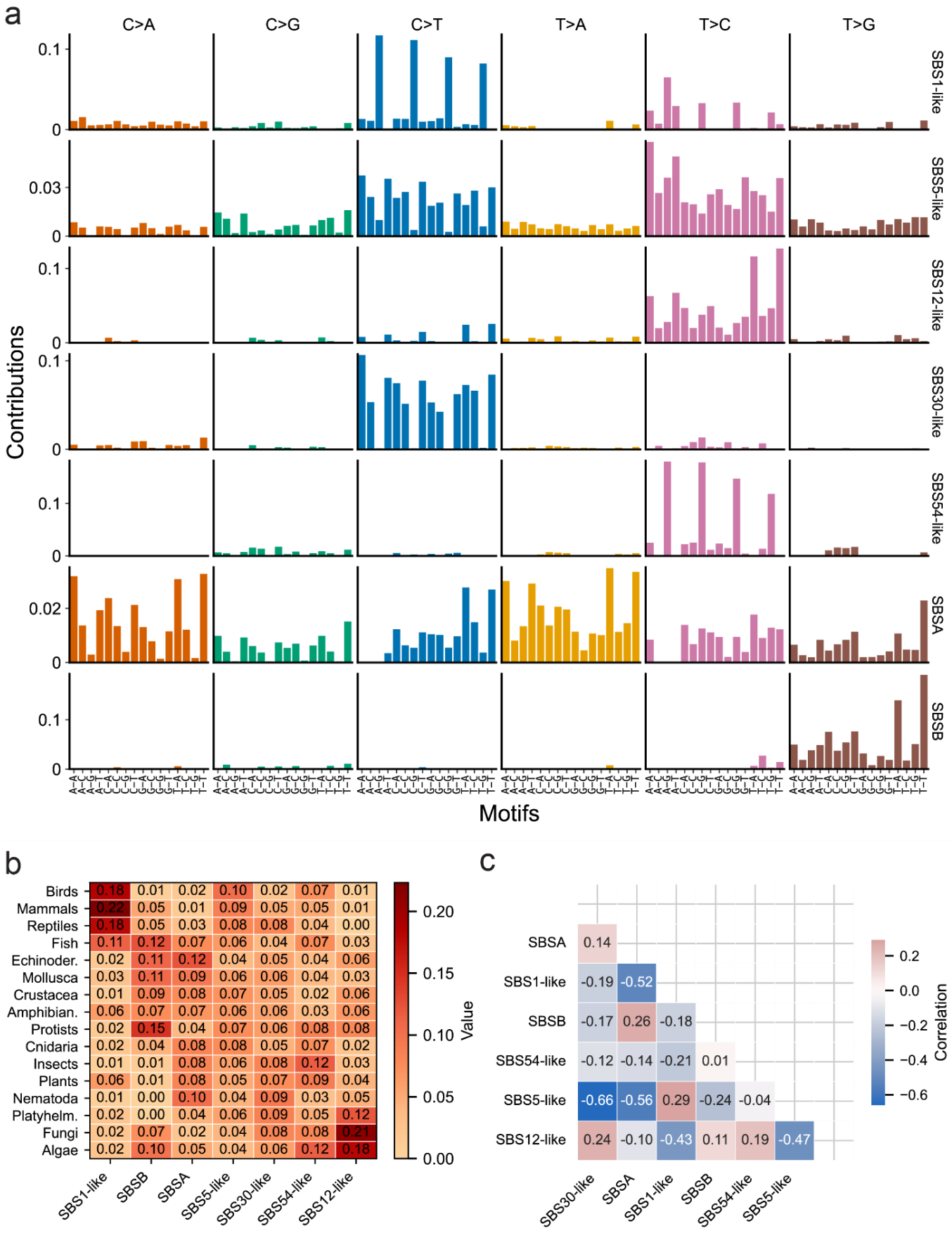

a
Mutation frequency
250
500
750
Reptilia
Lepidosauria
Aves
Mammalia
Actinopteri
Amphibia
Magnoliopsida
Pichiomycetes
Saccharomycetes
Coscinodiscophyceae
Leotiomycetes
Lecanoromycetes
Dothideomycetes
Sordariomycetes
Agaricomycetes
Eurotiomycetes
Kinetoplastea
Aconoidasida
Insecta
Bivalvia
Anthozoa
Malacostraca
0.00050
0.00025
b
UMAP2
UMAP1
Clade
Fungi
Plants
Birds
Fish
Insects
Reptiles
Mammals
Protists
Algae
Mollusca
Cnidaria
Echinodermata
Amphibians
Crustacea
Arthropods
Nematoda
Platyhelminthes
c
Pearson r
1.00
0.75
0.50
Bacillariophyceae
Oomycetes
Reptilia
Actinopteri
Aves
Mammalia
Cladistia
Lepidosauria
Amphibia
Hyperoartia
Phaeophyceae
Pucciniomycetes
Chromadorea
Pichiomycetes
Saccharomycetes
Dipodascomycetes
Taphrinomycetes
Leotiomycetes
Agaricomycetes
Eurotiomycetes
Lecanoromycetes
Dothideomycetes
Sordariomycetes
Coscinodiscophyceae
Lipomycetes
Magnoliopsida
Echinoidea
Cephalopoda
Anthozoa
Bivalvia
Gastropoda
Arachnida
Malacostraca
Aconoidasida
Branchiopoda
Kinetoplastea
Collembola
Thecostraca
Cestoda
Insecta
d
Observed r = 0.372
n = 5065
permutations = 10000
upper-tail p < 0.0001
Count
0
50
100
150
200
250
300
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Mantel correlation under permutations

Figure S10 Phylogenetic structure of normalized mutation spectra. a, Hierarchical clustering of clade-level mutation spectra based on pairwise spectral similarity. Rows correspond to clades and columns to mutation categories. The top heatmap shows mean normalized mutation spectra per clade. The bottom heatmap shows the same spectra further normalized across clades so that each mutation category sums to 1, highlighting relative differences among clades. **b,** UMAP embedding of normalized mutation spectra across species-branches, with points colored by clade. **c,** Pairwise similarity cluster map between mean normalized mutation spectra of classes. Color indicates correlation strength. **d,** Distribution of Mantel test statistics comparing spectral similarity and phylogenetic distance. The observed statistic is indicated relative to the permutation null distribution (n permutations shown).



213

214 **Figure S11 De novo evolutionary mutational signatures. a,** 96-category mutation spectra of the seven de novo

215 evolutionary signatures identified by unsupervised non-negative matrix factorization (NMF; SignatureAnalyzer).

216 Columns correspond to substitution classes and trinucleotide contexts; rows correspond to signatures. **b,** Relative

mean activity of each de novo signature across clades. For each signature, values indicate the fraction of total activity in each clade. **c,** Pairwise correlation matrix of de novo signature activities across species-branches. Cell values indicate correlation coefficients.

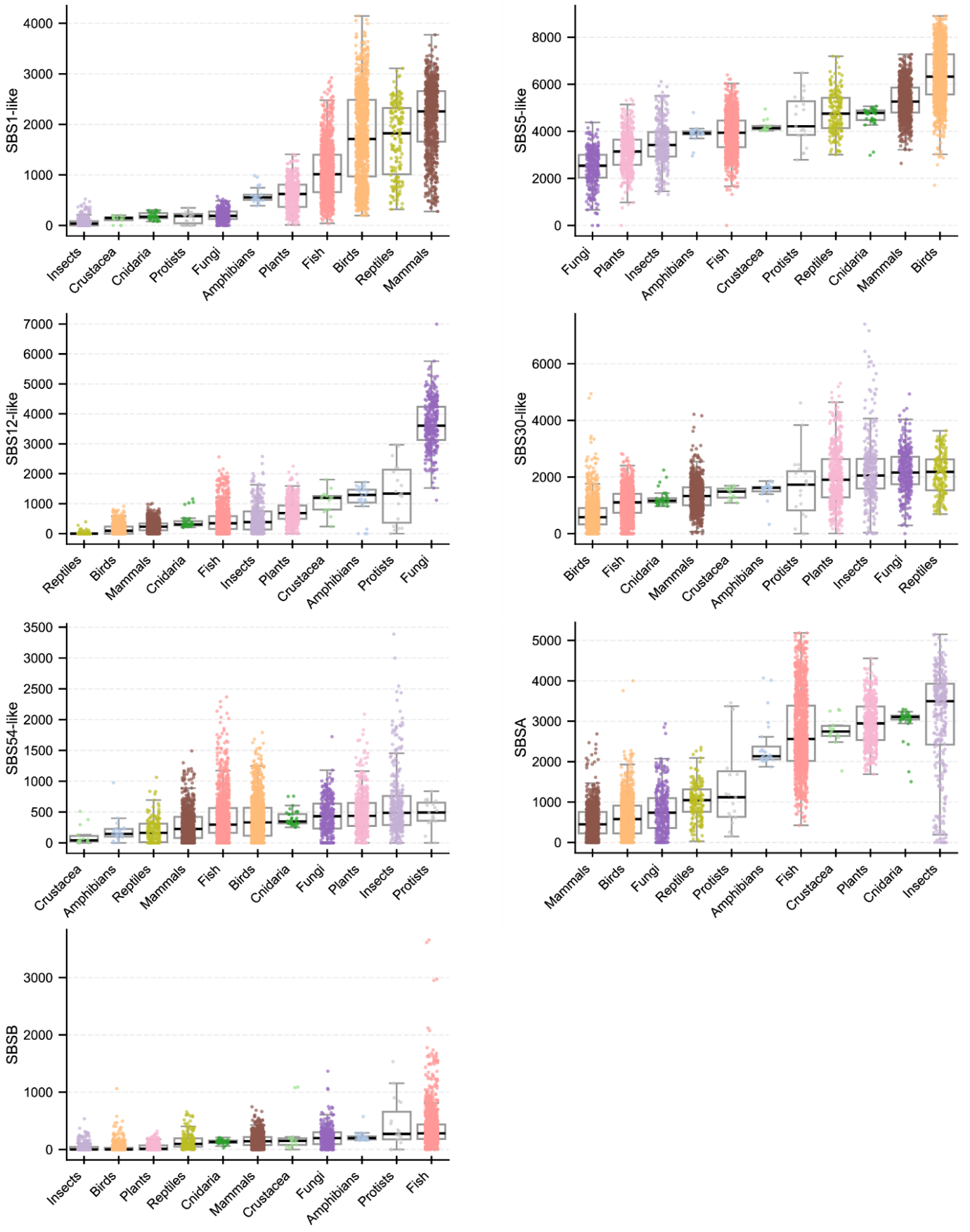
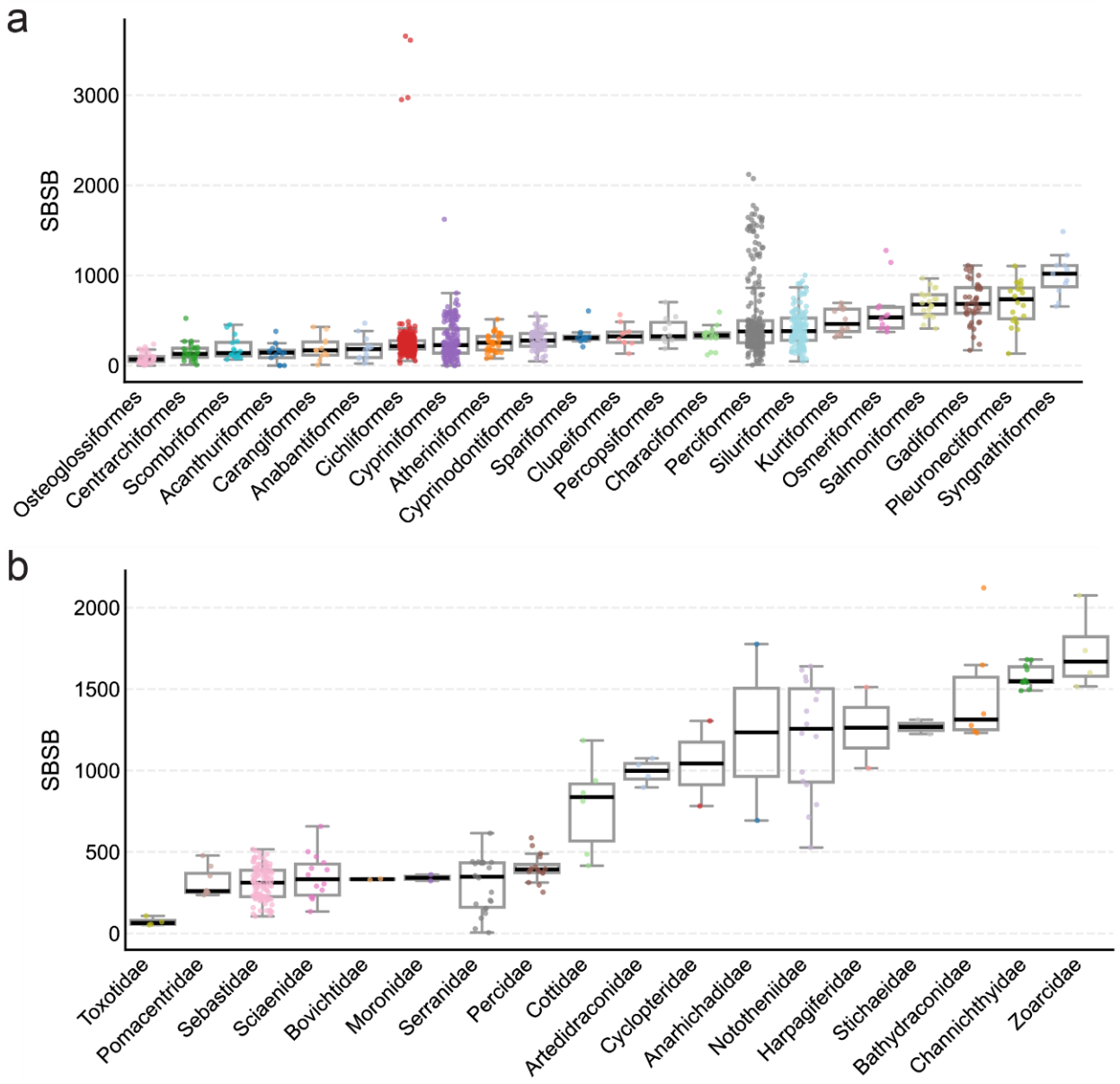


Figure S12 De novo mutational signature activity. Distribution of de novo signature activities across clades. Each panel corresponds to one signature; points represent species-branches and boxplots summarize per-clade distributions.



225

226 **Figure S13 SBSB signature activity. a,** Distribution of SBSB activity across fish orders. Points represent species-

227 branches, and boxplots summarize per-order distributions. **b,** Similar to **a.** across Perciformes fish families.

Supplementary Note 4

Genomic position and coverage analysis

Beyond genome-wide spectra, CORAL extracts positional information on read coverage and mutation density along the outgroup genome, enabling detection of local alterations in alignment depth and mutation distribution. Coverage is derived from the raw BAM files for each sister taxon and summarized along the outgroup genome using a sliding-window approach (100,000 bp), which facilitates direct comparison of coverage profiles between aligned species and highlights large-scale copy number and deletion events.

We demonstrate this potential using a triplet of *Amazona* species. Aligning *A. vittata* and *A. collaria* to *A. aestiva* (outgroup) revealed several notable patterns (Figure S14a). First, read coverage across the genome was close to the expected 2×, and local fluctuations in coverage were largely shared between sister taxa, reflecting either shared evolutionary changes or missing segments in the reference genome. Nonetheless, certain regions showed a doubling to nearly 4× in *A. collaria*, particularly on chromosome 1 (80–83 Mb and 120–122 Mb) and chromosome 6 (0–5 Mb and 24–36 Mb). The consistency of this doubling strongly suggests true biological duplications. This illustrates how read-based approaches similar to those used in cancer genomics can be applied in a comparative evolutionary context in order to identify large copy-number variants.

Examining the *Amazona* Z chromosome revealed a markedly different pattern. Unlike the autosomes, where both sister taxa exhibited nearly identical coverage, coverage along chromosome Z diverged substantially between species. This suggests more rapid sequence evolution and lineage-specific changes, consistent with accelerated sex-chromosome evolution driven by sex-biased inheritance [12], and the atypical sex-chromosome evolution reported in parrots [6].

Interpretation of such patterns warrants caution, as sequencing biases and differences in repeat handling may affect alignment. Nevertheless, the clear regional coverage changes observed here underscore the potential of this approach to detect large-scale genomic rearrangements across the tree of life.

Multi-species branch-wise phylogenetic detection

Although this work focused on pairwise sister-taxon comparisons, the same framework can be extended to multi-species alignments against a common reference. Aligning more than two species enables direct comparison of

orthologous sites across taxa, broadening inference from mutations along terminal branches to ancestral branches within the phylogenetic tree, and supporting fine-scale phylogenetic reconstruction of recent divergence history.

As a proof of concept, we examined several cat species (*Otocolobus manul*, *Felis catus*, *Felis chaus*, *Felis nigripes*, *Prionailurus bengalensis*, *P. iriomotensis*, and *P. viverrinus*). Their evolutionary relationships remain partly unresolved in *TimeTree5* (TTol5) [13] due to limited sequence data (Figure S14b). Aligning all species' pseudo-reads to *O. manul* as the reference, we extracted variable sites differing in at least one species and used these sites for phylogenetic reconstruction with PHYLIP [14]. This reconstruction resolved two ambiguous nodes: within Felis, F. nigripes diverged first, while F. chaus and F. catus formed a sister pair (Figure S14c). Within Prionailurus, P. bengalensis and P. iriomotensis appeared as sister taxa. Notably, a previous study proposed *F. nigripes* and *F. catus* as sister taxa, in contrast to our finding [15]. Although preliminary and requiring more stringent site-level filtering, this example illustrates how CORAL outputs can inform phylogenetic reconstruction, particularly among closely related species.

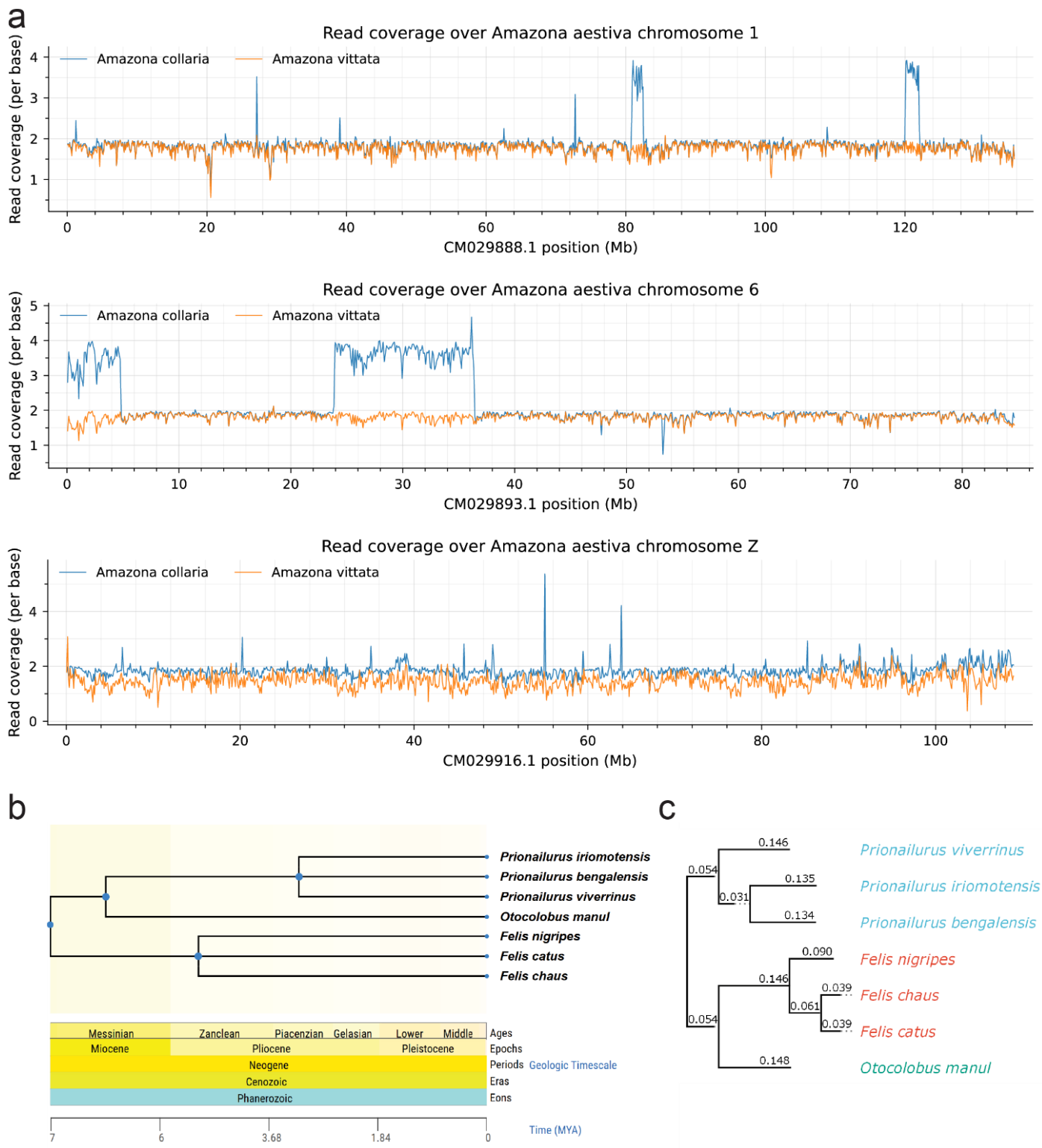


Figure S14 CORAL extended applications. a, Read coverage profiles for *Amazona collaria* and *A. vittata*, aligned to the A. aestiva reference genome across reference chromosomes 1 (top), 6 (middle), and Z (bottom). Smoothed read depth is shown along genomic position (100,000bp sliding window). **b,** Reference phylogenetic relationships among selected Felidae species based on TimeTree v5. Branch lengths reflect divergence-time estimates. **c,** Parsimony-based phylogenetic reconstruction inferred from CORAL multi-species alignment. Pseudo-reads from all species were

aligned to *Otocolobus manul* as a common reference, and variable sites differing in at least one species were extracted and used for tree reconstruction with PHYLIP. Branch lengths reflect relative substitution number.

Supplementary table legends

Supplementary Table 1 Lean CORAL database

Per-branch mutation data derived from species triplets. Each row represents a focal lineage from a sister-outgroup triplet and reports raw and normalized 96-category mutation counts, callable trinucleotide context abundances, taxonomy and divergence times, alignment and mutation-rate metrics, collapsed substitution summaries, de novo mutational signature activities, and external benchmarks and life-history traits.

Supplementary Table 2 CORAL alignment statistics

Pseudo-read alignment and filtering statistics for all species-branches (Sheet 1), including retained read fractions and filtering causes, and alignment runtime and genome size summaries for each species triplet (Sheet 2).

Supplementary Table 3 Taxonomic coverage summary

Counts of species-branches and unique species represented in the CORAL dataset, summarized by kingdom, phylum, and class, reported both before and after applying a ≥25% alignment coverage threshold.

Supplementary Table 4 Trait correlations with mutation rate

Spearman correlations between annual mutation rate and life-history traits computed on species-level means, reported for all species combined and separately for each clade.

Supplementary Table 5 Partial rank correlation (PRCC) analysis

Partial rank correlation coefficients between mutation rate and life-history traits, controlling for correlated predictors within predefined trait groups (G1-G4; each summarized is a separate sheet), reported for all species and for individual clades.

Supplementary Table 6 De novo mutational signatures

De novo mutational signatures inferred by signature analyzer, including signature profiles across 96 mutation channels (Sheet 1), per-branch signature activities (Sheet 2), and cosine similarity to COSMIC SBS signatures (Sheet 3).

Supplementary Table 7 Signature accumulation with divergence time

Correlations between total mutation burden or de novo signature burden and sister-taxa divergence time, reported for all species and by clade, using both Pearson and Spearman statistics.

Supplementary Table 8 Reference genomes and triplet definitions

NCBI reference genome assemblies considered in this study (Sheet 1) and the final set of species triplets used for CORAL analyses with associated TimeTree divergence times (Sheet 2).

Supplementary references

1. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://doi.org/10.48550/arXiv.1303.3997 (2013).
2. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
3. Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. https://escholarship.org/uc/item/1h3515gn (2014).
4. Vasimuddin, Md., Misra, S., Li, H. & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314–324 (2019).
5. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
6. Huang, Z. *et al.* Recurrent chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. *Nat. Commun.* **13**, 944 (2022).
7. Beichman, A. C. *et al.* Evolution of the Mutation Spectrum Across a Mammalian Phylogeny. *Mol. Biol. Evol.* **40**, msad213 (2023).
8. Blanchette, M. *et al.* Aligning Multiple Genomic Sequences With the Threaded Blockset Aligner. *Genome Res.* **14**, 708–715 (2004).
9. Foley, N. M. *et al.* A genomic timescale for placental mammal evolution. *Science* **380**, eabl8189 (2023).
10. Ho, S. Y. W., Phillips, M. J., Cooper, A. & Drummond, A. J. Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times. *Mol. Biol. Evol.* **22**, 1561–1568 (2005).
11. Ho, S. Y. W., Duchêne, S., Molak, M. & Shapiro, B. Time-dependent estimates of molecular evolutionary rates: evidence and causes. *Mol. Ecol.* **24**, 6007–6012 (2015).
12. Wang, Z. *et al.* Temporal genomic evolution of bird sex chromosomes. *BMC Evol. Biol.* **14**, 250 (2014).
13. Kumar, S. *et al.* TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol. Biol. Evol.* **39**, msac174 (2022).
14. Felsenstein, J. (1989). PHYLIP-Phylogeny Inference Package (Ver. 3.2). Cladistics, 5, 164-166 (1989).
15. Li, G., Davis, B. W., Eizirik, E. & Murphy, W. J. Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). *Genome Res.* **26**, 1–11 (2016).