Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal of Genetics and Genomics 40 (2013) 171−177

JGG

REVIEW

Pseudogenes: Pseudo or Real Functional Elements?

Wen Li [a,b], Wei Yang [a,b], Xiu-Jie Wang [a,*]

[a] State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China
[b] Graduate University of Chinese Academy of Sciences, Beijing 100101, China

Received 28 February 2013; revised 4 March 2013; accepted 4 March 2013
Available online 14 March 2013

ABSTRACT

Pseudogenes are genomic remnants of ancient protein-coding genes which have lost their coding potentials through evolution. Although broadly existed, pseudogenes used to be considered as junk or relics of genomes which have not drawn enough attentions of biologists until recent years. With the broad applications of high-throughput experimental techniques, growing lines of evidence have strongly suggested that some pseudogenes possess special functions, including regulating parental gene expression and participating in the regulation of many biological processes. In this review, we summarize some basic features of pseudogenes and their functions in regulating development and diseases. All of these observations indicate that pseudogenes are not purely dead fossils of genomes, but warrant further exploration in their distribution, expression regulation and functions. A new nomenclature is desirable for the currently called 'pseudogenes' to better describe their functions.

KEYWORDS: Pseudogene; Categorization; Origination; Function

INTRODUCTION

The word 'pseudogene' was first coined in 1977, when Jacq et al. (1977) reported the existence of a group of untranscribed genomic sequences homologous to the 5S DNA in *Xenopus laevis*. After that, pseudogenes have been identified to be widely existed in the genomes of most organisms, ranging from prokaryotes to eukaryotes (Hardison et al., 1979; Colbert et al., 1980; Proudfoot and Maniatis, 1980; Lee et al., 1983; Fischer and Maniatis, 1985; Harrison et al., 2001; Homma et al., 2002; Liu et al., 2004). Traditionally, pseudogenes were considered as the relics of ancient genes which do not possess real functions (Proudfoot, 1980; Zhang and Gerstein, 2004; Gray et al., 2006), as they have mutated from their parental genes therefore no longer encode proteins (Nishioka et al., 1980; Willecke et al., 1990; Jeffs and Ashburner, 1991). During recent two decades, especially with the broad applications of next-generation sequencing technologies, multiple classes of non-coding RNAs have been identified and extensive studies have revealed their essential functions in various biological processes, which have drawn great attention to the non-coding regions of genomes, including pseudogenes. Recent emerging evidence has confirmed that some pseudogenes have acquired diverse functions in regulating development and diseases (Korneev et al., 1999; Balakirev and Ayala, 2003; Kandouz et al., 2004; Tam et al., 2008; Podlaha and Zhang, 2010; Poliseno et al., 2010; Pink et al., 2011). Here, we review the identification, categorization, and functions of pseudogenes. Some examples related to their expression and functions are also included.

DEFINITION AND CLASSIFICATION OF PSEUDOGENES

Pseudogenes were originally defined as aberrant genes with high sequence similarity to the functional genes but have lost their coding ability, mainly due to the presence of premature stop-codons or frame shift (Proudfoot, 1980). Thus, they are considered as non-functional genomic elements which are not

* Corresponding author. Tel: +86 10 6480 6590, fax: +86 10 6480 6595.
E-mail address: xjwang@genetics.ac.cn (X.-J. Wang).

1673-8527/$ - see front matter Copyright © 2013, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, and Genetics Society of China. Published by Elsevier Limited and Science Press. All rights reserved.
http://dx.doi.org/10.1016/j.jgg.2013.03.003

under purifying selection. However, during the past two decades, a growing number of pseudogenes have been proven to possess important functions (Zhou et al., 1992; Korneev et al., 1999; Kandouz et al., 2004; Tam et al., 2008; Poliseno, 2012), suggesting that pseudogenes are not only evolutionary relics of genomes, but many may be functional elements (D'Errico et al., 2004; Zheng and Gerstein, 2007).

Pseudogenes can be generated by degeneration of single copy or duplicated protein-coding genes, or by retro-transposition of processed mRNAs back to the genome. According to their generation mechanisms, pseudogenes generally can be categorized into three groups, namely unitary pseudogenes, duplicated pseudogenes, and processed pseudogenes (Fig. 1). Unitary pseudogenes are derived from single copy functional genes, which accumulated spontaneous mutations during evolution and have lost their primary functions. Therefore, unitary pseudogenes have no paralogs in the same genome, but may have orthologs in relative species (Zhang et al., 2010). Duplicated pseudogenes are originated from uncompleted or mutated gene duplications. In these cases, one copy of the duplicated genes become pseudogenes, whereas the parental genes still retain their original functions (Lacy and Maniatis, 1980; Proudfoot and Maniatis, 1980). Both unitary and duplicated pseudogenes can also be called as unprocessed pseudogenes because they are derived directly from DNA sequences and maintain their original intron−exon structures and regulatory elements. To the contrary, the processed pseudogenes are formed by the retrotransposition of mRNA transcripts (Nishioka et al., 1980). As majority of these retrotransposed mRNAs are processed mature mRNAs, this group of processed pseudogenes usually contain the features of mRNAs and retroviruses, characterized by the presence of 3′ polyadenylation tags and flanking direct repeats but lacking of introns and 5′ regulatory sequences. Unlike duplicated pseudogenes which usually locate near their parental genes, processed pseudogenes are more likely to be found far away from their original genes or on different chromosomes (Vanin, 1985).

IDENTIFICATION OF PSEUDOGENES

The primary question regarding studies of pseudogenes is how to identify them from the genome. According to the definition of pseudogenes, their sequences should be highly homology to protein-coding genes but not under the conservation constrains. Therefore, computing the ratio of nucleotide nonsynonymous to synonymous substitutions (Ka/Ks) is a practical way to identify pseudogenes (Li et al., 1981). Unlike protein-coding genes, pseudogenes should be under neutral selection rather than purifying selection and their Ka/Ks ratios are expected to be equal to one (Wang et al., 2000; Betran et al., 2002; Torrents et al., 2003). With this criterion, thousands of pseudogenes have been identified in multiple species, including more than 8000 processed pseudogenes in the human genome (Table 1).

The second class of method considered more on the features of pseudogene classification rules as well as their relationships with parental genes. For instance, unitary pseudogenes are degenerated functional genes without paralogs in the same genome, thus are better to be identified by comparative genomics using other genomes as references (Zhang

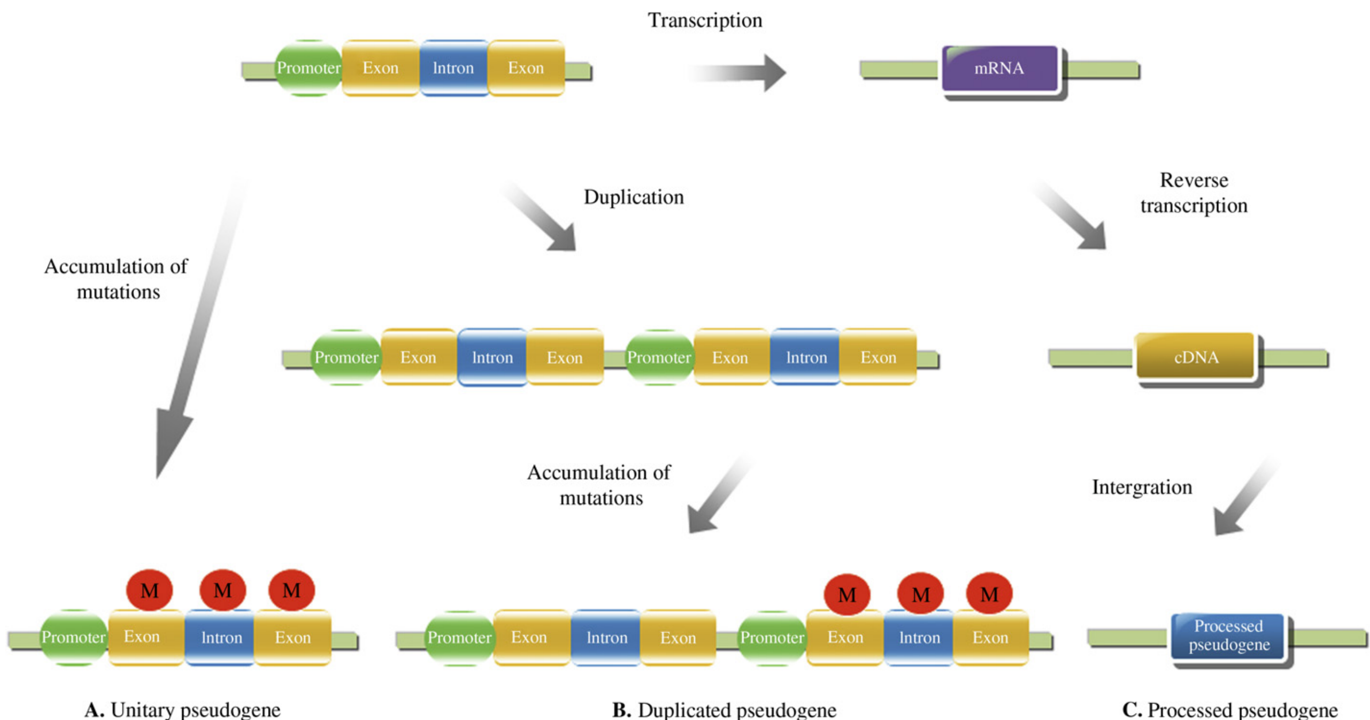

Fig. 1. Classification rules of pseudogenes.
A: unitary pseudogenes are originated from functional genes with the accumulation of degenerative mutations. **B:** duplicated pseudogenes are derived from one copy of duplicated genes. **C:** processed pseudogenes are generated by retrotransposition of mRNAs. Mutations are presented by red circles with the 'M' mark.

Table 1
Summary of annotated genes and pseudogenes in the genomes of nine species

Organism	No. of genes	No. of pseudogenes	No. of processed pseudogenes	Reference
E. coli str. K-12 substr. MG1655	4496[a]	95	0	Homma et al., 2002
E. coli O157:H7 str. Sakai	5460[a]	101	0	Homma et al., 2002
S. cerevisiae	6352[a]	221	0	Harrison et al., 2002
C. elegans	20,444[b]	2168	208	Harrison et al., 2001
D. melanogaster	14,408[b]	110	34	Harrison et al., 2003
G. gallus	16,382[b]	959[a]	51	Hillier et al., 2004
M. musculus	25,798[b]	9173	4847	Zhang et al., 2004
H. sapiens	34,580[b]	11,580	8298	Harrow et al., 2012
A. thaliana	27,819[b]	924[c]	376	Benovoy and Drouin, 2006

[a] Data from NCBI Genome (http://www.ncbi.nlm.nih.gov/genome); [b] Data from NCBI UniGene (http://www.ncbi.nlm.nih.gov/unigene/statistics); [c] Data from TAIR 10 (http://www.arabidopsis.org).

et al., 2010). On the other hand, processed pseudogenes should be intron-less and with polyadenylation sequences in most cases, and duplicated pseudogenes should have paralogous genes within the same genome.

Although pseudogenes are considered to be created by random or erroneous processes, some of them are indeed transcribed. Such expression feature has made transcribed pseudogenes easier to be identified than the untranscribed ones. With the increasing application of the high-throughput sequencing technology, genome-wide identification of expressed pseudogenes has been carried out in multiple human tissues and cancer samples, which have identified over 2000 pseudogenes with ubiquitous or cancer-specific expression patterns (Kalyana-Sundaram et al., 2012; Tonner et al., 2012). One difficulty in the identification of expressed pseudogenes by the sequencing method is to distinguish pseudogenes with only a few nucleotide differences from their parental genes. How to determine whether the sparse and rare nucleotide differences presented in the sequencing results is caused by degenerative mutation, individual genome difference or sequencing error is a major challenge faced by this method. Further improvement of sequencing accuracy and the development of more appropriate analysis methods are helpful for solving this problem.

ORIGINATION OF PSEUDOGENES

In theory, pseudogenes can be derived from any gene sequences of a genome. Yet differences have been observed on the frequency of different genes to produce pseudogenes. Housekeeping genes, genes highly expressed in germline cells and genes participating in basic metabolic regulations are the typical genes with multiple corresponding pseudogenes (Frederiksen et al., 1997; Zhang et al., 2003; Zhang et al., 2004; Pei et al., 2012). This phenomenon may be due to the high expression of these genes therefore they are more likely to accumulate mutations or be retrotransposed back to the genome. In addition to the expression level, the GC content of the genomic region where the pseudogenes were deposited will also affect the accumulation rate of mutations (Bustamante et al., 2002). The length of a gene is another factor affecting the generation of pseudogenes. It has been shown that long protein-coding genes tend to produce non-processed pseudogenes whereas short protein-coding genes tend to produce processed pseudogenes (Goncalves et al., 2000; Zhang et al., 2002; Khachane and Harrison, 2009b).

Some pseudogenes can also give birth to other pseudogenes. For example, the *PCNA* gene has two pseudogenes, *p1PCNA* and *pF2PCNA*, which are tandemly located within the human Chr4q24 region. Both of them are processed pseudogenes, but *pF2PCNA* has truncations on both ends as compared to *p1PCNA*. The sequences of both *p1PCNA* and *pF2PCNA* have a common five nucleotides deletion, and further phylogenetic study demonstrated that *pF2PCNA* may be originated from *p1PCNA* (Taniguchi et al., 1996).

The numbers of pseudogenes vary a lot among the genomes of different species, and had no correlation with the degree of complexity of the organisms (Table 1). Besides the consequence of natural selection and DNA deletion rate (Petrov et al., 1996), differences on the completeness of pseudogene identification may also be a reason for such variance among species.

REPORTED FUNCTIONS OF PSEUDOGENES

Many pseudogenes are transcribed in parallel as their parental genes (Fischer and Maniatis, 1985), or with their own tissue or temporal specific patterns (Elliman et al., 2006), indicating that their expression may be other than the transcriptional noise. However, expression may simply be the transcription byproduct of the neighboring genomic units and has no correlation with functions (Ebisuya et al., 2008). As the sequences of pseudogenes are supposed not to be under the constrain of natural selection, evolutionary conservation of sequences is a stronger indicator for the functional importance of some pseudogenes (Sudbrak et al., 2003; Harrison et al., 2005; Svensson et al., 2006; Khachane and Harrison, 2009a). *Pbcas4*, a transcribed unitary pseudogene which functions as microRNA (miRNA) decoy to regulate its parental gene, breast carcinoma amplified sequence 4 (*BCAS4*), is conserved between human and mouse. Other conserved pseudogene examples include pseudogenes derived from ataxia type 1 (*ATX1*) and ataxin 7-like 3 (*ATX7NL3*) (Svensson et al., 2006).

A recent systematic survey had identified 48 pseudogenes which are conserved in human, mouse, rat and dog, and also with detectable expression in human and dog (Marques et al., 2012). Another study also identified 68 human transcribed pseudogenes that are conserved in no less than two other mammals (Khachane and Harrison, 2009a). It is worth to note that sequence conservation may only be applied to pseudogenes with functional requirement at the sequence level, for those working at the structural level, there will be less conservation constrain on sequences.

In recent years, increasing lines of evidence have shown that some pseudogenes possess important functions in regulating the normal growth of an organism and the development of some diseases, especially in cancers. They can serve as antisense regulatory transcripts or miRNA decoys, produce small interfering RNAs (siRNAs), and encode short peptides or proteins.

Function as antisense transcripts

Some pseudogenes can function as antisense transcripts to regulate other genes (Fig. 2A), such as topoisomerase I (*TOPI*). The human genome contains two processed pseudogenes of *TOPI* (Kunze et al., 1989), one of which was reversely integrated into the genome, therefore its transcript is complementary to the *TOPI* mRNA in sequence (Zhou et al., 1992). Similar mechanism was found in the pseudogene of nitric oxide synthase (*NOS*), pseudo-*NOS*. The pseudo-*NOS* no longer encodes proteins due to the presence of premature stop-codons (Korneev et al., 1999), instead, it contains a region antisense to the mRNA of the neuronal transcribed NOS (*nNOS*) therefore can inhibit the translation of NOS by forming double-strand RNA-RNA duplex with the *nNOS* mRNA (Korneev et al., 1999). A later study has shown that the pseudo-*NOS* transcripts are transported from the cerebral ganglion to the synaptic zone of the buccal ganglion upon the learning process, indicating that they may play an active role in memory formation by regulating NO production (Korneev et al., 2013).

Serve as miRNA decoys

miRNAs are single-stranded short non-coding RNAs which mainly target the 3′-untranslated region (UTR) of mRNAs based on their sequence complementarities, and resulted in the silencing of mRNAs through either direct cleavage or translational repression at the post-transcriptional level. Consequently, pseudogenes sharing sequence similarities with miRNA targets can bind to miRNAs and serve as the decoys therefore to inhibit the function of miRNAs on their real targets (Fig. 2B). One of such examples is *PTENP1*, which is a processed pseudogene with high homology to phosphatase and tensin homolog deleted on chromosome ten (*PTEN*). The 3′UTR sequences of both *PTEN* and *PTENP1* have the target site of the same miRNA, therefore

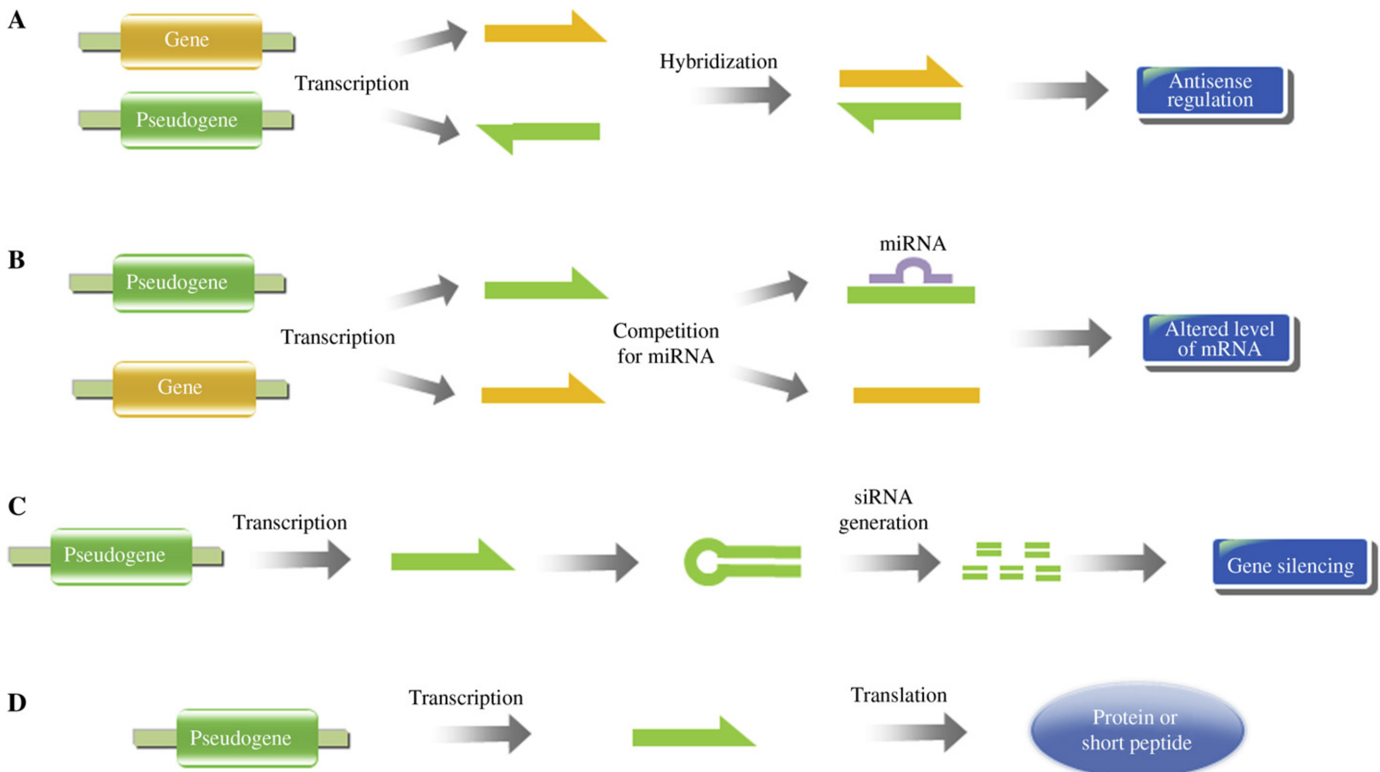

Fig. 2. Functional models of pseudogenes.
A: a reversely integrated pseudogene has transcripts complementary to the mRNAs of its parental gene, which can regulate the parental mRNA *via* antisense regulatory mechanism. **B:** transcripts of pseudogenes serve as miRNA decoys and inhibit the interaction of miRNAs with their real targets. **C:** transcripts of pseudogenes with inverted repeat sequences can form hairpin-shaped secondary structures and produce endogenous siRNAs. **D:** some pseudogenes retain or acquire the ability to encode short peptides or proteins, often with novel functions.

PTENP1 transcript can function as the decoy to interfere the binding between miRNA and *PTEN* mRNA (Poliseno et al., 2010). Another pseudogene functioning as miRNA decoy is the previously mentioned *Pbcas4*. Experiments in mouse neuroblastoma cells have shown that *Pbcas4* not only regulated the expression of its parental gene, but also dramatically affected the expression of other mRNAs sharing the same miRNA response elements (MREs) in their 3′UTRs (Marques et al., 2012).

Produce siRNAs

Endogenous small interfering RNAs (endo-siRNAs) can be generated from pseudogenes *via* two major mechanisms. One is from double-stranded RNA duplexes formed by mRNAs and their reversely transcribed pseudogenes; the other is from the inverted repeat regions of pseudogenes with hairpin-shaped secondary structures (Fig. 2C). Thus, pseudogene-derived siRNAs may be capable to regulate their parental genes *via* the RNA-interference mechanism. In 2008, two groups simultaneously reported pseudogene-derived siRNAs in mouse oocytes (Tam et al., 2008; Watanabe et al., 2008). The expression of these siRNAs was dependent on the small RNA biogenesis proteins Dicer and AGO2. Repression of siRNA expression was accompanied by an increased expression of transcripts with sequence complementary to these siRNAs, indicating that the expression of the transcripts was regulated by the pseudogene-derived siRNAs (Tam et al., 2008; Watanabe et al., 2008). Genome-wide analysis of siRNAs in *Arabidopsis* (Kasschau et al., 2007) and rice (Guo et al., 2009) suggested that pseudogene-derived siRNAs also exist in plants.

Encode short peptides or proteins

Some pseudogenes also have the ability to produce short peptides or proteins (Fig. 2D). Phosphoglycerate mutase family 3 (*PGAM3*) is the first known protein-coding pseudogene, which has only been found in the genome sequences of primates, including humans, chimpanzee and macaque (Betran et al., 2002). Similar to other processed pseudogenes, *PGAM3* was generated from *PGAM1* by retrotransposition and antisense to the Menkes disease gene (*MNK*) (Dierick et al., 1997). Although promoter-like sequence was found upstream of the *PGAM3* gene, it was thought not to have functional products (Dierick et al., 1997) until the polymorphism and expression data verified that *PGAM3* had the ability to produce proteins (Betran et al., 2002). Another expressed pseudogene with the ability to produce protein is the pseudogene of connexin 43 (ψ*Cx43*) in human. The *Cx43* gene only has one intron, and encodes a protein functioning as part of the gap junction channels. Although having lost the intron (Willecke et al., 1990; Fishman et al., 1991), ψ*Cx43* has a complete open reading frame and a regulatory element (Fishman et al., 1991). Studies have shown that ψ*Cx43* transcripts can be translated into a 43 kDa protein in tumor cells with the ability to inhibit cell growth (Kandouz et al., 2004). As encoding proteins is a typical feature of most functional genes, it should be reconsidered whether these pseudogenes should still be called as ‘pseudo’.

PSEUDOGENES RELATED TO DEVELOPMENT AND DISEASES

Similar to functional genes, some pseudogenes also have developmentally regulated expression patterns. By analyzing RNA-Seq data, Lin et al. (2011) reported 1371 pseudogenes with significant expression changes upon the early differentiation of human neurons. Among them, 1052 pseudogenes were down-regulated whereas the other 319 pseudogenes were up-regulated. The function of pseudogenes in regulating stem cell property has also been evidenced. It has been proven that the expression of a putative pseudogene transcript of the key pluripotent regulatory factor*Oct4* was able to promote cell proliferation whereas inhibit mesenchymal stem cell differentiation (Lin et al., 2007). Singh et al. (2012) also reported four *Oct4* pseudogenes, all of which were found to be expressed in buffalo embryonic stem cell-like cells.

Some pseudogenes are specifically expressed in certain cancers or diseases. *PTEN*, for instance, is a tumor suppressor gene whose mutation can cause various human cancers. It has been shown that the pseudogene of *PTEN*, *PTENP1*, can positively regulate the expression of *PTEN* and repress cell growth. In some human cancer cells, the *PTENP1* locus was selectively lost, resulting in decreased expression of *PTEN* and abnormal proliferation of cancer cells (Poliseno et al., 2010). Another example of known cancer-related pseudogene is the myosin light chain kinase pseudogene (*MYLKP1*), a duplicated pseudogene of *MYLK* which is highly expressed in carcinoma tissues and cell lines. The expression of *MYLKP1* can decrease the stability of *MYLK* mRNA at the post-transcriptional level and stimulate cell proliferation (Han et al., 2011). Besides cancers, pseudogenes also involve in the development of other diseases. One example is the high mobility group A1 pseudogene (*HMGA1-p*), whose expression can trigger the destabilization of *HMGA1* mRNA. As the HMGA1 protein participates in the regulation of the insulin receptor (INSR), the expression of *HMGA1-p* plays roles in the onset of type 2 diabetes (Chiefari et al., 2010).

Although the regulatory functions of pseudogenes seem to be striking, the functional studies of pseudogenes are still in its early stage. Controversial results have also been found for some pseudogenes. For example, the mouse *Makorin1-p1* pseudogene was originally reported to stabilize the mRNAs of *Makorin1* (Hirotsune et al., 2003; Lee, 2003; Podlaha and Zhang, 2004), but later proven to be non-transcribed and without sequence conservation (Gray et al., 2006; Kaneko et al., 2006). Due to the close sequence similarity of pseudogenes with their parental protein-coding genes, extra cautiousness has to be taken when studying the functions of pseudogenes.

CONCLUSIONS AND PERSPECTIVES

The great majority of known eukaryotic genomes are composed of non-coding sequences and parts of which are pseudogenes that had been ignored over a long period of time. With the improvement of sequencing technologies, people

started to pay attention to the identification and functional studies of pseudogenes.

From the evolution point of view, it seems to be economic and effective for a genome to acquire novel functions from the already existing sequences such as pseudogenes. Many studies are still required to gain a comprehensive understanding on pseudogenes and their functions, including but not limited to the development of accurate standard for the identification of pseudogenes, the investigations on the generation and regulatory mechanisms of pseudogenes, the studies of the expression patterns and functions of pseudogenes, especially during organism development and in diseases. With the help of the next-generation sequencing technology and other recently developed high-throughput experimental methods, genome-wide identification and functional studies of pseudogenes have become more feasible. As the importance of non-coding sequences has been more and more appreciated, it is expectable that many new features of pseudogenes will be discovered in the near future.

Furthermore, the word 'pseudogene' is a relative term. Pseudogenes are originally called 'pseudo' as compared to their parental genes. However, with the growing lines of evidence for the real functions of pseudogenes in different organisms, a new nomenclature which can appropriately reflect the features and functions of pseudogenes is desired.

ACKNOWLEDGEMENTS

This work was supported by the grants from the Ministry of Science and Technology of China (No. 2011CBA01101) to X.-J. W. and from the Chinese Academy of Sciences (Nos. XDA01020105, KSCX2-EW-R-01-03 and 2010-Biols-CAS-0303) to X.-J. W.

REFERENCES

Balakirev, E.S., Ayala, F.J., 2003. Pseudogenes: are they "junk" or functional DNA? Annu. Rev. Genet. 37, 123–151.

Benovoy, D., Drouin, G., 2006. Processed pseudogenes, processed genes, and spontaneous mutations in the *Arabidopsis* genome. J. Mol. Evol. 62, 511–522.

Betran, E., Wang, W., Jin, L., Long, M., 2002. Evolution of the phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene. Mol. Biol. Evol. 19, 654–663.

Bustamante, C.D., Nielsen, R., Hartl, D.L., 2002. A maximum likelihood method for analyzing pseudogene evolution: implications for silent site evolution in humans and rodents. Mol. Biol. Evol. 19, 110–117.

Chiefari, E., Iiritano, S., Paonessa, F., Le Pera, I., Arcidiacono, B., Filocamo, M., Foti, D., Liebhaber, S.A., Brunetti, A., 2010. Pseudogene-mediated posttranscriptional silencing of *HMGA1* can result in insulin resistance and type 2 diabetes. Nat. Commun. 1, 40.

Colbert, D.A., Knoll, B.J., Woo, S.L., Mace, M.L., Tsai, M.J., O'Malley, B.W., 1980. Differential hormonal responsiveness of the ovalbumin gene and its pseudogenes in the chick oviduct. Biochemistry 19, 5586–5592.

D'Errico, I., Gadaleta, G., Saccone, C., 2004. Pseudogenes in metazoa: origin and features. Brief Funct. Genomic Proteomic 3, 157–167.

Dierick, H.A., Mercer, J.F., Glover, T.W., 1997. A phosphoglycerate mutase brain isoform (*PGAM 1*) pseudogene is localized within the human Menkes disease gene (*ATP7 A*). Gene 198, 37–41.

Ebisuya, M., Yamamoto, T., Nakajima, M., Nishida, E., 2008. Ripples from neighbouring transcription. Nat. Cell. Biol. 10, 1106–1113.

Elliman, S.J., Wu, I., Kemp, D.M., 2006. Adult tissue-specific expression of a *Dppa3*-derived retrogene represents a postnatal transcript of pluripotent cell origin. J. Biol. Chem. 281, 16–19.

Fischer, J.A., Maniatis, T., 1985. Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes. Nucleic Acids Res. 13, 6899–6917.

Fishman, G.I., Eddy, R.L., Shows, T.B., Rosenthal, L., Leinwand, L.A., 1991. The human connexin gene family of gap junction proteins: distinct chromosomal locations but similar structures. Genomics 10, 250–256.

Frederiksen, S., Cao, H., Lomholt, B., Levan, G., Hallenberg, C., 1997. The rat 5S rRNA bona fide gene repeat maps to chromosome 19q12->qter and the pseudogene repeat maps to 12q12. Cytogenet. Cell Genet. 76, 101–106.

Goncalves, I., Duret, L., Mouchiroud, D., 2000. Nature and structure of human genes that generate retropseudogenes. Genome Res. 10, 672–678.

Gray, T.A., Wilson, A., Fortin, P.J., Nicholls, R.D., 2006. The putatively functional *Mkrn1-p1* pseudogene is neither expressed nor imprinted, nor does it regulate its source gene in trans. Proc. Natl. Acad. Sci. USA 103, 12039–12044.

Guo, X., Zhang, Z., Gerstein, M.B., Zheng, D., 2009. Small RNAs originated from pseudogenes: *cis*- or *trans*-acting? PLoS Comput. Biol. 5, e1000449.

Han, Y.J., Ma, S.F., Yourek, G., Park, Y.D., Garcia, J.G., 2011. A transcribed pseudogene of *MYLK* promotes cell proliferation. FASEB J. 25, 2305–2312.

Hardison, R.C., Butler 3rd, E.T., Lacy, E., Maniatis, T., Rosenthal, N., Efstratiadis, A., 1979. The structure and transcription of four linked rabbit beta-like globin genes. Cell 18, 1285–1297.

Harrison, P., Kumar, A., Lan, N., Echols, N., Snyder, M., Gerstein, M., 2002. A small reservoir of disabled ORFs in the yeast genome and its implications for the dynamics of proteome evolution. J. Mol. Biol. 316, 409–419.

Harrison, P.M., Echols, N., Gerstein, M.B., 2001. Digging for dead genes: an analysis of the characteristics of the pseudogene population in the *Caenorhabditis elegans* genome. Nucleic Acids Res. 29, 818–830.

Harrison, P.M., Milburn, D., Zhang, Z., Bertone, P., Gerstein, M., 2003. Identification of pseudogenes in the *Drosophila melanogaster* genome. Nucleic Acids Res. 31, 1033–1037.

Harrison, P.M., Zheng, D., Zhang, Z., Carriero, N., Gerstein, M., 2005. Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. Nucleic Acids Res. 33, 2374–2383.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., Barnes, I., Bignell, A., Boychenko, V., Hunt, T., Kay, M., Mukherjee, G., Rajan, J., Despacio-Reyes, G., Saunders, G., Steward, C., Harte, R., Lin, M., Howald, C., Tanzer, A., Derrien, T., Chrast, J., Walters, N., Balasubramanian, S., Pei, B., Tress, M., Rodriguez, J.M., Ezkurdia, I., van Baren, J., Brent, M., Haussler, D., Kellis, M., Valencia, A., Reymond, A., Gerstein, M., Guigo, R., Hubbard, T.J., 2012. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 22, 1760–1774.

Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A., Delany, M.E., Dodgson, J.B., 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432, 695–716.

Hirotsune, S., Yoshida, N., Chen, A., Garrett, L., Sugiyama, F., Takahashi, S., Yagami, K., Wynshaw-Boris, A., Yoshiki, A., 2003. An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. Nature 423, 91–96.

Homma, K., Fukuchi, S., Kawabata, T., Ota, M., Nishikawa, K., 2002. A systematic investigation identifies a significant number of probable pseudogenes in the *Escherichia coli* genome. Gene 294, 25–33.

Jacq, C., Miller, J.R., Brownlee, G.G., 1977. A pseudogene structure in 5S DNA of *Xenopus laevis*. Cell 12, 109–120.

Jeffs, P., Ashburner, M., 1991. Processed pseudogenes in *Drosophila*. Proc. Biol. Sci. 244, 151–159.

Kalyana-Sundaram, S., Kumar-Sinha, C., Shankar, S., Robinson, D.R., Wu, Y.M., Cao, X., Asangani, I.A., Kothari, V., Prensner, J.R., Lonigro, R.J., Iyer, M.K., Barrette, T., Shanmugam, A., Dhanasekaran, S.M., Palanisamy, N., Chinnaiyan, A.M., 2012. Expressed pseudogenes in the transcriptional landscape of human cancers. Cell 149, 1622–1634.

Kandouz, M., Bier, A., Carystinos, G.D., Alaoui-Jamali, M.A., Batist, G., 2004. *Connexin43* pseudogene is expressed in tumor cells and inhibits growth. Oncogene 23, 4763–4770.

Kaneko, S., Aki, I., Tsuda, K., Mekada, K., Moriwaki, K., Takahata, N., Satta, Y., 2006. Origin and evolution of processed pseudogenes that stabilize functional *Makorin1* mRNAs in mice, primates and other mammals. Genetics 172, 2421–2429.

Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Carrington, J.C., 2007. Genome-wide profiling and analysis of *Arabidopsis* siRNAs. PLoS Biol. 5, e57.

Khachane, A.N., Harrison, P.M., 2009a. Assessing the genomic evidence for conserved transcribed pseudogenes under selection. BMC Genomics 10, 435.

Khachane, A.N., Harrison, P.M., 2009b. Strong association between pseudogenization mechanisms and gene sequence length. Biology Direct 4, 38.

Korneev, S.A., Kemenes, I., Bettini, N.L., Kemenes, G., Staras, K., Benjamin, P.R., O'Shea, M., 2013. Axonal trafficking of an antisense RNA transcribed from a pseudogene is regulated by classical conditioning. Sci. Rep. 3, 1027.

Korneev, S.A., Park, J.H., O'Shea, M., 1999. Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an *NOS* pseudogene. J. Neurosci. 19, 7711–7720.

Kunze, N., Yang, G.C., Jiang, Z.Y., Hameister, H., Adolph, S., Wiedorn, K.H., Richter, A., Knippers, R., 1989. Localization of the active type I DNA topoisomerase gene on human chromosome 20q11.2-13.1, and two pseudogenes on chromosomes 1q23-24 and 22q11.2-13.1. Hum. Genet. 84, 6–10.

Lacy, E., Maniatis, T., 1980. The nucleotide sequence of a rabbit beta-globin pseudogene. Cell 21, 545–553.

Lee, J.S., Brown, G.G., Verma, D.P., 1983. Chromosomal arrangement of leghemoglobin genes in soybean. Nucleic Acids Res. 11, 5541–5553.

Lee, J.T., 2003. Molecular biology: complicity of gene and pseudogene. Nature 423, 26–28.

Li, W.H., Gojobori, T., Nei, M., 1981. Pseudogenes as a paradigm of neutral evolution. Nature 292, 237–239.

Lin, H., Shabbir, A., Molnar, M., Lee, T., 2007. Stem cell regulatory function mediated by expression of a novel mouse *Oct4* pseudogene. Biochem. Biophys. Res. Commun. 355, 111–116.

Lin, M., Pedrosa, E., Shah, A., Hrabovsky, A., Maqbool, S., Zheng, D., Lachman, H.M., 2011. RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. PLoS ONE 6, e23356.

Liu, Y., Harrison, P.M., Kunin, V., Gerstein, M., 2004. Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. Genome Biol. 5, R64.

Marques, A.C., Tan, J., Lee, S., Kong, L., Heger, A., Ponting, C.P., 2012. Evidence for conserved post-transcriptional roles of unitary pseudogenes and for frequent bifunctionality of mRNAs. Genome Biol. 13, R102.

Nishioka, Y., Leder, A., Leder, P., 1980. Unusual alpha-globin-like gene that has cleanly lost both globin intervening sequences. Proc. Natl. Acad. Sci. USA 77, 2806–2809.

Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X.J., Harte, R., Balasubramanian, S., Tanzer, A., Diekhans, M., Reymond, A., Hubbard, T.J., Harrow, J., Gerstein, M.B., 2012. The GENCODE pseudogene resource. Genome Biol. 13, R51.

Petrov, D.A., Lozovskaya, E.R., Hartl, D.L., 1996. High intrinsic rate of DNA loss in *Drosophila*. Nature 384, 346–349.

Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L., Carter, D.R., 2011. Pseudogenes: pseudo-functional or key regulators in health and disease? RNA 17, 792–798.

Podlaha, O., Zhang, J., 2004. Nonneutral evolution of the transcribed pseudogene *Makorin1-p1* in mice. Mol. Biol. Evol. 21, 2202–2209.

Podlaha, O., Zhang, J., 2010. Pseudogenes and their evolution. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd, Chichester (United Kingdom), pp. 1–8.

Poliseno, L., 2012. Pseudogenes: newly discovered players in human cancer. Sci. Signal. 5, re5.

Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., Pandolfi, P.P., 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465, 1033–1038.

Proudfoot, N., 1980. Pseudogenes. Nature 286, 840–841.

Proudfoot, N.J., Maniatis, T., 1980. The structure of a human alpha-globin pseudogene and its relationship to alpha-globin gene duplication. Cell 21, 537–544.

Singh, N., George, A., Sharma, R., Singla, S.K., Palta, P., Manik, R., Chauhan, M.S., Singh, D., 2012. Characterization of *POU5F1 (OCT4)* gene and its promoter in buffalo ESC-like cells identifies multiple transcription start sites and expression of four pseudogenes. Gene 491, 165–172.

Sudbrak, R., Reinhardt, R., Hennig, S., Lehrach, H., Gunther, E., Walter, L., 2003. Comparative and evolutionary analysis of the rhesus macaque extended MHC class II region. Immunogenetics 54, 699–704.

Svensson, O., Arvestad, L., Lagergren, J., 2006. Genome-wide survey for biologically functional pseudogenes. PLoS Comput. Biol. 2, e46.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., Hannon, G.J., 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature 453, 534–538.

Taniguchi, Y., Katsumata, Y., Koido, S., Suemizu, H., Yoshimura, S., Moriuchi, T., Okumura, K., Kagotani, K., Taguchi, H., Imanishi, T., Gojobori, T., Inoko, H., 1996. Cloning, sequencing, and chromosomal localization of two tandemly arranged human pseudogenes for the proliferating cell nuclear antigen (PCNA). Mamm. Genome 7, 906–908.

Tonner, P., Srinivasasainagendra, V., Zhang, S., Zhi, D., 2012. Detecting transcription of ribosomal protein pseudogenes in diverse human tissues from RNA-seq data. BMC Genomics 13, 412.

Torrents, D., Suyama, M., Zdobnov, E., Bork, P., 2003. A genome-wide survey of human pseudogenes. Genome Res. 13, 2559–2567.

Vanin, E.F., 1985. Processed pseudogenes: characteristics and evolution. Annu. Rev. Genet. 19, 253–272.

Wang, W., Zhang, J., Alvarez, C., Llopart, A., Long, M., 2000. The origin of the *Jingwei* gene and the complex modular structure of its parental gene, *yellow emperor*, in *Drosophila melanogaster*. Mol. Biol. Evol. 17, 1294–1301.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M.A., Sakaki, Y., Sasaki, H., 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 453, 539–543.

Willecke, K., Jungbluth, S., Dahl, E., Hennemann, H., Heynkes, R., Grzeschik, K.H., 1990. Six genes of the human connexin gene family coding for gap junctional proteins are assigned to four different human chromosomes. Eur. J. Cell Biol. 53, 275–280.

Zhang, Z., Carriero, N., Gerstein, M., 2004. Comparative analysis of processed pseudogenes in the mouse and human genomes. Trends Genet. 20, 62–67.

Zhang, Z.D., Frankish, A., Hunt, T., Harrow, J., Gerstein, M., 2010. Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. Genome Biol. 11, R26.

Zhang, Z., Gerstein, M., 2004. Large-scale analysis of pseudogenes in the human genome. Curr. Opin. Genet. Dev. 14, 328–335.

Zhang, Z., Harrison, P., Gerstein, M., 2002. Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome. Genome Res. 12, 1466–1482.

Zhang, Z., Harrison, P.M., Liu, Y., Gerstein, M., 2003. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res. 13, 2541–2558.

Zheng, D., Gerstein, M.B., 2007. The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they? Trends Genet. 23, 219–224.

Zhou, B.S., Beidler, D.R., Cheng, Y.C., 1992. Identification of antisense RNA transcripts from a human DNA topoisomerase I pseudogene. Cancer Res. 52, 4280–4285.