Genetics and Molecular Research

CpG island evolution in the mammalian *DHRS4* gene cluster and its role in the regulation of gene transcription

Z. Su, G. Liu, X. Song, B. Liang, X. Chang and D. Huang

Department of Cell Biology, Shantou University Medical College, Shantou, Guangdong, China

Corresponding author: D. Huang
E-mail: huangdy@stu.edu.cn

Genet. Mol. Res. 15 (2): gmr.15027752
Received September 29, 2015
Accepted December 23, 2015
Published June 3, 2016
DOI http://dx.doi.org/10.4238/gmr.15027752

ABSTRACT. The dehydrogenase/reductase (SDR family) member 4 (*DHRS4*) gene is copied during mammalian evolution; therefore, while only one *DHRS4* gene is expressed in the mouse genome, the gene cluster consists of two (*DHRS4* and *DHRS4L1*) and three (*DHRS4*, *DHRS4L2*, and *DHRS4L1*) copies in chimpanzees and humans, respectively. In this study, we explored the possible regulatory mechanism of the *DHRS4* gene cluster in mammalian evolution by analyzing the promoter sequence, methylation of CpG islands, and RNA expression of the *DHRS4* gene cluster in mice, chimpanzees, and humans by bioinformatics prediction, bisulfite sequencing PCR, and real-time reverse transcriptase-PCR. The results indicated that the *DHRS4* gene was actively expressed in the three model species. The RNA level of *DHRS4L1* was much lower than those of *DHRS4* and *DHRS4L2*, and expressed lower homologous sequence identity to *DHRS4* and *DHRS4L2*. *DHRS4L2*, the latest evolutionary copy of the *DHRS4* gene in mammals, received a high promoter prediction score, and was the only copy of the *DHRS4* gene cluster presenting hypermethylated CpG islands in the promoter region. An analysis of the relationship between

the promoter characteristics and RNA expression of the *DHRS4* gene cluster indicated that the development of CpG islands, in addition to the promoter sequence, during mammalian evolution could modulate the dose compensatory regulation of the copy number-varied *DHRS4* gene cluster.

Key words: Copy number variation; DNA methylation; Transcriptional regulation; *DHRS4* gene cluster; Natural antisense transcription

INTRODUCTION

The dehydrogenase/reductase member 4 gene (*DHRS4*) codes for short-chain dehydrogenase/reductases that participate in the retinoic acid, steroid, and xenobiotic carbonyl compounds in various species, associated with cell proliferation, differentiation, and tumorigenesis (Huang and Ichikawa, 1997; Matsunaga et al., 2008; Endo et al., 2009; Tang and Gudas, 2011). Mammals such as mice, pigs, and members of the *Canis* family possess only one copy of the *DHRS4* gene, while the chimpanzee genome harbors both the *DHRS4* gene and its copy *DHRS4L1*. On the other hand, humans express a *DHRS4* gene cluster comprising three homologous genes: *DHRS4*, *DHRS4L2*, and *DHRS4L1* (Su et al., 2010). The copied *DHRS4* gene cluster is arranged in tandem in chromosome 14 of the chimpanzee and human genome.

Gene duplication is a major source of copy number variations in genes; additionally, gene duplication contributes to the development of complex functions to combat a variety of conditions in various species (Conant and Wolfe, 2008; Elliott et al., 2013). For example, duplications or mutations in globulin reflect the metabolic demands and environmental oxygen availability during the different embryonic or post-birth life stages of an individual, or during the infection of erythrocytes by the malarial parasite (Lam and Jeffreys, 2006; Opazo et al., 2013); duplications in the *AQP7* gene, which codes for a family of water-selective membrane channels, is representative of the adaptations in thermoregulation and energy utilization through biological transmembrane transport (Dumas et al., 2007). Previous research has also suggested that mutations and duplications in the genes could also lead to the development of proteins that are harmful to the species; therefore, there is a high possibility that approximately 90% of the new copied genes is pseudogenized at the epigenetic level, affecting its functionality (Rodin and Riggs, 2003; Katju and Bergthorsson, 2013). Therefore, the fate of the vast majority of gene duplications is detrimental, leading to its inhibition or removal by natural selection.

We have previously reported that the natural antisense transcript *AS1DHS4* regulated the transcription of the *DHRS4* gene cluster through the recruitment of histone modifiers and DNA methyltransferases to control DNA methylation, histone acetylation, and histone methylation in the promoter region of the human *DHRS4* gene cluster (Li et al., 2012). This led us to investigate the various modulations in the *DHRS4* gene cluster in other species, containing different copy numbers of the *DHRS4* gene, and the various factors involved in the regulatory process. In this study, we analyzed mouse, chimpanzee, and human cell lines presenting different *DHRS4* gene copy numbers, to explore the transcriptional activity and possible transcription regulatory mechanism occurring in this gene during mammalian evolution, by bioinformatics analysis, reverse transcriptase quantitative PCR (RT-qPCR), and bisulfite sequencing PCR (BSP).

MATERIAL AND METHODS

Database and software

DNA and RNA sequences of the *DHRS4* gene cluster in mice, chimpanzees, and humans were acquired from the National Center for Biotechnology Information (NCBI) GenBank (http://www.ncbi.nlm.nih.gov) database. The Jellyfish software and the online Basic Local Alignment Search Tool (BLAST) were adapted to analyze the homologous identity of *DHRS4*, *DHRS4L2*, and *DHRS4L1*. The promoter site and activity of the *DHRS4* gene cluster were predicted by PROSCAN (http://www-bimas.cit.nih.gov/molbio/proscan/) (Prestridge, 1995, 2000). Promoter cutoff score was set as 53, and the sequence of approximately 2000 bp from transcriptional start sites were provided for the software online.

The possible epigenetic regulatory mechanism occurring in the *DHRS4* gene cluster was analyzed by investigating the status of the CpG islands in the *DHRS4* gene cluster in mice, chimpanzees, and humans. The CpG island data in the promoter region of the genes of interest was collected from the UCSC Genome browser (http://www.genome.ucsc.edu), which defined the CpG islands as being >500 bp long, containing >55% GC-repeats and an observed CG/expected CpG ratio of 0.65 (Takai and Jones, 2002; Zhao and Han, 2009).

Cell lines and cell culture

The mouse liver cell line NCTC 1469, human hepatocyte cell line HL-7702, and human neuroblastoma cell lines BE(2)-M17 and SK-N-SH were obtained from the cell bank of the Chinese Academy of Sciences (Shanghai, China). The chimpanzee skin fibroblast cell line WES was obtained from American Type Cell Culture association (ATCC, Manassas, VA, USA). The human HL-7702, BE(2)-M17, and SK-N-SH, and chimpanzee WES cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen, Carlsbad, CA, USA) supplemented with 10% (v/v) fetal bovine serum (Invitrogen), in a humidified 37°C incubator, containing 5% CO_2. The mouse liver cell line NCTC 1469 was maintained in DMEM supplemented with 10% (v/v) horse serum (Invitrogen) in humidified 37°C and with 5% CO_2.

RNA extraction and RT-qPCR

Total RNA was extracted from the mouse, chimpanzee, and human cell lines and prepared as described in a previous study (Su et al., 2010). In short, the RNA was extracted using TRIzol® reagent (Invitrogen) and reverse-transcribed with a QuantiTect Reverse Transcription kit (Qiagen, Venlo, Netherlands). qPCR was performed in a PRISM 7300 qPCR machine (Applied Biosystems, Foster City, CA, USA), after preparing the samples using Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen). All qPCR primers are listed in Table 1.

Bisulfite sequencing PCR for DNA methylation analysis

Genomic DNA was extracted using a QuickGene DNA tissue kit (Fujifilm, Tokyo, Japan), and the bisulfite conversion reaction was performed using an EpiTect Bisulfite kit (Qiagen, Venlo, Netherlands). PCR amplification of bisulfate-treated DNA was performed with primers listed in Table 2; the amplified products were then cloned and sequenced for further analysis.

Table 1. Primers used to detect RNA expression of the *DHRS4* gene cluster by reverse transcriptase quantitative polymerase chain reaction (RT-qPCR).

Name	Sequence(5'-3')	Application
mDHRS4-qF	CCCTTTGAACAGTCCCCTCC	RNA expression of mouse *DHRS4* gene
mDHRS4-qR	CGCTTGAGTCCAACCACCTA	
mGAPDH-qF	GCCCTTGAGCTAGGACTGGA	RNA expression of mouse *GAPDH* gene
mGAPDH-qR	CGTCTCTGGAACAGGGAGGA	
cDHRS4-qF	CTGGTCTGAGCCATGCACAA	RNA expression of chimpanzee *DHRS4* gene
cDHRS4-qR	TACCAGGGCCACCTTATTCG	
cDHRS4L1-qF	GCTGCTAGGCCACTGTGC	RNA expression of chimpanzee *DHRS4L1* gene
cDHRS4L1-qR	GCCACCTTATTTGTGAGCGG	
cGAPDH-qF	GCACCGTCAAGGCTGAGAAC	RNA expression of chimpanzee *GAPDH* gene
cGAPDH-qR	TGGTGAAGACGCCAGTGGA	
hDHRS4-qF	AGGCCTCTGTGCCCGGGCTTGGAAT	RNA expression of human *DHRS4L2* gene
hDHRS4-qR	TGCTGCTTCCGGCTGCTGACGACCA	
hDHRS4L1-qF	ATGCACAAGGCGCGGCTACG	RNA expression of human *DHRS4L1* gene
hDHRS4L1-qR	TCGGTGGAGGCCGTTACCAG	
hDHRS4L2-qF	AGGCCTCTGTGCCTGGGCACGGAAG	RNA expression of human *DHRS4L2* gene
hDHRS4L2-qR	TGCTGCTTCCGGCTGCTGACGACCA	
hGAPDH-qF	GCACCGTCAAGGCTGAGAAC	RNA expression of human *GAPDH* gene
hGAPDH-qR	TGGTGAAGACGCCAGTGGA	

Table 2. Primers used to detect DNA methylation in the *DHRS4* gene cluster by BSP.

Name	Sequence (5'-3')	Use
cDHRS4-BSPF	TTTGTAGGGGTTAGGGGAGG	Methylation status of CG island in chimpanzee *DHRS4*
cDHRS4-BSPR	CAACCAAAACCTCAAAAACTCC	
cDHRS4L1-BSPF	GTGGATAAGTAGATGTAGAAGGG	Methylation status of CG island in chimpanzee *DHRS4*-like gene
cDHRS4L1-BSPR	ACAACCAAAACCTCAAAAACTCC	
hDHRS4-UMF	TGAGTTTTATAGGTAATTTGAAT	Unmethylated CG island in human *DHRS4*
hDHRS4-UMR	CAAAAACTCCAACCAAAACA	
hDHRS4 -MF	CGAGTTTTATAGGTAATTTGAAC	Methylated CG island in human *DHRS4*
hDHRS4 -MR	AAAAACTCCGACCGAAAC	
hDHRS4L1-UMF	AAAAGGTGGGTAAGTAGATGTAGAAAG	Unmethylated CG island in human *DHRS4L1* gene
hDHRS4L1-UMR	CAAAAACTCCAACACCAACACTC	
hDHRS4L1- MF	AAAAGGTGGGTAAGTAGATGTAGAAAG	Methylated CG island in human *DHRS4L1* gene
hDHRS4L1- MR	CAAAAACTCCGACACCAACACTC	
hDHRS4L2-UMF	TGAGTTTTATAGGTAATTTGAAT	Unmethylated CG island in human *DHRS4L2* gene
hDHRS4L2-UMR	AAACCTCAAAAACTCCAAC	
hDHRS4L2-MR	AAACCTCAAAAACTCCGAC	Methylated CG island in human *DHRS4L2* gene
hDHRS4L2-MF	CGAGTTTTATAGGTAATTTGAAC	

RESULTS AND DISCUSSION

Three homologous copies of the *DHRS4* gene cluster have different expression levels

The structure of the *DHRS4* gene cluster in mice, chimpanzees, and humans recorded in the GenBank database are presented in Figure 1. Analysis of the homology between the *DHRS4* genes in the reference DNA and RNA of these three mammals using BLAST indicated that the *DHRS4*, *DHRS4L2*, and *DHRS4L1* genes are homologous copies in mice, chimpanzees, and humans. The mouse genome has only one copy of the *DHRS4* gene. *DHRS4L1* is paralogous to *DHRS4* in chimpanzees and humans, while *DHRS4L2*, expressed only in humans, has greater homologous identity to *DHRS4* than to *DHRS4L1*. These three species, expressing different copies of the *DHRS4* gene, provide a good model for the study of the evolutionary history and expression characteristics of duplicated genes.

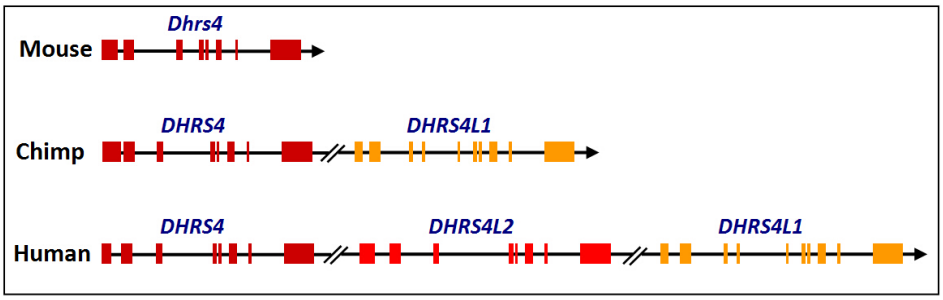

Figure 1. DNA copy number variation of the *DHRS4* gene cluster in mice, chimpanzees, and humans.

Copying of a gene is the process of gene duplication during evolution to increase the size of the genome and its resulting proteome in biological species belonging to higher orders. Recent reports have suggested that DHRS4 modulates the metabolism of vitamin A to retinoic acid, an import molecule that influences the signal transduction affecting cell differentiation during embryo development and tumorigenesis, in rabbits (Huang and Ichikawa, 1997; Matsunaga et al., 2008; Tang and Gudas, 2011). On the other hand, DHRS4 functions as an enzyme in pigs, dogs, and humans because of subtle differences in the substrate (Matsunaga et al., 2008; Endo et al., 2009). Studies have indicated that, in general, genes with copy number variations coding for signal transduction, extracellular biological processes, and some metabolic pathways (such as those affecting xenobiotic compounds) are enriched in different populations (Poptsova et al., 2013). In fact, a homologous copy of the *DHRS4* gene originated as *DHRS2* in species lower than the mammalian class. Unlike the *DHRS4* gene, *DHRS2* underwent rapid evolution, with the resulting protein reductase hep 27 presenting different substrate binding sites compared to DHRS4 (Gabrielli and Tofanelli, 2012). Additionally, DHRS4 is a vital enzyme existing in simple species as well as humans. The evolution to primates led to the development of an additional homologous copy of *DHRS4* (*DHRS4L1*), possibly to face the increasingly complex surroundings. This could also account for the evolution of *DHRS4L2* in humans. *DHRS4* gene duplication during evolution allows each species to code for a greater number of products that perform complex metabolic, or other unknown, functions.

Inherited copy number variations are known to modulate the expression of individual genes (Schlattl et al., 2011). In this study, we analyzed the expression of the *DHRS4* gene cluster with copy number variations in the mouse liver cell line NCTC 1469, chimpanzee skin fibroblast cell line WES, and human hepatocyte (HL-7702) and neuroblastoma (BE(2)-M17 and SK-N-SH) cell lines. RT-qPCR results indicated that the RNA level of *DHRS4* was higher than those of *DHRS4L1* and *DHRS4L2* in chimpanzees and humans; on the other hand, the expression of *DHRS4L1* and *DHRS4L2* were lower than those of the *DHRS4* gene cluster and *DHRS4* and *DHRS4L1*, respectively (Figure 2).

Previous studies have identified isoforms of the human DHRS4 protein (Song et al., 2007; Su et al., 2012); however, the proteins coded by *DHRS4L2* and *DHRS4L1* remain to be identified. Despite this, we did not exclude the possibility that duplicated *DHRS4* genes have other undetected functions, such as the regulation of expression or function of other genes by noncoding RNA. However, this theory remains to be substantiated. In conclusion, the RNA and protein expression level of the *DHRS4* gene cluster indicates that *DHRS4* is the most basic and important among the tested copies of the *DHRS4* gene in mammals.

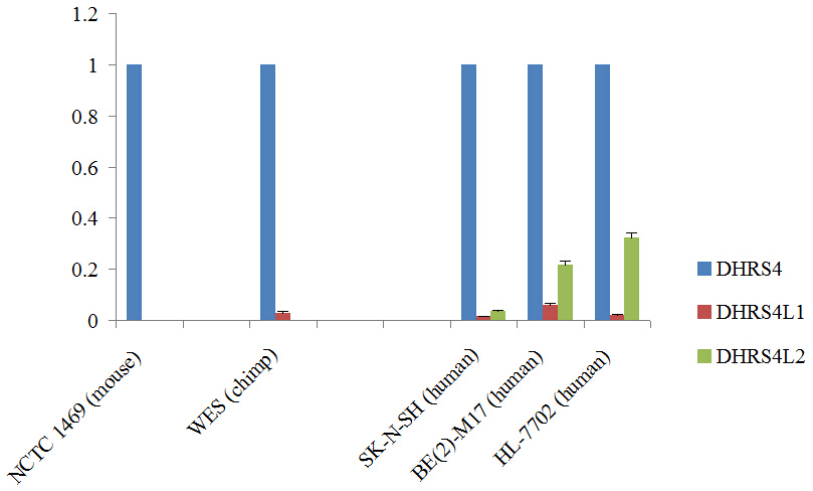

Figure 2. RNA expression of the *DHRS4* gene cluster relative to *GAPDH* in mouse, chimpanzee, and human cell lines, detected by RT-qPCR (means ± SD).

Promoter sequence and status of CpG islands could control the transcriptional activity of the *DHRS4* gene cluster

Different species present different copies and varying expression levels of the *DHRS4* gene cluster. In this study, we also attempted to identify the factors influencing the regulation of *DHRS4* gene cluster expression by predicting the promoter activity of the *DHRS4* gene cluster in mice, chimpanzees, and humans using the online platform PROSCAN.

Analysis of the gene cluster in the three species with PROSCAN led to a relatively high score for *DHRS4* and *DHRS4L2*; that is, the score of the *DHRS4L2* promoter was higher than that of *DHRS4* and *DHRS4L1* (lowest) in humans (Table 3).

Table 3. Promoter activity of the *DHRS4* gene cluster, predicted by PROSCAN.

		DHRS4	*DHRS4L1*	*DHRS4L2*
Mouse	Reverse			
	Forward	55.46		
Chimp	Reverse	53.62, 58.30	54.15	
	Forward	74.72	58.35	
Human	Reverse	53.85, 53.58	59.49, 55.91	53.45, 59.95
	Forward	72.74	57.59	79.89

Approximately 70% of the promoters have been correctly recognized by PROSCAN in a previous study (Prestridge, 1995). The PROSCAN prediction attributes the low expression of *DHRS4L1* RNA among the different *DHRS4* copies to the long heritage distance between *DHRS4* and *DHRS4L2* homologous identity. Interestingly, *DHRS4L2* received a high PROSCAN prediction score, despite the low *DHRS4L2* RNA expression compared to that of *DHRS4* in humans, indicating the role of other factors in suppressing the promoter activity of *DHRS4L2*.

Epigenetic regulation is a common method for the control of gene transcriptional activity without changing the DNA sequence; this includes CpG island methylation and

histone modification in the promoter region. Approximately 40% of genes contain CpG islands, which are often, but not always, found in the promoter regions. The CpG island criteria (Takai and Jones, 2002) indicate that these are present in each copy of *DHRS4* in chimpanzees and humans, and absent in mice. Bisulfate-treated methylation-specific PCR revealed that *DHRS4L2* is hypermethylated in humans, while *DHRS4* and *DHRS4L1* in humans and chimpanzees are hypomethylated (Figure 3). DNA modification by methylation is an important epigenetic mechanism that affects the spatial and temporal regulation of gene expression; in fact, promoter hypermethylation can silence the gene expression (Miranda and Jones, 2007).

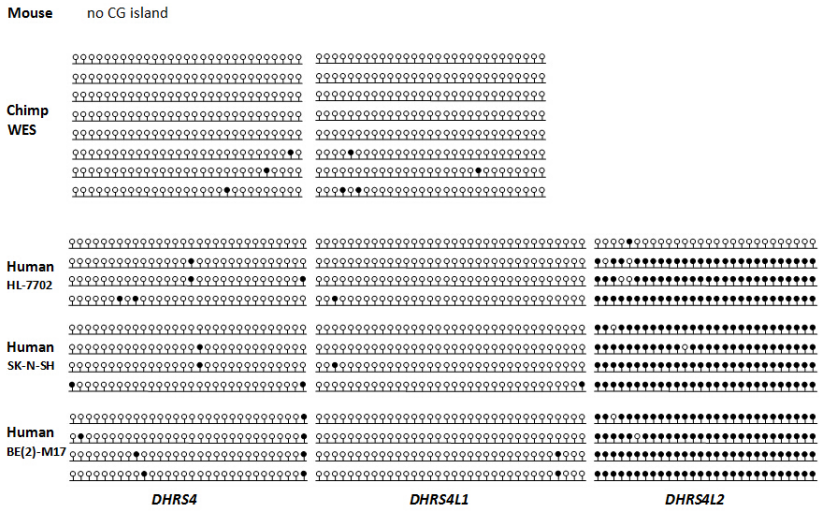

Figure 3. Methylation status of the CpG islands in the *DHRS4* gene cluster in mouse, chimpanzee, and human cell lines. Filled circles represent methylated CpG sites and open circles represent unmethylated CpG sites.

Therefore, hypermethylation of the CpG islands in the *DHRS4L2* promoter region could account for the high promoter score and low RNA expression of *DHRS4L2* compared to *DHRS4*. These results are consistent with the results of previous studies, wherein the promoter regions of young duplicates are hypermethylated, whereas those of old duplicates are generally hypomethylated (Keller and Yi, 2014). This is because, as a gene that has undergone duplication for a longer period would have developed a vital function for the organism through evolution or rapid mutation, the CpG island might not even exist in this region, similar to the development of *DHRS2* prior to the evolution of mammals.

On the one hand, gene duplication plays a vital role in the biological evolution of new functions and flexible regulation; on the other hand, it may also cause gene dosage imbalance, which could affect the survival or compatibility of the individual with the surroundings. Therefore, certain newly developed genes produce RNA or protein products and subsequently affect the functionality of an organism, while others simply undergo mutation and develop into pseudogenes (Rodin and Riggs, 2003). In this study, we speculated that the duplicated *DHRS4L1* has low homologous identity to *DHRS4* in chimpanzees and humans. PROSCAN predicted that the transcription activity of *DHRS4L1* is much lower than that of *DHRS4* even without the methylation of CpG islands in *DHRS4L1*; this was also validated by the RT-qPCR results. *DHRS4L2*, which has a high homologous identity to *DHRS4* in humans, showed a

high PROSCAN score for promoter activity, despite the inhibition of promoter activity by DNA methylation. Copy number variations associated with genome-wide DNA methylation are accumulated in gene clusters. DNA methylation contributes to dosage rebalance during mammalian evolution after gene duplication by inhibiting the transcription of copied genes. This hypermethylation of DNA protects the duplicates from "pseudogenization" (Chang and Liao, 2012; Mendizabal et al., 2014). However, we have been unable to predict the possible function of suppressed or active demethylated *DHRS4L2*. The *DHRS4* gene in mice does not express any CpG islands, because of the expression of only one copy of the gene and the established role of the resulting protein in the synthesis of retinoic acid. The homologous identity and methylation status of the *DHRS4* gene cluster is consistent with the theory that CpG islands are developed during the duplication of the *DHRS4* gene, and that CpG island methylation potentially regulates the dose balance of copy number variation during mammalian evolution. Epigenetic mechanisms other than DNA methylation, such as histone acetylation and methylation, are also observed throughout the evolutionary tree (Simonti and Capra, 2015); therefore, we theorized that complex factors besides the development and methylation of CpG islands could play a role in controlling the fate of the copied *DHRS4* gene cluster in mammals.

Antisense RNA might regulate the methylation of CpG islands in the *DHRS4* gene cluster

With regards to the epigenetic regulation of the *DHRS4* gene cluster, we have previously discovered that the natural antisense transcript of *DHRS4* could mediate the DNA methylation and histone modification by recruiting DNA methyltransferases and histone methyltransferase to sense *DHRS4* gene clusters in humans (Li et al., 2012). Approximately 5-30% of the transcriptional units in diverse eukaryotes have been found to harbor a cis-antisense transcript. Antisense regulation is an important regulation mechanism in the human genome, with partial antisense RNA being regarded as an evolved repressor (Chen et al., 2005). Genome-wide analyses have indicated the presence of a relationship between sense promoter methylation and antisense RNA expression; that is, the expression of a transcript from a methylated CpG island-bearing promoter has been reported to be conversely proportional to the expression of the opposite-strand transcript (Watanabe et al., 2010; Lin et al., 2013). Knockdown of antisense RNA of the human *DHRS4* gene cluster in a previous study resulted in the demethylation of *DHRS4L2* CpG islands; moreover, increased RNA expression indicated the presence of an association between the expression of natural antisense transcripts and CpG islands and the evolution of *DHRS4* copies, and was believed to be involved in the dosage balance of the gene cluster. PROSCAN results indicated the absence of any predicted promoter activity in the antisense strand of the mouse *DHRS4* gene. The analysis predicted a potential promoter for the antisense strand for each promoter region of the *DHRS4* gene cluster in human and chimpanzee genomes (Table 3). Therefore, we speculated that, antisense transcripts could develop alongside the formation of CpG islands during evolution, in addition to the positive correlation between the CpG islands and copy number variation of *DHRS4* gene cluster in mammals, resulting in their role in the regulatory mechanism (Figure 4).

In summary, our results indicate that the duplicated genes *DHRS4L2* and *DHRS4L1* are copied after the evolution of species to primates and humans. CpG islands and antisense transcripts are developed in the promoter region in association with gene duplication, which in turn allow for the regulation of DNA methylation and transcriptional activity, in order to

balance the dosage of *DHRS4* gene cluster in different species that express different copies of the *DHRS4* gene cluster. However, the exact function of the duplicates, and their role in gene regulation, remains to be verified.

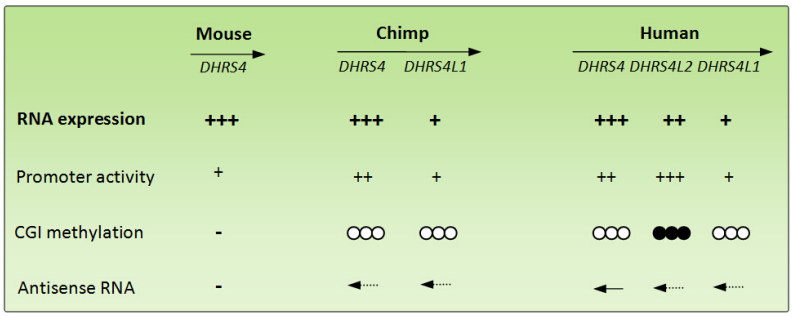

Figure 4. DNA methylation regulates the transcriptional activity of *DHRS4* copy genes in mammalian evolution, mediated by antisense RNA. Promoter activity prediction indicates that the activity of the *DHRS4L2* promoter is greater than that of *DHRS4* in humans. On the other hand, the RNA expression level (detected by RT-qPCR) of *DHRS4L2* is relatively lower than that of *DHRS4*. The CpG islands in *DHRS4* and *DHRS4L1* were hypomethylated, while those in *DHRS4L2* were hypermethylated.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research funded by the National Natural Science Foundation of China (#31100939 and #31371333) and the Natural Science Foundation of Guangdong Province (#S2012030006289).

REFERENCES

Chang AY and Liao BY (2012). DNA methylation rebalances gene dosage after mammalian gene duplications. *Mol. Biol. Evol.* 29: 133-144. http://dx.doi.org/10.1093/molbev/msr174

Chen J, Sun M, Hurst LD, Carmichael GG, et al. (2005). Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. *Trends Genet.* 21: 326-329. http://dx.doi.org/10.1016/j.tig.2005.04.006

Conant GC and Wolfe KH (2008). Probabilistic cross-species inference of orthologous genomic regions created by whole-genome duplication in yeast. *Genetics* 179: 1681-1692. http://dx.doi.org/10.1534/genetics.107.074450

Dumas L, Kim YH, Karimpour-Fard A, Cox M, et al. (2007). Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res.* 17: 1266-1277. http://dx.doi.org/10.1101/gr.6557307

Elliott KT, Cuff LE and Neidle EL (2013). Copy number change: evolving views on gene amplification. *Future Microbiol.* 8: 887-899. http://dx.doi.org/10.2217/fmb.13.53

Endo S, Maeda S, Matsunaga T, Dhagat U, et al. (2009). Molecular determinants for the stereospecific reduction of 3-ketosteroids and reactivity towards all-trans-retinal of a short-chain dehydrogenase/reductase (DHRS4). *Arch. Biochem. Biophys.* 481: 183-190. http://dx.doi.org/10.1016/j.abb.2008.11.014

Gabrielli F and Tofanelli S (2012). Molecular and functional evolution of human DHRS2 and DHRS4 duplicated genes. *Gene* 511: 461-469. http://dx.doi.org/10.1016/j.gene.2012.09.013

Huang DY and Ichikawa Y (1997). Purification and characterization of a novel cytosolic NADP(H)-dependent retinol oxidoreductase from rabbit liver. *Biochim. Biophys. Acta* 1338: 47-59. http://dx.doi.org/10.1016/S0167-4838(96)00183-5

Katju V and Bergthorsson U (2013). Copy-number changes in evolution: rates, fitness effects and adaptive significance. *Front. Genet.* 4: 273. http://dx.doi.org/10.3389/fgene.2013.00273

Keller TE and Yi SV (2014). DNA methylation and evolution of duplicate genes. *Proc. Natl. Acad. Sci. USA* 111: 5932-5937. http://dx.doi.org/10.1073/pnas.1321420111

Lam KW and Jeffreys AJ (2006). Processes of copy-number change in human DNA: the dynamics of alpha-globin gene deletion. *Proc. Natl. Acad. Sci. USA* 103: 8921-8927. http://dx.doi.org/10.1073/pnas.0602690103

Li Q, Su Z, Xu X, Liu G, et al. (2012). AS1DHRS4, a head-to-head natural antisense transcript, silences the DHRS4 gene cluster in cis and trans. *Proc. Natl. Acad. Sci. USA* 109: 14110-14115. http://dx.doi.org/10.1073/pnas.1116597109

Lin CH, Tsai ZT and Wang D (2013). Role of antisense RNAs in evolution of yeast regulatory complexity. *Genomics* 102: 484-490. http://dx.doi.org/10.1016/j.ygeno.2013.10.008

Matsunaga T, Endo S, Maeda S, Ishikura S, et al. (2008). Characterization of human DHRS4: an inducible short-chain dehydrogenase/reductase enzyme with 3beta-hydroxysteroid dehydrogenase activity. *Arch. Biochem. Biophys.* 477: 339-347. http://dx.doi.org/10.1016/j.abb.2008.06.002

Mendizabal I, Keller TE, Zeng J and Yi SV (2014). Epigenetics and evolution. *Integr. Comp. Biol.* 54: 31-42. http://dx.doi.org/10.1093/icb/icu040

Miranda TB and Jones PA (2007). DNA methylation: the nuts and bolts of repression. *J. Cell. Physiol.* 213: 384-390. http://dx.doi.org/10.1002/jcp.21224

Opazo JC, Butts GT, Nery MF, Storz JF, et al. (2013). Whole-genome duplication and the functional diversification of teleost fish hemoglobins. *Mol. Biol. Evol.* 30: 140-153. http://dx.doi.org/10.1093/molbev/mss212

Poptsova M, Banerjee S, Gokcumen O, Rubin MA, et al. (2013). Impact of constitutional copy number variants on biological pathway evolution. *BMC Evol. Biol.* 13: 19. http://dx.doi.org/10.1186/1471-2148-13-19

Prestridge DS (1995). Predicting Pol II promoter sequences using transcription factor binding sites. *J. Mol. Biol.* 249: 923-932. http://dx.doi.org/10.1006/jmbi.1995.0349

Prestridge DS (2000). Computer software for eukaryotic promoter analysis. *Methods Mol. Biol.* 130: 265-295.

Rodin SN and Riggs AD (2003). Epigenetic silencing may aid evolution by gene duplication. *J. Mol. Evol.* 56: 718-729. http://dx.doi.org/10.1007/s00239-002-2446-6

Schlattl A, Anders S, Waszak SM, Huber W, et al. (2011). Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions. *Genome Res.* 21: 2004-2013. PubMed http://dx.doi.org/10.1101/gr.122614.111

Simonti CN and Capra JA (2015). The evolution of the human genome. *Curr. Opin. Genet. Dev.* 35: 9-15. http://dx.doi.org/10.1016/j.gde.2015.08.005

Song XH, Liang B, Liu GF, Li R, et al. (2007). Expression of a novel alternatively spliced variant of NADP(H)-dependent retinol dehydrogenase/reductase with deletion of exon 3 in cervical squamous carcinoma. *Int. J. Cancer* 120: 1618-1626. http://dx.doi.org/10.1002/ijc.22306

Su Z, Li R, Song X, Liu G, et al. (2012). Identification of a novel isoform of DHRS4 protein with a nuclear localization signal. *Gene* 494: 161-167. http://dx.doi.org/10.1016/j.gene.2011.12.033

Su ZJ, Zhang QX, Liu GF, Song XH, et al. (2010). Bioinformatic analysis of the human DHRS4 gene cluster and a proposed mechanism for its transcriptional regulation. *BMC Mol. Biol.* 11: 43. http://dx.doi.org/10.1186/1471-2199-11-43

Takai D and Jones PA (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. USA* 99: 3740-3745. http://dx.doi.org/10.1073/pnas.052410099

Tang XH and Gudas LJ (2011). Retinoids, retinoic acid receptors, and cancer. *Annu. Rev. Pathol.* 6: 345-364. http://dx.doi.org/10.1146/annurev-pathol-011110-130303

Watanabe Y, Numata K, Murata S, Osada Y, et al. (2010). Genome-wide analysis of expression modes and DNA methylation status at sense-antisense transcript loci in mouse. *Genomics* 96: 333-341. http://dx.doi.org/10.1016/j.ygeno.2010.08.007

Zhao Z and Han L (2009). CpG islands: algorithms and applications in methylation studies. *Biochem. Biophys. Res. Commun.* 382: 643-645. http://dx.doi.org/10.1016/j.bbrc.2009.03.076