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 January 2, 2025

Key innovations are traits that enable and/or facilitate the radiation of lineages often in association with the invasion of new adaptive zones. Despite their importance for understanding phenotypic macroevolutionary dynamics in megadiverse lineages, these traits are usually treated in isolation ignoring the context within organismal anatomy and potential trait synergies. Furthermore, what is often interpreted as a key innovation might represent an evolutionary singular event at the origin of a new body plan affecting multiple morphological traits at once. Here we argue that the formation of the typical "wasp waist" of apocritan Hymenoptera represents such a macroevolutionary singular event. We employed a set of newly developed ontology-informed comparative tools to assess the evolution of the phenome of Hymenoptera based on a dataset including adult and larval characters. We showed that the phenotypic diversification process in Hymenoptera simultaneously affected multiple anatomical complexes and was characterized by fast evolutionary rates and strong directional selection at branches leading to major lineages——particularly Vespina and Apocrita——during the Late Triassic (201∼237 Ma). Additionally, evolutionary rates of adult and larval phenomes were decoupled suggesting that distinct phenotypic diversification processes characterized different ontogenetic stages in Hymenoptera. Lastly, no clear association was recovered between phenotypic and lineage diversification rates for any morphological complex of the adult or the larval phenome. These results suggest that some key innovations might have a delayed effect on linage diversification, as observed in other recent studies.

Phenomics | phenotypic innovation | ontologies | ontophylo | PARAMO

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 (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,

Author affiliations: [a]Finnish Museum of Natural History, Helsinki, Finland; [b]Natural History Museum of Denmark, 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

strong integration can also reduce morphological disparity by limiting 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.

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 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 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 latest time-calibrated phylogeny of Hymenoptera (2) and a comprehensive morphological dataset (346 discrete characters) including both adult (24) and larval (25) information.

We found that morphological evolution in Hymenoptera was marked by a fast and remarkable structural reorganization of the entire phenome affecting multiple anatomical complexes at once. Major clusters in the phenospace have been formed early on during the Hymenopteran evolution by strong directional selection, likely driven by the availability of open ecological niches. Evolution of larval and adult phenome is decoupled from each other and from diversification rates. These findings support a scenario of phenotypic diversification in Hymenoptera unfolding as a complex, multifaceted process, differently affecting distinct ontogenetic stages and multiple anatomical complexes but not immediately triggering an associated lineage diversification process.

Results

Rate evolutionary dynamics. 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. S8), with conspicuously higher rates estimated on branches leading to Vespina, Apocrita, and the clade including most Apocrita excluding Ichneumonoidea (Fig. 1A, arrowheads 1-3, respectively). For the larval phenome, higher rates were estimated on a few backbone branches preceding the emergence of Vespina/Apocrita (Fig. 1B, arrowhead 4). 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) (Fig. 1A; S1 Appendix, Figs S8, S10E), while for the larval phenome, higher rates were observed before the boundary between the Middle/Late Triassic (∼237 Ma) (Fig. 1B; S1 Appendix, Figs S8, S10G).

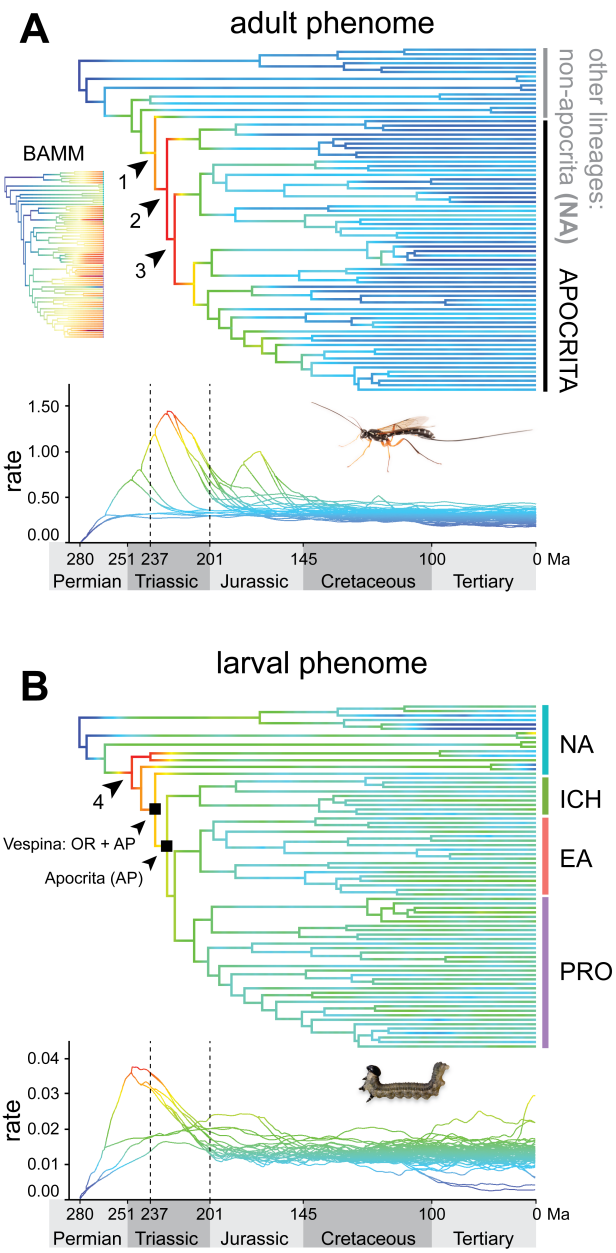

Fig. 1. Evolutionary rate dynamics inferred with ontophylo. *(A)* Branch-specific rates for the full adult phenome. *(B)* Same as (A) but for the larval phenome. In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rate-through-time variation for each branch (edgeplots); color gradient from blue to red indicates low to high rates respectively. Arrowheads (1-4) indicate some branches with conspicuously higher rates (see main text for explanations). The smaller indented tree in *(A)* shows net diversification rates inferred with BAMM obtained from (2). Vertical dashed lines indicate the limits of the Late Triassic. Legends: NA: non-Apocritan Hymenoptera; OR, Orussoidea; AP, Apocrita; ICH: Ichneumonoidea; EA: "Evaniomorpha"-Aculeata clade; PRO: Proctotrupomorpha. Image credits to Line Kræmer.

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 com-

plexes 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. S6). 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, Figs S6, S10A, C).

For all standard 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 S5, S7, S9, 10B,D,F,H).

Morphospace reconstructions. Reconstructed morphospaces for the full adult phenome were more concentrated in the standard analysis than in the null analysis when considering both ancestral and extant phenomic states (Fig. 3A-B; rightmost panel, gray and color-coded circles, respectively; note the different scales of the plots). On the other hand, morphospace occupation was not substantially different between standard and null analyses for the larval phenome (S1 Appendix, Figs S23-S24) 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 S19, S20 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 standard (S1 Appendix, Figs S19, S21) and null analyses (S1 Appendix, Figs S20, S22), 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 standard 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 analyses but accelerated and quickly became directional during the Late Triassic (225 Ma) on the path to Vespina and Apocrita in the standard analysis (Fig. 3A, arrowheads 1-2; S1 Appendix, Movie S1).

Similar results for the adult phenome were recovered when analyzing only sclerites (S1 Appendix, Figs S23-S24 and Movie S2); but less so when analyzing only muscles (S1 Appendix, Figs S23-S24 and Movie S3).

Macroevolutionary regimes. Results from the qualitative assessments of the adult and larval phenomes were complemented by our newly developed branch-specific metric to investigate regime of selection. In a nutshell, the core idea of the metric is to evaluate directionality in the evolution of multiple discrete traits (e.g., entire phenome) by calculating Hamming distances between states representing complex characters sampled in different points in time (e.g., start and end of a branch). In an analogy to the evolution of a continuous trait, non-directional evolution would be characterized by phenomic states gradually spreading in all directions (e.g., neutral evolution) or concentrated in a small portion of the morphospace (e.g., stabilizing selection), while directional selection would be characterized by directional walks in the morphospace (Fig. 3; analogy presented in the bottom-right diagram, 1-3 respectively). We emphasize that this metric refers to directional and stabilizing selection acting upon macroevolutionary scales. To assess significance of results, for each branch bootstrap resampling was used to obtain a distribution of differences between values calculated in standard and null analyses (see Materials and Methods, and S1 Appendix for details).

For the full adult phenome, differences were conspicuously higher and positive on the backbone branches leading to Vespina/Apocrita (Figs. 3A,D, orange-shaded branches, arrowheads 1-2 respectively; S1 Appendix, Figs S25A-B, S26A-B), thus indicating strong directional selection. Differences were less pronounced but also positive on branches leading to other major lineages of Hymenoptera (e.g., Ichneumonoidea, Proctotrupomorpha, and lineages within Aculeata). Differences were moderate-to-higher, but with negative values, on several lineages outside Vespina/Apocrita and long branches leading to extant taxa across Hymenoptera, thus indicating stabilizing selection (Figs. 3A,D, purple-shaded branches; S1 Appendix, Figs S25A-B, S26A-B). Results from individual anatomical complexes of the adult were more varied, but largely agree with that of the full phenome, particularly for mesosomal structures (e.g., mesonotum, mesopectus, metapectal-propodeal complex)

For the larval phenome, higher differences and positive values, thus strong directional selection, were observed on lineages outside Vespina/Apocrita, but conspicuously higher and negative within Tenthredinoidea (indicated with a star), and either weakly positive/negative or non-significant across most lineages within Apocrita (Fig. 3E; S1 Appendix, Figs S25C-D, S26C-D).

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 anatomical complex and the larval phenome (indicated with an asterisk) when considering raw rates (Fig. 4A-C; S1

DRAFT

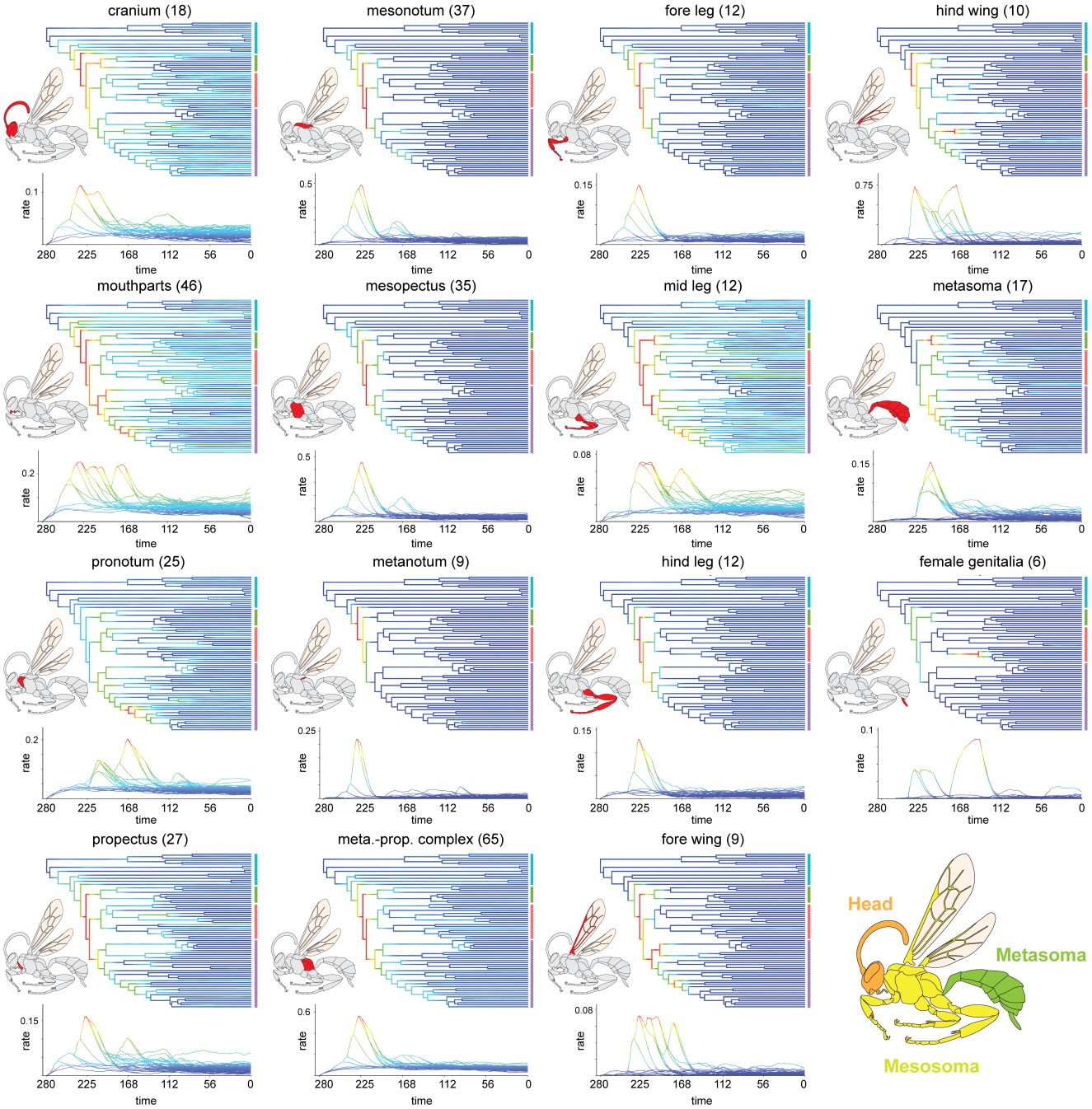

Fig. 2. Branch-specific evolutionary rate dynamics inferred with ontophylo. Each subpanel shows the result for an individual anatomical complex of the Hymenoptera anatomy (highlighted in red in the anatomical diagrams). In each subpanel, the top figure shows tree branches shaded by absolute rate values (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.

Appendix, Tables S2 and S4). Results based on rate contrasts (deltas) were more conservative, but still including mesosomal structures, with significant correlations observed only for the mesopectus, metapectal-propodeal complex, and hind leg (S1 Appendix, Tables S3 and S5).

Employing the same approach mentioned above, we also found that phenotypic rates of evolution were significantly correlated with our new metric of regime of selection, with high rate values associated with directional selection for all anatomical complexes of the adult phenome (individually: S1 Appendix, Table S6; or pooled together: Fig. 5C), but not for the larval phenome. These results support that, for this particular Hymenoptera dataset, the evolutionary rate dynamics can also be used to infer selection regimes, as previously suggested for other groups (13, 26).

Diversification rate correlations. Using the resampling-based approach, we also investigated the association of branch-

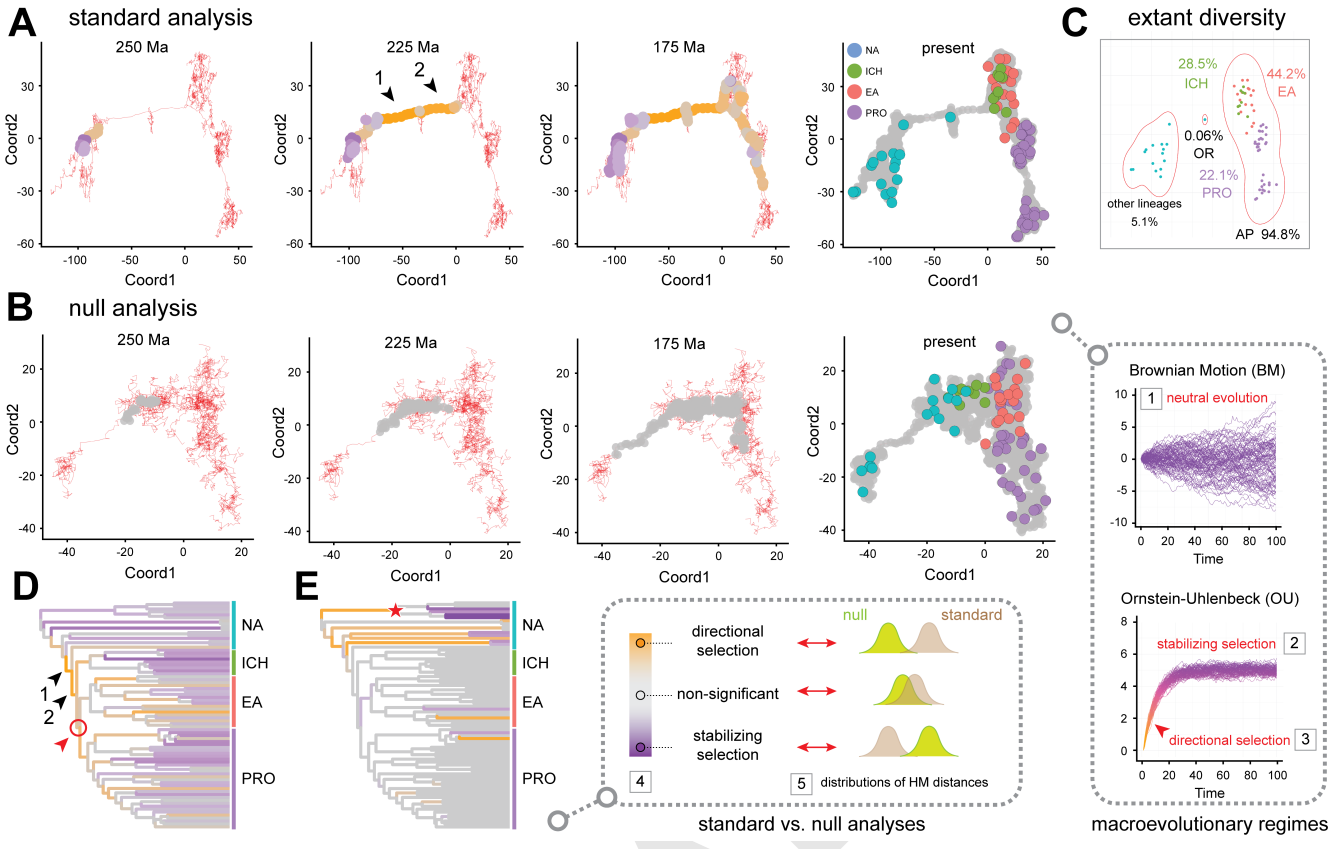

Fig. 3. Morphospaces reconstructed with ontophylo and macroevolutionary regimes. *(A)* Snapshots in different moments in time of the morphospace of the full adult phenome for the standard analysis; *(B)* Same as (A) but for the null analysis. *(C)* Present-day estimated relative diversity of major lineages of Hymenoptera. *(D)* Macroevolutionary regimes according to the new metric proposed here, calculated as the difference between the standard and null analyses for the full adult phenome. *(E)* Same as (D) but for the larval phenome. In (D-E) and the first three panels in (A), from left to right, orange color denotes directional selection; purple color denotes stabilizing selection; gray color denotes non-significant results. In all panels in (B) and the rightmost in (A), gray circles indicate ancestral phenomic states; other colors indicate states observed in extant lineages (as explained in the legends). Bottom-right box: analogy for interpreting the evolution of multiple discrete traits representing the entire phenome as a single continuous trait; non-directional walks in the morphospace would be indicative of either neutral evolution or stabilizing selection (1 or 2); directional walks would be indicative of directional selection (3); (4) explanation of colors in (D-E) and (5) corresponding distribution of Hamming distance values from the samples of stochastic maps in the standard and null analyses. Arrowheads and numbers (1-2) in (A) and (D) indicate some branches for which strong directional selection is supported (see main text for explanations). Red arrowhead and circle in (D) indicate the shift in lineage diversification rates supported in some of the BAMM analyses of (2). Red star indicates the clade of Tenthredinoidea. Phylogenetic legends: OR: Orussoidea; NA: non-Apocritan Hymenoptera; ICH: Ichneumonoidea; EA: "Evaniomorpha"-Aculeata clade; PRO: Proctotrupomorpha.

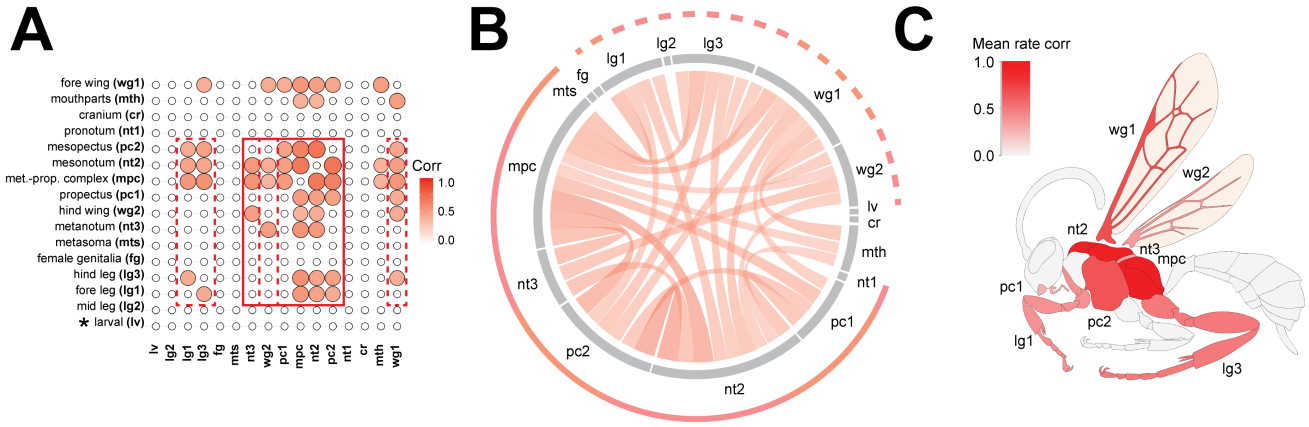

Fig. 4. Rate correlations among anatomical complexes (and larval phenome). *(A)* Correlogram. *(B)* Connectogram. *(C)* Anatomical diagram color-shaded by the mean correlation for each anatomical complex across all partitions (i.e., row means in the correlation matrix of the correlogram). Color intensity 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-C) as indicated in (A). Asterisk indicates the larval phenome.

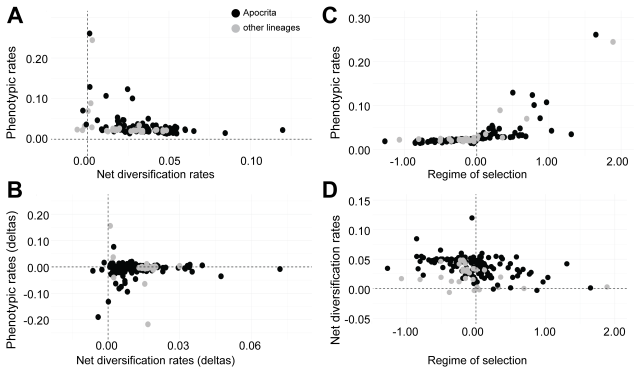

Fig. 5. Scatter plots showing the association between phenotypic and lineage diversification rates and selection regimes; data points correspond to mean values for each branch across the pooled information from all individual anatomical complexes. *(A)* Association between phenotypic and lineage diversification rates using raw rates. *(B)* Same as in (A), but using rate contrasts (deltas). *(C)* Association between phenotypic rates and regimes of selection. *(D)* Association between lineage diversification rates and regimes of selection.

specific rates of phenotypic evolution and regimes of selection with the branch-specific net diversification rates extracted from the BAMM analyses of Blaimer et al. (2). Based on the results using raw rates, significant negative correlations were observed for some anatomical complexes of the mesosoma (e.g., propectus, mesonotum, mesopectus, metapectal-propodeal complex; individually: S1 Appendix, Table S7; pooled together: Fig. 5A) but none when using rate contrasts (deltas) (individually: S1 Appendix, Table S8; pooled together: Fig. 5B). While intriguing, such results are not entirely surprising since many short internal branches inferred with low net diversification rates in the BAMM analyses from (2), for example the ones leading to Vespina and Apocrita, were inferred with higher phenotypic rates for several mesosomal anatomical complexes. Similar results were observed when contrasting regimes of selection and lineage diversification rates, with a few significant negative correlations also observed for some individual mesosomal structures (S1 Appendix, Table S9) but patterns are less clear when all complexes are pooled together (Fig. 5D).

Metrics of phenomic evolution. To describe different aspects of phenomic evolution, in addition to the metric to investigate regime of selection, we developed two additional metrics to assess (i) phenotypic innovation and (ii) phenotypic disparity using the information obtained from stochastic map samples (see Materials and Methods, and S1 Appendix for details).

Curves of phenotypic innovation through-time in the standard 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). Similar results were observed for all body regions and most anatomical complexes of the adult phenome (S1 Appendix, Figs S11A, C). Similar patterns were also observed for the disparity-through-time curves (Fig. 6E; S1 Appendix, Figs S18A, C). Furthermore, our results indicate that relatively more innovations were accumulated in Vespina/Apocrita than in the remaining Hymenoptera (S1 Appendix, Figs S12, S14, S16).

As for the larval phenome, both curves of phenotypic innovation and disparity-through-time in the standard analyses showed similar patterns, with slightly steeper accumulation of new character-states (Fig. 6D) and phenotypic divergence (Fig. 6F) before the boundary between the Middle/Late Triassic (∼237 Ma). Accumulation of innovation was more evenly distributed across Hymenoptera (S1 Appendix, Fig. S16).

All curves of phenotypic innovation and disparity-through-time showed similar patterns in the null analyses, with accumulation of new character-states and phenotypic divergence increasing gradually towards the present (Fig. 6C-F, dotted lines; S1 Appendix, Figs S11B, D; S18B, D). Likewise, accumulation of innovations was more evenly distributed across Hymenoptera (S1 Appendix, Figs S13, S15, S17).

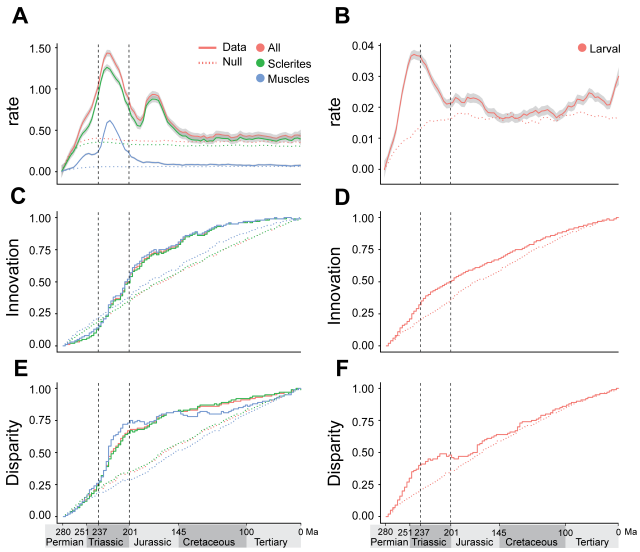

Fig. 6. Evolutionary rates and metrics of phenomic evolution. *(A)* Rate-through-time curve for the adult phenome (all characters, only sclerites, and only muscles). *(B)* Same as (A) but for the larval phenome. *(C)* Innovation-through-time curve for the adult phenome (all characters, only sclerites, 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 standard analyses and dotted lines results from the null analyses; vertical dashed lines indicate the limits of the Late Triassic.

Discussion

Phenome macroevolutionary dynamics. Notoriously, phenotypic rates in the adult and larvae were found to be decoupled from each other. This finding is well-aligned with the idea that the evolution of distinct life stages in Holometabola allowed larvae and adults to explore different ecological niches thus reducing intra-specific competition and potentially prompting lineage diversification (27). The absence of any observed correlation between the anatomical complexes of the adult and the larval phenome indicates that traits in these different life stages might be genetically and developmentally decoupled, thus allowing them to better-respond to divergent selection pressures associated with distinct ecological niches, as predicted by the "adaptive decoupling hypothesis" (28). Furthermore, in insects with complete metamorphosis (Holometabola), e.g., Hymenoptera, it has been suggested that modularity, associated with the independence of larval and adult phenotypes, may promote higher evolvability

compared to other lineages of insects with highly correlated life stages and incomplete metamorphosis (19) but this was never tested using contemporary methods. Here we found further support for the adaptive decoupling hypothesis based not only on the temporal and phylogenetically decoupling of phenotypic evolutionary rates but also distinctive selection regime patterns observed for larvae and adult. Nonetheless, we acknowledge that the larval dataset is relatively small and information on most lineages within Apocrita is lacking, thus additional studies are needed to further assess the generality of our conclusions.

Even more intriguing is the fact that no clear positive association between phenotypic and lineage diversification was found neither for any anatomical complex of the adult nor for the larval phenome. Although intriguing, such finding is not surprising. Some recent studies on different groups of vertebrates have found no clear association between phenotypic and lineage diversification rates, for instance in mammals (11) and snakes (12); other studies support a decoupling between the time of origin of major groups, usually associated with a burst of phenotypic innovation and increased rates, and the putative period of the adaptive radiation, as observed during the early evolution of lepidosaurs (13, 16), or the reverse, with increased phenotypic rates observed only later on, as observed in placental mammals (29), or even more complex scenarios, where different cranial modules exhibit different patterns as observed in birds (14). In an attempt to reconcile these various findings, a recent paper on cranial morphology of frogs proposed a new framework for interpreting phenotypic and lineage diversification rates (30) where adaptive and non-adaptive radiations are possible, among other scenarios.

While phenotypic rates are not correlated with diversification rates, we have demonstrated that rates for several anatomical complexes of the mesosoma are correlated among themselves. Correlations among traits resulting in their organization into semi-autonomous modules may promote or hinder adaptation depending on whether the vector of selection is aligned or not with the major axis of variation of the multidimensional phenotype (17). Theoretical predictions of the association between integration, evolutionary rates, and disparity for continuous traits evolving under a multivariate Brownian motion state that the patterns and magnitude of trait correlations constrain the shape of the morphospace while evolutionary rates can vary freely (17). On the one hand, some empirical results have shown that a modular pattern of organization of the anatomy facilitates mosaic evolution, with different cranial modules potentially evolving with different tempo and mode across lineages in birds (14) and placental mammals (18). On the other hand, empirical support for the relationship between disparity and within-module integration is less clear, as observed in birds (14) and frogs (10). Furthermore, the association between phenotypic rates and disparity is also not straightforward, since increased evolutionary rates have been observed in the morphologically homogeneous cranium of extant crocodyloids (31), for example, thus indicating rampant convergence to a limited set of phenotypes, or variation in a clade-dependent fashion, as observed in birds (32).

In these regards, our findings are remarkable. To our knowledge, it is the first time that an attempt was made to investigate the associations between phenotypic rates and integration, phenotypic disparity and novelty, and lineage diversification exploring a large-scale dataset of discrete traits including adult and larval characters and focusing on a mega-diverse order of insects, namely Hymenoptera. Furthermore, it is the first time that such associations are linked to potentially divergent selection regimes across lineages. In Hymenoptera, we found that the evolutionary singular event marking the transition from the ground-plan mesosomal morphology to the modified wasp-waisted morphology on the branch leading to Apocrita was characterized by fast evolutionary rates and strong directional selection in addition to a steeply increase in disparity across lineages and accumulation of evolutionary novelties. Such findings in conjunction with the observed pattern of correlations among mesosomal structures suggest that this evolutionary singular event constitutes a unique example of a scenario where integration among multiple discrete traits facilitates adaptive evolution. In this case, the selection pressures seem to be aligned with the major axis of variation. Nonetheless, it still remains to be addressed how interactions among multiple factors such as evolutionary rates, phenotypic integration, and selection regimes could have shaped the adaptive landscape in Hymenoptera at a macroevolutionary level.

Wasp waist as an evolutionary singularity. Arguably, the most important phenotypic transition along the evolutionary history of Hymenoptera was the formation of the "wasp waist" in Apocrita (33). Instead of a simple character-state transformation, the evolution of the wasp waist constituted a remarkable reorganization of the entire adult phenome of Hymenoptera. It involved the complete morpho-functional integration of the first abdominal segment into the thorax, a further development of the fusion of the first abdominal tergum with the metathorax (a putative ancestral condition in Holometabola; (34)), resulting in the formation of a new articulation and two new body regions: mesosoma (thorax + first abdominal segment) and metasoma (remaining abdomen) (35, 36). This apparently simple evolutionary change characterized a major morpho-functional shift in the articulation between thorax and abdomen resulting in the increased mobility of the third tagma likely associated with oviposition and parasitoid habit in Apocrita (35, 36).

We hypothesize that a possible consequence of the formation of the new wasp waist, in association with the emergence of parasitoidism, was an increased functional demand in the mesosoma for locomotion, since apocritan parasitoid adults are probably more active than other hymenopteran insects as they engage in more varied foraging behaviors in search for hosts. Our results corroborate this morpho-functional evolutionary hypothesis, since not only higher evolutionary rates were observed for several mesosomal anatomical complexes of the adult in the branches leading to Vespina/Apocrita (Fig. 1A) but also rates were correlated for most core-structural components (e.g., mesonotum, mesopectus, metapectal-propodeal complex; Fig. 4A-B, solid box and arch) and appendicular components (e.g., fore and hind legs, fore and hind wings; Fig. 4A-B, dashed boxes and arch). This morpho-functional evolutionary hypothesis is further corroborated by evidence from the internal anatomy of adults, which shows that evolutionary rates for muscle characters (e.g., mostly associated with the mesosoma) were also conspicuously higher in the branches leading to Vespina/Apocrita (S1

DRAFT

Appendix, Fig. S8). Thus the evolutionary transition of thorax/abdomen to mesosoma/metasoma in Hymenoptera is no less impressive than the dramatic transformations observed in other "evolutionary singular events" such as the evolution of wings and complete metamorphosis in pterygote and holometabolous insects (33); fin-to-limb transition in early tetrapodomorphs (13); fore-limb and other skeletal modifications associated with flight in birds (15); and loss of limbs and body elongation in snakes (12).

Apocritan wasp-waisted wasps make their first appearance in the fossil record in the Early Jurassic (e.g., (33); fig. 11.9), less than 200 Ma ago. Known pre-Cretaceous hymenopteran fossils are almost exclusively compressions, making it difficult to observe anatomical details around the mesosoma-metasomal boundary. Nevertheless, A.P. Rasnitsyn and co-workers (37, 38) based on comparatively well-preserved fossils from the Jurassic Haifanggou and Karabastau Formations claimed to observe a smooth transition from Xiphydrioidea and Orussoidea to Apocrita through a number of fossil taxa. They considered the wasp waist to have evolved independently among stem apocritan taxa up to three times (38). However, these hypotheses rely heavily on interpretations of the fossils, to the exclusion of most extant sources of phylogenetic information. They are difficult to reconcile with more detailed anatomical observations in extant taxa, and the phylogenetic results derived from increasingly comprehensive molecular datasets.

Phenomic adaptive landscape of Hymenoptera. Interpreting patterns of morphospace occupation through time in association with different evolutionary factors is not straightforward. On the one hand, revisionary studies have suggested that both magnitude and patterns of trait correlations influence the shape of morphospace while phenotypic rates do not (17). On the other hand, a few studies based on explicit mathematical models and simulations have demonstrated that morphospace structure can be explained by the interplay between phylogenetic structure and logistic growth (e.g., saturated lineage diversification) in a diversification process or by "developmental entrenchment" (39) or be a consequence of the geometrical properties of the morphospace itself (40).

Here, we try to offer a conciliatory interpretation of the macroevolutionary adaptive landscape in Hymenoptera based on our empirical results. The concentration of lineages in particular areas of the morphospace suggests the influence of phylogenetic structure, since lineages pertaining to the same major clades tend to occupy similar areas of the morphospace. Additionally, the observed concentration of lineages in the standard analyses relative to the null analyses is congruent with a scenario of developmental entrenchment/constraints also shaping morphospace, as suggested by correlations among evolutionary rates of several mesosomal structures, thus supporting trait integration as an important factor shaping the macroevolutionary landscape of Hymenoptera.

Conclusion

Our study shows that the phenotypic diversification in Hymenoptera was a complex, multifaceted process, affecting distinct ontogenetic stages (adults and larvae) and multiple anatomical complexes at once resulting in an evolutionary singular event that set the fundamental morphological background that likely enabled the later adaptive radiation of lineages within Apocrita. Additionally, our study demonstrates for the first time the effects of phenotypic integration in potentially facilitating directional evolutionary change with an empirical example including a dataset with hundreds of discrete traits for a megadiverse order of insects.

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 (41, 42), 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 in six groups according to the types of dependencies observed (Dataset S3, G1-6) and coded/recoded accordingly following the recommendations of (41–43) (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 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 manually annotated with terms from the Hymenoptera Anatomy Ontology (HAO) (44) (Dataset S3) (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).

Phenomic analyses. This section refers to the "standard analyses" (i.e., analyses conditioned on observed data at tips). Phenomic evolution was 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, 45). 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, standard model fitting was performed using functions implemented in the R package corHMM (46). For the dependent characters, transition rate matrices (Q matrices) were initially set using functions from the R package rphenoscate (43) and then parameters estimated with corHMM. ASR analyses employed stochastic character mapping (23) as implemented in

DRAFT

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 (44) 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 "standard analyses" described above with those from the "null analyses" (see section "Simulation analyses").

Metrics to assess phenomic evolution and selection regime. 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. XXXXX). 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. XXXXX). Selection regime was assessed for each map and each tree branch by calculating the Hamming distance between first and last amalgamated character-states sampled on that branch as a proxy for inferring directional selection (see explanatory diagrams in S1 Appendix, Fig. XXXXX). To assess the significance of Hamming distances calculated for each branch, bootstrap resampling (sample size = 1000, 10.000 replicates) was used to obtain a distribution of differences between values calculated in the standard and null analyses (see S1 Appendix for details). Significance of values was determined by the non-overlap of the 95% CI with zero.

Rate correlations. Mean rates were calculated for each branch of the phylogeny for each anatomical complex defined in L1 (15 anatomical partitions) and the larval 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 investigate for possible effects of non-independence of rates calculated for branches on the same path from root to tip, contrasts were also used. Rate contrasts (deltas) were calculated as the difference between the mean rates of sequential branches on the path from root to tip. 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 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 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 (47, 48). For each pairwise comparison, a stochastic map was randomly selected from the samples of both anatomical partitions, branch rates (raw or delta) 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 (raw or delta) from the standard 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 standard 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.

Simulation 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 standard 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 standard analyses, but not conditioned on character-states observed at the tree tips. Simulations used functions implemented in the R package phytools (49). 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 investigate rate evolutionary dynamics and morphospaces for all anatomical complexes, body regions, and adult and larval phenomes; to assess the significance of results from the metric of regime of selection; and to assess the significance and strength 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. B Ebenman, Niche differences between age classes and intraspecific competition in age-structured populations. *J. theoretical biology* **124**, 25–33 (1987).
28. B Ebenman, Evolution in organisms that change their niches during the life cycle. *The Am. Nat.* **139**, 990–1021 (1992).
29. TJ Halliday, et al., Rapid morphological evolution in placental mammals post-dates the origin of the crown group. *Proc. Royal Soc. B* **286**, 20182418 (2019).
30. G Morinaga, JJ Wiens, DS Moen, The radiation continuum and the evolution of frog diversity. *Nat. Commun.* **14**, 7100 (2023).
31. RN Felice, D Pol, A Goswami, Complex macroevolutionary dynamics underly the evolution of the crocodyliform skull. *Proc. Royal Soc. B* **288**, 20210919 (2021).
32. CR Cooney, et al., Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature* **542**, 344–347 (2017).
33. D Grimaldi, MS Engel, *Evolution of the Insects*. (Cambridge University Press), (2005).
34. F Friedrich, RG Beutel, The thoracic morphology of nannochorista (nannochoristidae) and its implications for the phylogeny of mecoptera and antliophora. *J. Zool. systematics evolutionary Res.* **48**, 50–74 (2010).
35. 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).
36. 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).
37. AP Rasnitsyn, Z Haichun, Early evolution of apocrita (insecta, hymenoptera) as indicated by new findings in the middle jurassic of daohugou, northeast china. *Acta Geol. Sinica-English Ed.* **84**, 834–873 (2010).
38. L Li, C Shih, AP Rasnitsyn, D Ren, New fossil ephialtitids elucidating the origin and transformation of the propodeal-metasomal articulation in apocrita (hymenoptera). *BMC Evol. Biol.* **15**, 1–17 (2015).
39. MR Pie, JS Weitz, A null model of morphospace occupation. *The Am. Nat.* **166**, E1–E13 (2005).
40. S Gavrilets, Dynamics of clade diversification on the morphological hypercube. *Proc. Royal Soc. London. Ser. B: Biol. Sci.* **266**, 817–824 (1999).
41. 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).
42. S Tarasov, New phylogenetic markov models for inapplicable morphological characters. *Syst. biology* **72**, 681–693 (2023).
43. DS Porto, et al., rphenoscate: An r package for semantics-aware evolutionary analyses of anatomical traits. *Methods Ecol. Evol.* **14**, 2531–2540 (2023).
44. MJ Yoder, I Mikó, KC Seltmann, MA Bertone, AR Deans, A gross anatomy ontology for hymenoptera. *PloS one* **5**, e15991 (2010).
45. S Tarasov, The invariant nature of a morphological character and character state: insights from gene regulatory networks. *Syst. biology* **69**, 392–400 (2020).
46. 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).
47. C Parins-Fukuchi, GW Stull, SA Smith, Phylogenomic conflict coincides with rapid morphological innovation. *Proc. Natl. Acad. Sci.* **118**, e2023058118 (2021).
48. 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).
49. 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 "standard 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 standard and null analyses are likely due to the common tree and/or models while those exclusive to the standard analyses are likely to be driven by observed data alone.

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 rates, and selection regimes. Results from standard 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 characters were removed since they were non-informative for the taxonomic sampling of this study: CH9, CH101, CH122, CH193, CH278, CH374. Additionally, characters CH17, CH352, CH361-362, CH369, CH384, CH388 were removed since they included polymorphic states, which can be difficult to model properly in case of dependencies; CH391, 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 221-236 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 222, 225, 231, 232, 233, and 236 were removed since they were non-informative for the taxonomic sampling of this study. The original character 228 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 can 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 of each other when the state observed in one character does not influence the probability of observing a given state of 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, CH 18: 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. 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 on 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 real morphological structures referred to as mandibles in real biological specimens, is part of some 'head'. This may seem a 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 manually annotated (by IM and DSP) with terms from the Hymenoptera Anatomy Ontology (HAO) (8) (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. Here, 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. (9). 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. (9), 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 (10) 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, 10)), 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* (11). 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 (12) 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 (Figure 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 (Figure 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 (Figure 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* (13). 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 into 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 different metrics to capture two key 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 bases in a DNA sequence. Along the branches of the phylogeny, character-state changes can occur at any 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 diagrams Fig. XXXXX).

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 diagrams Fig. XXXXX).

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 standard 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 neutral or non-directional evolution (see explanatory diagrams Fig. XXXXX).

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 S11. 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 S18. In both cases, values were rescaled between 0 and 1.

When contrasting standard and null analyses, qualitative assessments were used to compare metrics of phenomic evolution resulting in the curves shown in Fig. 6C-F and Figs. S11 and S18. 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 standard and null analyses (1000 branches each); the metric was calculated and the mean obtained for the standard 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 standard 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 was possible to investigate not only how the evolutionary dynamics of different portions of the phenome varied along the Hymenoptera phylogeny but also dissect their potential associations with lineage diversification rates. To investigate these processes, an assessment of potential correlations between phenotypic rates of evolution and between phenotypic and lineage diversification rates was performed. Furthermore, potential correlations between our new metric of regime of selection and phenotypic and lineage diversification rates were also investigated. 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. All comparisons between the phenotypic and lineage diversification rates were performed resulting in 16 pairs of comparisons. Additionally, all comparisons between the phenotypic rates and the metric of regime of selection calculated for the same 16 anatomical partitions (plus larval phenome) were also performed as well as between the metric of regime of selection and lineage diversification rates. Lineage diversification rates were obtained from the net lineage diversification rates extracted from the BAMM analyses of Blaimer et al. (9) (see R scripts in the Supporting Data for more details).

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 the metric of regime of selection. The second strategy used linear regressions including rates calculated from both the standard and null analyses. This 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 complex, 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. To account for possible non-independence of rates calculated in sequential branches on the same path from root to tip, rate contrasts (deltas) were also calculated. Rate contrasts were calculated as the difference between the mean rate of a branch `i+1` minus the mean rate of branch `i`, where `i = 1, 2, 3, ... n`; `i` is the index of a branch in the path from the root to the tip; and `n` is the total number of branches in the path.

In the first strategy, for each iteration, the Spearman's rank correlation was calculated between the branch rates/deltas 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 (14, 15).

In the second strategy, for each iteration and pair being compared (e.g., anatomical complex 1 and 2), branch rates/deltas were extracted from the stochastic maps obtained in the standard (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 coefficient was converted to obtain a value of Pearson's correlation. The rationale for 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 and 16 comparisons between phenotypic and lineage diversification rates using raw rates and deltas. In the case of the first strategy, resampling was also performed for the rates/deltas 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 standard 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 tree, character histories were simulated using the same models and parameters (i.e., Q matrices) fitted with *corHMM* for

the standard analyses, but not conditioned on the observed character-states at tips. Simulations used the function `sim.history` from the R package *phytools* (16). 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 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 (Fig. S2D), those based on the mean partition sizes were between 0.1 and 0.5 for the three partitions. However, when based on the observed partition sizes, the rate was conspicuously higher for the mesosoma ($0.7 \sim 0.8$) 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 (S2A-B: P4, P7, P10, P14-15), in others, rates seem to be higher (S2B: S2, S8) or lower (S2A-B: 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 when using raw rates (Table S2; Pval_prop: proportion of significant p-values equal to or greater than 0.95) and three pairs when using deltas (Table S3) for the approach based on permutations (strategy 1). All pairs recovered when using deltas were congruent with those recovered when using raw rates. As for the approach based on linear regressions (strategy 2), 27 pairs presented significant correlations between phenotypic rates when using raw rates (Table S4) and three when using deltas (Table S5). All pairs recovered with deltas were also congruent with those recovered with raw rates. 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 when using raw rates and between 0.46∼0.52 when using deltas in strategy 1 (Tables S2-3); between 0.39∼0.70 (0.39∼0.70 if uncorrected) when using raw rates and between 0.47∼0.52 (0.47∼0.52 if uncorrected) when using deltas in strategy 2 (Tables S4-5). Finally, when comparing correlation values obtained for the congruent pairs in the standard and null analyses (i.e., only for strategy 1), the distributions of values sampled in the permutation tests are clearly distinct (Fig. S4).

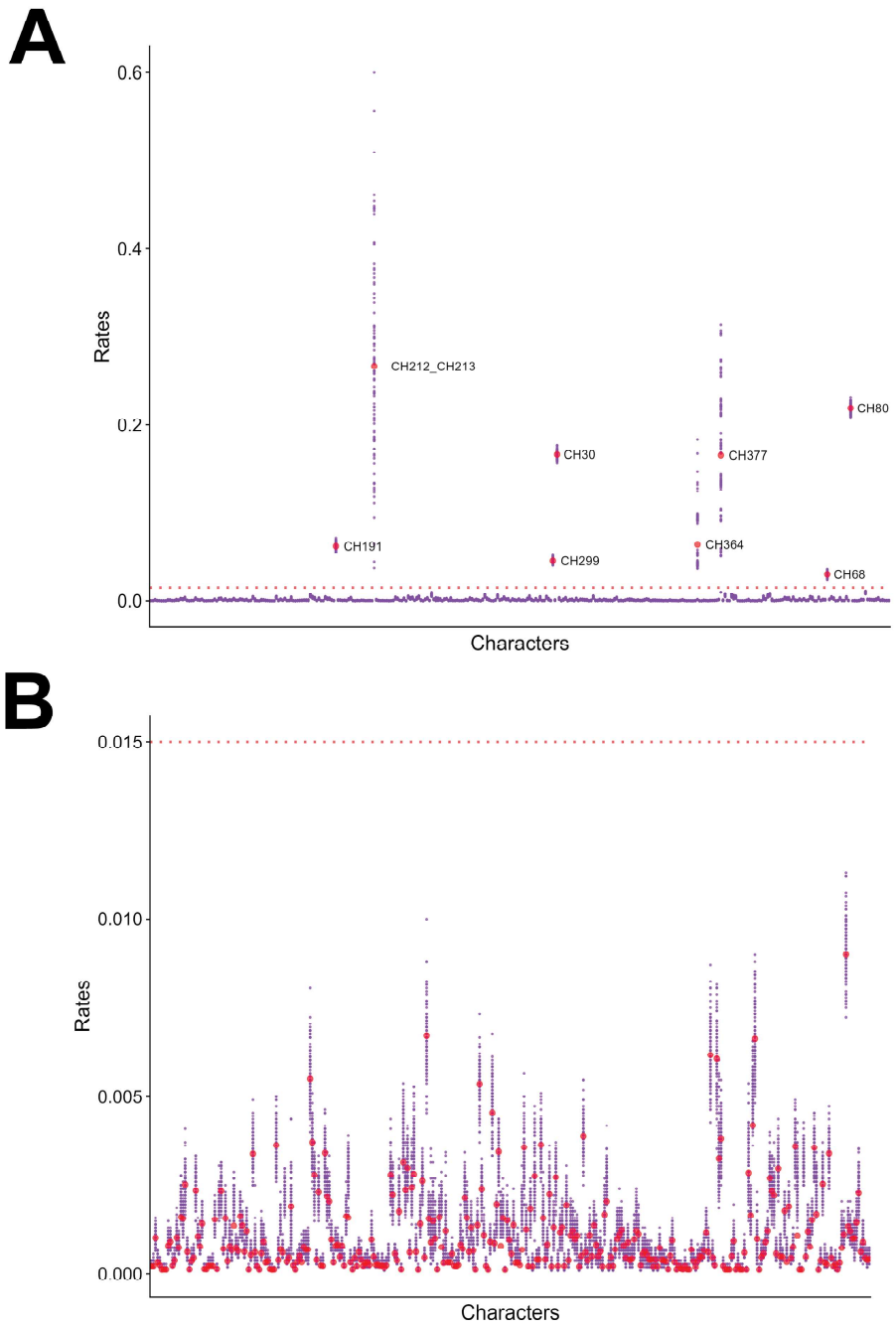

Fig. S1. Mean rates calculated from sampled stochastic maps for each individual character. *(A)* Dataset including all characters. *(B)* Dataset excluding outliers. Empirical rates calculated from observed changes on stochastic maps (purple dots). Mean rates estimated with *corHMM* (red dots; mean taken 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.

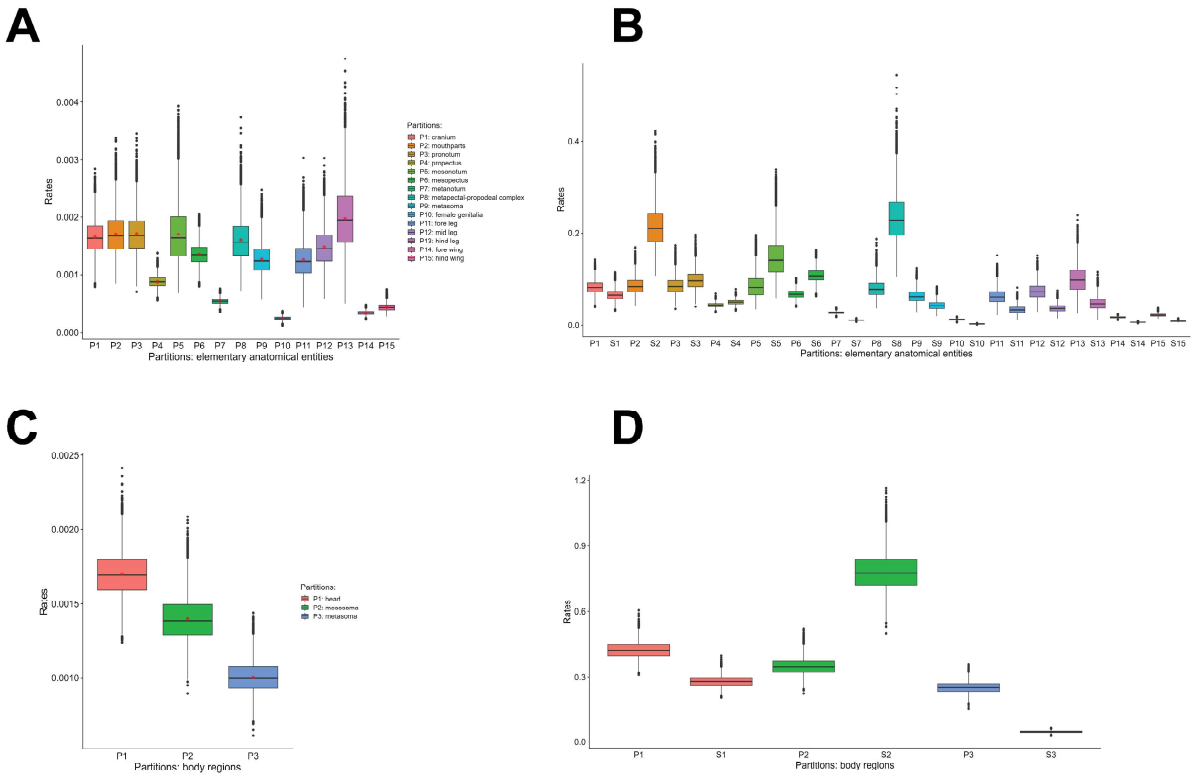

Fig. S2. Effect of partition size on estimated rates**.** *(A)* Mean rate of individual characters pertaining to each partition for each anatomical complex**.** *(B)* Mean partition rates rescaled by mean partition sizes (P1-15) and observed partition sizes (S1-15) for each anatomical complex**.** *(C)*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

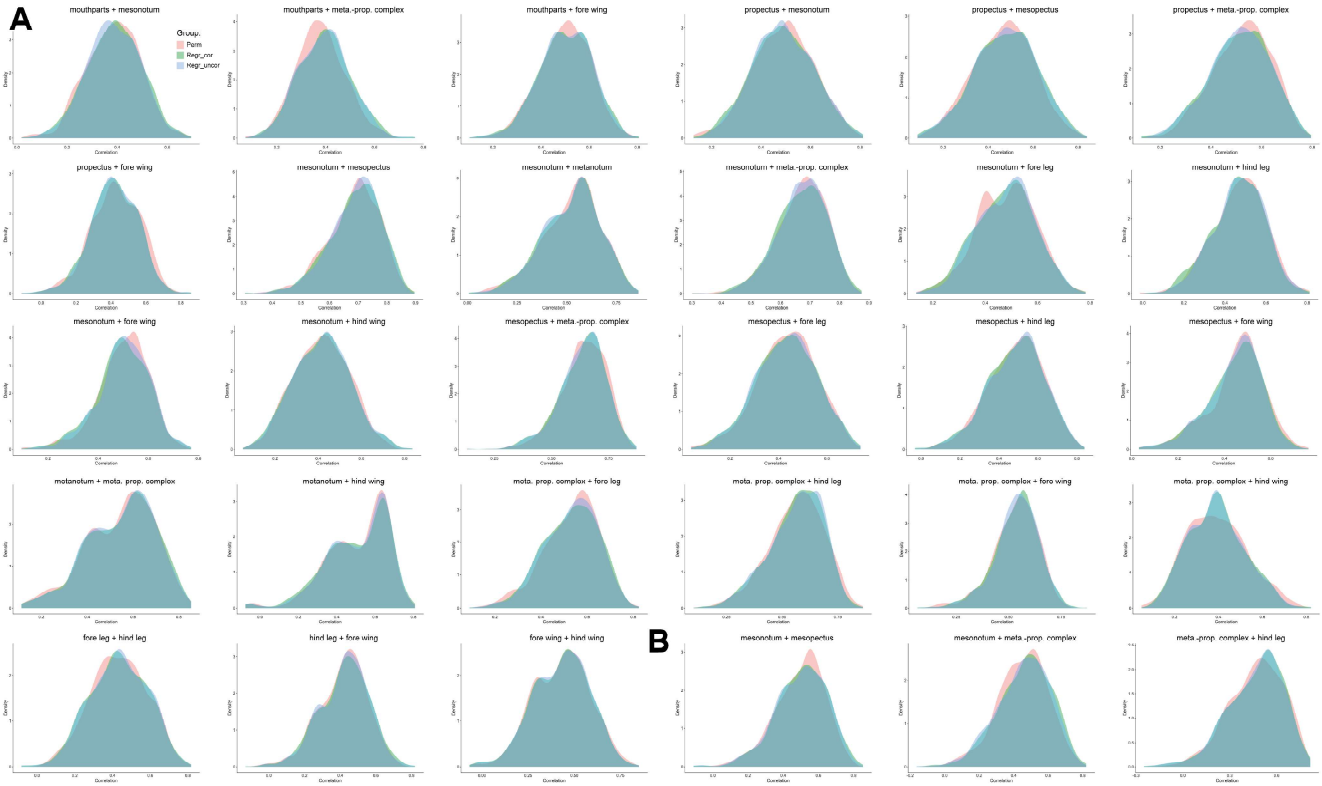

Fig. S3. Distributions of correlation coefficient values obtained through the resampling (permutation) and regression strategies for assessing correlations among the phenotypic rates of evolution**.** *(A)* Pairs showing significant correlations as obtained when using raw rates**.** *(B)* Same as (A), but obtained when using deltas**.** Legends: Perm: permutation strategy; Regr_cor: regression strategy using both rates from the standard and null analyses (corrected); Regr_uncor: regression strategy using only rates from the standard analysis (uncorrected)**.**

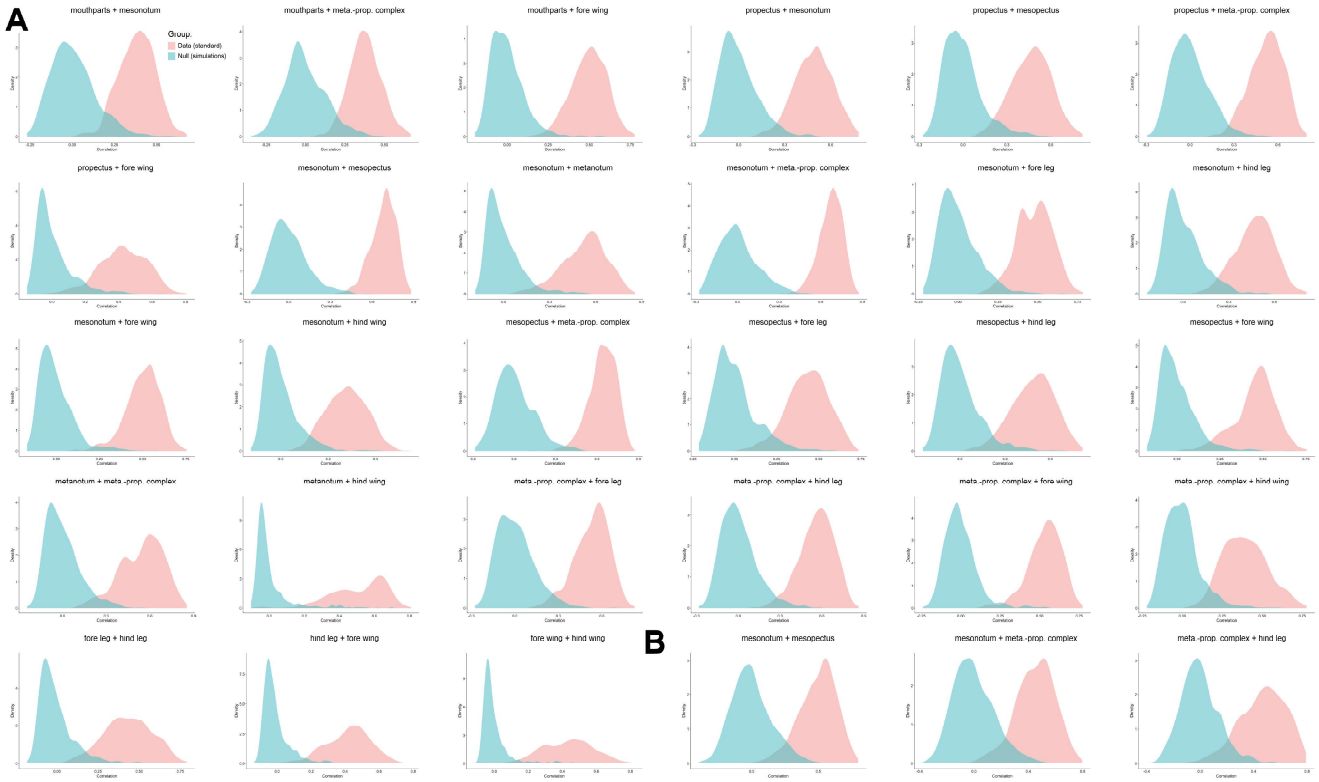

Fig. S4. Distributions of correlation coefficient values obtained through the regression (permutation) strategy. *(A)* Pairs showing significant correlations as obtained when using raw rates. *(B)* Same as (A), but obtained when using deltas.

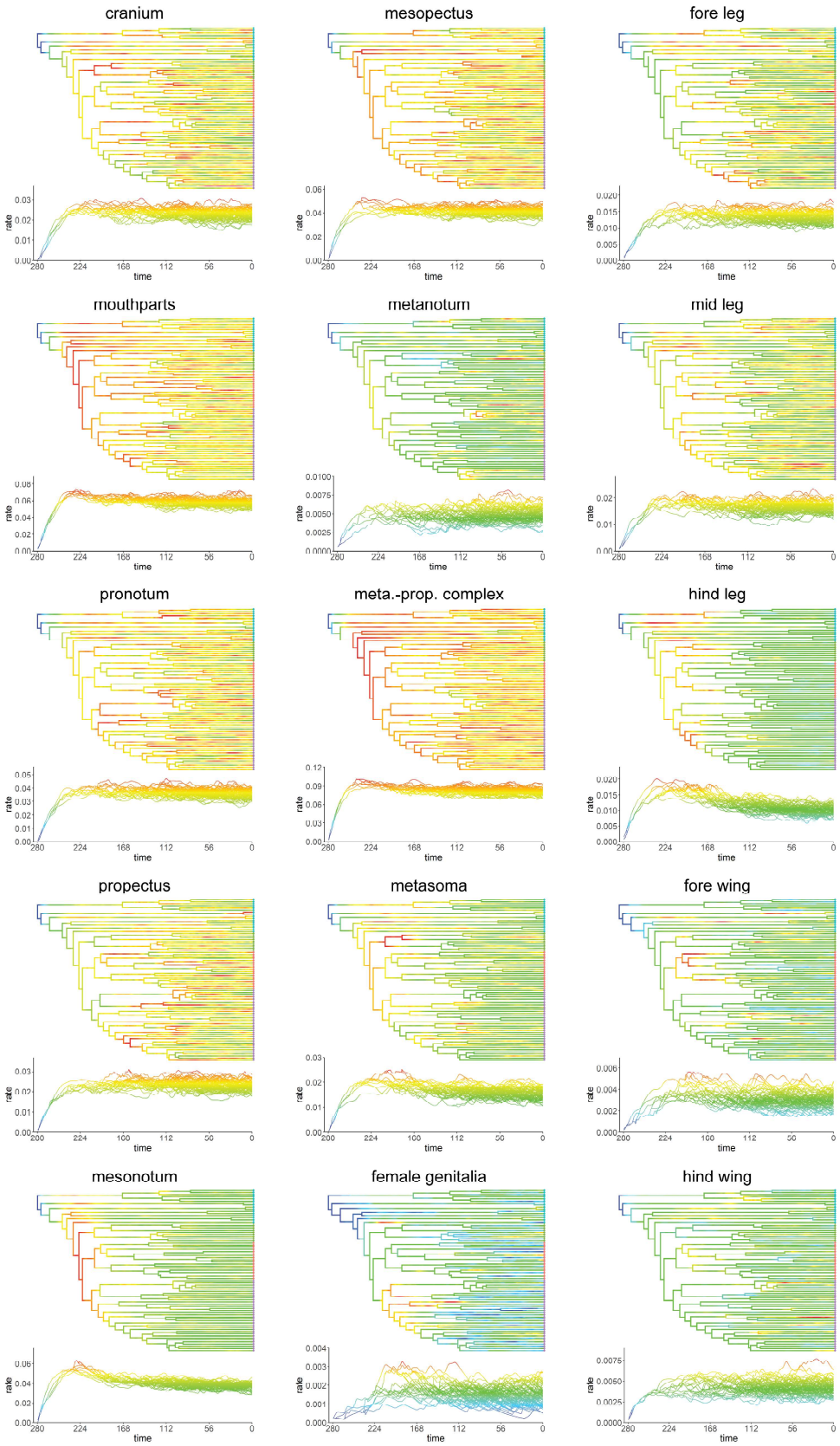

Fig. S5. Branch-specific rate dynamics inferred with ontophylo for the null analyses. Each subpanel shows the result for an elementary anatomical complex of the Hymenoptera anatomy. In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rates-through-time for each branch (edgeplots). Color gradient from blue to red indicates low to high rates respectively. Small filled circles at tips indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata clade; purple: Proctotrupomorpha.

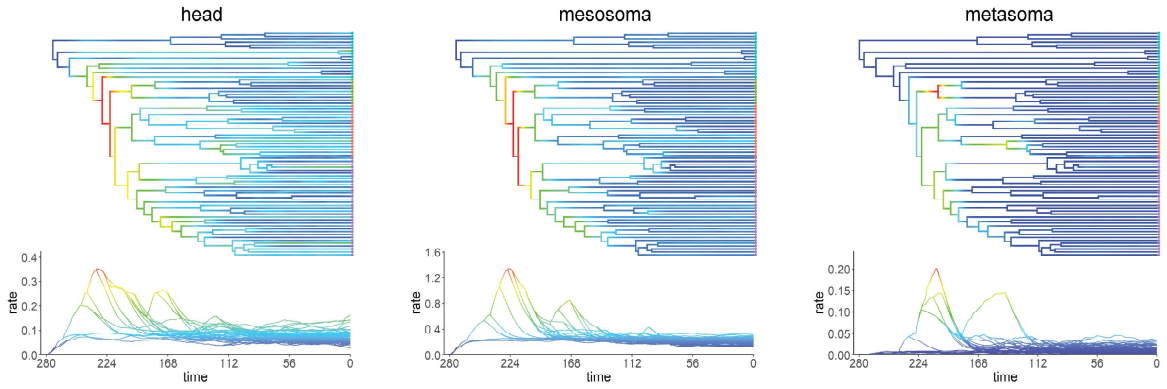

Fig. S6. Branch-specific rate dynamics inferred with ontophylo for the standard analyses**.** Each subpanel shows the result for a body region of the Hymenoptera anatomy. In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rates-through-time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

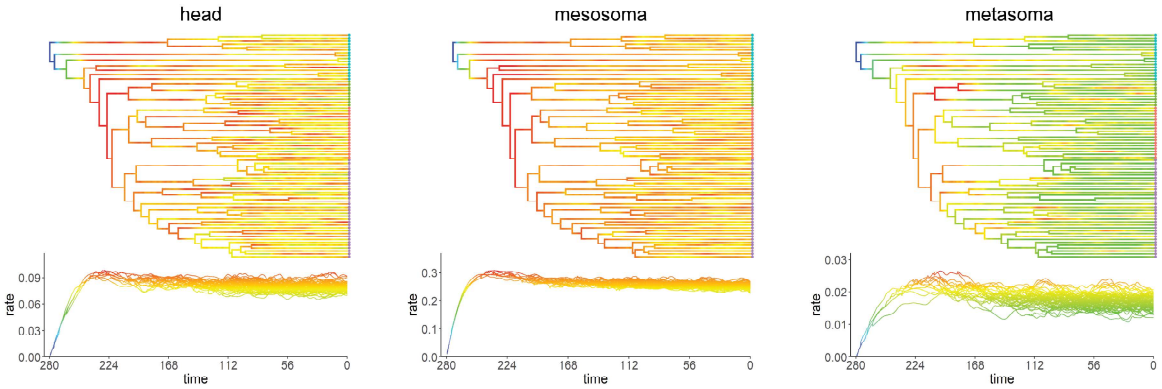

Fig. S7. Branch-specific rate dynamics inferred with ontophylo for the null analyses**.** Each subpanel shows the result for a body region of the Hymenoptera anatomy. In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rates-through-time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

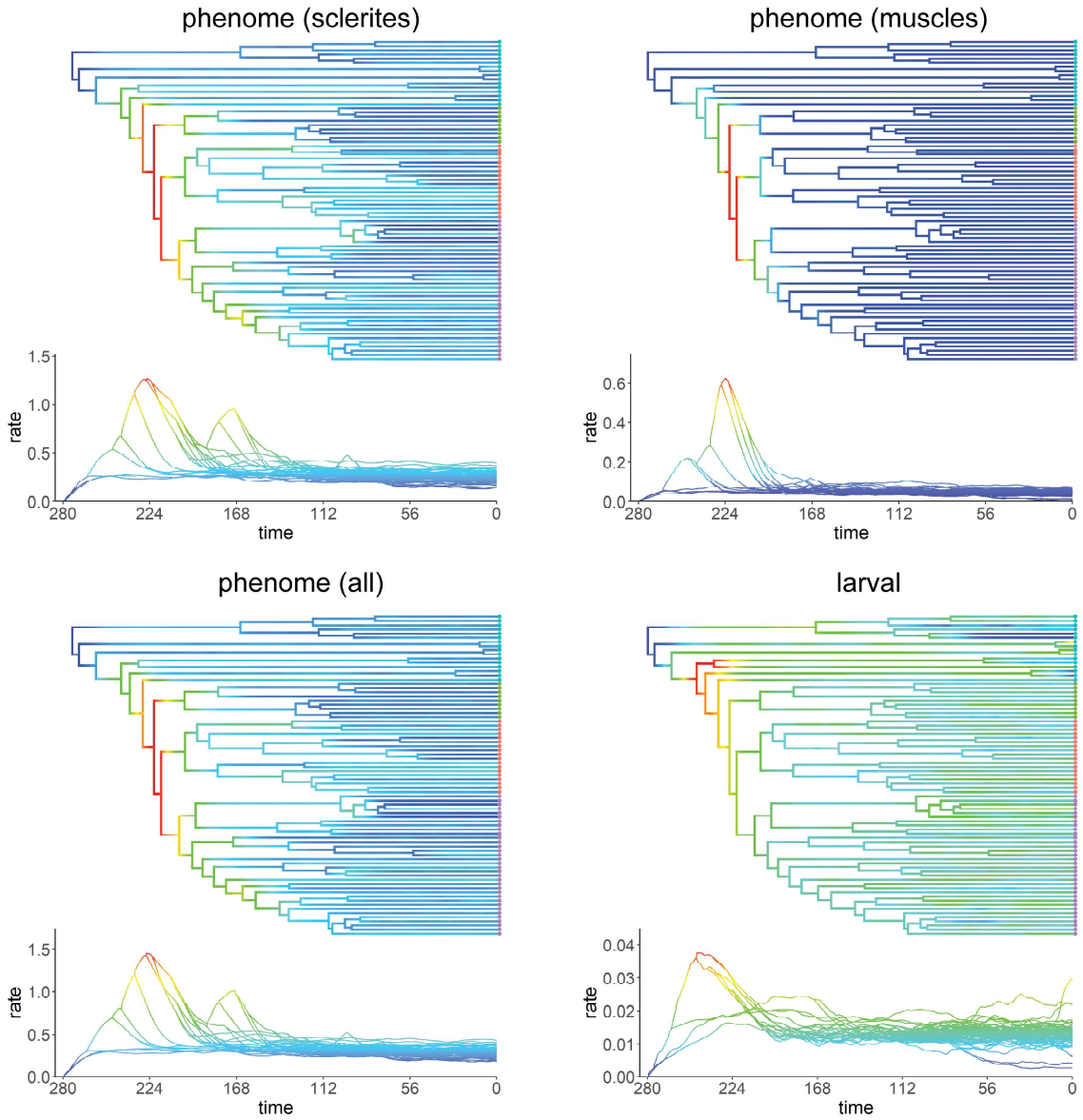

Fig. S8. Branch-specific rate dynamics inferred with ontophylo for the standard analyses. Subpanels show the results for adult and larval phenome of Hymenoptera. In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rates-through-time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

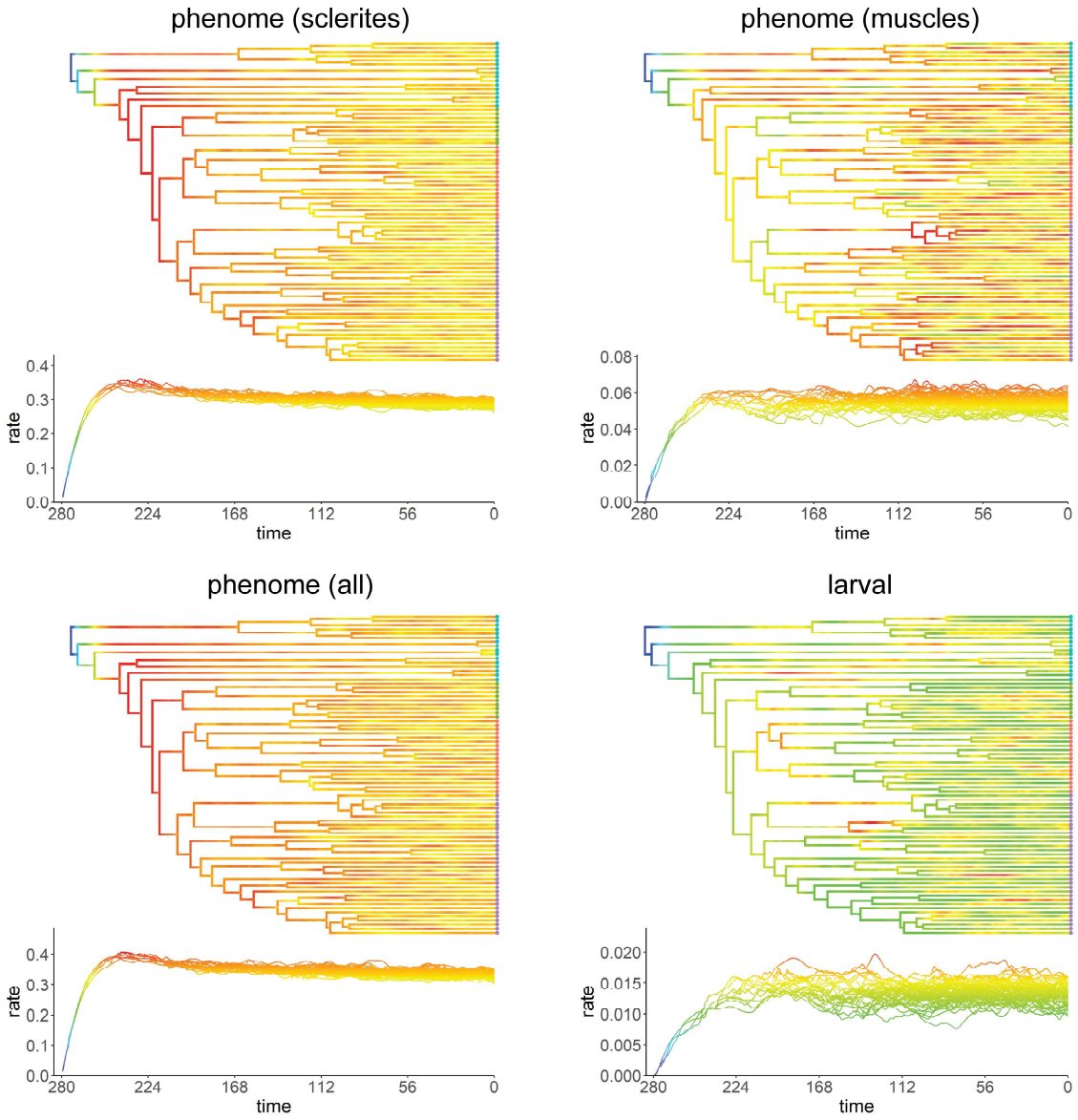

Fig. S9. Branch-specific rate dynamics inferred with ontophylo for the null analyses. Subpanels show the results for adult and larval phenome of Hymenoptera. In each subpanel, the top figure shows tree branches shaded by absolute rate values (contmaps); the bottom figure shows rates-through-time for each branch (edgeplots). Color gradient from blue to red indicates low to high rates respectively. Small filled circles at tips indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata clade; purple: Proctotrupomorpha.

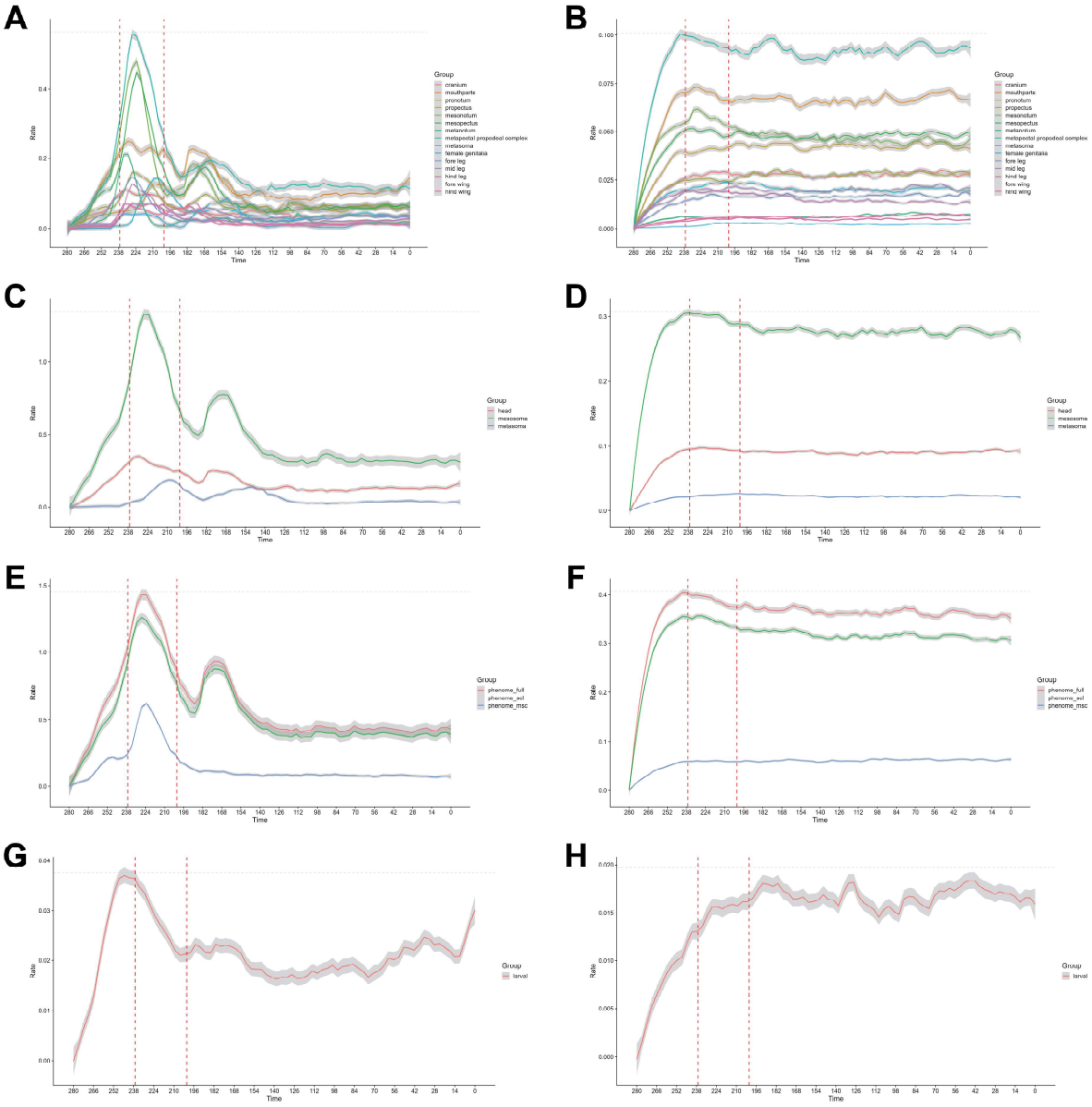

Fig. S10. Rate-through-time curves obtained from: *(A)* Anatomical complexes in the standard analyses (15 partitions). *(B)* Anatomical complexes in the null analyses. *(C)* Body regions in the standard analyses (3 partitions). *(D)* Body regions in the null analyses. *(E)* Adult phenome (all characters, only sclerites, and only muscles) in the standard analyses. *(F)* Adult phenome in the null analyses. *(G)* Larval phenome in the standard 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.

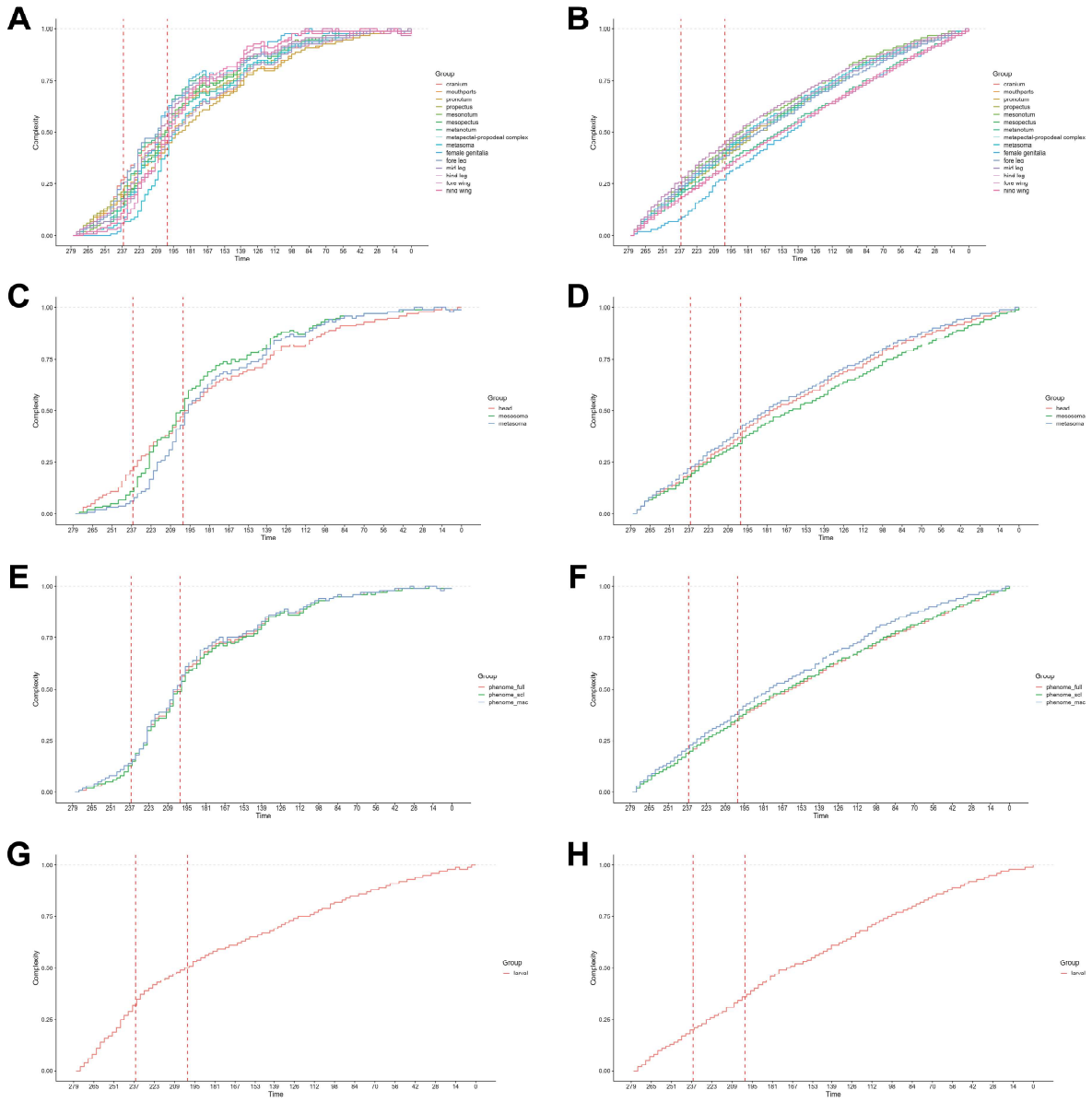

Fig. S11. Innovation-through-time curves obtained from: *(A)* Anatomical complexes in the standard analyses (15 partitions). *(B)* Anatomical complexes in the null analyses. *(C)* Body regions in the standard analyses (3 partitions). *(D)* Body regions in the null analyses. *(E)* Adult phenome (all characters, only sclerites, and only muscles) in the standard analyses. *(F)* Adult phenome in the null analyses. *(G)* Larval phenome in the standard 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.

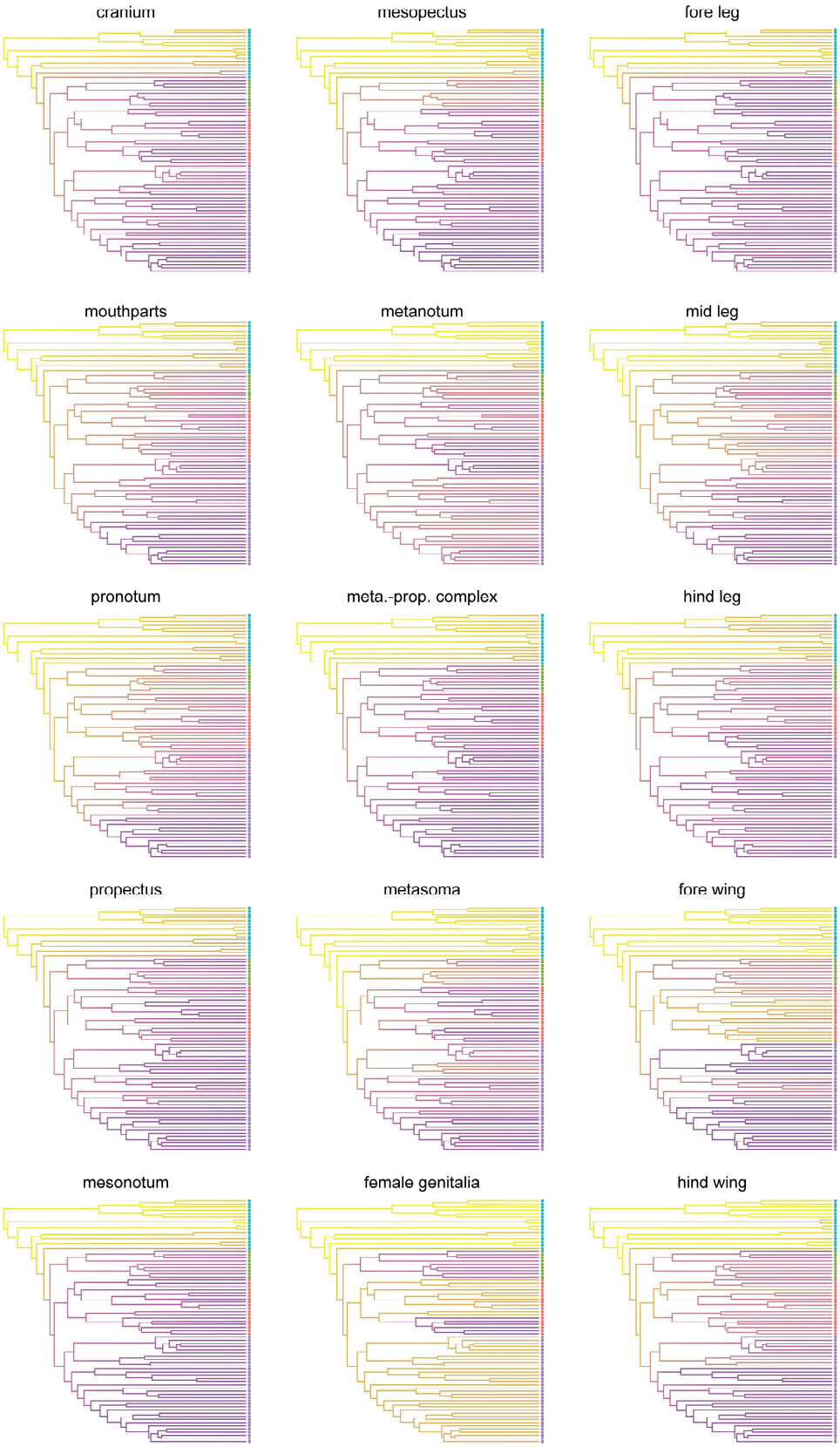

Fig. S12. Branch-specific innovation metric calculated from the samples of amalgamated stochastic maps in the standard analyses. Each subpanel shows the result for an elementary anatomical complex of the Hymenoptera anatomy. Purple color indicate higher innovation values. Small filled circles at tips indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata clade; purple: Proctotrupomorpha.

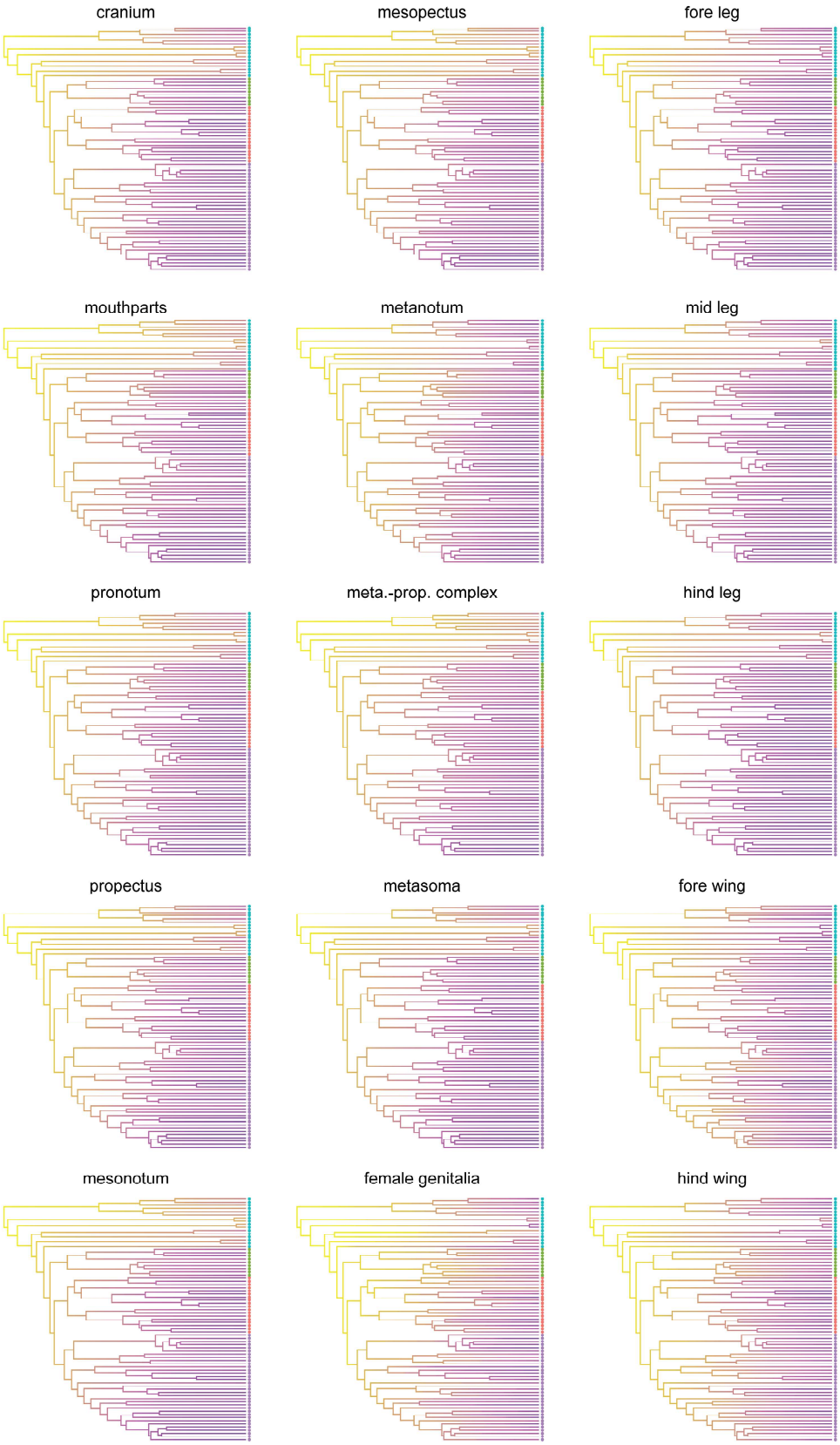

Fig. S13. Branch-specific innovation metric calculated from the samples of amalgamated stochastic maps in the null analyses. Each subpanel shows the result for an elementary anatomical complex of the Hymenoptera anatomy. Purple color indicate higher innovation values. Small filled circles at tips indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata clade; purple: Proctotrupomorpha.

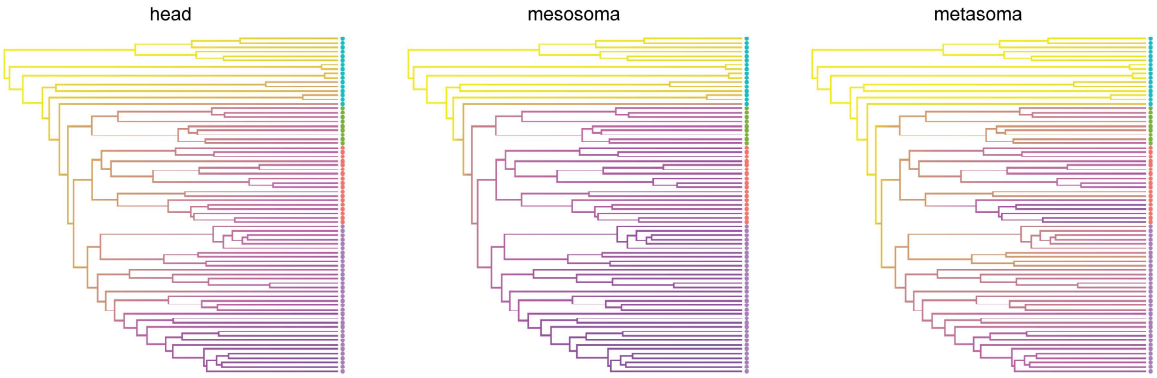

Fig. S14. Branch-specific innovation metric calculated from the samples of amalgamated stochastic maps in the standard analyses. Each subpanel shows the result for a body region of the Hymenoptera anatomy. Purple color indicate higher innovation valus. Small filled circles at tips indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata clade; purple: Proctotrupomorpha.

Fig. S15. 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 indicate higher innovation values. Small filled circles at tips indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata clade; purple: Proctotrupomorpha.

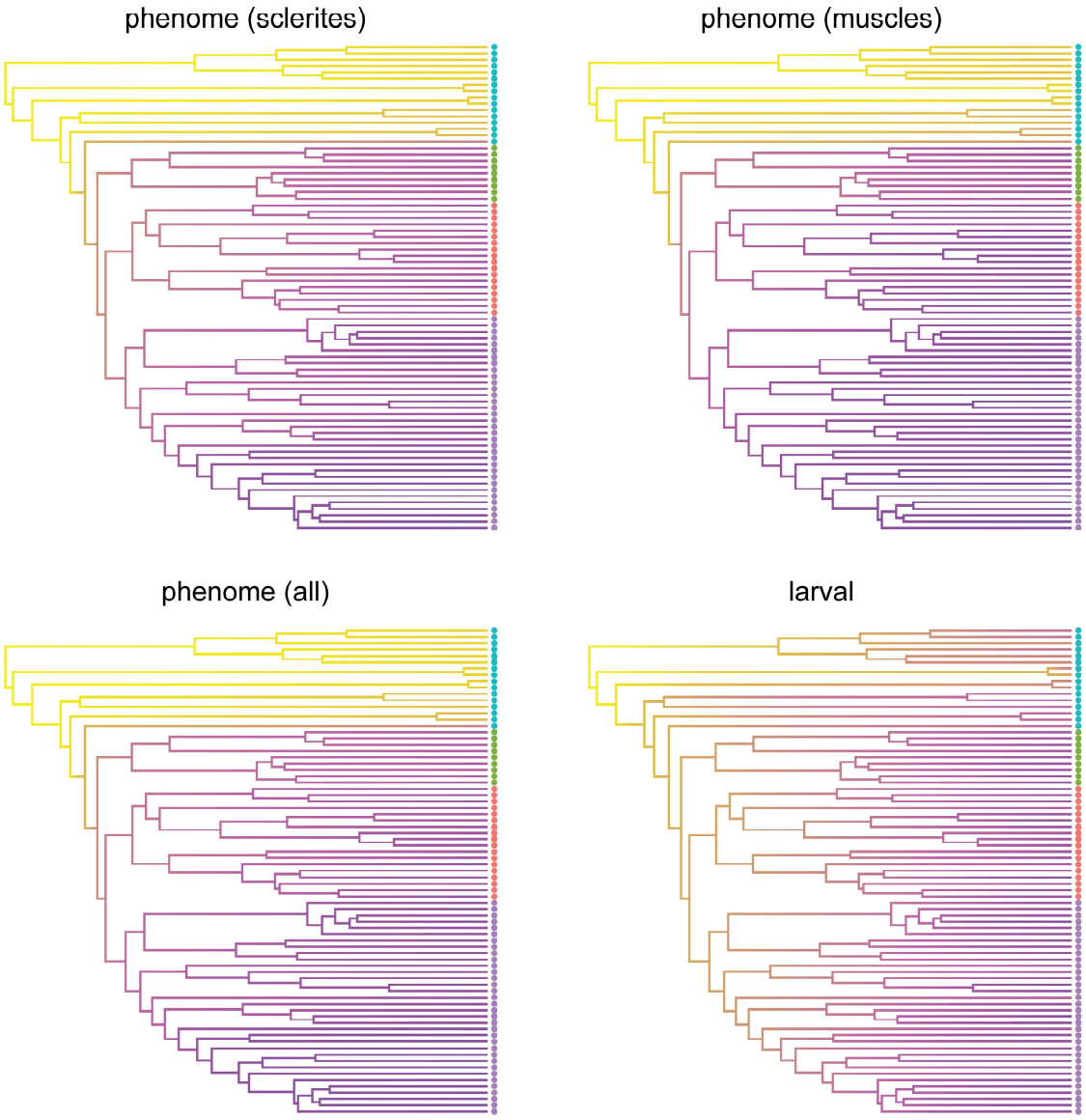

Fig. S16. Branch-specific innovation metric calculated from the samples of amalgamated stochastic maps in the standard analyses**.** Subpanels show the results for adult and larval phenome of Hymenoptera**.** Purple color indicate higher innovation values**.** Small filled circles at tips indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata clade; purple: Proctotrupomorpha.

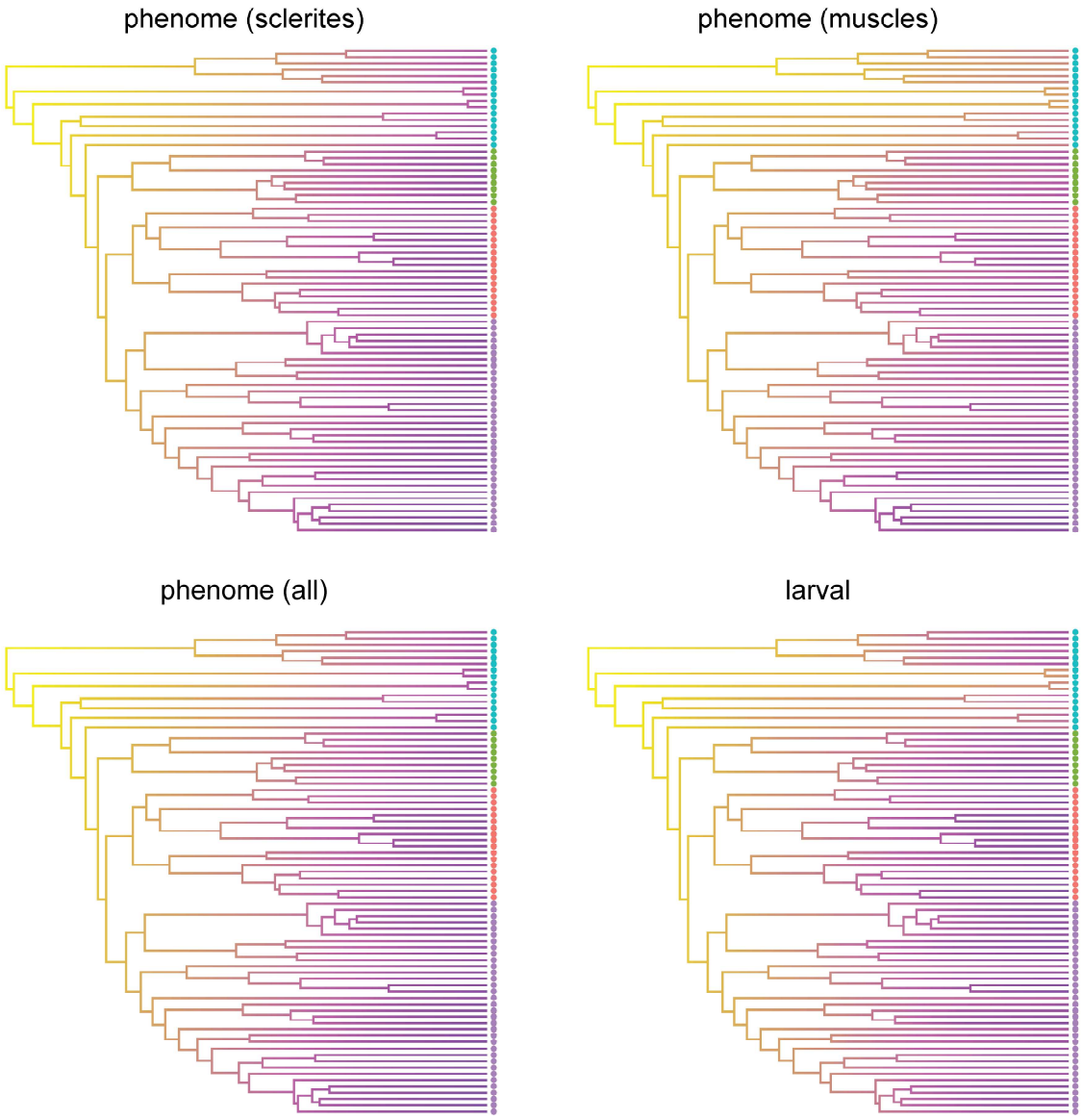

Fig. S17. Branch-specific innovation metric calculated from the samples of amalgamated stochastic maps in the null analyses. Subpanels show the results for adult and larval phenome of Hymenoptera. Purple color indicate higher innovation values. Small filled circles at tips indicate different lineages: cyan: non-Apocrita, green: Ichneumonoidea; red: "Evaniomorpha"-Aculeata clade; purple: Proctotrupomorpha.

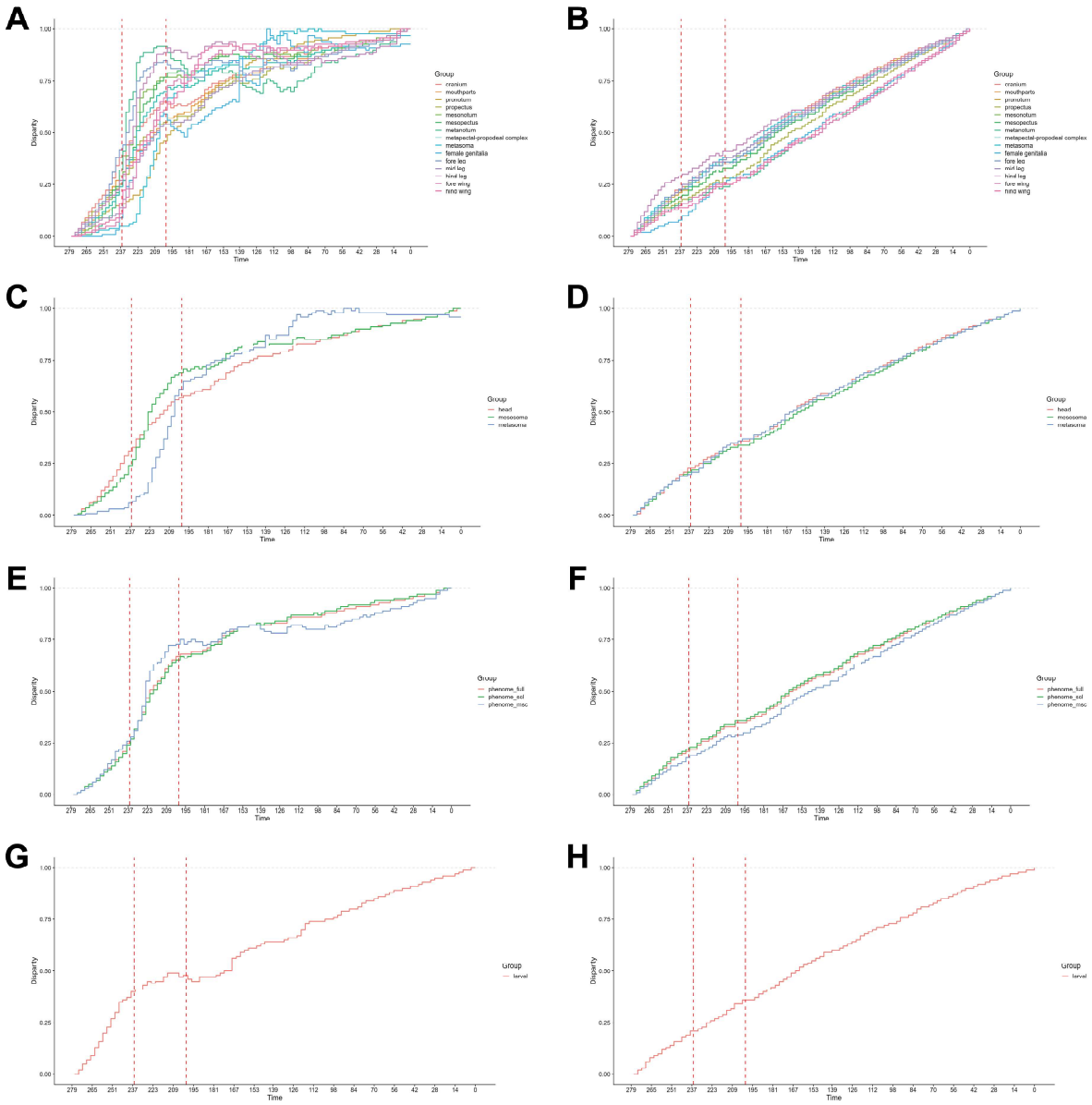

Fig. S18. Disparity-through-time curves obtained from: *(A)* Anatomical complexes in the standard analyses (15 partitions)**.** *(B)* Anatomical complexes in the null analyses**.** *(C)* Body regions in the standard analyses (3 partitions)**.** *(D)* Body regions in the null analyses**.** *(E)* Adult phenome (all characters, only sclerites, and only muscles) in the standard analyses**.** *(F)* Adult phenome in the null analyses**.** *(G)* Larval phenome in the standard 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**.**

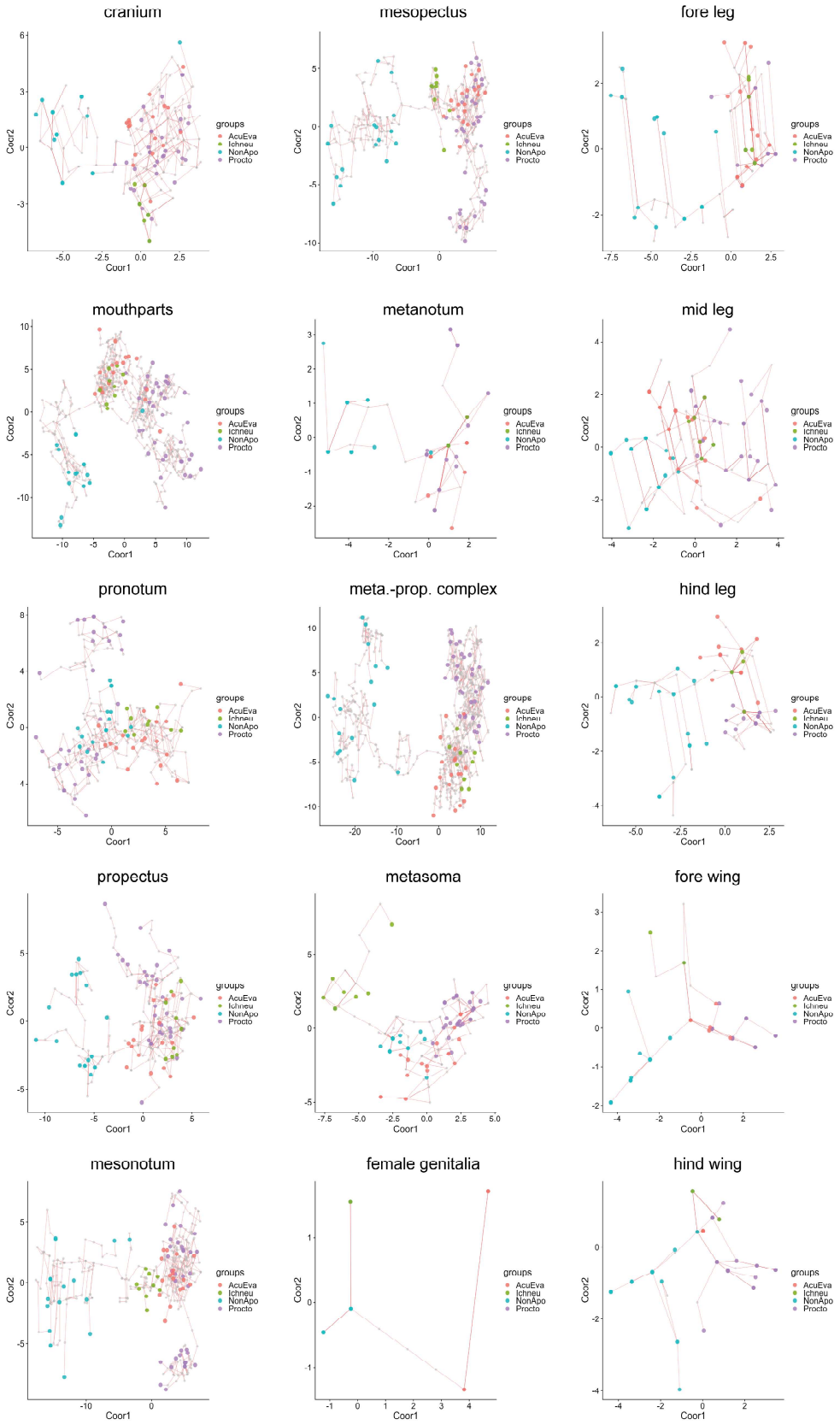

Fig. S19. Morphospaces reconstructed with ontophylo for the standard analyses. Each subpanel shows the result for an elementary anatomical complex of the Hymenoptera anatomy. Small, grey circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

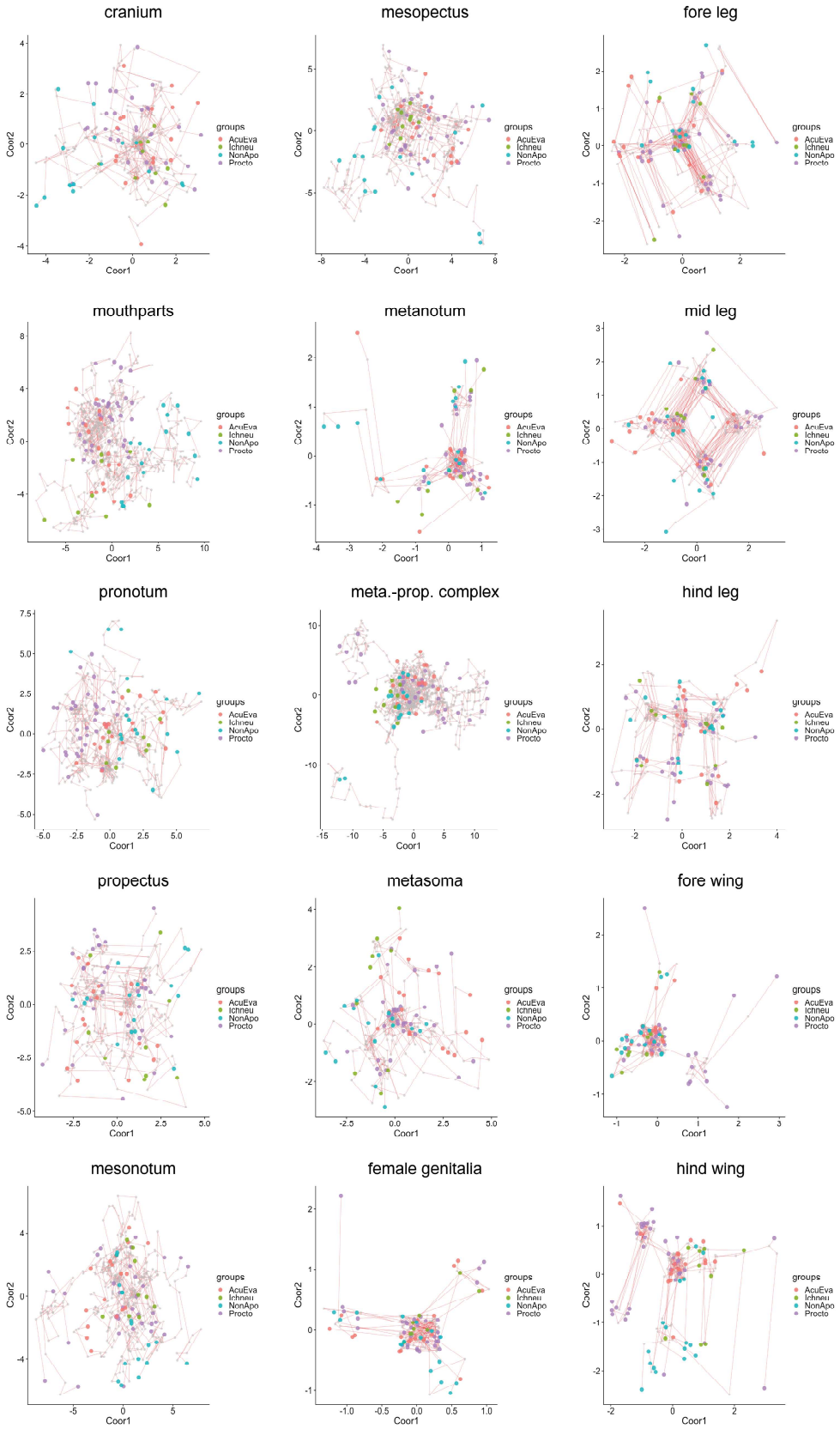

Fig. S20. Morphospaces reconstructed with ontophylo for the null analyses. Each subpanel shows the result for an elementary anatomical complex of the Hymenoptera anatomy. Small, grey circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

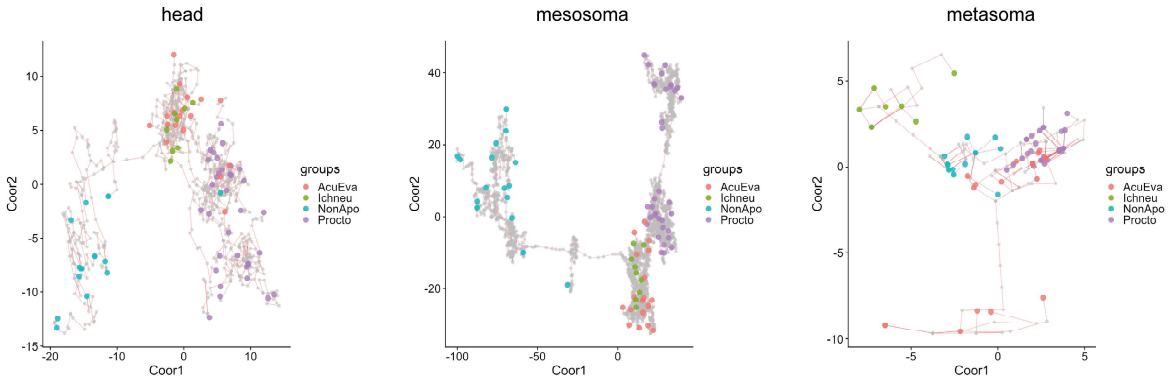

Fig. S21. Morphospaces reconstructed with ontophylo for the standard analyses. Each subpanel shows the result for a body region of the Hymenoptera anatomy. Small, grey circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

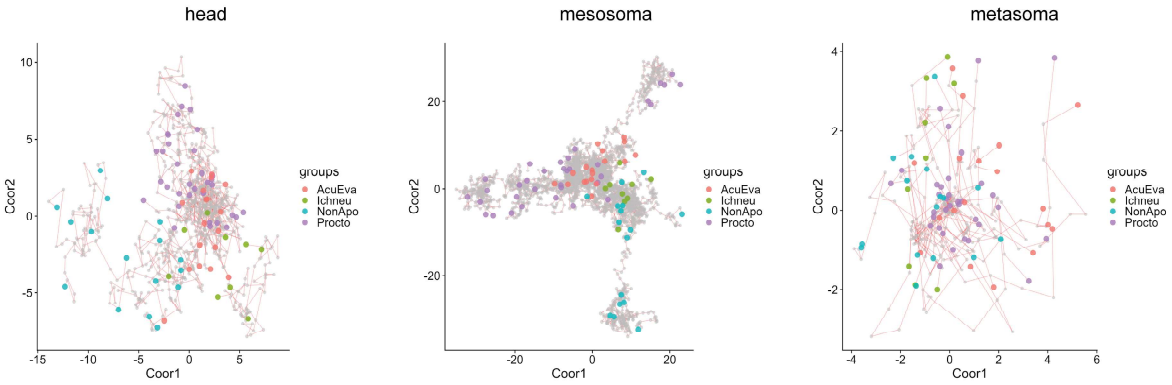

Fig. S22. Morphospaces reconstructed with ontophylo for the null analyses. Each subpanel shows the result for a body region of the Hymenoptera anatomy. Small, grey circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

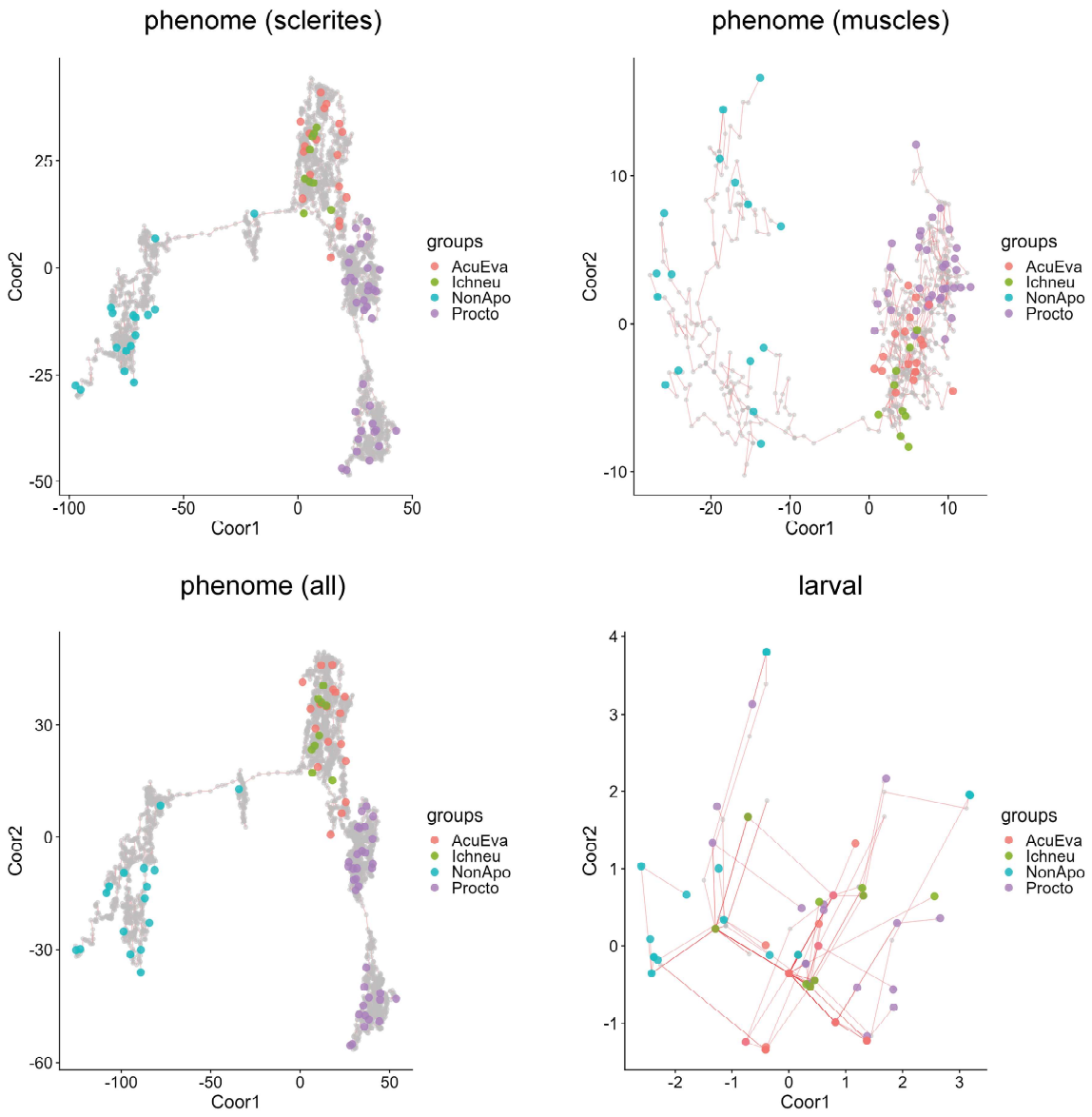

Fig. S23. Morphospaces reconstructed with ontophylo for the standard analyses. Subpanels show the results for adult and larval phenome of Hymenoptera. Small, grey circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

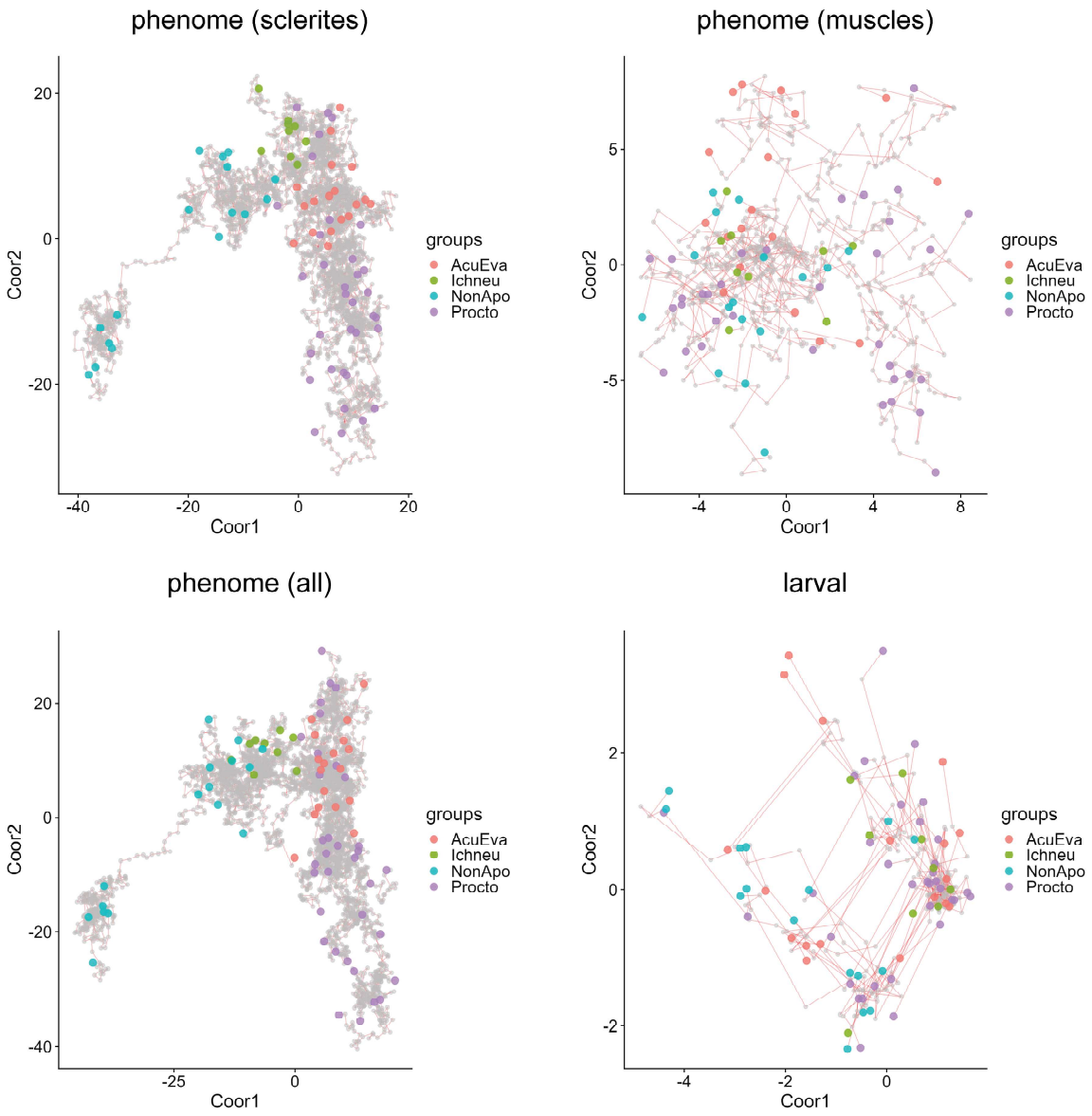

Fig. S24. Morphospaces reconstructed with ontophylo for the null analyses**.** Subpanels show the results for adult and larval phenome of Hymenoptera**.** Small, grey circles indicate ancestral phenomic states; large, color-coded circles indicate tip phenomic states; red lines connecting circles indicate steps between sampled states in the path from root to tips.

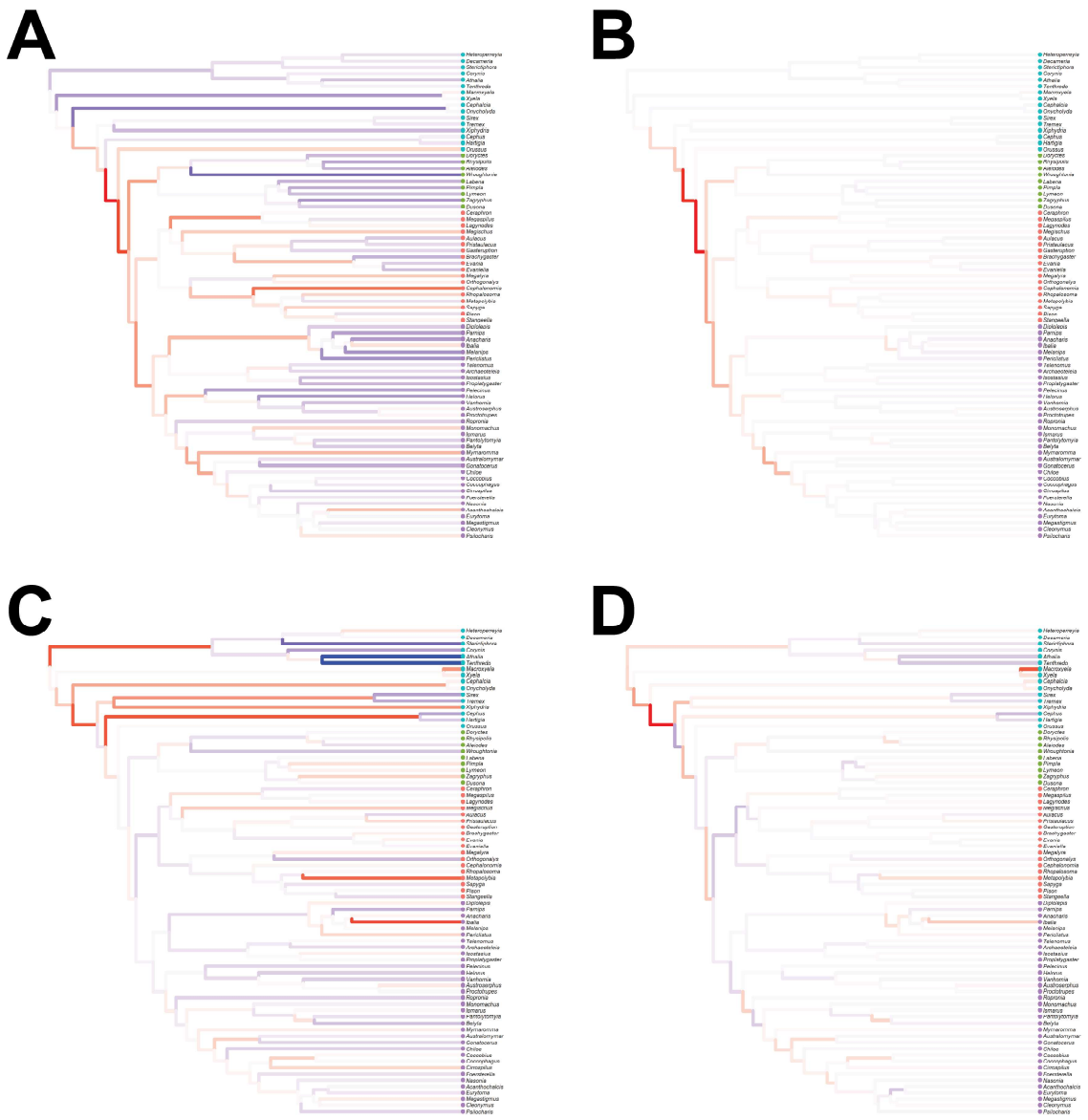

Fig. S25. Phenomic metric to assess regime of selection. *(A)* Full adult phenome (raw values). *(B)* Full adult phenome (normalized relative to branch lengths). *(C)* Larval phenome (raw values). *(D)* Larval phenome (normalized relative to branch lengths). Tree branches shaded according to the difference between the metric calculated in the standard and null analyses. Red color indicates positive values (i.e., directional selection); blue color indicates negative values (i.e., stabilizing selection); light-grey color indicates non-significant values.

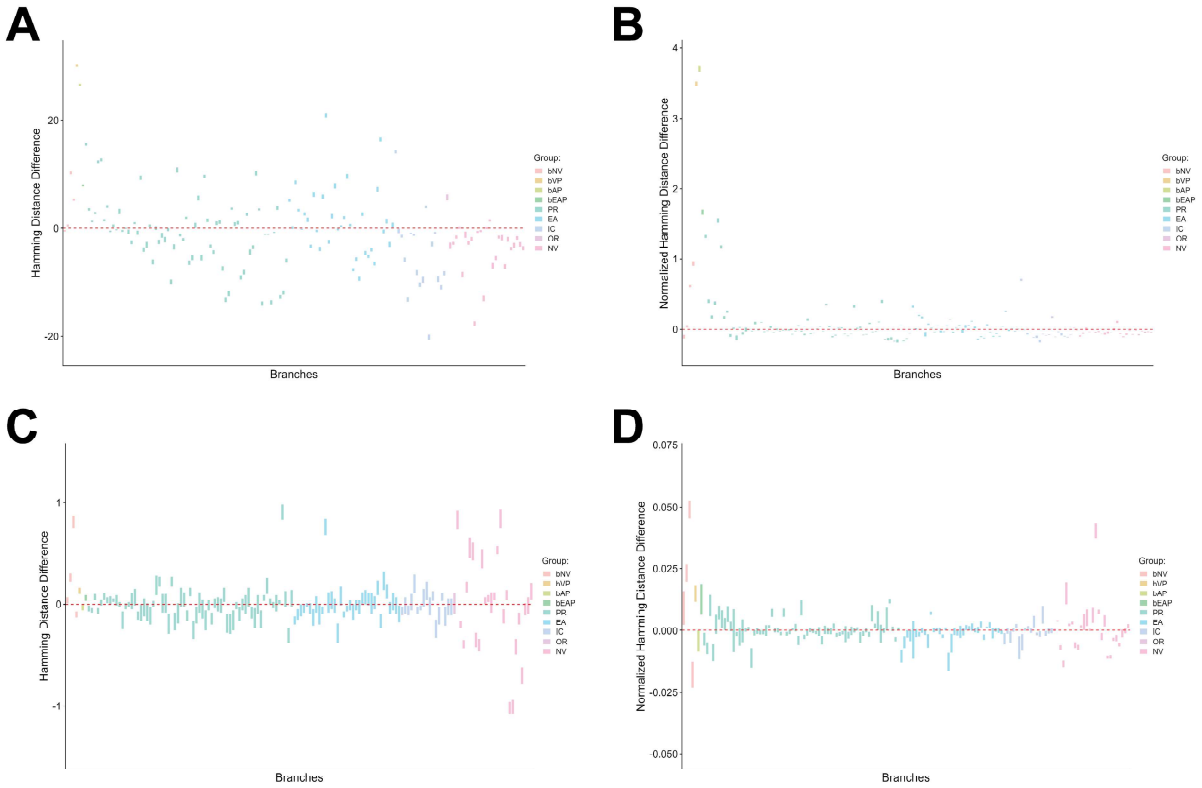

Fig. S26. Phenomic metric to assess regime of selection. *(A)* Full adult phenome (raw values). *(B)* Full adult phenome (normalized relative to branch lengths). *(C)* Larval phenome (raw values). *(D)* Larval phenome (normalized relative to branch lengths). Boxplots show the difference between the metric calculated in the standard 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 clade; PRO: Proctotrupomorpha; bNV: backbone non-Vespina; VP: backbone Vespina; AP: backbone Apocrita; EAP: backbone Apocrita non-Ichneumonoidea.

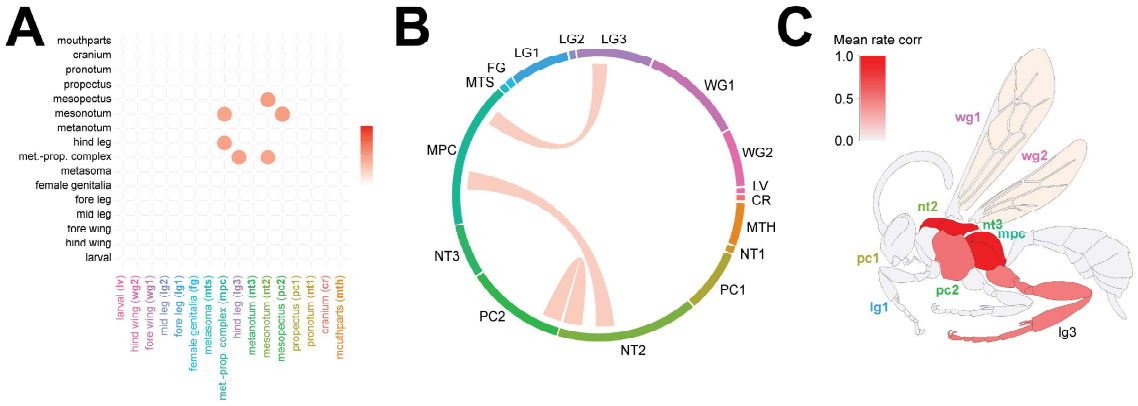

Fig. S27. Rate contrast (delta) correlations among anatomical complexes (and larval phenome). *(A)* Correlogram. *(B)* Connectogram. *(C)* Anatomical diagram color-shaded by the mean correlation for each anatomical complex across all partitions (i.e., row means in the correlation matrix of the correlogram). Color intensity in (A-C) and thickness of connections in (B) indicate strength of correlations (all positive). Abbreviations of anatomical complexes in (B-C) as indicated in (A).

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 using raw rates.

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	0.00	0.00	0.16	0.95
mouthparts + mesonotum	0.19	0.39	0.57	0.00	0.00	0.01	0.99
mouthparts + mesopectus	0.09	0.33	0.54	0.00	0.00	0.14	0.95
mouthparts + metapectal-propodeal complex	0.19	0.38	0.58	0.00	0.00	0.01	0.99
mouthparts + fore wing	0.28	0.51	0.69	0.00	0.00	0.00	1.00
mouthparts + hind wing	0.11	0.36	0.61	0.00	0.00	0.09	0.95
propectus + mesonotum	0.24	0.50	0.72	0.00	0.00	0.00	1.00
propectus + mesopectus	0.20	0.48	0.72	0.00	0.00	0.01	1.00
propectus + metapectal-propodeal complex	0.30	0.53	0.71	0.00	0.00	0.00	1.00
propectus + fore leg	0.10	0.36	0.62	0.00	0.00	0.12	0.95
propectus + hind leg	0.12	0.41	0.66	0.00	0.00	0.08	0.97
propectus + fore wing	0.13	0.43	0.67	0.00	0.00	0.07	0.97
mesonotum + mesopectus	0.48	0.69	0.83	0.00	0.00	0.00	1.00
mesonotum + metanotum	0.21	0.54	0.77	0.00	0.00	0.01	0.99
mesonotum + metapectal-propodeal complex	0.49	0.68	0.81	0.00	0.00	0.00	1.00
mesonotum + fore leg	0.25	0.49	0.68	0.00	0.00	0.00	1.00
mesonotum + hind leg	0.21	0.48	0.70	0.00	0.00	0.01	0.99
mesonotum + fore wing	0.26	0.51	0.67	0.00	0.00	0.00	1.00
mesonotum + hind wing	0.18	0.42	0.65	0.00	0.00	0.02	0.99
mesopectus + metanotum	0.09	0.40	0.66	0.00	0.00	0.13	0.95
mesopectus + metapectal-propodeal complex	0.42	0.65	0.80	0.00	0.00	0.00	1.00
mesopectus + fore leg	0.16	0.44	0.65	0.00	0.00	0.02	0.99
mesopectus + hind leg	0.19	0.50	0.74	0.00	0.00	0.01	0.99
mesopectus + fore wing	0.18	0.47	0.66	0.00	0.00	0.02	0.99
metanotum + metapectal-propodeal complex	0.21	0.57	0.78	0.00	0.00	0.00	1.00
metanotum + fore leg	0.15	0.46	0.75	0.00	0.00	0.03	0.98
metanotum + hind wing	0.18	0.55	0.72	0.00	0.00	0.02	0.98
metapectal-propodeal complex + fore leg	0.26	0.55	0.73	0.00	0.00	0.00	1.00
metapectal-propodeal complex + hind leg	0.29	0.57	0.77	0.00	0.00	0.00	1.00
metapectal-propodeal complex + fore wing	0.29	0.54	0.70	0.00	0.00	0.00	1.00
metapectal-propodeal complex + hind wing	0.16	0.38	0.67	0.00	0.00	0.03	0.99
fore leg + hind leg	0.13	0.44	0.70	0.00	0.00	0.05	0.98
hind leg + fore wing	0.12	0.43	0.65	0.00	0.00	0.08	0.97
fore wing + hind wing	0.13	0.45	0.70	0.00	0.00	0.06	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 resampling approach to assess correlations among phenotypic rates of evolution using rate contrasts (deltas).

Pair	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
mesonotum +mesopectus	0.18	0.52	0.73	0.000	0.000	0.017	0.99
mesonotum + metapectal-propodeal complex	0.12	0.46	0.70	0.000	0.000	0.071	0.97
metapectal-propodeal complex + hind leg	0.09	0.49	0.75	0.000	0.000	0.120	0.96

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

Table S4. Results from the regression approach to assess correlations among phenotypic rates of evolution using raw rates.

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 standard 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 standard 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 standard 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 standard analyses.

Table S5. Results from the regression approach to assess correlations among phenotypic rates of evolution using rate contrasts (deltas).

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]
mesonotum + mesopectus	0.98	0.000	0.000	0.029	0.18	0.52	0.75	0.17	0.51	0.75
mesonotum + metapectal-propodeal complex	0.97	0.000	0.000	0.078	0.15	0.47	0.72	0.14	0.47	0.71
metapectal-propodeal complex + hind leg	0.95	0.000	0.000	0.290	0.08	0.49	0.74	0.07	0.49	0.74

[1-7]Column descriptions as in Table S4.

Table S6. Results from the permutation tests to assess correlations between phenotypic rates and regimes of selection.

Pair	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
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

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

Table S7. Results from the permutation tests to assess correlations between phenotypic and lineage diversification rates using raw rates.

Pair	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
cranium	-0.31	-0.19	-0.08	0.000	0.019	0.331	0.62
mouthparts	-0.34	-0.24	-0.16	0.000	0.002	0.044	0.99
pronotum	-0.27	-0.16	-0.04	0.001	0.042	0.602	0.57
propectus	-0.33	-0.25	-0.16	0.000	0.001	0.060	0.96
mesonotum	-0.36	-0.30	-0.24	0.000	0.000	0.003	1.00
mesopectus	-0.38	-0.31	-0.23	0.000	0.000	0.004	1.00
metanotum	-0.23	-0.16	-0.13	0.003	0.042	0.130	0.64
metapectal-propodeal complex	-0.37	-0.30	-0.24	0.000	0.000	0.004	1.00
metasoma	-0.28	-0.20	-0.10	0.000	0.013	0.224	0.83
female genitalia	-0.21	-0.18	-0.12	0.007	0.022	0.122	0.90
fore leg	-0.33	-0.20	-0.11	0.000	0.014	0.188	0.76
mid leg	-0.22	-0.12	-0.01	0.004	0.136	0.898	0.25
hind leg	-0.29	-0.19	-0.07	0.000	0.018	0.433	0.72
fore wing	-0.30	-0.25	-0.20	0.000	0.002	0.015	1.00
hind wing	-0.30	-0.23	-0.20	0.000	0.004	0.016	1.00
larval	-0.20	-0.04	0.09	0.013	0.469	0.979	0.09

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

Table S8. Results from the permutation tests to assess correlations between phenotypic and lineage diversification rates (deltas).

Pair	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
cranium	-0.10	0.03	0.13	0.113	0.620	0.966	0.02
mouthparts	-0.10	0.01	0.13	0.109	0.733	0.985	0.01
pronotum	-0.09	0.03	0.10	0.206	0.599	0.990	0.00
propectus	-0.09	0.04	0.10	0.202	0.566	0.986	0.00
mesonotum	-0.06	0.03	0.10	0.216	0.679	0.984	0.00
mesopectus	-0.03	0.07	0.13	0.102	0.394	0.956	0.00
metanotum	-0.14	-0.04	0.06	0.081	0.567	0.963	0.01
metapectal-propodeal complex	-0.04	0.04	0.13	0.103	0.634	0.995	0.00
metasoma	-0.09	0.02	0.11	0.173	0.572	0.979	0.00
female genitalia	-0.14	-0.03	0.04	0.089	0.634	0.946	0.02
fore leg	-0.11	0.05	0.16	0.045	0.376	0.933	0.04
mid leg	-0.11	0.05	0.15	0.054	0.463	0.978	0.02
hind leg	-0.09	0.08	0.14	0.089	0.320	0.953	0.00
fore wing	0.01	0.08	0.10	0.229	0.351	0.892	0.00
hind wing	-0.01	0.04	0.09	0.266	0.655	0.938	0.00
larval	-0.18	0.01	0.12	0.033	0.509	0.981	0.04

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

Table S9. Results from the permutation tests to assess correlations between lineage diversification rates and regimes of selection.

Pair	Corr_lw[1]	Corr_md[2]	Corr_up[3]	Pval_lw[4]	Pval_md[5]	Pval_up[6]	Pval_prop[7]
cranium	-0.18	-0.15	-0.12	0.028	0.067	0.136	0.25
mouthparts	-0.18	-0.16	-0.13	0.028	0.057	0.104	0.36
pronotum	-0.27	-0.24	-0.22	0.000	0.002	0.009	1.00
propectus	-0.26	-0.24	-0.22	0.000	0.003	0.010	1.00
mesonotum	-0.27	-0.25	-0.23	0.000	0.001	0.005	1.00
mesopectus	-0.44	-0.42	-0.40	0.000	0.000	0.000	1.00
metanotum	-0.11	-0.09	-0.07	0.175	0.282	0.419	0.00
metapectal-propodeal complex	-0.29	-0.27	-0.25	0.000	0.001	0.003	1.00
metasoma	-0.03	-0.01	0.01	0.712	0.898	0.993	0.00
female genitalia	-0.16	-0.15	-0.14	0.039	0.063	0.096	0.15
fore leg	-0.09	-0.06	-0.03	0.298	0.468	0.684	0.00
mid leg	-0.04	-0.01	0.02	0.632	0.879	0.994	0.00
hind leg	-0.19	-0.15	-0.12	0.019	0.059	0.139	0.38
fore wing	-0.30	-0.28	-0.26	0.000	0.000	0.003	1.00
hind wing	-0.29	-0.28	-0.26	0.000	0.000	0.003	1.00
larval	-0.20	-0.14	-0.08	0.011	0.077	0.331	0.33

[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. BB Blaimer, et al., Key innovations and the diversification of hymenoptera. *Nat. Commun.* **14**, 1212 (2023).
10. 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).
11. 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).
12. JP Huelsenbeck, R Nielsen, JP Bollback, Stochastic mapping of morphological characters. *Syst. biology* **52**, 131–158 (2003).
13. D Greene, S Richardson, E Turro, ontologyx: a suite of r packages for working with ontological data. *Bioinformatics* **33**, 1104–1106 (2017).
14. C Parins-Fukuchi, GW Stull, SA Smith, Phylogenomic conflict coincides with rapid morphological innovation. *Proc. Natl. Acad. Sci.* **118**, e2023058118 (2021).
15. 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).
16. LJ Revell, phytools 2.0: an updated r ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**, e16505 (2024).