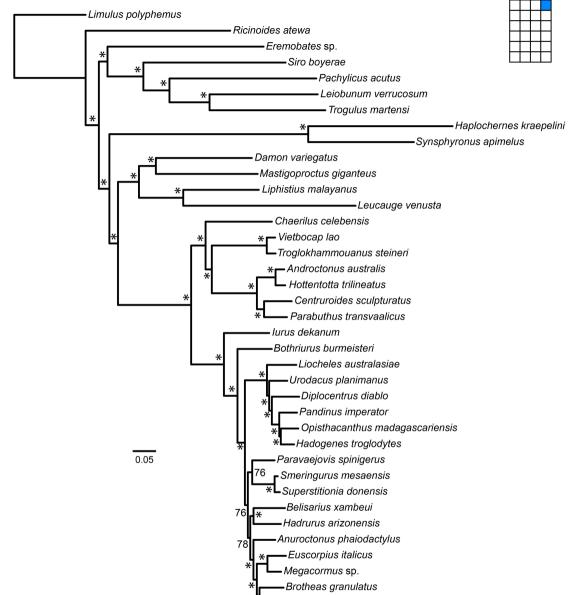
Figure S1. Tree topology based on maximum likelihood analysis of 136 orthologs (Matrix 1). Numbers on nodes indicate bootstrap resampling frequencies. - Limulus polyphemus Ricinoides atewa - Eremobates sp. Siro boyerae Pachylicus acutus Troqulus martensi Leiobunum verrucosum Mastigoproctus giganteus <50 Damon variegatus Leucauge venusta Liphistius malayanus Synsphyronus apimelus 97 Haplochernes kraepelini Chaerilus celebensis Troglokhammouanus steineri Vietbocap lao Centruroides sculpturatus 96 Parabuthus transvaalicus Androctonus australis Hottentotta trilineatus lurus dekanum Bothriurus burmeisteri × Superstitionia donensis Smeringurus mesaensis Anuroctonus phaiodactylus Brotheas granulatus Scorpiops sp. 0.05 Euscorpius italicus Megacormus sp. 98 Hadrurus arizonensis Belisarius xambeui Paravaejovis spinigerus 63 Urodacus planimanus 76 Liocheles australasiae 26.337 Diplocentrus diablo 52 93.0% occupancy Hadogenes troglodytes Opisthacanthus madagascariensis Pandinus imperator 136 orthologs

Figure S2. Tree topology based on maximum likelihood analysis of 599 orthologs (Matrix 2). Numbers on nodes indicate bootstrap resampling frequencies. Limulus polyphemus Ricinoides atewa - Eremobates sp. 64 Siro boverae Pachylicus acutus Leiobunum verrucosum Trogulus martensi Mastigoproctus giganteus 98 Damon variegatus Leucauge venusta Liphistius malayanus Haplochernes kraepelini Synsphyronus apimelus 93 Chaerilus celebensis Vietbocap lao Troglokhammouanus steineri <50 Parabuthus transvaalicus · Centruroides sculpturatus - Hottentotta trilineatus Androctonus australis lurus dekanum Bothriurus burmeisteri Superstitionia donensis Smeringurus mesaensis Anuroctonus phaiodactylus Euscorpius italicus Megacormus sp. Scorpiops sp. 0.05 Brotheas granulatus Belisarius xambeui Hadrurus arizonensis Paravaejovis spinigerus Urodacus planimanus Liocheles australasiae 128,842 Diplocentrus diablo 91 86.9% occupancy Opisthacanthus madagascariensis Pandinus imperator Hadogenes troglodytes 599 orthologs

Figure S3. Tree topology based on maximum likelihood analysis of 1,557 orthologs (Matrix 3). Numbers on nodes indicate bootstrap resampling frequencies. Limulus polyphemus Ricinoides atewa - Eremobates sp. 91 Siro boyerae Pachylicus acutus Trogulus martensi Leiobunum verrucosum Haplochernes kraepelini 91 Synsphyronus apimelus Mastigoproctus giganteus Damon variegatus Liphistius malayanus Leucauge venusta Chaerilus celebensis Vietbocap lao Troglokhammouanus steineri - Hottentotta trilineatus Androctonus australis Parabuthus transvaalicus Centruroides sculpturatus lurus dekanum Bothriurus burmeisteri Smeringurus mesaensis Superstitionia donensis Anuroctonus phaiodactylus Megacormus sp. Euscorpius italicus Scorpiops sp. 0.05 Brotheas granulatus Belisarius xambeui Hadrurus arizonensis Paravaejovis spinigerus - Liocheles australasiae Urodacus planimanus 346,260 Diplocentrus diablo 79 80.1% occupancy Hadogenes troglodytes Opisthacanthus madagascariensis Pandinus imperator

1,557 orthologs

**Figure S4.** Tree topology based on maximum likelihood analysis of 5,025 orthologs (Matrix 4). Numbers on nodes indicate bootstrap resampling frequencies.



Scorpiops sp.

Figure S5. Tree topologies based on species tree methods, analysis of gene trees from Matrix 1. (a) STAR method; (b) NJst method; (c) MP-est method

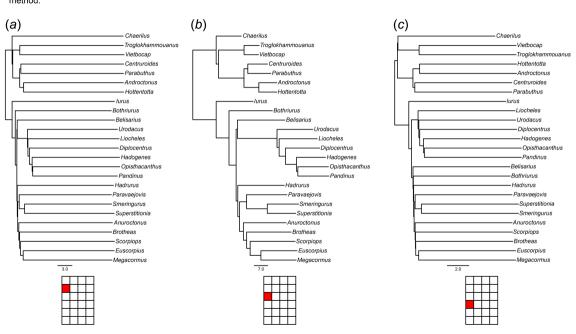
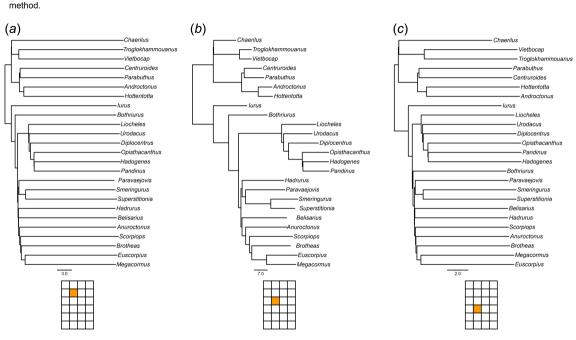


Figure S6. Tree topologies based on species tree methods, analysis of gene trees from Matrix 2. (a) STAR method; (b) NJst method; (c) MP-est



**Figure S7.** Tree topology based on maximum likelihood analysis of 131 orthologs (Matrix 5). Numbers on nodes indicate bootstrap resampling frequencies.

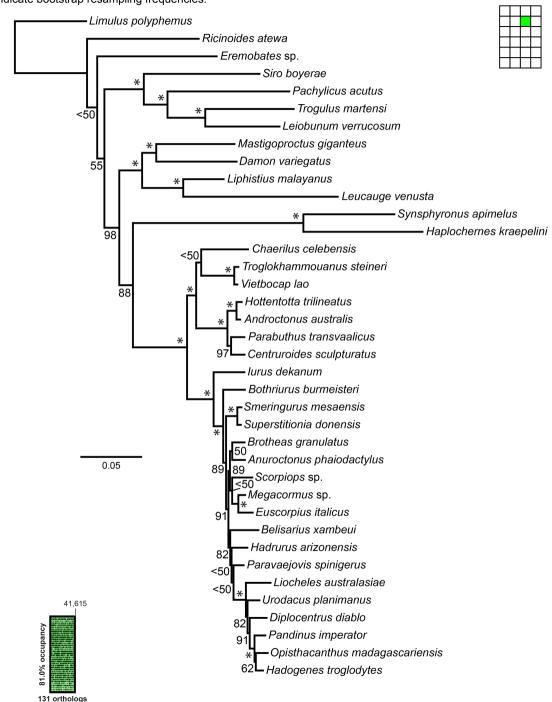


Figure S8. Tree topology based on maximum likelihood analysis of 453 orthologs (Matrix 6). Numbers on nodes indicate bootstrap resampling frequencies. Limulus polyphemus Ricinoides atewa - Eremobates sp. Siro boverae Pachylicus acutus Troqulus martensi 92 Leiobunum verrucosum Leucauge venusta Liphistius malayanus < 50 Mastigoproctus giganteus Damon variegatus - Haplochernes kraepelini Synsphyronus apimelus 98 Chaerilus celebensis Troglokhammouanus steineri Vietbocap lao < 50 Androctonus australis Hottentotta trilineatus Parabuthus transvaalicus Centruroides sculpturatus · lurus dekanum Bothriurus burmeisteri Superstitionia donensis Smeringurus mesaensis Anuroctonus phaiodactylus Brotheas granulatus 0.05 99 Scorpiops sp. Megacormus sp. Euscorpius italicus Belisarius xambeui Hadrurus arizonensis Paravaejovis spinigerus 99 Urodacus planimanus Liocheles australasiae 98,370 Diplocentrus diablo 58 87.1% occupancy Opisthacanthus madagascariensis Hadogenes troglodytes - Pandinus imperator 453 orthologs

Figure S9. Tree topology based on maximum likelihood analysis of 2,580 orthologs (Matrix 7). Numbers on nodes indicate bootstrap resampling frequencies. Limulus polyphemus ·Ricinoides atewa · Eremobates sp. 65 Siro boyerae Pachylicus acutus Trogulus martensi Leiobunum verrucosum Synsphyronus apimelus 91 - Haplochernes kraepelini Leucauge venusta Liphistius malayanus Mastigoproctus giganteus Damon variegatus Chaerilus celebensis Troglokhammouanus steineri 99 Vietbocap lao Androctonus australis Hottentotta trilineatus Centruroides sculpturatus Parabuthus transvaalicus -lurus dekanum Bothriurus burmeisteri Smeringurus mesaensis Superstitionia donensis Anuroctonus phaiodactylus Scorpiops sp. Brotheas granulatus Euscorpius italicus 0.05 Megacormus sp. 80 Belisarius xambeui Hadrurus arizonensis Paravaejovis spinigerus 78 Liocheles australasiae 79 Urodacus planimanus Diplocentrus diablo 89 Pandinus imperator Hadogenes troglodytes Opisthacanthus madagascariensis 565,648 72.0t% occupancy

Figure S10. Tree topology based on maximum likelihood analysis of 500 orthologs (Matrix 8). Numbers on nodes indicate bootstrap resampling frequencies. Limulus polyphemus Ricinoides atewa Eremobates sp. Siro boyerae Pachylicus acutus Leiobunum verrucosum Trogulus martensi Damon variegatus Mastigoproctus giganteus 62 - Leucauge venusta Liphistius malayanus Synsphyronus apimelus Haplochernes kraepelini Chaerilus celebensis Vietbocap lao Troglokhammouanus steineri Hottentotta trilineatus Androctonus australis Centruroides sculpturatus Parabuthus transvaalicus lurus dekanum Bothriurus burmeisteri Smeringurus mesaensis Superstitionia donensis Anuroctonus phaiodactylus Brotheas granulatus 0.05 Euscorpius italicus Megacormus sp. Belisarius xambeui Scorpiops sp. Hadrurus arizonensis 96 Paravaejovis spinigerus Urodacus planimanus Liocheles australasiae 118.851 Diplocentrus diablo 76 81.7% occupancy Hadogenes troglodytes 99 Pandinus imperator Opisthacanthus madagascariensis 500 orthologs

Figure S11. Tree topology based on maximum likelihood analysis of 500 orthologs (Matrix 9). Numbers on nodes indicate bootstrap resampling frequencies. Limulus polyphemus Ricinoides atewa - Eremobates sp. < 50 Siro boverae - Pachylicus acutus Troqulus martensi Leiobunum verrucosum Liphistius malayanus <<del>5</del>0 - Leucauge venusta Damon variegatus Mastigoproctus giganteus Haplochernes kraepelini Synsphyronus apimelus 90 Chaerilus celebensis 93 Troglokhammouanus steineri Vietbocap lao <50 · Androctonus australis Hottentotta trilineatus Centruroides sculpturatus Parabuthus transvaalicus lurus dekanum Bothriurus burmeisteri Superstitionia donensis Smeringurus mesaensis Anuroctonus phaiodactylus Euscorpius italicus 0.05 Megacormus sp. Scorpiops sp. Brotheas granulatus 93 Hadrurus arizonensis Belisarius xambeui Paravaejovis spinigerus 92 Urodacus planimanus 93 Liocheles australasiae 118.260 Diplocentrus diablo 65 80.1% occupancy Opisthacanthus madagascariensis Hadogenes troglodytes Pandinus imperator 500 orthologs

Figure S12. Tree topology based on maximum likelihood analysis of 577 orthologs (Matrix 10). Numbers on nodes indicate bootstrap resampling frequencies. Limulus polyphemus Ricinoides atewa - Eremobates sp. Siro boverae Pachylicus acutus Leiobunum verrucosum Trogulus martensi Mastigoproctus giganteus Damon variegatus Leucauge venusta Liphistius malayanus Synsphyronus apimelus Haplochernes kraepelini <50 Chaerilus celebensis <50 Vietbocap lao Troglokhammouanus steineri <50 Centruroides sculpturatus Parabuthus transvaalicus Hottentotta trilineatus Androctonus australis lurus dekanum Bothriurus burmeisteri Smeringurus mesaensis Superstitionia donensis Anuroctonus phaiodactylus Megacormus sp. 0.05 Euscorpius italicus Scorpiops sp. Brotheas granulatus Belisarius xambeui 99 Hadrurus arizonensis Paravaejovis spinigerus - Liocheles australasiae Urodacus planimanus 109,149 Diplocentrus diablo 75 78.5% occupancy Pandinus imperator Hadogenes troglodytes Opisthacanthus madagascariensis

577 orthologs

Figure S13. Tree topology based on Bayesian inference analysis of 136 orthologs (Matrix 1). Numbers on nodes indicate posterior probabilities. Unlabelled nodes correspond to posterior probabilities of 1.00. Limulus polyphemus Ricinoides atewa - Eremobates sp. 0.65 Siro boyerae Pachylicus acutus Troqulus martensi Leiobunum verrucosum Liphistius malayanus 0.58 Leucauge venusta Mastigoproctus giganteus Damon variegatus Synsphyronus apimelus - Haplochernes kraepelini Chaerilus celebensis 0.97 Vietbocap lao Troglokhammouanus steineri  $0.99^{-}$ Parabuthus transvaalicus Centruroides sculpturatus Androctonus australis Hottentotta trilineatus lurus dekanum Bothriurus burmeisteri Superstitionia donensis Smeringurus mesaensis Anuroctonus phaiodactylus Brotheas granulatus 0.05 Scorpiops sp. - Megacormus sp. Euscorpius italicus - Hadrurus arizonensis 0.99 Belisarius xambeui Paravaejovis spinigerus Urodacus planimanus Liocheles australasiae 26.337 Diplocentrus diablo 93.0% occupancy Hadogenes troglodytes Pandinus imperator Opisthacanthus madagascariensis 136 orthologs

Figure S14. Tree topology based on Bayesian inference analysis of 599 orthologs (Matrix 2). Numbers on nodes indicate posterior probabilities. Unlabelled nodes correspond to posterior probabilities of 1.00. Limulus polyphemus Ricinoides atewa - Eremobates sp. 0.55 Siro boyerae Pachylicus acutus Trogulus martensi Leiobunum verrucosum Liphistius malayanus 0.93 Leucauge venusta Mastigoproctus giganteus Damon variegatus Synsphyronus apimelus - Haplochernes kraepelini Chaerilus celebensis Vietbocap lao Troglokhammouanus steineri Parabuthus transvaalicus Centruroides sculpturatus Androctonus australis Hottentotta trilineatus lurus dekanum Bothriurus burmeisteri Superstitionia donensis Smeringurus mesaensis Anuroctonus phaiodactylus Brotheas granulatus 0.05 Scorpiops sp. - Megacormus sp. Euscorpius italicus Hadrurus arizonensis Belisarius xambeui Paravaejovis spinigerus Urodacus planimanus Liocheles australasiae 128,842 Diplocentrus diablo 86.9% occupancy Opisthacanthus madagascariensis Pandinus imperator Hadogenes troglodytes 599 orthologs

Figure S15. Tree topology based on maximum likelihood analysis of 67 orthologs (Matrix 11). Numbers on nodes indicate bootstrap resampling frequencies. - Limulus polyphemus Ricinoides atewa Eremobates sp. <50 Siro boyerae Pachylicus acutus Troqulus martensi Leiobunum verrucosum Mastigoproctus giganteus <50 Damon variegatus Leucauge venusta Liphistius malayanus Haplochernes kraepelini Synsphyronus apimelus Chaerilus celebensis 80 Vietbocap lao Troglokhammouanus steineri 99 98 Androctonus australis Hottentotta trilineatus Parabuthus transvaalicus × 98 Centruroides sculpturatus lurus dekanum Bothriurus burmeisteri Superstitionia donensis Smeringurus mesaensis Anuroctonus phaiodactylus 53 Brotheas granulatus 0.05 93 93 Scorpiops sp. Euscorpius italicus Megacormus sp. 55 Belisarius xambeui Hadrurus arizonensis <50 Paravaejovis spinigerus Urodacus planimanus 63 Liocheles australasiae 13,540 Diplocentrus diablo 56 92.7% occupancy Hadogenes troglodytes 95 Pandinus imperator Opisthacanthus madagascariensis 67 orthologs

Figure S16. Tree topology based on maximum likelihood analysis of 280 orthologs (Matrix 12). Numbers on nodes indicate bootstrap resampling frequencies. - Limulus polyphemus Ricinoides atewa Eremobates sp. 69 Siro boyerae Pachylicus acutus Leiobunum verrucosum Trogulus martensi Mastigoproctus giganteus 72 Damon variegatus Leucauge venusta Liphistius malayanus Haplochernes kraepelini Synsphyronus apimelus Chaerilus celebensis <50 Vietbocap lao Troglokhammouanus steineri <50 Centruroides sculpturatus Parabuthus transvaalicus Androctonus australis Hottentotta trilineatus lurus dekanum Bothriurus burmeisteri Smeringurus mesaensis Superstitionia donensis Paravaejovis spinigerus Liocheles australasiae 0.05 Urodacus planimanus Diplocentrus diablo <50 Opisthacanthus madagascariensis Pandinus imperator 99 - Hadogenes troglodytes Scorpiops sp. .95 Hadrurus arizonensis Belisarius xambeui 63,274 87 Anuroctonus phaiodactylus 86.9% occupancy Brotheas granulatus 95 Megacormus sp. 97 Euscorpius italicus 280 orthologs

Figure S17. Tree topology based on maximum likelihood analysis of 689 orthologs (Matrix 13). Numbers on nodes indicate bootstrap resampling frequencies. - Limulus polyphemus Ricinoides atewa - Eremobates sp. Siro boyerae Pachylicus acutus Troqulus martensi Leiobunum verrucosum Leucauge venusta 75 Liphistius malayanus Mastigoproctus giganteus Damon variegatus Synsphyronus apimelus - Haplochernes kraepelini Chaerilus celebensis Troglokhammouanus steineri Vietbocap lao <50 Centruroides sculpturatus Parabuthus transvaalicus Hottentotta trilineatus Androctonus australis lurus dekanum Bothriurus burmeisteri Smeringurus mesaensis Superstitionia donensis Anuroctonus phaiodactylus Megacormus sp. 0.05 Euscorpius italicus · Scorpiops sp. Brotheas granulatus Belisarius xambeui Hadrurus arizonensis Paravaejovis spinigerus Liocheles australasiae Urodacus planimanus 162,067 Diplocentrus diablo 80 80.4% occupancy Opisthacanthus madagascariensis Pandinus imperator Hadogenes troglodytes 689 orthologs

Figure S18. Tree topology based on maximum likelihood analysis of 1,725 orthologs (Matrix 14). Numbers on nodes indicate bootstrap resampling frequencies. Limulus polyphemus Ricinoides atewa - Eremobates sp. Siro boyerae Pachylicus acutus Leiobunum verrucosum

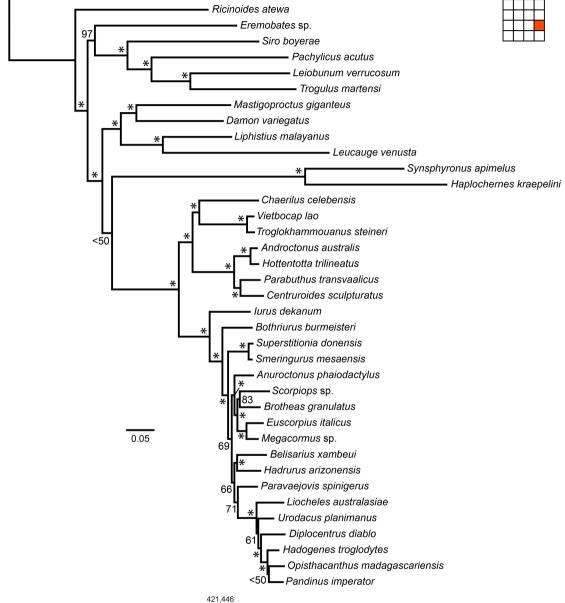


Figure S19. Tree topology based on maximum likelihood analysis of 136 orthologs (Matrix 1) under a mixed branch length model of heterotachy. Limulus polyphemus Haplochernes kraepelini Synsphyronus apimelus Ricinoides atewa Eremobates sp. Siro boyerae Pachylicus acutus - Leiobunum verrucosum Trogulus martensi Mastigoproctus giganteus Damon variegatus Liphistius malayanus - Leucauge venusta Chaerilus celebensis Vietbocap lao Troglokhammouanus steineri Hottentotta trilineatus Androctonus australis Centruroides sculpturatus - Parabuthus transvaalicus lurus dekanum · Liocheles australasiae Urodacus planimanus - Diplocentrus diablo Hadogenes troglodytes Opisthacanthus Pandinus imperator 0.10 Belisarius xambeui Paravaejovis spinigerus Hadrurus arizonensis - Bothriurus burmeisteri Smeringurus mesaensis Superstitionia donensis Anuroctonus phaiodactylus 26.337 Brotheas granulatus 93.0% occupancy Scorpiops sp. Megacormus sp. Euscorpius italicus

136 orthologs

Figure S20. Tree topology based on maximum likelihood analysis of 67 orthologs (Matrix 11) under a mixed branch length model of heterotachy. Limulus polyphemus Synsphyronus apimelus Haplochernes kraepelini Ricinoides atewa Eremobates sp. Siro boyerae - Pachylicus acutus Leiobunum verrucosum Trogulus martensi Liphistius malayanus Leucauge venusta - Damon variegatus Mastigoproctus giganteus Chaerilus celebensis Vietbocap lao Troglokhammouanus steineri Parabuthus transvaalicus Centruroides sculpturatus Hottentotta trilineatus Androctonus australis lurus dekanum · Liocheles australasiae Urodacus planimanus Diplocentrus diablo Hadogenes troglodytes Pandinus imperator 0.10 Opisthacanthus Bothriurus burmeisteri Smeringurus mesaensis Superstitionia donensis Paravaejovis spinigerus Hadrurus arizonensis - Anuroctonus phaiodactylus Belisarius xambeui 13,540 Brotheas granulatus 92.7% occupancy Scorpiops sp. Megacormus sp. Euscorpius italicus 67 orthologs

Figure S21. (a) Distributions of patristic distances for all terminals from the node that excludes *Limulus polyphemus*, *Ricinoides atewa*, and *Eremobates* sp. Red highlight indicates scorpion terminals. (b) Inset: Distributions of same patristic distances, portrayed as histograms. Scorpions are depicted with red, and outgroups with blue.

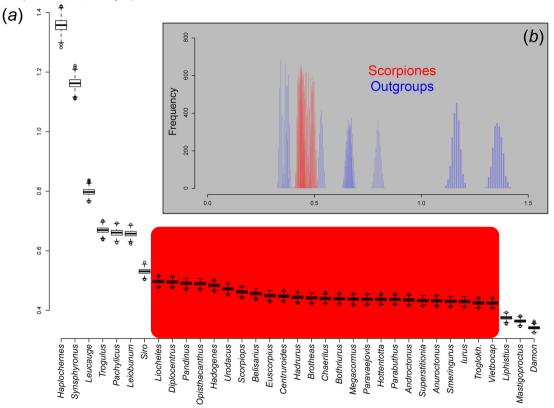
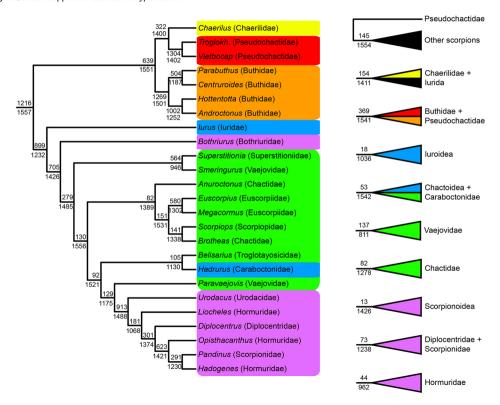
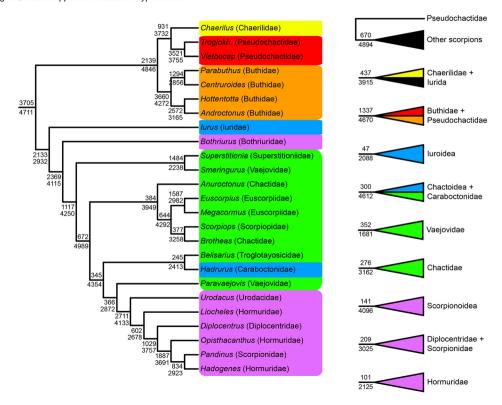


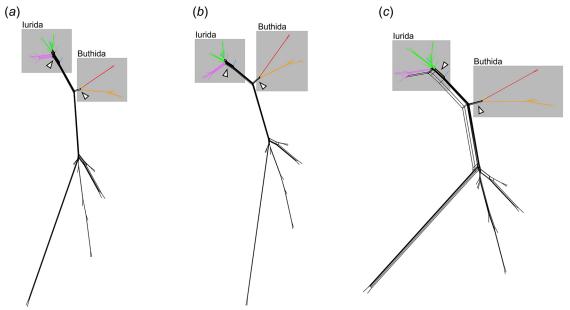
Figure S22. Tree topology based on maximum likelihood analysis of 1,557 orthologs, showing relative support from gene trees in Matrix 3. For each node, the upper number indicates the count of congruent gene trees; the lower number indicates the count of potentially informative genes. Right: Counts mapped for alternative hypotheses.



**Figure S23.** Tree topology based on maximum likelihood analysis of 1,557 orthologs, showing relative support from gene trees in Matrix 4. For each node, the upper number indicates the count of congruent gene trees; the lower number indicates the count of potentially informative genes. Right: Counts mapped for alternative hypotheses.



**Figure S24.** Supernetwork representation of quartets derived from individual ML gene trees, for three different data sets: (a) Matrix 2, (b) Matrix 3, and (c) Matrix 4. Phylogenetic conflict is represented by reticulations. Arrowheads indicate reticulations at the base of Buthida and within lurida. Edge lengths correspond to quartet frequencies.



**Figure S25.** Paucity of morphological characters consistent with the molecular tree topology and informative of interfamilial relationships. Shading of squares indicates character states of morphological characters. Scoring of morphological characters is provided in table S3.

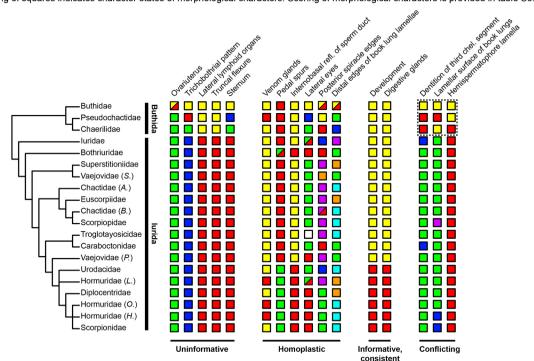


Figure S26. (a) Design of permutation-based test for random distribution of missing data. Parsimony scores are calculated for each gene, coding absence as state 0 and presence as state 1. Empirical cases of two orthogroups (#5016 and #5018) with 19 sequences present and different parsimony scores. (b) Missing data are determined to be non-randomly (e.g., #5018) or randomly (e.g., #5016) distributed for each gene, relative to a null distribution generated from reshuffling terminals 1,000 times. (c) Cumulative number of genes with non-randomly distributed missing data (NDMD) and number of taxa by each orthogroup, ranked in order of taxon occupancy (most to least). Colours indicate key data sets analysed in this study. p(NDMD) indicates proportion of each supermatrix that is comprised of genes with non-randomly distributed missing data. (d) Cladogram indicating 50% majority rule consensus of bootstrap resampling replicates from 4,008-gene supermatrix with randomly distributed missing data. Numbers on nodes in indicate bootstrap resampling frequencies.

