Chance and Necessity: Emerging Introns in Intronless Retrogenes

Shengjun Tan, *State Key Laboratory of Integrated Management of Pest Insects and Rodents and Key Laboratory of the Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China*

Zhenglin Zhu, *School of Life Sciences, Chongqing University, Chongqing, China*

Tao Zhu, *MOE Key Laboratory for Biodiversity Science and Ecological Engineering and Beijing Key Laboratory of Gene Resource and Molecular Development, College of Life Sciences, Beijing Normal University, Beijing, China*

Rigen Te, *State Key Laboratory of Integrated Management of Pest Insects and Rodents and Key Laboratory of the Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China*

Yong E Zhang, *State Key Laboratory of Integrated Management of Pest Insects and Rodents and Key Laboratory of the Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China*

Advanced article

Article Contents

Online posting date: 15th August 2014

Retrogenes are duplicated genes generated via retroposition, which were conventionally believed to contain no introns. However, emerging data showed that a significant number of retrogenes do have introns. Thus, these genes represent an attractive system to study how new genes evolve exon–intron structure. Comparison between parental genes and retrogenes revealed that retrogenes mainly evolve chimeric structures by fusing with local host genes or recruiting pre-existing intergenic sequences. Additionally, retrogenes could gain introns by inheriting introns of parental genes or by transforming parental exonic sequences. The functional necessity on intron gain in retrogenes remains largely elusive although limited data suggest that newborn introns play regulatory roles, enable exon shuffling and alternative splicing. Accumulation of population genomic data may help to understand which evolutionary force shapes the fixation of introns in both retrogenes and *de novo* originated genes given the same intron birth process acts on both type of new genes.

eLS subject area: Evolution & Diversity of Life

How to cite:
Tan, Shengjun; Zhu, Zhenglin; Zhu, Tao; Te, Rigen; and Zhang, Yong E (August 2014) Chance and Necessity: Emerging Introns in Intronless Retrogenes. In: eLS. John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0022886

Introduction

New gene origination attracted the interest of biologists back to the 1930s (Muller, 1936) and emerging evidence in the era of genomics showed that new gene acts as one major driver of phenotypic evolution (Chen *et al*., 2013). Among various mechanisms leading to the birth of new genes, retroposition or reverse transcription of messenger RNA (mRNA) of parental genes is special in that the loss of introns could serve as a natural hallmark to differentiate old parental genes and newly derived retrogenes (Brosius, 2003; Cardoso-Moreira and Long, 2012). This is possibly the most important reason why many early characterised new genes such as *jingwei* (Long and Langley, 1993) or *sphinx* (Wang *et al*., 2002) are all retroposed duplicated genes or retrogenes. **See also**: Processed Pseudogenes and Their Functional Resurrection in the Human and Mouse Genomes; Pseudogenes and Their Evolution

Since retrogenes were generally believed to be intronless, screening of paralogous pairs including both multi-exonic gene and single-exon gene becomes a standard routine in the identification of retrogenes (Meisel *et al*., 2009; Zhang *et al*., 2011). However, after carefully examining the structure of retrogenes, numerous reports (e.g. Fablet *et al*., 2009; Zhu *et al*., 2009; Zhang *et al*., 2014) discovered a surprising fact that an appreciable proportion of retrogenes do have introns. Considering that most extensively studied retrogenes are often quite young, the origination mechanism of newly evolved introns could be reliably inferred because sequence features surrounding them do

not have enough time to degenerate (Fablet *et al*., 2009; Zhu *et al*., 2009; Zhang *et al*., 2014).

In a bigger picture, the timing, mechanism and cause of intron evolution is a long-term active research area in the whole field of molecular evolution given the logical relationship between intron and multiple essential concepts (e.g. exon shuffling) (William Roy and Gilbert, 2006). The related reports often exclusively focused on relatively old proteins, for example, proteins emerging before the split of human and mouse, and found that coding regions tend to lose but not gain introns (Roy *et al*., 2003). From this aspect, the emergence of introns in both coding sequence (CDS) and untranslated region (UTR) of retrogenes will not only help to understand the evolution of new gene structure, but also provide novel insight into the overall topic of intron evolution. **See also**: Intron Loss and Gain; Introns: Movements

In this review, the authors began with a brief introduction on how introns can be created in general and why they get originated. Then, they focused on how introns arose mechanistically in retrogenes. After that, they explained the cause of intron evolution and discussed how random factors and the functional reasons acted in the process of intron birth. The review ended with another type of new genes, *de novo* genes or genes originated from previously noncoding deoxyribonucleic acid (DNA) (Levine *et al*., 2006), given they appeared to be subject to similar intron birth process like retrogenes. **See also**: Evolutionary Origin of Orphan Genes

General Mechanisms and Causes of Intron Gain

A plenty of mechanistic models have been proposed to account for the origination of introns including intron transposition, intronization, transposon insertion and so on (Roy, 2004; William Roy and Gilbert, 2006; Irimia *et al*., 2008; Yenerall and Zhou, 2012). Among these models, intron transposition appears to be the most related mechanism to retroposition. Specifically, intron transposition occurs when an intron spliced out of the transcript is reverse spliced into another mRNA or its own mRNA (Sharp, 1985). Then, this mRNA with the new intron is reverse transcribed into an intron-containing complementary DNA (cDNA) (similar to the generation of retrogenes), and transfers the new intron to another gene by gene conversion (Sharp, 1985). Thus, compared to the origination of retrogenes which removed all introns by taking mRNA as the template of duplication, intron transposition gains introns by additional reverse splicing stage.

Introns might emerge under the interaction of evolutionary force and functional necessity. On the one hand, eukaryote species especially those with small effective population size could accumulate introns under genetic drift (Lynch, 2007; Li *et al*., 2009). In this model, introns are assumed to have deleterious effects, for example, delay in mature mRNA production. On the other hand, the birth of introns could result in beneficial effects such as removal of premature termination codons (PTC) (Catania and Lynch, 2008), enhancement of transcription efficiency or transcript stability (Wang *et al*., 2007; Rose *et al*., 2008; Catania and Lynch, 2013) and providing new protein products by alternative splicing (Barrett *et al*., 2012).

Retrogenes Evolve New Exon–Intron Structures Mainly through Chimerism

In the context of retrogene study, gene structure change between retrogene and the corresponding parental gene is routinely surveyed. This is different from general studies on intron evolution, which often exclusively relied on the orthologous alignments of relatively old proteins (e.g. Roy *et al*., 2003). This unique angle leads to the discovery of a novel mechanism where retrogenes often evolve new exon–intron structures by recruiting flanking and previously intergenic sequences or fusing with host genes (Vinckenbosch *et al*., 2006; Fablet *et al*., 2009). The reason for retrogenes to evolve such chimeric structure is possibly the loss of original regulatory sequences during the retroposition process. Indeed, Vinckenbosch *et al*. (2006) found a significant excess of transcribed retrocopies near other genes or within introns in the human genome, demonstrating that it is an effective way to prevent the fate of ''death on arrival'' (Brosius, 2005).

As one of the first identified young genes, *jingwei* in *Drosophila* represented a type of chimerism mediated by gene fusion (Long *et al*., 2003). Specifically, an *Adh* mRNA molecule was retroposed into the third intron of *yande* (**Figure 1a**). Combined with the first three coding exons of *yande*, the retrogene evolved into a new gene *jingwei* that is translated as a chimeric protein. The downstream region of *yande* degenerated afterwards. Thus, this chimeric retrogene showcased an effective way of exon shuffling, that is, evolving new functions by combining the domains from two genes (Gilbert, 1978). Certainly, retrogenes could also evolve by fusing with the UTR region of their host genes. For example, *MIP-2A* was retroposed to the first intron of the host gene, *ZNF547*, and then fused to the 5′ UTR exon of the later gene (Vinckenbosch *et al*., 2006). Different from fusing with a pre-existing gene, retrogene can also recruit nearby intergenic sequences and form a new chimeric gene structure with *de novo* function. For example, *sphinx* (Wang *et al*., 2002) and *PMCHL1* (Courseaux and Nahon, 2001) recruited a group of exons and introns into their transcription units from upstream (**Figure 1c**) or downstream region, respectively.

In the case of fusion, the original intronic structure of host gene certainly facilitates the origination of new chimeric introns. The exact mechanistic details on how recruitment occurs remain elusive. This process may be

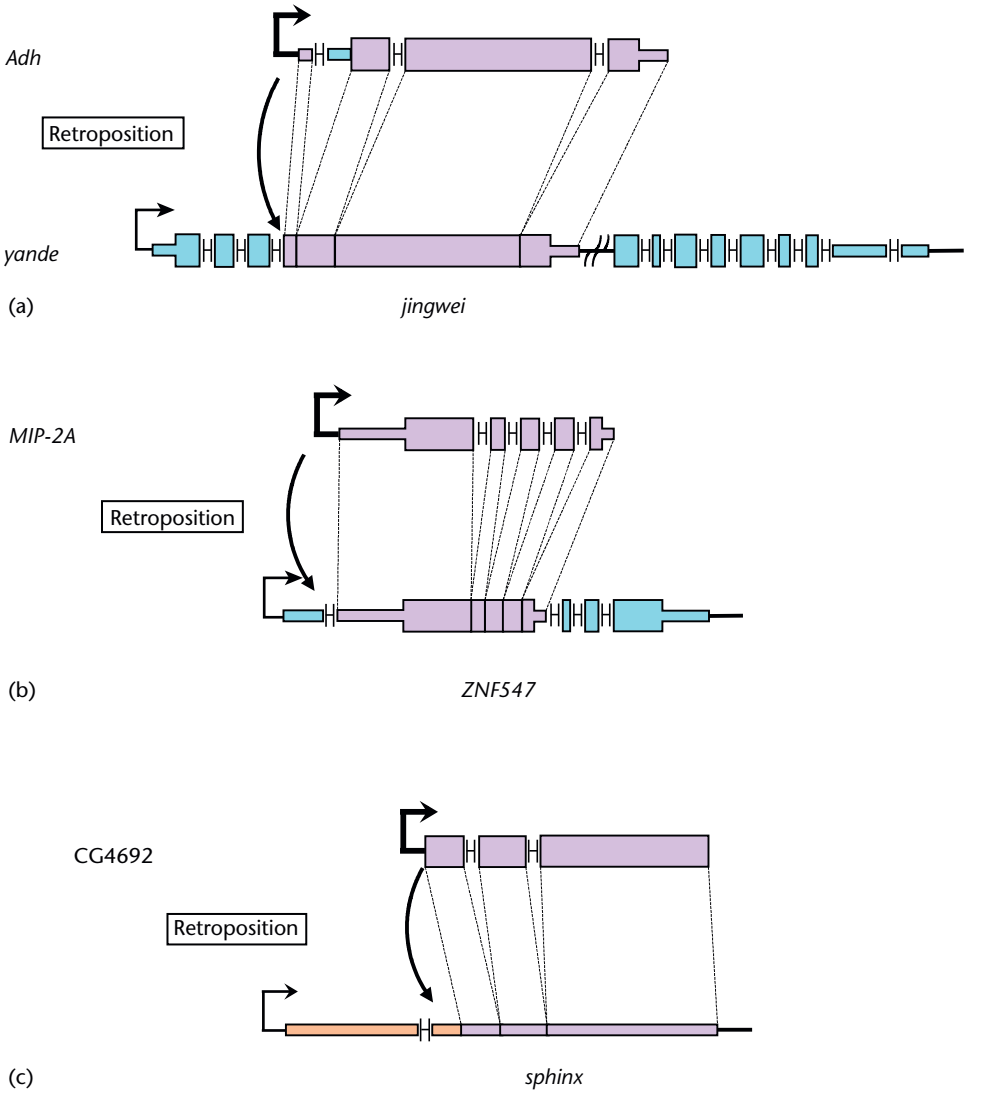

Figure 1 How chimerism occurs. The top and bottom part indicate parental gene and retrogene, respectively. Thicker boxes represent coding exons, while thinner boxes represent UTRs. 'H'-like tags represent introns. The retroposed regions are marked in purple, while other regions are marked in blue (genic) or orange (intergenic region). The sequence correspondence between parental and retrogene is marked with dotted lines. Semi-rectangle lines with arrows indicate the direction of transcription. (a) The retrogene *jingwei* was fused with the neighbouring gene *yande.* The other region of *yande,* including nine exons and nine introns is degenerated. (b) The retrogene was inserted into the intron between UTR and coding exon, and fused with 5′ UTR later. (c) The noncoding gene *sphinx* recruits two exons and one intron from the 5′ flanking intergenic region.

largely *de novo* although a simple proto-splice site may exist before retroposition (Fablet *et al*., 2009).

Chimeric Retrogenes are Widespread in both Plants and Animals

Fast accumulation of sequencing data enabled genome-wide identification of chimeric retrogenes in multiple species (**Table 1**). A pioneering effort in rice found 380 retrogenes (42%) recruiting new exons from flanking regions, 73 of them could be verified by either full-length cDNA or expressed sequence tag (EST) evidences (Wang *et al*., 2006). A significant proportion of chimeric genes were also found in other plants: 13% (12/90) retrogenes in *Populus* and 23% (19/83) in *Arabidopsis* (Zhu *et al*., 2009).

Analogously, genome-wide surveys of chimeric retrogenes were performed in animals. First, 29 out of 3590 human retrocopies including both retrogenes and processed pseudogenes were demonstrated to recruit introns, which were proved by both full-length mRNA sequences and spliced ESTs (Fablet *et al*., 2009). Most (26/29) introns are embedded in the 5′ UTRs and the majority (25/29) have evolved new 5′ exon–intron structures by recruiting flanking intergenic sequences. With relatively relaxed identification

Table 1 An overview of the numbers of retrocopies with introns in the eukaryotic genomes

Mechanism	No. of retrocopies[a] with introns	No. of retrocopies	Species	References
Chimerism	380	1235	Rice	Wang *et al*. (2006)
	8	150	Rice	Sakai *et al*. (2011)
	12	106	*Populus*	Zhu *et al*. (2009)
	19	83	*Arabidopsis*	Zhu *et al*. (2009)
	28[b]	3590	Human	Fablet *et al*. (2009)
	95	652	Zebrafish	Fu *et al*. (2010)
	9	195	Zebrafish	Chen *et al*. (2011)
	16	398	Western clawed frog	Chen *et al*. (2011)
Intronization	11	106	*Populus*	Zhu *et al*. (2009)
	18	83	*Arabidopsis*	Zhu *et al*. (2009)
	10[c]	3436	Human	Szcześniak *et al*. (2011), Kang *et al*. (2012)

[a]We used the term 'retrocopy' rather than 'retrogene' to refer to all retroposed sequence including both functional retrogene and processed pseudogene.
[b]29 Retrogenes were found to recruite new intron, among which 28 were involved in chimerism.
[c]Two genes were found by Szcześniak *et al*. (2011) and the other eight by Kang *et al*. (2012).

parameters, more than 12% human retrogenes were found to evolve introns, which situate in the 5′ UTR, 3′ UTR or CDS region (Baertsch *et al*., 2008).

The uncertainty of the actual proportion of chimeric retrogenes also exists in other system. Taking zebrafish as an example, 14.6% (95/652) of retrogenes appear to be chimeric with most of them (84/95) recruiting coding exons along with introns. However, one thing should be noted that only 4.6% (9/195, **Table 1**) chimeric retrocopies in the zebrafish genome were detected in another study (Chen *et al*., 2011). All these inconsistency may root in the difference of the underlying *in silico* pipelines, for example, how retrogenes were identified, how exon–intron structures were estimated, etc. Both software development and gene annotation refinement is needed to generate a consensus view of how pervasive chimerism is for retrogenes. The later point is especially noteworthy considering the general low quality of young gene annotation (Zhang *et al*., 2012).

Inheritance of Parental Introns

Parental gene and retrogene comparison also revealed that retrogene could directly inherit introns from their parental genes. One early study identified two *Arabidopsis* retrogenes (AT1G63210 and AT5G56720) with each sharing one intron from their corresponding parental genes (Zhang *et al*., 2005). In mouse, the preproinsulin I gene was retroposed from the ancestral preproinsulin II gene but kept one parental intron in the 5′ UTR (**Figure 2a**; Katju, 2012). A recent survey extracted 0, 7, 12 and 6 similar cases in fruitfly, human, rice and *Arabidopsis*, respectively, by comparing the location of intron sites of parental and retroposed genes (Zhang *et al*., 2014). For example, retrogene LOC_Os05g39720.1 retained one intron from its parental gene LOC_Os01g61080.1 (**Figure 2b**). At least three possible mechanisms can explain this phenomenon. First, retrogenes represent alternative isoforms with introns encoded by parental loci. Second, the parental mRNA is only partially processed and some introns have not been spliced out when the following reverse transcription process happened (de Boer *et al*., 2014). In other words, the transcript was 'caught in the act' of splicing. Finally, the intron-containing parental gene partially recombines with the intronless retrocopy and transfers one or several introns to the retrogene, similar to the intron transfer mechanism (Yenerall and Zhou, 2012).

Differentiation of these three possibilities appears feasible in theory. The existence of parental transcripts encoding intron retention form is consistent with the first model, which is shown by cases in (Zhang *et al*., 2014). Partial splicing may favour intron retention in the 3′ terminal given the splicing order from 5′ to 3′, whereas recombination mediated mechanism will leave a long tract of high sequence similarity between parental and retroposed genes. The relative contribution of these three mechanisms remains unknown.

Intronization

Apart from the aforementioned mechanisms of intron gain in the context of new gene evolution, that is, chimerism and inheritance, intronization also contributed to origination of new introns in retrogenes. Intronization, referring to a process that some genes convert internal exonic sequences into introns, was first observed by comparing orthologous genes across *Caenorhabditis* species (Irimia *et al*., 2008). Later it was indicated that retrogene could gain new introns via the same mechanism.

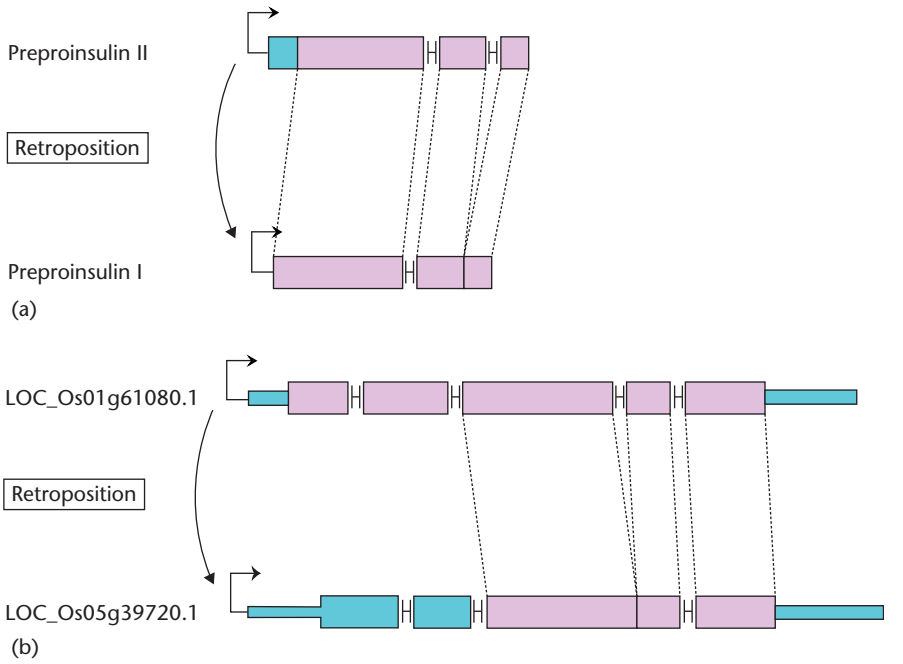

Figure 2 How intron inheritance occurs. The figure convention follows **Figure 1**. In case of preproinsulin I (a) and LOC_Os05g39720.1 (b) one intron appeared to be inherited from the corresponding parental gene, respectively. For LOC_Os05g39720.1, this retrocopy was also fused with the flanking region to form a chimeric gene.

In plants, 29 retrogenes were observed to undergo intronization (**Table 1**). As shown in **Figure 3a**, a previously exonic region of parental gene AT1G66860 in *Arabidopsis* turned into a new intron in the retrogene AT1G15040 (Zhu *et al*., 2009). This was induced by *de novo* point mutations from AC to GT, which generated the splicing donor site of the new intron. The splicing acceptor site was shared between the retrogene and its corresponding parental gene. Intronization in retrogenes facilitated by cryptic splicing sites were reported in 10 cases of human (Szcześniak *et al*., 2011; Kang *et al*., 2012). As a retroduplicate of HSP90AA1, HSP90AA4P has three new introns: two were shared between HSP90AA1 and HSP90AA4P in terms of splicing site sequences; the other one evolved by *de novo* mutation (**Figure 2b**). *In silico* prediction indicates that the splicing efficiency improved for these two cryptic splicing sites after retroposition (Kang *et al*., 2012).

Consistent with intronization, events occurred between orthologous worm genes (Irimia *et al*., 2008), 8 of 10 introns did not disrupt the reading frame given the net length difference induced by intronization is multiple of 3 bp (Kang *et al*., 2012). Although the sample size is so small, this trend indicates that natural selection may maintain the open reading frame during the evolution of new exon–intron structures.

Sequence Insertion

General studies on intron gain in orthologous genes across species found that insertion-based mechanism such as intron transposition or transposon insertion is dominant (Li *et al*., 2009; Zhang *et al*., 2010; Torriani Stefano *et al*., 2011). By contrast, aforementioned novel introns in retrogenes mainly evolved by chimerism followed by intronization (Fablet *et al*., 2009; Zhu *et al*., 2009; Kang *et al*., 2012). Actually, no well-characterised case of intron insertion has been reported in retrogenes. Except for the small number of retrogenes, this inconsistency may also be accounted for by the difference of evolutionary ages. Specifically, retrogenes tend to be younger and smaller in size. Thus, older gene surveyed in comparative genomics studies provides big mutational target for sequence insertion. This hypothesis could be tested in future by analysing intron gain patterns between retrogenes of different ages.

Functional Necessity of Intron Gain for Retrogenes

In the context of retrogene evolution, studies of introns tend to be limited to identification and mechanisms (Zhu *et al*., 2009; Szcześniak *et al*., 2011; Kang *et al*., 2012; Zhang *et al*., 2014), the cause of intron gain remains largely unknown. Certainly, just like non-retroposed genes, the existence of introns enables exon shuffling or alternative splicing in retrogenes. The functional role of introns for exon shuffling is self-evident especially for chimeric retrogenes such as *jingwei* or *sphinx* (**Figure 1** and **Figure 4a**). Primate-specific retrogene *RNF113B* and the

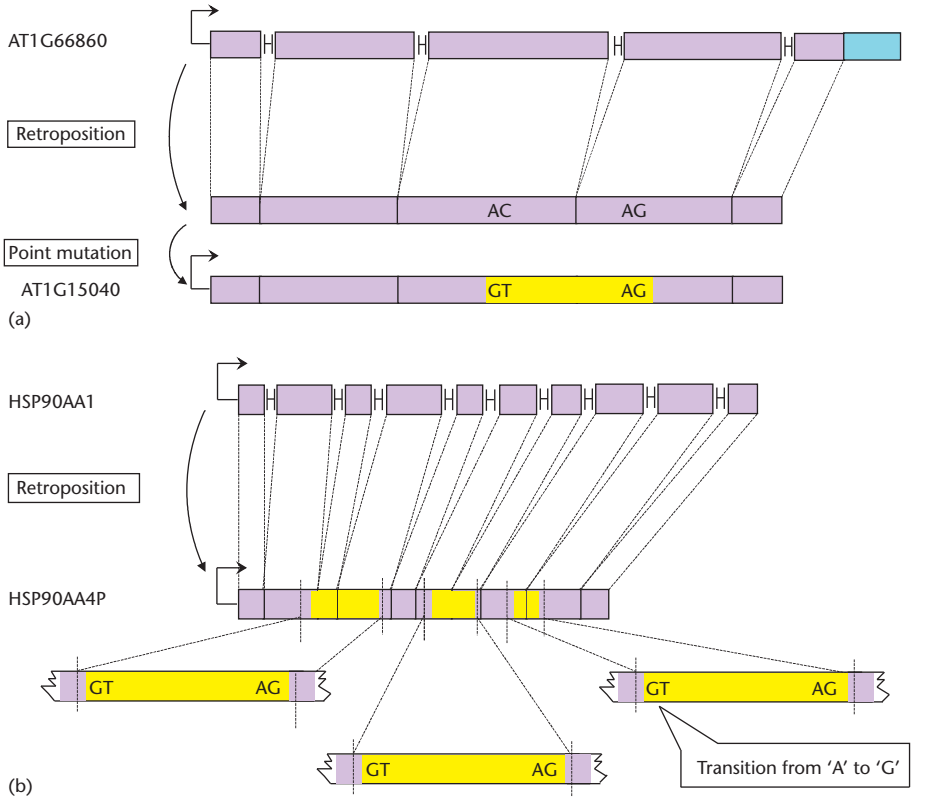

Figure 3 How intronization occurs. The figure convention follows **Figure 1** except that the newly evolved intronic regions are shown in yellow. (a) The retrogene AT1G15040 (*Arabidopsis*) gained an intron after point mutations from 'AC' to 'GT', acting as the splicing donor site. (b) In retrogene HSP90AA4P (human), three new introns were generated by intronization. There is no mutation at the splice sites in the two introns near the 5′ terminus, whereas one transition from 'A' to 'G' (indicated in red) at the splice sites occurred in the intron near the 3′ terminus.

aformentioned *Arabidopsis* retrogene AT1G15040 further demonstrated the significance of introns for alternative splicing (Zhu *et al*., 2009; Szcześniak *et al*., 2011). In both cases, retrogenes encode both spliced form and intron retention form after the intronization event in coding region (**Figure 4b**) (Zhu *et al*., 2009; Szcześniak *et al*., 2011). Actually, the concurrence of two isoforms is also consistent with PTC-driven intronization model (Catania and Lynch, 2008). In other words, the cause of intronization is not to drive alternative splicing, but to serve as compensation of pre-existing PTC mutations (**Figure 4c**; Catania and Lynch, 2008). **See also**: Alternative Splicing: Cell-type-specific and Developmental Control; Alternative Splicing in the Human Genome and its Evolutionary Consequences

Until now only one work directly analysed the functional consequence of retrogene introns (Fablet *et al*., 2009). It is argued that the emergence of introns in 5′ UTR by chimerism is mainly to recruit remote promoters while preventing PTCs (**Figure 4d**). Thus, as expected, retrogenes with 5′ UTR introns generally have broader transcription profiles compared to the remaining retrogenes. By contrast, 3′ UTR introns may lead to downregulation of expression levels via nonsense-mediated decay (Fablet *et al*., 2009). Therefore, introns born in UTRs appear to mainly play regulatory roles.

Conclusion and Open Question

Emerging efforts depict a vivid picture on how retrogenes conventionally believed intronless recruit introns. Depending on the opportunities offered by the genomic context and retroposed sequence itself, retrogenes can evolve novel introns by chimerism, intronization or directly inheriting retained introns from corresponding parental genes (**Table 1**). There is no clear evidence that external sequence insertion can create introns in retrogenes. It is possible that two or more mechanisms participate in multi-intron origination in a single retrogene. For example, the aforementioned rice retrogene LOC_Os05g39720.1 is subject to both chimerism and inheritance of parental introns (**Figure 2b**; Zhang *et al*., 2014). The joint action of chimerism and intronization is also observed in *Arabidopsis* and *Populus* (Zhu *et al*., 2009). Thus, molecular tinkering process (Jacob, 1977) enables the rapid evolution of exon–intron structure of newborn retrogene.

Like retroposed new genes, *de novo* genes, that is, genes born in noncoding DNAs (Levine *et al*., 2006), also evolve rapidly in terms of gene structure. A recent populational survey in fruitfly discovered 142 strain-specific *de novo* genes with 61 (43%) of which gains at least two exons (Zhao *et al*., 2014). Analogously, the majority of

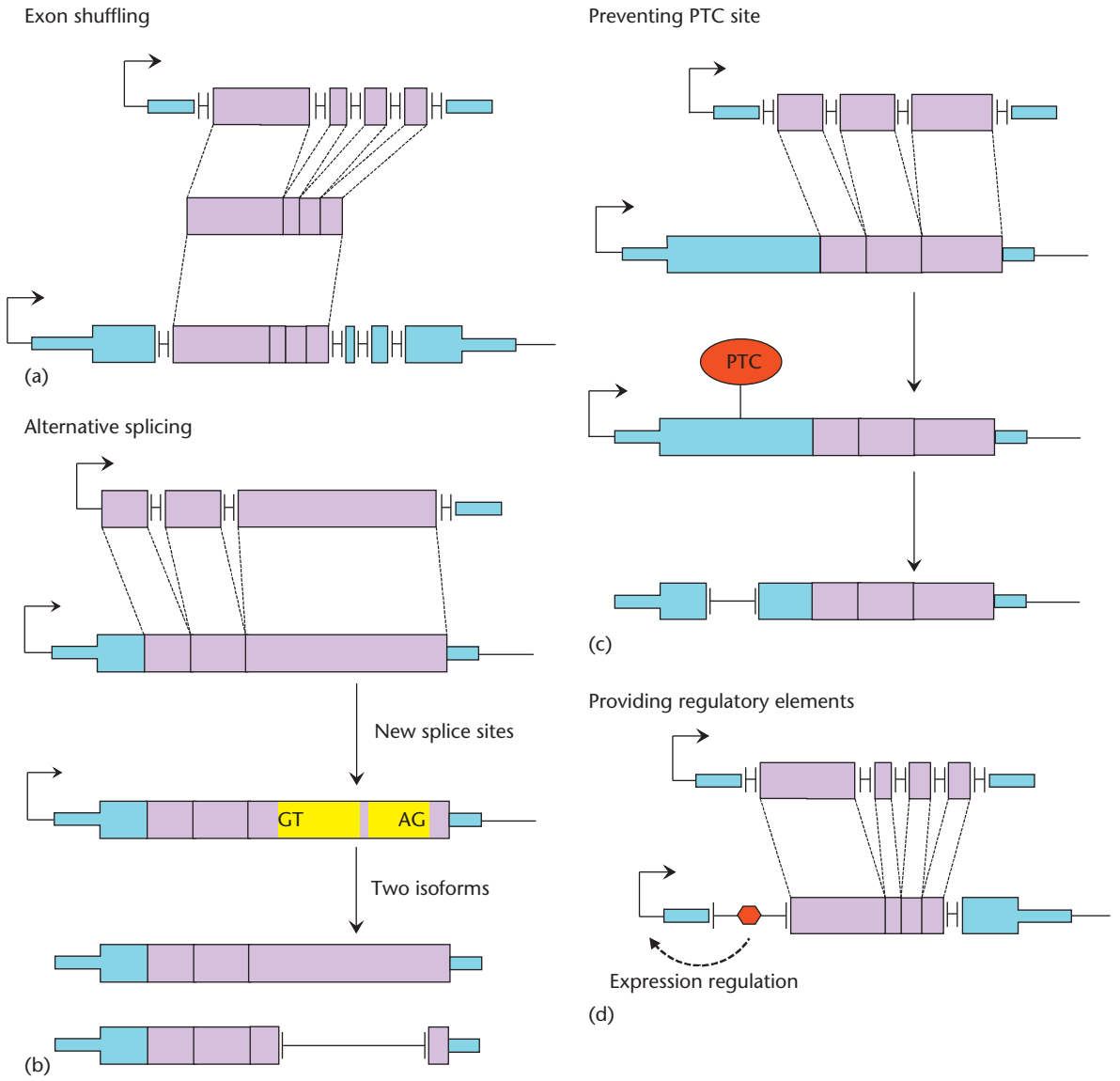

Figure 4 Models on why retrogene needs to evolve introns. (a) The insertion of a retrocopy to the intronic region of a gene (or recruiting the nearby sequences) will result in an exon shuffling event. (b) The occurrence of new splice sites or the activation of previously cryptic sites will generate two or more possible splice variants. (c) The coding potential will be disrupted by a PTC mutation in a long exonic region unless the PTC site is spliced out by a new intron. (d) Retrogene intron can recruit remote promoters and other regulatory elements, leading to regulation of its expression.

hominoid-specific *de novo* genes gain at least one intron (Xie *et al*., 2012). This line of data is consistent with the prevalence of chimerism in retrogene evolution, which not only corroborates the feasibility of *de novo* emergence of introns, but also demonstrates the significant role that new genes play as a target in intron evolution.

Although models like chimerism or intronization are well formulated in the context of intron evolution in retrogenes (Fablet *et al*., 2009; Zhu *et al*., 2009), the evolutionary cause and the functional necessity of intron gain is poorly characterised except a couple of cases documented in (Fablet *et al*., 2009). This survey supports that intron gain in UTRs does have functional consequence, but whether the fixation of these introns is selectively favored in the beginning is still unknown. Hence, as did for the intron presence and absence polymorphism studies in the case of *jingwei* which demonstrated adaptive intron loss (Llopart *et al*., 2002), more populational genomics is required, especially considering the flood of resequencing data in recent years. In parallel, transcription or translation assay in (Fablet *et al*., 2009) could be routinely implemented on the population level to infer the functionality of these newborn introns.

Acknowledgements

We are grateful to Dr Deng-Ke Niu for his critical reading of the manuscript. Authors were supported by grants from the

National Key Basic Research Program of China 973 program (2013CB531200) and 1000 Young Talents Program.

References

Baertsch R, Diekhans M, Kent WJ, Haussler D and Brosius J (2008) Retrocopy contributions to the evolution of the human genome. *BMC Genomics* **9**: 466.

Barrett L, Fletcher S and Wilton S (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences* **69**(21): 3613–3634.

de Boer M, van Leeuwen K, Geissler J *et al*. (2014) Primary immunodeficiency caused by an exonized retroposed gene copy inserted in the CYBB gene. *Human Mutation* **35**(4): 486–496.

Brosius J (2003) The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* **118**(2–3): 99–116.

Brosius J (2005) Echoes from the past – are we still in an RNP world? *Cytogenetic and Genome Research* **110**(1–4): 8–24.

Cardoso-Moreira M and Long M (2012) The origin and evolution of new genes. *Methods in Molecular Biology (Clifton, NJ)* **856**: 161–186.

Catania F and Lynch M (2008) Where do introns come from? *PLoS Biology* **6**(11): 2354–2361.

Catania F and Lynch M (2013) A simple model to explain evolutionary trends of eukaryotic gene architecture and expression. *BioEssays* **35**(6): 561–570.

Chen M, Zou M, Fu BD *et al*. (2011) Evolutionary patterns of RNA-based duplication in non-mammalian chordates. *PLoS One* **6**(7): e21466.

Chen SD, Krinsky BH and Long MY (2013) New genes as drivers of phenotypic evolution. *Nature Reviews Genetics* **14**(9): 645–660.

Courseaux A and Nahon J-L (2001) Birth of two chimeric genes in the Hominidae lineage. *Science* **291**(5507): 1293–1297.

Fablet M, Bueno M, Potrzebowski L and Kaessmann H (2009) Evolutionary origin and functions of retrogene introns. *Molecular Biology and Evolution* **26**(9): 2147–2156.

Fu B, Chen M, Zou M, Long M and He S (2010) The rapid generation of chimerical genes expanding protein diversity in zebrafish. *BMC Genomics* **11**(1): 657.

Gilbert W (1978) Why genes in pieces? *Nature* **271**(5645): 501–501.

Irimia M, Rukov JL, Penny D *et al*. (2008) Origin of introns by intronization of exonic sequences. *Trends in Genetics: TIG* **24**(8): 378–381.

Jacob F (1977) Evolution and tinkering. *Science* **196**(4295): 1161–1166.

Kang L-F, Zhu Z-L, Zhao Q, Chen L-Y and Zhang Z (2012) Newly evolved introns in human retrogenes provide novel insights into their evolutionary roles. *BMC Evolutionary Biology* **12**(1): 128.

Katju V (2012) In with the old, in with the new: the promiscuity of the duplication process engenders diverse pathways for novel gene creation. *International Journal of Evolutionary Biology* **2012**: 24.

Levine MT, Jones CD, Kern AD, Lindfors HA and Begun DJ (2006) Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proceedings of the National Academy of Sciences of the USA* **103**(26): 9935–9939.

Li W, Tucker AE, Sung W, Thomas WK and Lynch M (2009) Extensive, recent intron gains in *Daphnia* populations. *Science* **326**(5957): 1260–1262.

Llopart A, Comeron JM, Brunet FG, Lachaise D and Long M (2002) Intron presence-absence polymorphism in *Drosophila* driven by positive Darwinian selection. *Proceedings of the National Academy of Sciences of the USA* **99**(12): 8121–8126.

Long M, Betran E, Thornton K and Wang W (2003) The origin of new genes: glimpses from the young and old. *Nature Reviews Genetics* **4**(11): 865–875.

Long MY and Langley CH (1993) Natural-selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* **260**(5104): 91–95.

Lynch M (2007) *The Origins of Genome Architecture*. Sunderland, MA: Sinauer Associates.

Meisel RP, Han MV and Hahn MW (2009) A complex suite of forces drives gene traffic from *Drosophila* X chromosomes. *Genome Biology and Evolution* **1**: 176–188.

Muller HJ (1936) Bar duplication. *Science* **83**(2161): 528–530.

Rose AB, Elfersi T, Parra G and Korf I (2008) Promoter-proximal introns in *Arabidopsis thaliana* are enriched in dispersed signals that elevate gene expression. *Plant Cell Online* **20**(3): 543–551.

Roy S (2004) The origin of recent introns: transposons? *Genome Biology* **5**(12): 1–4.

Roy SW, Fedorov A and Gilbert W (2003) Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. *Proceedings of the National Academy of Sciences of the USA* **100**(12): 7158–7162.

Sakai H, Mizuno H, Kawahara Y *et al*. (2011) Retrogenes in rice (*Oryza sativa* L. ssp. *japonica*) exhibit correlated expression with their source genes. *Genome Biology and Evolution* **3**: 1357–1368.

Sharp PA (1985) On the origin of RNA splicing and introns. *Cell* **42**(2): 397–400.

Szcześniak MW, Ciomborowska J, Nowak W, Rogozin IB and Makałowska I (2011) Primate and rodent specific intron gains and the origin of retrogenes with splice variants. *Molecular Biology and Evolution* **28**(1): 33–37.

Torriani Stefano FF, Stukenbrock Eva H, Brunner Patrick C, McDonald Bruce A and Croll D (2011) Evidence for extensive recent intron transposition in closely related fungi. *Current Biology* **21**(23): 2017–2022.

Vinckenbosch N, Dupanloup I and Kaessmann H (2006) Evolutionary fate of retroposed gene copies in the human genome. *Proceedings of the National Academy of Sciences of the USA* **103**(9): 3220–3225.

Wang HF, Feng LA and Niu DK (2007) Relationship between mRNA stability and intron presence. *Biochemical and Biophysical Research Communications* **354**(1): 203–208.

Wang W, Brunet FG, Nevo E and Long M (2002) Origin of sphinx, a young chimeric RNA gene in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **99**(7): 4448–4453.

Wang W, Zheng H, Fan C *et al*. (2006) High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* **18**(8): 1791–1802.

William Roy S and Gilbert W (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nature Reviews Genetics* **7**(3): 211–221.

Xie C, Zhang YE, Chen JY *et al*. (2012) Hominoid-specific *de novo* protein-coding genes originating from long non-coding RNAs. *PLoS Genetics* **8**(9): e1002942.

Yenerall P and Zhou L (2012) Identifying the mechanisms of intron gain: progress and trends. *Biology Direct* **7**(1): 29.

Zhang C, Gschwend AR, Ouyang Y and Long M (2014) Evolution of gene structural complexity: an alternative-splicing based model accounts for intron-containing retrogenes. *Plant Physiology* doi:10.1104/pp.113.231696.pdf.

Zhang L-Y, Yang Y-F and Niu D-K (2010) Evaluation of models of the mechanisms underlying intron loss and gain in *Aspergillus* fungi. *Journal of Molecular Evolution* **71**(5–6): 364–373.

Zhang Y, Wu Y, Liu Y and Han B (2005) Computational identification of 69 retroposons in *Arabidopsis*. *Plant Physiology* **138**(2): 935–948.

Zhang YE, Landback P, Vibranovski M and Long M (2012) New genes expressed in human brains: implications for annotating evolving genomes. *BioEssays* **34**(11): 982–991.

Zhang YE, Vibranovski MD, Krinsky BH and Long MY (2011) A cautionary note for retrocopy identification: DNA-based duplication of intron-containing genes significantly contributes to the origination of single exon genes. *Bioinformatics* **27**(13): 1749–1753.

Zhao L, Saelao P, Jones CD and Begun DJ (2014) Origin and spread of *de novo* genes in *Drosophila melanogaster* populations. *Science* **343**(6172): 769–772.

Zhu Z, Zhang Y and Long M (2009) Extensive structural renovation of retrogenes in the evolution of the *Populus* genome. *Plant Physiology* **151**(4): 1943–1951.

Further Reading

Bai Y, Casola C and Betran E (2008) Evolutionary origin of regulatory regions of retrogenes in *Drosophila*. *BMC Genomics* **9**(1): 241.

Bai Y, Casola C, Feschotte C and Betran E (2007) Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in *Drosophila*. *Genome Biology* **8**(1): R11.

Chorev M and Carmel L (2012) The function of introns. *Frontiers in Genetics* **3**: 55.

Fedorov A, Roy S, Fedorova L and Gilbert W (2003) Mystery of intron gain. *Genome Research* **13**(10): 2236–2241.

Fica SM, Tuttle N, Novak T *et al*. (2013) RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**(7475): 229–234.

Le Hir H, Nott A and Moore MJ (2003) How introns influence and enhance eukaryotic gene expression. *Trends in Biochemical Sciences* **28**(4): 215–220.

Rogozin I, Carmel L, Csuros M and Koonin E (2012) Origin and evolution of spliceosomal introns. *Biology Direct* **7**(1): 11.

Strobel SA (2013) Biochemistry: metal ghosts in the splicing machine. *Nature* **503**(7475): 201–202.

Yao J, Truong DM and Lambowitz AM (2013) Genetic and biochemical assays reveal a key role for replication restart proteins in group II intron retrohoming. *PLoS Genetics* **9**(4): e1003469.