Evolutionary Conservation of Five Pax6 Homologs under Strong Purifying Selection in the Horseshoe Crab Species Cluster

Tanay Dakarapu and Markus Friedrich *

Posted Date: 5 September 2023

doi: 10.20944/preprints202309.0246.v1

Keywords: Pax6; Xiphosura; eye evolution; chelicerates; visual system; horseshoe crab; subfuctionalization; gene family evolution; gene duplication; gene regulatory network evolution; whole genome duplication

Review

Evolutionary Conservation of Five *Pax6* Homologs under Strong Purifying Selection in the Horseshoe Crab Species Cluster

Tanay Dakarapu [1] and Markus Friedrich [1,2,*]

[1] Department of Biological Sciences, Wayne State University, 5047 Gullen Mall, Detroit, MI 48202; tanaydak@umich.edu
[2] Department of Ophthalmological, Visual, and Anatomical Sciences, Wayne State University, School of Medicine, 540 East Canfield Avenue, Detroit, MI 48201; friedrichm@wayne.edu
* Correspondence: friedrichm@wayne.edu; Tel.: 001 313 577 9612.

Abstract: Horseshoe crabs represent the most ancestral chelicerate lineage characterized by marine ecology and the possession of lateral compound eyes. While considered living fossils, recent studies reported an unusual number of *Pax6* transcription factor genes in the Atlantic horseshoe crab *Limulus polyphemus*. *Pax* genes encode a family of ancient metazoan transcription factors, which comprise seven subfamilies. Among these, the members of the *Pax6* subfamily confer critical functions in the development of the head, the visual system, and further body plan components. Arthropods are characterized by two *Pax6* subfamily homologs that were discovered in *Drosophila* and named *eyeless* (*ey*) and *twin of eyeless* (*toy*). However, whole genome sequence searches uncovered three homologs of *ey* and two homologs of *toy* in *L. polyphemus*. These numbers are explained by the occurrence of three whole genome duplications (WGD) in the lineage to the last common ancestor of *L. polyphemus* and its closely related three additional members of the extant horseshoe crab species cluster. Here we report comparative sequence evidence for the functional conservation of the five *L. polyphemus Pax6* transcription factor homologs. Our analyses reveal that all paralogs are conserved in the approximately 135 million-year-old horseshoe crab species cluster and that they evolve under strong purifying selection. These findings identify subfunctionalization as the likeliest post-WGD outcome for the five *Pax6* homologs. While awaiting confirmation by gene expression studies, this scenario would reconcile the discrepancy between the WGD-expanded gene repertoires and phenotypic stasis in the horseshoe crabs.

Keywords: *Pax6*; Xiphosura; eye evolution; chelicerates; visual system; horseshoe crab; subfuctionalization; gene family evolution; gene duplication; gene regulatory network evolution; whole genome duplication

1. Introduction

Together with Myriapoda and Pancrustacea, chelicerates constitute one of the three monophyletic subphyla of the arthropods. Best known for fear-instilling terrestrial taxa like harvestmen, scorpions, and spiders [1], chelicerate biodiversity also includes two ancient marine lineages, i.e. sea spiders and horseshoe crabs (Xiphosura). Among these two, horseshoe crabs stand out by the conservation of canonical arthropod compound eyes [2], as part of an overall body plan design that is famous for qualifying as a "living fossil" [3]. Equally important, the Atlantic horseshoe crab *Limulus polyphemus* has played a pivotal role in early vision research [2].

Two vision-related gene families received exceptional attention in *L. polyphemus*. This includes opsins, the GPCR family of light-sensitive transmembrane receptors [4], and the developmental transcription factor subfamily *Pax6* [5]. The first arthropod homologs of the *Pax* gene family were discovered in the fruit fly Drosophila melanogaster, which possesses two homologs of *Pax6* named *eyeless* (*ey*) and *twin of eyeless* (*toy*) [6]. Comparative analyses of the past 15 years revealed that the presence of two *Pax6* genes in *Drosophila* and other arthropods, including species from Pancrustacea,

Myriapoda, and Chelicerata, is due to a gene duplication that must have taken place over 500 million years ago in the arthropod stem lineage [7]. Given this deep conservation of singleton homologs of *ey* and *toy* in arthropods, it came as a surprise to find not only one *Pax6* ortholog in *L. polyphemus* that was originally uncovered by gene cloning efforts [5] but three orthologs of *ey* and two orthologs of *toy* in the whole genome assembly of *L. polyphemus* [8].

While unexpected given the generally limited range of *Pax* transcription factor gene family size variation in arthropods, these findings align with the comparative genomic reconstruction of three whole genome duplication (WGD) events in the xiphosuran stem lineage. Initial evidence of this exceptional trajectory emerged from the analysis of the first genome assemblies for the Atlantic horseshoe crab *L. polyphemus* [4,9]. In addition to *L. polyphemus*, extant Xiphosura includes the Mangrove horseshoe crab *Carcinoscorpius rotundicauda*, the Chinese horseshoe crab *Tachypleus tridentatus*, and the Indo-Pacific horseshoe crab *Tachypleus gigas* [10,11]. Comparative genomic analyses timed the earliest split of *L. polyphemus* from the other three species to approximately 135 million years ago and the split of *C. rotundicauda* from the last common ancestor of *T. tridentatus* and *T. gigas* to about 50 million years ago [12,13].

The comparative analysis of the organization of the homeobox (Hox) gene clusters in *L. polyphemus*, *C. rotundicauda*, and *T. tridentatus* produced evidence of at least one WGD in the xiphosuran lineage to the last common ancestor of extant horseshoe crabs [4,9]. Subsequent analyses of high-quality whole genome assemblies for all four xiphosuran species, however, produced evidence of a total of three WGDs in the xiphosuran lineage [14–16]. The genomes of the members of the horseshoe crab species cluster thus stand out by having been molded by one more WGD compared to the two rounds of WGD in early vertebrates [17–20] and two more rounds of WGD compared to the independent WGD discovered for terrestrial Chelicerates [21], i.e. the Arachnopulmonata (Figure 1).

While these insights shed light into the cause for the large number of *Pax6* homologs in horseshoe crab species, it is unknown which evolutionary forces led to their final numbers and whether these paralogous loci serve novel or ancestral functions. Novel functions can be acquired by a process that is referred to as neofunctionalization following gene duplication [22,23]. Alternatively, the duplication of highly pleiotropic genes frequently results in the differential inheritance of ancestral functions in the descendant gene duplicates [22,23]. These basic scenarios can be distinguished by comparative analysis of substitution rates both at the amino acid and nucleotide sequence levels. Neofunctionalized paralogs, for instance, are characterized by transiently increased amino acid substitution rates as a result of positive selection on function optimizing amino acid residue replacements [24]. This acceleration is effectively measured by substitution rate comparisons between of sister gene duplicates. As one paralog continues to confer ancestral function and hence evolves under purifying selection, the neofunctionalized sister paralog will be characterized by a significantly higher relative substitution rate. In the case of subfunctionalization, by contrast, sister paralogs are less likely to differ dramatically in their protein sequence substitution rates. However, the release of constraints resulting from the reduction of pleiotropy can lead to moderate but detectable transient increases in protein substitution rates [25]. Last but not least, paralogous genes can also maintain genetically redundant functions, thereby buffering the impacts of environmental or gene regulatory variation [26,27]. One example of this is the deeply conserved redundant specification of the ocular segment region by *ey* and *toy* in arthropods [7].

An alternative approach to probing for innovative versus conservative sequence evolution trajectories in the comparison of non-synonymous versus synonymous sequence changes at the nucleotide sequence level [28]. Grounded in the neutral theory of molecular evolution [29], strong positive selection on proteins is predicted to lead to a higher rate in the fixation of non-synonymous nucleotide substitutions (dN) compared to the base rate of silent substitution accumulation (dS). By the same token, strong purifying selection on protein sequences is indicated by a lower dN compared to dS.

In this study, we pursued two aims. First, we asked whether all five *Pax6* paralogs discovered in *L. polyphemus* were conserved in other horseshoe crab species. Our second aim was to probe for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2023 doi:10.20944/preprints202309.0246.v1

sequence evidence of innovative versus conservative gene duplication outcomes by analyzing dN/dS ratios along the horseshoe crab species cluster tree. Our efforts revealed the conservation of *Pax6* gene subfamily size in the horseshoe crabs and that all five xiphosuran *Pax6* homologs have evolved under strong purifying selection identifying subfunctionalization as their likeliest post-WGD trajectories.

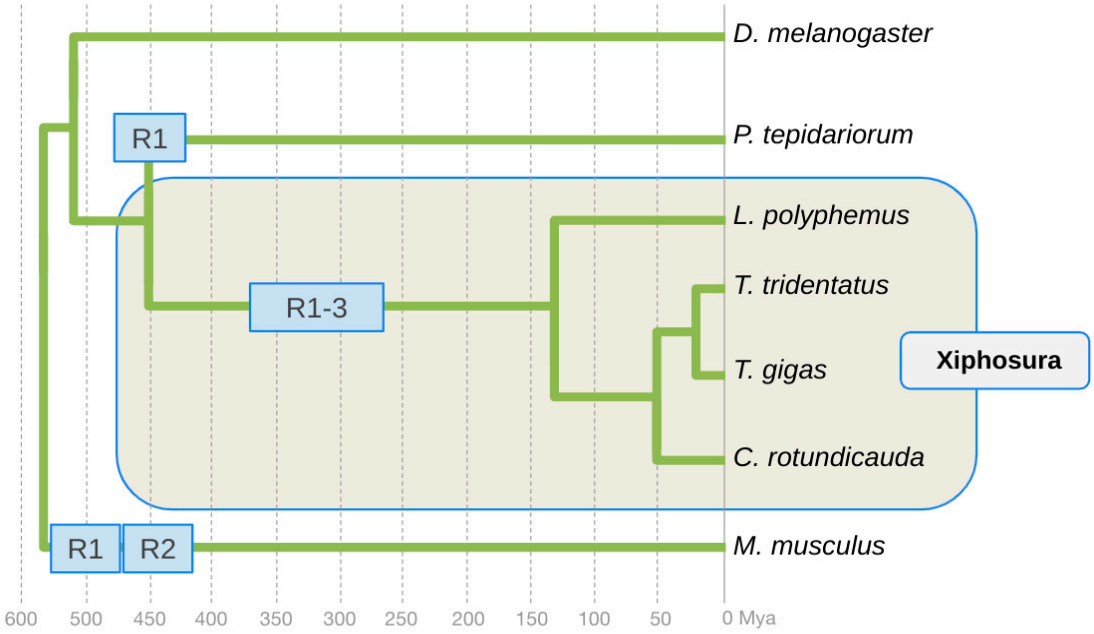

Figure 1. Phylogenetic framework of whole genome duplication events. Horseshoe crab species relationships and divergence times based on [13]. Approximate timing of the WGD event in the Arachnopulmonata represented by *P. tepidariorum* based on [21]. Approximate timing of the WDG events in the early vertebrate lineage based on [17].

2. Materials and Methods

Genome assemblies

Homolog searches were conducted in the GenBank assembly GCF_000517525.1 of *L. polyphemus*, GenBank assembly GCA_011833715.1 for the IMCB_SINMHF_001 isolate of *C. rotundicauda*, GenBank assembly GCA_004102145.1 for *T. tridentatus* isolate BBG1, and GenBank assembly GCA_014155125.1 for *T. gigas* isolate IMCB_SINCHF_001.

Homolog searches

BLASTn and BLASTp searches were conducted in the NCBI genome assembly user interfaces.

Gene model annotation

Genome models were hand-annotated in downloaded chromosome and scaffold text files.

Gene tree reconstruction

We generated a multiple sequence alignment of the conceptual protein sequences with T-Coffee at default settings [30]. Given the high degree of overall sequence conservation, the alignment used as direct input for gene tree estimation with Randomized Accelerated Maximum Likelihood (RAxML) [31] as made available in the CIPRES Science Gateway online user interface [32]. We applied the Jones-Taylor-Thornton (JTT) model of amino acid sequence evolution and implemented gamma-distributed substitution rates across sites with 3 categories.

 doi:10.20944/preprints202309.0246.v1

dN/dS analysis

dN/dS ratios were computed using the online Ka/Ks Calculation tool (https://services.cbu.uib.no/tools/kaks) [33,34]. We submitted complete open reading frame nucleotide sequences and the species tree (Figure 1) and applied the discrete Grantham submatrix option without codon bias.

3. Results

Conservation of three homologs of *ey* in horseshoe crabs

Previous BLAST searches identified three distinct homologs of *ey* in *L. polyphemus*, which contrasted with the conservation of a singleton homolog of *ey* throughout all other arthropods investigated so far, including other chelicerates [8]. To probe for the conservation of all three *L. polyphemus ey* homologs in other horseshoe crab species, we searched the genome assemblies of *C. rotundicauda*, *T. tridentatus*, and *T. gigas* by tBLASTn with the *L. polyphemus ey1*, *ey2*, and *ey3* homologs as queries. Candidate homologs were re-BLASTed against reference protein databases of *L. polyphemus* and *D. melanogaster* to filter false positive orthology assignments. These efforts and preliminary gene tree analyses identified 1:1 orthologs of *L. polyphemus ey1*, *ey2*, and *ey3* in all of the other three horseshoe crab species (Figure 2 and Table S1). Of note, all of the putative *ey* homologs were characterized by the previously described lysine residue at position 64 of the Pax domain, which is diagnostic for *ey* throughout arthropods (Text document S1) [7,8].

Subsequently, we hand annotated all of the relevant loci to produce complete conceptual transcript, open reading frame, and protein sequences for *ey1*, *ey2*, and *ey3* of *C. rotundicauda*, *T. tridentatus*, and *T. gigas* (Text document S2 - 4). The multiple sequence alignment of the protein sequences revealed a high level of sequence conservation among the xiphosuran *ey1*, *ey2*, and *ey3* orthologs (Text document S1). Most sequence differences stemmed from single amino acid residue replacements. The *ey3* homolog of *L. polyphemus* stood out by a two amino acid long deletion and a three amino acid residue long alanine string insertion. Both of these larger scale differences mapped to sequence variable regions of *ey* based on outgroup comparisons with the singleton *ey* orthologs of the common house spider *Parasteatoda tepidariorum* and the red flour beetle *Tribolium castaneum* (Text document S1).

The global gene tree analysis confirmed the presumed orthology relationships of the individual *ey* homologs but did not resolve the relationship between the *ey1*, *ey2*, and *ey3* ortholog clusters (Figure 2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2023 doi:10.20944/preprints202309.0246.v1

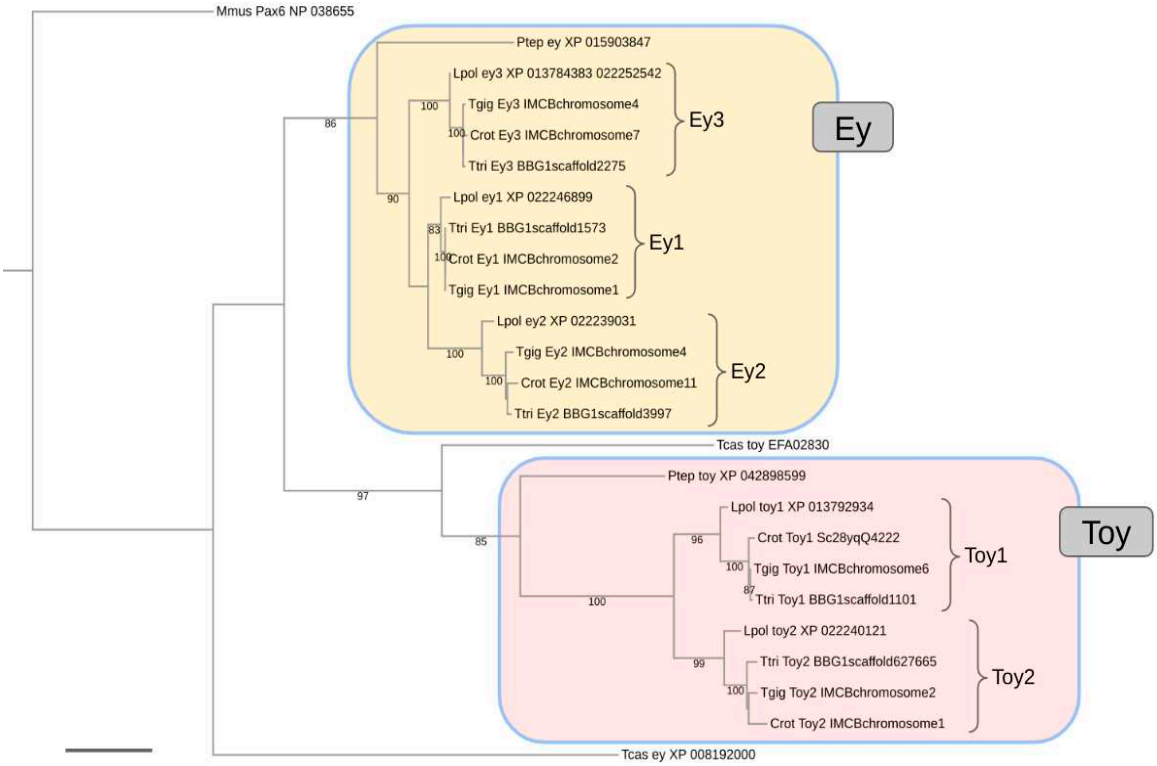

Figure 2. Maximum likelihood gene tree of xiphosuran *Pax6* homologs. Numbers at branches represent nonparametric bootstrap support from 100 replications. Scale bar corresponds to 0.1 substitutions per amino acid site. Species name abbreviations: Crot = *C. rotundicauda*, Lpol = *L. polyphemus*, Ptep = *P. tepidariorum* (Araneae) Tcas = *T. castaneum* (Coleoptera), Tgig = *T. gigas*, Ttri = *T. tridentatus*).

Conservation of two homologs of *toy* in horseshoe crabs

Previous studies also identified two homologs of *toy* in *L. polyphemus* [8]. To probe for their conservation in the horseshoe crab species assembly, we proceeded in the same way as in the investigation of the conservation of the three homologs of *ey*. As a result of this effort, we found that also both *L. polyphemus* homologs of *toy* were conserved in *C. rotundicauda*, *T. tridentatus*, and *T. gigas* (Table S1; Text document S1-3). We further found that all *toy* homologs were consistently characterized by the previously described diagnostic arginine residue at position 64 of the Pax domain [7,8]. The global gene tree analysis confirmed the presumed orthology relationships of all individual *toy* homologs (Figure 2).

As in the case of the *ey* homologs, multiple sequence alignment of the xiphosuran *toy* protein sequences revealed a high level of sequence conservation (Text document S1). Most sequence differences stemmed from single amino acid residue replacements. The *toy1* homolog of *T. gigas* was characterized by the deletion of a proline residue in a sequence variable region C-terminal of the homeodomain (Text document S1). Moreover, the *toy1* homolog of *L. polyphemus* was characterized by the insertion of three amino acid residues in a variable sequence region even further C-terminal of the before-mentioned *T. gigas* variable site (Text document S1).

Strong purifying selection on all horseshoe crab *Pax6* homologs

The conservation of complete open reading frames for all xiphosuran *Pax6* transcription factor homologs provided strong evidence of their continued functionality in the horseshoe species cluster. To gain further insights into the functionalization trajectories of these critical regulatory genes following three WGDs and the three speciation events (Figure 1), we took advantage of the overall

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 September 2023 doi:10.20944/preprints202309.0246.v1

high conservation at the nucleotide sequence level to probe for selection pressure by dN/dS ratio analyses (Figure 3).

Consistent with strong functional conservation, all five paralogs were characterized by dN/dS values below zero. Averaged across the entire phylogeny of the xiphosuran species tree, the dN/dS ratios of *toy1* and *toy2* were very similar with values ranging between 0.08 and 0.12, respectively (Table S2). Compared to these points of reference, the dN/dS values of 0.02 for *ey1* and 0.04 for *ey3* suggested a slightly higher strength of purifying selection compared to *toy1* and *toy2*. The *ey2* homolog, finally, stood out with the highest across phylogeny dN/dS value of 0.22, indicating a possibly weaker degree of purifying selection.

Overall, these findings corroborated the functionality of all five *Pax6* transcription factor paralogs in the Xiphosura and were best compatible with subfunctionalization or conserved functional redundancy following gene family expansion as a result of WGD.

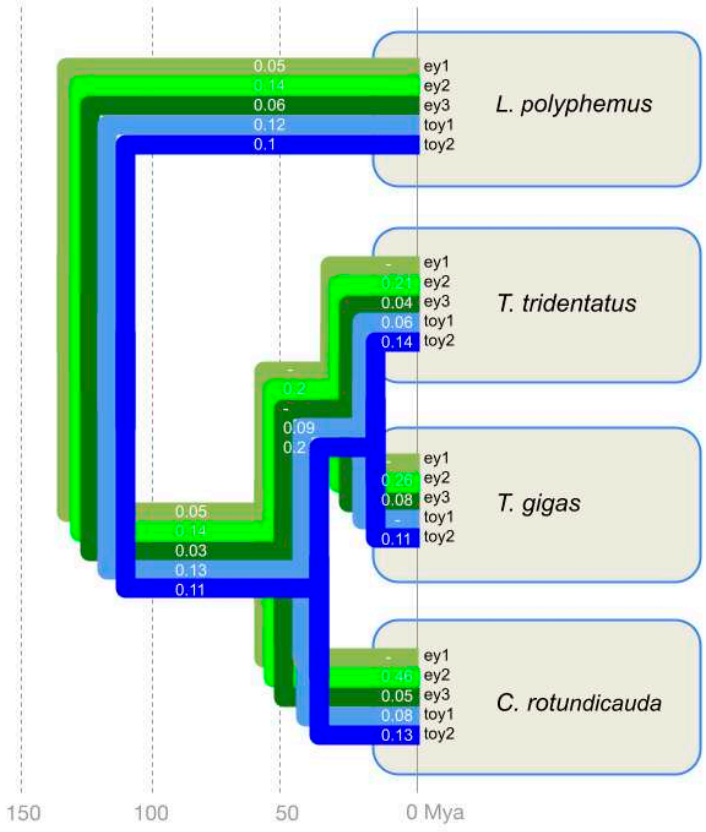

Figure 3. Selection pressure on the five xiphosuran Pax6 homologs along quantified by dN/dS ratios.

4. Discussion

As part of the bilaterian gene toolkit [35], most members of the *Pax* transcription factor gene family perform essential regulatory gene network functions during embryonic and postembryonic tissue and cell specification processes [36]. Successful execution of these pivotal functions depends on precisely regulated expression in time and tissue context. A systematic analysis in *Drosophila* demonstrated the critical effects of these variables as the misexpression of *Pax* genes leads to dramatic morphological phenotypes in the developing head [37]. In the specific case of the arthropod *Pax* transcription factor paralogs *ey* and *toy*, these aspects are highlighted by the extreme nature of lack of function and gene misexpression consequences. Both, *ey* and *toy* mutant *Drosophila* strains are characterized by depletion of prominent components of the peripheral visual system, i.e. the compound eyes and ocelli, respectively [38]. Opposite outcomes are observed as the consequence of

 doi:10.20944/preprints202309.0246.v1

misexpressing *ey* and *toy*, which famously causes the formation of extra-compound eye retinal tissues in antennae and leg appendages [39].

A third critical variable underlying proper *Pax* transcription factor function is defined by the level of expression, i.e. dosage. This variable can be conceived to be particularly relevant in the context of gene duplication. This is because the complete duplication of the cis-regulatory content of a gene locus is predicted to result in the duplication of overall expression levels in the absence of dosage-compensating mechanisms or preceding allelic differentiation. Considering these implications, it is of little surprise to find that most chelicerates possess only a single homolog of each *ey* and *toy* despite the duplication of gene content through one round of WGD in the early Arachnopulmonata (Figure 1) [8,40–42]. This outcome aligns with the assumption that dosage level increases of *ey* and *toy* are likely to result in deleterious fitness consequences, which were selected against by positive selection on post-WGD individuals with deletion mutations removing one sister paralog of each *ey* and *toy*.

Given the evidence of not only one but three WGD events in the early xiphosuran lineage (Figure 1) [16], the existence of higher numbers of *ey* and *toy* homologs in extant horseshoe crabs fits expectations. Indeed, the fact that none of the *ey* and *toy* paralogs are tandem linked but are, instead, spread out over different chromosomes or scaffolds is consistent with their origins in the wake of WGDs. Applying the terminology of ohnologs for paralogous genes that originated via WGD [43], it is safe to conclude that *toy1* and *toy2* represent ohnologs with the respect to each other and that the same is the case for *ey1*, *ey2*, and *ey3*.

And yet, given the significance of dosage for proper *Pax* gene function, the existence of three ohnologs of *ey* and two ohnologs of *toy* in the Xiphosura still constitutes a challenging conundrum. Their endurance in the genomes of extant horseshoe crabs raises two questions. The first is how the likely fitness-reducing effects of gene duplication on expression levels were resolved. The second question is which mechanisms explain the subsequent long-term conservation of the persisting paralogs. Of course, all the while posing these questions, the five xiphosuran *Pax6* genes also represent indirect evidence of the assumed fitness-reducing effects resulting from *Pax* homolog number increases. This is because the three WGD events in the early xiphosuran lineage must have spawned a total of eight ohnologs for each *ey* and *toy*, the majority of which were selected against to extinction.

From the copious research on *ey* and *toy* in insects and arachnids [6,40–42,44–46], we know that both ancient *Pax6* paralogs are critical factors in a large number of cell- and tissue-specification processes, i.e. that they are highly pleiotropic [47]. This background leads to the straightforward prediction that the xiphosuran *ey* and *toy* homologs diversified by inheriting non-overlapping components of the highly pleiotropic functional landscape of their ancestral singleton precursor homologs. This scenario is very likely the answer to the puzzling early finding of the presumed lack of *Pax6* homolog expression in the developing visual organs of *L. polyphemus*, which was based on the assumption of the exclusive presence of a singleton *Pax6* homolog [5]. Subsequent annotation efforts identified the *Pax6* homolog molecularly cloned by Blackburn et al. (2008) as the representative of the xiphosuran *toy1* homolog [8] (Table S1). At this point, it is reasonable to hypothesize that one of the *ey1*, *ey2*, or *ey3* paralogs is expressed in the early ocular segment and later in the peripheral visual organ precursor cell populations, while the other paralogs may be differentially involved in brain or insulin receptor specification [48]. It is further tempting to speculate that this is more likely the case for *ey1* or *ey3* given the lower strength of purifying selection on *ey2* based on the findings in our dN/dS analysis. Tissue expression studies, ideally conducted by embryonic whole-mount hybridization as conducted by Blackburn et al. (2008), are now warranted to probe these predictions. The second possibility that awaits testing by expression studies is that some of the five xiphosuran *Pax6* transcription factors engage redundantly in specific gene regulatory networks. This scenario is predicted by the deeply conserved redundant roles of *ey* and *toy* in the specification of the ocular segment during arthropod embryonic development [7].

The comparison with the functionalization fates of *ey* and *toy* in other arthropod lineages is critically informative to rank the exceptionality of the preservation and functionalization trajectories

of the five horseshoe crab *Pax6* genes. Similarly instructive is the comparison with the evolutionary fates of *Pax6* transcription factor homologs that resulted from the two rounds of WGD in the vertebrate stem lineage (Figure 1). In this group, three paralog descendants of the *Pax6* singleton ancestral precursor locus have remained conserved in select lineages, thereby defining what can be referred to as the vertebrate *Pax4/6/10* subfamily [49,50]. The existence of maximally three post-WGD descendants in the well-sampled vertebrate lineages implies the early extinction of at least one *Pax4/6/10* group member. In contemporary vertebrates, *Pax6* is most pervasively conserved while *Pax4* and *Pax10* are only preserved in a small number of taxa [49,50]. Combined, these data paint a similar picture of paralog losses and subfunctionalization in comparison to the fates of *ey* and *toy* paralogs after three rounds of WGD in the horseshoe crabs.

Our study complements previous analyses of the effects of the xiphosuran WGDs on other gene families, i.e. opsins [4], the Hox gene complex members [9,14,15], the components of the JAK-STAT signaling pathway [12], and micro RNAs [16,51]. Combined, these efforts provide ample testimony of the unexpectedly eventful genomic history of horseshoe crabs, which will continue to serve as an important paradigm case for genome evolution studies.

Looking ahead, the next obvious question is how the duplication and conservation histories of other *Pax* transcription factor subfamilies compare to the *Pax6* subgroup in the horseshoe crabs. Our ongoing analyses indicate a wide range of preservation outcomes with maximum subfamily sizes of up to seven members (Kassem et al., in preparation). We, therefore, predict that the horseshoe crab lineage will emerge as particularly useful to study the roles of constraints and opportunities in the evolutionary diversification of *Pax* transcription factors besides serving as an exceptionally well-defined sample case for understanding the complex consequences of WGD [52,53].

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Table S1: Genomic locations of xiphosuran Pax6 homologs; Table S2: dN/dS values; Text document S1: Protein sequence alignment; Text document S2: Annotated xiphosuran Pax6 loci; Text document S3: Protein sequences.

Author Contributions: Conceptualization, supervision, writing—review and editing, M.F; Investigation, data curation, and analysis, T.D.

Funding: This research received no external funding.

Data Availability Statement: All data are provided in the Supplementary Materials.

Acknowledgments: We thank the members of the Friedrich lab horseshoe crab genomics group for their comments and encouragement, and the reviewers for their time and diligence.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sharma, P.P. Chelicerates. *Curr. Biol.* **2018**, *28*, R774–R778.
2. Battelle, B.-A. The Eyes of Limulus Polyphemus (Xiphosura, Chelicerata) and Their Afferent and Efferent Projections. *Arthropod Struct. Dev.* **2006**, *35*, 261–274.
3. Kin, A.; Błażejowski, B. The Horseshoe Crab of the Genus Limulus: Living Fossil or Stabilomorph? *PLoS One* **2014**, *9*, e108036.
4. Battelle, B.-A.; Ryan, J.F.; Kempler, K.E.; Saraf, S.R.; Marten, C.E.; Warren, W.C.; Minx, P.J.; Montague, M.J.; Green, P.J.; Schmidt, S.A.; et al. Opsin Repertoire and Expression Patterns in Horseshoe Crabs: Evidence from the Genome of Limulus Polyphemus (Arthropoda: Chelicerata). *Genome Biol. Evol.* **2016**, *8*, 1571–1589.
5. Blackburn, D.C.; Conley, K.W.; Plachetzki, D.C.; Kempler, K.; Battelle, B.-A.; Brown, N.L. Isolation and Expression of Pax6 and Atonal Homologues in the American Horseshoe Crab, Limulus Polyphemus. *Dev. Dyn.* **2008**, *237*, 2209–2219.
6. Czerny, T.; Halder, G.; Kloter, U.; Souabni, A.; Gehring, W.J.; Busslinger, M. Twin of Eyeless, a Second Pax-6 Gene of Drosophila, Acts Upstream of Eyeless in the Control of Eye Development. *Mol. Cell* **1999**, *3*, 297–307.
7. Friedrich, M. Ancient Genetic Redundancy of Eyeless and Twin of Eyeless in the Arthropod Ocular Segment. *Dev. Biol.* **2017**, *432*, 192–200.
8. Friedrich, M. Coming into Clear Sight at Last: Ancestral and Derived Events during Chelicerate Visual System Development. *Bioessays* **2022**, *44*, e2200163.

9. Kenny, N.J.; Chan, K.W.; Nong, W.; Qu, Z.; Maeso, I.; Yip, H.Y.; Chan, T.F.; Kwan, H.S.; Holland, P.W.H.; Chu, K.H.; et al. Ancestral Whole-Genome Duplication in the Marine Chelicerate Horseshoe Crabs. *Heredity* 2016, *116*, 190–199.
10. Tanacredi, J.T.; Botton, M.L.; Smith, D. *Biology and Conservation of Horseshoe Crabs*; Springer Science & Business Media, 2009; ISBN 9780387899596.
11. Lamsdell, J.C. The Phylogeny and Systematics of Xiphosura. *PeerJ* **2020**.
12. Zhou, Y.; Liang, Y.; Yan, Q.; Zhang, L.; Chen, D.; Ruan, L.; Kong, Y.; Shi, H.; Chen, M.; Chen, J. The Draft Genome of Horseshoe Crab Tachypleus Tridentatus Reveals Its Evolutionary Scenario and Well-Developed Innate Immunity. *BMC Genomics* **2020**, *21*, 137.
13. Obst, M.; Faurby, S.; Bussarawit, S.; Funch, P. Molecular Phylogeny of Extant Horseshoe Crabs (Xiphosura, Limulidae) Indicates Paleogene Diversification of Asian Species. *Mol. Phylogenet. Evol.* **2012**, *62*, 21–26.
14. Shingate, P.; Ravi, V.; Prasad, A.; Tay, B.-H.; Garg, K.M.; Chattopadhyay, B.; Yap, L.-M.; Rheindt, F.E.; Venkatesh, B. Chromosome-Level Assembly of the Horseshoe Crab Genome Provides Insights into Its Genome Evolution. *Nat. Commun.* **2020**, *11*, 2322.
15. Shingate, P.; Ravi, V.; Prasad, A.; Tay, B.-H.; Venkatesh, B. Chromosome-Level Genome Assembly of the Coastal Horseshoe Crab (Tachypleus Gigas). *Mol. Ecol. Resour.* **2020**, *20*, 1748–1760.
16. Nong, W.; Qu, Z.; Li, Y.; Barton-Owen, T.; Wong, A.Y.P.; Yip, H.Y.; Lee, H.T.; Narayana, S.; Baril, T.; Swale, T.; et al. Horseshoe Crab Genomes Reveal the Evolution of Genes and microRNAs after Three Rounds of Whole Genome Duplication. *Commun Biol* **2021**, *4*, 83.
17. Simakov, O.; Marlétaz, F.; Yue, J.-X.; O'Connell, B.; Jenkins, J.; Brandt, A.; Calef, R.; Tung, C.-H.; Huang, T.-K.; Schmutz, J.; et al. Deeply Conserved Synteny Resolves Early Events in Vertebrate Evolution. *Nat Ecol Evol* **2020**, *4*, 820–830.
18. Ohno, S. Evolution by Gene Duplication. **1970**.
19. Sacerdot, C.; Louis, A.; Bon, C.; Berthelot, C.; Roest Crollius, H. Chromosome Evolution at the Origin of the Ancestral Vertebrate Genome. *Genome Biol.* **2018**, *19*, 166.
20. Nakatani, Y.; Shingate, P.; Ravi, V.; Pillai, N.E.; Prasad, A.; McLysaght, A.; Venkatesh, B. Reconstruction of Proto-Vertebrate, Proto-Cyclostome and Proto-Gnathostome Genomes Provides New Insights into Early Vertebrate Evolution. *Nat. Commun.* **2021**, *12*, 4489.
21. Schwager, E.E.; Sharma, P.P.; Clarke, T.; Leite, D.J.; Wierschin, T.; Pechmann, M.; Akiyama-Oda, Y.; Esposito, L.; Bechsgaard, J.; Bilde, T.; et al. The House Spider Genome Reveals an Ancient Whole-Genome Duplication during Arachnid Evolution. *BMC Biol.* **2017**, *15*, 62.
22. Birchler, J.A.; Yang, H. The Multiple Fates of Gene Duplications: Deletion, Hypofunctionalization, Subfunctionalization, Neofunctionalization, Dosage Balance Constraints, and Neutral Variation. *Plant Cell* **2022**, *34*, 2466–2474.
23. Kuzmin, E.; Taylor, J.S.; Boone, C. Retention of Duplicated Genes in Evolution. *Trends Genet.* **2022**, *38*, 59–72.
24. Weadick, C.J.; Chang, B.S.W. Complex Patterns of Divergence among Green-Sensitive (RH2a) African Cichlid Opsins Revealed by Clade Model Analyses. *BMC Evol. Biol.* **2012**, *12*, 206.
25. Des Marais, D.L.; Rausher, M.D. Escape from Adaptive Conflict after Duplication in an Anthocyanin Pathway Gene. *Nature* **2008**, *454*, 762–765.
26. Vavouri, T.; Semple, J.I.; Lehner, B. Widespread Conservation of Genetic Redundancy during a Billion Years of Eukaryotic Evolution. *Trends Genet.* **2008**, *24*, 485–488.
27. Nowak, M.A.; Boerlijst, M.C.; Cooke, J.; Smith, J.M. Evolution of Genetic Redundancy. *Nature* **1997**, *388*, 167–171.
28. Liberles, D.A.; Kolesov, G.; Dittmar, K. Understanding Gene Duplication through Biochemistry and Population Genetics. *Evolution after gene duplication* **2010**, 1–21.
29. Kimura, M. The Neutral Theory of Molecular Evolution. *Sci. Am.* **1979**, *241*, 98–100, 102, 108 passim.
30. Notredame, C.; Higgins, D.G.; Heringa, J. T-Coffee: A Novel Method for Fast and Accurate Multiple Sequence Alignment. *J. Mol. Biol.* **2000**, *302*, 205–217.
31. Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313.
32. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In Proceedings of the 2010 Gateway Computing Environments Workshop (GCE); IEEE, 2010; pp. 1–8.
33. Roth, C.; Liberles, D.A. A Systematic Search for Positive Selection in Higher Plants (Embryophytes). *BMC Plant Biol.* **2006**, *6*, 12.
34. Liberles, D.A. Evaluation of Methods for Determination of a Reconstructed History of Gene Sequence Evolution. *Mol. Biol. Evol.* **2001**, *18*, 2040–2047.
35. Carroll, S.B.; Grenier, J.K.; Weatherbee, S.D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*; John Wiley & Sons: Oxford, 2013;.

36. Chi, N.; Epstein, J.A. Getting Your Pax Straight: Pax Proteins in Development and Disease. *Trends Genet.* **2002**, *18*, 41–47.
37. Kronhamn, J.; Frei, E.; Daube, M.; Jiao, R.; Shi, Y.; Noll, M.; Rasmuson-Lestander, A. Headless Flies Produced by Mutations in the Paralogous Pax6 Genes Eyeless and Twin of Eyeless. *Development* **2002**, *129*, 1015–1026.
38. Blanco, J.; Pauli, T.; Seimiya, M.; Udolph, G.; Gehring, W.J. Genetic Interactions of Eyes Absent, Twin of Eyeless and Orthodenticle Regulate Sine Oculis Expression during Ocellar Development in Drosophila. *Dev. Biol.* **2010**, *344*, 1088–1099.
39. Halder, G.; Callaerts, P.; Gehring, W.J. Induction of Ectopic Eyes by Targeting Expression of the Eyeless Gene in Drosophila. *Science* **1995**, *267*, 1788.
40. Baudouin-Gonzalez, L.; Harper, A.; McGregor, A.P.; Sumner-Rooney, L. Regulation of Eye Determination and Regionalization in the Spider Parasteatoda Tepidariorum. *Cells* **2022**, *11*, doi:10.3390/cells11040631.
41. Schomburg, C.; Turetzek, N.; Schacht, M.I.; Schneider, J.; Kirfel, P.; Prpic, N.-M.; Posnien, N. Molecular Characterization and Embryonic Origin of the Eyes in the Common House Spider Parasteatoda Tepidariorum. *Evodevo* **2015**, *6*, 15.
42. Samadi, L.; Schmid, A.; Eriksson, B.J. Differential Expression of Retinal Determination Genes in the Principal and Secondary Eyes of Cupiennius Salei Keyserling (1877). *Evodevo* **2015**, *6*, 16.
43. Singh, P.P.; Isambert, H. OHNOLOGS v2: A Comprehensive Resource for the Genes Retained from Whole Genome Duplication in Vertebrates. *Nucleic Acids Res.* **2020**, *48*, D724–D730.
44. Yang, X.; Weber, M.; Zarinkamar, N.; Posnien, N.; Friedrich, F.; Wigand, B.; Beutel, R.; Damen, W.G.M.; Bucher, G.; Klingler, M.; et al. Probing the Drosophila Retinal Determination Gene Network in Tribolium (II): The Pax6 Genes Eyeless and Twin of Eyeless. *Dev. Biol.* **2009**, *333*, 215–227.
45. Luan, Q.; Chen, Q.; Friedrich, M. The Pax6 Genes Eyeless and Twin of Eyeless Are Required for Global Patterning of the Ocular Segment in the Tribolium Embryo. *Dev. Biol.* **2014**, *394*, 367–381.
46. Halder, G.; Callaerts, P.; Flister, S.; Walldorf, U.; Kloter, U.; Gehring, W.J. Eyeless Initiates the Expression of Both Sine Oculis and Eyes Absent during Drosophila Compound Eye Development. *Development* **1998**, *125*, 2181–2191.
47. Cvekl, A.; Callaerts, P. PAX6: 25th Anniversary and More to Learn. *Exp. Eye Res.* **2017**, *156*, 10–21.
48. Clements, J.; Hens, K.; Francis, C.; Schellens, A.; Callaerts, P. Conserved Role for the Drosophila Pax6 Homolog Eyeless in Differentiation and Function of Insulin-Producing Neurons. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 16183–16188.
49. Feiner, N.; Meyer, A.; Kuraku, S. Evolution of the Vertebrate Pax4/6 Class of Genes with Focus on Its Novel Member, the Pax10 Gene. *Genome Biol. Evol.* **2014**, *6*, 1635–1651.
50. Manousaki, T.; Feiner, N.; Begemann, G.; Meyer, A.; Kuraku, S. Co-Orthology of Pax4 and Pax6 to the Fly Eyeless Gene: Molecular Phylogenetic, Comparative Genomic, and Embryological Analyses. *Evol. Dev.* **2011**, *13*, 448–459.
51. Peterson, K.J.; Beavan, A.; Chabot, P.J.; McPeek, M.A.; Pisani, D.; Fromm, B.; Simakov, O. MicroRNAs as Indicators into the Causes and Consequences of Whole-Genome Duplication Events. *Mol. Biol. Evol.* **2022**, *39*, doi:10.1093/molbev/msab344.
52. Aase-Remedios, M.E.; Ferrier, D.E.K. Improved Understanding of the Role of Gene and Genome Duplications in Chordate Evolution With New Genome and Transcriptome Sequences. *Frontiers in Ecology and Evolution* **2021**, *9*, doi:10.3389/fevo.2021.703163.
53. Holland, L.Z.; Ocampo Daza, D. A New Look at an Old Question: When Did the Second Whole Genome Duplication Occur in Vertebrate Evolution? *Genome Biol.* 2018, *19*, 209.