Widespread retention of ohnologs in key developmental gene families following whole genome duplication in arachnopulmonates

Amber Harper[1], Luis Baudouin Gonzalez[1], Anna Schönauer[1], Michael Seiter[2], Michaela Holzem[1,3], Saad Arif[1,4], Alistair P. McGregor[1,4*], Lauren Sumner-Rooney[5,*]

[1]Department of Biological and Medical Sciences, Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, OX3 0BP, United Kingdom.

[2]Department of Evolutionary Biology, Unit Integrative Zoology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.

[3]Division of Signalling and Functional Genomics, German Cancer Research Centre (DKFZ), Heidelberg, Germany and Department of Cell and Molecular Biology, Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany.

[4]Centre for Functional Genomics, Oxford Brookes University, Oxford, OX3 0BP, United Kingdom.

[5]Oxford University Museum of Natural History, University of Oxford, Oxford OX1 3PW, United Kingdom.

*Corresponding authors: amcgregor@brookes.ac.uk (APM) and lauren.sumner-rooney@oum.ox.ac.uk (LSR).

Abstract

Whole genome duplications (WGD) have occurred multiple times in the evolution of animals, including in the lineages leading to vertebrates, teleosts, horseshoe crabs and arachnopulmonates. These dramatic genomic events initially produce a wealth of new genetic material, which is generally followed by extensive gene loss. It appears that developmental genes such as homeobox genes, signalling pathway components and microRNAs, however, tend to be more frequently retained in duplicate following WGD (ohnologs). These not only provide the best evidence for the occurrence of WGD, but an opportunity to study its evolutionary implications. Although these genes are relatively well studied in the context of vertebrate WGD, genomic and transcriptomic data for independent comparison in other groups are scarce, with patchy sampling of only two of the five extant arachnopulmonate orders. To improve our knowledge of developmental gene repertoires, and their evolution since the

arachnopulmonate WGD, we sequenced embryonic transcriptomes from two additional spider species and two whip spider species and surveyed them for three important gene families: Hox, Wnt and frizzled. We report extensive retention of ohnologs in all four species, further supporting the arachnopulmonate WGD hypothesis. Thanks to improved sampling we were able to identify patterns of likely ohnolog retention and loss within spiders, including apparent differences between major clades. The two amblypygid species have larger ohnolog repertoires of these genes than both spiders and scorpions; including the first reported duplicated *Wnt1/wg*, the first *Wnt10* recovered in an arachnid, and broad retention of frizzled genes. These insights shed light on the evolution of the enigmatic whip spiders, highlight the importance of the comparative approach within lineages, and provide substantial new transcriptomic data for future study.

Introduction

The duplication of genetic material is widely accepted to be an important contributor to the evolution of morphological and physiological innovations (Ohno 1970; Zhang 2003). The most dramatic example of this is whole genome duplication (WGD), when gene copy numbers are doubled and retained paralogs or ohnologs can then share ancestral functions (subfunctionalisation) and/or evolve new roles (neofunctionalization) (Ohno 1970; Force et al. 1999; Lynch and Conery 2000). The occurrence of two rounds (2R) of WGD in the early evolution of vertebrates has long been associated with their taxonomic and morphological diversity (e.g. Ohno 1970; Holland et al. 1994; Dehal and Boore 2005; Holland 2013a), and a subsequent 3R in teleosts is linked to their success as the most diverse vertebrate group (e.g. Meyer and Schartl 1999; Glasauer and Neuhauss 2014). However, this remains controversial and difficult to test (Donoghue and Purnell 2005) and in several animal lineages there is not a clear association between WGD and diversification (Mark Welch et al. 2008; Flot et al. 2013; Havlak et al. 2014; Kenny et al. 2016; Nong et al. 2020). Along with vertebrates, chelicerates also appear to be hotspots of WGD, with up to three rounds reported in horseshoe crabs (Kenny et al. 2016; Nong et al. 2020), one in the ancestor of arachnopulmonates (spiders, scorpions, and their allies) (Schwager et al. 2017), and potentially two further rounds within the spider clade Synspermiata (Král et al. 2019). Chelicerates demonstrate a highly variable body plan, occupy a wide range of habitats and ecological niches, and have evolved a variety of biologically important innovations such as venoms and silks (Schwager et al. 2015). They therefore offer an excellent opportunity for comparison with vertebrates

concerning the implications of WGD for morphological and taxonomic diversity, and genome evolution in its wake.

Over the past 20 years, the house spider *Parasteatoda tepidariorum* has emerged as a model species to study the impacts of WGD on arachnid evolution and development. Genomic and functional developmental studies have found retained ohnologs of many important genes, with evidence for neo- and subfunctionalisation in many of these compared to single-copy orthologs in arachnids lacking WGD (Janssen et al. 2015; Leite et al. 2016, 2018; Turetzek et al. 2016, 2017; Schwager et al. 2017; Baudouin-Gonzalez et al. 2020). Work on the scorpions *Centruroides sculpturatus* and *Mesobuthus martensii* has consistently complemented findings in *P. tepidariorum*, with genomic studies recovering many ohnologs in common with spiders (Di et al. 2015; Sharma et al. 2015; Leite et al. 2018). Although high quality genome assemblies are required for the analysis of synteny between gene duplicates, these remain relatively scarce in arachnids. Work on the *P. tepidariorum, Ce. sculpturatus* and *Me. martensii* genomes has been complemented by targeted studies of individual gene families and transcriptomic surveys (e.g. Schwager et al. 2007; Sharma et al. 2012; Leite et al. 2018; Gainett and Sharma 2020). Combined with phylogenetic analyses, the identification of large-scale gene duplications using transcriptomics can provide evidence of WGD events and their timing in the history of arachnid evolution. Although transcriptomes can yield variant sequences of individual genes, from different alleles or individuals in mixed samples, these are generally easy to filter out from truly duplicated loci owing to substantial sequence divergence in the latter. They also offer the double-edged sword of capturing gene expression, rather than presence in the genome; pseudogenised or silenced duplicates are not detected, but neither are functional genes if they are not expressed at the sampled timepoint or tissue. Such studies have produced strong additional evidence for an ancestral WGD, with patterns of duplication coinciding with our expectations for arachnopulmonate ohnologs (Clarke et al. 2014, 2015; Sharma et al. 2015; Turetzek et al. 2017; Leite et al. 2018; Gainett and Sharma 2020; Gainett et al. 2020).

Comparison of WGD events among arachnopulmonates, horseshoe crabs and vertebrates indicates that despite extensive gene loss following duplication events, certain gene families are commonly retained following duplication (Holland et al. 1994; Schwager et al. 2007, 2017; Kuraku and Meyer 2009; Di et al. 2015; Sharma et al. 2015; Kenny et al. 2016; Leite et al. 2016, 2018). These typically include

genes from the conserved developmental 'toolkit' of transcription factors (TFs), cell signalling ligands and receptors, and microRNAs (Erwin 2009). Among these, several have stood out as focal points in the study of gene and genome duplications. The Hox gene group of homeobox genes regulate the identity of the body plan along the antero-posterior axis of all bilaterian animals (McGinnis and Krumlauf 1992; Abzhanov et al. 1999; Carroll et al. 2005; Pearson et al. 2005; Hueber and Lohmann 2008; Holland 2013b). Four clusters of these key developmental genes were retained after 1R and 2R in vertebrates (Holland et al. 1994; Meyer and Schartl 1999; Kuraku and Meyer 2009; Pascual-Anaya et al. 2013), and the arachnopulmonate WGD is evident in the almost universal retention of Hox gene duplicates in sequenced genomes, with two ohnologs of all ten arthropod Hox genes in the scorpion *M. martensii* (Di et al. 2015; Leite et al. 2018), all except *Hox3* being represented by two copies in *Ce. sculpturatus* (Leite et al. 2018), and all except *fushi tarazu* (*ftz*) in *P. tepidariorum* (Schwager et al. 2017). Systematic studies of Hox gene expression patterns in the latter demonstrated that all nine pairs of Hox paralogs exhibit signs of sub- or neofunctionalization (Schwager et al. 2017). This high level of retention and functional divergence lends strong support to the importance of Hox gene duplication in the evolution of the arachnopulmonate body plan, and further consolidates the position of this family as a key indicator of WGD.

In addition to TFs, the ligands and receptors of some signalling pathways of the developmental toolkit (e.g. Hedgehog, Wnt, TGF-ß, NHR) also demonstrate higher copy numbers in vertebrates and other groups subject to WGD, including arachnopulmonates (Holland et al. 1994; Meyer and Schartl 1999; Shimeld 1999; Pires-daSilva and Sommer 2003; Cho et al. 2010; Janssen et al. 2010, 2015; Hogvall et al. 2014). The Wnt signal transduction pathway plays many important roles during arthropod development, including segmentation and patterning of the nervous system, eyes and gut (Erwin 2009; Murat et al. 2010). In the canonical pathway, Wnt ligands bind to transmembrane receptors, such Frizzled, to trigger translocation of ß-catenin to the nucleus and mediate regulation of gene expression (Cadigan and Nusse 1997; Hamilton et al. 2001; Logan and Nusse 2004; van Amerongen and Nusse 2009). There are thirteen subfamilies of Wnt genes found in bilaterians, as well as multiple receptor families and downstream components. In contrast to the extensive retention of Hox ohnologs following WGD, Wnt duplicates in *P. tepidariorum* appear to be restricted to *Wnt7* and *Wnt11*, with the remaining eight subfamilies represented by single genes (Janssen et al. 2010). However, these are the only reported Wnt gene duplications in arthropods despite several recent surveys (Bolognesi et al. 2008;

Murat et al. 2010; Hayden and Arthur 2013; Meng et al. 2013; Hogvall et al. 2014; Janssen and Posnien 2014; Holzem et al. 2019), and beyond *P. tepidariorum* no other arachnopulmonates have been systematically searched. Similarly, duplications within the four *frizzled* gene subfamilies appear to be restricted to arachnopulmonates among arthropods, wherein only *fz4* is duplicated in both *P. tepidariorum* and *M. martensii* (Janssen et al. 2015).

Several Wnt families have also been retained after the 1R and 2R events in vertebrates, for example there are two copies each of *Wnt2, Wnt3, Wnt5, Wnt7, Wnt8, Wnt9*, and *Wnt10* in humans (Miller 2001; Janssen et al. 2010). However, no subfamilies are represented by three or four copies in humans and so there is some consistency with arachnopulmonates in that the Wnts may be more conservative markers of WGD, to be used in combination with Hox and other homeobox genes.

The extensive and consistent retention of key developmental genes like Hox genes apparent in *P. tepidariorum* and *Ce. sculpturatus*, and Wnt genes in *P. tepidariorum*, strongly support the occurrence of an ancestral WGD in arachnopulmonates. However, data are only available for a handful of species so far, resulting in very patchy taxonomic sampling. For example, only *P. tepidariorum* and *Pholcus phalangioides* have been comprehensively surveyed for homeobox genes among spiders (Leite et al. 2018), omitting the large and derived retrolateral tibial apophysis (RTA) clade, which includes jumping spiders, crab spiders and other free hunters, and the systematic identification of Wnt genes has been restricted to only *P. tepidariorum*. Spiders and scorpions are by far the most speciose of the arachnopulmonates, and there may be additional diversity in their repertoires of these important developmental gene families of which we are not yet aware.

In addition, and perhaps more urgently, only two of the five arachnopulmonate lineages have dominated the field thus far; sufficient genomic information for comparison is lacking beyond spiders and scorpions. Also represented in Arachnopulmonata are the amblypygids (whip spiders), relatively understudied and enigmatic animals comprising around 190 extant species. They exhibit highly derived morphology of the pedipalps, which are adapted to form raptorial appendages, and of the first pair of walking legs, which are antenniform and can comprise more than 100 segments (Weygoldt 2009). Despite the scarcity of transcriptomic or genomic data for amblypygids (whip spiders) (Garb et al. 2018; though see Gainett and Sharma 2020; and Gainett et al. 2020 for recent advances), their widely accepted position within Arachnopulmonata implies that they were also subject to an ancestral WGD. A recent survey of

the *Phrynus marginemaculatus* transcriptome supported this in the recovery of multiple duplicate Hox and leg gap genes (Gainett and Sharma 2020). Particularly given the derived nature of their appendages, this group could shed substantial light on genomic and morphological evolution following WGD.

To better understand the genomic consequences of WGD in a greater diversity of arachnopulmonate lineages, we sequenced *de novo* embryonic transcriptomes from two spiders belonging to the derived RTA clade and two amblypygids. We surveyed Hox, Wnt and frizzled genes in these species and existing genomic and transcriptomic resources for as examples for comparison with other arachnids, both with and without an ancestral WGD, improving sampling at both the order and sub-order levels.

Materials and methods

Embryo collection, fixation and staging

Embryos of mixed ages were collected from captive females of the amblypygids *Charinus acosta* (Charinidae; parthenogenetic, collected at one day, one month and two months after the appearance of egg sacs) and *Euphrynichus bacillifer* (Neoamblypygi: Phrynichidae; mated, collected at approximately 30% of development), the wolf spider *Pardosa amentata* (collected in Oxford, UK) and mixed stage embryos of the jumping spider *Marpissa muscosa* (kindly provided by Philip Steinhoff and Gabriele Uhl) and stored in RNAlater.

Transcriptomics

Total RNA was extracted from mixed aged embryos, pooled by species, of *C. acosta, E. bacillifer, Pa. amentata* and *M. muscosa* using QIAzol according to the manufacturer's instructions (QIAzol Lysis Reagent, Qiagen). Illumina libraries were constructed using a TruSeq RNA sample preparation kit (including polyA selection) and sequenced using the Illumina NovaSeq platform (100bp PE) by Edinburgh Genomics. Quality of raw reads was assessed using FastQC v0.11.9 (Andrews 2010). Erroneous k-mers were corrected using rCorrector (default settings, Song and Florea 2015) and unfixable read pairs (from low-expression homolog pairs or containing too many errors) were discarded using a custom Python script (available at https://github.com/harvardinformatics/TranscriptomeAssemblyTools/blob/master/FilterUncorrectabledPEfastq.py courtesy of Adam Freeman). Adapter sequences were identified and removed and low

quality ends (phred score cut-off = 5) trimmed using TrimGalore! v0.6.5 (available at https://github.com/FelixKrueger/TrimGalore). *De novo* transcriptome assembly was performed using only properly paired reads with Trinity v2.10.0 (Haas et al. 2013) using default settings. Transcriptome completeness was evaluated on the longest isoform per gene using BUSCO v4.0.2 (Seppey et al. 2019) along with the arachnid database (arachnida_odb10 created on 2019-11-20; 10 species, 2934 BUSCOs) and the arthropod database (arthropoda_odb10 created on 2019-11-20; 90 species, 1013 BUSCOs).

Identification of gene candidates

To identify Wnt, frizzled and Hox gene candidates in *C. acosta*, *E. bacillifer*, *Pa. amentata* and *M. muscosa*, we performed TBLASTN (version 2.6.0+) searches of the assembled embryonic transcriptome, using Wnt and frizzled protein sequences previously identified in *P. tepidariorum* (Janssen et al. 2010, 2015) and homeodomain protein sequences of the *Drosophila melanogaster* Hox genes from HomeoDB2 (Zhong and Holland 2011). Existing Wnt and Frizzled protein predictions were collected for *Ce. sculpturatus* from NCBI (PRJNA422877, Supplementary Data Tables 1-3). Predicted protein sequences were obtained using the Translate ExPASy online tool (https://web.expasy.org/translate/) and the standard genetic code.

Phylogenetic analysis

Transcript identity was confirmed by reciprocal BLAST and the construction of maximum likelihood trees. Amino acid sequences of Hox, Wnt and frizzled family genes of known identity from selected arthropods (*Bicyclus anynana, Bombyx mori, Daphnia pulex, Drosophila melanogaster, Parage aegeria, Strigamia maritima,* and *Tribolium castaneum*) and an onychophoran (*Euperipatoides kanangrensis*) were retrieved from NCBI (Accession numbers: Supplementary Tables 1-3). Alignments of full protein sequences were performed in Clustal Omega using default parameters (Goujon et al. 2010; Sievers et al. 2011). Maximum likelihood trees were generated from whole-sequence alignments to assign genes to families and study the relationship between candidate duplicates. Phylogenetic analyses were performed in IQ-Tree (v2.0.3, Nguyen et al. 2015) using ModelFinder to identify optimal substitution models (VT+F+R10 for Hox, LG+R8 for *Wnt*, JTT+5 for *fz*; Kalyaanamoorthy et al. 2017) and 1000 bootstrap replicates. Trees were visualised in FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Identification of duplicate genes

Where more than one sequence was identified as a potential candidate for a single gene, several factors were examined to eliminate the possibility that they were isoforms, individual variants, or fragments of the same gene. Nucleotide and protein alignments (Supplementary Data Files 1-4) were inspected and if sequences did not overlap, the shorter one was discarded. If sequences showed very high similarity from alignments, or potential duplicates resolved as short-branch sister pairs in phylogenetic analysis, as expected from isoforms or individual variants, the shorter sequence was discarded. Candidate duplicates that passed these tests were also BLASTed back against source transcriptomes to confirm their origin.

Results and Discussion

Transcriptome assemblies

To further study the outcomes of WGD in the ancestor of arachnopulmonates we carried out RNA-Seq on embryos of two further spider species, *Pa. amentata* and *M. muscosa*, and two species of amblypygids, *C. acosta* and *E. bacillifer*.

RNA-Seq for the four species produced between 222,479,664 and 272,844,971 raw reads, reduced to 211,848,357 and 260,853,757 after processing. Trinity assembled between 184,142 and 316,021 transcripts in up to 542,344 isoforms (Table 1). Contig N50 ranged from 592 bp in *M. muscosa* to 978 bp in *E. bacillifer*, and from 1461 bp (*M. muscosa*) to 2671 bp (*E. bacillifer*) in the most highly expressed genes (representing 90% of total normalised expression) (Table 1).

Transcriptomes were found to be between 83.7% (*C. acosta*) and 89.4% (*E. bacillifer*) complete according to BUSCO scores compared to the arthropod database, with between 3.5% and 9.5% duplicated BUSCOs. Compared to the arachnid databases, transcriptomes were 82%-90.1% complete for single-copy BUSCOs and contained between 5.3%-12.9% duplicated BUSCOs (Table 1).

To explore the extent of duplication in these arachnopulmonates we then surveyed the copy number of Hox, Wnt and Frizzled genes in their transcriptomes in comparison to other arachnids. It is important to note that the absence of genes recovered from transcriptomes does not eliminate the possibility that they are present in the genome, as the transcriptomes will only capture genes expressed at the relevant point in development. Mixed-stage embryonic samples may yield more transcripts for the same reason.

Duplication of Hox genes in spiders and amblypygids

The Hox gene repertoires for the two spiders are largely consistent with previous surveys of the *P. tepidariorum* genome (Figure 1), which has two copies of all genes except for *ftz*. There are three exceptions: the recovery of a single copy of *Deformed* (*Dfd*) in *M. muscosa*, the apparent absence of *ftz* in *Pa. amentata*, and the presence of one, rather than two, copies of *Hox3* in both species (Figure 1). Perhaps more comparable is previous analysis of the embryonic transcriptome of the synspermiatan spider *Ph. phalangioides*, which detected single transcripts of *labial* (*lab*)*, ftz* and *Ultrabithorax* (*Ubx*), but two copies of the remaining Hox genes (*proboscipedia, Hox3, Dfd, Sex combs reduced, Antennapedia, Abdominal-A, And abdominal-B*; Leite et al. 2018). From these combined data it seems likely that having a single copy of *ftz* is common across all spiders, but both copies are retained in all other arachnopulmonates studied to date (Figures 1-2). This is consistent with the loss of one copy of *ftz* in the common ancestor of all spiders, following WGD. The absence of *Hox3* duplicates in *M. muscosa* and *Pa. amentata* could indicate a lineage-specific loss in the RTA clade, which unites salticids, lycosids and their allies. Indeed, only one copy of *Hox3* has been recovered in *Cupiennius salei*, a ctenid also belonging to the RTA clade (Schwager et al. 2007). Other apparent losses, of *Dfd-A* in *M. muscosa* and *Ubx* and *lab-A* in *Ph. phalangioides*, may be specific to Salticidae and Synspermiata/Pholcidae, respectively, if they are absent from the genome.

Both amblypygids exhibit extensive duplication of Hox genes, in line with expectations following the arachnopulmonate WGD (Figures 1-2). *Charinus acosta* appears to have two copies of all surveyed Hox genes except for *pb.* We recovered single copies of *pb, Scr* and *Ubx* in *E. bacillifer*, but two copies of *lab, zen, Dfd, ftz, Antp, abdA* and *abdB*. The absence of a second copy of *pb* in both *C. acosta* and *E. bacillifer*, which are relatively distantly related within Amblypygi, suggests a loss in the common ancestor of all amblypygids. This is also supported by a recent survey of *Phrynus marginemaculatus*, which recovered a single copy of *pb* but duplicates of all other Hox genes (Gainett and Sharma 2020). Embryos of *C. acosta* were collected at multiple stages of development, supporting the hypothesis that this may be a true loss, rather than absence of expression at a particular developmental stage. However, the apparent additional absence of *Scr* and *Ubx* duplicates in *E. bacillifer* could equivocally indicate lineage-specific losses or absence of expression at a single timepoint.

The duplication of Hox genes is consistent among the three arachnopulmonate orders studied to date, and specific repertoires appear to be fairly conserved at the order level (this study; Schwager et al. 2007, 2017; Cao et al. 2013; Di et al. 2015; Leite et al. 2018). Given that this is the level at which overall body plans are conserved, this is perhaps not surprising. The potential loss of a *Hox3* duplicate in the spider RTA clade (in *M. muscosa* and *Pa. amentata*, and *Cu. salei*, Schwager et al. 2007) is an unusual example of infraorder variation in Hox repertoires. Although initial analyses found that the expression patterns of the two *Hox3* ohnologs overlapped in *P. tepidariorum* (Schwager et al. 2017), both duplicates were still expressed. As other intraorder losses of Hox genes were only observed in single species from embryonic transcriptomes, it would be premature to conclude that they are genuinely absent from the genome.

Thanks to the relatively conserved expression patterns of Hox genes along the anterior-posterior axis of chelicerates, we can begin to make tentative inferences about the possible macroevolutionary implications of duplication and loss. For example, an anticipated duplicate of *pb* has been lost in both amblypygids but persists in spiders and scorpions. In spiders, both *pb* paralogs are expressed in the pedipalp and leg-bearing segments, separated temporally (Schwager et al. 2017). Given the highly derived nature of the raptorial pedipalps and the antenniform first pair of walking legs in amblypygids, it is perhaps surprising that this duplicate was not retained. However, this might indicate that other Hox genes expressed in the anterior prosomal segments (e.g. *lab*, *zen*, or *Dfd*) may contribute to these morphological innovations. A good candidate for future study might be *lab*: a single ortholog is expressed in both the pedipalps and the first walking leg in the harvestman *Phalangium opilio* (Sharma et al. 2012), and expression patterns and experimental manipulation provide evidence for functional divergence between the two *lab* paralogs, also expressed in the pedipalps and first walking legs, in *P. tepidariorum* (Pechmann et al. 2015; Schomburg et al. 2020).

Wnt gene repertoires exhibit both between- and within-lineage diversity

Consistent with previous studies of *P. tepidariorum*, we found representatives of ten *Wnt* subfamilies in *M. muscosa* and *Pa. amentata* transcriptomes, with all three spiders lacking *Wnt3*, *Wnt9*, and *Wnt10*. The absence of *Wnt3* (in both amblypygids and both spiders) is consistent with all other protostomes (Janssen et al. 2010; Murat et al. 2010; Hogvall et al. 2014), but the absence of *Wnt9* and *Wnt10* in spiders indicates losses in the spider ancestor. We did not recover duplicates of *Wnt2, Wnt8-10, Wnt16*

or *WntA* in any arachnopulmonate lineage. This suggests loss shortly after WGD in the common ancestor of all arachnopulmonates.

Both *M. muscosa* and *Pa. amentata* expressed two copies of *Wnt7* and *Wnt11*, in line with *P. tepidariorum*, as well as a second copy of *Wnt4* that is absent in *P. tepidariorum*.

Representation of the Wnt subfamilies in our amblypygid transcriptomes is higher than any other arachnid studied to date, including those with high-quality genome assemblies (Janssen et al. 2010; Hogvall et al. 2014; Holzem et al. 2019). We recovered transcripts from twelve out of thirteen subfamilies in *C. acosta* (missing *Wnt3*) and eleven in *E. bacillifer* (missing *Wnt3* and *Wnt9*) (Figure 3). Two copies of *Wnt1/wg, Wnt4* and *Wnt7* were recovered for both species, with an additional duplicate of *Wnt6* in *C. acosta.* Unlike all other arachnopulmonates surveyed, we only identified a single *Wnt11* gene in amblypygids, suggesting a lineage-specific loss following WGD. Most of the duplicate Wnt genes identified in our analysis appear to be likely ohnologs; conclusive confirmation of this requires synteny analysis of fully sequenced genomes, but the relationships between paralogs resolved by phylogenetic analysis generally do not support more recent tandem duplications. Duplicates of *Wnt7* were previously identified in *P. tepidariorum* (Janssen et al. 2010), and are recovered in this study from all four transcriptomes and existing sequence data from *Ce. sculpturatus*. These paralogs did not resolve as sister pairs in phylogenetic analyses and sequence similarity between paralogs was low (61-73%, Supplementary Data File 2). *Wnt7* ortholog groups formed two well-supported clades within spiders and amblypygids, suggesting retention of *Wnt7* ohnologs in these groups following the arachnopulmonate WGD (Figure 4). The two *Wnt7* sequences from *Ce. sculpturatus* also demonstrated low similarity (66%, Supplementary Data File 2) but did not resolve in separate clades and may indicate a lineage-specific duplication. *Wnt11* duplicates were recovered from transcriptomes of *M. muscosa* and *Pa. amentata*, and from the published genome of *Ce. sculpturatus*. These formed two separate and well-supported clades (78% and 95%, Figure 4), also including genomic sequences from *P. tepidariorum* duplicates. Sequence similarity between paralogs was very low (40-50%, Supplementary Data File 2); combined with their phylogenetic placement, we conclude that this reflects the likely retention of ohnologs following WGD. Only the amblypygids appear not to have retained two copies of *Wnt11* following ancestral duplication. Paired *Wnt4* paralogs detected in *M. muscosa* and *Pa. amentata* form well supported clades with duplicates recovered in the amblypygids (bootstrap ≥ 96%; Figure 4)

and show substantial sequence divergence within species (56-65% similarity, Supplementary Table 2), indicating that they are again likely to represent retained ohnologs following the arachnopulmonate WGD, despite being lost in the lineage to *P. tepidariorum*.

We have identified two copies of *Wnt1/wg* in both the amblypygid transcriptomes and in the previously published genome of the scorpion *Ce. sculpturatus*. To the best of our knowledge this is the first time a duplication of *Wnt1/wg* has been reported in any animal surveyed to date. Since this is highly unusual it requires critical interpretation. We can eliminate the possibility of individual variation in *C. acosta,* as embryos are produced by parthenogenesis and are therefore clones, and in *Ce. sculpturatus,* as the sequences were recovered from a single individual's published genome (Supplementary Table 2). Sequence similarity between paralogs was low (72-76%, Supplementary Data File 2), even compared to similarity between *Wnt1* orthologs at the order level (e.g. 91% between *M. muscosa* and *Pa. amentata*), reducing the likelihood that we are detecting allelic variation within individuals. We also inspected nucleotide alignments and found lower paralog sequence similarity than evident from amino acid sequences (65-69%, Supplementary Data File 3), indicating synonymous evolution. Although synteny analysis is required for conclusive confirmation, our phylogenetic analysis indicates that the amblypygid duplicates are likely to be ohnologs retained from the arachnopulmonate WGD, as they form separate, well-supported clades with other arachnid *Wnt1*s (bootstrap values ≥79%; Figure 4). The resolution of the *Ce. sculpturatus Wnt1* paralogs had lower support and their relationship is therefore more ambiguous. The current placement of *Cs-Wnt1-2* as sister to *Cs-Wnt1-1+(Ca-Wnt1-1+Eb-Wnt1-1*) lends support to a lineage-specific duplication, but support for this topology is middling (75%, Figure 4), and it is noteworthy that the two *Ce. sculpturatus* sequences are recovered from different genomic scaffolds (see Supplementary Table 2 for accession numbers).

The presence of *Wnt10* in both amblypygids is also intriguing because it is absent from all other arachnids surveyed so far. These sequences were placed within an arthropod *Wnt10* clade with high bootstrap support (100%; Figure 4), according to our phylogenetic analysis. Whether this indicates multiple losses of *Wnt10* in all other arachnid lineages, the recovery of a lost *Wnt10* in amblypygids, or the co-option of another gene, is unclear. Insights from other groups, such as harvestmen, will shed further light on this in future.

In contrast to the widespread retention of Hox duplicates, these new data indicate that the retention of duplicate Wnt genes is less common and restricted to certain subfamilies. Our understanding of specific Wnt functions among arthropods is more limited than that of Hox genes, but *Wnt* expression patterns in *P. tepidariorum* are available for a tentative comparison. For example, the two *Wnt7* paralogs show clear functional divergence in *P. tepidariorum*, with *Wnt7-1* expressed in the segment addition zone (SAZ) and *Wnt7-2* at the base of the appendages with some signal in the head lobes (Janssen et al. 2010). Whether this separation is consistent in all arachnopulmonates remains to be determined, but levels of sequence divergence between newly identified ohnologs are similar to that of *P. tepidariorum* (62%).

Conversely, previous attempts to characterise the expression patterns of *Wnt11* paralogs in *P. tepidariorum* only detected expression of *Wnt11-2* (Janssen et al. 2010). Given the retention of *Wnt11-1* in both spiders and scorpions, and the considerable divergence between paralogous sequences, *Wnt11* could be a good candidate for sub- or neofunctionalization, but the role of *Wnt11-1* remains unknown. Similarly, *Wnt4* expression in *P. tepidariorum* is restricted to a few cells at the posterior edge of the germ band, towards the end of embryogenesis (Janssen et al. 2010). These authors noted that this was in stark contrast to *Platynereis dumerilii,* where *Wnt4* is expressed in segments, the ventral midline and the SAZ (*Wnt4* is absent in insects, preventing a closer phylogenetic comparison). If *Wnt4* has an ancestrally complex role, as suggested by *Pl. dumerilii,* the very restricted expression of *Wnt4* in *P. tepidariorum* could be the result of subfunctionalisation followed by loss of one paralog, which is apparently retained in RTA-clade spiders and amblypygids (Figures 3-4). The expression patterns of *Wnt4-2* in these groups will help to clarify this in future. Alternatively, as insects have lost *Wnt4* entirely, there may simply be reduced *Wnt4* functionality across arthropods. However, this hypothesis stands at odds with not only the retention of both ohnologs in two large clades, but their detectable expression during development as evidenced by RNA-Seq.

The discovery of duplicate *Wnt1/wg* is particularly exciting: duplicates of this Wnt gene have not yet been detected in any other metazoans, even following multiple rounds of WGD in vertebrates and teleosts (see https://web.stanford.edu/group/nusselab/cgi-bin/wnt/vertebrate). Horseshoe crabs, which have also undergone multiple WGD and retain multiple Hox clusters (Kenny et al. 2016; Nong et al. 2020; Shingate et al. 2020), have not yet been systematically surveyed for Wnt genes. Embryonic

transcriptomes so far have only recovered one copy of *Wnt1* (Chen et al. 2016), but this should be a priority for future study. In arthropods *Wnt1/wg* performs a wide variety of roles, including in segment polarisation and in appendage and nervous system development (Murat et al. 2010) and has an accordingly complex expression pattern in *P. tepidariorum*, appearing in the L1 and L2 segments, limb buds, and dorsal O2 and O3 segments (Janssen et al. 2010). In theory, therefore, there is ample potential for subfunctionalisation. Functional analysis of *Wnt1/wg* duplicates in amblypygids and scorpions will no doubt prove extremely interesting in the future.

Frizzled *duplicates are retained in amblypygids, but not in spiders*

The transcriptomes of the spiders *M. muscosa* and *Pa. amentata* contained orthologs of *fz1, fz2* and *fz4*, but *fz3* was absent (Figures 5-6). The same subfamilies are represented in the *P. tepidariorum* genome, but a single copy of *fz3* was identified in *Ph. phalangioides* (Janssen et al. 2015); thus, entelegyne spiders may universally lack *fz3* but it was likely present in the ancestor of all spiders. Analysis of the *M. muscosa* transcriptome also returned a second copy of *fz2*, which is not shared by any other arachnid to date. These two paralogs form a well-supported clade (95%, Figure 6), indicating that this is the result of a lineage-specific tandem duplication followed by rapid sequence divergence in *fz2-2* (sequence similarity 53%, Supplementary Data File 4). Although two copies of *fz4* were identified in *P. tepidariorum* (Janssen et al. 2015), we only detected single copies in transcriptomes from *M. muscosa* and *Pa. amentata*, and only one was recovered from an embryonic transcriptome of *Ph. phalangioides* (Janssen et al. 2015). However, Janssen et al. (2015) demonstrated that the expression of the two *fz4* paralogs in *P. tepidariorum* is separated temporally. Therefore, we might not expect to detect both transcripts in embryos of similar stages, and an additional copy may be present in the genome.

Both amblypygid species have a large repertoire of *frizzled* genes compared to other arachnids, expressing all four orthology groups with two copies each of *fz1, fz3* and *fz4* (Figures 5-6). Duplicates of *fz1* and *fz3* appear to be unique to amblypygids. The *fz1* duplicates could be ohnologs retained from the arachnopulmonate WGD, as they form separate clades with the *fz1* genes of other arachnopulmonates and exhibit reasonable sequence divergence (support values ≥98%, paralog sequence similarity 76-77%; Figure 6). The origin of the *fz3* duplication is less clear; although the four amblypygid sequences form two separate clades, they are separated only by the placement of *Cs-fz3*,

which has low support (35%; Figure 6). Therefore we cannot yet confirm the timing of the duplication. Two copies of *fz4* are also found in the genomes of *P. tepidariorum* and the scorpions *Me. martensii* (Janssen et al. 2015) and *Ce. sculpturatus*. Sequences from *P. tepidariorum* and *Ce. sculpturatus* resolve in separate clades with other arachnopulmonate sequences (*fz4-1*; support value 62%; Figure 6) or together with *S. maritima* (*fz4-2*, support value 99%; Figure 6). They also demonstrate considerable sequence divergence (44% similarity in *P. tepidariorum* and 57% in *Ce. sculpturatus*; Supplementary Data File 4). We propose that these duplicates are probably retained from the ancestral WGD. The position of duplicates in the amblypygids, however, is less well resolved. Paralogs do not form within-species clades (Figure 6) and have fairly low sequence similarity (56-59%, Supplementary Data File 4), nor do all four sequences form an amblypygid clade. The four *fz4* genes were all placed within the same clade as *Pt-fz4-1* and *Cs-fz4-1*, and pairs of orthologs diverge sequentially (Figure 6). This conflicts with an origin in WGD. Ortholog pairs (e.g. *Ca-fz4-1* and *Eb-fz4-1*) returned 100% bootstrap support, but deeper relationships were more ambiguous (support values ≥ 51%). As a result, we cannot conclusively identify their origin but we hypothesise that these *fz4* duplicates reflect a lineage-specific duplication in either the ancestor of amblypygids or that of Pedipalpi (the larger clade to which amblypygids belong).

Previous studies of spiders, scorpions, and ticks indicated that *frizzled* repertoires in these groups are restricted to three or four copies, often with incomplete representation of the four orthology groups. Analysis of the new transcriptomes for the spiders *M. muscosa* and *Pa. amentata* is consistent with this pattern, albeit with a unique duplication of *fz2* in the jumping spider. We also recovered a single copy of *fz2* in *Ce. sculpturatus*, which was missing from previous work on *Me. martensii* (Janssen et al. 2015). The absence of *fz2* in the latter could result from a lineage-specific loss or an issue with genome assembly. In contrast, all four *frizzled* subfamilies were recovered in both amblypygid species, with three of these present in duplicate. Based on our data, it appears that the *frizzled* repertoire of amblypygids is around twice the size of all other arachnids and may have followed a very different evolutionary trajectory to spiders and scorpions following WGD. The expanded repertoire of frizzled genes in amblypygids is intriguing since they have also retained most Wnts and indeed several Wnt subfamilies are duplicated, and therefore it is possible that some specialised ligand-receptor interactions have evolved compared to other arthropods (Wu and Nusse 2002). However, although frizzled genes encode key receptors for Wnt ligands, they have other Wnt-independent functions, so

the expansion of the frizzled gene repertoire could be related to the evolution of alternative signalling roles (Janssen et al. 2015; Yu et al. 2020).

Conclusions: arachnopulmonate genome evolution in the wake of WGD

Our new transcriptome data and phylogenetic analyses provide the most comprehensive survey of Hox, Wnt and frizzled genes in arachnids to date, and substantially improve the density and breadth of taxonomic sampling for key developmental genes in Arachnopulmonata. We have identified intraorder variation at the level of major clades in spiders, which could help us better understand their morphological evolution. In new data for a third arachnopulmonate lineage, the amblypygids, we find additional evidence supporting an ancestral WGD and are better able to reconstruct the chronology of gene duplications and losses in spiders and scorpions. These transcriptomic resources are among the very first available for amblypygids and will aid future investigations of this fascinating group. We also find evidence of consistent evolutionary trajectories in Hox and Wnt gene repertoires across three of the five arachnopulmonate orders, with inter-order variation in the retention of specific paralogs.

By improving taxonomic coverage within the spider lineage we are better able to polarise some loss/duplication events and identify potential new trends within the spiders, particularly illustrating separations between synspermiatan and entelegyne spiders, and between the derived RTA clade and other spiders. Despite being unable to ultimately conclude that some missing transcripts reflect genuine genomic losses, it appears that the evolution of these developmental genes in spiders is more complicated than we thought. It may be that these gene repertoires are genuinely more variable within spiders than they are in amblypygids or scorpions; spiders are by far the most taxonomically diverse arachnopulmonate order, and the apparent diversity of repertoires may simply reflect this. Conversely, the higher apparent intraorder diversity of gene repertoires may be an artefact of increased sampling in spiders (up to four or five species for specific gene families) compared to the one or two available resources for scorpions and amblypygids; we may detect more diversity within these groups with increased sampling. Nonetheless, we see two notable trends within spiders, outlined below.

First, we see several characters that appear to unite the RTA clade, which contains almost half of all extant spider species (World Spider Catalog 2019), having diversified rapidly following its divergence from the orb weavers (Garrison et al. 2016; Fernández et al. 2018; Shao and Li 2018). *M. muscosa* and *Pa. amentata* both exhibit the apparent loss of *Hox3* and *fz4* paralogs and the retention of a *Wnt4*

duplicate, in contrast to *P. tepidariorum* and *Ph. phalangioides*. Although frizzleds and Wnts have not been surveyed in *Cu. salei*, also a member of the RTA clade, previous studies of Hox genes have so far only recovered a single copy of *Hox3* (Schwager et al. 2007). The identification of genetic trends potentially uniting this group is exciting, even if the macroevolutionary implications are unclear: as described above, the possible functions of a *Wnt4* paralog are elusive in the context of very specific *Wnt4* expression in *P. tepidariorum*. Members of the RTA clade are very derived compared to other araneomorph spiders, both morphologically (e.g. male pedipalp morphology and sophisticated eyes) and ecologically (most are wandering hunters), and their rapid diversification would align with clade-specific genetic divergence (Garrison et al. 2016; Fernández et al. 2018; Shao and Li 2018).

Second, although data are only available for a single representative of the plesiomorphic clade Synspermiata, *Ph. phalangioides*, these suggest lineage-specific losses of *lab* and *Ubx* paralogs and the only example of *fz3* found in spiders so far. The presence of *fz3* is consistent with other arachnopulmonate groups and suggests that it was present in the spider ancestor and only lost in the more derived entelegyne lineages (as seen in *P. tepidariorum, M. muscosa* and *Pa. amentata*). If *lab* and *Ubx* duplicates are indeed absent from the genome of *Ph. phalangioides*, this unusual loss of Hox genes represents an interesting divergence between these two major groups. Synspermiatan spiders are separated from other spiders by their relatively simpler genitalia and the absence of a cribellum, which was putatively present in ancestral spiders but lost in Synspermiata (Michalik and Ramírez 2014). Although they are unlikely to be directly responsible, the divergence in gene repertoires we see between *Ph. phalangioides* and the other spider lineages might provide a starting point for understanding these important morphological differences.

The amblypygids emerge as a key group of interest for studying the impacts of WGD owing to their high levels of ohnolog retention. Our transcriptomes, from representatives of two major clades, provide new evidence supporting a WGD in the ancestor of arachnopulmonates and demonstrating widespread retention of ohnologs in three major families of developmental genes (consistent with the retention of many duplicated regulators of eye development in other species; Gainett et al. 2020). In all three gene families we studied, repertoires were largest in the amblypygid species. This was particularly the case in *C. acosta*, which belongs to the less speciose and more plesiomorphic infraorder Charinidae within living Amblypygi. Although this study represents just two amblypygid species and three gene families,

this appears to contradict widespread predictions of diversification with the duplication of important developmental genes such as Hox (e.g. Van De Peer et al. 2009). Of particular interest are the amblypygid Wnt gene repertoires. We have identified from their transcriptomes, and from the published genome of *Ce. sculpturatus*, the first reported duplicates of *Wnt1/wg* in any animal, as well as the first reported *Wnt10* in any arachnid. Future functional studies of these genes and their expression during development will be critical to understanding the evolutionary impacts of these unusual components of amblypygid gene repertoires. Amblypygids also represent a potential model group for studying the evolution of arthropod body plans, owing to the unusual and derived morphology of the pedipalps and especially the first walking legs. Thanks to a substantial existing body of work on anterior-posterior patterning, segmentation and appendage development in spiders and other arachnids, we may have a chance to crack the genetic underpinnings of these dramatic evolutionary innovations (Pechmann et al. 2009; Sharma et al. 2012, 2014; Turetzek et al. 2016, 2017; Schwager et al. 2017; Baudouin-Gonzalez et al. 2020; Schomburg et al. 2020).

Finally, our analysis of existing genomic data for *Ce. sculpturatus* has recovered several Wnt and Frizzled gene duplications, similar to spiders and amblypygids. However, in contrast to those groups, our phylogenies have sometimes supported within-lineage duplication in *Ce. sculpturatus*, as opposed to the retention of ohnologs following WGD, even when these are observed in spiders and amblypygids. This was the case for *Wnt1/wg, Wnt6, Wnt7,* and potentially *fz2* (Figures 4,6). However, levels of sequence similarity in these cases were comparable for *Ce. sculpturatus* paralogs and amblypygid and spider ohnologs, when we might expect within-lineage duplicates to show higher similarity. The resolution of the paralogous sequences in our phylogenetic analyses could be confounded by the early-branching position of scorpions within Arachnopulmonata, which means paralogs would be expected to appear towards the bottom of ortholog clades and are more vulnerable to movement.

Overall, our new data provide further evidence of an ancestral arachnopulmonate WGD, identify evolutionary patterns within gene families following WGD, reveal new diversity in spider gene repertoires, better contextualise existing data from spiders and scorpions, and broaden the phylogenetic scope of available data for future researchers. However, other arachnid groups, both with and without ancestral WGD, require further study. Thelyphonids (vinegaroons or whip scorpions) and schizomids form a clade with amblypygids (Pedipalpi) and should also have been subject to the arachnopulmonate

WGD. Future work on these groups will shed light on the unusual patterns of gene retention we find in both major clades of amblypygids. Pseudoscorpions especially require urgent attention; due to their uncertain phylogenetic placement and lack of genomic data, we don't currently know whether they have also been subject to the arachnopulmonate WGD. Finally, to better contextualise the genomic changes that occur following the arachnopulmonate WGD, we require further data on arachnids without WGD, namely harvestmen and ticks. The ability to compare rates of sequence divergence, within-lineage gene duplication, and, eventually, functional properties of developmental genes in these groups will provide critical comparative data for arachnopulmonates.

527 **Acknowledgements**

528 This work was supported by the John Fell Fund, University of Oxford (award 0005632 to LSR), a NERC
529 grant (NE/T006854/1) to LSR and APM, and a Leverhulme Trust grant (RPG-2016-234) to APM. AH
530 was funded by a BBSRC DTP studentship. The authors are very grateful to Philip Steinhoff and Gabriele
531 Uhl (University of Greifswald) for providing embryos of *M. muscosa* and to Simon Ellis (University of
532 Oxford) for IT support.

533

References

Abzhanov, A., A. Popadic, and T. C. Kaufman. 1999. Chelicerate Hox genes and the homology of arthropod segments. Evol. Dev. 1:77–89.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data.

Baudouin-Gonzalez, L., A. Schoenauer, A. Harper, G. Blakeley, M. Seiter, S. Arif, L. Sumner-Rooney, S. Russell, P. P. Sharma, and P. Alistair. 2020. The evolution of Sox gene repertoires and regulation of segmentation in arachnids. BioRXiv, doi: https://doi.org/10.1101/2020.06.04.133389.

Beermann, A., R. Pruhs, R. Lutz, and R. Schroder. 2011. A context-dependent combination of wnt receptors controls axis elongation and leg development in a short germ insect. Development 138:2793–805.

Bolognesi, R., A. Beermann, L. Farzana, N. Wittkopp, R. Lutz, G. Balavoine, S. J. Brown, and R. Schröder. 2008. *Tribolium* Wnts: Evidence for a larger repertoire in insects with overlapping expression patterns that suggest multiple redundant functions in embryogenesis. Dev. Genes Evol. 218:193–202.

Brown, S. J., J. P. Fellers, T. D. Shippy, E. A. Richardson, M. Maxwell, J. J. Stuart, and R. E. Denell. 2002. Sequence of the *Tribolium castaneum* homeotic complex: The region corresponding to the Drosophila melanogaster Antennapedia complex. Genetics 160:1067–1074.

Cadigan, K. M., and R. Nusse. 1997. Wnt signaling: A common theme in animal development. Genes Dev. 11:3286–3305.

Cao, Z., Y. Yu, Y. Wu, P. Hao, Z. Di, Y. He, Z. Chen, W. Yang, Z. Shen, X. He, J. Sheng, X. Xu, B. Pan, J. Feng, X. Yang, W. Hong, W. Zhao, Z. Li, K. Huang, T. Li, Y. Kong, H. Liu, D. Jiang, B. Zhang, J. Hu, Y. Hu, B. Wang, J. Dai, B. Yuan, Y. Feng, W. Huang, X. Xing, G. Zhao, X. Li, Y. Li, and W. Li. 2013. The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. Nat. Commun. 4. Nature Publishing Group.

Carroll, S., J. Grenier, and S. Weatherbee. 2005. From DNA to diversity. Blackwell Publishing Ltd, London.

Chen, M., C. Wang, W. Wang, G. Ji, B. Hu, M. Du, G. Liu, Z. Li, W. Wang, X. Lin, W. Zheng, and J. Chen. 2016. De novo assembly and characterization of early embryonic transcriptome of the horseshoe crab *Tachypleus tridentatus*. PLoS One 11:1–18.

Chipman, A. D., D. E. K. Ferrier, C. Brena, J. Qu, D. S. T. Hughes, R. Schröder, M. Torres-Oliva, N. Znassi, H. Jiang, F. C. Almeida, C. R. Alonso, Z. Apostolou, P. Aqrawi, W. Arthur, J. C. J. Barna, K. P. Blankenburg, D. Brites, S. Capella-Gutiérrez, M. Coyle, P. K. Dearden, L. Du Pasquier, E. J. Duncan, D. Ebert, C. Eibner, G. Erikson, P. D. Evans, C. G. Extavour, L. Francisco, T. Gabaldón, W. J. Gillis, E. A. Goodwin-Horn, J. E. Green, S. Griffiths-Jones, C. J. P. Grimmelikhuijzen, S. Gubbala, R. Guigó, Y. Han, F. Hauser, P. Havlak, L. Hayden, S. Helbing, M. Holder, J. H. L. Hui, J. P. Hunn, V. S. Hunnekuhl, L. R. Jackson, M. Javaid, S. N. Jhangiani, F. M. Jiggins, T. E. Jones, T. S. Kaiser, D. Kalra, N. J. Kenny, V. Korchina, C. L. Kovar, F. B. Kraus, F. Lapraz, S. L. Lee, J. Lv, C. Mandapat, G. Manning, M. Mariotti, R. Mata, T. Mathew, T. Neumann, I. Newsham, D. N. Ngo, M. Ninova, G. Okwuonu, F. Ongeri, W. J. Palmer, S. Patil, P. Patraquim, C. Pham, L. L. Pu, N. H. Putman, C. Rabouille, O. M. Ramos, A. C. Rhodes, H. E. Robertson, H. M. Robertson, M. Ronshaugen, J. Rozas, N. Saada, A. Sánchez-Gracia, S. E. Scherer, A. M. Schurko, K. W. Siggens, D. N. Simmons, A. Stief, E. Stolle, M. J. Telford, K. Tessmar-Raible, R. Thornton, M. van der Zee, A. von Haeseler, J. M. Williams, J. H. Willis, Y. Wu, X. Zou, D. Lawson, D. M. Muzny, K. C. Worley, R. A. Gibbs, M. Akam, and S. Richards. 2014. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede *Strigamia maritima*. PLoS Biol. 12.

Cho, S. J., Y. Vallès, V. C. Giani, E. C. Seaver, and D. A. Weisblat. 2010. Evolutionary dynamics of the *wnt* gene family: A lophotrochozoan perspective. Mol. Biol. Evol. 27:1645–1658.

583 Clarke, T. H., J. E. Garb, C. Y. Hayashi, P. Arensburger, and N. A. Ayoub. 2015. Spider
584 Transcriptomes Identify Ancient Large-Scale Gene Duplication Event Potentially Important in
585 Silk Gland Evolution. Genome Biol. Evol. 7:1856–1870.

586 Clarke, T. H., J. E. Garb, C. Y. Hayashi, R. A. Haney, A. K. Lancaster, S. Corbett, and N. A. Ayoub.
587 2014. Multi-tissue transcriptomics of the black widow spider reveals expansions, co-options, and
588 functional processes of the silk gland gene toolkit. BMC Genomics 15:1–17.

589 Dehal, P., and J. L. Boore. 2005. Two rounds of whole genome duplication in the ancestral vertebrate.
590 PLoS Biol. 3.

591 Di, Z., Y. Yu, Y. Wu, P. Hao, Y. He, H. Zhao, Y. Li, G. Zhao, X. Li, W. Li, and Z. Cao. 2015. Genome-
592 wide analysis of homeobox genes from *Mesobuthus martensii* reveals Hox gene duplication in
593 scorpions. Insect Biochem. Mol. Biol. 61:25–33. Elsevier Ltd.

594 Ding, X., J. Liu, L. Zheng, J. Song, N. Li, H. Hu, X. Tong, and F. Dai. 2019. Genome-wide
595 identification and expression profiling of Wnt family genes in the silkworm, Bombyx mori. Int. J.
596 Mol. Sci. 20:1–13.

597 Donoghue, P. C. J., and M. A. Purnell. 2005. Genome duplication, extinction and vertebrate evolution.
598 Trends Ecol. Evol. 20:312–319.

599 Erwin, D. H. 2009. Early origin of the bilaterian developmental toolkit. Philos. Trans. R. Soc. B Biol.
600 Sci. 364:2253–2261.

601 Ferguson, L., F. Marlétaz, J. M. Carter, W. R. Taylor, M. Gibbs, C. J. Breuker, and P. W. H. Holland.
602 2014. Ancient expansion of the Hox cluster in Lepidoptera generated four homeobox genes
603 implicated in extra-embryonic tissue formation. PLoS Genet. 10.

604 Fernández, R., R. J. Kallal, D. Dimitrov, J. A. Ballesteros, M. A. Arnedo, G. Giribet, and G. Hormiga.
605 2018. Phylogenomics, Diversification Dynamics, and Comparative Transcriptomics across the
606 Spider Tree of Life. Curr. Biol. 28:1489-1497.e5.

607 Flot, J. F., B. Hespeels, X. Li, B. Noel, I. Arkhipova, E. G. J. Danchin, A. Hejnol, B. Henrissat, R.
608 Koszul, J. M. Aury, V. Barbe, R. M. Barthélémy, J. Bast, G. A. Bazykin, O. Chabrol, A. Couloux,
609 M. Da Rocha, C. Da Silva, E. Gladyshev, P. Gouret, O. Hallatschek, B. Hecox-Lea, K. Labadie,
610 B. Lejeune, O. Piskurek, J. Poulain, F. Rodriguez, J. F. Ryan, O. A. Vakhrusheva, E. Wajnberg,
611 B. Wirth, I. Yushenova, M. Kellis, A. S. Kondrashov, D. B. M. Welch, P. Pontarotti, J.
612 Weissenbach, P. Wincker, O. Jaillon, and K. Van Doninck. 2013. Genomic evidence for ameiotic
613 evolution in the bdelloid rotifer *Adineta vaga*. Nature 500:453–457.

614 Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of
615 duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.

616 Gainett, G., J. A. Ballesteros, C. R. Kanzler, J. T. Zehms, J. M. Zern, S. Aharon, E. Gavish-Regev,
617 and P. P. Sharma. 2020. How spiders make their eyes: Systemic paralogy and function of retinal
618 determination network homologs in arachnids. BioRXiv 1–13.

619 Gainett, G., and P. P. Sharma. 2020. New genomic resources and toolkits for developmental study of
620 whip spiders (Amblypygi) provide insights into arachnid genome evolution and antenniform leg
621 patterning. BioRixv 1–39.

622 Garb, J. E., P. P. Sharma, and N. A. Ayoub. 2018. Recent progress and prospects for advancing
623 arachnid genomics. Curr. Opin. Insect Sci. 25:51–57. Elsevier Inc.

624 Garrison, N. L., J. Rodriguez, I. Agnarsson, J. A. Coddington, C. E. Griswold, C. A. Hamilton, M.
625 Hedin, K. M. Kocot, J. M. Ledford, and J. E. Bond. 2016. Spider phylogenomics: untangling the
626 Spider Tree of Life. PeerJ 4.

627 Glasauer, S. M. K., and S. C. F. Neuhauss. 2014. Whole-genome duplication in teleost fishes and its
628 evolutionary consequences. Mol. Genet. Genomics 289:1045–1060.

629 Goujon, M., H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern, and R. Lopez. 2010. A new
630 bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 38:695–699.

631 Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D.
632 Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R.
633 Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, and A. Regev.
634 2013. De novo transcript sequence reconstruction from RNA-Seq: reference generation and
635 analysis with Trinity.

636 Hamilton, F. S., G. N. Wheeler, and S. Hoppler. 2001. Difference in XTcf-3 dependency accounts for
637 change in response to β-catenin-mediated Wnt signalling in Xenopus blastula. Development
638 128:2063–2073.

639 Havlak, P., N. H. Putnam, J.-X. Yue, H. J. Brockmann, C. W. Nossa, K. Y. Vincent, and J. Lv. 2014.
640 Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient
641 whole genome duplication. Gigascience 3:1–21.

642 Hayden, L., and W. Arthur. 2013. Expression patterns of Wnt genes in the venom claws of
643 centipedes. Evol. Dev. 15:365–372.

644 Hogvall, M., A. Schönauer, G. E. Budd, A. P. McGregor, N. Posnien, and R. Janssen. 2014. Analysis
645 of the Wnt gene repertoire in an onychophoran provides new insights into the evolution of
646 segmentation. Evodevo 5:1–15.

647 Holland, L. Z. 2013a. Evolution of new characters after whole genome duplications: Insights from
648 amphioxus. Semin. Cell Dev. Biol. 24:101–109. Elsevier Ltd.

649 Holland, P. W. H. 2013b. Evolution of homeobox genes. Wiley Interdiscip. Rev. Dev. Biol. 2:31–45.

650 Holland, P. W. H., J. Garcia-Fernandez, N. A. Williams, and A. Sidow. 1994. Gene duplications and
651 the origins of vertebrate development. Development 120:125–133.

652 Holzem, M., N. Braak, O. Brattström, A. P. McGregor, and C. J. Breuker. 2019. Wnt gene expression
653 during early embryogenesis in the nymphalid butterfly *Bicyclus anynana*. Front. Ecol. Evol. 7.

654 Hueber, S. D., and I. Lohmann. 2008. Shaping segments: Hox gene function in the genomic age.
655 BioEssays 30:965–979.

656 Janssen, R., M. Le Gouar, M. Pechmann, F. Poulin, R. Bolognesi, E. E. Schwager, C. Hopfen, J. K.
657 Colbourne, G. E. Budd, S. J. Brown, N. M. Prpic, C. Kosiol, M. Vervoort, W. G. Damen, G.
658 Balavoine, and A. P. McGregor. 2010. Conservation, loss, and redeployment of Wnt ligands in
659 protostomes: Implications for understanding the evolution of segment formation. BMC Evol. Biol.
660 10:1–21.

661 Janssen, R., and N. Posnien. 2014. Identification and embryonic expression of *Wnt2, Wnt4, Wnt5* and
662 *Wnt9* in the millipede *Glomeris marginata* (Myriapoda: Diplopoda). Gene Expr. Patterns 14:55–
663 61.

664 Janssen, R., A. Schönauer, M. Weber, N. Turetzek, M. Hogvall, G. E. Goss, N. H. Patel, A. P.
665 McGregor, and M. Hilbrant. 2015. The evolution and expression of panarthropod frizzled genes.
666 Front. Ecol. Evol. 3:1–14.

667 Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. Von Haeseler, and L. S. Jermiin. 2017.
668 ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14:587–
669 589.

670 Kenny, N. J., K. W. Chan, W. Nong, Z. Qu, I. Maeso, H. Y. Yip, T. F. Chan, H. S. Kwan, P. W. H. H.
671 Holland, K. H. Chu, and J. H. L. L. Hui. 2016. Ancestral whole-genome duplication in the marine
672 chelicerate horseshoe crabs. Heredity (Edinb). 116:190–199.

673 Král, J., M. Forman, T. Kořínková, A. C. Reyes Lerma, C. R. Haddad, J. Musilová, M. Řezáč, I. M.
674 Ávila Herrera, S. Thakur, A. S. Dippenaar-Schoeman, F. Marec, L. Horová, and P. Bureš. 2019.

675 Insights into the karyotype and genome evolution of haplogyne spiders indicate a polyploid
676 origin of lineage with holokinetic chromosomes. Sci. Rep.

677 Kuraku, S., and A. Meyer. 2009. The evolution and maintenance of Hox gene clusters in vertebrates
678 and the teleost-specific genome duplication. Int. J. Dev. Biol. 53:765–773.

679 Leite, D. J., L. Baudouin-Gonzalez, S. Iwasaki-Yokozawa, J. Lozano-Fernandez, N. Turetzek, Y.
680 Akiyama-Oda, N. M. Prpic, D. Pisani, H. Oda, P. P. Sharma, and A. P. McGregor. 2018.
681 Homeobox gene duplication and divergence in arachnids. Mol. Biol. Evol. 35:2240–2253.

682 Leite, D. J., M. Ninova, M. Hilbrant, S. Arif, S. Griffiths-Jones, M. Ronshaugen, and A. P. McGregor.
683 2016. Pervasive microRNA duplication in chelicerates: Insights from the embryonic microRNA
684 repertoire of the spider *Parasteatoda tepidariorum*. Genome Biol. Evol. 8:2133–2144.

685 Logan, C. Y., and R. Nusse. 2004. The Wnt signaling pathway in development and disease. Annu.
686 Rev. Cell Dev. Biol. 20:781–810.

687 Lynch, M., and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes.
688 Science (80-.). 290:1151–1155.

689 Mark Welch, D. B., J. L. Mark Welch, and M. Meselson. 2008. Evidence for degenerate tetraploidy in
690 bdelloid rotifers. Proc. Natl. Acad. Sci. U. S. A. 105:5145–5149.

691 McGinnis, W., and R. Krumlauf. 1992. Homeobox genes and axial patterning. Cell 68:283–302.

692 Meng, F., I. Braasch, J. B. Phillips, X. Lin, T. Titus, C. Zhang, and J. H. Postlethwait. 2013. Evolution
693 of the eye transcriptome under constant darkness in *Sinocyclocheilus* cavefish. Mol. Biol. Evol.
694 30:1527–1543.

695 Meyer, A., and M. Schartl. 1999. Gene and genome duplications in vertebrates: The one-to-four (-to-
696 eight in fish) rule and the evolution of novel gene functions. Curr. Opin. Cell Biol. 11:699–704.

697 Michalik, P., and M. J. Ramírez. 2014. Evolutionary morphology of the male reproductive system,
698 spermatozoa and seminal fluid of spiders (Araneae, Arachnida) - Current knowledge and future
699 directions. Arthropod Struct. Dev. 43:291–322.

700 Miller, J. R. 2001. Protein family review The Wnts. Gene 1–15.

701 Murat, S., C. Hopfen, and A. P. McGregor. 2010. The function and evolution of Wnt genes in
702 arthropods. Arthropod Struct. Dev. 39:446–452. Elsevier Ltd.

703 Nguyen, L. T., H. A. Schmidt, A. Von Haeseler, and B. Q. Minh. 2015. IQ-TREE: A fast and effective
704 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–
705 274.

706 Nong, W., Z. Qu, Y. Li, T. Barton-Owe, A. Y. P. Wong, H. Y. Yip, H. T. Lee, S. Narayana, T. Baril, T.
707 Swale, J. Cao, T. F. Chan, H. S. Kwan, N. S. Ming, G. Panagiotou, P.-Y. Qian, J.-W. Qui, K. Y.
708 Yip, N. Ismail, S. Pati, A. John, S. S. Tobe, W. G. Bendena, S. G. Cheung, A. Hayward, and J.
709 H. L. Hui. 2020. Horseshoe crab genomes reveal the evolutionary fates of genes and
710 microRNAs after three rounds (3R) of whole genome duplication. BioRXiv 1–29.

711 Ohno, S. 1970. Evolution by Gene Duplication. Springer, New York.

712 Pascual-Anaya, J., S. D'Aniello, S. Kuratani, and J. Garcia-Fernàndez. 2013. Evolution of Hox gene
713 clusters in deuterostomes Hox content in invertebrate deuterostomes.

714 Pearson, J. C., D. Lemons, and W. McGinnis. 2005. Modulating Hox gene functions during animal
715 body patterning. Nat. Rev. Genet. 6:893–904.

716 Pechmann, M., A. P. McGregor, E. E. Schwager, N. M. Feitosa, and W. G. M. Damen. 2009. Dynamic
717 gene expression is required for anterior regionalization in a spider. Proc. Natl. Acad. Sci. U. S.
718 A. 106:1468–1472.

719 Pechmann, M., E. E. Schwager, N. Turetzek, and N. M. Prpic. 2015. Regressive evolution of the
720 arthropod tritocerebral segment linked to functional divergence of the Hox gene labial. Proc. R.
721 Soc. B Biol. Sci. 282.

722 Pires-daSilva, A., and R. J. Sommer. 2003. The evolution of signalling pathways in animal
723 development. Nat. Rev. Genet. 4:39–49.

724 Schomburg, C., N. Turetzek, and N. M. Prpic. 2020. Candidate gene screen for potential interaction
725 partners and regulatory targets of the Hox gene *labial* in the spider *Parasteatoda tepidariorum*.
726 Dev. Genes Evol. 230:105–120. Development Genes and Evolution.

727 Schwager, E. E., A. Schönauer, D. J. Leite, P. P. Sharma, and A. P. McGregor. 2015. Chelicerata.
728 Pp. 99–139 *in* A. Wanninger, ed. Evolutionary Developmental Biology of Invertebrates 3:
729 Ecdysozoa I: Non-Tetraconata. Springer-Verlag, Wien.

730 Schwager, E. E., M. Schoppmeier, M. Pechmann, and W. G. M. Damen. 2007. Duplicated Hox genes
731 in the spider *Cupiennius salei*. Front. Zool. 4.

732 Schwager, E. E., P. P. Sharma, T. Clarke, D. J. Leite, T. Wierschin, A. D. Buffry, H. Chao, H. Dinh, H.
733 Doddapaneni, S. Dugan, S. L. Lee, I. Maeso, S. C. Murali, D. M. Muzny, R. Nunes, C. Wolff, K.
734 C. Worley, G. Bucher, R. A. Gibbs, J. Coddington, H. Oda, S. Richards, and A. P. Mcgregor.
735 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid
736 evolution. BMC Biol. 15:1–27. BMC Biology.

737 Seppey, M., M. Manni, and E. M. Zdobnov. 2019. BUSCO: Assessing Genome Assembly and
738 Annotation Completeness. P. *in* M. Kollmar, ed. Gene Prediction. Methods in Molecular Biology,
739 Vol 1962. Humana, New York, NY.

740 Shao, L., and S. Li. 2018. Early Cretaceous greenhouse pumped higher taxa diversification in spiders.
741 Mol. Phylogenet. Evol. 127:146–155. Elsevier.

742 Sharma, P. P., M. A. Santiago, E. González-Santillán, L. Monod, and W. C. Wheeler. 2015. Evidence
743 of duplicated Hox genes in the most recent common ancestor of extant scorpions. Evol. Dev.
744 17:347–355.

745 Sharma, P. P., E. E. Schwager, C. G. Extavour, and G. Giribet. 2012. Hox gene expression in the
746 harvestman *Phalangium opilio* reveals divergent patterning of the chelicerate opisthosoma. Evol.
747 Dev. 14:450–463.

748 Sharma, P. P., E. E. Schwager, C. G. Extavour, and W. C. Wheeler. 2014. Hox gene duplications
749 correlate with posterior heteronomy in scorpions. Proc. R. Soc. B Biol. Sci. 281.

750 Shimeld, S. M. 1999. Gene function, gene networks and the fate of duplicated genes. Semin. Cell
751 Dev. Biol. 10:549–553.

752 Shingate, P., V. Ravi, A. Prasad, B. H. Tay, K. M. Garg, B. Chattopadhyay, L. M. Yap, F. E. Rheindt,
753 and B. Venkatesh. 2020. Chromosome-level assembly of the horseshoe crab genome provides
754 insights into its genome evolution. Nat. Commun. 11. Springer US.

755 Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert,
756 J. Söding, J. D. Thompson, and D. G. Higgins. 2011. Fast, scalable generation of high-quality
757 protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7.

758 Song, L., and L. Florea. 2015. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq
759 reads. Gigascience 4:1–8. GigaScience.

760 Stauber, M., H. Jäckle, and U. Schmidt-Ott. 1999. The anterior determinant bicoid of *Drosophila* is a
761 derived Hox class 3 gene. Proc. Natl. Acad. Sci. U. S. A. 96:3786–3789.

762 Turetzek, N., S. Khadjeh, C. Schomburg, and N. M. Prpic. 2017. Rapid diversification of homothorax
763 expression patterns after gene duplication in spiders. BMC Evol. Biol. 17:1–12. BMC
764 Evolutionary Biology.

765 Turetzek, N., M. Pechmann, C. Schomburg, J. Schneider, and N. M. Prpic. 2016. Neofunctionalization
766 of a duplicate dachshund gene underlies the evolution of a novel leg segment in arachnids. Mol.
767 Biol. Evol. 33:109–121.

768 van Amerongen, R., and R. Nusse. 2009. Towards an integrated view of Wnt signaling in
769 development. Development 136:3205–3214.

770 Van De Peer, Y., S. Maere, and A. Meyer. 2009. The evolutionary significance of ancient genome
771 duplications. Nat. Rev. Genet. 10:725–732. Nature Publishing Group.

772 Weygoldt, P. 2009. Evolutionary morphology of whip spiders: towards a phylogenetic system
773 (Chelicerata: Arachnida: Amblypygi)*. J. Zool. Syst. Evol. Res. 34:185–202.

774 Wu, C. H., and R. Nusse. 2002. Ligand receptor interactions in the Wnt signaling pathway in
775 *Drosophila*. J. Biol. Chem. 277:41762–41769.

776 Yu, J. J. ., A. Maugarny-Cales, S. Pelletier, C. Alexandre, Y. Bellaiche, J.-P. Vincent, and I. J.
777 McGough. 2020. Frizzled-dependent Planar Cell Polarity without Wnt Ligands. bioRxiv
778 2020.05.23.108977.

779 Zhang, J. 2003. Evolution by gene duplication: An update. Trends Ecol. Evol. 18:292–298.

780 Zhong, Y. F., and P. W. H. Holland. 2011. HomeoDB2: Functional expansion of a comparative
781 homeobox gene database for evolutionary developmental biology. Evol. Dev. 13:567–568.

782 2019. World Spider Catalog. Natural History Museum Bern.

783

Table 1. Assembly metrics for transcriptomes of *Charinus acosta, Euphrynichus bacillifer, Marpissa muscosa* and *Pardosa amentata*.

* Based on longest Isoform per gene

** Same as contig N50 but based on top most highly expressed genes that represent 90% of the total normalized expression data

*** The number of genes for which Ex90N50 is calculated

[1] 10 species, n=2934 BUSCOs; C=Complete [D=Duplicated], F=Fragmented, M=Missing.

[2] 90 species, n=1013 BUSCOs; C=Complete [D=Duplicated], F=Fragmented, M=Missing.

Species	Raw Reads	Processed Reads	#Trinity Genes	#Trinity Isoforms	Contig N50 (bases)*	Ex90N50 (bases)**	#Ex90N50 genes***	Arachnid BUSCO Scores (C[D],F,M)[1]	Arthropod BUSCO Scores (C[D],F,M)[2]
C. acosta	272,844,971	260,853,757	237,678	334,267	896	2406	31,012	94.9%[12.9%],1.0%,4.1%	93.2%[9.5%],1.5%,5.3%,
E. bacillifer	249,938,618	239,034,000	184,142	285,861	978	2671	22,647	93.8%[7.1%],1.5%,4.7%	92.9%[3.5%],0.7%,6.4%
Pa. amentata	266,764,548	256,911,378	316,021	542,344	652	1758	38,423	95.4%[5.3%],0.9%,3.7%	92.9%[4.5%],1.2%,5.9%
M. muscosa	222,479,664	211,848,357	276,943	473,878	592	1461	46,196	94.7%[6.1%],1.3%,4.0%	90.8%[3.9%],1.9%,7.3%

Figures

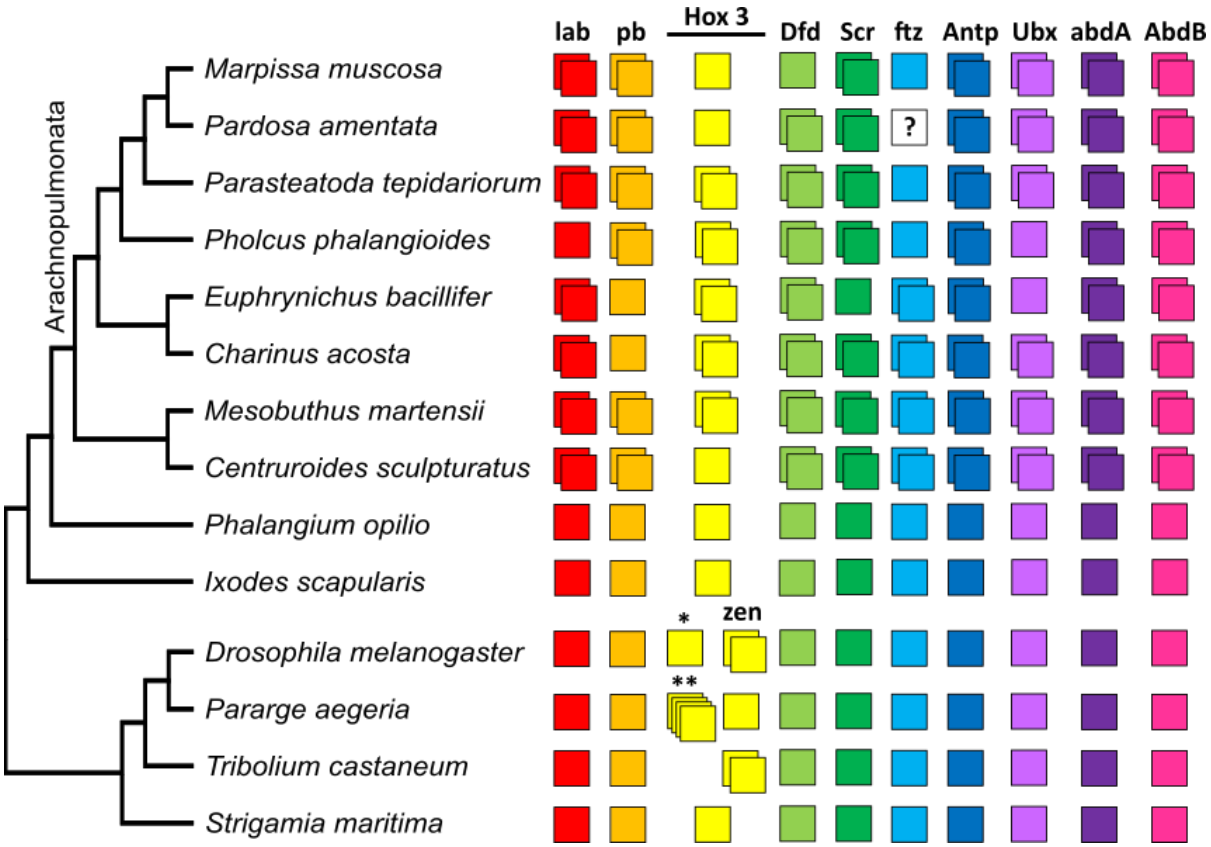


Figure 1. Repertoires of Hox genes in arachnids and other selected arthropods. Hox genes are represented by coloured boxes with duplicated Hox genes indicated by overlapping boxes. *ftz* was not found in the *Pa. amentata* transcriptome. Figure includes Hox repertoires previously surveyed in the arachnids *P. tepidariorum* (Schwager et al. 2017; Leite et al. 2018), *Ce. sculpturatus* (Schwager et al. 2017), *Me. martensii* (Di et al. 2015), *Pha. opilio* (Sharma et al. 2012), *I. scapularis* (all genomes) and *Ph. phalangioides* (embryonic transcriptome; Leite et al. 2018), the myriapod *S. maritima* (Chipman et al. 2014) and the insects *D. melanogaster, T. castaneum* (Zhong and Holland 2011) and *Par. aegeria* (Ferguson et al. 2014). The insect *Hox3* homolog *zen* has undergone independent tandem duplications in *T. castaneum* to yield *zen* and *zen2*; in cyclorrhaphan flies to yield *zen* and *bicoid* (*) (Stauber et al. 1999); and in the genus *Drosophila* to yield *zen2* (Brown et al. 2002). The nymphalid butterfly *P. aegeria* is representative of most species of ditrysian Lepidoptera which possess four distinct *Hox3* genes termed Special homeobox genes (*ShxA*, *ShxB*, *ShxC* and *ShxD*) (**) and the canonical *zen* gene (Ferguson et al. 2014).

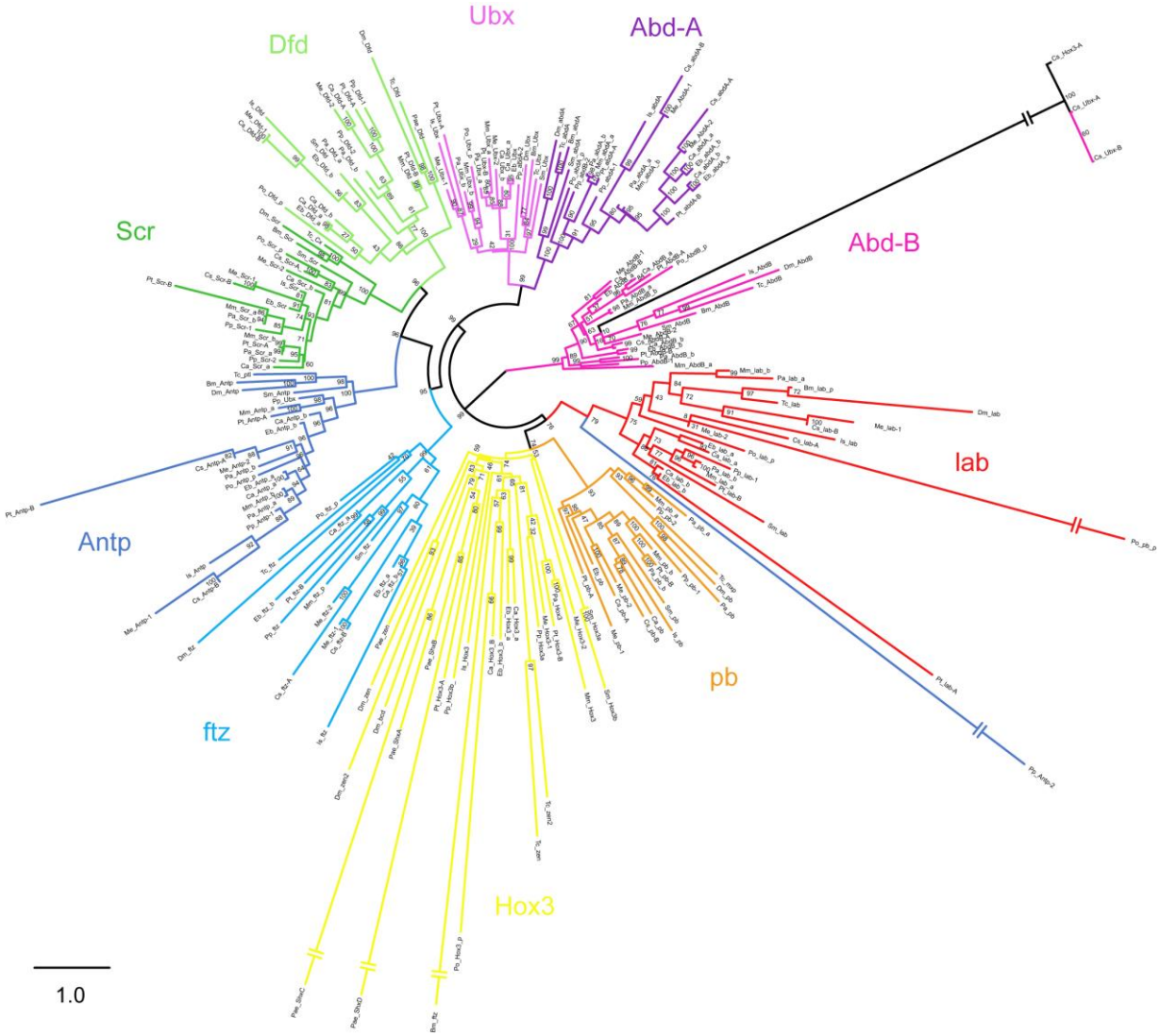


Figure 2 Maximum likelihood phylogeny of Hox amino acid sequences. The Hox genes are shown as different colours (after Figure 1). Panarthropods included: *M. muscosa* (Mm), *Pa. amentata* (Pa), *P. tepidariorum* (Pt), *Ph. phalangioides* (Pp), *E. bacillifer* (Eb), *C. acosta* (Ca), *Me. martensii* (Me), *Ce. sculpturatus* (Cs), *Pha. opilio* (Po), *I. scapularis* (Is), *D. melanogaster* (Dm), *Par. aegeria* (Pae), *T. castaneum* (Tc), and *S. maritima* (Sm). Node labels indicate ultrafast bootstrap support values. See Supplementary Table 1 for accession numbers, Supplementary Data File 1 for amino acid sequence alignments.

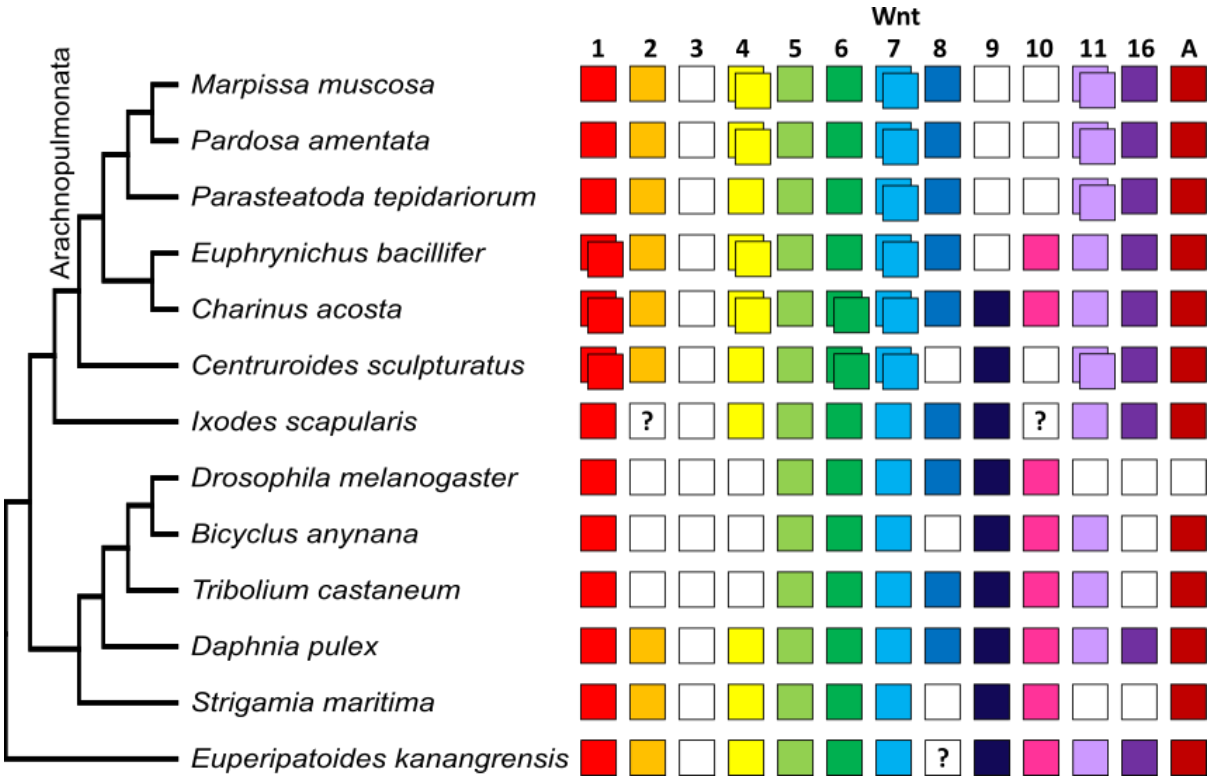


Figure 3 Repertoires of Wnt subfamilies in arthropods and an onychophoran. The Wnt subfamilies (1-11, 16 and A) are represented by coloured boxes with duplicated genes represented by overlapping boxes and putatively lost subfamilies indicated by white boxes. Question marks indicate subfamilies that have not been found but were probably not detected due to the lack of an available genomes or genome assembly quality. Figure includes Wnt repertoires recovered in this study and previously surveyed in the arachnids *P. tepidariorum* and *I. scapularis* (Janssen et al. 2010); the insects *D. melanogaster, T. castaneum* (Bolognesi et al. 2008) and *B. anynana* (Ding et al. 2019; Holzem et al. 2019); the crustacean *Da. pulex* (Janssen et al. 2010); the myriapod *S. maritima* (Chipman et al. 2014); and the onychophoran *Eu. kanangrensis* (Hogvall et al. 2014).

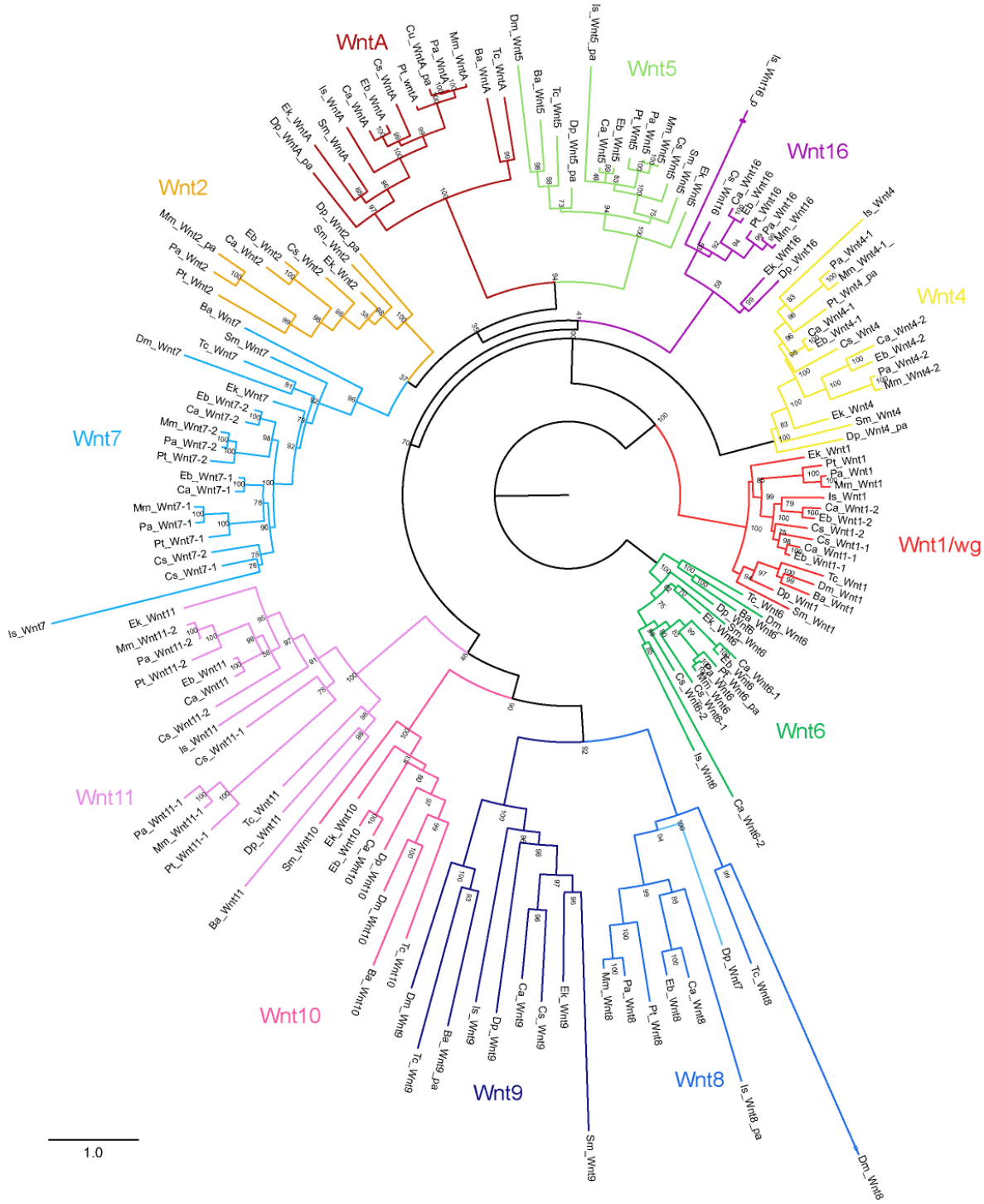


Figure 4. Maximum likelihood phylogeny of Wnt amino acid sequences. The 12 Wnt subfamilies are shown as different colours (after Figure 3). Panarthropods included: *P. tepidariorum* (Pt), *Cu. salei* (Cu), *Pa. amentata* (Pa), *M. muscosa* (Mm), *C. acosta* (Ca), *E. bacillifer* (Eb), *Ce. sculpturatus* (Cs), *I. scapularis* (Is), *D. melanogaster* (Dm), *B. anynana* (Ba), *T. castaneum* (Tc), *Da. pulex* (Dp), *S. maritima* (Sm), and *Eu. kanangrensis* (Ek). Node labels indicate ultrafast bootstrap support values. See Supplementary Table 2 for accession numbers, Supplementary Data File 2 for amino acid sequence alignments, and Supplementary Data File 3 for nucleotide sequence alignments of *Wnt1/wg* duplicates in *C. acosta, Ce. sculpturatus* and *E. bacillifer*.

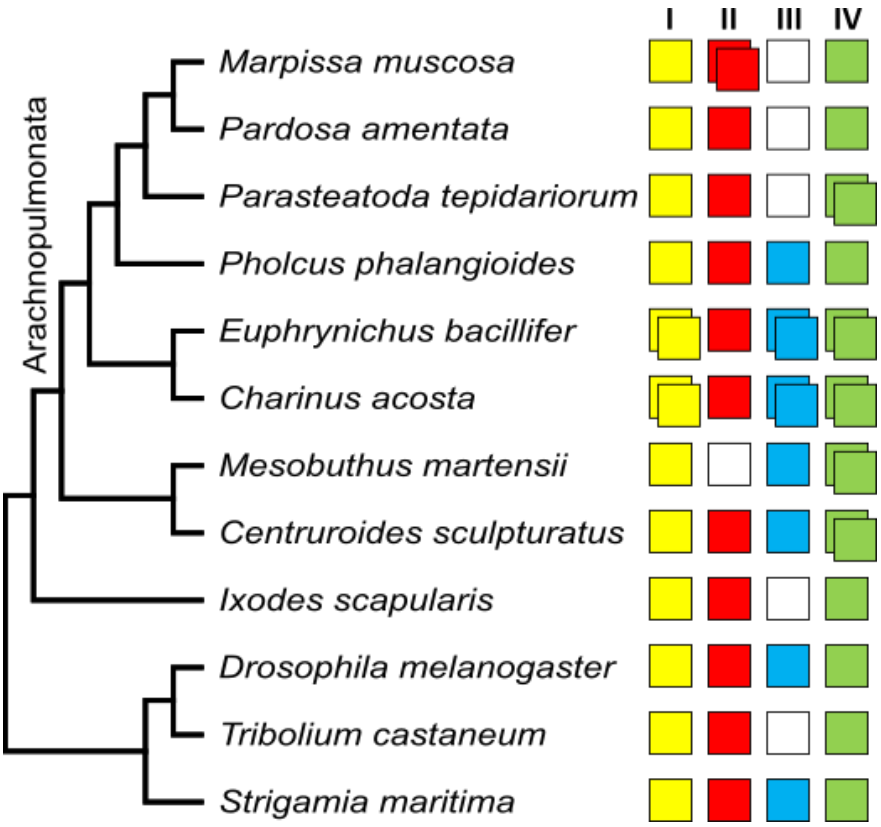


Figure 5. Repertoire of *frizzled* genes in arachnids and other selected arthropods. The four *frizzled* orthology groups (FzI, FzII, FzIII, and FzIV) are represented by coloured boxes, with duplicated genes represented by overlapping boxes and gene loss represented by a white box. Figure includes *frizzled* repertoires previously surveyed in the arachnids *P. tepidariorum*, *Ce. sculpturatus, Me. martensii*, *I. scapularis* (all genomes), and *Ph. phalangioides* (embryonic transcriptome; Janssen et al. 2015); the myriapod *S. maritima* (Janssen et al. 2015); and the insects *D. melanogaster* and *T. castaneum* (Beermann et al. 2011).

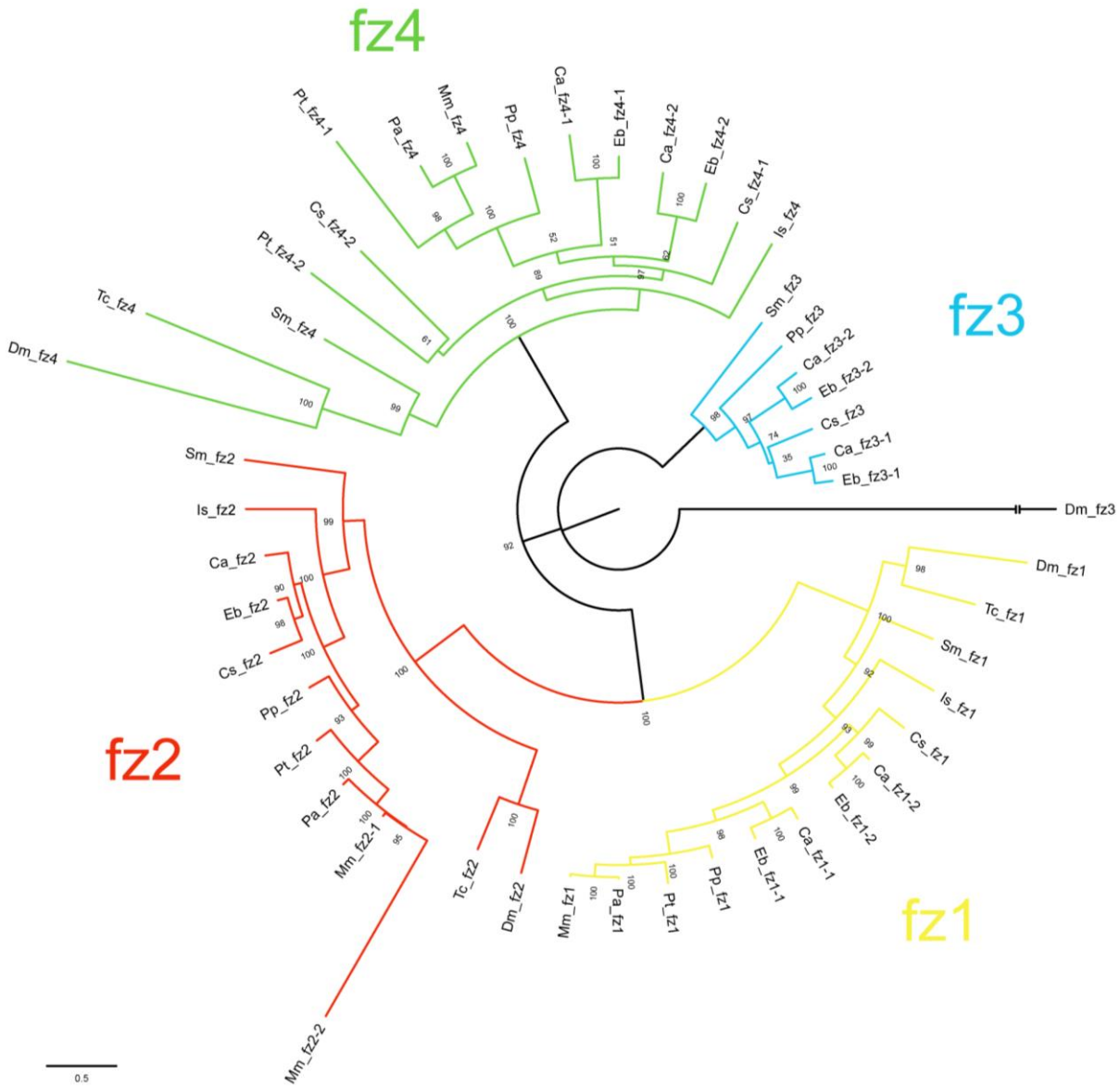


Figure 6 Maximum likelihood phylogeny of Frizzled proteins. The *frizzled* genes are shown as different colours (after Figure 5). Panarthropods included: *M. muscosa* (Mm), *Pa. amentata* (Pa), *P. tepidariorum* (Pt), *Ph. phalangioides* (Pp), *E. bacillifer* (Eb), *C. acosta* (Ca), *Me. martensii* (Me), *I. scapularis* (Is), *D. melanogaster* (Dm), *T. castaneum* (Tc), and *S. maritima* (Sm). Node labels indicate ultrafast bootstrap support values. See Supplementary Table 3 for accession numbers and Supplementary Data File 4 for alignments.

Supplementary Data and Tables

Supplementary Data File 1 Alignment of full Hox protein sequences, Phylip format.

Supplementary Data File 2 Alignment of full Wnt protein sequences, Phylip format.

Supplementary Data File 3 Alignment of *Wnt1* duplicate nucleotide sequences in *Charinus acosta, Euphrynichus bacillifer* and *Centruroides sculpturatus*, Phylip format.

Supplementary Data File 4 Alignment of full Frizzled protein sequences, Phylip format.

Supplementary Table 1 Protein accession numbers for Hox gene sequences used in this study.

Species	Hox genes	Protein accession number
Centruroides sculpturatus (Cs)		**SupFile1** (Leite et al. 2018) Predicted protein sequences were obtained using the Translate ExPASy online tool (https://web.expasy.org/translate/).
Mesobuthus martensii (Mm)		**Supplementary file 2: Classification of scorpion homeobox genes**. (Di et al. 2015)
Parasteatoda tepidariorum (Pt)		**SupFile1** (Leite et al. 2018) Predicted protein sequences were obtained using the Translate ExPASy online tool (https://web.expasy.org/translate/).
	Ubx-B	XP_021004342.1
	abd-A	XP_015921999.1
Pholcus phalangoides (Pp)		**SupFile1** (Leite et al. 2018) Predicted protein sequences were obtained using the Translate ExPASy online tool (https://web.expasy.org/translate/).
Ixodes scapularis (Is)		**SupFile1** (Leite et al. 2018)
Phalangium opilio (Po)	lab	CCH51000.1
	pb	CCH51001.1
	Hox3	CCH51002.1
	Dfd	CCH51003.1
	Scr	CCH51004.1
	ftz	CCH51005.1
	Antp	CCH51006.1
	Ubx	CCH51007.1
	abdA	CCH51008.1
	AbdB	CCH51009.1
Bombyx mori (Bm)	lab	BAC99310.1
	Scr	NP_001037339.1
	ftz	NP_001037528.2
	Antp	NP_001037319.1
	Ubx	NP_001107632.1
	abdA	NP_001166808.1
	AbdB	NP_001139700.1
Drosophila melanogaster (Dm)	pb	NP_996163.1
	lab	NP_001246953.1
	bcd	NP_731111.1
	zen	NP_476793.1
	zen2	NP_476794.1
	Dfd	NP_477201.1
	Scr	NP_001368995.1

	ftz	NP_477498.1
	Antp	NP_996175.1
	Ubx	NP_996219.1
	abdA	NP_001247145.1
	AbdB	NP_001303474.1
Pararge aegeria (Pae)	pb	AIB07898.1
	zen	AIB07903.1
	Dfd	AIB07904.1
	ShxA	AIB07900.1
	ShxB	AIB07899.1
	ShxC	AIB07901.1
	ShxD	AIB07902.1
Strigamia maritima (Sm)		**Table S30. Details of the manually annotated genes of *S. maritima*.** (Chipman et al. 2014)
Tribolium castaneum (Tc)	mxp/pb	NP_001107807.1
	lab	NP_001107762.1
	zen	NP_001036813.1
	zen2	NP_001038090.1
	Dfd	NP_001034510.1
	Cx/Scr	NP_001034523.1
	ftz	NP_001034539.1
	ptl/Antp	NP_001034505.1
	Ubx	NP_001034497.1
	abdA	NP_001034518.1
	AbdB	NP_001034519.1

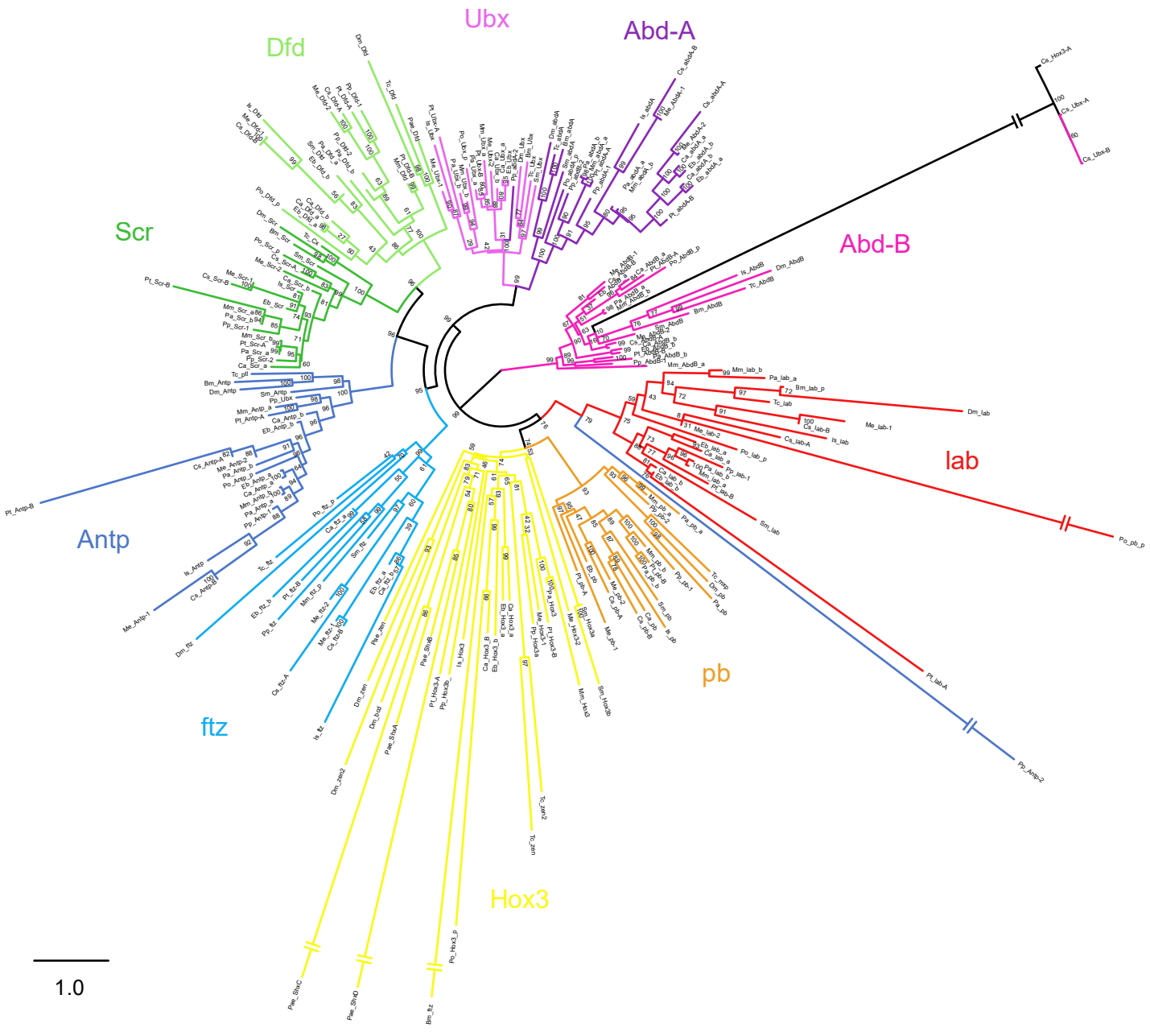
Supplementary Table 2. Protein accession numbers for all Wnt sequences used in this study. Note that the protein accession number could not be found for *Ixodes scapularis Wnt8*, the partial sequence from Janssen et al. (2010) was used in the Maximum likelihood tree of Wnt amino acid sequences (*).

Species	Wnt subfamily	Protein accession number
Centruroides sculpturatus (Cs)	Wnt1-1	XP_023219341.1
	Wnt1-2	XP_023228816.1
	Wnt2	XP_023224164.1
	Wnt4	XP_023222233.1
	Wnt5	XP_023215183.1
	Wnt6-1	XP_023219342.1
	Wnt6-2	XP_023228802.1
	Wnt7-1	XP_023215187.1
	Wnt7-2	XP_023220709.1
	Wnt9	XP_023228817.1
	Wnt11-1	XP_023240364.1
	Wnt11-2	XP_023228131.1
	Wnt16	XP_023224171.1
	WntA	XP_023218523.1
Cupiennius salei (Cu)	WntA (partial)	ADR79167.1
Parasteatoda tepidariorum (Pt)	Wnt1	XP_015906154.1
	Wnt2	NP_001310740.1
	Wnt4 (partial)	ADR79163.1
	Wnt5	NP_001310745.1
	Wnt6 (partial)	ADR79164.1
	Wnt7_1	NP_001310739.1
	Wnt7_2	NP_001310746.1
	Wnt8	ACH88002.1
	Wnt11_1	XP_015920223.1
	Wnt11_2	XP_015916686.1
	Wnt16	NP_001310769.1
Ixodes scapularis (Is)	Wnt1	XP_002407192.2
	Wnt4	XP_002436043.2
	Wnt5 (partial)	EEC11679.1
	Wnt6	EEC06108.1
	Wnt7	EEC17948.1
	Wnt8 (partial)	*
	Wnt9	EEC10449.1
	Wnt11	XP_002434188.1
	Wnt16 (partial)	EEC05882.1
	WntA	EEC02958.1
Bicyclus anynana (Ba)	Wnt1	XP_023955185.1
	Wnt5	XP_023937318.1
	Wnt6	XP_023955186.1

	Wnt7	XP_023935071.1
	Wnt9	XP_023953537.1
	Wnt10	XP_023955187.1
	Wnt11	XP_023934504.1
	WntA	XP_023937297.1
Daphnia pulex (Dp)	Wnt1	EFX86386.1
	Wnt2 (partial)	EFX87200.1
	Wnt4 (partial)	EFX72339.1
	Wnt5 (partial)	EFX66479.1
	Wnt6	EFX86167.1
	Wnt7	EFX66449.1
	Wnt8	EFX83364.1
	Wnt9	EFX86385.1
	Wnt10	EFX86388.1
	Wnt11	EFX77586.1
	Wnt16	EFX82994.1
	WntA (partial)	EFX69968.1
Drosophila melanogaster (Dm)	Wnt1	NP_523502.1
	Wnt5	NP_476924.1
	Wnt6	NP_609108.3
	Wnt7	NP_476810.1
	Wnt8/D	NP_650272.1
	Wnt9	NP_476972.2
	Wnt10	NP_609109.3
Euperipatoides kanangrensis (Ek)	Wnt1 (partial)	ABY60732.1
	Wnt2	CDI40099.1
	Wnt4	CDI40100.1
	Wnt5	CDI40101.1
	Wnt6	CDI40102.1
	Wnt7	CDI40103.1
	Wnt9	CDI40104.1
	Wnt10	CDI40105.1
	Wnt11	CDI40106.1
	Wnt16	CDI40107.1
	WntA	CDI40108.1
Tribolium castaneum (Tc)	Wnt1	NP_001107822.1
	Wnt5	XP_974684.1
	Wnt6	NP_001164137.1
	Wnt7	XP_008196351.1
	Wnt8/D	XP_971439.1
	Wnt9	XP_015835609.1
	Wnt10	XP_015835532.1
	Wnt11	XP_015835988.1
	WntA	KYB26594.1

Supplementary Table 3. Protein accession numbers for Frizzled sequences used in this study.

Species	Frizzled genes	Protein accession number
Centruroides sculpturatus (Cs)	fz1	XP_023229313.1
	fz2	XP_023229226.1
	fz3	XP_023221477.1
	fz4-1	XP_023220057.1
	fz4-2	XP_023233147.1
Parasteatoda tepidariorum (Pt)	fz1	XP_015922960.1
	fz2	XP_015922948.1
	fz4-1	XP_015911110.1
	fz4-2	XP_015910102.1
Ixodes scapularis (Is)	fz1	EEC05379.1
	fz2	EEC03613.1
	fz4	XP_002402968.1
Strigamia maritima (Sm)	fz1 (SMAR014833)	**Table S30. Details of the manually annotated genes of *S. maritima*.** (Chipman et al. 2014)
	fz2 (SMAR012389)	
	fz3 (SMAR007293)	
	fz4 (SMAR009650)	
Drosophila melanogaster (Dm)	fz1	CAA38458.1
	fz2	AAC47273.1
	fz3	ABW09320.1
	fz4	NP_511068.2
Tribolium castaneum (Tc)	fz1	EFA04653.1
	fz2	EFA01325.1
	fz4	EFA09255.1


Ubx
Abd-A
Dfd
Scr
Abd-B
lab
Antp
pb
ftz
Hox3
1.0