Supplementary Information

Myriapod genomes reveal ancestral horizontal gene transfer and hormonal gene loss in millipedes

So, Nong, and Xie et al

Supplementary Figures

- Figure 1. Macrosynteny between *Thereuonema tuberculata* (centipede) and eight other myriapods.
- Figure 2. Macrosynteny between *Trigoniulus corallinus* (millipede) and eight other myriapods.
- Figure 3. Macrosynteny between *Helicorthomorpha holstii* (millipede) and eight other myriapods.
- Figure 4. Number of shared orthogroups among 9 myriapods and 24 outgroup species.
- Figure 5. GO enrichment analysis of conserved gene families.
- Figure 6. KEGG enrichment analysis of conserved gene families.
- Figure 7. KOG enrichment analysis of conserved gene families.
- Figure 8. Schematic summary of Hox and ParaHox gene arrangement in the genome of centipede *R. immarginata*.
- Figure 9. Cladogram of HOXL genes
- Figure 10. Phylogenetic tree of glucose/arabinose dehydrogenase (GDH)
- Figure 11. Phylogenetic tree of non-ribosomal peptide synthetase (NRPS)
- Figure 12. Phylogenetic tree of glycoside hydrolase family 16 protein (GH16)
- Figure 13. Phylogenetic tree of efflux RND transporter permease subunit
- Figure 14. Phylogenetic tree of NADH dehydrogenase
- Figure 15. Phylogenetic tree of AzlD domain-containing protein
- Figure 16. Phylogenetic tree of anaerobic sulfatase maturase
- Figure 17. Phylogenetic tree of alpha-2-macroglobulin
- Figure 18. Phylogenetic tree of SYLF domain-containing protein
- Figure 19. Microsynteny of glucose/arabinose dehydrogenase (GDH; WP_189008864.1)
- Figure 20. Microsynteny of glucose/arabinose dehydrogenase (GDH; WP_146884959.1)
- Figure 21. Microsynteny of non-ribosomal peptide synthetase (NRPS; WP_096595152.1)
- Figure 22. Microsynteny of glycoside hydrolase family 16 protein (GH16; AQQ75061.1)
- Figure 23. Microsynteny of glycoside hydrolase family 16 protein (GH16; WP_052600908.1)
- Figure 24. Microsynteny of efflux RND transporter permease subunit (WP_163176792.1)
- Figure 25. Microsynteny of NADH dehydrogenase (WP_034862301.1)
- Figure 26. Microsynteny of AzlD domain-containing protein (WP_095524423.1)
- Figure 27. Microsynteny of anaerobic sulfatase maturase (SBW02910.1)
- Figure 28. Microsynteny of alpha-2-macroglobulin (OYY43986.1)
- Figure 29. Microsynteny of SYLF domain-containing protein (RPH48231.1)

- Figure 30. Tissue differential expression of horizontal transferred WP_189008864.1 Glucose/arabinose dehydrogenases in *T. tuberculata*.
- Figure 31. Phylogenetic tree of acetyl-CoA C-acetyltransferase (ACAT)
- Figure 32. Phylogenetic tree of hydroxymethylglutaryl-CoA synthase (HMGCS)
- Figure 33. Phylogenetic tree of hydroxymethylglutaryl-CoA reductase (HMGCR)
- Figure 34. Phylogenetic tree of mevalonate kinase (MK) and phosphomevalonate kinase (PMK)
- Figure 35. Phylogenetic tree of diphosphomevalonate decarboxylase (DPMD)
- Figure 36. Phylogenetic tree of isopentenyl-diphosphate delta-isomerase (IPPI)
- Figure 37. Phylogenetic tree of farnesyl pyrophosphate synthase (FPPS)
- Figure 38. Phylogenetic tree of farnesyltransferase beta (FNTB)
- Figure 39. Phylogenetic tree of ste24 endopeptidase (ste24)
- Figure 40. Phylogenetic tree of protein-S-isoprenylcysteine O-methyltransferase (ste14/ICMT)
- Figure 41. Phylogenetic tree of prenylcysteine oxidase (PCYOX1)
- Figure 42. Phylogenetic tree of aldehyde dehydrogenase 3 (ALDH3)
- Figure 43. Phylogenetic tree of juvenile hormone acid methyltransferase (JHAMT)
- Figure 44. Phylogenetic tree of juvenile hormone epoxide hydrolase (JHEH)
- Figure 45. Phylogenetic tree of juvenile hormone diol kinase (JHDK)
- Figure 46. Phylogenetic tree of cytosolic juvenile hormone binding protein (cJHBP)/glyoxalase domain containing protein 4 (GLOD4)
- Figure 47. Phylogenetic tree of ultraspiracle (USP) and ecdysteroid receptor (EcR)
- Figure 48. Phylogenetic tree of steroid receptor coactivator (SRC)
- Figure 49. Phylogenetic tree of methoprene receptor (Met)
- Figure 50. Phylogenetic tree of heat shock protein 83 (Hsp83)
- Figure 51. Phylogenetic tree of allatotropin receptor (ATR)
- Figure 52. Phylogenetic tree of allatostatin receptor (ASR)
- Figure 53. Phylogenetic tree of S-adenosylmethionine synthase (MAT)
- Figure 54. Phylogenetic tree of S-adenosylhomocysteinase (AHC)
- Figure 55. Phylogenetic tree of adenosine kinase (AK)
- Figure 56. Sequence alignment of centipede juvenile hormone acid methyltransferase (JHAMT)
- Figure 57. Microsynteny of JHAMT in centipede genomes
- Figure 58. Hormone measurement of sesquiterpenoids

- Figure 59. Myriapod repeat location plots
- Figure 60. Myriapod repeat landscape plots
- Figure 61. Repeat density karyotype plot for Helicorthomorpha holstii
- Figure 62. Repeat density karyotype plot for Trigoniulus corallinus
- Figure 63. Repeat density karyotype plot for Thereuonema tuberculata
- Figure 64. Bar chart showing the pattern by which different myriapod species share individual TE families

Supplementary Discussion - Genome size and transposable element content



Supplementary Figure 1. Macrosynteny between Trigoniulus corallinus (millipede) and eight other myriapods.



Supplementary Figure 2. Macrosynteny between Trigoniulus corallinus (millipede) and eight other myriapods.



Supplementary Figure 3. Macrosynteny between Helicorthomorpha holstii (millipede) and eight other myriapods.

Helicorthomorpha holstii	6603	6295	6555	6739	4866	5274	5812	6077	5968	6552	6526	7071	6879	6519	6769	6903	7178	6015	6997	7674	7315	8513	9833	10730	
Niponia nodulosa	6819	6399	6791	7449	4931	5387	5853	6121	7487	6724	6616	7479	7093	6734	6955	7183	7788	6132	7247	8857	7431	12067	19313	9833	
Anaulaciulus tonginus	6926	6493	6882	7500	4945	5407	5906	6211	6840	6855	6759	7591	7315	6944	7169	7465	7980	6264	7465	8613	7756	16048	12067	8513	
Trigoniulus corallinus	6695	6373	6649	6607	4865	5271	5824	6113	5409	6640	6618	7091	7071	6650	6840	7104	7184	6087	7204	7121	8800	7756	7431	7315	
Glomeris maerens	6634	6274	6617	6905	4835	5247	5778	6028	6067	6609	6572	7229	7017	6688	6890	7171	7523	6125	7235	10606	7121	8613	8857	7674	
Rhysida immarginata	6976	6566	6999	6979	4907	5389	5922	6246	5675	6807	6835	7403	7369	7120	7163	8113	8563	6679	10595	7235	7204	7465	7247	6997	
Strigamia maritima	5979	5685	5965	5905	4410	4793	5204	5482	4915	5869	5870	6240	6214	5899	6069	6508	6752	7570	6679	6125	6087	6264	6132	6015	
Lithobius niger	7101	6605	7194	7240	4944	5443	5915	6255	5984	6965	6846	7624	7537	7264	7260	8457	13412	6752	8563	7523	7184	7980	7788	7178	
Thereuonema tuberculata	7004	6539	7003	7015	4834	5313	5816	6156	5654	6780	6670	7370	7358	7017	7035	10793	8457	6508	8113	7171	7104	7465	7183	6903	
Centruroides sculpturatus	6792	6380	6744	6766	4858	5280	5821	6122	5562	6655	6936	7560	7522	7157	9142	7035	7260	6069	7163	6890	6840	7169	6955	6769	
Stegodyphus mimosarum	6483	6110	6494	6534	4769	5182	5684	5943	5437	6498	6717	7402	7375	9225	7157	7017	7264	5899	7120	6688	6650	6944	6734	6519	Ratio
Tachypleus tridentatus	7008	6550	6987	6973	4921	5392	5910	6214	5666	6841	7041	12730	14694	7375	7522	7358	7537	6214	7369	7017	7071	7315	7093	6879	0.4
Carcinoscorpius rotundicauda	7035	6604	7025	7084	4946	5431	5949	6250	5850	6894	7044	14827	12730	7402	7560	7370	7624	6240	7403	7229	7091	7591	7479	7071	0.2 0.1
lxodes scapularis	6511	6151	6443	6389	4817	5163	5703	5983	5340	6450	8763	7044	7041	6717	6936	6670	6846	5870	6835	6572	6618	6759	6616	6526	0.0
Litopenaeus vannamei	6637	6251	6526	6514	4821	5185	5850	6126	6439	9257	6450	6894	6841	6498	6655	6780	6965	5869	6807	6609	6640	6855	6724	6552	
Eriocheir sinensis	5549	5177	5525	5874	4180	4462	4871	5073	9964	6439	5340	5850	5666	5437	5562	5654	5984	4915	5675	6067	5409	6840	7487	5968	
Tribolium castaneum	6030	5817	5965	5934	4605	4979	6187	7283	5073	6126	5983	6250	6214	5943	6122	6156	6255	5482	6246	6028	6113	6211	6121	6077	
Drosophila melanogaster	5718	5558	5655	5661	4528	4843	6992	6187	4871	5850	5703	5949	5910	5684	5821	5816	5915	5204	5922	5778	5824	5906	5853	5812	
Hypsibius dujardini	5401	5182	5362	5382	4362	6979	4843	4979	4462	5185	5163	5431	5392	5182	5280	5313	5443	4793	5389	5247	5271	5407	5387	5274	
Caenorhabditis elegans	4937	4791	4918	4928	6532	4362	4528	4605	4180	4821	4817	4946	4921	4769	4858	4834	4944	4410	4907	4835	4865	4945	4931	4866	
Biomphalaria straminea	7684	6895	8448	13858	4928	5382	5661	5934	5874	6514	6389	7084	6973	6534	6766	7015	7240	5905	6979	6905	6607	7500	7449	6739	
Magallana hongkongensis	8003	7021	12910	8448	4918	5362	5655	5965	5525	6526	6443	7025	6987	6494	6744	7003	7194	5965	6999	6617	6649	6882	6791	6555	
Homo sapiens	7435	8513	7021	6895	4791	5182	5558	5817	5177	6251	6151	6604	6550	6110	6380	6539	6605	5685	6566	6274	6373	6493	6399	6295	
Branchiostoma floridae	10202	7435	8003	7684	4937	5401	5718	6030	5549	6637	6511	7035	7008	6483	6792	7004	7101	5979	6976	6634	6695	6926	6819	6603	
all have been al	oridae se	piens	Jensis stra	minea el	egans du	ardini	Jaster aste	meum sir	ensis van	namei	ularis undi	cauda nide	ntatus	sarum ulpt	Jratus	culata	niger	ritima	jinata ma	erens corr	ulinus tor	iginus not	ulosa oha	nolstii	
oranchiosion.	Hon	nonsinphi	alaria	laboli. Hyr	Sibiu	a mer Tribol	NUM E	loche Hoper	Benz M	odes-	15 rout	leus trodyp	us n. oid	Jes su one	matt	Linne	gamu physic	Jain. GK	Inen-	niulus Anault	Ciulus N	ponicontr	omori		
Ø. 1	N30.	Ø.,	00		Oru			v	Card	ino	Ģ	3 ¹⁰⁵ (5 ^{811.} 1	ne.						`		He.			

Supplementary Figure 4. Number of shared orthogroups among 9 myriapods and 24 outgroup species.



Supplementary Figure 5. GO enrichment analysis of conserved gene families. The number of annotated gene families in each group are listed below. Enrich groups with Benjamini and Hochberg method (BH) adjusted p value ≤ 0.20 are coloured red to blue, others are in grey. Ratio are calculated by number of gene families over total number of annotated gene families of specific GO term. Top 20 enriched groups are shown.



Supplementary Figure 6. KEGG enrichment analysis of conserved gene families. The number of annotated gene families in each group are listed below. Enrich groups with Benjamini and Hochberg method (BH) adjusted p value ≤ 0.20 are coloured red to blue, others are in grey. Ratio are calculated by number of gene families over total number of annotated gene families of specific KEGG term. Top 20 enriched groups are shown.



Supplementary Figure 7. KOG enrichment analysis of conserved gene families. The number of annotated gene families in each group are listed below. Enrich groups with Benjamini and Hochberg method (BH) adjusted p value ≤ 0.20 are coloured red to blue, others are in grey. Ratio are calculated by number of gene families over total number of annotated gene families of specific KOG term.



Supplementary Figure 8. Schematic summary of Hox and ParaHox gene arrangement in the genome of centipede *R. immarginata*.



Supplementary Figure 9. Cladogram of HOXL genes. The phylogenetic tree was constructed with the maximum likelihood method in IQ-Tree with ultrafast bootstrap of 1000 times (-B 1000 -bnni --alrt 1000). LG+I+G4 was chosen as the best-fit model according to BIC value. The tree was midpoint rooted using function 'midpoint()' in R package 'phangorn' v2.7.1. The phylogenetic tree was visualized using function 'ggtree()' in R package 'ggtree' v3.0.2, with 'branch.length = "none"' to show the cladogram. Only bootstrap values larger than 80% are indicated for clarity. Bilaterian homeobox genes from *Branchiostoma floridae* and *Drosophila melanogaster* are used as bilaterian representatives. Ttu, *T. tuberculate*; Rim, *R. immarginata*; Lni, *L. niger;* Sma, *S. maritima*; Gma, *G. maerens*; Hho, *H. holstii,* Nno, *N. niponia*, Tco, *T. corallinus,* Ato, *A. tonginus.*



1

Supplementary Figure 10. Phylogenetic tree of glucose/arabinose dehydrogenase (GDH). Tree topology is displayed according to the Maximum Likelihood (ML) method (WAG+G4). Bootstrapping was conducted on each node with 1000 replicates on both ML (black) and NJ (blue) algorithms. Only bootstrap values larger than 80% are indicated for clarity. Ttu, *T. tuberculate*; Rim, *R. immarginata*; Lni, *L. niger*; SMAR, *S. maritima*.



Supplementary Figure 11. Phylogenetic tree of non-ribosomal peptide synthetase (NRPS). Tree topology is displayed according to the Maximum Likelihood (ML) method (WAG+G4). Bootstrapping was conducted on each node with 1000 replicates on both ML (black) and NJ (blue) algorithms. Only bootstrap values larger than 80% are indicated for clarity. Ttu, *T. tuberculate*; Rim, *R. immarginata*; Lni, *L. niger*.



Supplementary Figure 12. Phylogenetic tree of glycoside hydrolase family 16 protein (GH16). Tree topology is displayed according to the Maximum Likelihood (ML) method (WAG+R3). Bootstrapping was conducted on each node with 1000 replicates on both ML (black) and NJ (blue) algorithms. Only bootstrap values larger than 80% are indicated for clarity. Hho, *H. holstii*; Nno, *N. nodulosa*; Tco, *T. corallinus*; Jul, *A. tonginus*.



Supplementary Figure 13. Phylogenetic tree of efflux RND transporter permease subunit. Tree topology is displayed according to the Maximum Likelihood (ML) method (LG+F+R3). Bootstrapping was conducted on each node with 1000 replicates on both ML (black) and NJ (blue) algorithms. Only bootstrap values larger than 80% are indicated for clarity. Gma, *G. maerens*; Nno, *N. nodulosa*; Jul, *A. tonginus*.



Supplementary Figure 14. Phylogenetic tree of NADH dehydrogenase. Tree topology is displayed according to the Maximum Likelihood (ML) method (WAG+R3). Bootstrapping was conducted on each node with 1000 replicates on both ML (black) and NJ (blue) algorithms. Only bootstrap values larger than 80% are indicated for clarity. Hho, *H. holstii*; Nno, *N. nodulosa*; Jul, *A. tonginus*.



Supplementary Figure 15. Phylogenetic tree of AzlD domain-containing protein. Tree topology is displayed according to the Maximum Likelihood (ML) method (JTT+G4). Bootstrapping was conducted on each node with 1000 replicates on both ML (black) and NJ (blue) algorithms. Only bootstrap values larger than 80% are indicated for clarity. Hho, *H. holstii*; Nno, *N. nodulosa*; Jul, *A. tonginus*.



Supplementary Figure 16. Phylogenetic tree of anaerobic sulfatase maturase. Tree topology is displayed according to the Maximum Likelihood (ML) method (WAG+G4). Bootstrapping was conducted on each node with 1000 replicates on both ML (black) and NJ (blue) algorithms. Only bootstrap values larger than 80% are indicated for clarity. Nno, *N. nodulosa*; Jul, *A. tonginus*.



0.5

Supplementary Figure 17. Phylogenetic tree of alpha-2-macroglobulin. Tree topology is displayed according to the Maximum Likelihood (ML) method (WAG+F+G4). Bootstrapping was conducted on each node with 1000 replicates on both ML (black) and NJ (blue) algorithms. Only bootstrap values larger than 80% are indicated for clarity. Hho, *H. holstii*; Nno, *N. nodulosa*; Jul, *A. tonginus*.



Supplementary Figure 18. Phylogenetic tree of SYLF domain-containing protein. Tree topology is displayed according to the Maximum Likelihood (ML) method (LG+R3). Bootstrapping was conducted on each node with 1000 replicates on both ML (black) and NJ (blue) algorithms. Only bootstrap values larger than 80% are indicated for clarity. Hho, *H. holstii*; Nno, *N. nodulosa*; Jul, *A. tonginus*.



Supplementary Figure 19. Microsynteny of glucose/arabinose dehydrogenase (GDH; WP_189008864.1).



Supplementary Figure 20. Microsynteny of glucose/arabinose dehydrogenase (GDH; WP_146884959.1).



Supplementary Figure 21. Microsynteny of non-ribosomal peptide synthetase (NRPS; WP_096595152.1).



Supplementary Figure 22. Microsynteny of glycoside hydrolase family 16 protein (GH16; AQQ75061.1).



Supplementary Figure 23. Microsynteny of glycoside hydrolase family 16 protein (GH16; WP_052600908.1).



Supplementary Figure 24. Microsynteny of efflux RND transporter permease subunit (WP_163176792.1).



Supplementary Figure 25. Microsynteny of NADH dehydrogenase (WP_034862301.1).



Supplementary Figure 26. Microsynteny of AzlD domain-containing protein (WP_095524423.1).



Supplementary Figure 27. Microsynteny of anaerobic sulfatase maturase (SBW02910.1).



Supplementary Figure 28. Microsynteny of alpha-2-macroglobulin (OYY43986.1).



Supplementary Figure 29. Microsynteny of SYLF domain-containing protein (RPH48231.1).



Supplementary Figure 30. Tissue differential expression of horizontal transferred WP_189008864.1 Glucose/arabinose dehydrogenases in adult female *T. tuberculata* (n=3 biologically independent *T. tuberculata*). Data are presented as mean values +/- SEM. Individual data points are present on each bar. The vertical error bars indicate the standard error of mean (SEM) among the data points in each experiment.



Supplementary Figure 31. Phylogenetic tree of acetyl-CoA C-acetyltransferase (ACAT). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan diglyceride acyltransferase (DGAT).



Supplementary Figure 32. Phylogenetic tree of hydroxymethylglutaryl-CoA synthase (**HMGS**). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan fatty acid synthase (FAS).



Supplementary Figure 33. Phylogenetic tree of hydroxymethylglutaryl-CoA reductase (**HMGR**). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan aldo-keto reductase (AKR1).



Supplementary Figure 34. Phylogenetic tree of mevalonate kinase (MK) and phosphomevalonate kinase (PMK). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity.



Supplementary Figure 35. Phylogenetic tree of diphosphomevalonate decarboxylase (**DPMD**). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using malonyl-CoA decarboxylase (MLYCD).



Supplementary Figure 36. Phylogenetic tree of isopentenyl-diphosphate delta-isomerase (**IPPI**). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1 (HSD3B1).



Supplementary Figure 37. Phylogenetic tree of farnesyl pyrophosphate synthase (FPPS). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan geranyl diphosphate synthase (GPPS).



Supplementary Figure 38. Phylogenetic tree of farnesyltransferase beta (FNTB). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan geranylgeranyltransferase (FNTA).



Supplementary Figure 39. Phylogenetic tree of ste24 endopeptidase (ste24). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using protozoan M48 metalloprotease.



Supplementary Figure 40. Phylogenetic tree of protein-S-isoprenylcysteine O-methyltransferase (ICMT). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan steroid 5 alpha-reductase 1 (SRD5A1).



Supplementary Figure 41. Phylogenetic tree of prenylcysteine oxidase (PCYOX1). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan selenium binding protein 1 (SELENBP1).



Supplementary Figure 42. Phylogenetic tree of aldehyde dehydrogenase 3 (ALDH3). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using arthropod aldehyde dehydrogenase 7 (ALDH7).



Supplementary Figure 43. Phylogenetic tree of juvenile hormone acid methyltransferase (JHAMT). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using arthropod farnesoic acid methyltransferase (FAMeT).



Supplementary Figure 44. Phylogenetic tree of juvenile hormone epoxide hydrolase (**JHEH**). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using mammalian epoxide hydrolase (EPHX).



Supplementary Figure 45. Phylogenetic tree of juvenile hormone diol kinase (JHDK). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan hydroxylysine kinase (HYKK).



Supplementary Figure 46. Phylogenetic tree of cytosolic juvenile hormone binding protein (cJHBP)/glyoxalase domain containing protein 4 (GLOD4). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using plant glyoxalase (GLXI).



Supplementary Figure 47. Phylogenetic tree of ultraspiracle (USP) and ecdysteroid receptor (EcR). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity.



Supplementary Figure 48. Phylogenetic tree of steroid receptor coactivator (SRC). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using other bHLH-PAS containing family proteins.



Supplementary Figure 49. Phylogenetic tree of methoprene receptor (Met). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using other bHLH-PAS containing family proteins.



Supplementary Figure 50. Phylogenetic tree of heat shock protein 83 (Hsp83). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using heat shock protein 70 (Hsp70).



Supplementary Figure 51. Phylogenetic tree of allatotropin receptor (ATR). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan arginine vasopressin receptor 1A (AVPR).



Supplementary Figure 52. Phylogenetic tree of allatostatin receptor (ASR). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan glutamate receptor (GluR).



Supplementary Figure 53. Phylogenetic tree of S-adenosylmethionine synthase (MAT). Tree is constructed with the neighbor-Joining (NJ) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan geranylgeranyltransferase (FNTA). The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site.



Supplementary Figure 54. Phylogenetic tree of S-adenosylhomocysteinase (AHC). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan aminomethyltransferase (AMT).



Supplementary Figure 55. Phylogenetic tree of adenosine kinase (AK). Tree is constructed with the maximum likelihood (ML) method with 1000 bootstrap replicates. Only bootstrap values larger than 80% are indicated for clarity. Tree rooted using metazoan ketohexokinase (KHK).



Supplementary Figure 56. Sequence alignment of centipede juvenile hormone acid methyltransferase (JHAMT). Alignment shows a highly conserved S-adenosyl methionine (SAM) binding domain among the genes. Ttu, *T. tuberculata*; Rim, *R. immarginata*; Lni, *L. niger*; Sma, *S. maritima*.



Supplementary Figure 57. Microsteny of JHAMT in centipede genomes. Ttu, *T. tuberculate*; Rim, *R. immarginata*; Lni, *L. niger*; Sma, *S. maritima*; LRTOMT, leucine-rich repeat-containing protein 51; HP, hypothetical protein; RNF121, RING finger protein 121; FUCA, alpha-L-fucosidase; AGPAT1, 1-acyl-sn-glycerol-3-phosphate acyltransferase alpha; JHAMT, juvenile hormone acid O-methyltransferase; C4orf33, UPF0462 protein C4orf33 homolog; POSTN, periostin; KLHL5, kelch-like protein 5; EMC, ER membrane protein complex subunit; TGFBI, transforming growth factor-beta-induced protein ig-h3; CETN, centrin; TPR, tetratricopeptide repeat protein; GGH, gamma-glutamyl hydrolase; SQSTM1, sequestosome-1; DDX49, ATP-dependent RNA helicase DDX49; Shrm, protein Shroom; UBN1, ubinuclein; CA10, carbonic anhydrase-related protein 10; TRP, transient-receptor-potential-like protein, RTN4IP1, reticulon-4-interacting protein 1 homolog, mitochondrial-like; PTK, tyrosine-protein kinase; MTH2, G-protein coupled receptor Mth2; AC78C, adenylyl cyclase 78C/adenylate cyclase type 8.



Supplementary Figure 58. Hormone measurement of sesquiterpenoids farnesoic acid and methyl farnesoate, in biologically independent adult *T. tuberculata* (male<1.5cm, n=5; male>2.0cm, n=5; female<2.0cm, n=5; female>2.5cm, n=5) and *H. holstii* (male, n=4; female, n=4). Ttu, *T. tuberculata*; Hho, *H. holstii*. Data are presented as mean values +/- SEM. Individual data points are present on each bar. The vertical error bars indicate the standard error of mean (SEM) among the data points in each experiment.



Supplementary Figure 59. Myriapod repeat location plots. Left-hand plots indicate the amount of each repeat class identified in each genomic compartment in base pairs (bp). Right-hand plots indicate the relative proportion of each genomic compartment occupied by each repeat type.



Supplementary Figure 60. Myriapod repeat landscape plots.



Supplementary Figure 61. Repeat density karyotype plot for Helicorthomorpha holstii.



Supplementary Figure 62. Repeat density karyotype plot for *Trigoniulus corallinus*.



Supplementary Figure 63. Repeat density karyotype plot for *Thereuonema tuberculata*.



Supplementary Figure 64. Bar chart showing the pattern by which different myriapod species share individual TE families. Original figure can be found in Figshare: "sharedRepeatsSetmap_page-0001.zip" <u>https://doi.org/10.6084/m9.figshare.15088722</u>

Supplementary Discussion - Genome size and transposable element content

The transposable elements (TEs) of Myriapoda are poorly described. Here, we provide TE annotations, TE libraries, and a comparative TE analysis for nine genomes from across myriapod diversity (four centipedes, and five millipedes) to describe the abundance, diversity, and genomic distribution of myriapod TEs. We identify considerable variability in the proportion of the genome composed of TEs in myriapods, ranging from 17.84% in the millipede *G. maerens*, to 51.67% in the millipede *T. corallinus*, as well as the relative quantity of TEs present in myriapod genomes, ranging from 26.74Mb in the millipede *G. maerens*, to 1609.95Mb in the centipede *L. niger* (Supplementary Data 2). Overall, millipede genomes have a higher mean TE content (millipede $\overline{x} = 43.17\%$, centipede $\overline{x} = 31.86\%$), but centipede genomes have a much higher relative quantity of TEs (millipede $\overline{x} = 129.14$ Mb, centipede $\overline{x} = 945.78$ Mb).

Myriapod genomes contain representation from all major TE types except SINEs which are absent or only present <1% (except *T. corallinus* 3.43%). Retroelements comprise 3.00 - 26.69% of total genomic content, while DNA TEs comprise 2.05 - 20.34%, and rolling-circle elements comprise 0.17 - 5.22% (Supplementary Data 2). There is a relatively high proportion of unclassified repeats present in myriapod genomes (6.27 - 23.98%, Supplementary Data 2), as expected given that TEs are poorly explored in the group. DNA TEs and unclassified elements are often most abundant, while a few genomes instead show large proportions of LINEs (*A. tonginus, T. corallinus*) or LTR TEs (*L. niger, S. maritima*) (Supplementary Data 2, Figure 2A).

Considering TE distribution within myriapod genomes, most insertions occur in intergenic regions (non-coding regions >20Kb from annotated genes), except *S. maritima*, where the majority occur within gene flanks (<20kb from annotated genes) (Supplementary Figure 59-60). The genomic compartment with the second highest proportion of repeats is generally gene flanks (except *S. maritimia*). As expected, relatively few TE insertions are present within gene exons, but there is a much greater presence of TEs within gene introns in many genomes (Supplementary Figure 59-60).

For three myriapod genomes with chromosome-level assemblies, we were able to examine karyotype TE density plots (*H. holstii, T. tuberculata, T. corallinus,* Supplementary Figures 61-63). The karyotype plots for *T. tuberculata* and *T. corallinus* suggest a relatively even spread of TEs across chromosomes (Supplementary Figures 59-60 & 62-63). However,

the karyotype plot for *H. holstii* indicates the presence of several TE hotspots: SczTNLB_3097, high abundance of all TEs except SINEs over the entire short chromosome; SczTNLB_7440, high abundance of all TEs except SINEs at the start (0-2Mb); SczTNLB_6789, dense cluster of LINEs, LTR TEs, and DNA TEs off-centre (18-20Mb); SczTNLB_5799, cluster of all TEs except SINEs at the start (0-2Mb); SczTNLB_6657, dense clusters of DNA TEs at the start (0-1Mb) and end (38Mb); SczTNLB_3777, cluster of all TEs except SINEs at the end (16Mb); SczTNLB_2740, cluster of DNA TEs at the start (0Mb) and end (16Mb); SczTNLB_5985, dense cluster of all TEs except SINEs at the start of DNA TEs at the start (0Mb), and off-centre (7Mb) (Supplementary Figure 61). It is currently unclear what drives the accumulations of TEs in genomic hotspots in the millipede *H. holstii* compared to the millipede *T. coralinus* and the centipede *T. tuberculata*.