

Supplementary Figures - Molecular evolutionary trends and biosynthesis pathways in the Oribatida revealed by the genome of *Archegozetes longisetosus*

Adrian Brückner, Austen A. Barnett, Prashant Bhat, Igor A. Antoshechkin and Sheila A. Kitchen

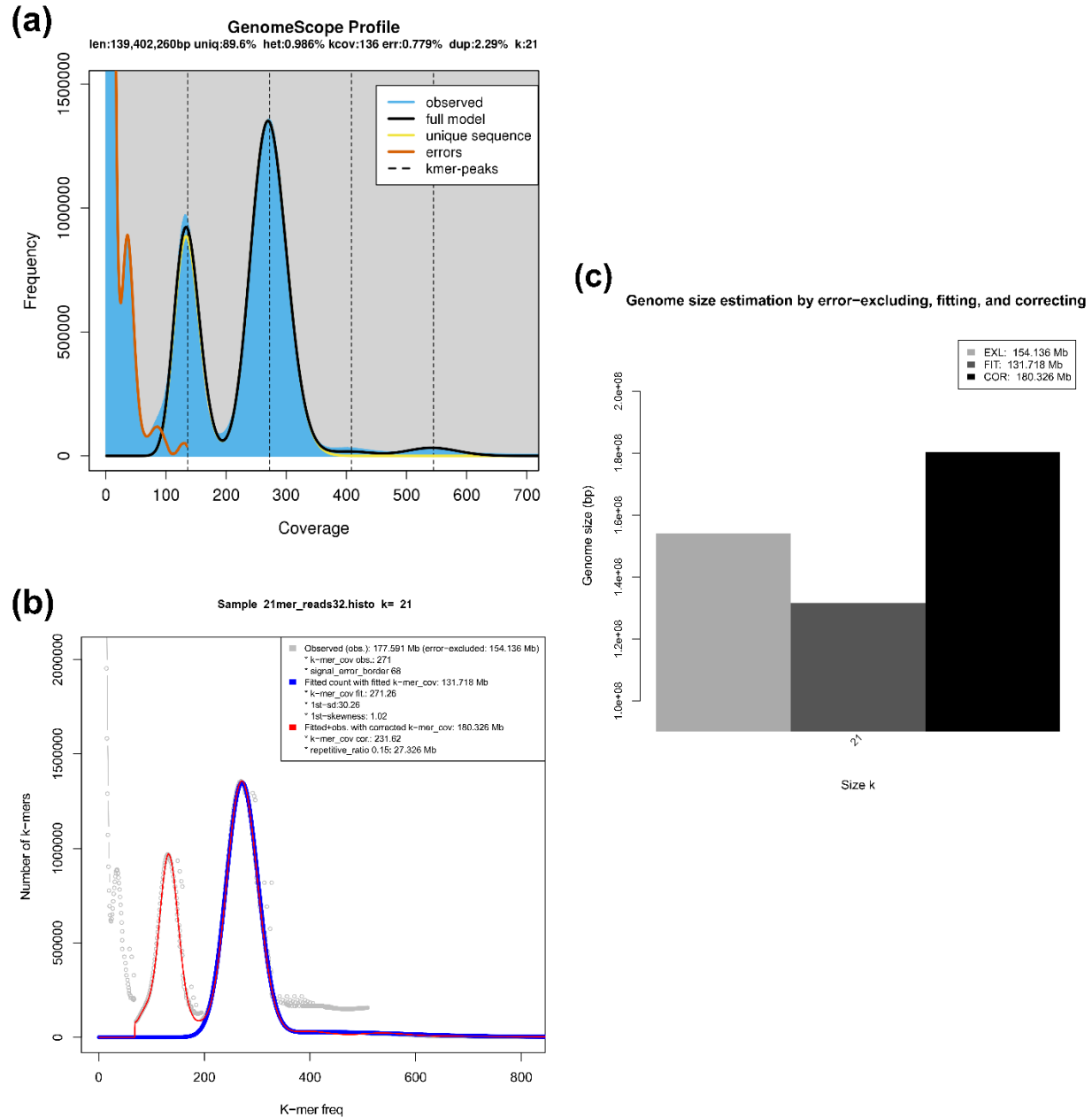


Figure S1 Results of *in silico* genome size estimations based on jellyfish *k-mer* counting using **a**: GenomeScope v1.0 and **b** and **c**: the findGSE v0.1.0 R package (Sun *et al.*, 2018).

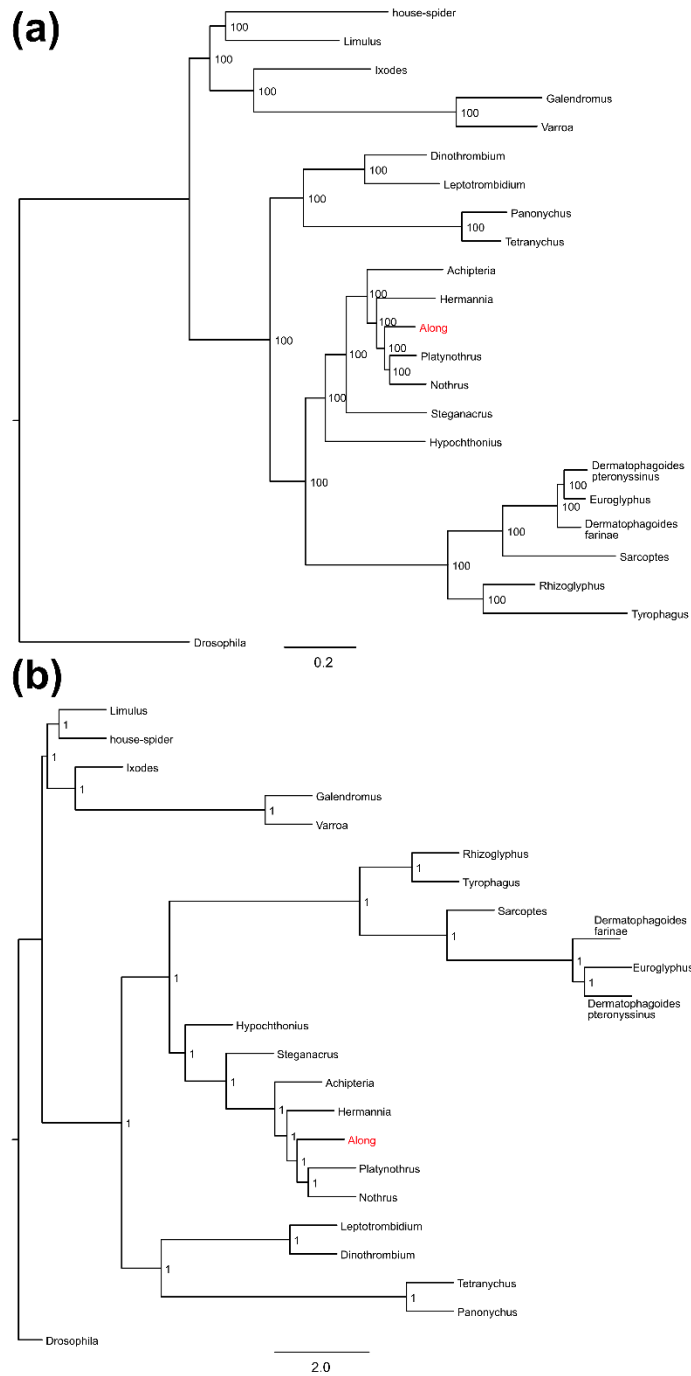


Figure S2 Phylogenetic placement of *Archegozetes longisetosus* among other chelicerates. **a:** Maximum likelihood phylogeny based on concatenation of 1,121 orthologs. Branch lengths unit is substitutions per site and the node values reflect bootstrap supports. **b:** Coalescence species tree reconstruction of the 1,121 filtered orthogroups. Branch lengths are presented in coalescent units (differences in the 1,121 gene trees) and the node values reflect the local posterior probabilities.

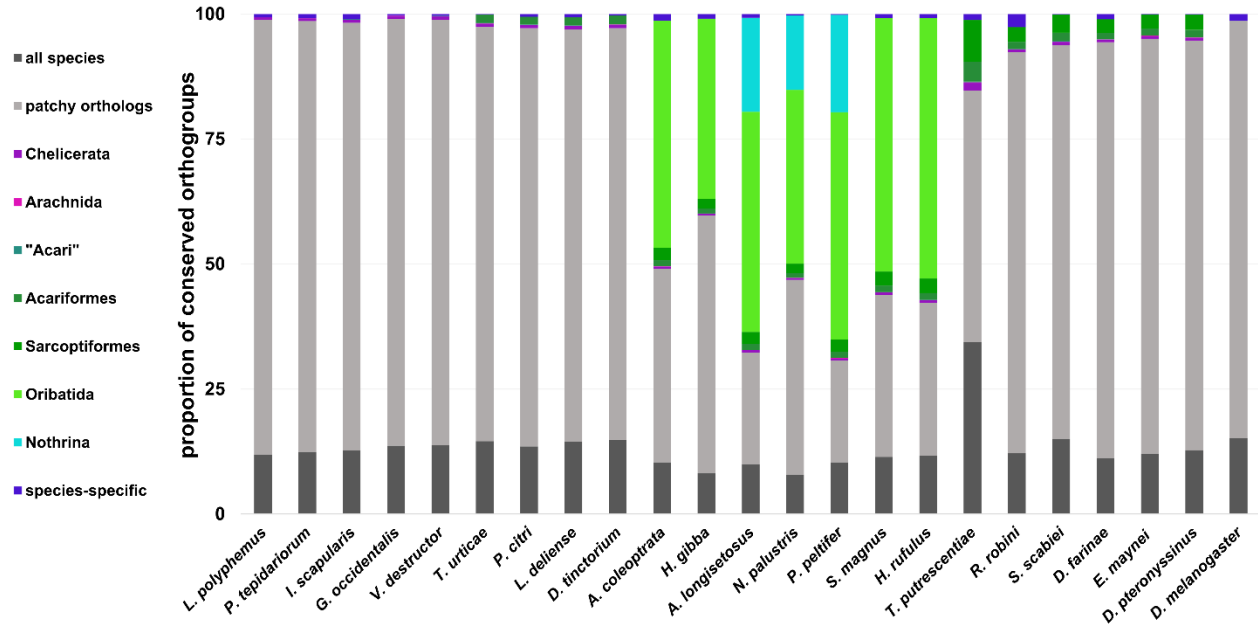
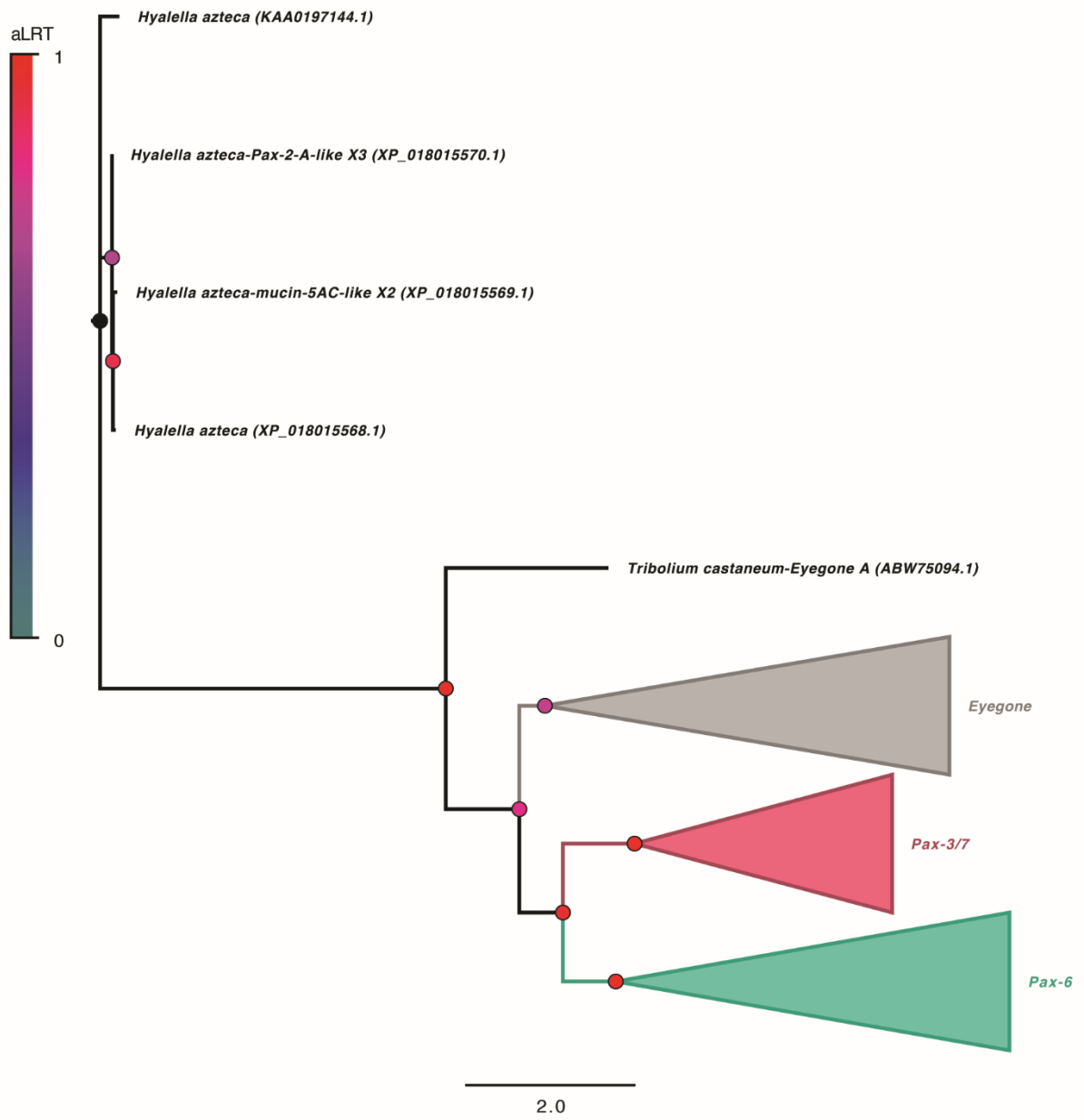


Figure S3 Comparisons of protein-coding genes of 23 arthropod species, including *Archezogetes*.

The bar charts show the proportion of orthogroup conservation with each species (see insert legend) based on OrthoFinder clustering.

Figure S4 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Paired protein orthologs. **(a)** Maximum likelihood tree showing the relationship of the Eyegone, Pax-3/7, and Pax-6 clades as collapsed subtrees. **(b)** The un-collapsed clade in **A** showing the phylogenetic relationships of selected Eyegone proteins and the putative *A. longisetosus* Eyegone ortholog. **(c)** The un-collapsed clade in **A** showing the phylogenetic relationships of selected Pax-3/7 proteins and the putative *A. longisetosus* Pax-3/7 orthologs. **(d)** The un-collapsed clade in **A** showing the phylogenetic relationships of selected Pax-6 proteins and the putative *A. longisetosus* Pax-6 ortholog. All *A. longisetosus* orthologs are in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

(a)



(b)



(c)



(d)



Pax-6

Figure S5 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Eyes absent (Eya) protein ortholog and selected metazoan Eya proteins. The *A. longisetosus* ortholog is in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

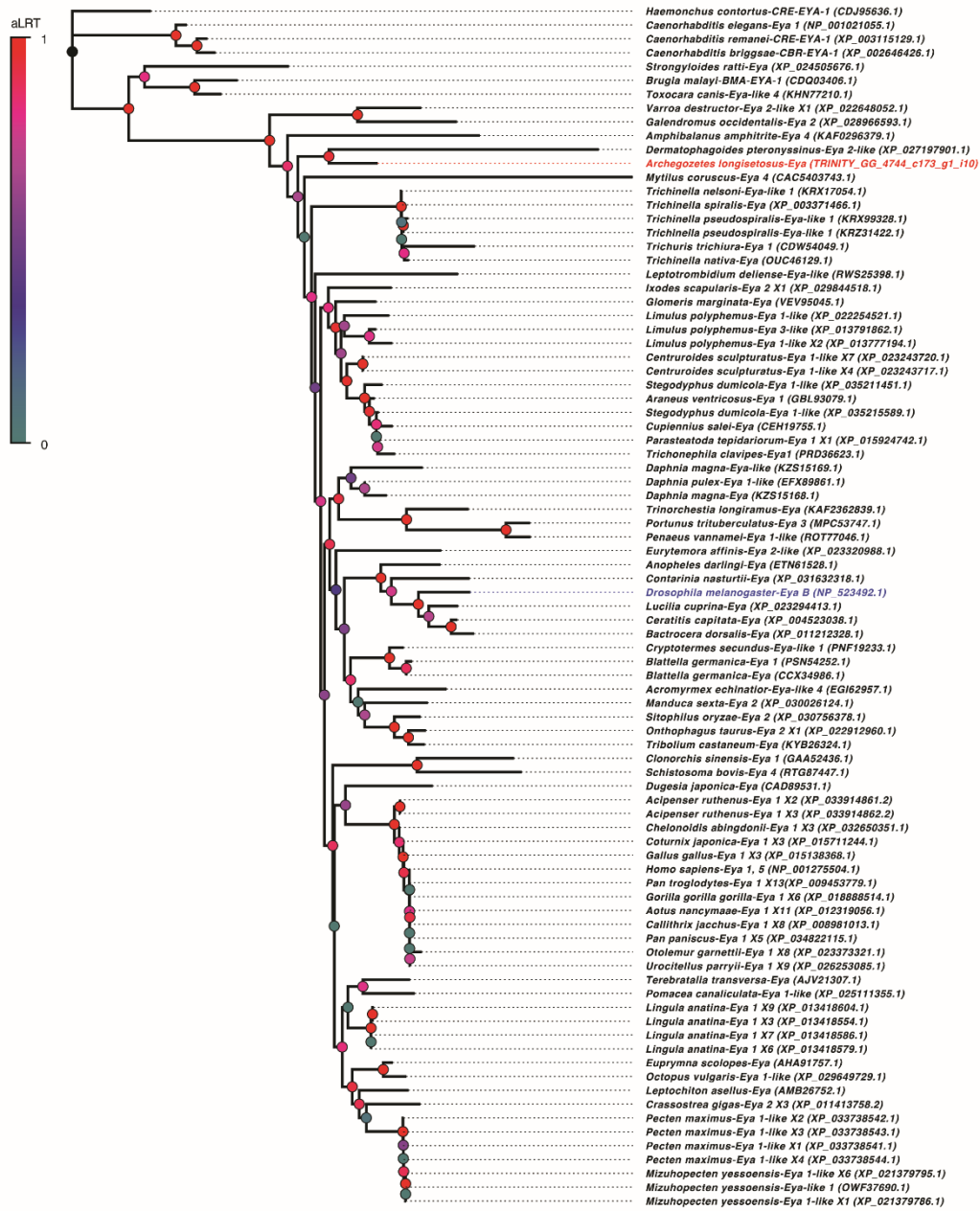


Figure S6 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Hairy protein ortholog and selected metazoan Hairy proteins. Hairy/E(spl) proteins were used as an outgroup. The *A. longisetosus* ortholog is in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

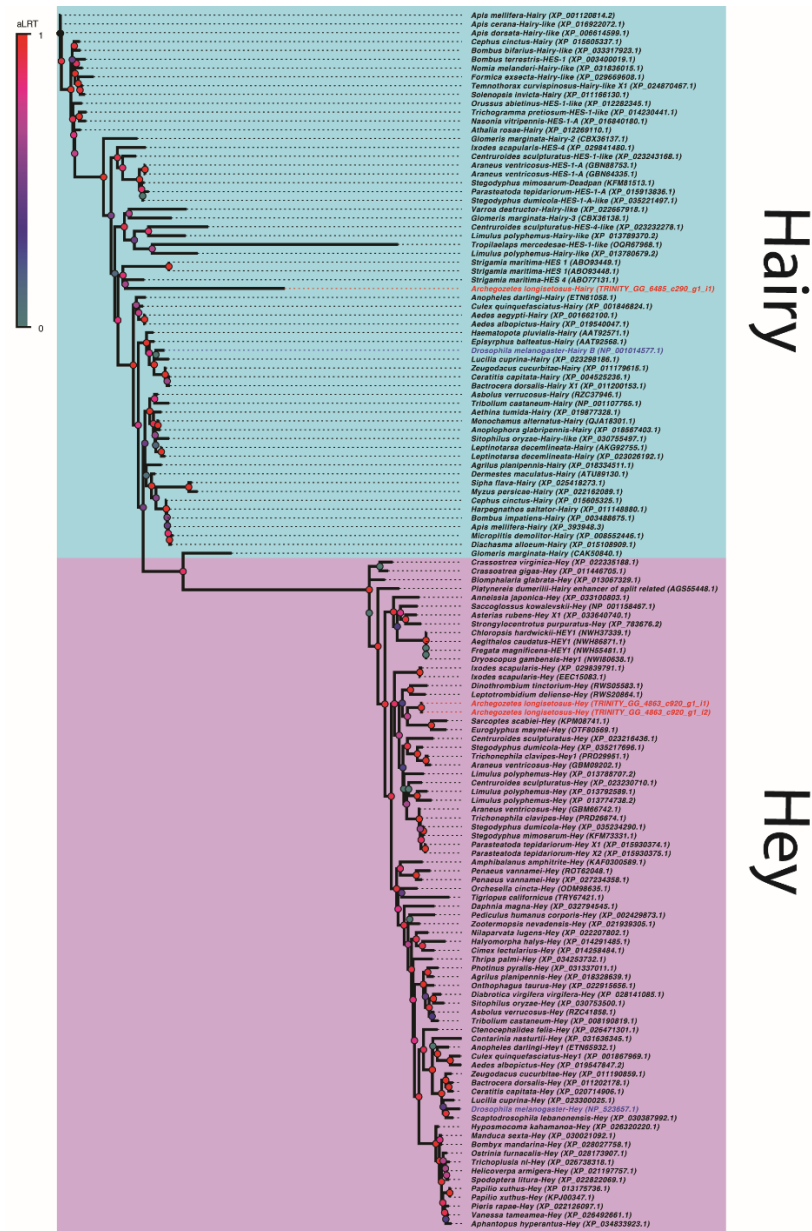


Figure S7 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Omb, T-box H15, and TBX1 protein orthologs and selected metazoan T-box proteins. All *A. longisetosus* orthologs are in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

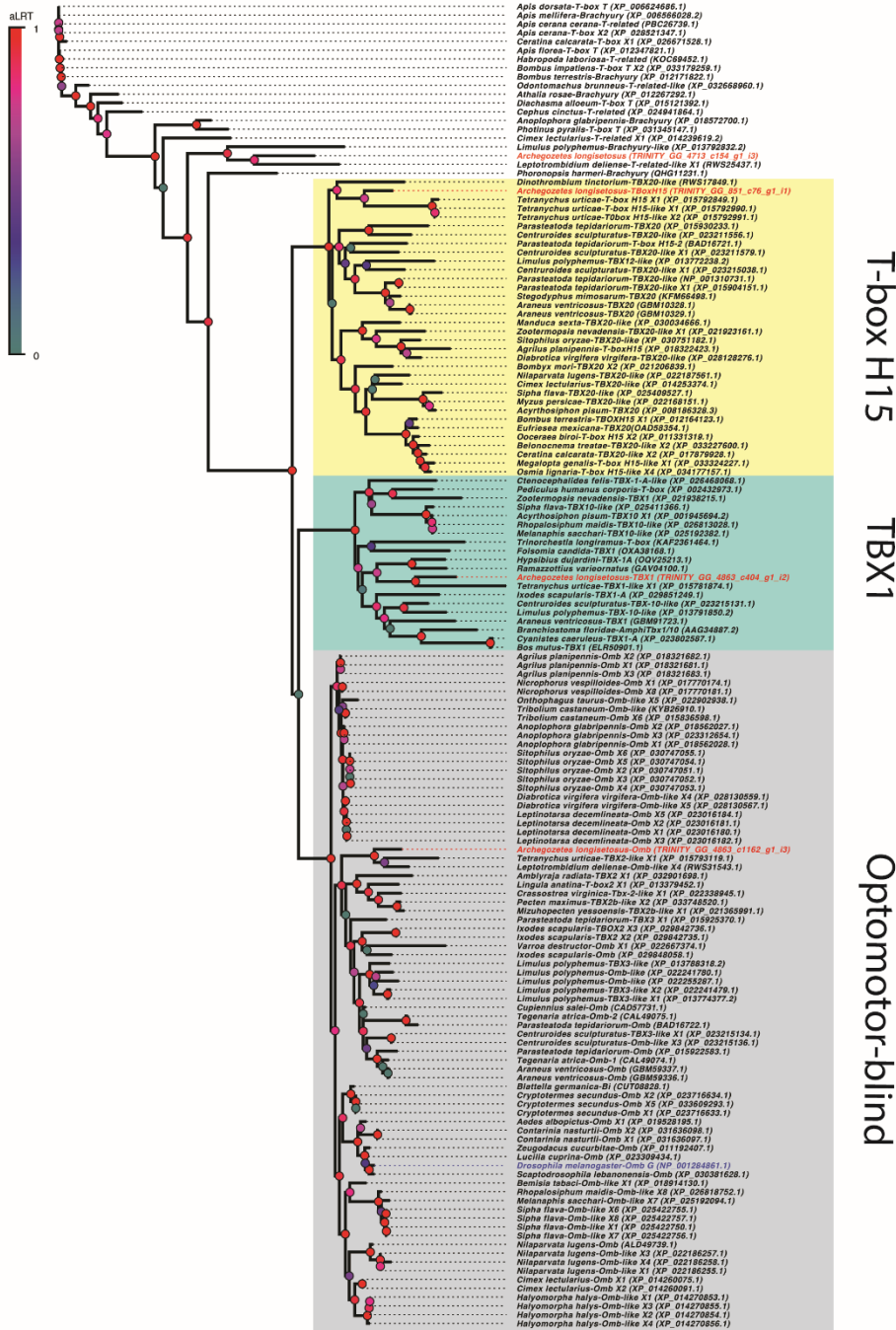


Figure S8 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Runt protein ortholog and selected arthropod Runt proteins. The *A. longisetosus* ortholog is in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

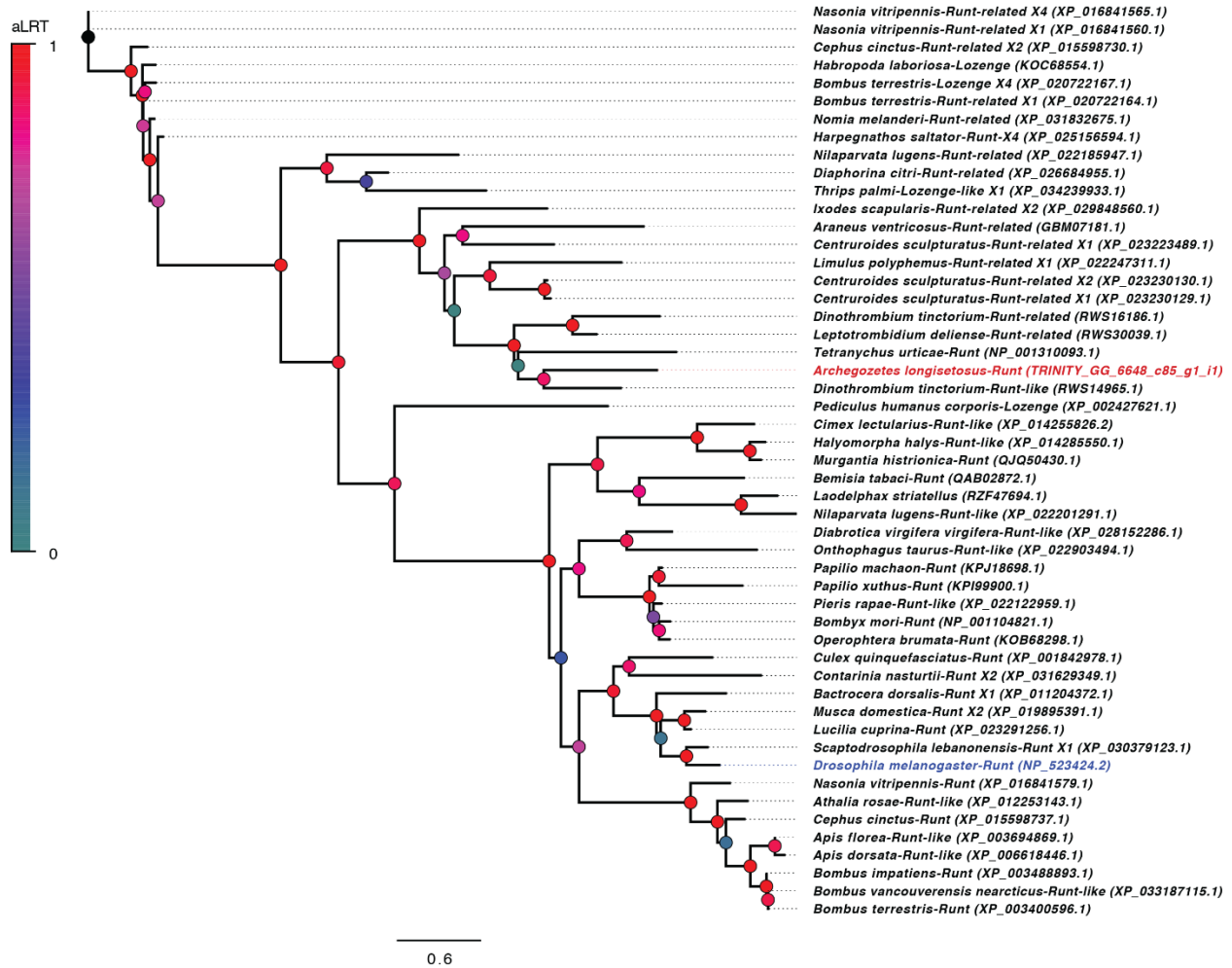


Figure S9 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Six family protein orthologs and selected metazoan Six family proteins. All *A. longisetosus* orthologs are in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.

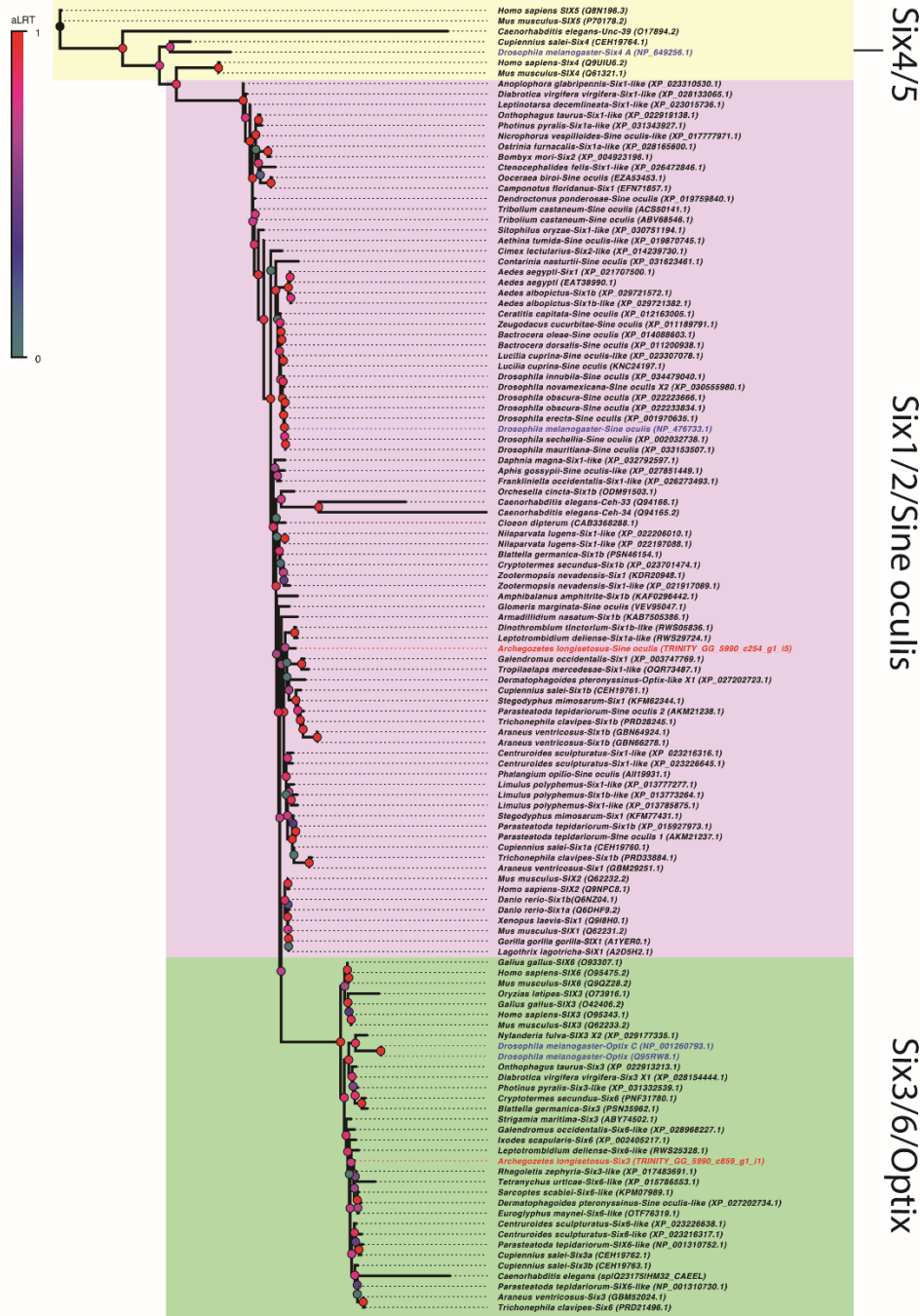


Figure S10 Maximum likelihood phylogenetic analyses of the *A. longisetosus* Sp-family protein orthologs and selected metazoan Sp-family proteins. All *A. longisetosus* orthologs are in red, and the *D. melanogaster* orthologs are in blue. Node support was calculated using the approximate likelihood ratio (aLRT) method and is represented by the color of each node. All taxa are represented by their species names, gene names if given, and their NCBI accession numbers.



Figure S12 Unrooted maximum-likelihood phylogenetic trees of cell-wall degrading enzymes based on the alignment of amino acid sequences. Branch lengths unit is substitutions per site and the node values reflect bootstrap supports. *Archeozetes* sequences are highlighted in red.

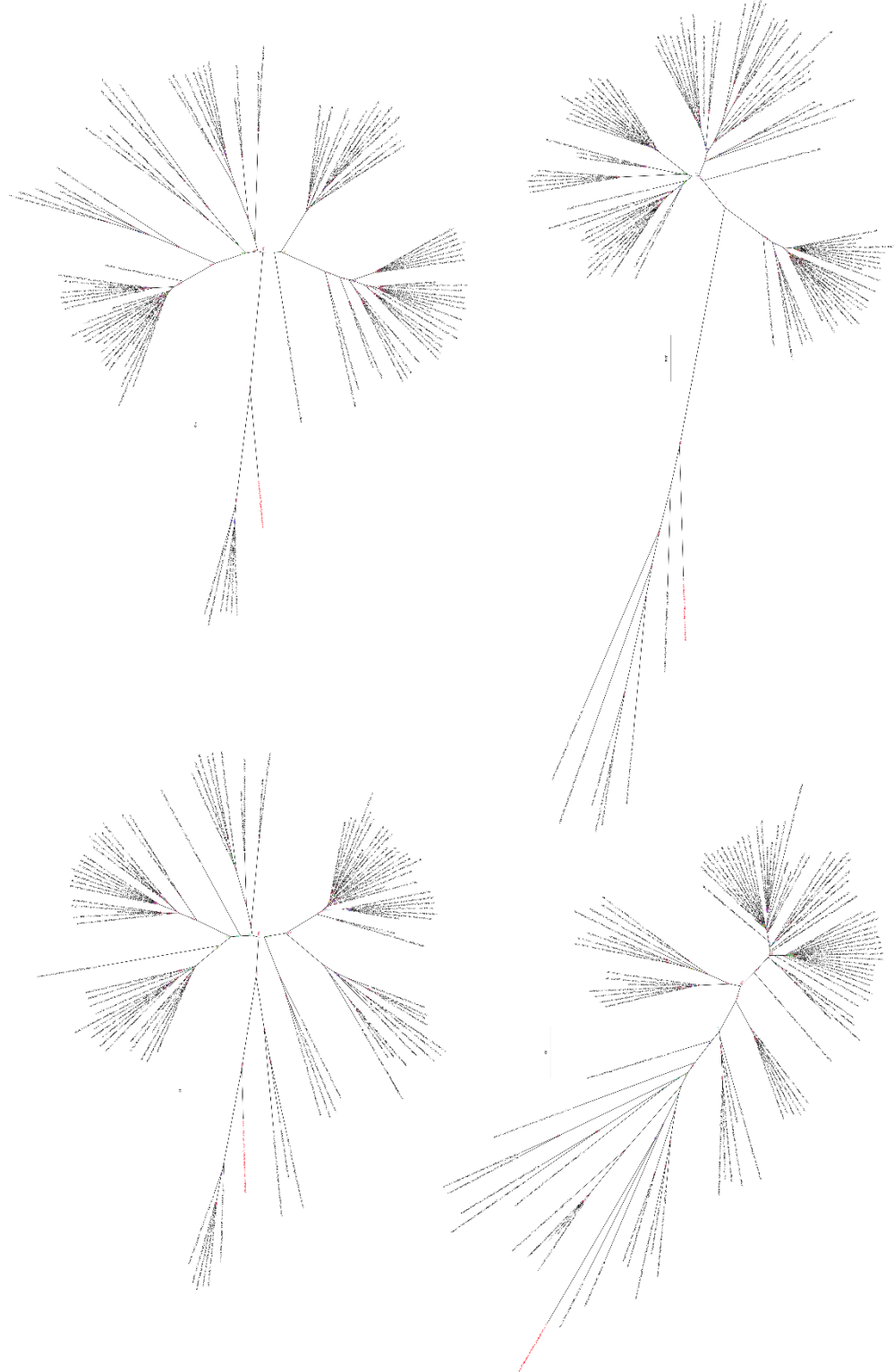


Figure S13 Maximum likelihood phylogeny of ionotropic receptors and ionotropic glutamate receptors of *Archeogozetes* (Along), *Dinothrombium* (Dt), *Leptothrombidium* (Ld), *Tetranychus* (Tu) and *Drosophila* (Dmel). The tree was rooted to the middle point. Branch lengths unit is substitutions per site and the node values reflect bootstrap supports.

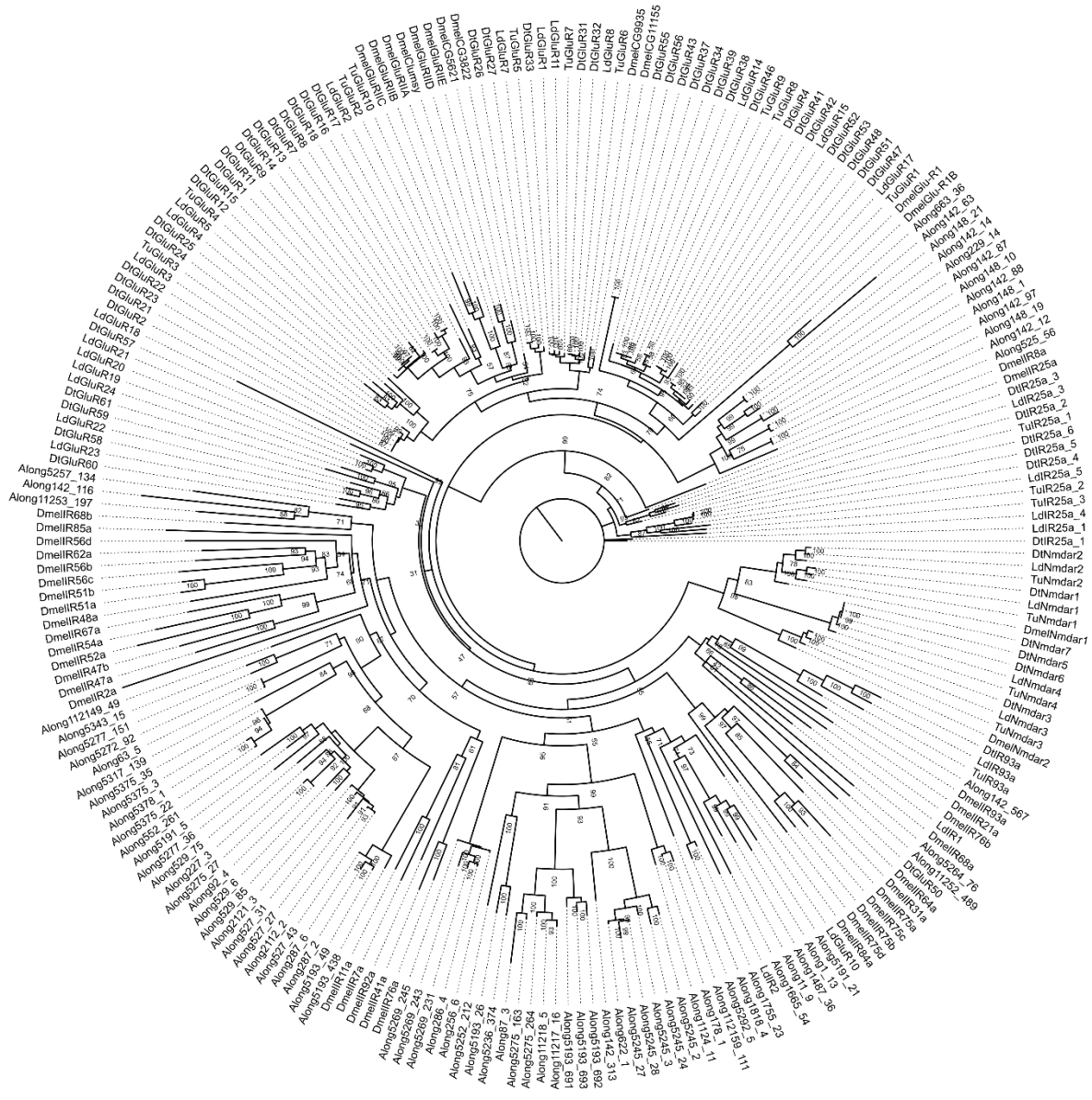


Figure S14 Maximum likelihood phylogenetic tree of gustatory receptors of *Archegozetes* (Along), *Ixodes* (Is), *Tropilaelaps* (Tm), *Metaseiulus* (Mocc) and *Drosophila* (Dmel). The tree was rooted to the middle point. Branch lengths unit is substitutions per site and the node values reflect bootstrap supports.



Figure S15 Unrooted maximum-likelihood tree of glycoside hydrolase family 48 members (GH_48) from *Streptomyces* bacteria and HGT genes from other arthropods as well as *Archezogetes*. Branch lengths unit is substitutions per site and the node values reflect bootstrap supports.

