Untangling the Evolutionary Dynamics of the Phenome in Megadiverse Hymenoptera

Diego S. Porto[a,1], **Lars Vilhelmsen**[b], **István Mikó**[c], **and Sergei Tarasov**[a]

This manuscript was compiled on February 28, 2025

Hymenoptera is a megadiverse insect order, yet the drivers of its massive radiation remain poorly understood. Here, we employ novel comparative phylogenetic methods that integrate computational ontologies with ancestral state reconstructions of 346 discrete morphological characters to investigate the tempo and mode of phenome evolution in Hymenoptera across both adults and larvae. We examine the interplay between phenotypic rates, selection, modularity, and lineage diversification over deep time. Our analyses reveal that both larval and adult phenomes experienced an early burst scenario during the Late Permian – Triassic. However, their evolution appears decoupled, which enhanced the overall evolvability of Hymenoptera. Moreover, major phenotypic transitions in the order are complex, multifaceted processes driven by directional selection. One such event—the emergence of the wasp waist—involved a fundamental reorganization of the adult phenome. This dramatic transformation shifted net diversification from negative to positive during the Triassic, thereby enabling the subsequent survival and explosive diversification of the apocritan lineage.

Phenomics | innovations | ontologies | ontophylo | larva | diversification | selection | morphospace

DRAFT

Hymenoptera is a megadiverse insect order with over 150,000 described species, including ants, social wasps, and bees (1). The remarkable diversity of hymenopteran morphology, physiology, and behavior has allowed this group to occupy diverse ecological niches since the crown lineage originated around 280 million years ago (Ma) (2). Despite advances in the understanding of Hymenoptera's phylogenetic relationships [(2–6); reviewed in (7)], the main drivers of their massive radiation still remain elusive. Four key innovations—wasp waist, stinger apparatus, parasitoidism, and secondary phytophagy—have been proposed to explain this evolutionary success. A recent study (2) found that only secondary phytophagy, which evolved independently in several lineages during the Cretaceous (around 80–105 Ma), significantly impacted species diversification. The roles of other putative key innovations remain unclear. Nonetheless, it was suggested that such innovations might have provided the phenotypic foundation for subsequent adaptations and diversification within Hymenoptera (2).

Are individual traits or key innovations sufficient to explain the evolution of entire phenomes? Organismal populations respond to different evolutionary forces (e.g., directional or stabilizing selection) through changes in their genotypes weighted by the fitness effects on their multidimensional phenotypes (8, 9). Some studies on vertebrates, for example, indicate that phenotypic evolution and diversification rarely stem from isolated traits alone, emphasizing the need to examine whole phenomes or major morphological/functional complexes in light of lineage-specific selection pressures and trait integration (10–12). Major shifts in anatomy and ecology often coincide with brief bursts of directional selection and increased phenotypic rates, as seen in tetrapods (13), mammals (11), birds (14, 15), and snakes (16). However, none of these studies have found a significant and/or direct correlation between the phenotypic/ecological shifts and lineage diversification rates (11, 12).

An important factor influencing phenotypic evolution is modularity, arising from trait integration into semi-independent units (modules), which may variously impact the tempo and mode of morphological evolution (17). For instance, on the one hand, high integration within distinct cranial regions allows them to evolve at different rates and modes across major avian (14) and mammal lineages (18). On the other hand, strong integration can limit the phenotypic space available for selection, slowing down evolution (10, 11, 14, 17). In insects, modularity is proposed to facilitate evolvability in Holometabola, where larval and adult stages are semi-independent, unlike the more integrated life stages of other insects (19). However, this hypothesis of evolutionary independence between different ontogenetic stages remains untested with modern methods.

Author affiliations: [a]Finnish Museum of Natural History, Helsinki, Finland; [b]Natural History Museum of Denmark, SCIENCE, University of Copenhagen, Copenhagen, Denmark; [c]University of New Hampshire, Durham, New Hampshire, USA

[1]To whom correspondence should be addressed. E-mail: diego.sassoporto@helsinki.fi

To date, no studies have focused widely on phenome evolution across the species-rich insects. Here, we examine the tempo and mode of phenome evolution in Hymenoptera, which due to their stunning diversity, represent an ideal case study for such an investigation. We address the following questions: (1) What are the evolutionary rates of adult and larval phenomes, and of major anatomical complexes of the adult phenome, across lineages and time? (2) Are adult and larval phenomes shaped by the same or different selective pressures and are these associated with phenotypic and lineage diversification rates? (3) Do major anatomical complexes of the adult phenome exhibit distinct modular patterns? (4) How do rates and selective pressures influence lineage diversification, disparity, and novelty? and (5) How has the interplay of these factors shaped the phenotypic landscape over macroevolutionary timescales?

To address these questions, we employ newly developed comparative methods tailored to discrete phenomic data (20, 21). These methods integrate computational ontologies (see (22) for an overview) with ancestral character state reconstructions (23), allowing inference of rate variation, selective regimes, and evolutionary dynamics at unprecedented detail. We applied these methods to the most recent time-calibrated phylogeny of Hymenoptera (2) and to a comprehensive morphological dataset for both adults and larvae comprising 346 discrete characters (24, 25).

Results

Rate evolutionary dynamics. Our primary approach for assessing dynamics of phenomes involves amalgamating the original 346 individual characters at three levels of the anatomical hierarchy—namely, the elementary anatomical complexes, the main body regions, and the entire phenome (21). Each amalgamated character is treated as a single entity for which we evaluate various dynamic metrics. For example, an amalgamated cranial character comprises 18 individual characters (S1 Appendix, Table S1).

Across the hymenopteran phylogeny, overall rate patterns for the adult phenome were similar when analyzing all characters (Fig. 1A), only sclerites, and only muscles (S1 Appendix, Fig. S6), with conspicuously higher rates estimated on branches leading to Vespina, Apocrita, and the clade including most Apocrita excluding Ichneumonoidea (Fig. 1A). For the larval phenome, higher rates were estimated on a few backbone branches preceding the emergence of Vespina/Apocrita (Fig. 1B). When considering rate variation through time, for the adult phenome (all characters, only sclerites, and only muscles), higher rates were observed in the Late Triassic (201∼237 Ma) (Figs. 1C and 6A), while for the larval phenome, higher rates were observed before the boundary between the Middle/Late Triassic (∼237 Ma) (Figs. 1E and 6B).

Rate patterns of individual anatomical complexes were more variable, but largely agreed with those observed for the full adult phenome. For example, for most anatomical complexes of the adult (e.g. cranium, mouthparts, mesonotum, mesopectus, and metapectal-propodeal complex), higher rates were estimated on branches leading to Vespina and Apocrita, but also other major lineages such as Ichneumonoidea, the "Evaniomorpha"-Aculeata clade, and Proctotrupomorpha (Fig. 2). For some complexes, however, higher rates were estimated only on branches leading to a few major clades such as Ichneumonoidea (metasoma), Aculeata (female genitalia), and lineages within Proctotrupomorpha (fore- and hind wings). Similar results were obtained for body regions, particularly the mesosoma (S1 Appendix, Fig. S7). Rate variation through time followed similar patterns to that observed for the whole phenome, particularly for the mesonotum, mesopectus, and metapectal-propodeal complex (anatomical complexes) and mesosoma (body regions), with higher rates observed in the Late Triassic (201∼237 Ma) (S1 Appendix, Fig. S20A, C, E).

For all empirical analyses, results from their respective null analyses do not suggest that evolutionary models nor phylogenetic structure influenced the observed rate patterns inferred from data for the adult/larval phenomes, body regions, or anatomical complexes (S1 Appendix, Figs. S8-10, S20B, D, F, H).

Morphospace reconstructions. Reconstructed morphospaces for the full adult phenome were more concentrated in the empirical analysis than in the null analysis (Fig. 3A-B; note the difference in scale between the plots). On the other hand, morphospace occupation was not substantially different between empirical and null analyses for the larval phenome (S1 Appendix, Figs. S11, S14) and for some anatomical complexes of the adult (e.g., female genitalia, fore- and hind wings, mid and hind legs), usually the partitions including a smaller number of amalgamated characters (S1 Appendix, Figs. S13, S16 and Table S1). For other anatomical complexes of the adult (e.g., mouthparts, mesonotum, mesopectus, and metapectal-propodeal complex) and all body regions of the adult, the overall occupied area of the morphospace was also similar between the empirical (S1 Appendix, Figs. S12-13) and null analyses (S1 Appendix, Figs. S15-16), but there was substantially more phylogenetic structure in the former. These were usually the partitions including a larger number of amalgamated characters (S1 Appendix, Table S1).

Morphospace occupation-through-time for the full adult phenome also differed between empirical and null analyses (Fig. 3A-B, sequential subpanels from left to right). Walks in the morphospace were initially restricted and slow during the Early Triassic (250 Ma) in both the empirical and null analyses but accelerated and quickly became directional during the Late Triassic (225 Ma) on the path to Vespina and Apocrita in the empirical analysis (Fig. 3A; S1 Appendix, Movie S1).

Similar results for the adult phenome were recovered when analyzing only sclerites (S1 Appendix, Movie S2); but less so when analyzing only muscles (S1 Appendix, Movie S3).

Selection regimes. Results from the qualitative assessments of the adult and larval phenomes were complemented by our new method for estimating branch-specific selection regimes (quantified by the S statistic) at macroevolutionary scales. In short, the S statistic characterizes selection regimes as follows: positive S indicates directional selection, negative S indicates stabilizing selection, and non-significant S values imply diffusive evolution, analogous to Brownian motion (Fig. 3E).

For the full adult phenome, significantly positive S values—indicative of strong directional selection—were inferred on the backbone branches leading to Vespina/Apocrita (Fig. 3A,

DRAFT

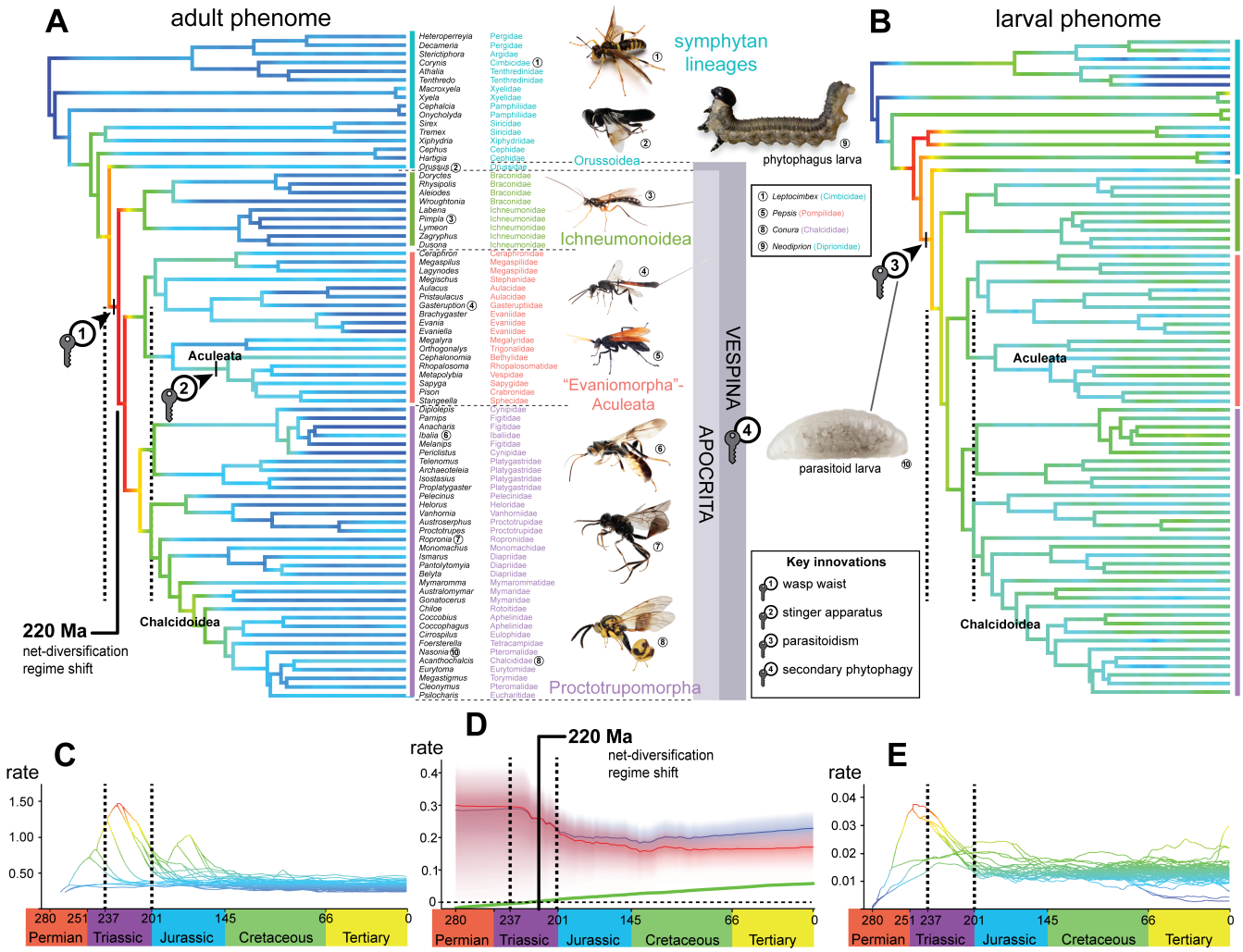

Fig. 1. Evolutionary rate dynamics inferred with ontophylo. *(A)* Branch-specific rates for the full adult phenome. *(B)* Same as (A) but for the larval phenome. *(C)* Rate-through-time variation for each branch (edgeplots) for the full adult phenome. *(D)* Rate-through-time variation for net diversification (green), speciation (blue), and extinction (red) as inferred in the BAMM analyses of Blameir et al. (2). *(E)* Same as (C) but for the larval phenome. Vertical dashed lines indicate the limits of the Late Triassic. Image credits to David Cheung (5), Hiroshi Nakamine (1), and Line Kræmer (3,4,9).

C, orange-shaded branches; S1 Appendix, Figs. S17A and S19A). Directional selection, albeit less pronounced, was also estimated on branches leading to other major Hymenoptera lineages (e.g., Ichneumonoidea, Proctotrupomorpha, and lineages within Aculeata). Conversely, negative S values, suggesting stabilizing selection, were estimated on several lineages outside Vespina/Apocrita as well as on long branches leading to extant taxa across Hymenoptera (Fig. 3A, C, purple-shaded branches; S1 Appendix, Figs. S17A, S19A). Results from the analyses of individual anatomical complexes were more variable but largely consistent with those obtained for the full adult phenome, particularly for mesosomal structures (e.g., mesonotum, mesopectus, and the metapectal-propodeal complex; S1 Appendix, Fig. S18).

For the larval phenome, strong directional selection was observed among several symphytan lineages, while stabilizing selection was detected within Tenthredinoidea (e.g., Argidae, Cimbicidae, and Tenthredinidae). Moreover, non-significant S estimates—indicative of diffusive evolution—were found across most lineages within Apocrita (Fig. 3D; S1 Appendix, Figs. S17B, S19B).

Phenotypic rate correlations. To assess correlations among phenotypic rates of evolution and search for putative semi-independent evolutionary units (i.e., evolutionary modules), we performed pairwise comparisons between all individual anatomical complexes of the adult and the larval phenome. We hypothesize that groups of highly correlated anatomical structures should also exhibit highly correlated evolutionary dynamics. We employed a resampling-based approach to estimate the degree and significance of correlations between branch-specific rates (see Materials and Methods, and S1 Appendix for details).

From all possible pairwise comparisons (15 anatomical complexes + larval phenome: 120 pairs in total), significant correlations were observed between many anatomical complexes of the mesosoma (e.g., fore- and hind wings, fore- and hind legs, propectus, mesonotum, mesopectus, metapectal-propodeal complex) but no correlation between any adult

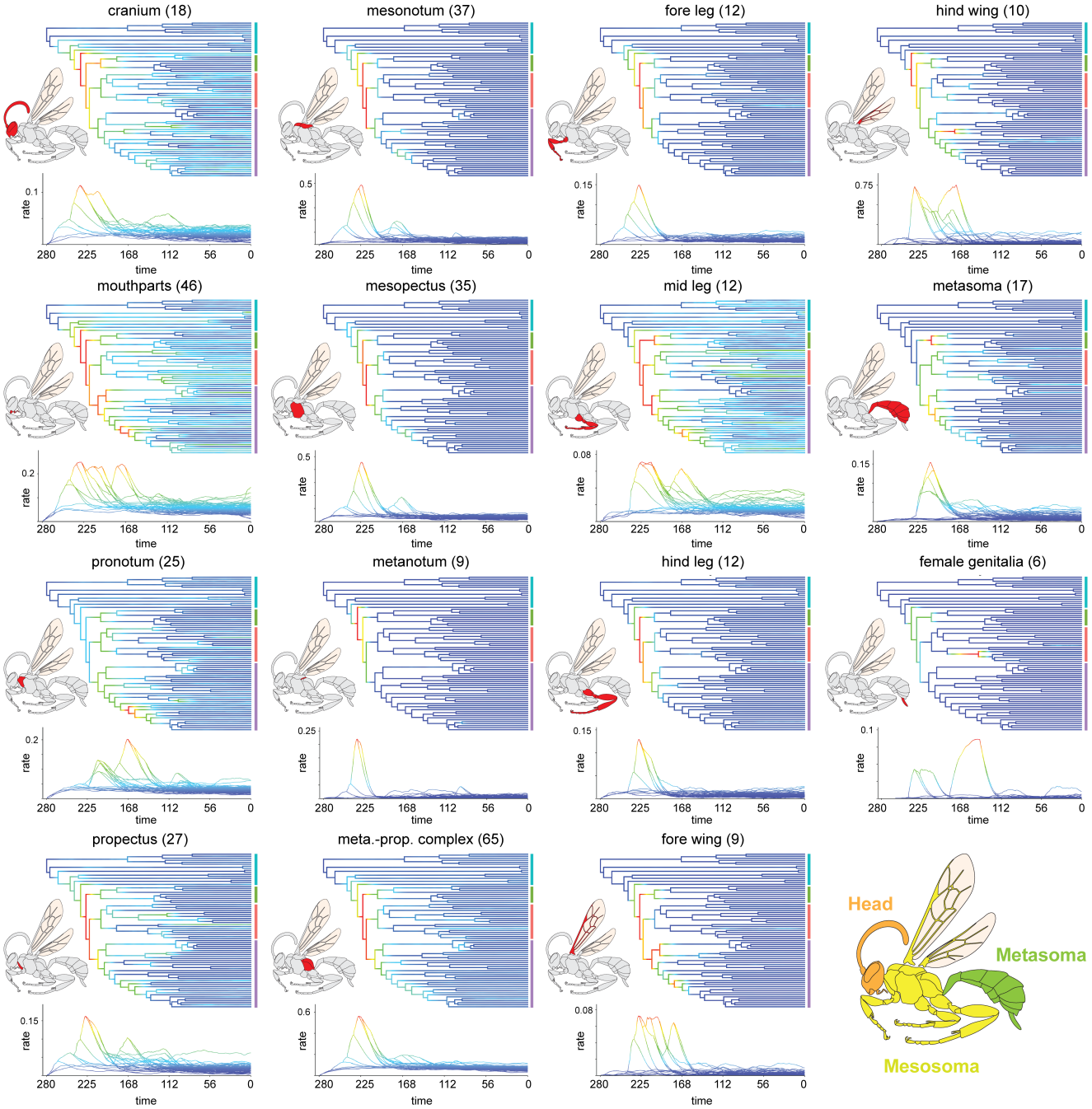

Fig. 2. Branch-specific evolutionary rate dynamics inferred with ontophylo. Each subpanel shows the result for an individual anatomical complex of the Hymenoptera anatomy (highlighted in red in the anatomical diagrams). In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rate-through-time variation for each branch (edgeplots); color gradient from blue to red indicates low to high rates respectively. Color codes of main clades as in Fig. 1. Values between parentheses indicate the total number of amalgamated characters in each anatomical complex. Bottom-right diagram shows the three main body regions of the Hymenoptera (Apocrita) anatomy shaded with different colors.

anatomical complex and the larval phenome was identified (Fig. 4A-C; S1 Appendix, Table S2).

Employing the same approach as described above, we found that phenotypic rates of evolution were significantly correlated with our new selection regime statistic (S). In particular, high evolutionary rate values were generally associated with directional selection across all anatomical complexes of the adult phenome—whether analyzed individually (S1 Appendix, Table S10) or pooled together (Fig. 5C). The same was observed when analyzing the full adult phenome (S1 Appendix, Table S10).

In contrast, no such correlation was observed for the larval phenome (S1 Appendix, Table S10). Although our S statistic is designed to decouple phenotypic rates from evolutionary regimes, our dataset revealed a strong correlation between them, a finding that is consistent with previous studies (13, 26).

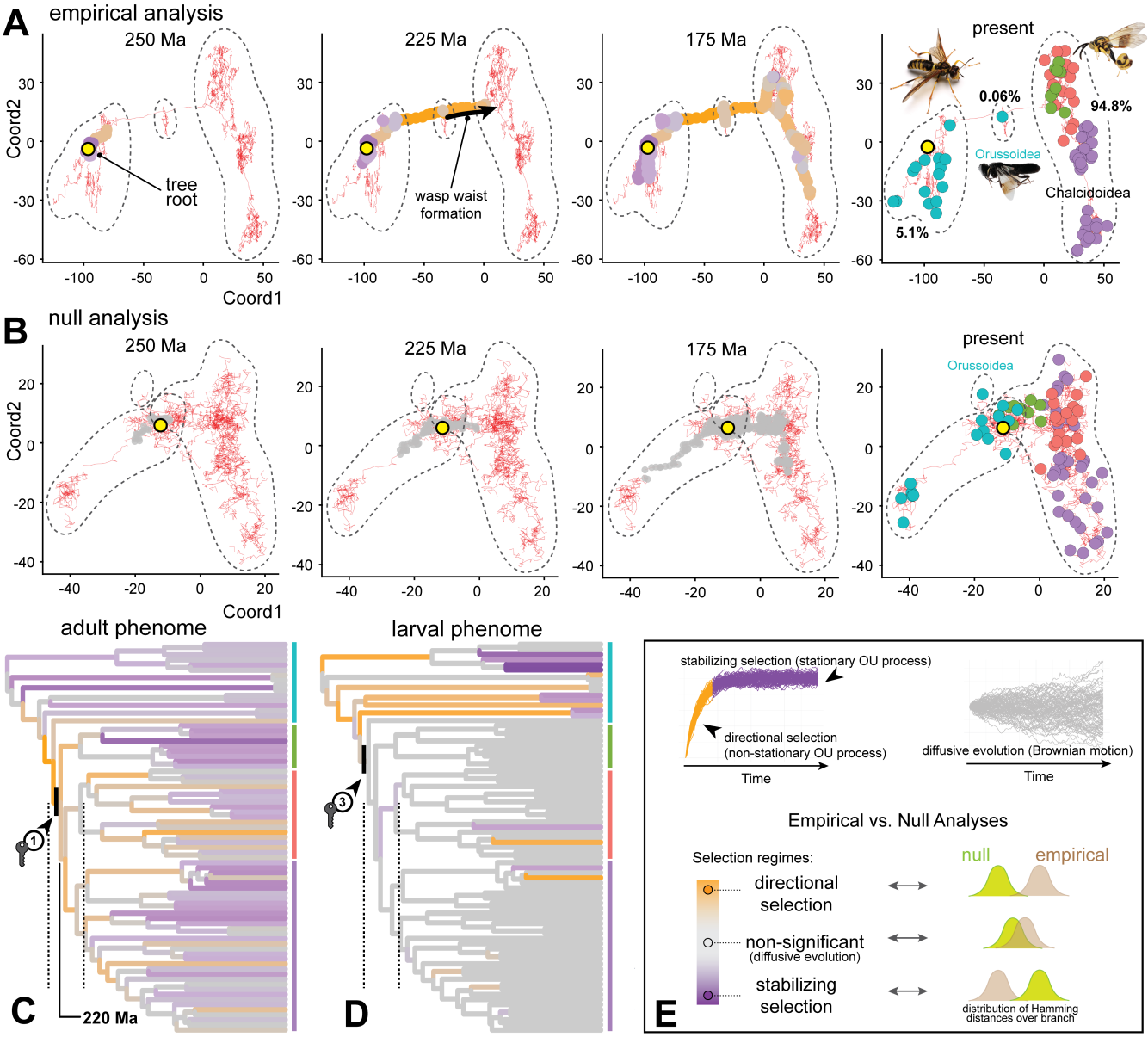

Fig. 3. Morphospaces reconstructed with ontophylo and macroevolutionary selection regimes. *(A)* Snapshots at different moments in time of the morphospace of the full adult phenome for the empirical analysis. *(B)* Same as (A) but for the null analysis. *(C)* Macroevolutionary selection regimes for the full adult phenome plotted onto tree branches according to the new metric proposed here (see main text and S1 Appendix for detailed explanation). *(D)* Same as (C) but for the larval phenome. *(E)* Analogy for interpreting the evolution of multiple discrete traits as a single continuous trait and its relation to our new metric to assess macroevolutionary selection regimes. In (A-B), dashed contour lines indicate symphytan lineages, Orussoidea, and Apocrita, from left to right, respectively; in the right-most panels, numbers indicate present-day estimated relative diversity and color codes for taxa as indicated in Fig. 1. In (C-D) and the first three panels in (A-B), from left to right, color codes for the macroevolutionary selection regimes as indicated in (E). Key innovations indicated in (C-D) as explained in Fig. 1. Vertical dashed lines indicate the limits of the Late Triassic. Image credits as in Fig. 1.

Diversification rate correlations. Using the same resampling-based approach mentioned above, we also investigated the association between branch-specific rates of phenotypic evolution and selection regimes with branch-specific net diversification, extinction, and speciation rates, which were extracted from the most recent BAMM analysis of Hymenoptera (2). We observed significant negative correlations between phenotypic rates and lineage net diversification rates in the adult phenome, particularly for certain mesosomal anatomical complexes (e.g., propectus, mesonotum, mesopectus, and the metapectal-propodeal complex; see S1 Appendix, Table S4 for individual complexes and the full adult phenome and Fig. 5A for a pooled analysis of all anatomical complexes). Similar patterns were found when contrasting selection regimes with lineage net diversification rates, with significant negative correlations also detected for several mesosomal complexes (see S1 Appendix, Table S7 for individual complexes and the full adult phenome and Fig. 5D for a pooled analysis of all anatomical complexes).

In contrast, no correlations were found between any anatomical complexes or the full adult phenome and speciation or extinction rates (Fig. 6B, speciation; S1 Appendix, Tables S5-6), nor were any correlations observed for the larval phenome across all diversification measures, including

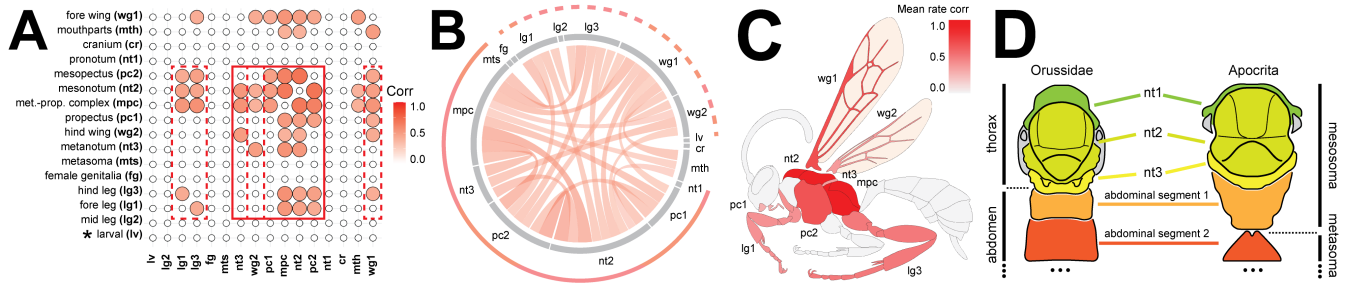

Fig. 4. Rate correlations among anatomical complexes of the adult (and larval phenome). *(A)* Correlogram. *(B)* Connectogram. *(C)* Anatomical diagram color-shaded by the mean correlation for each anatomical complex across all partitions (i.e., row means from the correlation matrix used in the correlogram). *(D)* Anatomical diagrams (dorsal view) illustrating the conditions of the thorax and abdomen before (left) and after (right) the wasp waist formation. Color intensity in (A-C) and thickness of connections in (B) indicate strength of correlations (all positive). Solid boxes in (A) and arch in (B) indicate core structural mesosomal structures; dashed boxes and arch indicate appendicular mesosomal structures. Abbreviations of anatomical complexes in (B-D) as indicated in (A). Asterisk in (A) indicates the larval phenome.

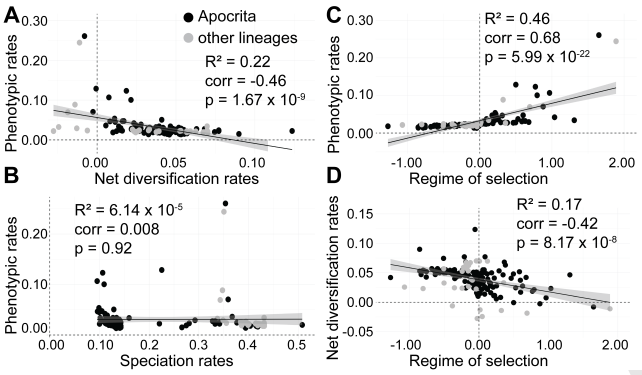

Fig. 5. Scatter plots showing the association between phenotypic and lineage diversification rates and selection regimes; data points correspond to mean values for each branch across the pooled information from all individual anatomical complexes. *(A)* Association between phenotypic and lineage diversification rates (net diversification). *(B)* Same as (A), but for speciation. *(C)* Association between phenotypic rates and selection regimes. *(D)* Association between lineage diversification rates (net diversification) and selection regimes. In each subfigure, R^2 denotes the coefficient of determination, corr denotes the Pearson's correlation, and p denotes the p-value for the linear regressions.

net diversification, speciation and extinction (S1 Appendix, Tables S4-6).

Finally, correlations between selection regimes and speciation or extinction rates were observed only for two anatomical complexes (cranium and pronotum), and between the full adult phenome and speciation (S1 Appendix, Tables S8-9).

Metrics of phenomic evolution. To describe different aspects of phenomic evolution, in addition to the metric for assessing selection regimes, we developed two additional metrics: one for quantifying phenotypic innovation and another for measuring phenotypic disparity. These metrics are computed using information obtained from stochastic map samples (see Materials and Methods and S1 Appendix for details).

Curves of phenotypic innovation through-time in the empirical analyses showed steeply accumulation of new character states during the Late Triassic (201∼237 Ma) for the adult phenome when analyzing all characters, only sclerites, and only muscles (Fig. 6C; S1 Appendix, Fig. S21E). Similar results were observed for all body regions and most anatomical complexes of the adult phenome (S1 Appendix, Fig. S21A, C). Similar patterns were also observed for the disparity-through-time curves (Fig. 6E; S1 Appendix, Fig. S22A, C, E). Furthermore, our results indicate that relatively more innovations were accumulated in Vespina/Apocrita than in the remaining Hymenoptera (S1 Appendix, Figs. S23-25).

As for the larval phenome, both curves of phenotypic innovation and disparity-through-time in the empirical analyses showed similar patterns, with slightly steeper accumulation of new character states (Fig. 6D; S1 Appendix, Fig. S21G) and phenotypic divergence (Fig. 6F; S1 Appendix, Fig. S22G) before the boundary between the Middle/Late Triassic (∼237 Ma). Accumulation of innovation was more evenly distributed across Hymenoptera (S1 Appendix, Fig. S23).

In contrast, in the null analyses, all curves of phenotypic innovation and disparity through-time exhibited distinct patterns, with the accumulation of new character states and phenotypic divergence increasing gradually toward the present (Fig. 6C–F, dotted lines; S1 Appendix, Figs. S21B, D, F, H and S22B, D, F, H). Similarly, the accumulation of innovations was more evenly distributed across Hymenoptera (S1 Appendix, Figs. S26–28).

Discussion

Phenotypic rates and macroevolutionary regimes. Our analyses indicate that both larval and adult phenomes—including most anatomical complexes, major body regions, sclerites, and muscles—follow an early burst pattern, with markedly elevated rates during the Late Permian – Middle Triassic (Figs. 1, 2, 6). These accelerated dynamics coincide with episodes of strong directional selection and significant increases in phenotypic innovation and disparity. Ultimately, these processes culminated in two key morphological transitions (2): the emergence of the wasp waist in adults and the parasitoid body plan in larvae (Figs. 1 and 4D). We interpret these events as the outcome of newly available ecological niches after the Permian-Triassic mass extinction event (27), which were rapidly exploited by novel phenotypes undergoing intense selection.

Interestingly, our results also show that the phenotypic rates and selection regimes for adult and larval phenomes are decoupled (Figs. 1, 3, 4, 6). This finding aligns with the long-standing hypothesis that the independent evolution of larvae and adults in Holometabola facilitated the exploration of distinct ecological niches by reducing intraspecific competition between different life stages (28), and enhancing evolvability

DRAFT

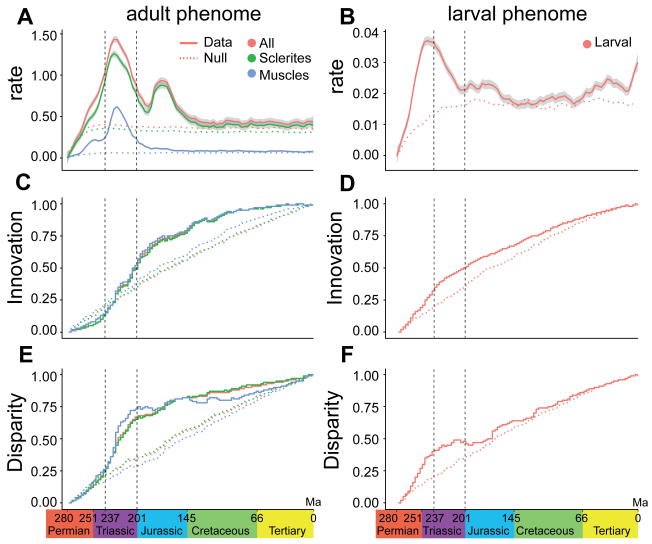

Fig. 6. Evolutionary rates and metrics of phenomic evolution. *(A)* Rate-through-time curve for the adult phenome (all characters, only sclerites, and only muscles). *(B)* Same as (A) but for the larval phenome. *(C)* Innovation-through-time curve for the adult phenome (all characters, only sclerites, and only muscles). *(D)* Same as (C) but for the larval phenome. *(E)* Disparity-through-time curve for the adult phenome (all characters, only sclerites, and only muscles). *(F)* Same as (E) but for the larval phenome. For each curve (i.e., rate, innovation, and disparity), solid lines indicate results from the empirical analyses and dotted lines results from the null analyses; vertical dashed lines indicate the limits of the Late Triassic.

(19). In contrast, species with incomplete metamorphosis (non-Holometabola) exhibit more correlated life stages due to similar morphologies, which is thought to constrain their evolution (19). Our study is the first to demonstrate the evolutionary decoupling in Holometabola (at least in Hymenoptera) through comparative analyses.

Furthermore, our findings indicate that the adult phenome evolves in a modular fashion (Fig. 4). Many anatomical complexes display independent phenotypic rates (Fig. 2), whereas several regions of the mesosoma exhibit significant inter-correlations associated with high evolutionary rates (S1 Appendix, Tables S2-3). Such modularity promotes adaptation through mosaic evolution, analogous to patterns observed in the cranial modules of birds and mammals, where different regions evolve at distinct tempos and modes (14, 18). This pattern is consistent with theoretical predictions (17) suggesting that while the magnitude of trait correlation may constrain the morphospace, it does not necessarily limit the evolutionary rates of individual modules.

Adaptive landscape of Hymenoptera. Mathematical models suggest that the geometry of empirical morphospaces can often emerge from simple processes—such as random walks—without invoking additional mechanisms (29, 30). To test this, we compared null morphospace reconstructions, representing a purely diffusive random walk on a morphological hypercube (30), with our empirical reconstructions using observed data (Fig. 3). For larvae, the empirical and null morphospaces were nearly indistinguishable (S1 Appendix, Figs. S11 and S14), likely because our dataset—despite being the most comprehensive available—comprises a relatively limited number of original characters.

In adults, however, the null morphospace produced clusters similar to the empirical data, indicating that phylogeny and phenotypic rates, without invoking additional mechanisms, partially account for the observed morphological diversity in Hymenoptera (Fig. 3B). However, the empirical reconstruction has revealed markedly distinct lineage clusters (Fig. 3A). These clusters appear to result from repeated pulses of strong directional selection that expanded the morphospace beyond the predictions of the null model, representing pivotal adaptive events. The most pronounced pulse (Fig. 3A), occurring from the Late Permian to the Late Triassic, led to the emergence of the highly speciose Apocrita (94.8% of extant species), whereas the symphytan lineages, which have drifted little in morphospace since the origin of Hymenoptera, remain species-poor (5.2% of extant species). Thus, in Hymenoptera, directional selection represents a key mechanism that has shaped the geometry of morphospace and, consequently, the observed diversity of phenotypes.

Phenotypic Revolutions and Singularity. The pulses of directional selection during the Triassic triggered two major evolutionary singular events—ensembles of key innovations—that fundamentally reorganized both adult and larval phenomes.

The first event involved the transition of larvae from a caterpillar-like morphology with elongated bodies, clearly demarcated integumental annulation, and well-developed appendages to a grub-like morphology with compact bodies, reduced integumental annulation, and reduced appendages, reflecting a shift from an active herbivorous lifestyle to a specialized parasitoid mode of life (Fig. 1B) (31).

The second event was the emergence of the wasp waist in Apocrita (Fig. 1A, 4D). The formation of the wasp waist involved the complete morpho-functional integration of the first abdominal segment into the thorax, including the fusion of the first abdominal tergum with the metathorax. This integration established a new articulation and resulted in two distinct body regions: the mesosoma (comprising the thorax and the first abdominal segment) and the metasoma (the remaining abdomen) (32, 33). This transition marked a major shift in thorax-abdomen articulation, enhancing its mobility in association with oviposition and emergence of parasitoid behavior in Apocrita (32, 33).

The evolutionary magnitude of both the wasp waist and the larval parasitoid body plan is comparable to other major phenotypic singularities, such as the transformation of the fore wings into elytra in coleopterans and hind wings into halters in dipterans (34), the fin-to-limb shift in early tetrapodomorphs (13), and limb loss with body elongation in snakes (12).

Phenomes and Diversification. The phenotypic rates in adult Hymenoptera were found to be negatively correlated with net diversification rates (Fig. 5A; S1 Appendix, Table S4). We interpret this association as likely coincidental, arising from (1) a shared across-lineage trend—globally decreasing phenotypic rates alongside monotonically increasing net diversification rates—and (2) the absence of significant correlations between phenotypic and either speciation or extinction rates for all anatomical complexes or the full adult phenome (S1 Appendix, Tables S5-6).

More intriguing, however, is the shift in net diversification regime that occurred around 220 Ma during the transition from symphytan to early apocritan lineages (Fig. 1C, D,

DRAFT

E). This event coincided with strong directional selection across multiple anatomical complexes and the formation of the wasp waist (Fig. 2; S1 Appendix, Fig. S18). During this period, both speciation and extinction rates declined—with extinction dropping more sharply—resulting in a shift from negative to positive net diversification regime, as evidenced by both net diversification–through–time and lineage–specific rate plots (Fig. 1D; S1 Appendix, Fig. S29).

It seems unlikely to us that this shift—accompanied by directional selection and major morphological changes—is purely coincidental. A plausible explanation is that these singular events, which have profoundly changed the evolutionary trajectory of Hymenoptera, simply lack sufficient replication to yield a statistically detectable correlation (35).

Nonetheless, the overall decoupling of diversification is intriguing and consistent with previous work (2), which found no strong association between speciation/extinction rates and key innovations in Hymenoptera (except for secondary phytophagy among some lineages in Apocrita). A similar pattern has been observed in lepidosaurs (13, 16). However, this lack of overall correlation is not universal; studies in other taxa have reported a positive association between phenomic rates and species diversification (11, 12, 14).

It is quite possible that diversification in Hymenoptera is closely linked to biochemical, physiological, or behavioral traits that may have facilitated the exploitation of new ecological opportunities or promoted reproductive isolation (36). Such factors may not be directly captured by the skeletomuscular phenotypes examined here, and their potential influence require further investigation.

Materials and Methods

Phenomic dataset. Adult characters were obtained from the phylogenetic character matrix of Sharkey et al. (24). The original matrix contained 401 characters scored for 98 taxa. From the original dataset, 21 taxa and 18 characters were removed (see S1 Appendix for details). The resultant adult matrix contained 383 characters scored for 77 taxa (Dataset S1). Larval characters were obtained from the phylogenetic character matrix of Ronquist et al. (25). From the original matrix, only characters 221-236 (16 characters) corresponded to larval characters. The original characters were modified and recoded (see S1 Appendix for details) resulting in 11 characters scored for 77 taxa (Dataset S2).

Character coding and dependencies. As discussed elsewhere (37, 38), dependencies are common in morphological matrices and should be modeled properly by combining dependent characters into composite characters to avoid misleading results in analyses of ancestral state reconstruction (ASR). For the purpose of this study, characters from the adult matrix were classified into six groups according to the types of dependencies observed (Dataset S3, G1-6) and coded/recoded accordingly following the recommendations of (37–39) (see S1 Appendix for details). Likewise, some larval characters showed dependencies and were recoded as well (Dataset S4) (see S1 Appendix for details).

Character annotations. Organismal anatomy is a hierarchy of nested sets of anatomical entities (=morphological structures). By annotating the anatomical entities referred to in phylogenetic morphological characters with terms from an anatomy ontology (anatomical concepts), the knowledge about the hierarchy of morphological structures can be automatically retrieved and sets of characters grouped based on anatomical complexes or body regions. The new comparative methods employed in this study utilize the information from ontology annotations to cluster characters based on the anatomical hierarchy. Characters from the adult matrix were annotated with terms from the Hymenoptera Anatomy Ontology (HAO) (40) (Dataset S3) using ontoFAST (41) (see S1 Appendix for details). Larval characters were not annotated with specific ontology terms and were all treated as a single cluster.

Time-calibrated tree. ASR analyses utilized a dated phylogeny modified from topology C1 of Blaimer et al. (2). The original tree contained 765 hymenopteran species and six outgroup taxa. This tree was pruned to match the information available for adult (Dataset S1) and larval characters (Dataset S2) resulting in a final sampling of 77 taxa (Dataset S5) (see S1 Appendix for details).

Empirical phenomic analyses. Empirical analyses referring to phenomic evolution were assessed using the ontophylo workflow (21), a new set of comparative methods developed for combining (=amalgamating) ASR of multiple individual characters and analyzing their combined character histories. In a nutshell, the workflow allows combining individual characters into complex characters to represent the joint evolution of different anatomical complexes, body regions, or the entire organismal phenome (20, 42). The workflow consists of the three main steps of the PARAMO pipeline (20) and the ontophylo analyses to estimate evolutionary rate dynamics and reconstruct morphospaces through time (see S1 Appendix for details). For non-dependent characters, empirical model fitting was performed using functions implemented in the R package corHMM (43). For the dependent characters, transition rate matrices (Q matrices) were initially set using functions from the R package rphenoscate (39) and then parameters estimated with corHMM. ASR analyses employed stochastic character mapping (23) as implemented in corHMM. Adult characters were clustered at three different levels of the anatomical hierarchy: L1: anatomical entities (15 anatomical partitions), L2: body regions (3 anatomical partitions), and L3: the entire phenome. Larval characters were clustered as a single partition (6 chars) (see S1 Appendix for details). Character amalgamations were performed using functions implemented in the R package ontophylo (21). Amalgamations for L1 and L2 used the information about the anatomical hierarchy from the HAO (40) ontology (see S1 Appendix for details). Additionally, for the adult phenome, separate analyses were performed including all characters, only sclerites, and only muscles.

To assess potential biases due to partition size, a resampling approach was employed to obtain comparable samples from each partition and then the mean rates of individual characters and individual partitions were calculated and contrasted (see S1 Appendix for details). To evaluate the power and biases of the inference, we contrasted the results from the empirical analyses described above with those from the null analyses described below.

Estimating selection regimes. The recently proposed method of estimating selection regimes from discrete characters on phylogenetic trees (13) has significantly advanced our ability to assess macroevolutionary dynamics. This approach relies on comparing inferred rates to a mean background rate—both estimated from the data (44) and inspired by similar molecular methods using d_n/d_s. However, unlike molecular evolution, where the background rate can be inferred from neutral loci, the concept of a "background rate" in discrete morphology is more elusive because we do not know a priori which characters evolve neutrally.

To correct for the unobserved neutral evolution, we developed a novel method that quantifies the trajectory of phenotypes through the phenotypic space over time by contrasting null and empirical ancestral state reconstructions. This approach is motivated by continuous characters: if explicit reconstructions of trait trajectories were available, one could distinguish directional from stabilizing selection by examining those trajectories (Fig. 3E). Similarly, a directional walk in the discrete phenotypic space of amalgamated characters would correspond to directional selection, whereas orbiting around a focal point would indicate stabilizing selection. In contrast, an undirected walk through discrete phenotypes is analogous to a discrete bounded diffusion process, corresponding to Brownian motion in the continuous setting (i.e., diffusive evolution in Fig. 3E).

DRAFT

To assess these modes of evolution—and thus infer the underlying selection regime—we define the selection statistic S as:

$$S = \frac{\bar{H}_{\text{emp}} - \bar{H}_0}{\sigma},$$

where $\bar{H}_{\text{emp}}$ is the mean Hamming distance computed from the empirical reconstructions and $\bar{H}_0$ is the mean Hamming distance from the null model (i.e., reconstructions not conditioned on the observed data). For each branch, the Hamming distance is calculated as $H(c_s, c_e)$, with c_s and c_e representing the character state at the beginning and end of the branch, respectively. The mean Hamming distances $\bar{H}_{\text{emp}}$ and $\bar{H}_0$ are estimated from the samples of the stochastic maps (see the explanatory diagram in S1 Appendix, Fig. S5C).

Within this framework, a positive S indicates directional selection (i.e., a systematic directional shift in phenotype), whereas a negative S suggests stabilizing selection (i.e., a tendency to remain near a phenotypic optimum). To assess the significance of our S estimates, we use bootstrap resampling to generate a distribution of differences between the empirical and null values (sample size = 1000, 10,000 replicates; see S1 Appendix for details). Significance is determined by the non-overlap of the 95% confidence interval with zero. If there is no overlap, the S estimate is considered significant; if there is overlap, the result is interpreted as evidence for diffusive evolution, indicating that no specific selection regime can be reliably inferred.

Metrics to assess phenomic evolution. To calculate the metrics to assess phenomic evolution (and regime of selection), amalgamated stochastic maps sampled for each anatomical complex (L1), body region (L2), and the entire phenome (L3) were used. Each amalgamated map comprised the combined histories of multiple discrete characters clustered in a given anatomical partition (e.g., as defined in L1-L3). Metrics were calculated using all maps in a sample for a given partition. Phenotypic innovation was calculated for each map and each root-to-tip path as the number of new unique individual character states accumulated. The number of new unique states was obtained as the sum across all positions in the vector representing the amalgamated character state (see explanatory diagram in S1 Appendix, Fig. S5A). Phenotypic disparity was calculated for each map and each time slice (based on a predefined number of time bins) as the mean Hamming distance between amalgamated character states sampled across all lineages (see explanatory diagram in S1 Appendix, Fig. S5B).

Rate correlations. Mean rates were calculated for each branch of the phylogeny for each anatomical complex defined in L1 (15 anatomical partitions), the larval and the full adult phenome. Since ontophylo utilizes a modified Markov kernel for density estimation, allowing rates to vary along tree branches, we opted instead for using the raw data from the amalgamated stochastic maps to calculate branch rates. Branch rates were calculated simply as the number of observed character-state transitions divided by the branch length. Rates were calculated individually for each branch and tree in a sample of stochastic maps. To assess correlations among phenotypic rates, all pairwise comparisons between anatomical complexes (+ larval phenome) were performed. To check for possible correlations with lineage diversification rates, the phenotypic rates were also compared with the branch-specific net diversification, speciation, and extinction rates extracted from the BAMM results of Blaimer et al. (2) (see S1 Appendix for details). Additionally, correlations between phenotypic rates and our new branch-specific metric of regime of selection and between regime of selection and lineage diversification rates (net diversification, speciation, and extinction) were also investigated. Correlations were initially assessed by employing two strategies (see S1 Appendix for details). The first strategy was a resampling approach similar to the one used in (45, 46). For each pairwise comparison, a stochastic map was randomly selected from the samples of both anatomical partitions, branch rates were obtained, and Spearman's rank correlation was calculated (i.e., observed correlation). Then, permutations were performed by reshuffling branches (1000 replicates), re-calculating correlations, and calculating the proportion of values equal to or greater than the observed correlation value to obtain a p-value. For each pairwise comparison, 1000 stochastic maps were resampled. The second strategy was to perform linear regression including not only the branch rates from the empirical analyses but also the branch rates from the null analyses as additional variables to account for possible common effects due to shared evolutionary models and phylogeny. This strategy was employed only to asses correlations among phenotypic rates and to contrast with the results from the first strategy. A similar analytical setting was used, as described above, with 1000 iterations in total. For each iteration, the p-value and angular coefficient between the two rate variables of the empirical analyses were extracted and the beta coefficient converted to obtain the value of Pearson's correlation. For both strategies, significance of correlation was determined by p-values equal to or below the threshold of 0.05 in 95% of the iterations or more.

Null analyses. Phenomic analyses were also performed employing unconditioned evolutionary models ("null analyses"). The simulation analyses were used to sample 'null' evolutionary scenarios to assess the baseline effects of phylogenetic structure and evolutionary models on rate dynamics and morphospace reconstructions. Simulations used the same settings as the empirical analyses except for the ancestral state reconstructions. Instead, character histories were sampled using stochastic character mapping, using the same models and parameters fitted with corHMM for the empirical analyses, but not conditioned on character states observed at the tree tips. Simulations used functions implemented in the R package phytools (47). Additionally, the new metrics to assess phenomic evolution and regime of selection were also calculated for the null character histories. Null analysis outputs were used to examine the evolutionary rate dynamics and morphospace organization for all anatomical complexes, body regions, and both adult and larval phenomes. These outputs also enabled us to assess the significance of our selection regime statistic and to evaluate the strength and significance of rate correlations (see S1 Appendix for details).

Data availability. All data and scripts necessary for reproducing the analyses and figures of this manuscript are available on Zenodo (XXXXX).

ACKNOWLEDGMENTS. DSP was supported by the Research Council of Finland (grant 362624). ST was also supported by the Research Council of Finland (grants 339576 and 346294).

DRAFT

1. JT Huber, Biodiversity of hymenoptera. *Insect biodiversity: science society* pp. 419–461 (2017).
2. BB Blaimer, et al., Key innovations and the diversification of hymenoptera. *Nat. Commun.* **14**, 1212 (2023).
3. RS Peters, et al., Evolutionary history of the hymenoptera. *Curr. Biol.* **27**, 1013–1018 (2017).
4. MG Branstetter, et al., Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. *Curr. Biol.* **27**, 1019–1025 (2017).
5. P Tang, et al., Mitochondrial phylogenomics of the hymenoptera. *Mol. Phylogenetics Evol.* **131**, 8–18 (2019).
6. S Wutke, et al., Phylogenomics and biogeography of sawflies and woodwasps (hymenoptera, symphyta). *Mol. Phylogenetics Evol.* **199**, 108144 (2024).
7. YM Zhang, S Bossert, T Spasojevic, Evolving perspectives in hymenoptera systematics: Bridging fossils and genomes across time. *Syst. Entomol.* (2024).
8. D Houle, Numbering the hairs on our heads: the shared challenge and promise of phenomics. *Proc. Natl. Acad. Sci.* **107**, 1793–1799 (2010).
9. D Houle, DR Govindaraju, S Omholt, Phenomics: the next challenge. *Nat. reviews genetics* **11**, 855–866 (2010).
10. C Bardua, et al., Evolutionary integration of the frog cranium. *Evolution* **74**, 1200–1215 (2020).
11. A Goswami, et al., Attenuated evolution of mammals through the cenozoic. *Science* **378**, 377–383 (2022).
12. PO Title, et al., The macroevolutionary singularity of snakes. *Science* **383**, 918–923 (2024).
13. TR Simões, SE Pierce, Sustained high rates of morphological evolution during the rise of tetrapods. *Nat. Ecol. & Evol.* **5**, 1403–1414 (2021).
14. RN Felice, A Goswami, Developmental origins of mosaic evolution in the avian cranium. *Proc. Natl. Acad. Sci.* **115**, 555–560 (2018).
15. MS Lee, A Cau, D Naish, GJ Dyke, Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562–566 (2014).
16. TR Simões, O Vernygora, MW Caldwell, SE Pierce, Megaevolutionary dynamics and the timing of evolutionary innovation in reptiles. *Nat. Commun.* **11**, 3322 (2020).
17. RN Felice, M Randau, A Goswami, A fly in a tube: macroevolutionary expectations for integrated phenotypes. *Evolution* **72**, 2580–2594 (2018).
18. A Goswami, et al., Developmental origin underlies evolutionary rate variation across the placental skull. *Philos. Transactions Royal Soc. B* **378**, 20220083 (2023).
19. AS Yang, Modularity, evolvability, and adaptive radiations: a comparison of the hemi-and holometabolous insects. *Evol. & development* **3**, 59–72 (2001).
20. S Tarasov, I Mikó, MJ Yoder, JC Uyeda, Paramo: A pipeline for reconstructing ancestral anatomies using ontologies and stochastic mapping. *Insect Syst. Divers.* **3**, 1 (2019).
21. DS Porto, J Uyeda, I Mikó, S Tarasov, ontophylo: Reconstructing the evolutionary dynamics of phenomes using new ontology-informed phylogenetic methods. *Methods Ecol. Evol.* **15**, 290–300 (2024).
22. AE Thessen, et al., Transforming the study of organisms: Phenomic data models and knowledge bases. *PLoS computational biology* **16**, e1008376 (2020).
23. JP Huelsenbeck, R Nielsen, JP Bollback, Stochastic mapping of morphological characters. *Syst. biology* **52**, 131–158 (2003).
24. MJ Sharkey, et al., Phylogenetic relationships among superfamilies of hymenoptera. *Cladistics* **28**, 80–112 (2012).
25. F Ronquist, et al., A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst. biology* **61**, 973–999 (2012).
26. J Baker, A Meade, M Pagel, C Venditti, Positive phenotypic selection inferred from phylogenies. *Biol. J. Linnean Soc.* **118**, 95–115 (2016).
27. CC Labandeira, The fossil record of insect extinction: new approaches and future directions. *Am. Entomol.* **51**, 14–29 (2005).
28. B Ebenman, Evolution in organisms that change their niches during the life cycle. *The Am. Nat.* **139**, 990–1021 (1992).
29. MR Pie, JS Weitz, A null model of morphospace occupation. *The Am. Nat.* **166**, E1–E13 (2005).
30. S Gavrilets, Dynamics of clade diversification on the morphological hypercube. *Proc. Royal Soc. London. Ser. B: Biol. Sci.* **266**, 817–824 (1999).
31. A Polaszek, L Vilhemsen, Biodiversity of hymenopteran parasitoids. *Curr. Opin. Insect Sci.* **56**, 101026 (2023).
32. L Vilhelmsen, Before the wasp-waist: comparative anatomy and phylogenetic implications of the skeleto-musculature of the thoraco-abdominal boundary region in basal hymenoptera (insecta). *Zoomorphology* **119**, 185–221 (2000).
33. L Vilhelmsen, I Miko, L Krogmann, Beyond the wasp-waist: structural diversity and phylogenetic significance of the mesosoma in apocritan wasps (insecta: Hymenoptera). *Zool. J. Linnean Soc.* **159**, 22–194 (2010).
34. D Grimaldi, MS Engel, *Evolution of the Insects*. (Cambridge University Press), (2005).
35. JC Uyeda, R Zenil-Ferguson, MW Pennell, Rethinking phylogenetic comparative methods. *Syst. Biol.* **67**, 1091–1109 (2018).
36. G Morinaga, JJ Wiens, DS Moen, The radiation continuum and the evolution of frog diversity. *Nat. Commun.* **14**, 7100 (2023).
37. S Tarasov, Integration of anatomy ontologies and evo-devo using structured markov models suggests a new framework for modeling discrete phenotypic traits. *Syst. biology* **68**, 698–716 (2019).
38. S Tarasov, New phylogenetic markov models for inapplicable morphological characters. *Syst. biology* **72**, 681–693 (2023).
39. DS Porto, et al., rphenoscate: An r package for semantics-aware evolutionary analyses of anatomical traits. *Methods Ecol. Evol.* **14**, 2531–2540 (2023).
40. MJ Yoder, I Mikó, KC Seltmann, MA Bertone, AR Deans, A gross anatomy ontology for hymenoptera. *PloS one* **5**, e15991 (2010).
41. S Tarasov, I Mikó, MJ Yoder, ontofast: An r package for interactive and semi-automatic annotation of characters with biological ontologies. *Methods Ecol. Evol.* **13**, 324–329 (2022).
42. S Tarasov, The invariant nature of a morphological character and character state: insights from gene regulatory networks. *Syst. biology* **69**, 392–400 (2020).
43. JM Beaulieu, BC O'Meara, MJ Donoghue, Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Syst. biology* **62**, 725–737 (2013).
44. TR Simões, N Greifer, J Barido-Sottani, SE Pierce, Evophylo: An r package for pre-and postprocessing of morphological data from relaxed clock bayesian phylogenetics. *Methods Ecol. Evol.* **14**, 1981–1993 (2023).
45. C Parins-Fukuchi, GW Stull, SA Smith, Phylogenomic conflict coincides with rapid morphological innovation. *Proc. Natl. Acad. Sci.* **118**, e2023058118 (2021).
46. Y Asar, H Sauquet, SY Ho, Evaluating the accuracy of methods for detecting correlated rates of molecular and morphological evolution. *Syst. Biol.* **72**, 1337–1356 (2023).
47. LJ Revell, phytools 2.0: an updated r ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**, e16505 (2024).

DRAFT

Supporting Information Text

Extended Materials and Methods

Overview of analytical strategy. The general framework adopted in the present study was based on the comparison between two sets of analyses. The first set, hereafter referred to as the "empirical analyses", included all analyses based on stochastic mapping obtained with evolutionary models conditioned on the character states observed at the tips of the phylogeny. The second set, hereafter referred to as the "null analyses", included all analyses based on stochastic mapping unconditioned on character-state data. The rationale for the comparison was to assess the influence of phylogeny and evolutionary models on the interpretation of phenome evolution in Hymenoptera. Patterns shared between the empirical and null analyses are likely due to the common tree and/or models while those exclusive to the empirical analyses are likely to be driven by observed data.

The two sets of analyses included the estimation of evolutionary rates, reconstruction of morphospaces, calculation of metrics of phenome evolution (e.g., phenotypic innovation and disparity) and regime of selection, and assessment of correlations among phenotypic rates, diversification, and selection regimes. Results from empirical and null analyses were contrasted qualitatively and/or quantitatively, as explained in the relevant sections below.

Phenomic dataset. The phenomic dataset employed in this study included morphological characters of adults and larvae of Hymenoptera. Adult characters were obtained from the phylogenetic character matrix of Sharkey et al. (1). The original matrix contained 401 characters scored for 98 taxa. The following 21 taxa were removed since they do not have correspondences in the dated phylogeny used in this study: *Chalybion*, *Cales*, *Paramblynotus*, *Poecilopsilus*, *Maaminga*, *Pseudofoenus*, *Mymaromella*, *Orussobaius*, *Archaeoteleia*, *Platygaster*, *Exallonyx*, *Syntexis*, *Schlettererius*, *Atomacera*, *Runaria*, *Monoctenus*, *Notofenusa*, *Taeniogonalys*, *Dasymutilla*, *Scolia*, and *Derecyrta*. Accordingly, the following original characters were removed since they were non-informative for the taxonomic sampling of this study: CH9, CH101, CH122, CH193, CH278, and CH374. Additionally, characters CH17, CH352, CH361-362, CH369, CH384, and CH388 were removed since they included polymorphic states, which can be difficult to model properly in case of dependencies; CH391 and CH398-401 were also removed since they included male-specific or non-morphological information. The resultant adult matrix contained 383 characters scored for 77 taxa (Dataset S1). Missing data (characters coded as '?' or '-') across characters in the adult matrix ranged from 10% to 91%, and across taxa from 4% (e.g., *Evania*) to 94% (e.g., *Tenthredo*), around 30% across the entire matrix.

Larval characters were obtained from the phylogenetic character matrix of Ronquist et al. (2). From the original matrix, characters CH221-236 correspond to larval characters. These characters were extracted for the same subset of 77 taxa of the adult matrix. Character scores were obtained from the information available for congeneric species in the original matrix or manually scored by one of the authors of the present study (LV). The original characters CH222, CH225, CH231, CH232, CH233, and CH236 were removed since they were non-informative for the taxonomic sampling of this study. The original character CH228 was split into two new characters: (character 1) Larval thoracic legs: (0) absent, (1) present; (character 2) Larval thoracic legs: (0) fully developed, (1) reduced, (2) vestigial. The resultant larval matrix contained 11 characters scored for 77 taxa (Dataset S2). Missing data across characters in the larval matrix ranged from 0% to 91%, and across taxa from 0% (e.g., *Xyela*) to 91% (most taxa), around 75% across the entire matrix.

Character coding and dependencies. Morphological characters usually refer to distinct anatomical entities (e.g., antenna or fore wing) or different properties of the same anatomical entity (e.g., shape or color of the antenna). Two characters are said to be independent from each other when the state observed in one character does not influence the probability of observing a given state in another character. For example, the shape of the antenna (e.g., states: filiform or serrate), in principle, does not influence its color (e.g., states: black or yellow). Dependent characters, however, do influence each other. The absence or presence of the antenna should influence any possible property it may have; if the antenna is absent, then no shape or color can be observed. As discussed elsewhere (3–5), dependencies are common in morphological matrices and should be modeled properly by combining dependent characters into composite characters to avoid misleading results in analyses of ancestral state reconstruction (ASR).

Characters from the adult matrix were classified into six groups according to the types of dependencies observed (Dataset S3, *G1-6*).

G1 included all non-dependent characters (306 in total). This group of characters can be analyzed without further processing. For example, CH50: Glossal fringe: (0) absent or (1) present; and CH71: Shape of lacinia: (0) rounded or (1) acuminate.

G2 included all characters with simple non-hierarchical dependencies (34 in total): a single binary character controlling a single dependent character (binary or with more states). Each pair of dependent characters was combined into a composite character following the coding scheme appropriate for the embedded dependency Markov model of the qualitative type (ED-ql) as discussed in (4). For example, CH18: Occipital carina: (0) absent or (1) present; and CH19: Occipital carina configuration: (0) reaching hypostome ventrally, (1) not reaching hypostome and not continuous medially, (2) continuous ventrally of occipital foramen. These can be combined into a composite character CH18+19 with the states (0) absent, (1) present, with carina reaching hypostome ventrally, and so on.

G3 included characters with complex non-hierarchical dependencies (12 in total): a single binary character controlling two binary dependent characters. In this case, for each triple of characters, the two binary dependent characters are independent from each other but each individually dependent on the controlling character. They were first combined following the coding scheme for a structured Markov model for independent characters as discussed in (3, 4). The resultant composite character was then combined with the controlling character using the coding scheme for ED-ql models as described above. For example,

CH64: Third maxillary palpomere: (0) absent or (1) present; CH69: Sensillum on third maxillary palpomere: (0) absent or (1) present; and CH70: Third maxillary palpomere: (0) cylindrical, (1) enlarged broader than the following palpomere and triangular. CH69 and CH70 are controlled by CH64. Thus, the former two can be first combined into a composite character CH69+70 with the states (0) sensillum absent and palpomere cylindrical, (1) sensillum absent and palpomere enlarged, and so on. Then all three characters can be combined into a final composite character CH64+CH69+CH70 with the states (0) palpomere absent, (1) palpomere present, with sensillum absent and cylindrical shape, and so on.

G4 included controlling characters of groups with multiple non-hierarchical dependencies (4 in total) or non-binary controlling characters (3 in total). For example, the controlling character CH25 has several dependent characters (CH23-24, CH26, CH32-35) and CH227 has more than two states. Characters from this group were removed either because combining groups with several dependent states would result in models with too many rate parameters to be properly estimated or because there are no models available to properly account for non-binary controlling characters. The removal of characters from *G4* resulted in the 'release' of several dependent characters (see comments below).

G5 included all dependent characters 'released' from the controlling characters of *G4* (21 in total). In this case, inapplicable states (coded as '-') were recoded as missing information (coded as '?') and then treated as non-dependent characters (as those from *G1*).

Finally, *G6* included characters with complex hierarchical dependencies (3 in total). These include only one group comprising a chain of dependencies between characters CH199, CH200, and CH201. These were also removed from the final adult matrix.

Likewise, some larval characters showed dependencies, comprising the following groups: CH1-2 and CH6-8 (i.e., character ids in reference to the larval matrix) (Dataset S4). Characters from the first group received the same treatment as the ones from *G1* and those from the second as the ones from *G3*.

Character annotations. Organismal anatomy is a hierarchy of nested sets of anatomical entities (=morphological structures). In this regard, anatomical entities can be grouped based on the anatomical regions they are part of. For example, in the Hymenoptera's (and other insects') anatomy, the anatomical entities cranium, labrum, mandible, maxilla, labium, and antenna are all part of the body region head. Biological knowledge about the hierarchy of anatomical entities of organisms can be formalized using anatomy ontologies—structured controlled vocabularies of anatomical concepts and their relationships in a given model organism or taxonomic group (6, 7). Through mathematical logic, anatomy ontologies can formalize, for example, that the anatomical concept 'mandible' has a relationship of the type '*part_of*' to the anatomical concept 'head'. Thus, every instance of the concept 'mandible', the morphological structure referred to as mandible in real biological specimens, is part of some 'head'. This may seem trivial piece of knowledge for experienced insect morphologists, but it is crucial for translating anatomical data to a machine-readable format.

By annotating the anatomical entities referred to in morphological characters with terms from an anatomy ontology (anatomical concepts), the knowledge about the hierarchy of morphological structures can be automatically retrieved and sets of characters grouped based on anatomical complexes or body regions. The new comparative methods employed in this study utilize the information from ontology annotations to cluster characters based on the anatomical hierarchy. Characters from the adult matrix were annotated with terms from the Hymenoptera Anatomy Ontology (HAO) (8) using *ontoFAST* (9). A complete list of the ontology annotations is also available in Dataset S3.

In the present study, most phylogenetic characters refer to single anatomical entities. For example, CH75: Pronotum, shape anteriorly: protruding or reduced, can be annotated with the HAO term 'pronotum' (`HAO_0001565`). In these cases, characters were annotated by simply selecting the most appropriate terms from the HAO based on term definitions and prior expert knowledge (IM, LV, and DSP). Terms were selected trying to keep a good balance between granularity (i.e., more general or more specific terms) and number of character annotations. More granular annotations (specific terms) allow investigating several anatomical entities, but each resultant anatomical cluster will include few characters. Less granular annotations (general terms) allow investigating few anatomical entities, but with more characters in each cluster. All annotations resulting in clusters with less than five characters were replaced by more general terms.

In these regards, a few characters were annotated to alternative general terms. Characters CH91-94 refer to the prepectus and CH96-98 to the anterior thoracic spiracle. The specific terms 'prepectus' and 'anterior thoracic spiracle' are both available in HAO (`HAO_0000811` and `HAO_0000582` respectively), but they result in very small clusters of characters. The HAO defines prepectus as "the intersegmentalia that is located on the mesopectus-pronotum intersegmental membrane" and the anterior thoracic spiracle as "the spiracle that is located on the border of the pronotum and mesopleuron". Neither can be directly associated with a particular anatomical entity compatible with the annotations already used at the adopted lowest level of the anatomical hierarchy. Thus, characters CH91-94 and CH96-98 were arbitrary annotated with the term 'pronotum' (`HAO_0001565`). A similar situation occurred for characters CH174-175 (refers to the mesobasalare), CH186 (refers to the prospinasternal apodeme), CH210 (refers to the hind wing tegula), and CH217 (refers to the metabasalare). CH174-175 and CH186 were annotated with the term 'mesopectus' (`HAO_0000557`), CH210 with the term 'hind wing' (`HAO_0000400`), and CH217 with the term 'metapectal-propodeal complex' (`HAO_0000604`).

Finally, some characters explicitly refer to more than one anatomical entity, as is the case for most muscle characters (CH197, CH289-351). In these cases, a primary annotation referring to the muscle origin (Dataset S3, column *onto_id1*) and a secondary annotation referring to its insertion (Dataset S3, column *onto_id2*) were provided. Nonetheless, since the site of muscle origin is more likely to move from one sclerite to another than the site of insertion, the latter was adopted as the final annotation employed to guide the comparative analyses (Dataset S3, column *onto_id*), with a few exceptions being characters explicitly stating the muscle origins as the main focus.

Larval characters were not annotated with specific ontology terms and were all treated as a single cluster.

Time-calibrated tree. ASR analyses utilized a dated phylogeny modified from topology C1 of Blaimer et al. (10). The original tree contained 765 hymenopteran species and six outgroup taxa. This tree was pruned to match the information available for adult (Dataset S1) and larval characters (Dataset S2) resulting in a final sampling of 77 taxa (Dataset S5). From the phylogeny of Blaimer et al. (10), correspondences at the generic level were found for 55 taxa in the original matrix of Sharkey et al (1); at the subfamilial level, for 12 taxa (i.e., different genera representing the same subfamily); and at the familial level, for 10 taxa (i.e., different genera and subfamily representing the same major clade in a family).

Phenomic analyses. Phenome evolution was assessed using the *ontophylo* workflow (7), a new set of comparative methods developed for assembling ASR of multiple individual characters and analyzing their combined character histories. The workflow integrates the *PARAMO* pipeline (11) with new methods and tools to estimate evolutionary rates and reconstruct the evolutionary dynamics of groups of characters representing individual anatomical entities, morphological complexes, body regions, or the whole phenome of organisms. The workflow consists of the three main steps of the *PARAMO* pipeline, as summarized below (also described in (7, 11)), and the additional *ontophylo* analyses, summarized in the next sections ('evolutionary rate estimation' and 'morphospace reconstruction').

Step 1 comprises character coding and annotation with ontology terms, as described in detail in the previous sections 'character coding and dependencies' and 'character annotations'.

Step 2 comprises model fitting and inference of individual evolutionary histories for every character (i.e., adult and larval matrices) through ASR analyses.

Model fitting and ASR were performed using functions implemented in the R package *corHMM* (12). Standard models (i.e., equal rates [ER], symmetrical [SYM], and all rates different [ARD]) were fitted for the non-dependent adult characters pertaining to groups *G1* and *G5*, and larval characters CH3-5, CH9-11. Custom models accounting for dependencies (e.g., embedded dependency models [ED]) were fitted for the sets of dependent adult characters in groups *G2* and *G3*, and larval characters CH1-2, CH6-8. For the dependent characters, transition rate matrices (Q matrices) were initially set using functions from the R package *rphenoscate* (5) and then parameters estimated with *corHMM*. ASR analyses employed stochastic character mapping (13) as implemented through the function 'makeSimmap' adopting the best fit models for each character and equal initial root probabilities for all states. For each character, 100 samples of stochastic maps were taken.

To ensure characters with poorly estimated rates do not affect downstream analyses, a quality control was introduced before moving to the next step. For adult characters, mean rate approximations were calculated for each sampled stochastic map for each character by dividing the total number of observed changes by the total sum of branch lengths (Fig. S1, purple dots). Taking a conservative approach, all characters with mean rates above 10× the upper limit of the interquartile range (IQR), calculated across all characters in the adult matrix, were flagged as outliers. This resulted in the removal of eight characters (Fig. S1A, characters above red dashed line): CH30, CH68, CH80, CH191, CH212+213, CH299, CH364, and CH377. Additionally, mean rates were also calculated directly from the Q matrices estimated with *corHMM* using the values of the main diagonal (see R scripts in the Supporting Data). Outlier characters detected from *corHMM* rates based on the same IQR threshold matched exactly the ones mentioned above (see R scripts in the Supporting Data). For the remaining characters, the *corHMM* rates were plotted against the approximations calculated from the observed changes in stochastic maps to check if results were consistent (Fig. S1B, red dots). After combining the sets of dependent characters (step 1) and removing outliers, the final adult matrix contained 340 characters for downstream analyses.

For larval characters, *corHMM* rates were used to detect outliers, but adopting a lower threshold of 1.5× the IQR. There is considerably more missing information in the larval matrix (∼75%), thus greater chances of obtaining misleading rate estimates due to lack of data. Two outlier characters were detected and removed: CH1+2 and CH11. After combining dependent characters (step 1) and removing outliers, the final larval matrix contained only 6 characters for downstream analyses.

Step 3 comprises character amalgamation (i.e., combination or clustering of characters). For thoroughly assessing phenome evolution, adult characters were clustered at three different levels of the anatomical hierarchy (Table S1): (*L1*) anatomical entities, (*L2*) body regions, and (*L3*) the whole (adult) phenome.

L1 includes 15 anatomical partitions based on the primary character annotations (Dataset S3): P1: cranium (18 chars), P2: mouthparts, (46 chars) P3: pronotum (25 chars), P4: propectus (27 chars), P5: mesonotum (37 chars), P6: mesopectus (35 chars), P7: metanotum (9 chars), P8: metapectal-propodeal complex (65 chars), P9: metasoma (17 chars; excluding female genitalia), P10: female genitalia (6 chars), P11: fore leg (12 chars), P12: mid leg (12 chars), P3: hind leg (12 chars), P14: fore wing (9 chars), and P15: hind wing (10 chars).

L2 includes three anatomical partitions based on the main body regions of the hymenopteran apocritan anatomy: P1: head (64 chars), P2: mesosoma (253 chars), and P3: metasoma (23 chars).

L3 comprises a single partition with all 340 characters from the final adult matrix. In addition to the partition including all characters (full), two alternative partition schemes were evaluated for L3: (scl) including only characters from sclerites (Dataset S3, 280 characters tagged as 'sclerites') and (msc) including only characters from muscles (Dataset S3, 60 characters tagged as 'muscles').

Larval characters were clustered as a single partition (6 chars).

Character amalgamations were performed using functions implemented in the R package *ontophylo* (7) and a series of custom R scrips (Supporting Data). Amalgamations for the L2 level used the information about the anatomical hierarchy from the HAO (8) imported as an OBO file into R using the package *ontologyIndex* (14). The resolution parameter used in the

discretization of stochastic maps, a necessary step in the amalgamation process, was set to 500 as suggested in (7) to keep a good balance between computational time and fine-tuned temporal resolution.

Evolutionary rate estimation. *ontophylo* models rates as a phylogenetic non-Homogeneous Poisson Process (pNHPP), thus allowing heterogeneity within and across branches of the phylogeny (7). It takes stochastic maps as input data and makes minimal assumptions about rate distributions, allowing the state transitions directly observed onto tree branches and the total number of changes to inform rate estimation. Rates were estimated using the amalgamated stochastic maps obtained from the anatomical partitions defined for adult (*L1-3*) and larval characters.

Estimating rates with *ontophylo* involves a series of steps (see R scripts in the Supporting Data). First, the amalgamated stochastic maps from a given anatomical partition were processed by merging identical adjacent state bins from the discretized tree branches. Then, information on mapped states and changes were extracted from all paths, from root to tips, for all branches across the 100 samples of amalgamated stochastic maps and pooled. The bandwidth, a parameter used for the kernel density estimation (KDE) implemented in *ontophylo*, was estimated directly from the path data using the option '*bw.nrd*' of the bandwidth estimator. Finally, KDE was performed using the path data and bandwidth, normalized, smoothed, and scaled by the total number of observed changes to obtain the final rate estimates for all branches of the phylogeny.

As mentioned in (7), overall rates are influenced by partition size (i.e., number of amalgamated characters and states). To assess the effect of partition size and obtain confidence intervals, bootstrap resampling was used. Rate values from the main diagonal of the Q matrices estimated with *corHMM* were amassed to create a sampling pool for all characters in each anatomical partition in *L1* (P1-15) and *L2* (P1-3). Then, 1000 rounds of resampling with replacement was performed. In each round, N values were resampled and the mean was calculated. N was obtained from the mean partition size, calculated based on the total number of states in each partition, across all partitions in *L1* (N = 50) and L2 (N = 250). The mean rate of an individual character pertaining to that partition was then rescaled by the mean partition size and observed partition sizes in *L1* and *L2* to allow for direct comparisons.

Morphospace reconstruction. *ontophylo* reconstructs the evolutionary dynamics of amalgamated characters by first reducing the dimensionality of morphological data and representing it onto a 2D space (i.e., 'morphospace'), and then stacking morphospaces obtained from multiple time slices to produce an animation. Moreover, the fully-stacked morphospace or individual time slices can be visualized.

Morphospaces were reconstructed using the amalgamated stochastic maps obtained from the anatomical partitions defined for adult *(L1-3)* and larval characters. For each partition, one sample of amalgamated map was taken and figures/animations produced. Nonetheless, when calculating innovation and disparity metrics (see next section), all 100 stochastic maps obtained for each partition were used. Multidimensional scaling (MDS) was used for dimensionality reduction, as implemented in *ontophylo* (see R scripts in the Supporting Data).

Metrics to assess phenome evolution and regime of selection. Phenome evolution was assessed by two metrics to capture two different aspects of phenotypic change at macroevolutionary scale.

The first aspect was the phenotypic innovation of lineages. An amalgamated character is the combination of two or more individual characters. It can be conceptualized as a vector with N positions, where each position corresponds to one of the N individual characters before amalgamation, similar to the individual bases in a DNA sequence. Along the branches of the phylogeny, character-state changes can occur at any given position of the character vector. Whenever a change occurs to a new character-state never observed before at a given position, the innovation score increases by one. The sum of all innovations across all positions of the amalgamated character accumulated in a lineage along the path from root to tip of the phylogeny is used here as a metric of innovation (see explanatory diagram in Fig. S5A).

The second aspect of phenome evolution considered here was the phenotypic disparity among lineages. At any given point in time in the phylogeny, different lineages can have different state combinations for the individual characters constituting an amalgamated character (hereafter referred to as phenomic state). For example, a phenomic state representing five individual binary characters of the 'head' can have states `01011`, `00011`, and `11111` in three different lineages present at a given point in time. It is possible to use a distance metric, for example Hamming distance, to assess how dissimilar these three phenomic states are. In the previous example, the Hamming distances between the first phenomic state to the second and the third ones is 1 and 2, respectively. The mean Hamming distance among all phenomic states observed in the lineages present at a given point in time is used here as a metric of disparity (see explanatory diagram in Fig. S5B).

To assess the regime of selection on each branch of the phylogeny, a new metric is proposed here. For each branch, the Hamming distance between the first and last observed phenomic state was used to evaluate if evolution along that path resulted in more similar or dissimilar phenotypes (e.g., shorter or longer distances in the morphospace, respectively). When comparing results from the empirical and null analyses, if distances calculated in the former are significative and longer than the ones calculated in the latter (i.e., large positive values), then directional selection can be inferred. Otherwise, if distances are significative and shorter (i.e., large negative values) stabilizing selection can be inferred. Non-significative differences (e.g., small positive or negative values) indicate diffusive evolution (see explanatory diagram in Fig. S5C).

To calculate the different metrics of phenomic evolution and selection regimes, the raw information from stochastic map samples of amalgamated characters was used. For example, for the amalgamated character representing the anatomical complex 'head', the three metrics were calculated for each one of the 100 stochastic maps sampled. The three metrics were calculated for all anatomical complexes (*L1*), body regions (*L2*), adult phenome (*L3*), and larval phenome. To facilitate calculations of

the metrics of phenomic evolution, amalgamated stochastic maps were split into discrete time bins (100 in total). For the innovation metric, innovation was calculated as the cumulative sum along each path from root to tip for each lineage and then each time bin was summarized as the mean value across all branches represented at a given time slice to obtain the step-wise curves shown in Fig. 6C-D and Fig. S21. For the disparity metric, the mean Hamming distance calculated for each bin as explained above was used to obtain the step-wise curves shown in Fig. 6E-F and Fig. S22. In both cases, values were rescaled between 0 and 1.

When contrasting empirical and null analyses, qualitative assessments were used to compare metrics of phenomic evolution resulting in the curves shown in Fig. 6C-F and Figs. S21 and S22. On the other hand, a quantitative assessment was used to evaluate the metric of regime of selection. The significance of values was assessed through bootstrap resampling (see R scripts in the Supporting Data for more details). For each branch `br_n` of the tree (where n = 1, 2, 3, ... total number of branches), a sample of 1000 `br_n` branches was taken from the pool of 100 `br_n` branches (with replacement) obtained from the stochastic maps of the empirical and null analyses (1000 branches each); the metric was calculated and the mean obtained for the empirical and null sets; and then the difference between the means was calculated. In total, 10.000 replicates were performed for each branch to obtain distributions of differences between mean values calculated in empirical and null analyses. For each branch, the significance of the metric was determined by the 95% CI of its difference distribution not overlapping with zero.

Rate correlations. By assessing branch-specific rates across multiple anatomical complexes it is possible to investigate not only the evolutionary dynamics of different portions of the phenome along the Hymenoptera phylogeny but also their potential associations among themselves, with lineage diversification rates, and selection regimes. To investigate these associations, an assessment of potential correlations among phenotypic rates and between phenotypic and lineage diversification rates was performed. Furthermore, potential correlations between our new metric for assessing regime of selection and phenotypic and lineage diversification rates were also investigated.

For investigating correlations among phenotypic rates, all pairwise comparisons between the phenotypic rates estimated for the elementary anatomical complexes of *L1* (15 anatomical partitions) plus the larval phenome were performed (16 partitions in total) resulting in 120 pairs of comparisons.

For investigating correlations between phenotypic and lineage diversification rates, all comparisons between the phenotypic rates mentioned above and lineage diversification rates extracted from the BAMM analyses of Blaimer et a. (10) were performed resulting in 16 pairs of comparisons (see R scripts in the Supporting Data for more details). Additionally, correlations between phenotypic rates for the full adult phenome (*L3*, all characters) and lineage diversification rates were also investigated resulting in one additional pair of comparison. Analyses were performed using the net diversification rates extracted from the BAMM analyses as well as only speciation and extinction rates resulting in a total of 54 comparisons.

For investigating correlations between lineage diversification rates and selection regimes, all comparisons mentioned above (16 elementary anatomical partitions, larval and full adult phenome) were performed against net diversification, speciation, and extinction rates resulting in an additional 54 comparisons. Finally, correlations between phenotypic rates and selection regimes were performed for the same anatomical partitions, larval and full adult phenomes resulting in 17 comparisons.

Initially, two strategies were explored to investigate rate correlations. The first strategy used a resampling-based approached with permutation tests to compare phenotypic rates, diversification rates, and selection regimes. The second strategy used linear regressions including rates calculated from both the empirical and null analyses. The second strategy was employed only to asses correlations among phenotypic rates and to contrast with the results from the first strategy. In each comparison, an stochastic map was randomly sampled from the pool of maps of each anatomical partition, the mean rate was calculated for each branch, and one of the two strategies mentioned above was applied. Each comparison was repeated 1000 times.

Since *ontophylo* utilizes a modified Markov kernel for density estimation allowing rates to vary along branches as well as applying smoothing to the final rates, we opted instead for using the raw data from the amalgamated stochastic maps to calculate mean branch rates. Rates were calculated simply as the number of observed character-state transitions divided by branch length. Rates were calculated individually for each branch and tree in the samples of stochastic maps.

In the first strategy, for each iteration, the Spearman's rank correlation was calculated between the branch rates of the two anatomical complexes, then a permutation test was performed by reshuffling the branches 1000 times (replicates) each time calculating a new correlation value, and finally a p-value was obtained by calculating the proportion of times the new correlations were equal to or greater than the initial observed correlation. This strategy is similar to the one used in (15, 16).

In the second strategy, for each iteration and pair being compared (e.g., anatomical complex 1 and 2), branch rates were extracted from the stochastic maps obtained in the empirical (data) analyses (e.g., D1: anatomical complex 1 and D2: anatomical complex 2) and null (simulation) analyses (e.g., S1: anatomical complex 1 and S2: anatomical complex 2). Then, a linear regression was performed (formula: $D1 \sim D2 + S1 + S2$), the p-value and the angular coefficient between D1 and D2 was obtained, and finally the angular coefficient was used to obtain a value of Pearson's correlation. The rationale for also including the rate variables of the null analyses (S1 and S2) was to remove possible confounding effects due to the shared tree and evolutionary models between the pairs. The regression was repeated for 1000 iterations for each comparison.

In summary, the application of both strategies resulted in 1000 correlation values (Spearman and Pearson) and 1000 p-values for each one of the 120 pairwise comparisons among phenotypic rates. In the case of the first strategy, resampling was also performed separately for the rates obtained from the null analyses as well. Significance of correlation values was determined by p-values equal to or below the threshold of 0.05 in 95% of the iterations or more. Only congruent results between the two strategies were discussed in the main text.

Simulation analyses. The null analyses used the same settings as the empirical analyses except for the following. Instead of performing standard stochastic mapping with evolutionary models conditioned on the character states observed at the tips of the phylogeny, character histories were simulated using the same models and parameters (i.e., Q matrices) fitted with *corHMM* for the empirical analyses, but not conditioned on the observed character states at the tips. Simulations used the function `sim.history` from the R package *phytools* (17). The null analysis outputs were used to evaluate rate evolutionary dynamics and morphospaces for all anatomical complexes, body regions, adult and larval phenomes; assess the significance of results from the metric of regime of selection; and assess the significance and strength of phenotypic rate correlations.

Extended Results

Effects of partition size on rates. When comparing individual characters (Fig. S2A), the mean rates for most anatomical complexes (*L1*) varied between 0.001∼0.002; that for the hind leg (P13) was slightly higher, around 0.002; and those for the propectus (P4), metanotum (P7), female genitalia (P10), fore- and hind wing (P14-15) were slightly lower, below 0.001. When comparing rescaled rates based on the mean partition sizes and observed partition sizes (Fig. S2B: P1-15 and S1-15 respectively), rates for most anatomical complexes varied between 0.05 and 0.2; the rates based on observed partition sizes were higher, more than 0.2, for the mouthparts (S2) and metapectal-propodeal complex (S8); and conspicuously lower, below 0.05, for the metanotum (S7), metasoma (S9), female genitalia (S10), fore-, mid-, and hind leg (S11-13), fore- and hind wing (S14-15). As for the body regions (*L2*), individual character rates (Fig. S2C) were higher for the head (0.002∼0.0015), followed by mesosoma (0.0015∼0.001), and metasoma (below 0.001). When comparing rescaled rates, those based on the mean partition sizes (Fig. S2D: P1-3) were between 0.1 and 0.5 for the three partitions. However, when based on the observed partition sizes (Fig. S2D: S1-3), the rate was conspicuously higher for the mesosoma (0.7 ∼ 0.8) and lower for the metasoma (below 0.1).

Note that for both, anatomical complexes and body regions, rescaled rates based on observed partition sizes were usually higher for the partitions including a larger number of elementary characters (e.g., mouthparts: 46 characters; metapectal-propodeal complex: 65 characters) and lower for the partitions including fewer characters (e.g., fore-, mid- and hind leg: 12 characters each). As suggested in (7) caution is advised when comparing and interpreting the absolute rates estimated with *ontophylo* across different anatomical complexes. While in some instances it seems that rates are indeed low due to underlying evolutionary processes (Fig. S2A-B: P4, P7, P10, P14-15), in others, rates seem to be higher (Fig. S2B: S2, S8) or lower (Fig. S2B: S11-13) due to differences in the number of characters (and states) being amalgamated.

Strategies for assessing rate correlations. From all 120 possible pairwise comparisons between 15 morphological complexes and the larval phenome (i.e., 16 partitions in total), 34 pairs presented significant correlations between phenotypic rates of evolution (Table S2; Pval_prop: proportion of significant p-values equal to or greater than 0.95) for the approach based on permutations (strategy 1). As for the approach based on linear regressions (strategy 2), 27 pairs presented significant correlations (Table S3). All pairs recovered from the regression approach (strategy 2) were congruent with those from the permutation approach (strategy 1). For all 27 congruent pairs, distributions of correlation coefficients estimated using both strategies (strategy 1: Spearman's rank correlation; strategy 2: corrected and uncorrected Pearson's correlation; see R scripts in the Supporting Data for more details) were almost identical (Fig. S3) with median values ranging between 0.38∼0.69 in strategy 1 (Table S2) and between 0.39∼0.70 (0.39∼0.70 if uncorrected) in strategy 2 (Table S3). Finally, when comparing correlation values obtained for the congruent pairs in the empirical and null analyses (i.e., only for strategy 1), the distributions of values sampled in the permutation tests are clearly distinct (Fig. S4).

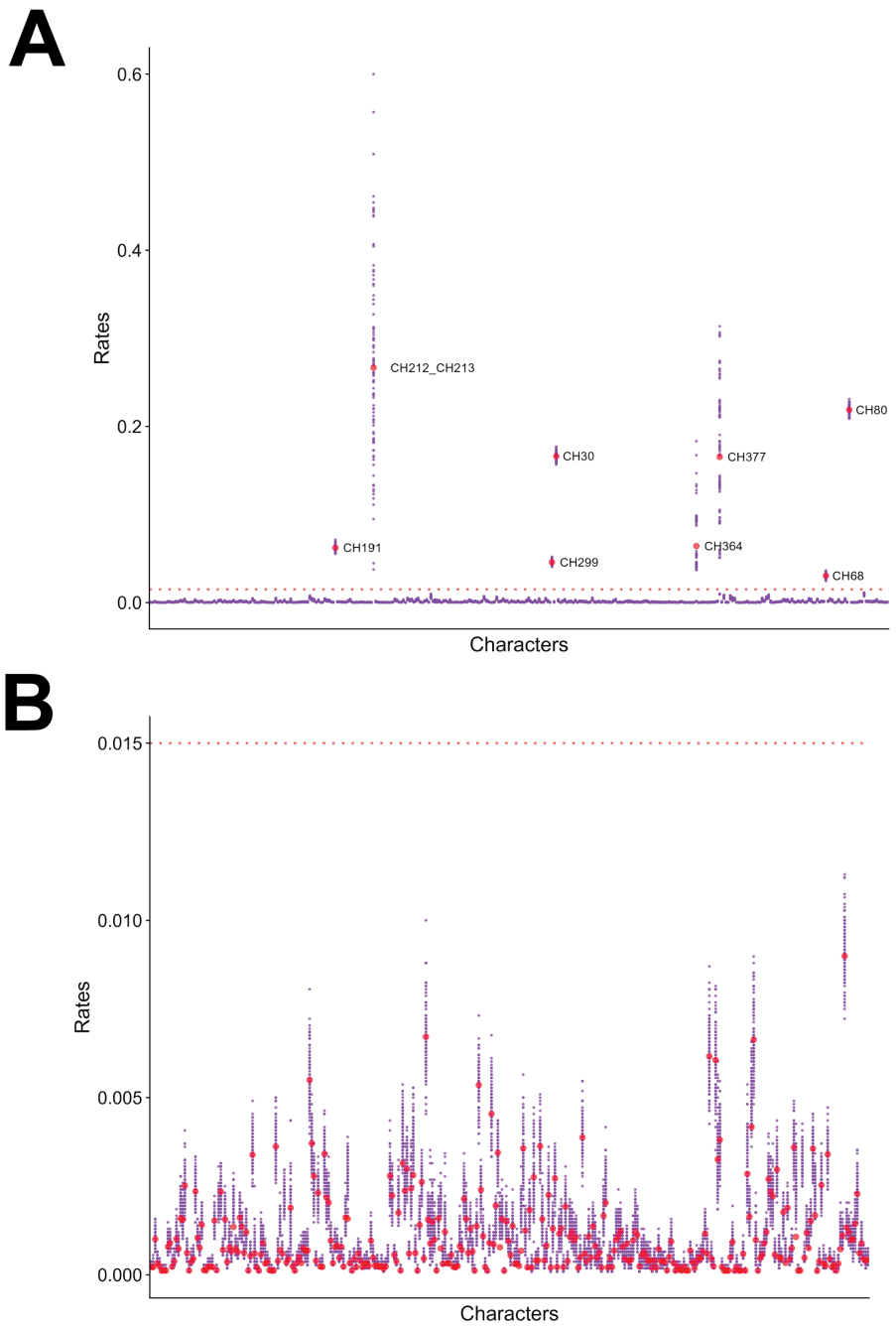

Fig. S1. Mean rates calculated from sampled stochastic maps for each individual character. *(A)* Dataset including all characters. *(B)* Dataset excluding outliers. Empirical rates calculated from observed changes on stochastic maps (purple dots). Mean rates estimated with *corHMM* (red dots; mean taken from the main diagonal of fitted Q matrices). Red dashed line indicates 10× the upper limit of interquartile range (IQR). Values above the red dashed line in (A) are putative outliers.

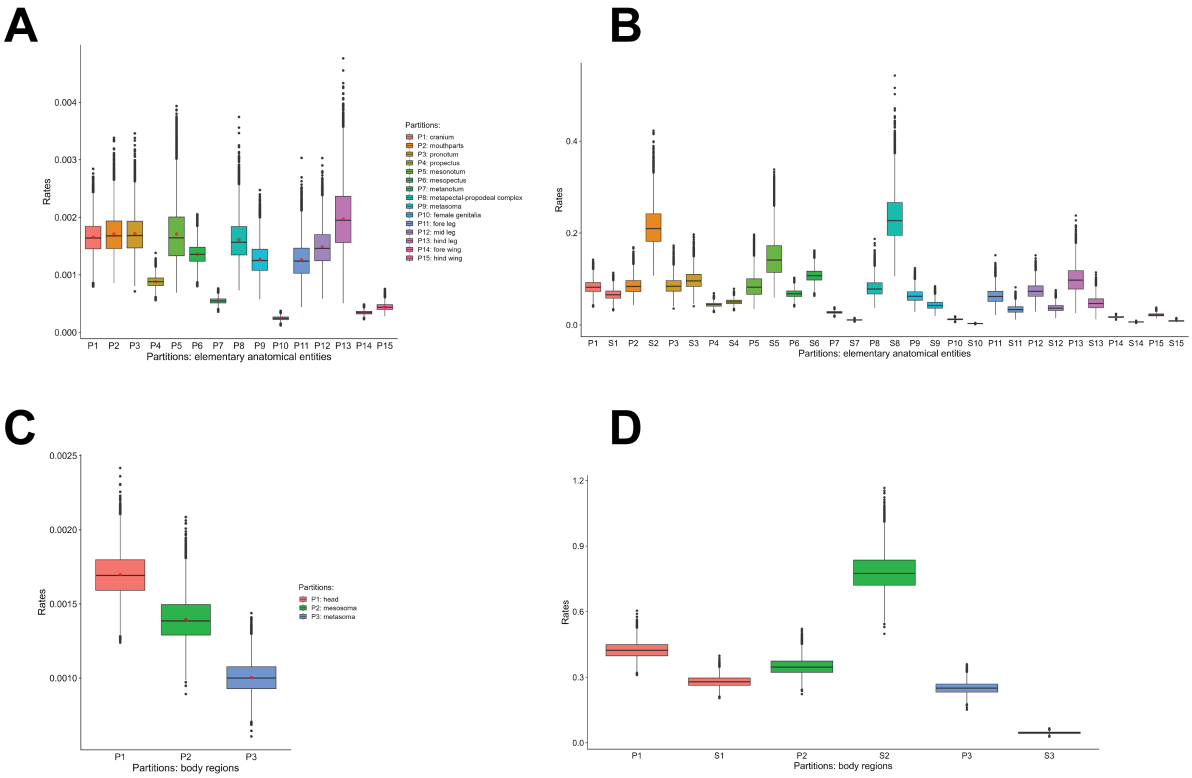

Fig. S2. Effect of partition size on estimated rates. *(A)* Mean rate of individual characters pertaining to each partition for each anatomical complex. *(B)* Mean partition rates rescaled by mean partition sizes (P1-15) and observed partition sizes (S1-15) for each anatomical complex. *(C)* Mean rate of individual characters pertaining to each partition for each body region. *(D)* Mean partition rates rescaled by mean partition sizes (P1-3) and observed partition sizes (S1-3) for each body region.

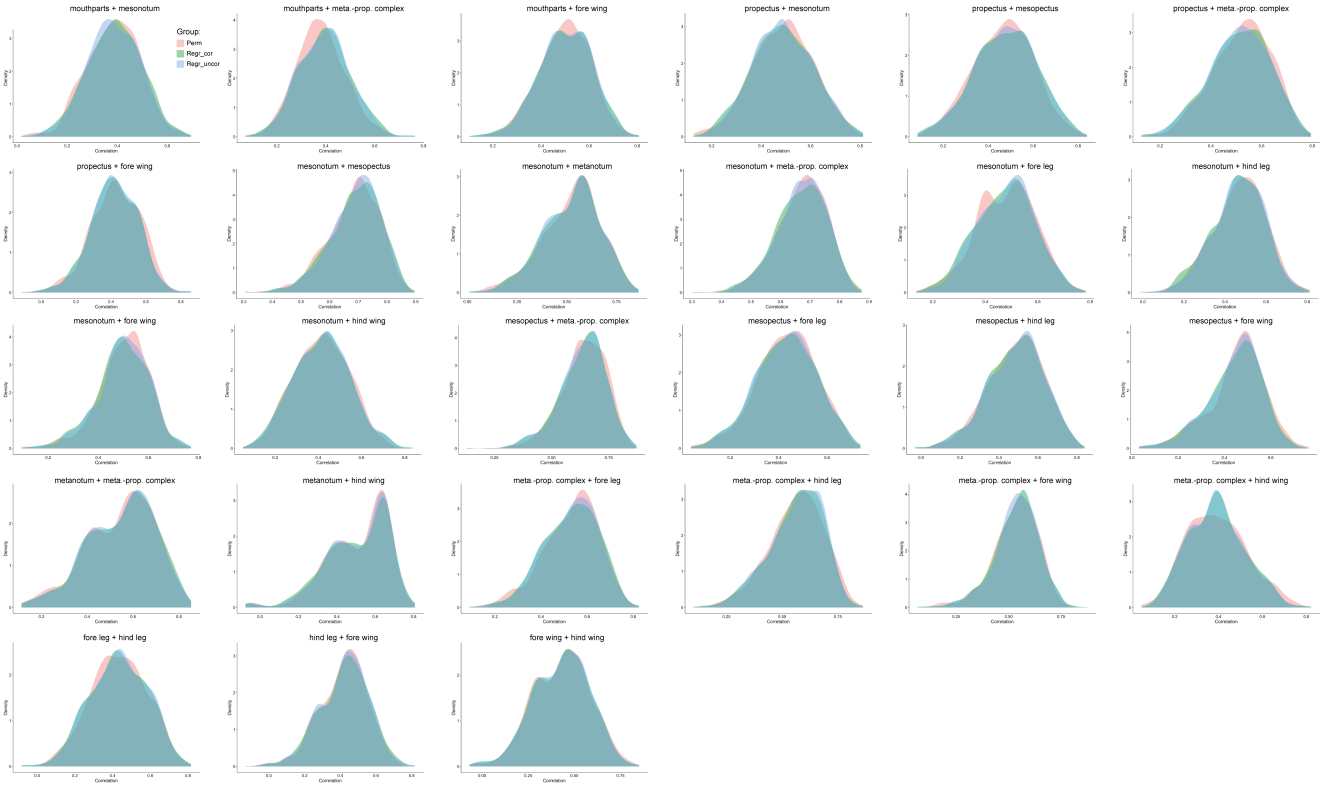

Fig. S3. Distributions of correlation coefficient values obtained through the resampling (permutation) and regression strategies for assessing correlations among the phenotypic rates of evolution. Only pairs with significant correlations are shown. Legends: Perm: permutation strategy; Regr_cor: regression strategy using both rates from the empirical and null analyses (corrected); Regr_uncor: regression strategy using only rates from the empirical analysis (uncorrected).

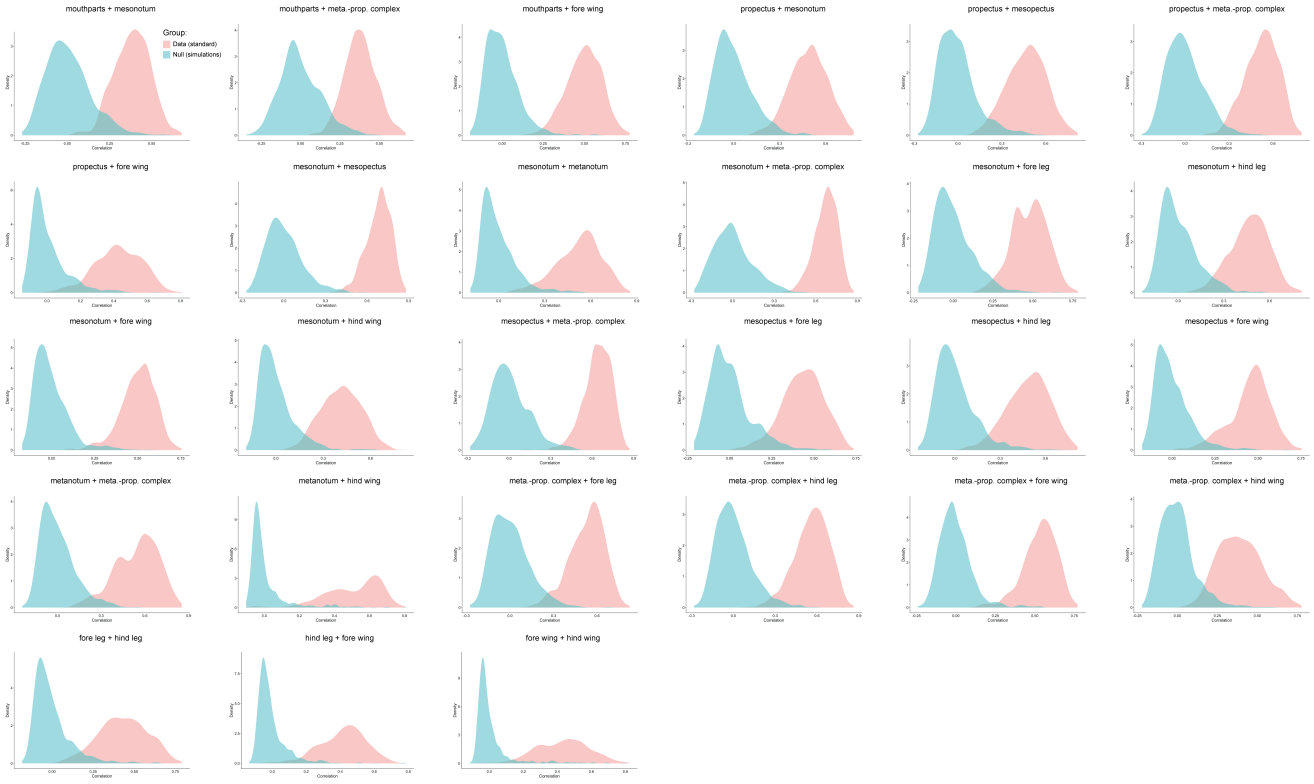

Fig. S4. Distributions of correlation coefficient values obtained through the resampling (permutation) strategy. Only pairs with significant correlations are shown.

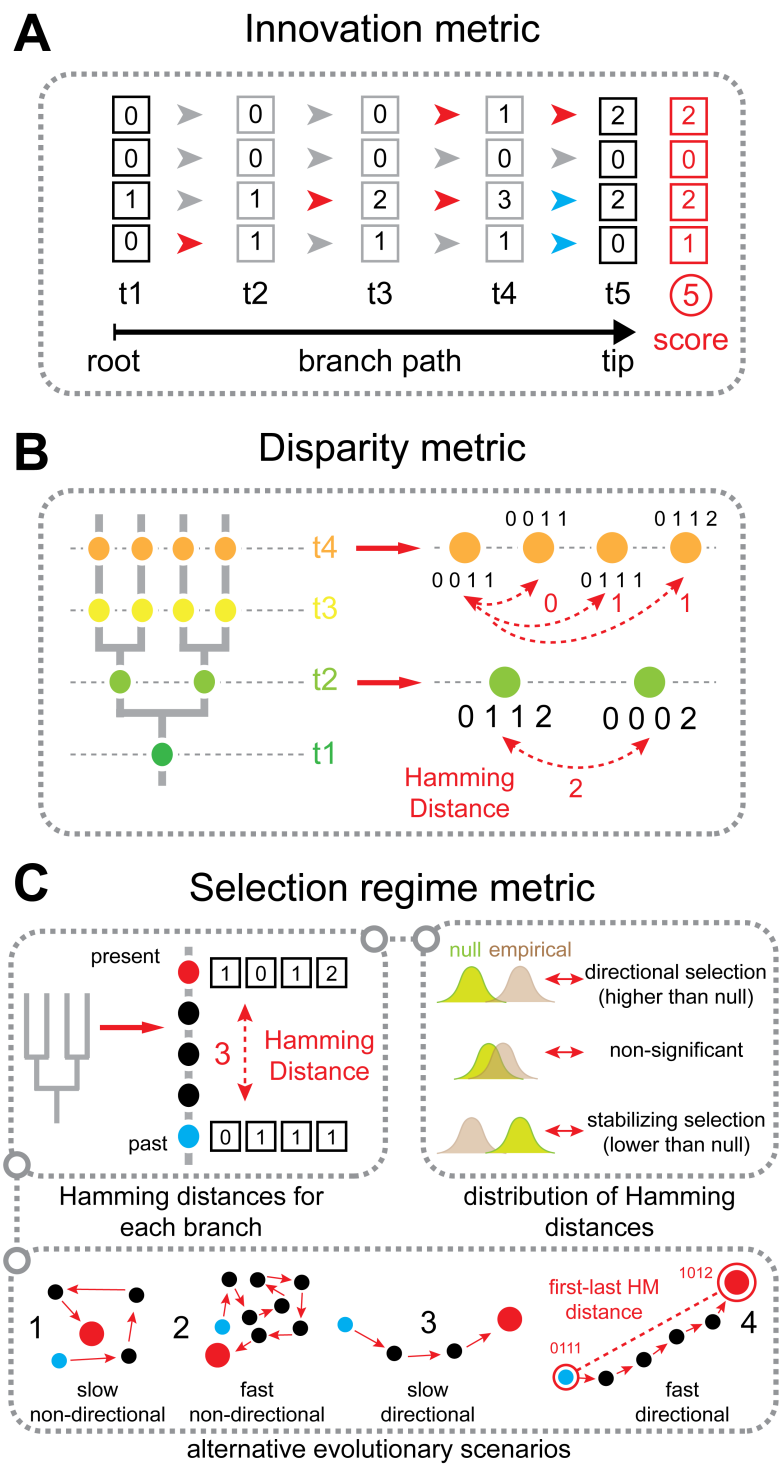

Fig. S5. Diagrams explaining the metrics to assess phenomic evolution and regime of selection. *(A)* Phenotypic innovation. *(B)* Phenotypic disparity. *(C)* Regime of selection. Each column (in (A)) or row (in (C)) indicates an amalgamated character (group of boxes); each individual box (gray or black outlines) indicates an individual character; for instance, the example shows the amalgamation of four individual characters. In (A), gray arrowhead indicates no character-state transitions; red arrowhead indicates transitions to new unique character states never observed before along the path from root-to-tip for a given lineage; blue arrowhead indicates transitions to character states already observed (i.e., does not count for the innovation score); each box with a red outline indicates the innovation score for an individual character (i.e., sum of new unique character-state transitions) and the circle indicates the total score for the amalgamated character. In (B), filled-circles with different colors indicate phenomic states (character-state combinations representing an amalgamated character) sampled at a given time slice. In (C), filled-circles with different colors indicate sequential phenomic states sampled in a given tree branch; blue and red circles indicate first and last sampled states, respectively; the bottom box shows four possible evolutionary scenarios (1-4) for phenomic walks in the morphospace and their interpretation in terms of evolutionary rates and selection regimes.

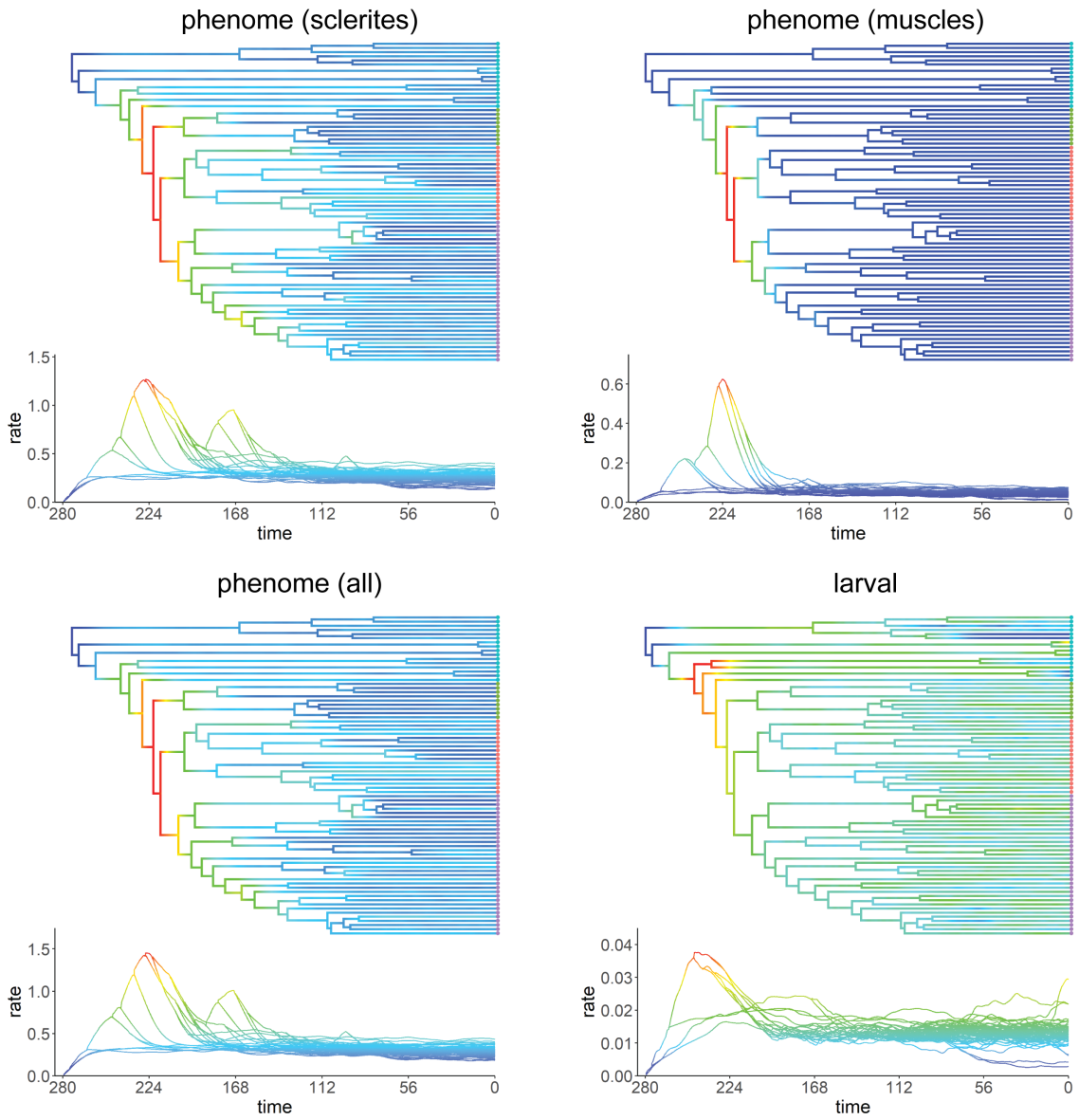

Fig. S6. Branch-specific rate dynamics inferred with *ontophylo* for the empirical analyses. Subpanels show the results for adult and larval phenome of Hymenoptera. In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rates through time for each branch (edgeplots). Color gradient from blue to red indicates low to high rates respectively. Small filled circles at the tips of the tree indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata; purple: Proctotrupomorpha.

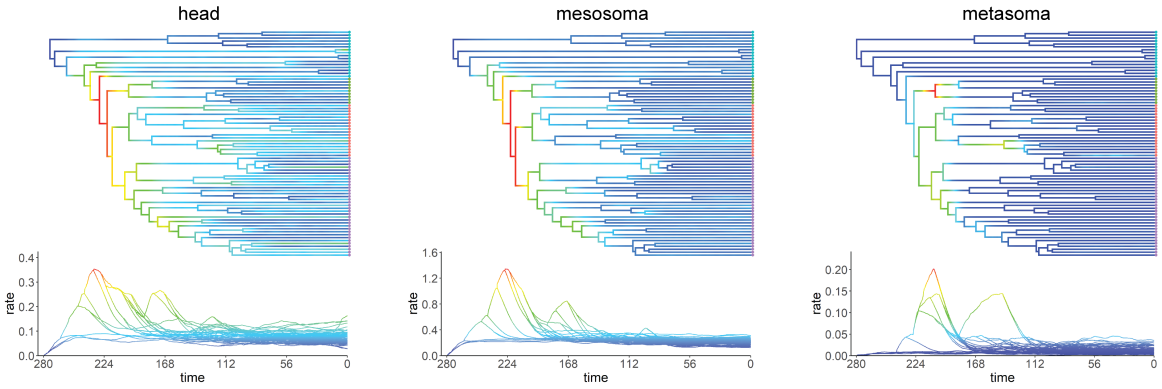

Fig. S7. Branch-specific rate dynamics inferred with *ontophylo* for the empirical analyses. Each subpanel shows the result for a body region of the Hymenoptera anatomy. In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rates through time for each branch (edgeplots). Color gradient from blue to red indicates low to high rates respectively. Small filled circles at the tips of the tree indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: “Evaniomorpha”-Aculeata; purple: Proctotrupomorpha.

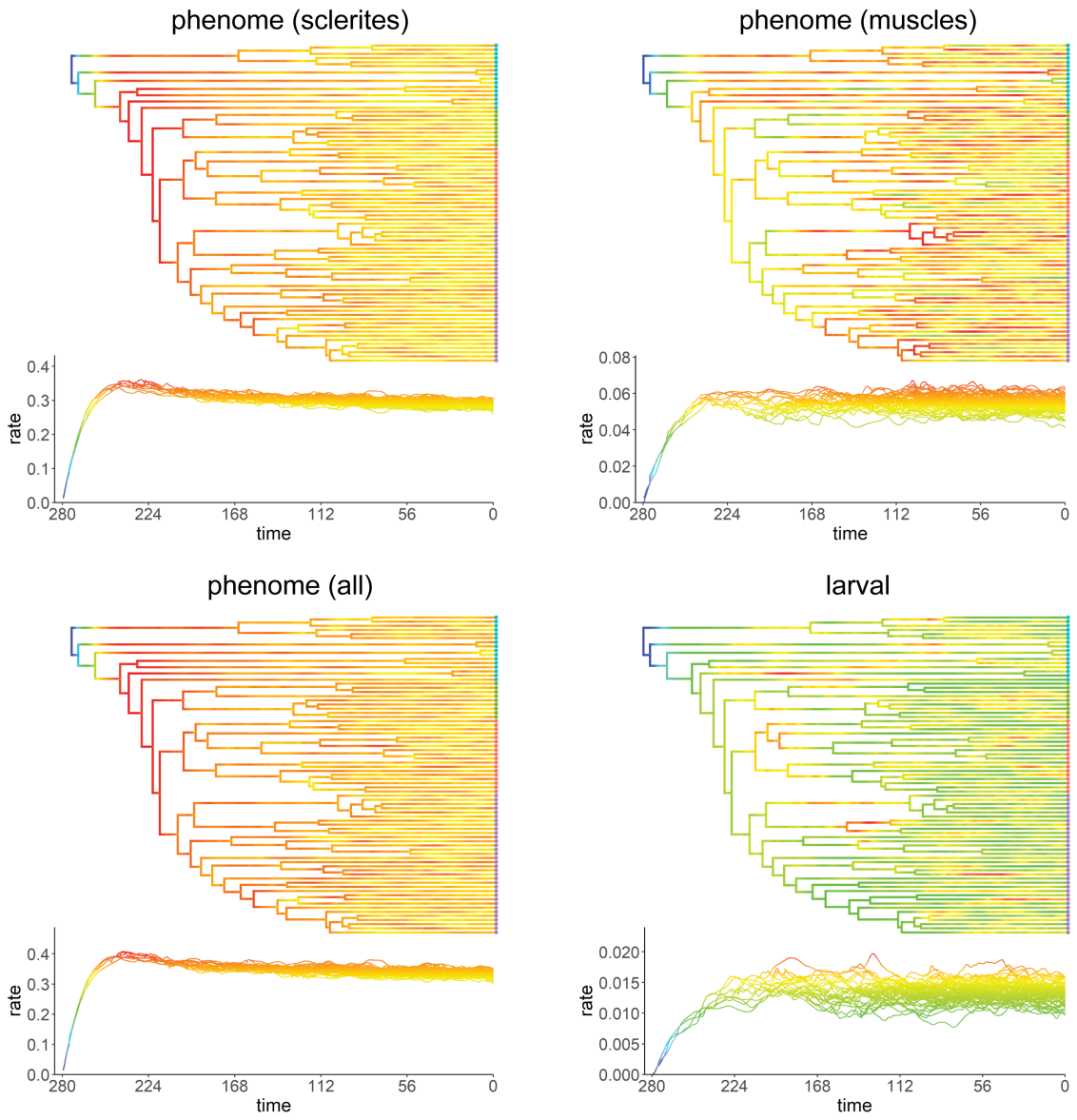

Fig. S8. Branch-specific rate dynamics inferred with *ontophylo* for the null analyses. Subpanels show the results for adult and larval phenome of Hymenoptera. In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rates through time for each branch (edgeplots). Color gradient from blue to red indicates low to high rates respectively. Small filled circles at the tips of the tree indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata; purple: Proctotrupomorpha.

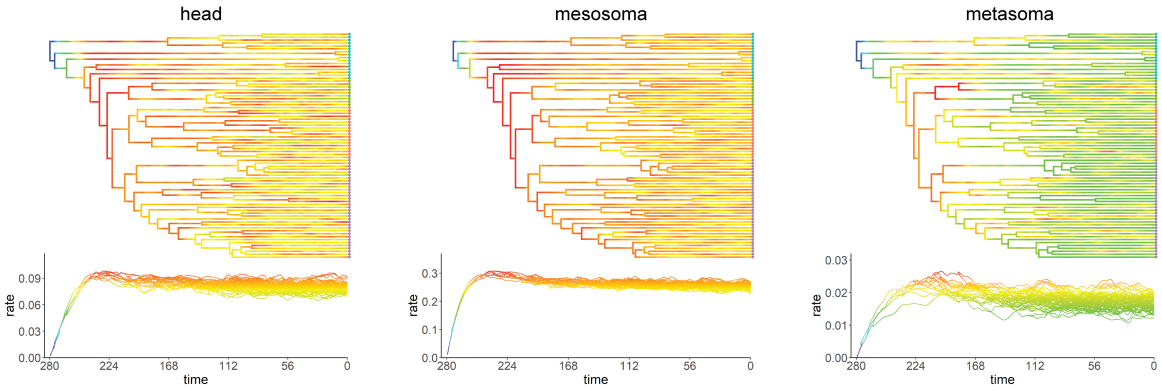

Fig. S9. Branch-specific rate dynamics inferred with *ontophylo* for the null analyses. Each subpanel shows the result for a body region of the Hymenoptera anatomy. In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rates through time for each branch (edgeplots). Color gradient from blue to red indicates low to high rates respectively. Small filled circles at the tips of the tree indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: “Evaniomorpha”-Aculeata; purple: Proctotrupomorpha.

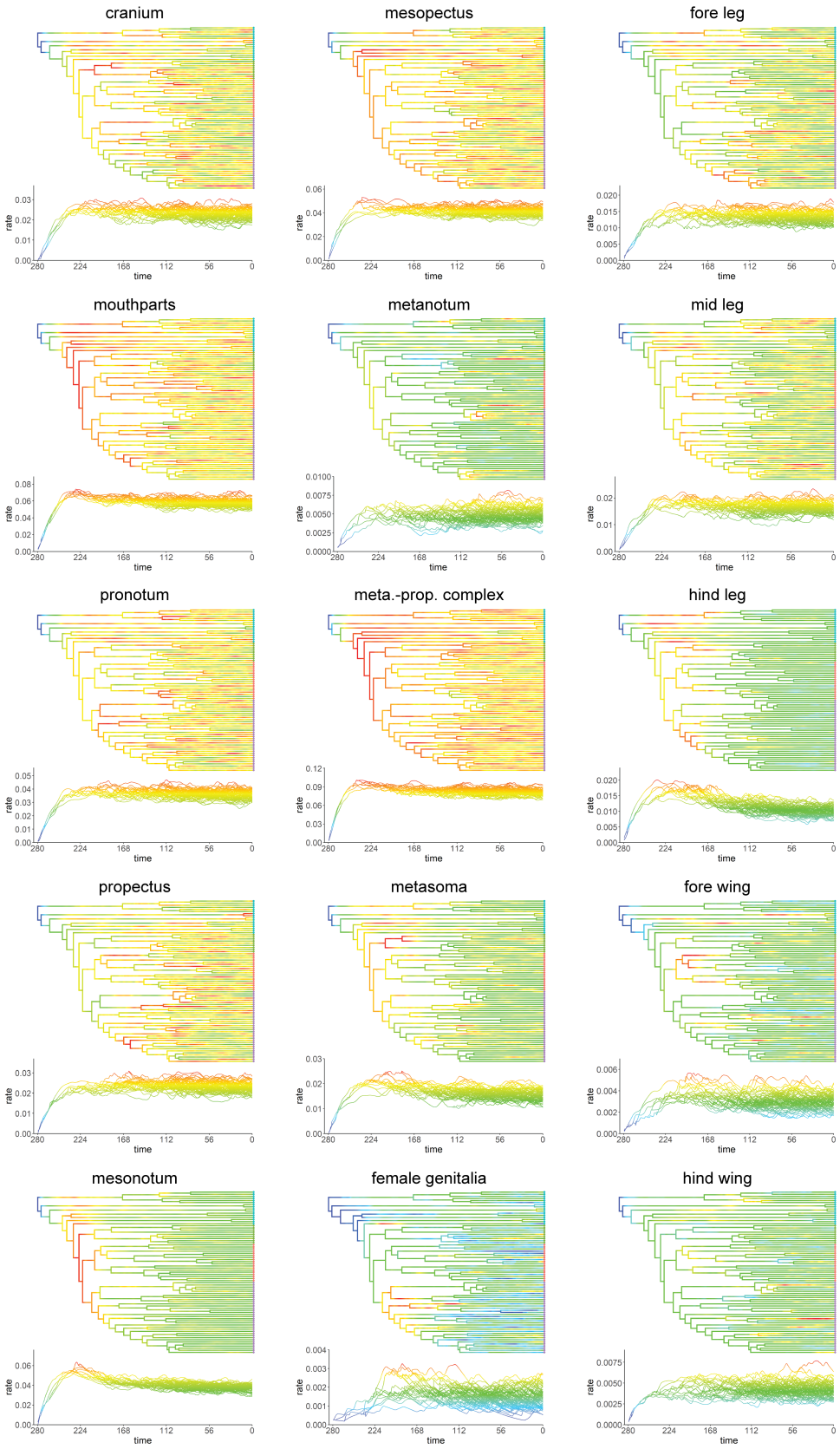

Fig. S10. Branch-specific rate dynamics inferred with *ontophylo* for the null analyses. Each subpanel shows the result for an elementary anatomical complex of the Hymenoptera anatomy. In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rates through time for each branch (edgeplots). Color gradient from blue to red indicates low to high rates respectively. Small filled circles at the tips of the tree indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: “Evaniomorpha”-Aculeata; purple: Proctotrupomorpha.

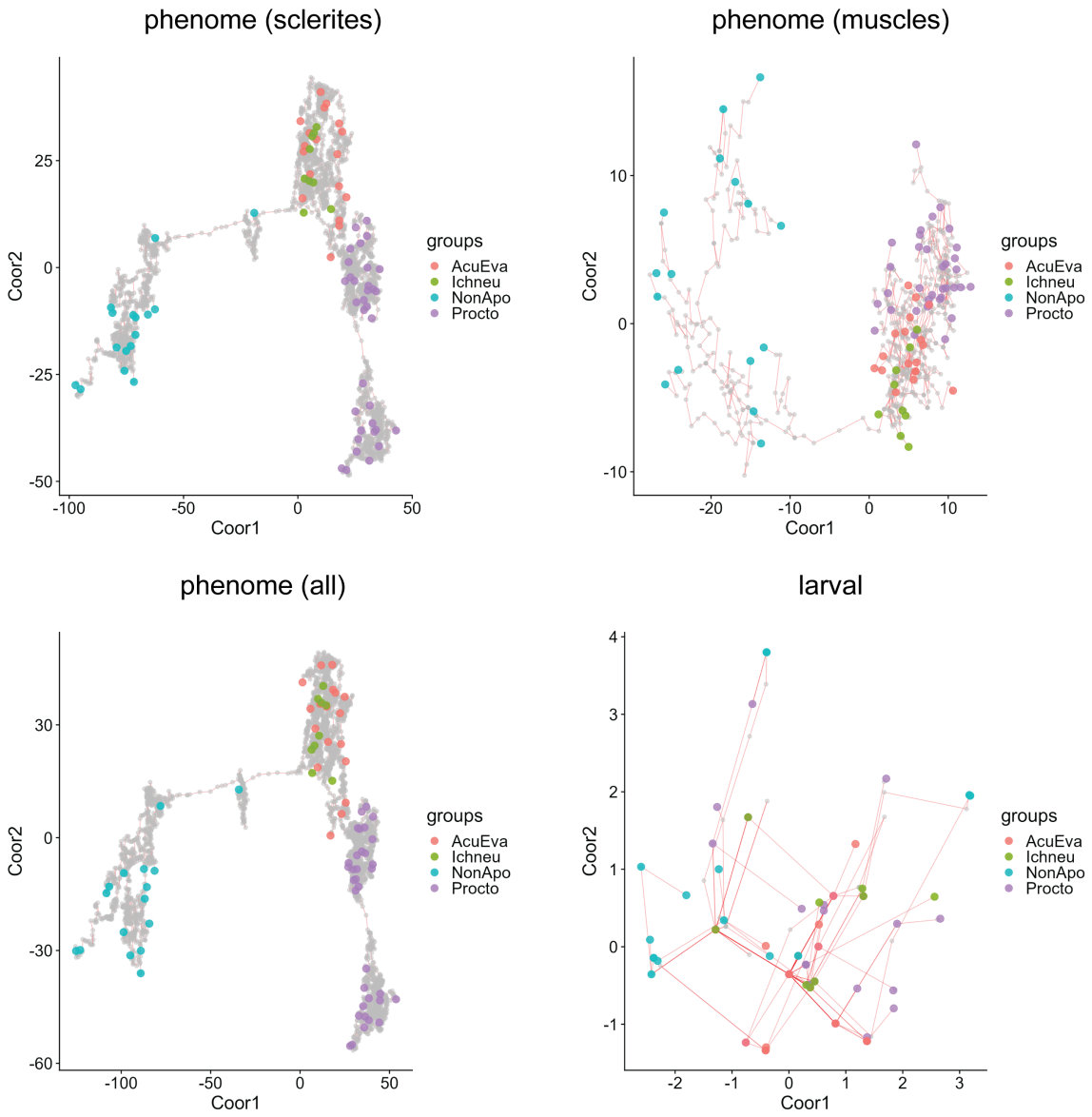

Fig. S11. Morphospaces reconstructed with *ontophylo* for the empirical analyses. Subpanels show the results for adult and larval phenome of Hymenoptera. Small, gray circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

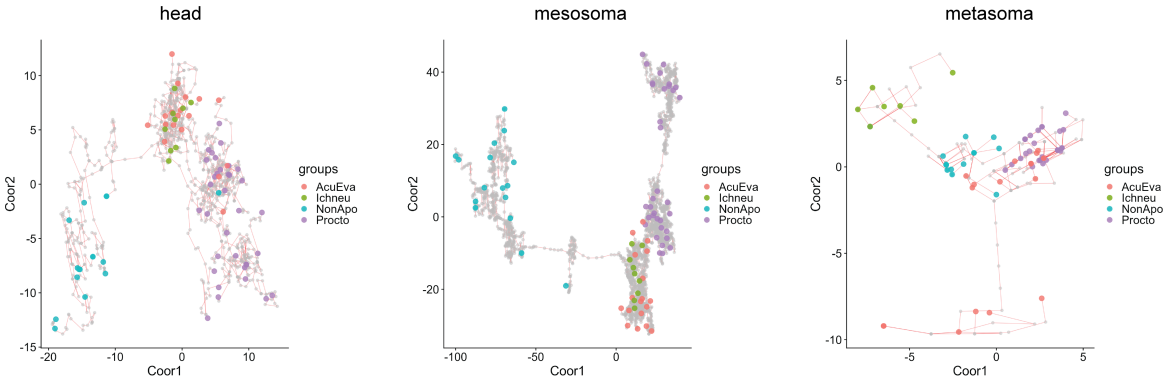

Fig. S12. Morphospaces reconstructed with *ontophylo* for the empirical analyses. Each subpanel shows the result for a body region of the Hymenoptera anatomy. Small, gray circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

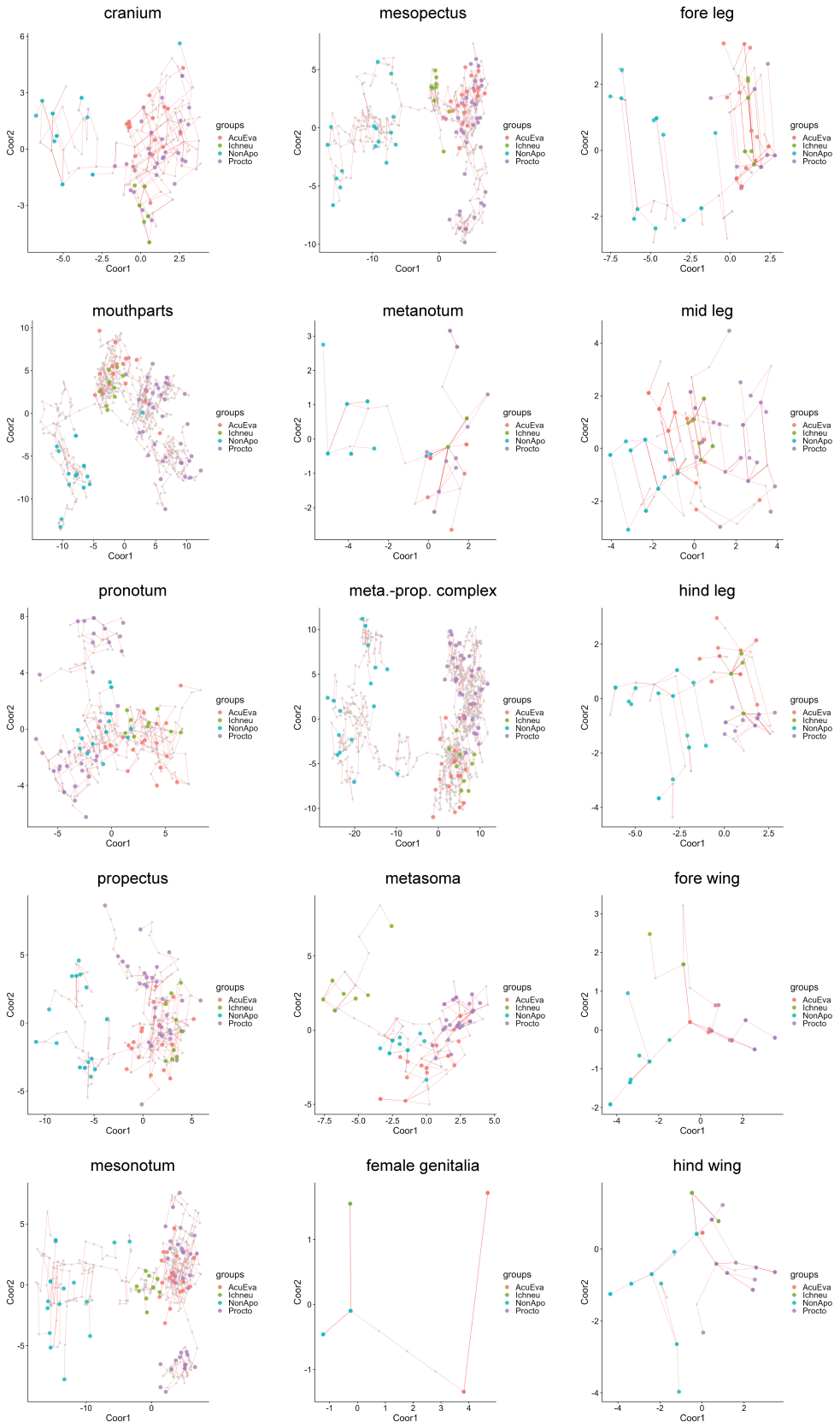

Fig. S13. Morphospaces reconstructed with *ontophylo* for the empirical analyses. Each subpanel shows the result for an elementary anatomical complex of the Hymenoptera anatomy. Small, gray circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

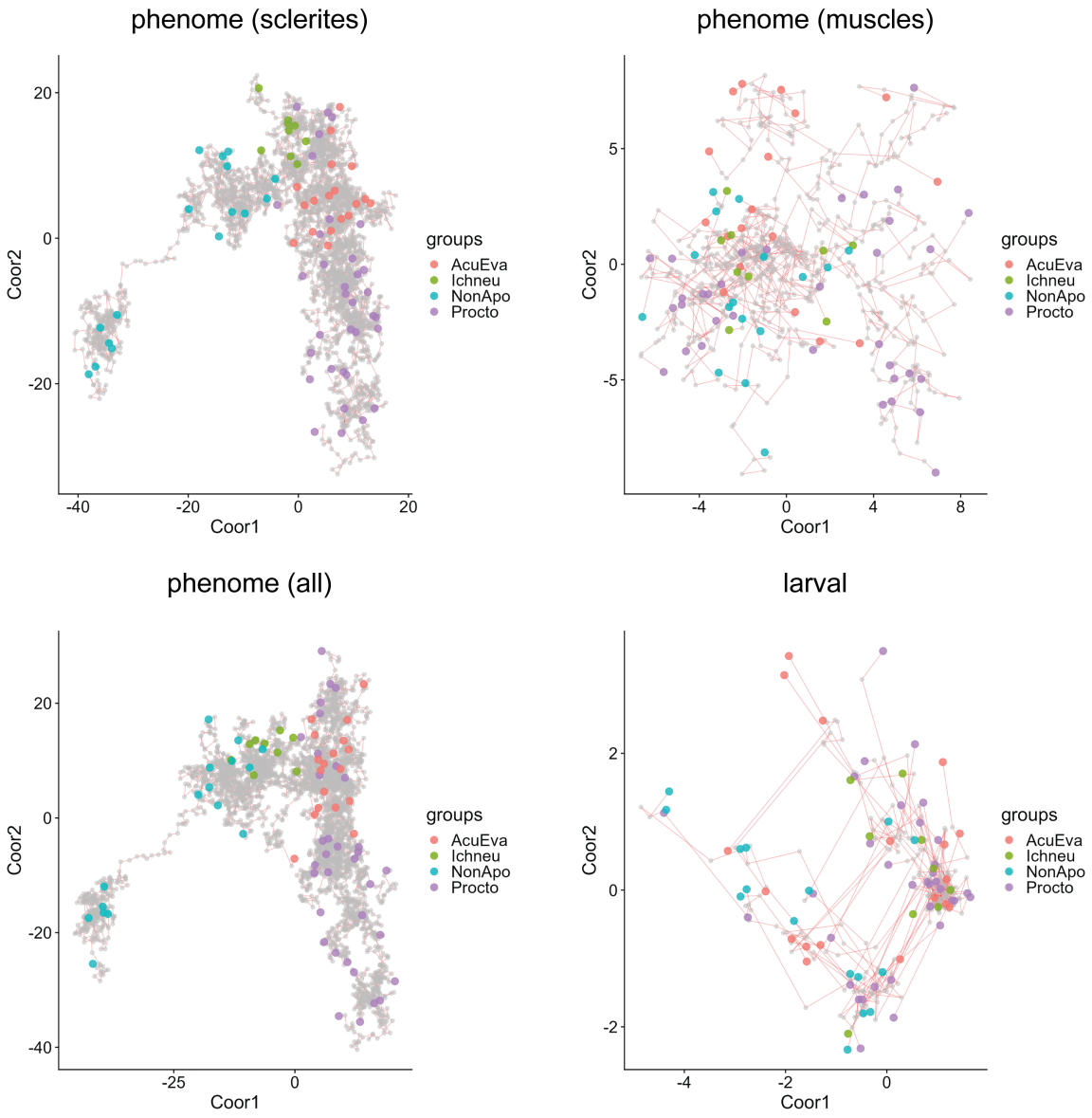

Fig. S14. Morphospaces reconstructed with *ontophylo* for the null analyses. Subpanels show the results for adult and larval phenome of Hymenoptera. Small, gray circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

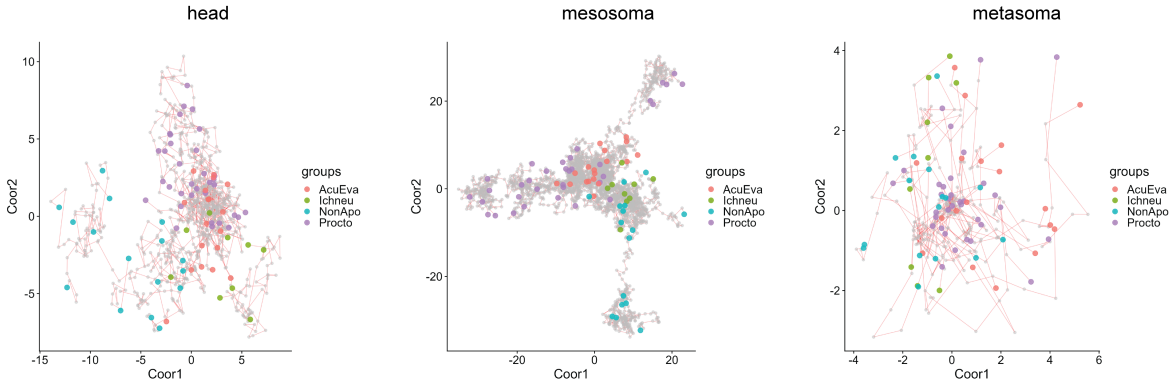

Fig. S15. Morphospaces reconstructed with *ontophylo* for the null analyses. Each subpanel shows the result for a body region of the Hymenoptera anatomy. Small, gray circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

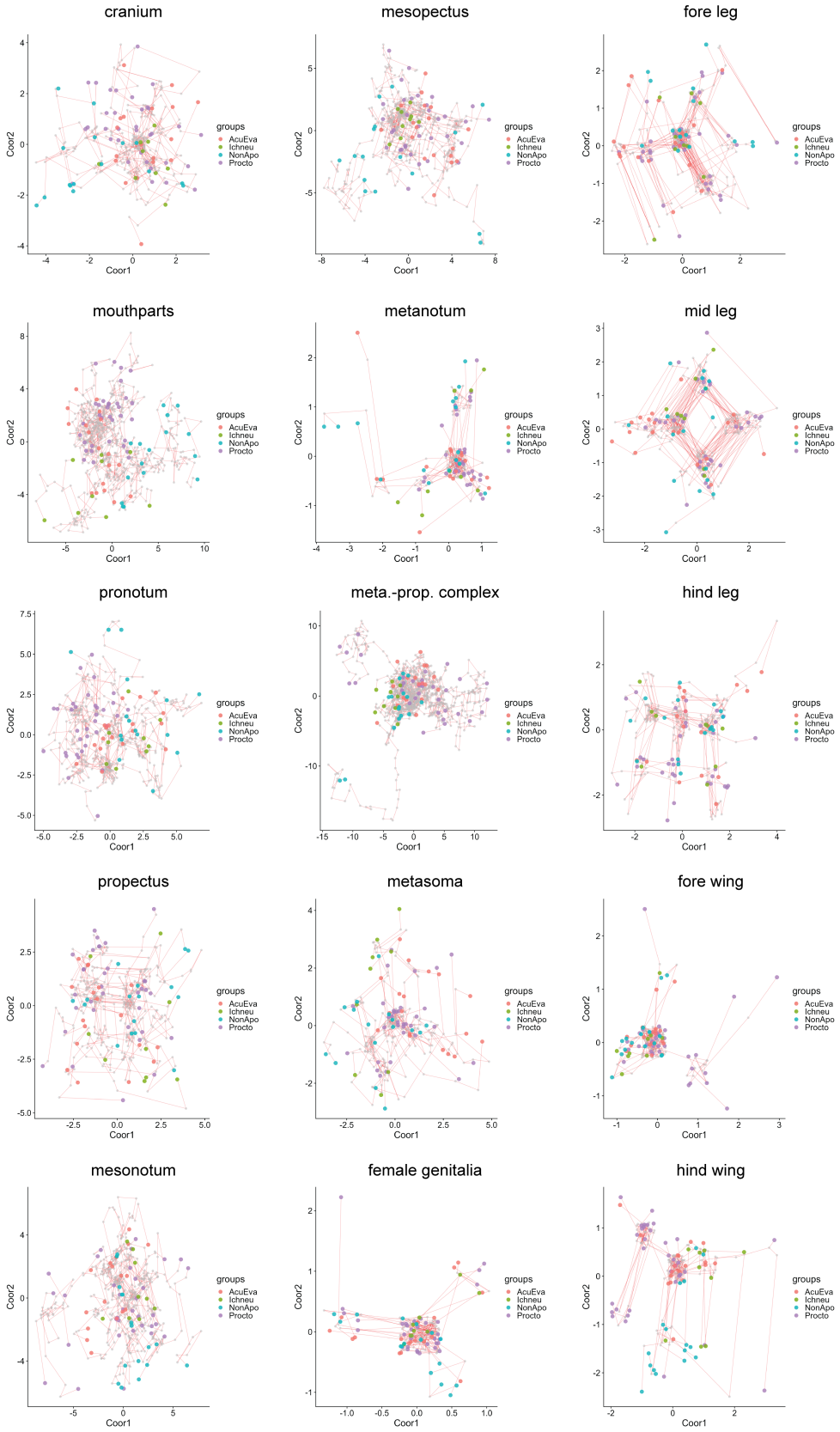

Fig. S16. Morphospaces reconstructed with *ontophylo* for the null analyses. Each subpanel shows the result for an elementary anatomical complex of the Hymenoptera anatomy. Small, gray circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

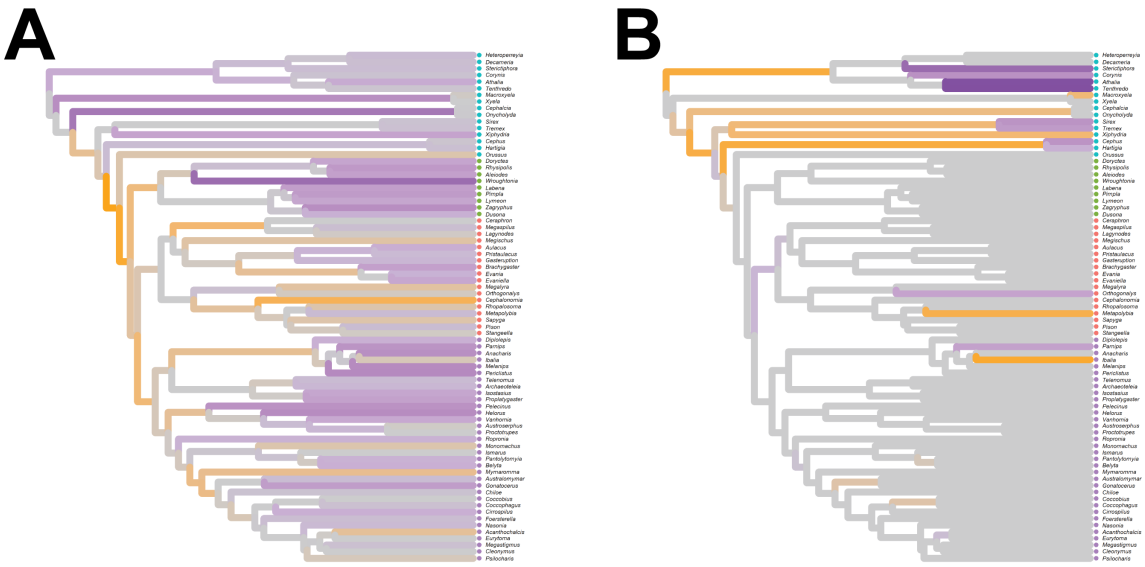

Fig. S17. Metric to assess regime of selection. *(A)* Full adult phenome. *(B)* Larval phenome. Tree branches shaded according to the difference between the metric calculated in the empirical and null analyses. Orange color indicates positive values (i.e., directional selection); purple color indicates negative values (i.e., stabilizing selection); gray color indicates non-significant values.

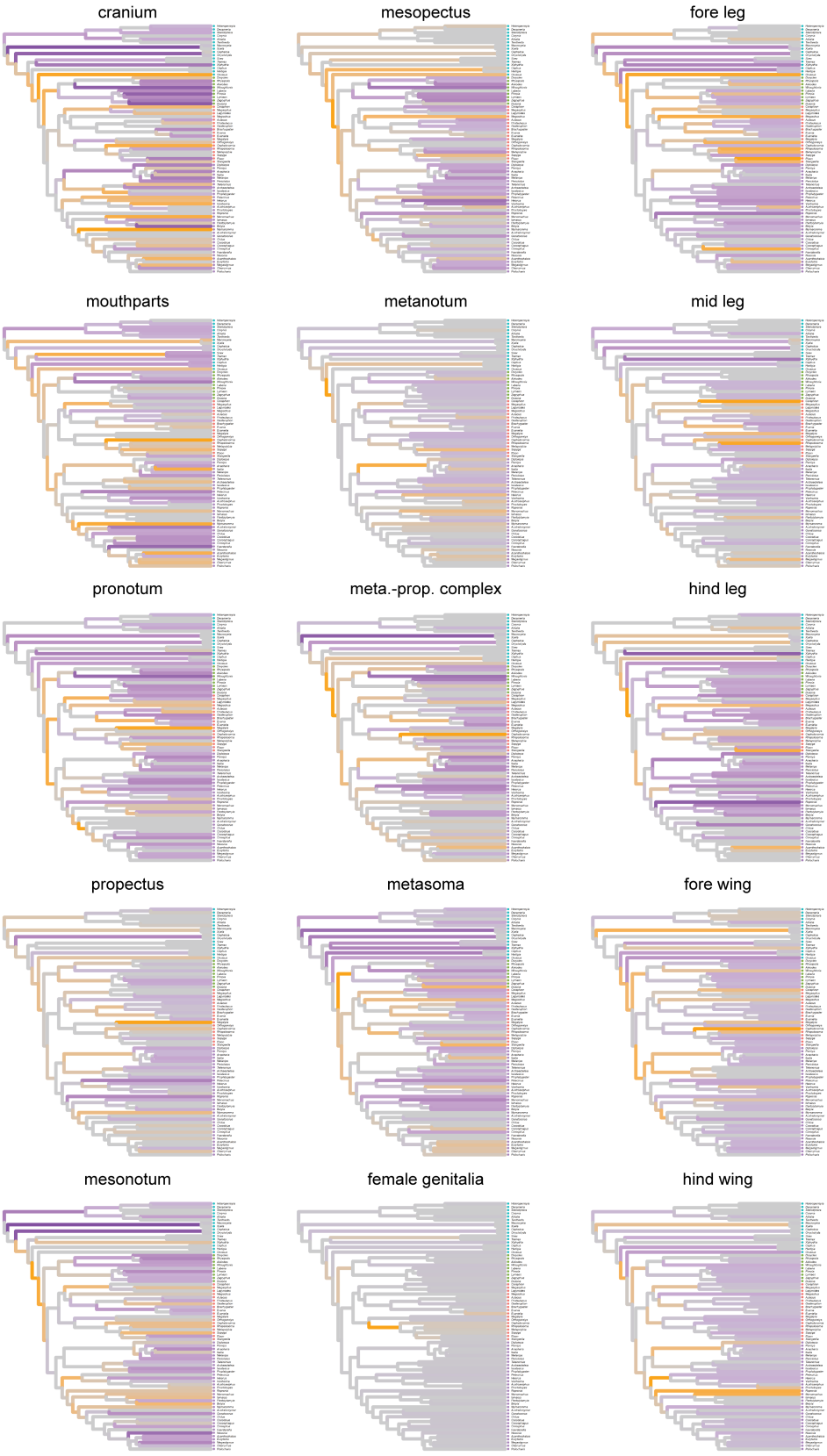

Fig. S18. Metric to assess regime of selection. Each subpanel shows the result for an elementary anatomical complex of the Hymenoptera anatomy. Tree branches shaded according to the difference between the metric calculated in the empirical and null analyses. Orange color indicates positive values (i.e., directional selection); purple color indicates negative values (i.e., stabilizing selection); gray color indicates non-significant values.

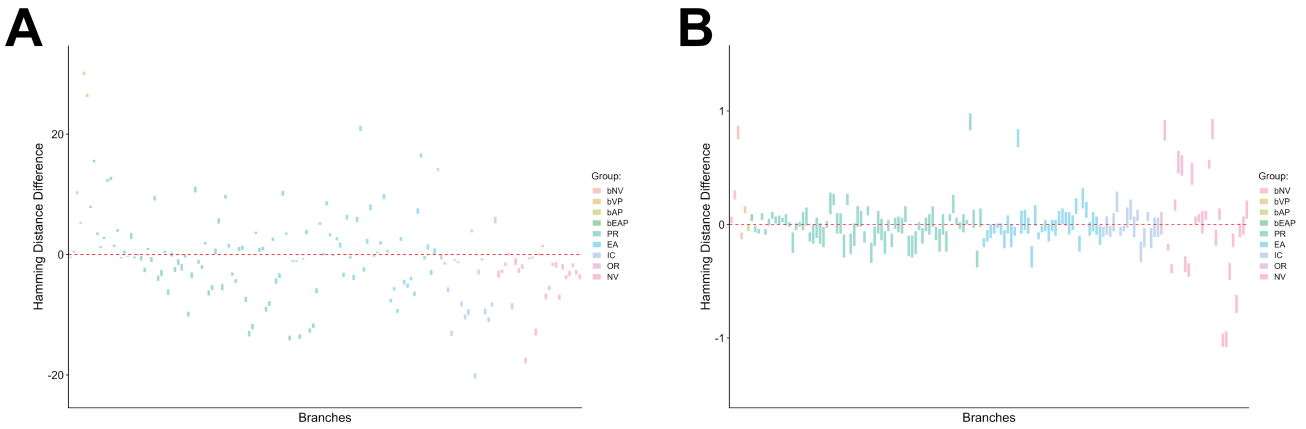

Fig. S19. Metric to assess regime of selection. *(A)* Full adult phenome. *(B)* Larval phenome. Boxplots show the difference between the metric calculated in the empirical and null analyses. Horizontal red dashed line indicates non-significant values. Values above the line are positive (i.e., directional selection); those below are negative (i.e., stabilizing selection). Boxplots shaded according to the branch groups: NV: non-Vespina; OR: Orussoidea; NA: non-Apocrita; ICH: Ichneumonoidea; EA: "Evaniomorpha"-Aculeata; PRO: Proctotrupomorpha; bNV: backbone non-Vespina; VP: backbone Vespina; AP: backbone Apocrita; EAP: backbone Apocrita non-Ichneumonoidea.

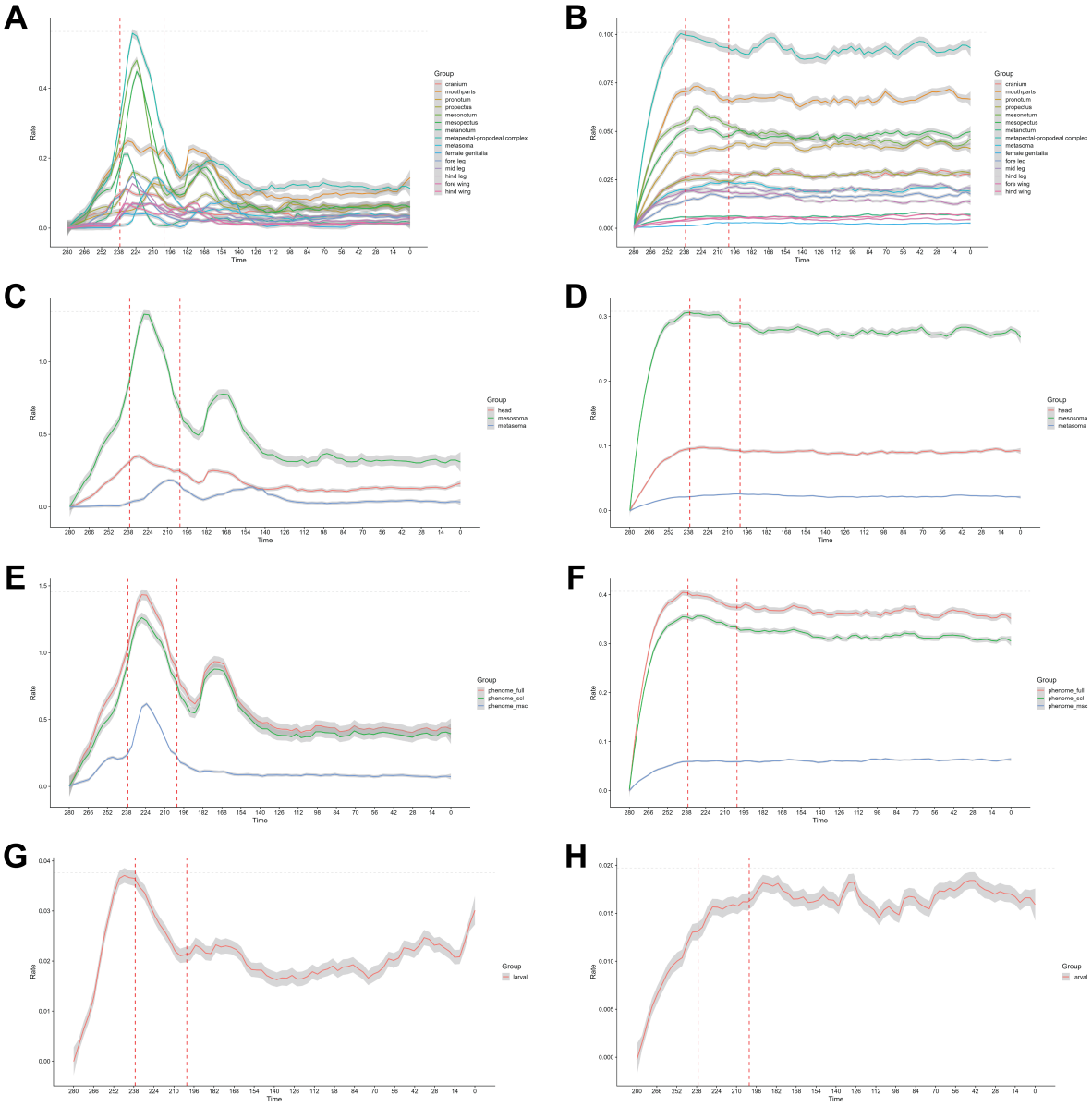

Fig. S20. Rate-through-time curves obtained from: *(A)* Anatomical complexes in the empirical analyses (15 partitions). *(B)* Anatomical complexes in the null analyses. *(C)* Body regions in the empirical analyses (3 partitions). *(D)* Body regions in the null analyses. *(E)* Adult phenome (all characters, only sclerites, and only muscles) in the empirical analyses. *(F)* Adult phenome in the null analyses. *(G)* Larval phenome in the empirical analyses. *(H)* Larval phenome in the null analyses. For each curve, rate values are a summary (max value) across all lineages represented at a given time bin; vertical red dashed lines indicate the limits of the Late Triassic.

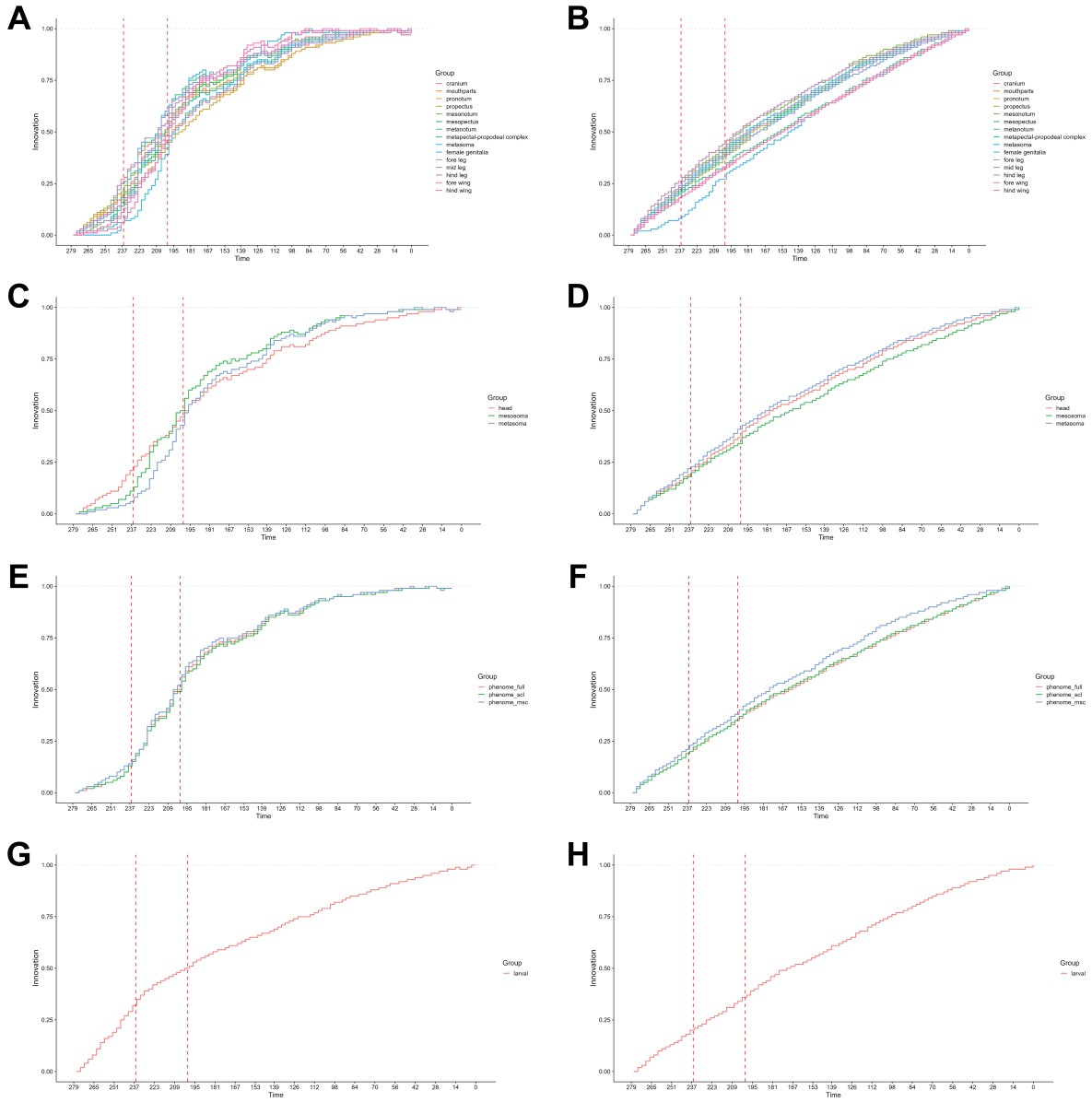

Fig. S21. Innovation-through-time curves obtained from: *(A)* Anatomical complexes in the empirical analyses (15 partitions). *(B)* Anatomical complexes in the null analyses. *(C)* Body regions in the empirical analyses (3 partitions). *(D)* Body regions in the null analyses. *(E)* Adult phenome (all characters, only sclerites, and only muscles) in the empirical analyses. *(F)* Adult phenome in the null analyses. *(G)* Larval phenome in the empirical analyses. *(H)* Larval phenome in the null analyses. For each curve, innovation values are a summary (mean value) across all lineages represented at a given time bin; values were rescaled between 0 and 1; vertical red dashed lines indicate the limits of the Late Triassic.

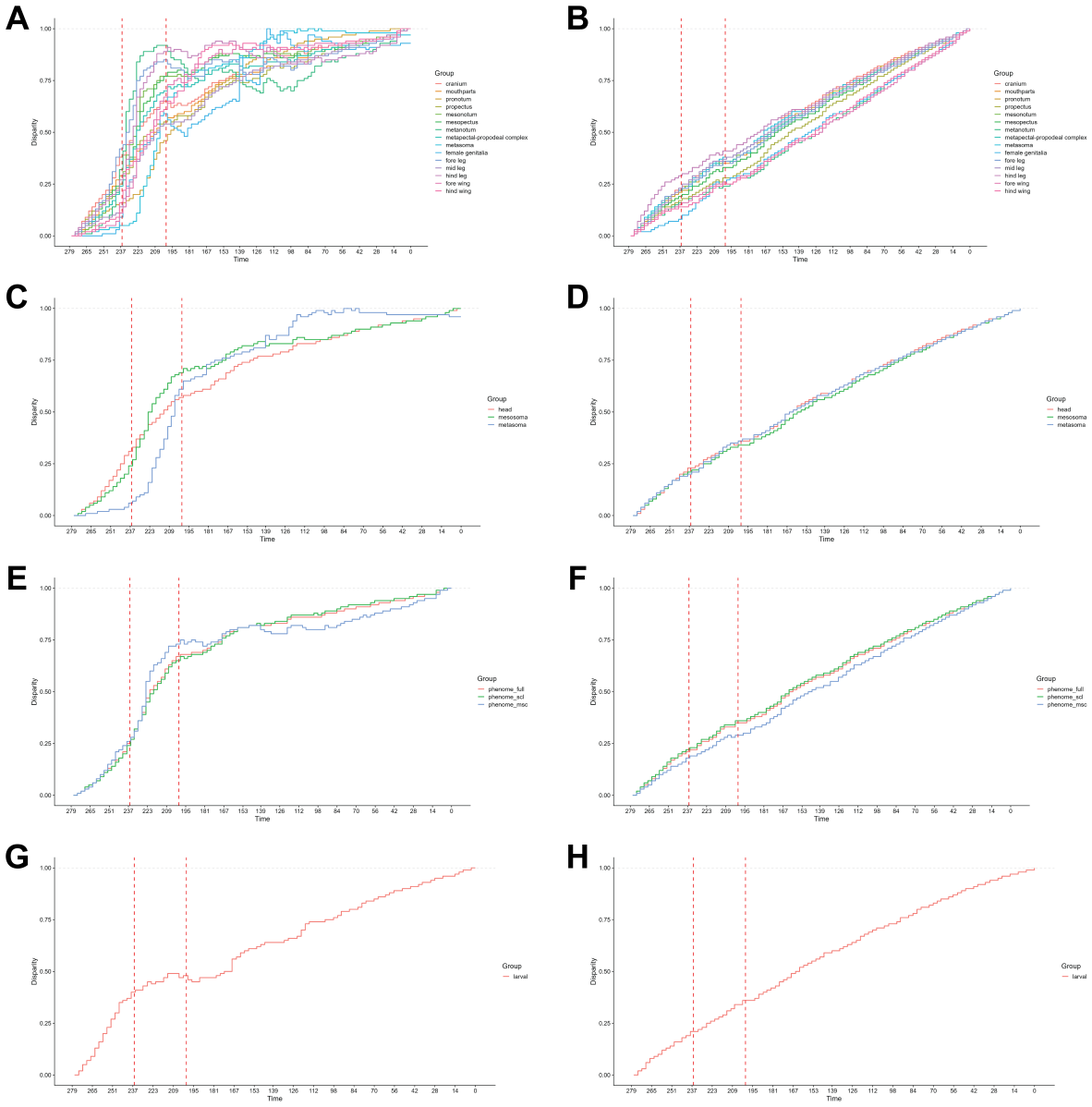

Fig. S22. Disparity-through-time curves obtained from: *(A)* Anatomical complexes in the empirical analyses (15 partitions). *(B)* Anatomical complexes in the null analyses. *(C)* Body regions in the empirical analyses (3 partitions). *(D)* Body regions in the null analyses. *(E)* Adult phenome (all characters, only sclerites, and only muscles) in the empirical analyses. *(F)* Adult phenome in the null analyses. *(G)* Larval phenome in the empirical analyses. *(H)* Larval phenome in the null analyses. For each curve, disparity values are calculated among all lineages (mean value) represented at a given time bin; values were rescaled between 0 and 1; vertical red dashed lines indicate the limits of the Late Triassic.

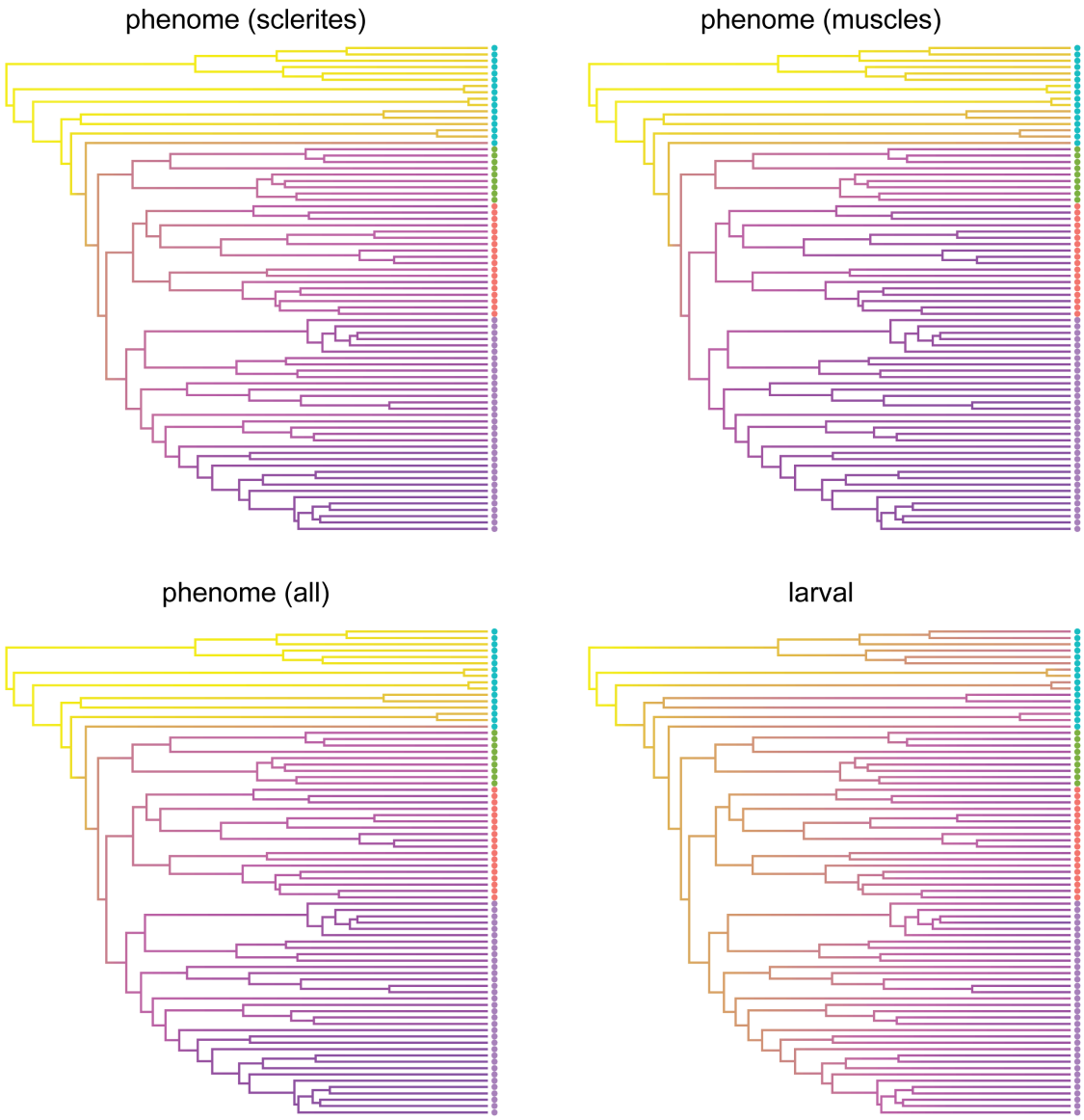

Fig. S23. Branch-specific innovation metric calculated from the samples of amalgamated stochastic maps in the empirical analyses. Subpanels show the results for adult and larval phenome of Hymenoptera. Purple color indicates higher innovation values. Small filled circles at the tips of the tree indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata; purple: Proctotrupomorpha.

Fig. S24. Branch-specific innovation metric calculated from the samples of amalgamated stochastic maps in the empirical analyses. Each subpanel shows the result for a body region of the Hymenoptera anatomy. Purple color indicates higher innovation values. Small filled circles at the tips of the tree indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: “Evaniomorpha”-Aculeata; purple: Proctotrupomorpha..

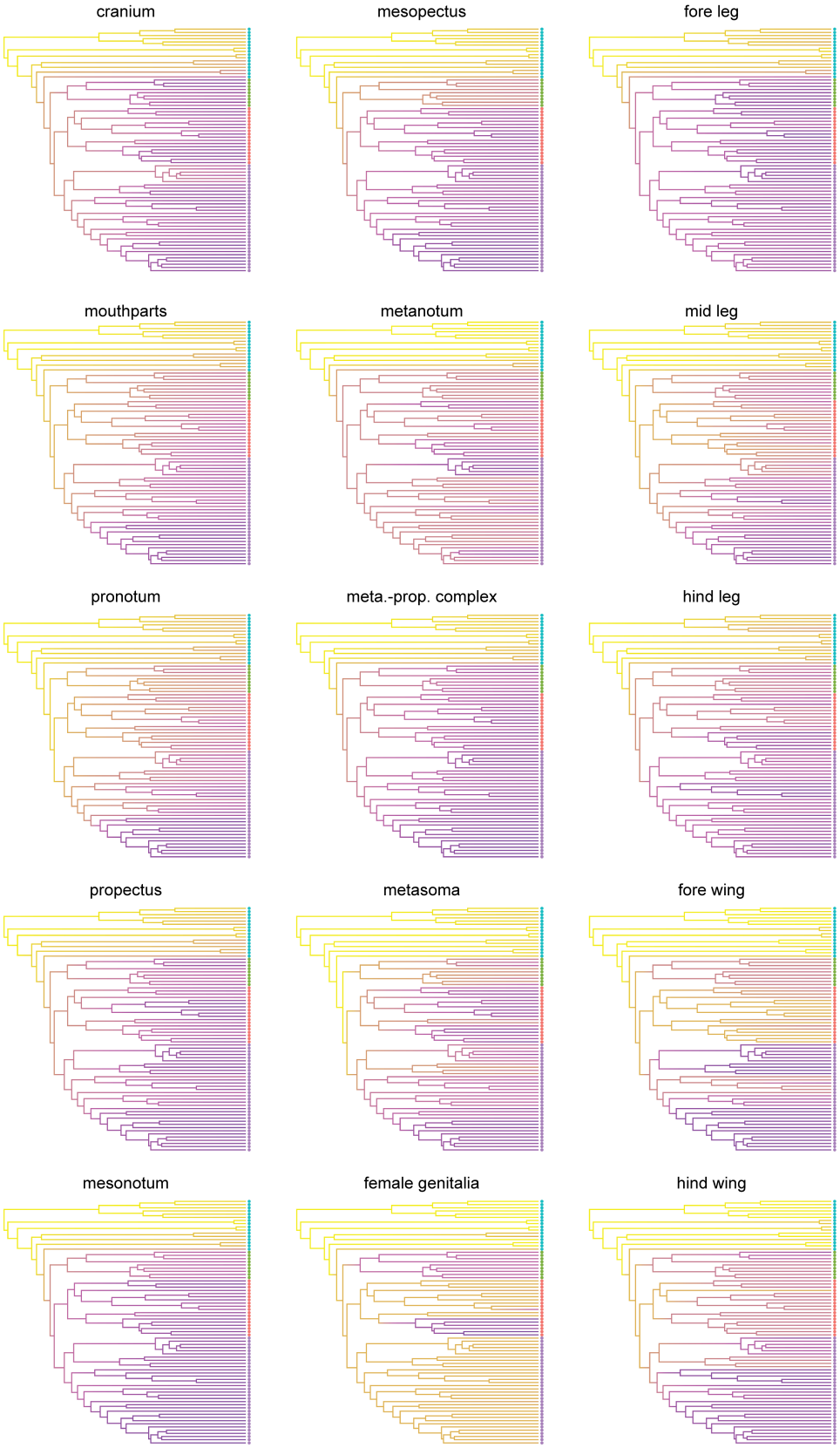

Fig. S25. Branch-specific innovation metric calculated from the samples of amalgamated stochastic maps in the empirical analyses. Each subpanel shows the result for an elementary anatomical complex of the Hymenoptera anatomy. Purple color indicates higher innovation values. Small filled circles at the tips of the tree indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata; purple: Proctotrupomorpha.

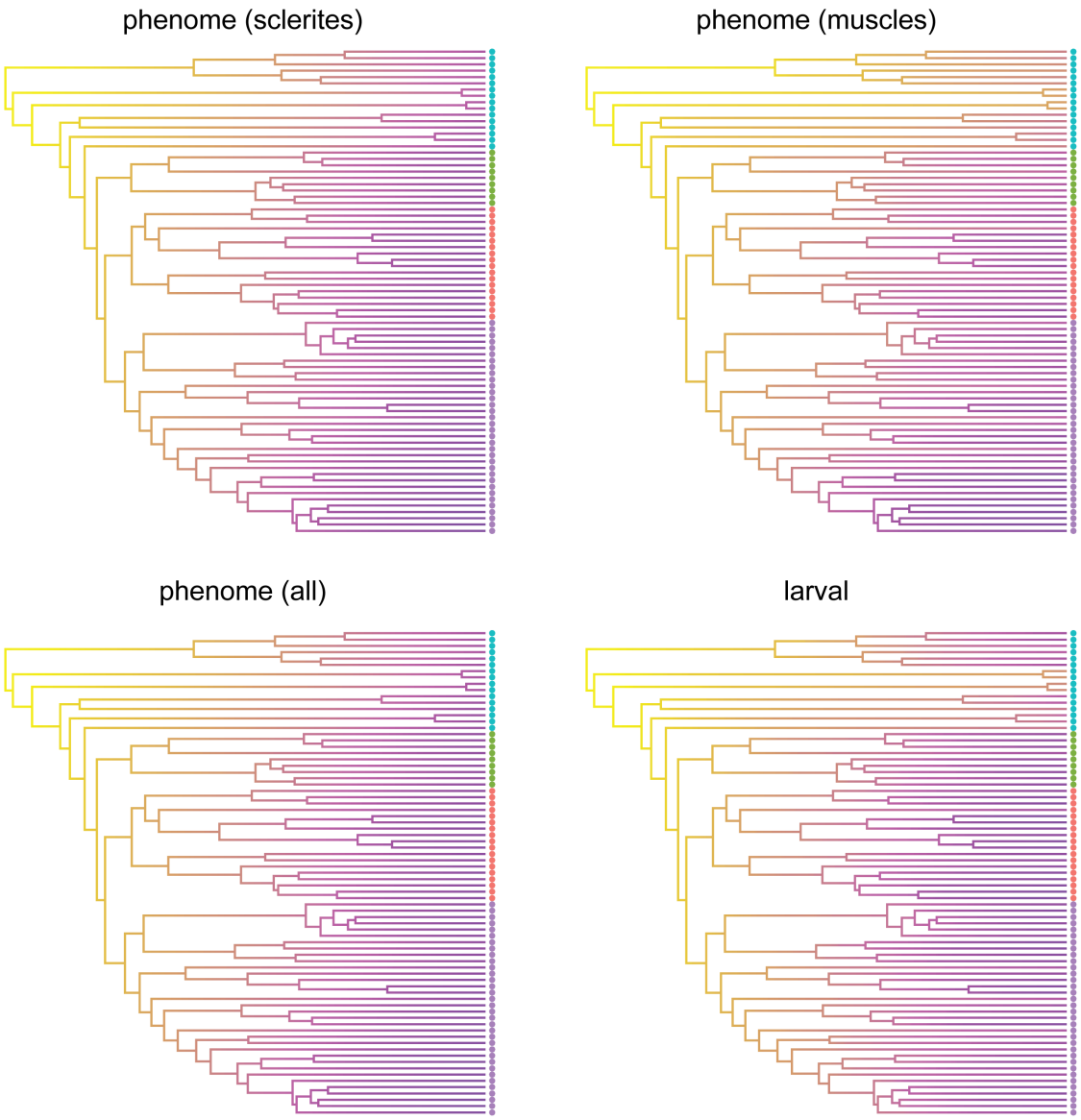

Fig. S26. Branch-specific innovation metric calculated from the samples of amalgamated stochastic maps in the null analyses. Subpanels show the results for adult and larval phenome of Hymenoptera. Purple color indicates higher innovation values. Small filled circles at the tips of the tree indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata; purple: Proctotrupomorpha.

Fig. S27. Branch-specific innovation metric calculated from the samples of amalgamated stochastic maps in the null analyses. Each subpanel shows the result for a body region of the Hymenoptera anatomy. Purple color indicates higher innovation values. Small filled circles at the tips of the tree indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: “Evaniomorpha”-Aculeata; purple: Proctotrupomorpha.

Fig. S28. Branch-specific innovation metric calculated from the samples of amalgamated stochastic maps in the null analyses. Each subpanel shows the result for an elementary anatomical complex of the Hymenoptera anatomy. Purple color indicates higher innovation values. Small filled circles at the tips of the tree indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata; purple: Proctotrupomorpha.

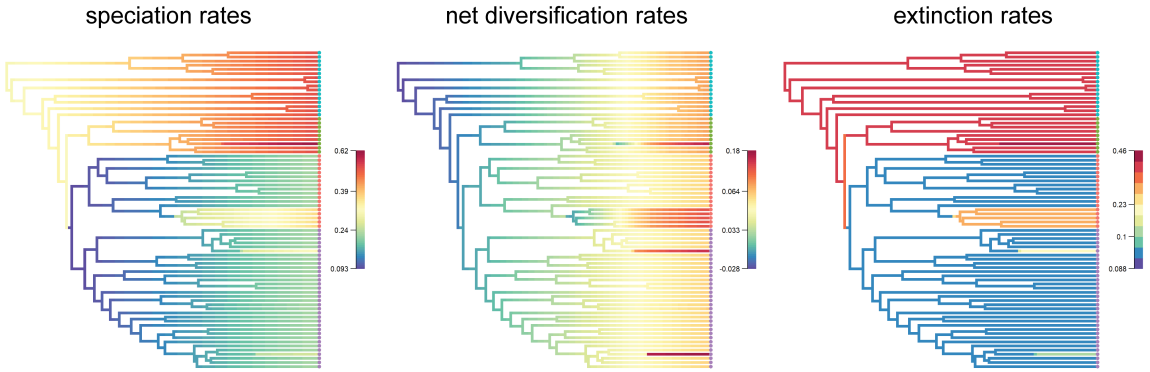

Fig. S29. Net diversification, speciation, and extinction rates as inferred in the BAMM analyses of Blameir et al. (10) plotted for the modified phylogeny used in this study.

Table S1. Hierarchical levels of the hymenopteran anatomy investigated in this study (L1-L3).

Hierarchical level	N_chars[1]
Anatomical complexes (L1)	
cranium	18
mouthparts	46
pronotum	25
propectus	27
mesonotum	37
mesopectus	35
metanotum	9
metapectal-propodeal complex	65
metasoma[2]	17
female genitalia	6
fore leg	12
mid leg	12
hind leg	12
fore wing	9
hind wing	10
Body regions (L2)	
head	64
mesosoma	253
metasoma	23
Adult phenome (L3)	
All	340
Only sclerites	280
Only muscles	60

[1]Number of individual characters in each anatomical partition (after combining dependent characters); [2]Excluding female genitalia.

Table S2. Results from the resampling approach to assess correlations among phenotypic rates of evolution.

Pair	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
mouthparts + propectus	0.08	0.34	0.56	0.000	0.000	0.163	0.95
mouthparts + mesonotum	0.19	0.39	0.57	0.000	0.000	0.011	0.99
mouthparts + mesopectus	0.09	0.33	0.54	0.000	0.000	0.137	0.95
mouthparts + metapectal-propodeal complex	0.19	0.38	0.58	0.000	0.000	0.010	0.99
mouthparts + fore wing	0.28	0.51	0.69	0.000	0.000	0.000	1.00
mouthparts + hind wing	0.11	0.36	0.61	0.000	0.000	0.090	0.95
propectus + mesonotum	0.24	0.50	0.72	0.000	0.000	0.001	1.00
propectus + mesopectus	0.20	0.48	0.72	0.000	0.000	0.007	1.00
propectus + metapectal-propodeal complex	0.30	0.53	0.71	0.000	0.000	0.000	1.00
propectus + fore leg	0.10	0.36	0.62	0.000	0.000	0.121	0.95
propectus + hind leg	0.12	0.41	0.66	0.000	0.000	0.079	0.97
propectus + fore wing	0.13	0.43	0.67	0.000	0.000	0.065	0.97
mesonotum + mesopectus	0.48	0.69	0.83	0.000	0.000	0.000	1.00
mesonotum + metanotum	0.21	0.54	0.77	0.000	0.000	0.007	0.99
mesonotum + metapectal-propodeal complex	0.49	0.68	0.81	0.000	0.000	0.000	1.00
mesonotum + fore leg	0.25	0.49	0.68	0.000	0.000	0.001	1.00
mesonotum + hind leg	0.21	0.48	0.70	0.000	0.000	0.007	0.99
mesonotum + fore wing	0.26	0.51	0.67	0.000	0.000	0.000	1.00
mesonotum + hind wing	0.18	0.42	0.65	0.000	0.000	0.016	0.99
mesopectus + metanotum	0.09	0.40	0.66	0.000	0.000	0.134	0.95
mesopectus + metapectal-propodeal complex	0.42	0.65	0.80	0.000	0.000	0.000	1.00
mesopectus + fore leg	0.16	0.44	0.65	0.000	0.000	0.022	0.99
mesopectus + hind leg	0.19	0.50	0.74	0.000	0.000	0.009	0.99
mesopectus + fore wing	0.18	0.47	0.66	0.000	0.000	0.015	0.99
metanotum + metapectal-propodeal complex	0.21	0.57	0.78	0.000	0.000	0.004	1.00
metanotum + fore leg	0.15	0.46	0.75	0.000	0.000	0.028	0.98
metanotum + hind wing	0.18	0.55	0.72	0.000	0.000	0.015	0.98
metapectal-propodeal complex + fore leg	0.26	0.55	0.73	0.000	0.000	0.000	1.00
metapectal-propodeal complex + hind leg	0.29	0.57	0.77	0.000	0.000	0.000	1.00
metapectal-propodeal complex + fore wing	0.29	0.54	0.70	0.000	0.000	0.000	1.00
metapectal-propodeal complex + hind wing	0.16	0.38	0.67	0.000	0.000	0.029	0.99
fore leg + hind leg	0.13	0.44	0.70	0.000	0.000	0.050	0.98
hind leg + fore wing	0.12	0.43	0.65	0.000	0.000	0.078	0.97
fore wing + hind wing	0.13	0.45	0.70	0.000	0.000	0.060	0.97

[1-3]Lower boundary, median, and upper boundary of the 95% confidence interval for the distribution of correlation values; [4-6]Lower boundary, median, and upper boundary of the 95% confidence interval for the p-values obtained in the permutation tests; [7]Proportion of significant p-values.

Table S3. Results from the regression approach to assess correlations among phenotypic rates of evolution.

Pair	Pval_prop[1]	Pval_lw[2]	Pval_md[3]	Pval_up[4]	Corr_c_lw[5]	Corr_c_md[6]	Corr_c_up[7]	Corr_u_lw[8]	Corr_u_md[9]	Corr_u_up[10]
mouthparts + mesonotum	0.98	0.000	0.000	0.036	0.17	0.39	0.59	0.17	0.39	0.58
mouthparts + metapectal-propodeal complex	0.99	0.000	0.000	0.018	0.19	0.39	0.61	0.19	0.40	0.59
mouthparts + fore wing	1.00	0.000	0.000	0.001	0.27	0.50	0.71	0.29	0.51	0.70
propectus + mesonotum	1.00	0.000	0.000	0.002	0.25	0.49	0.73	0.26	0.49	0.72
propectus + mesopectus	0.98	0.000	0.000	0.022	0.19	0.48	0.74	0.19	0.48	0.73
propectus + metapectal-propodeal complex	1.00	0.000	0.000	0.001	0.27	0.52	0.72	0.27	0.52	0.72
propectus + fore wing	0.97	0.000	0.000	0.091	0.14	0.42	0.66	0.14	0.42	0.67
mesonotum + mesopectus	1.00	0.000	0.000	0.000	0.49	0.70	0.83	0.49	0.70	0.83
mesonotum + metanotum	0.99	0.000	0.000	0.007	0.22	0.54	0.77	0.22	0.55	0.77
mesonotum + metapectal-propodeal complex	1.00	0.000	0.000	0.000	0.48	0.67	0.82	0.50	0.68	0.81
mesonotum + fore leg	1.00	0.000	0.000	0.002	0.25	0.48	0.68	0.27	0.49	0.68
mesonotum + hind leg	0.99	0.000	0.000	0.018	0.18	0.47	0.68	0.21	0.47	0.70
mesonotum + fore wing	1.00	0.000	0.000	0.001	0.26	0.50	0.68	0.27	0.51	0.68
mesonotum + hind wing	0.98	0.000	0.000	0.046	0.16	0.42	0.69	0.17	0.42	0.68
mesopectus + metapectal-propodeal complex	1.00	0.000	0.000	0.000	0.39	0.65	0.81	0.40	0.65	0.81
mesopectus + fore leg	0.98	0.000	0.000	0.045	0.17	0.43	0.67	0.17	0.44	0.66
mesopectus + hind leg	0.98	0.000	0.000	0.042	0.17	0.50	0.74	0.17	0.50	0.74
mesopectus + fore wing	0.98	0.000	0.000	0.026	0.18	0.46	0.64	0.18	0.47	0.64
metanotum + metapectal-propodeal complex	0.99	0.000	0.000	0.006	0.22	0.57	0.79	0.22	0.57	0.79
metanotum + hind wing	0.98	0.000	0.000	0.047	0.15	0.53	0.73	0.17	0.54	0.73
metapectal-propodeal complex + fore leg	1.00	0.000	0.000	0.000	0.28	0.54	0.73	0.29	0.54	0.73
metapectal-propodeal complex + hind leg	1.00	0.000	0.000	0.000	0.30	0.57	0.75	0.30	0.58	0.75
metapectal-propodeal complex + fore wing	1.00	0.000	0.000	0.000	0.29	0.55	0.73	0.30	0.55	0.72
metapectal-propodeal complex + hind wing	0.97	0.000	0.000	0.056	0.16	0.39	0.65	0.16	0.38	0.65
fore leg + hind leg	0.97	0.000	0.000	0.104	0.14	0.43	0.70	0.14	0.44	0.69
hind leg + fore wing	0.97	0.000	0.000	0.101	0.14	0.43	0.68	0.13	0.43	0.65
fore wing + hind wing	0.96	0.000	0.000	0.132	0.12	0.44	0.69	0.13	0.44	0.70

[1] Proportion of significant p-values obtained from the regressions including both rates from the empirical and null analyses. [2-4]Lower boundary, median, and upper boundary of the 95% confidence interval for the p-values obtained in the regressions including both rates from the empirical and null analyses; [5-7]Lower boundary, median, and upper boundary of the 95% confidence interval for the distribution of (corrected) correlation values obtained from the regressions including both rates from the empirical and null analyses; [8-10]Lower boundary, median, and upper boundary of the 95% confidence interval for the distribution of (uncorrected) correlation values obtained from the regressions including only rates from the empirical analyses.

Table S4. Results from the permutation tests to assess correlations between phenotypic and lineage diversification rates (net diversification).

Partition	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
Anatomical complexes							
cranium	-0.34	-0.23	-0.10	0.000	0.004	0.210	0.81
mouthparts	-0.44	-0.31	-0.22	0.000	0.000	0.006	1.00
pronotum	-0.32	-0.19	-0.05	0.000	0.019	0.501	0.71
propectus	-0.39	-0.28	-0.18	0.000	0.000	0.021	1.00
mesonotum	-0.42	-0.35	-0.25	0.000	0.000	0.001	1.00
mesopectus	-0.43	-0.34	-0.27	0.000	0.000	0.001	1.00
metanotum	-0.31	-0.22	-0.17	0.000	0.006	0.043	0.99
metapectal-propodeal complex	-0.43	-0.35	-0.27	0.000	0.000	0.001	1.00
metasoma	-0.30	-0.22	-0.12	0.000	0.008	0.146	0.89
female genitalia	-0.22	-0.18	-0.09	0.005	0.023	0.285	0.87
fore leg	-0.39	-0.24	-0.14	0.000	0.003	0.105	0.94
mid leg	-0.26	-0.14	-0.03	0.000	0.074	0.706	0.41
hind leg	-0.31	-0.21	-0.10	0.000	0.009	0.218	0.74
fore wing	-0.35	-0.30	-0.23	0.000	0.000	0.004	1.00
hind wing	-0.34	-0.27	-0.21	0.000	0.000	0.009	1.00
larval	-0.26	-0.04	0.14	0.002	0.428	0.954	0.10
Adult phenome							
phenome_full	-0.50	-0.44	-0.41	0.000	0.000	0.000	1.00

[1-7]Column descriptions as in Table S2.

Table S5. Results from the permutation tests to assess correlations between phenotypic and lineage diversification rates (speciation).

Partition	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
Anatomical complexes							
cranium	-0.11	0.00	0.11	0.118	0.657	0.993	0.00
mouthparts	-0.16	-0.08	0.01	0.062	0.340	0.947	0.02
pronotum	-0.23	-0.13	-0.05	0.004	0.120	0.553	0.21
propectus	-0.09	0.00	0.10	0.194	0.712	0.983	0.00
mesonotum	-0.03	0.02	0.07	0.356	0.722	0.987	0.00
mesopectus	-0.04	0.02	0.07	0.401	0.788	0.995	0.00
metanotum	0.04	0.09	0.14	0.102	0.263	0.567	0.00
metapectal-propodeal complex	-0.01	0.04	0.11	0.189	0.596	0.988	0.00
metasoma	-0.15	-0.07	0.02	0.065	0.370	0.917	0.01
female genitalia	0.01	0.05	0.08	0.335	0.567	0.898	0.00
fore leg	0.04	0.13	0.22	0.006	0.123	0.614	0.19
mid leg	-0.11	0.00	0.14	0.082	0.582	0.971	0.00
hind leg	-0.02	0.11	0.21	0.009	0.185	0.915	0.13
fore wing	-0.07	-0.04	0.02	0.384	0.640	0.982	0.00
hind wing	-0.10	-0.04	0.04	0.223	0.652	0.988	0.00
larval	-0.09	0.04	0.18	0.023	0.524	0.951	0.07
Adult phenome							
phenome_full	-0.03	0.01	0.04	0.635	0.885	0.993	0.00

[1-7]Column descriptions as in Table S2.

Table S6. Results from the permutation tests to assess correlations between phenotypic and lineage diversification rates (extinction).

Partition	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
Anatomical complexes							
cranium	-0.07	0.04	0.15	0.060	0.587	0.958	0.01
mouthparts	-0.11	-0.02	0.06	0.200	0.676	0.972	0.00
pronotum	-0.20	-0.09	-0.01	0.012	0.245	0.884	0.10
propectus	-0.04	0.06	0.15	0.060	0.460	0.966	0.02
mesonotum	0.03	0.08	0.14	0.081	0.309	0.737	0.00
mesopectus	0.03	0.07	0.13	0.115	0.370	0.748	0.00
metanotum	0.09	0.13	0.18	0.027	0.106	0.280	0.14
metapectal-propodeal complex	0.05	0.11	0.17	0.036	0.191	0.540	0.07
metasoma	-0.12	-0.04	0.07	0.152	0.575	0.990	0.00
female genitalia	0.02	0.08	0.11	0.181	0.325	0.769	0.00
fore leg	0.09	0.17	0.27	0.000	0.039	0.284	0.59
mid leg	-0.09	0.03	0.16	0.051	0.557	0.981	0.03
hind leg	-0.02	0.14	0.26	0.001	0.088	0.818	0.32
fore wing	-0.03	0.02	0.07	0.414	0.797	0.993	0.00
hind wing	-0.06	0.01	0.08	0.332	0.828	0.995	0.00
larval	-0.12	0.05	0.21	0.011	0.438	0.983	0.06
Adult phenome							
phenome_full	0.05	0.08	0.11	0.158	0.303	0.540	0.00

[1-7]Column descriptions as in Table S2.

Table S7. Results from the permutation tests to assess correlations between regime of selection and lineage diversification rates (net diversification).

Partition	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
Anatomical complexes							
cranium	-0.20	-0.18	-0.16	0.011	0.027	0.058	0.94
mouthparts	-0.25	-0.23	-0.21	0.001	0.004	0.011	1.00
pronotum	-0.29	-0.27	-0.25	0.000	0.001	0.004	1.00
propectus	-0.27	-0.26	-0.24	0.000	0.001	0.005	1.00
mesonotum	-0.30	-0.28	-0.26	0.000	0.000	0.003	1.00
mesopectus	-0.42	-0.40	-0.38	0.000	0.000	0.000	1.00
metanotum	-0.16	-0.14	-0.13	0.045	0.080	0.127	0.05
metapectal-propodeal complex	-0.34	-0.32	-0.30	0.000	0.000	0.001	1.00
metasoma	-0.08	-0.07	-0.05	0.312	0.421	0.544	0.00
female genitalia	-0.14	-0.13	-0.12	0.076	0.108	0.149	0.00
fore leg	-0.11	-0.08	-0.06	0.198	0.311	0.480	0.00
mid leg	-0.08	-0.06	-0.03	0.324	0.504	0.736	0.00
hind leg	-0.20	-0.16	-0.13	0.013	0.043	0.100	0.62
fore wing	-0.32	-0.30	-0.29	0.000	0.000	0.001	1.00
hind wing	-0.32	-0.30	-0.28	0.000	0.000	0.001	1.00
larval	-0.23	-0.18	-0.13	0.003	0.024	0.112	0.81
Adult phenome							
phenome_full	-0.42	-0.41	-0.40	0.000	0.000	0.000	1.00

[1-7]Column descriptions as in Table S2.

Table S8. Results from the permutation tests to assess correlations between regime of selection and lineage diversification rates (speciation).

Partition	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
Anatomical complexes							
cranium	-0.31	-0.28	-0.25	0.000	0.000	0.003	1.00
mouthparts	-0.19	-0.17	-0.15	0.015	0.033	0.062	0.90
pronotum	-0.26	-0.23	-0.21	0.000	0.004	0.012	1.00
propectus	-0.05	-0.03	-0.01	0.541	0.742	0.942	0.00
mesonotum	-0.19	-0.17	-0.15	0.016	0.035	0.065	0.87
mesopectus	-0.01	0.02	0.05	0.581	0.802	0.987	0.00
metanotum	-0.08	-0.06	-0.04	0.309	0.459	0.632	0.00
metapectal-propodeal complex	-0.10	-0.08	-0.06	0.203	0.319	0.471	0.00
metasoma	-0.18	-0.16	-0.15	0.026	0.047	0.079	0.61
female genitalia	-0.03	-0.02	-0.01	0.693	0.832	0.956	0.00
fore leg	0.09	0.11	0.14	0.086	0.175	0.304	0.00
mid leg	-0.19	-0.16	-0.13	0.021	0.054	0.119	0.44
hind leg	0.09	0.13	0.16	0.045	0.126	0.274	0.04
fore wing	-0.11	-0.09	-0.06	0.183	0.296	0.443	0.00
hind wing	-0.16	-0.14	-0.12	0.045	0.082	0.139	0.06
larval	-0.06	0.01	0.07	0.355	0.767	0.985	0.00
Adult phenome							
phenome_full	-0.20	-0.19	-0.17	0.012	0.023	0.037	1.00

[1-7]Column descriptions as in Table S2.

Table S9. Results from the permutation tests to assess correlations between regime of selection and lineage diversification rates (extinction).

Partition	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
Anatomical complexes							
cranium	-0.28	-0.25	-0.23	0.000	0.002	0.007	1.00
mouthparts	-0.16	-0.13	-0.11	0.055	0.100	0.170	0.01
pronotum	-0.21	-0.19	-0.16	0.007	0.021	0.051	0.97
propectus	-0.01	0.02	0.04	0.620	0.836	0.993	0.00
mesonotum	-0.14	-0.12	-0.11	0.076	0.128	0.207	0.00
mesopectus	0.07	0.09	0.11	0.162	0.271	0.421	0.00
metanotum	-0.06	-0.04	-0.01	0.475	0.661	0.868	0.00
metapectal-propodeal complex	-0.05	-0.03	0.00	0.539	0.742	0.949	0.00
metasoma	-0.17	-0.15	-0.13	0.036	0.065	0.108	0.21
female genitalia	-0.01	0.01	0.02	0.814	0.942	0.999	0.00
fore leg	0.10	0.12	0.15	0.058	0.127	0.236	0.01
mid leg	-0.18	-0.15	-0.12	0.028	0.070	0.160	0.23
hind leg	0.12	0.15	0.19	0.018	0.060	0.145	0.38
fore wing	-0.06	-0.03	-0.01	0.495	0.697	0.906	0.00
hind wing	-0.11	-0.09	-0.07	0.172	0.271	0.397	0.00
larval	-0.02	0.04	0.10	0.211	0.612	0.976	0.00
Adult phenome							
phenome_full	-0.13	-0.11	-0.10	0.118	0.163	0.212	0.00

[1-7]Column descriptions as in Table S2.

Table S10. Results from the permutation tests to assess correlations between phenotypic rates and regime of selection.

Partition	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
Anatomical complexes							
cranium	0.21	0.32	0.40	0.000	0.000	0.009	1.00
mouthparts	0.35	0.43	0.49	0.000	0.000	0.000	1.00
pronotum	0.41	0.49	0.55	0.000	0.000	0.000	1.00
propectus	0.29	0.37	0.45	0.000	0.000	0.001	1.00
mesonotum	0.48	0.54	0.58	0.000	0.000	0.000	1.00
mesopectus	0.43	0.52	0.57	0.000	0.000	0.000	1.00
metanotum	0.43	0.59	0.63	0.000	0.000	0.000	1.00
metapectal-propodeal complex	0.47	0.52	0.57	0.000	0.000	0.000	1.00
metasoma	0.34	0.46	0.55	0.000	0.000	0.000	1.00
female genitalia	0.53	0.74	0.89	0.000	0.000	0.000	1.00
fore leg	0.26	0.38	0.45	0.000	0.000	0.002	1.00
mid leg	0.17	0.28	0.39	0.000	0.000	0.034	0.98
hind leg	0.27	0.42	0.53	0.000	0.000	0.001	1.00
fore wing	0.42	0.55	0.61	0.000	0.000	0.000	1.00
hind wing	0.38	0.55	0.61	0.000	0.000	0.000	1.00
larval	0.04	0.15	0.28	0.000	0.071	0.606	0.41
Adult phenome							
phenome_full	0.63	0.65	0.68	0.000	0.000	0.000	1.00

[1-7]Column descriptions as in Table S2.

References

1. MJ Sharkey, et al., Phylogenetic relationships among superfamilies of hymenoptera. *Cladistics* **28**, 80–112 (2012).
2. F Ronquist, et al., A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst. biology* **61**, 973–999 (2012).
3. S Tarasov, Integration of anatomy ontologies and evo-devo using structured markov models suggests a new framework for modeling discrete phenotypic traits. *Syst. biology* **68**, 698–716 (2019).
4. S Tarasov, New phylogenetic markov models for inapplicable morphological characters. *Syst. biology* **72**, 681–693 (2023).
5. DS Porto, et al., rphenoscate: An r package for semantics-aware evolutionary analyses of anatomical traits. *Methods Ecol. Evol.* **14**, 2531–2540 (2023).
6. AE Thessen, et al., Transforming the study of organisms: Phenomic data models and knowledge bases. *PLoS computational biology* **16**, e1008376 (2020).
7. DS Porto, J Uyeda, I Mikó, S Tarasov, ontophylo: Reconstructing the evolutionary dynamics of phenomes using new ontology-informed phylogenetic methods. *Methods Ecol. Evol.* **15**, 290–300 (2024).
8. MJ Yoder, I Mikó, KC Seltmann, MA Bertone, AR Deans, A gross anatomy ontology for hymenoptera. *PloS one* **5**, e15991 (2010).
9. S Tarasov, I Mikó, MJ Yoder, ontofast: An r package for interactive and semi-automatic annotation of characters with biological ontologies. *Methods Ecol. Evol.* **13**, 324–329 (2022).
10. BB Blaimer, et al., Key innovations and the diversification of hymenoptera. *Nat. Commun.* **14**, 1212 (2023).
11. S Tarasov, I Mikó, MJ Yoder, JC Uyeda, Paramo: A pipeline for reconstructing ancestral anatomies using ontologies and stochastic mapping. *Insect Syst. Divers.* **3**, 1 (2019).
12. JM Beaulieu, BC O'Meara, MJ Donoghue, Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Syst. biology* **62**, 725–737 (2013).
13. JP Huelsenbeck, R Nielsen, JP Bollback, Stochastic mapping of morphological characters. *Syst. biology* **52**, 131–158 (2003).
14. D Greene, S Richardson, E Turro, ontologyx: a suite of r packages for working with ontological data. *Bioinformatics* **33**, 1104–1106 (2017).
15. C Parins-Fukuchi, GW Stull, SA Smith, Phylogenomic conflict coincides with rapid morphological innovation. *Proc. Natl. Acad. Sci.* **118**, e2023058118 (2021).
16. Y Asar, H Sauquet, SY Ho, Evaluating the accuracy of methods for detecting correlated rates of molecular and morphological evolution. *Syst. Biol.* **72**, 1337–1356 (2023).
17. LJ Revell, phytools 2.0: an updated r ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**, e16505 (2024).