

Article

Pattern of New Gene Origination in a Special Fish Lineage, the Flatfishes

Haorong Li [†] , Chunyan Chen [†], Zhongkai Wang, Kun Wang, Yongxin Li and Wen Wang ^{*}

School of Ecology and Environment, Northwestern Polytechnical University, Xi'an 710072, China; lainnorton@outlook.com (H.L.); chenchunyan@nwpu.edu.cn (C.C.); 15834186753@163.com (Z.W.); wangkun@nwpu.edu.cn (K.W.); yxli28science@sina.com (Y.L.)

^{*} Correspondence: wenwang@nwpu.edu.cn

[†] These authors contributed equally to this work.

Abstract: Origination of new genes are of inherent interest of evolutionary geneticists for decades, but few studies have addressed the general pattern in a fish lineage. Using our recent released whole genome data of flatfishes, which evolved one of the most specialized body plans in vertebrates, we identified 1541 (6.9% of the starry flounder genes) flatfish-lineage-specific genes. The origination pattern of these flatfish new genes is largely similar to those observed in other vertebrates, as shown by the proportion of DNA-mediated duplication (1317; 85.5%), RNA-mediated duplication (retrogenes; 96; 6.2%), and de novo-origination (128; 8.3%). The emergence rate of species-specific genes is 32.1 per Mya and the whole average level rate for the flatfish-lineage-specific genes is 20.9 per Mya. A large proportion (31.4%) of these new genes have been subjected to selection, in contrast to the 4.0% in primates, while the old genes remain quite similar (66.4% vs. 65.0%). In addition, most of these new genes (70.8%) are found to be expressed, indicating their functionality. This study not only presents one example of systematic new gene identification in a teleost taxon based on comprehensive phylogenomic data, but also shows that new genes may play roles in body planning.

Keywords: flatfishes; phylogenomic context; origin; new genes



Citation: Li, H.; Chen, C.; Wang, Z.; Wang, K.; Li, Y.; Wang, W. Pattern of New Gene Origination in a Special Fish Lineage, the Flatfishes. *Genes* **2021**, *12*, 1819. <https://doi.org/10.3390/genes12111819>

Academic Editor: Esther Betran

Received: 24 October 2021

Accepted: 17 November 2021

Published: 19 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

New genes may play significant roles in evolution of organisms [1,2]. The term “new gene” refers to a novel genetic unit that originated in a specific lineage [2]. As they only exist in certain species or evolutionary lineages, new genes are also called species or lineage-specific genes [3]. The main origin mechanisms of new genes include DNA-mediated duplication, RNA-mediated duplication and de novo origination, which can be identified from their sequence features [4]. With large number of genomes being sequenced, comparative genomic methods can be used to identify new genes based on the species phylogeny [5,6]. The synteny-based pipeline (SBP) method [7] is suited for recently duplicated genes' identification, whereas the protein-family-based methods are useful for ancient new genes [8,9]. The establishment of these method provides us an opportunity to understand the pattern of origin of new genes in a lineage.

So far, many studies on new genes have focused on the model organisms including primates, rodents and *Drosophila* [1,10–13]. However, there are few researches on new genes in non-model vertebrates, and no study using phylogenomic data has been conducted to investigate new genes in a lineage of teleosts [14]. In addition, teleosts usually have an additional genome-wide duplication event, compared to other vertebrates, which may result in more duplicate genes and some of them may have functions in these fishes [15,16]. Among teleosts, flatfishes have evolved a very unique body plan, especially for its asymmetric body structure has been of great interest [17,18]. Recently, a relatively complete set of genome data of flatfishes and their outgroup species were published by us [19,20],

providing a valuable opportunity to comprehensively identify new genes and investigate their possible roles in the evolution of flatfishes.

In this study, we defined the gene age using *Platichthys stellatus* (starry flounder) as the focal species representing the real flatfishes (Pleuronectoidei), which has the highest quality of genome assembly in flatfishes. Traditionally flatfishes include Pleuronectoidei and Psettodoidei species, but here we only consider the Pleuronectoidei, because Pleuronectoidei and Psettodoidei do not form a monophyletic group and Psettodoidei has only one species, *Psettodeserumei* [19]. We reconstructed the phylogenetic tree and estimated the divergence time in order to identify new genes by the SBP method. The results show that the flatfish-specific new genes have similar evolutionary pattern to those identified in primates, and perhaps due to longer divergence time (73.7 Mya) higher proportion of new genes of flatfishes were under functional constraint. The divergence time between the reference species (*P. stellatus*) and its closest species (*Paralichthys olivaceus*) is around 41.8 Mya. In addition, we observed some new genes were asymmetrically expressed in the two body sides during the metamorphosis of the starry flounder, indicating their possible roles in the development of special body plan of flatfishes.

2. Materials and Methods

2.1. Data Acquisition

Four published high-quality Pleuronectoidei genome assemblies, including starry flounder (*P. stellatus*, GCA_016801935.1), Japanese flounder (*P. olivaceus*, GCF_001970005.1), turbot (*Scophthalmus maximus*) and tongue sole (*Cynoglossus semilaevis*, GCF_000523025.1) [19], were collected to represent the species in Pleuronectoidei. The genome assembly of turbot was downloaded from the website (<https://denovo.cnag.cat/turbot>) [21]. The genome assembly of two Carangarians, which are the closest species of Pleuronectoidei, including sailfish (*Istiophorus platypterus*, GCA_016859345.1) and swordfish (*Xiphias gladius*, GCF_016859285.1) [20], and two close Perciformes species, yellow perch (*Perca flavescens*, GCF_004354835.1) [22] and Arkansas darter (*Etheostoma cragini*, GCF_013103735.1) [23], were retrieved from NCBI and used as outgroup species.

2.2. Phylogenetic Reconstruction

After collecting the protein-coding genes in the eight species, the longest transcript for each gene was retained. All-vs-all BLAST alignment was used to find the relationship of reciprocal best BLAST hit among genes, which were used to construct the phylogenetic tree by RAxML (v8.2.10) [24] with the PROTGAMMAAUTO model. The fourfold degenerate synonymous sites (4dTVs) were extracted and prepared from the genes to estimate the divergence time of species using the MCMCtree model in PAML (v4.4) [25] with the calibration of several fossil records downloaded from the TimeTree website (<http://www.timetree.org> accessed on 25 February 2021).

2.3. Identification of New Genes

Syntenic-based pipeline (SBP) [7] was used to identify the new genes of Pleuronectoidei. Considering the quality of genome assembly, we selected the chromosome-level starry flounder genome generated by the third-generation sequencing [18] as the reference in the subsequent analysis. The Lastz software (v1.04.00) [26] was used to align each collected genome to the reference genome. Parameters used in Lastz were set differently according to the divergence time of the query species with starry flounder, which refers to the different species alignment parameters with human in UCSC (<https://genome.ucsc.edu/index.html>, accessed on 25 February 2021). Then the reciprocal alignment results were generated following the steps in Zhang et al. [27]. The branch of new gene origination was inferred by the distribution of their position and existence in whole genome reciprocal alignment results. The age of each gene can be marked using branches 0–5. The genes that exist in all species on the phylogenetic tree were defined as the oldest genes and the branch is 0, while the branch 5 indicates genes only existing in the starry flounder. If there are

multiple exons or transcripts of a gene, the age of the oldest exon or transcript was used to represent the gene [7]. In addition, genes with more than 70% overlapping sequences with repetitive sequence regions or are not located on the chromosomes were excluded. To detect possible horizontal gene transfer, especially for those non-duplicated de novo genes, the NR (Non-Redundant Protein Sequence) database of NCBI was used to align flatfish new genes with the genes of microorganisms. None of the new genes is homologous to that of microorganisms, and for all of them homologous sequences can be found in the outgroup teleosts, suggesting that there was no horizontal gene transfer event happened in the origination of flatfish new genes.

2.4. Inference of the Origin Mechanism

The origin mechanisms of new genes were mainly classified into DNA-mediated duplicate genes, RNA-mediated duplicate genes (retrogenes), and de novo genes. The method in the previous studies [27,28] was used to analyze the origin mechanism of new genes. To further check the absence of homologous genes of de novo genes in outgroup species, NR database and Pfam databases were further checked.

2.5. Selection Analysis of New Genes

Selection signal of duplicate genes was calculated by comparing their closest paralogous genes in the starry flounder. The relationship between duplicate genes and their paralogous copies was obtained by BLAST (v2.2.26). The MUSCLE (v3.8.3) [29] software was used to align their protein sequences, and then pal2nal [30] was used to convert the results to codon level alignment. The PAML (v4.9e) [25] software was used to calculate the Ka/Ks value. The likelihood ratio (LRT) test was executed to calculate p -value with the assumption of $Ka/Ks = 0.5$ [31]. According to Betrán et al. [31], a paralogous Ka/Ks ratio significantly (by likelihood ratio test; LRT) lowers than 0.5 is considered as evidence for evolutionary constraint on both copies. The genes with $Ka > 0.5$, $Ks > 5$ or the Ks value beyond the 1.5-fold interquartile range in the Ks distribution were filtered. The candidate paralogous genes with outlier Ks values may not represent real paralogous genes. Duplicate genes with $Ka/Ks < 0.5$ and p -value < 0.05 were considered to be under negative selection.

We used the codeml of PAML4.9 [25] package to calculate the Ka/Ks values and performed the likelihood ratio test (LRT) with the assumption of $Ka/Ks = 1$ to test whether a de novo gene is functionally constrained by natural selection, we calculated the Ka/Ks ratio and p -value between a de novo gene and its homologous genes in Japanese flounder genome, if there exists. If $Ka/Ks < 1$ and p -value < 0.05 , respectively, they were considered to be under selective pressure. To do this, we first used BLAST (v2.2.26) to align de novo genes to Japanese flounder genome and extended 10 Kb of upstream and downstream of homologous regions. Next, we used protein sequence of de novo gene as template to predict open reading frame (ORF) of homologous sequences by Exonerate (v2.2.0) [32]. The longest ORF without stop codon or frame shifts was selected.

2.6. Gene Expression in Tissues of Starry Flounder

We downloaded the published multi-tissue RNA-sequencing (RNA-seq) data of starry flounder from NCBI, which have two replicate data for male and female brain, gonad, liver and muscle (PRJNA556158), respectively. We also downloaded RNA-seq data for the eyes and body tissues at the post-metamorphosis stage of starry flounder larvae (PRJNA592732), including the left-eye type's tissues around the eye, the left-eye type's around both eyes, the right-eye type's tissues around the eye and the right-eye type's around both eyes. Unfortunately, these RNA-seq data of starry flounder were not generated from left and right side of an individual, rather both side tissues from an individual were mixed, although left type and right type individuals were sampled separately. Therefore, these data cannot be used to look at the asymmetrical expression of new genes, and can be only used to check if a new gene was expressed or not. The adaptors and low-quality bases were removed with Trim-galore (v0.5.0) [33]. Hisat2 (v2.2.1) [34] was used to align the RNA-seq reads to

the starry flounder genome with default parameters. Then the Stringtie (v2.1.4) [35] was used to quantify the transcripts to obtain the fragments per kilobase million (FPKM) values and confirm the expression of new genes.

2.7. Asymmetrical Expression Analysis

The published multi-tissue RNA-seq data of Japanese flounder was downloaded from NCBI with accession number of PRJNA632737. There are 72 RNA-seq libraries which were produced by the method of orthogonal design, including four developmental stages around metamorphosis (pre-metamorphic larva; pro-metamorphic larva; metamorphic climax larva; post-metamorphic larva), three tissues (eye; muscle; skin), and two sides (left side; right side). Each sample was constructed and sequenced for three biological replicates. Hisat2 (v2.2.1) [34] was used to align the RNA-seq reads to the Japanese flounder genome with parameters—very-sensitive. The Stringtie (v2.1.4) [35] software was used to quantify the transcripts to obtain FPKM values for the 200 Pleuronectoidei-specific new genes (branch 2-4) which are also present in the Japanese flounder. The raw count values generated by Stringtie were used for the differential expression analysis of the left and right side data of each stages by the R package Deseq2 (v1.28.1) [36]. It is noteworthy that the genome of the Japanese flounder was assembled using short next generation sequencing reads [37] and, thus, was not good enough to be used as the reference genome when we were identifying new genes.

3. Results

3.1. New Genes Emerged in Pleuronectoidei

We re-constructed the phylogenetic tree and estimated the divergence time in order to identify new genes by the SBP method [7,27,28]. The divergence time between the reference species (*P. stellatus*) and its closest species is around 41.8 million year ago (Mya), and in the flatfishes with outgroups is 73.7 Mya. Then, there are 1541 Pleuronectoidei-specific new genes (assigned on branch 2–5; Supplementary Tables S2 and S3) were identified, which account for 6.9% of the genes that were located on chromosomes of the starry flounder (Figure 1). The species-specific genes emergence rate is 32.1 Mya and the whole average level rate for the Pleuronectoidei lineage genes is 20.9 per Mya. There are 1317 (85.5%) DNA-mediated duplicates, 96 (6.2%) RNA-mediated duplicates (retrogenes) and 128 (8.3%) de novo genes. We did not detect any new genes originated from horizontal gene transfer. The proportions of different categories of new genes are consistent with previous results in mammals [7,28,38]. Indicating the pattern of new gene origination in Pleuronectoidei is largely similar to other animal taxa [39] and that the ancient fish-specific genome duplication event may not retain many duplicates in the modern teleosts [39,40].

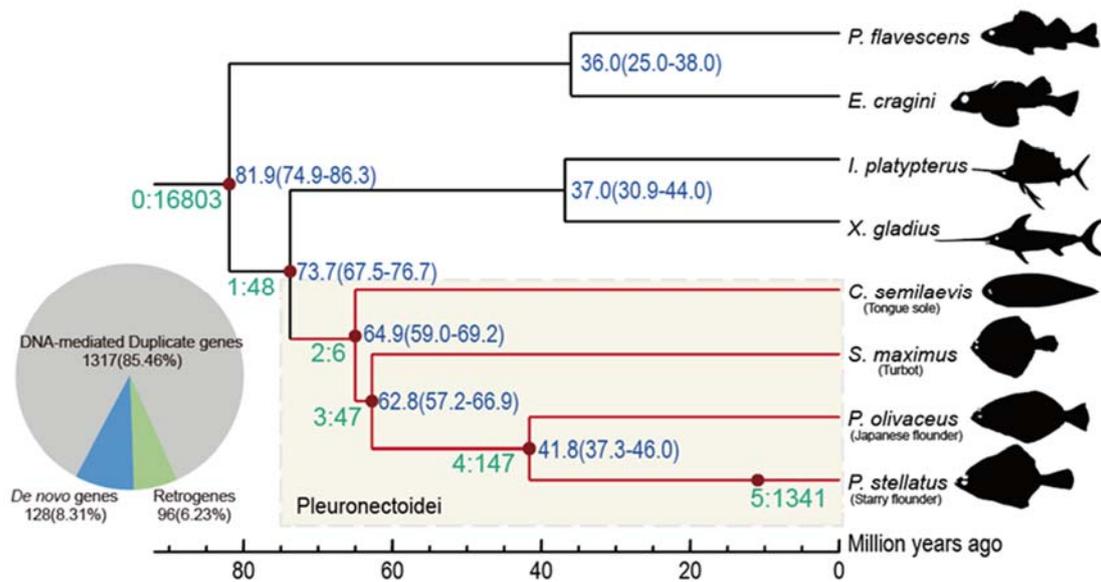


Figure 1. Branch view of the distribution of new genes. In SBP, a total of six distinct age groups (branch 0–5) were specified based on the phylogenetic context. Genes postdating *P. stellatus*-*C. semilaevis* split (branch 2–5) are referred as Pleuronectoidei-specific genes, marked with light yellow block. The number and proportion for each origin mechanism of the 1541 Pleuronectoidei-specific genes are plotted in a pie chart.

3.2. Most New Genes Are Expressed and Some under Natural Selection

About 74.0% (1046 out of all the 1413 duplicate genes, including both the DNA and RNA mediated duplications) of new duplicate genes are found to be expressed in at least one tissue (FPKM > 0.5; Figure 2B). After filter genes with outlier K_s , 308 of the 1413 duplicate genes can be conducted using paralog K_a/K_s analysis. Among the 308 genes, 128 (31.4%) genes were shown to be under negative selection ($K_a/K_s < 0.5$; p -value < 0.05) (Figure 2A), of which 101 genes were also expressed. In a previous study on primates [7], the proportion of old genes under negative selection was significantly higher than the new duplicate genes (65.0% in old genes and 4.0% in new genes). In our study, the proportion of old genes (branches 0–1) under negative selection was 66.4% (1299 genes under negative selection/1964 old genes), but the proportion of new duplicate genes under selection in our study is much higher, which may be explained by the deeper divergence of Pleuronectoidei (~73 Mya [15]) compared to the primates (~43 Mya [7]), or the new duplicate genes in Pleuronectoidei might have been under higher selection pressure.

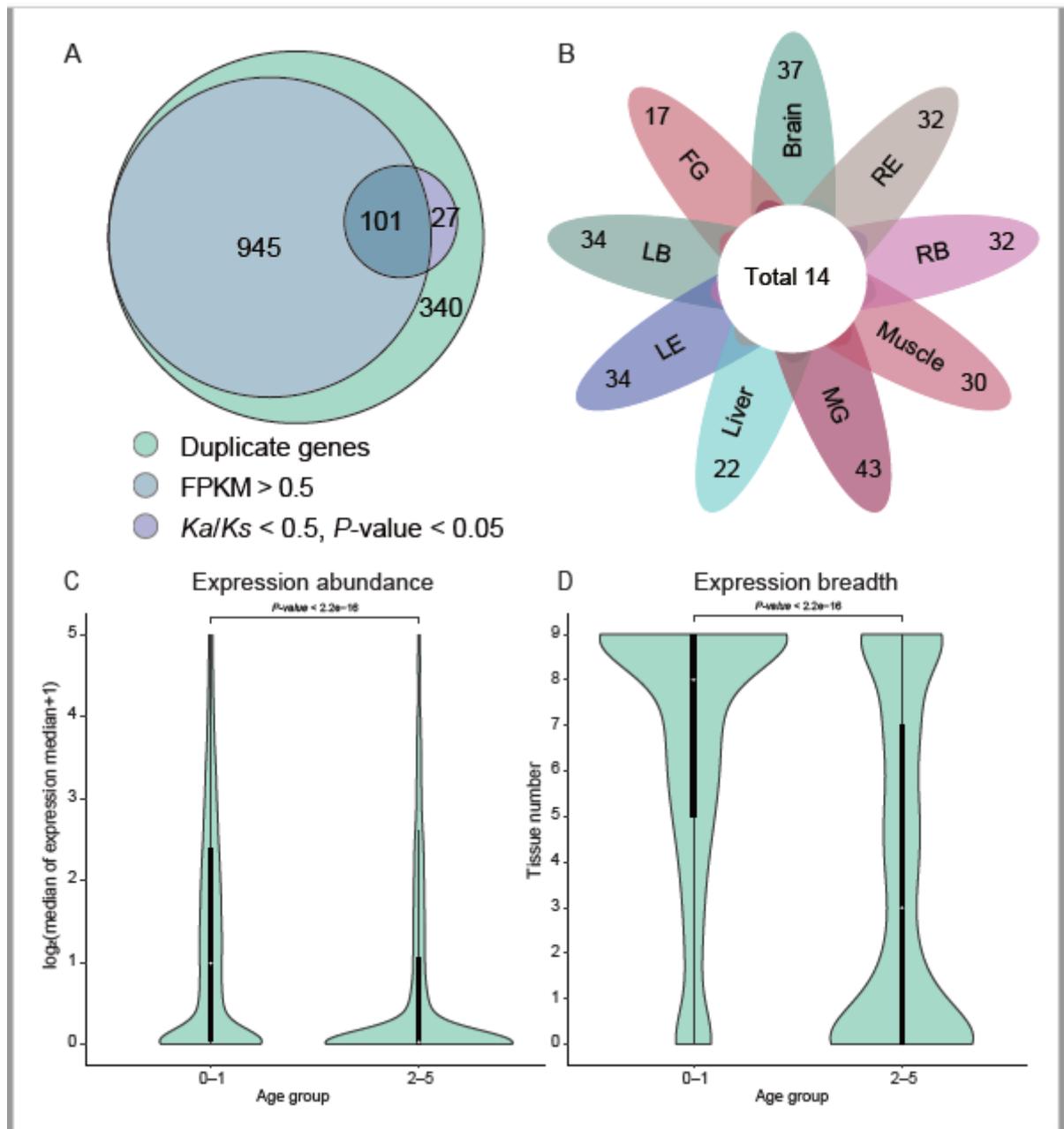


Figure 2. Expression of new genes. **(A)** Expression evidence of new genes. New genes with evidence of expression mean expressed in at least one tissue (FPKM > 0.5). Genes with significant paralog $Ka/Ks < 0.5$ are 128 (101 + 27). **(B)** Expression of de novo genes in different tissues. Each petal represents the number of genes expressed in the tissue (left-eye type's tissues around eyes (LB), left-eye type's two eyes (LE), right-eye type's tissues around eyes (RB), right-eye type's two eyes (RE), female gonad (FG), male gonad (MG)). Fourteen de novo genes are expressed in all the studied tissues. **(C,D)** show violin plots of expression abundance and breadth of Pleuronectoidei-specific genes across indicated age groups (log₂-based median of the median expression in 9 starry flounder tissues and the numbers of tissues where genes are expressed (FPKM > 0.5)). In each case, the violin curve indicates the probability density of the data, the black bar in the center shows the interquartile range and the white dot shows the median. The genes were divided into two age groups. Results of Wilcoxon tests of the significance of differences between the age groups are also presented.

There are 128 de novo genes in Pleuronectoidei, of which 45 genes were found to be expressed in the analysis of the transcriptome (FPKM > 0.5), accounting for 35.2% of the de novo genes in Pleuronectoidei. We counted the expression of all de novo genes in various tissues (Figure 2C), and found that the number of genes expressed in all tis-

sues is 14, the number of genes expressed in only a single tissue is 5, and the others expressed in two or more tissues. The expression of de novo genes varied in different tissues, with the highest number of 61 genes expressed in male gonad and lowest number of 19 genes expressed in female gonad. The result of more de novo genes expressed in male gonads is consistent with the observation in other animals by the previous study [7], but the specific functions of these de novo genes need further investigation. Among the 128 de novo genes, through Ka/Ks analysis with the orthologous genes in Japanese flounder, we found only two genes have been under significant negative selection, including gene *evm.model.Hic_chr_10.1034* ($Ka/Ks = 0.16$; p -value = 1.4×10^{-2}) and gene *evm.model.Hic_chr_16.310* ($Ka/Ks = 0.13$; p -value = 2.8×10^{-5}). Among them, the gene *evm.model.Hic_chr_10.1034* was expressed in all tissues (Figure S1), which might have evolved into a housekeeping gene in Pleuronectoidei.

By analyzing RNA-seq data in nine tissues of starry flounder, we found that gene transcription profiles changed across different gene age groups. The median expression of young genes (Pleuronectoidei-specific genes; branch 2–5) is close to 0 FPKM, while the median of nearly half of the old genes (branch 0–1) is higher than 1.0 FPKM (Figure 2C). In the expression profiles of various tissues, the young genes (branch 2–5) are expressed with a median less than 1.0 FPKM in 3 tissues, while the old gene (branch 0–1) are expressed with a median higher than 2.0 FPKM in 8 tissues. This age-related expression trend was also observed in the results of previous studies in primate- or rodent-specific genes [4,7].

3.3. Asymmetric Expression of Some New Genes May Play Roles in the Formation of the Unique Body Plan of Flatfishes

To investigate the possible roles of new genes on the asymmetric body plan of the Pleuronectoidei, unfortunately, we lacked two sides' transcriptome data of starry flounder from same individuals (see Methods). We thus tentatively used the transcriptome data from Japanese flounder for differential expression analysis for the 200 Pleuronectoidei-specific new genes which are also present in Japanese flounder (branch 2–4; Figure 3A). By comparing the expression level of left and right side in each stage, we found that the stage and tissue containing the most number of differentially expressed genes (DEGs) was the pro-metamorphic larva of muscle (Supplementary Figure S2). In other tissues, only one or zero new genes were significantly differentially expressed. In these DEGs at pro-metamorphic larva of muscle, nine new genes were significantly differentially expressed (p -value < 0.05, Table S1), eight genes are DNA-mediated new duplicate genes, and one gene is de novo gene (*evm.model.Hic_chr_6.523*) which has no functional annotation information yet.

Among the new genes with significantly higher expression on the right side, one gene is *Hipk1* (Figure 3B), which is related to the regulation of eyeball size, lens formation and retinal morphogenesis [41], and we found 14 genes that belonged to *Hipk* gene family in the Pleuronectoidei-specific new genes. The *Hipk* gene family were also found to have flatfish-specific expansion in previous study [37]. Besides, two new genes associated with cell proliferation were also highly expressed on the right side, including *Nlrc3-like* [42] (Figure 3C) and *Trim25* [43] (Figure 3D). Asymmetric expression of these three new genes may have potential functions in the formation of the asymmetry of the left and right sides of the Japanese flounder.

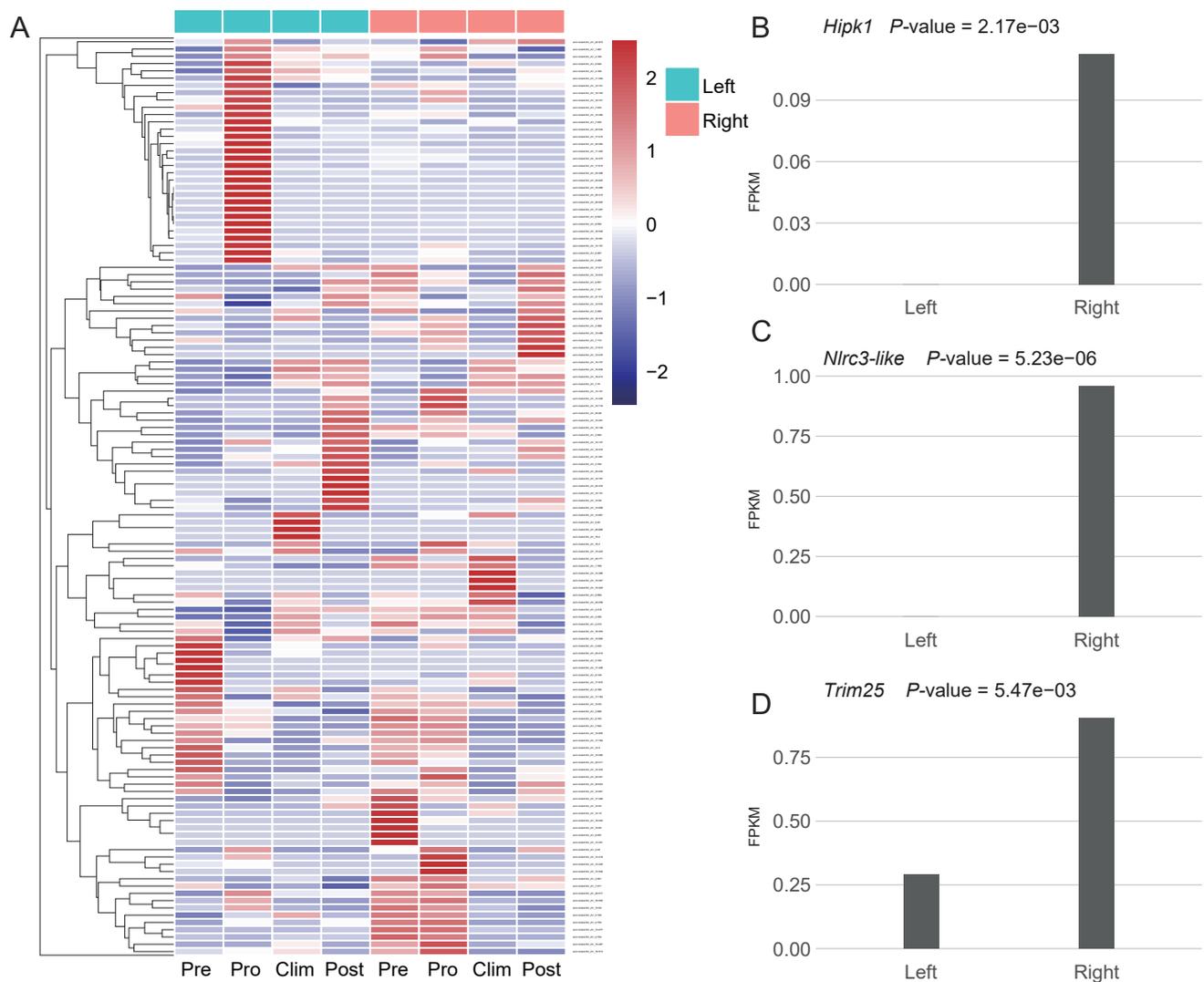


Figure 3. New genes expressed in Japanese flounder. (A) Heatmap of expression of 200 Pleuronectoidei-specific new genes which are present in Japanese flounder in left and right side muscle at different development stages, Pre, pre-metamorphic larva; Pro, pro-metamorphic larva; Clim, metamorphic climax larva; Post, post-metamorphic larva. Genes were clustered by hclust [44]. (B–D) Expression level and p -value of genes *Hipk1*, *Nlrc3-like* and *Trim25* at the pro-metamorphic larva of muscle.

4. Discussion

Our results show that origin of new genes in flatfishes shares similar pattern and mechanism as observed in other animals including mammals [7,28,45]. They originated mainly by DNA-mediated duplicates (85.5%), with some were from RNA-mediated duplicates (retrogenes) (6.2%) or de novo genes (8.3%), indicating that different animal taxa share similar patterns of new gene origination. In fishes, it is somewhat unexpected that the proportions of these categories of new genes are similar to previous results observed in mammals [7,28,38], suggesting that the ancient teleost-specific genome duplication event did not retain many duplicates in the modern teleosts.

In this study we found a high proportion (1341/1541 = 87%) of species-specific genes in the focal reference starry flounder, which corresponds to a new gene emergence rate of 32.1 genes per Mya, in contrast to the whole average level rate for the Pleuronectoidei lineage (20.9 genes per Mya), indicating rapid birth and death rate of genes in the flatfish lineage. A previous study has found a higher zebra fish-specific gene duplication rate (9.0 genes per Mya) than Tetraodon (2.6 genes per Mya) [39]. The authors explained this phenomenon as distant divergence between the two fishes, and especially that zebra fish

might have experienced more recent duplications [39]. In the case of the starry flounder, we also suspect that the long divergence time (41.8 Mya) between it and its closest species (the Japanese flounder) may partially lead to identification of more species-specific genes. However, it is possible that the higher genome assembly quality of the starry flounder than other flatfishes may artificially resulted in identification of some starry flounder-specific genes. Future studies with denser representative species and higher genome assemblies will clarify this problem.

It is interesting that the proportions of new genes under selection between Pleuronectoidei species and primates are highly different (31.4% vs. 4%) while the old genes remain quite similar (66.4% vs. 65%) [7]. This may have resulted from three possible reasons. First, flatfishes might have received stronger selection due to adaptation to the seafloor habitats. Second, the divergence time between Pleuronectoidei species and outgroup (73.7 Mya) is much earlier than that between primates and other mammals (43.0 Mya) [46], which thus allows us to see more selection-retained functional new genes in flatfishes. The third reason may be because that the focal reference species (the starry flounder) in this study diverged from its closest species (the Japanese flounder) 41.8 Mya, which is much earlier than that between human and chimpanzee (6.0 Mya), many species-specific new genes in the starry flounder have evolved substantial functions and thus are under selection. In other words, many new genes identified in the primate study [7] are very young human-specific genes that have not evolved substantial functions and may disappear in the future. We think the third reason is the most important factor that has led to the observation of much higher proportion of new genes under selection in flatfishes.

In our transcriptome analysis, our results show that the new genes in fishes also tend to express in the male gonad, further supporting the conjecture that new genes tend to evolve in male reproduction-related functions [2,7,47]. The male gonad expressed de novo gene *evm.model.Hic_chr_10.1034* is under negative selection, and can be partially (Identities = 24/95, 25%) aligned to deoxynucleotidyltransferase terminal-interacting protein 2 (*Dnttip2*). This gene may have relationship with chromatin remodeling and gene transcription [48]. This ubiquitously expressed de novo new gene deserves more functional studies in the future to demonstrate how a de novo gene can become a seemingly essential gene.

Although only limited transcriptome data available from left vs. right sides' tissues during the metamorphosis stages of a founder, we were still able to observe some new genes show asymmetric expression between the left and right tissues during metamorphic stages, which suggests they might be involved in the asymmetric body plan formation in flatfishes. In the future, collecting transcriptome data of left and right side of the metamorphosis stage from the reference species, the starry flounder, and especially more experimental validation work on these new genes, will provide more information to test the roles of these new genes in the asymmetric body plan of flatfishes.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes12111819/s1>, Table S1: New genes with significant differential expression between the left and right sides, Table S2: The branch and number of genes in starry flounder, Table S3: The list of the 1541 Pleuronectoidei-specific genes, Figure S1: Expression profiles of gene *Hic_chr_10.1034* in different tissues, Figure S2: The Principal Component Analysis (PCA) plot of muscle expression profiles of different samples of all genes in the Japanese flounder, including left and right side at different development stages. Pre, pre-metamorphic larva; Pro, pro-metamorphic larva; Clim, metamorphic climax larva; Post, post-metamorphic larva.

Author Contributions: Conceptualization, W.W. and H.L.; Methodology, C.C.; Writing—Original Draft Preparation, H.L., C.C.; Writing—Review & Editing, W.W., Y.L., C.C.; Visualization, Z.W.; Resources, Y.L., K.W.; Investigation, W.W., Y.L.; Funding acquisition W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Talents Team Construction Fund of Northwestern Polytechnical University (NWPU) to W.W.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. Zhang, Y.E.; Long, M.Y. New genes contribute to genetic and phenotypic novelties in human evolution. *Curr. Opin. Genet. Dev.* **2014**, *29*, 90–96. [[CrossRef](#)] [[PubMed](#)]
2. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Res.* **2010**, *20*, 1313–1326. [[CrossRef](#)] [[PubMed](#)]
3. Long, M.; Betrán, E.; Thornton, K.; Wang, W. The origin of new genes: Glimpses from the young and old. *Nat. Rev. Genet.* **2003**, *4*, 865–875. [[CrossRef](#)] [[PubMed](#)]
4. Zhang, Y.E.; Landback, P.; Vibranovski, M.; Long, M.Y. New genes expressed in human brains: Implications for annotating evolving genomes. *Bioessays* **2012**, *34*, 982–991. [[CrossRef](#)] [[PubMed](#)]
5. Cardoso-Moreira, M.; Long, M. The origin and evolution of new genes. *Methods Mol. Biol. Clifton N. J.* **2012**, *856*, 161–186. [[CrossRef](#)]
6. Ranz, J.M.; Parsch, J. Newly evolved genes: Moving from comparative genomics to functional studies in model systems. *Bioessays* **2012**, *34*, 477–483. [[CrossRef](#)]
7. Shao, Y.; Chen, C.Y.; Shen, H.; He, B.Z.; Yu, D.Q.; Jiang, S.; Zhao, S.L.; Gao, Z.Q.; Zhu, Z.L.; Chen, X.; et al. GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Res.* **2019**, *29*, 682–696. [[CrossRef](#)]
8. Domazet-Loso, T.; Tautz, D. Phylostratigraphic tracking of cancer genes suggests a link to the emergence of multicellularity in metazoa. *BMC Biol.* **2010**, *8*, 66. [[CrossRef](#)] [[PubMed](#)]
9. Capra, J.A.; Williams, A.G.; Pollard, K.S. ProteinHistorian: Tools for the comparative analysis of eukaryote protein origin. *PLoS Comput. Biol.* **2012**, *8*, e1002567. [[CrossRef](#)]
10. Villanueva-Canas, J.L.; Ruiz-Orera, J.; Agea, M.I.; Gallo, M.; Andreu, D.; Alba, M.M. New Genes and Functional Innovation in Mammals. *Genome Biol. Evol.* **2017**, *9*, 1886–1900. [[CrossRef](#)]
11. Huminiecki, L.; Wolfe, K.H. Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse. *Genome Res.* **2004**, *14*, 1870–1879. [[CrossRef](#)] [[PubMed](#)]
12. Zhang, W.; Landback, P.; Gschwend, A.R.; Shen, B.; Long, M. New genes drive the evolution of gene interaction networks in the human and mouse genomes. *Genome Biol.* **2015**, *16*, 202. [[CrossRef](#)]
13. Shiao, M.S.; Khil, P.; Camerini-Otero, R.D.; Shiroishi, T.; Moriwaki, K.; Yu, H.T.; Long, M. Origins of new male germ-line functions from X-derived autosomal retrogenes in the mouse. *Mol. Biol. Evol.* **2007**, *24*, 2242–2253. [[CrossRef](#)]
14. Cheng, Y.; Shang, D.; Luo, M.; Huang, C.; Lai, F.; Wang, X.; Xu, X.; Ying, R.; Wang, L.; Zhao, Y.; et al. Whole genome-wide chromosome fusion and new gene birth in the *Monopterus albus* genome. *Cell Biosci.* **2020**, *10*, 67. [[CrossRef](#)]
15. Meyer, A.; Scharl, M. Gene and genome duplications in vertebrates: The one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.* **1999**, *11*, 699–704. [[CrossRef](#)]
16. Zhang, Y.B.; Liu, T.K.; Jiang, J.; Shi, J.; Liu, Y.; Li, S.; Gui, J.F. Identification of a Novel Gig2 Gene Family Specific to Non-Amniote Vertebrates. *PLoS ONE* **2013**, *8*, 14. [[CrossRef](#)]
17. Friedman, M. The evolutionary origin of flatfish asymmetry. *Nature* **2008**, *454*, 209–212. [[CrossRef](#)] [[PubMed](#)]
18. Schreiber, A.M. Flatfish: An asymmetric perspective on metamorphosis. *Curr. Top. Dev. Biol.* **2013**, *103*, 167–194. [[CrossRef](#)]
19. Lu, Z.M.; Gong, L.; Ren, Y.D.; Chen, Y.J.; Wang, Z.K.; Liu, L.Q.; Li, H.R.; Chen, X.Q.; Li, Z.Z.; Luo, H.R.; et al. Large-scale sequencing of flatfish genomes provides insights into the polyphyletic origin of their specialized body plan. *Nat. Genet.* **2021**, *53*, 742–751. [[CrossRef](#)]
20. Wu, B.; Feng, C.; Zhu, C.; Xu, W.; Yuan, Y.; Hu, M.; Yuan, K.; Li, Y.; Ren, Y.; Zhou, Y.; et al. The Genomes of Two Billfishes Provide Insights into the Evolution of Endothermy in Teleosts. *Mol. Biol. Evol.* **2021**, *38*, 2413–2427. [[CrossRef](#)]
21. Figueras, A.; Robledo, D.; Corvelo, A.; Hermida, M.; Pereiro, P.; Rubiolo, J.A.; Gómez-Garrido, J.; Carreté, L.; Bello, X.; Gut, M.; et al. Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): A fish adapted to demersal life. *DNA Res.* **2016**, *23*, 181–192. [[CrossRef](#)] [[PubMed](#)]
22. Feron, R.; Zahm, M.; Cabau, C.; Klopp, C.; Roques, C.; Bouchez, O.; Eché, C.; Valière, S.; Donnadiou, C.; Haffray, P.; et al. Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca flavescens*. *Mol. Ecol. Resour.* **2020**, *20*, 531–543. [[CrossRef](#)] [[PubMed](#)]
23. Reid, B.N.; Moran, R.L.; Kopack, C.J.; Fitzpatrick, S.W. Rapture-ready darters: Choice of reference genome and genotyping method (whole-genome or sequence capture) influence population genomic inference in *Etheostoma*. *Mol. Ecol. Resour.* **2021**, *21*, 404–420. [[CrossRef](#)] [[PubMed](#)]
24. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]
25. Yang, Z.H. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)]
26. Harris, R.S. Improved Pairwise Alignment of Genomic DNA. Ph.D. Thesis, The Pennsylvania State University, State College, PA, USA, 2007.

27. Zhang, Y.E.; Vibranovski, M.D.; Krinsky, B.H.; Long, M.Y. Age-dependent chromosomal distribution of male-biased genes in *Drosophila*. *Genome Res.* **2010**, *20*, 1526–1533. [CrossRef]
28. Zhang, Y.E.; Vibranovski, M.D.; Landback, P.; Marais, G.A.; Long, M. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS Biol.* **2010**, *8*, e1000494. [CrossRef]
29. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]
30. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **2006**, *34*, W609–W612. [CrossRef]
31. Betran, E.; Thornton, K.; Long, M. Retroposed new genes out of the X in *Drosophila*. *Genome Res.* **2002**, *12*, 1854–1859. [CrossRef]
32. Slater, G.S.; Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* **2005**, *6*, 31. [CrossRef] [PubMed]
33. Kechin, A.; Boyarskikh, U.; Kel, A.; Filipenko, M. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *J. Comput. Biol.* **2017**, *24*, 1138–1143. [CrossRef] [PubMed]
34. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [CrossRef] [PubMed]
35. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [CrossRef]
36. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 38. [CrossRef]
37. Shao, C.; Bao, B.; Xie, Z.; Chen, X.; Li, B.; Jia, X.; Yao, Q.; Ortí, G.; Li, W.; Li, X.; et al. The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat. Genet.* **2017**, *49*, 119–124. [CrossRef]
38. Magadum, S.; Banerjee, U.; Murugan, P.; Gangapur, D.; Ravikesavan, R. Gene duplication as a major force in evolution. *J. Genet.* **2013**, *92*, 155–161. [CrossRef]
39. Blomme, T.; Vandepoele, K.; De Bodt, S.; Simillion, C.; Maere, S.; Van de Peer, Y. The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* **2006**, *7*, 12. [CrossRef]
40. Lu, J.; Peatman, E.; Tang, H.; Lewis, J.; Liu, Z. Profiling of gene duplication patterns of sequenced teleost genomes: Evidence for rapid lineage-specific genome expansion mediated by recent tandem duplications. *BMC Genom.* **2012**, *13*, 246. [CrossRef]
41. Inoue, T.; Kagawa, T.; Inoue-Mochita, M.; Isono, K.; Ohtsu, N.; Nobuhisa, I.; Fukushima, M.; Tanihara, H.; Taga, T. Involvement of the Hipk family in regulation of eyeball size, lens formation and retinal morphogenesis. *FEBS Lett.* **2010**, *584*, 3233–3238. [CrossRef]
42. Zha, L.H.; Zhou, J.; Tan, Y.; Guo, S.; Zhang, M.Q.; Li, S.; Yan, P.; Yu, Z.X. NLRC3 inhibits PDGF-induced PSMCs proliferation via PI3K-mTOR pathway. *J. Cell. Physiol.* **2020**, *235*, 9557–9567. [CrossRef] [PubMed]
43. Takayama, K.I.; Suzuki, T.; Tanaka, T.; Fujimura, T.; Takahashi, S.; Urano, T.; Ikeda, K.; Inoue, S. TRIM25 enhances cell growth and cell survival by modulating p53 signals via interaction with G3BP2 in prostate cancer. *Oncogene* **2018**, *37*, 2165–2180. [CrossRef] [PubMed]
44. Kolde, R. Pheatmap: Pretty Heatmaps. 2015. Available online: <https://rdrr.io/cran/pheatmap/> (accessed on 25 February 2021).
45. Zhou, Q.; Zhang, G.; Zhang, Y.; Xu, S.; Zhao, R.; Zhan, Z.; Li, X.; Ding, Y.; Yang, S.; Wang, W. On the origin of new genes in *Drosophila*. *Genome Res.* **2008**, *18*, 1446–1455. [CrossRef] [PubMed]
46. Schneider, H.; Sampaio, I. The systematics and evolution of New World primates—A review. *Mol. Phylog. Evol.* **2015**, *82*, 348–357. [CrossRef]
47. Vinckenbosch, N.; Dupanloup, I.; Kaessmann, H. Evolutionary fate of retroposed gene copies in the human genome. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3220–3225. [CrossRef]
48. Edmiston, J.S.; Archer, K.J.; Scian, M.J.; Joyce, A.R.; Zedler, B.K.; Murrelle, E.L. Gene expression profiling of peripheral blood leukocytes identifies potential novel biomarkers of chronic obstructive pulmonary disease in current and former smokers. *Biomarkers* **2010**, *15*, 715–730. [CrossRef]