ORIGINAL ARTICLE

Acquisition of inverted GSTM exons by an intron of primate GSTM5 gene

Yong Wang and Frederick CC Leung

The human GSTM gene family is composed of five gene members, GSTM1–5, and plays an important role in detoxification. In this study, the human GSTM5 gene was found to have a long inverted repeat (LIR) in intron 5. The LIR is able to form a stem-loop structure with a 31-bp stem and a 9-nt loop. The intronic LIR was also identified in other primates but not in non-primates. The human and chimpanzee LIRs had undergone compensating mutations that make the stem loop more stable, suggesting a functional role for the LIR. Sequence homology showed that the LIR was actually a part of inverted exons acquired by the intron. Results of phylogenetic analysis indicate that the inverted exons were derived from exon 5 of GSTM4 and exon 5 of GSTM1. The intronic LIR and inverted GSTM exons can probably introduce complexity in the expression of GSTM gene family.

Journal of Human Genetics (2009) 54, 271–276; doi:10.1038/jhg.2009.23; published online 20 March 2009

Keywords: GSTM; inverted repeat; primate; intronic stem loop

INTRODUCTION

Inverted repeats (IRs) are unstable motifs capable of inducing recombination, gene amplifications and rearrangements in a genome.^{1–5} On the other hand, a considerable number of IRs are functional elements in eukaryotes. Long IRs (LIRs; >22 bp for one copy) in microRNA can fold into a hairpin. Processed by a dicer protein, the hairpin eventually becomes a small interfering RNA (siRNA) for RNA interference.^{6,7} In other experiments, intronic IRs were shown to affect exon–intron splicing efficiency and determine alternative exon splicing.^{8,9} More intriguingly, we found that some intronic LIRs are primate-specific, and probably critical in the evolution toward primates.¹⁰ In this study, we report one of the cases: an intronic LIR in primate GSTM5 gene.

The GSTM gene family contains five genes, GSTM1–5 in humans, and encodes one of the eight distinct classes of glutathione transferases (GST).^{11–13} The enzyme produced by GSTM genes functions in the detoxification of electrophilic compounds, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress, by conjugation with glutathione.¹⁴ The five human GSTM genes are organized in a gene cluster on chromosome 1p13.^{3,12,13} and are well known to be highly polymorphic.^{15–17} About 50% of the human population carries polymorphic deletions for GSTM1 gene (Xu *et al.*¹⁸ and the references therein). The variants of the genes have been tightly linked to susceptibility to carcinogens and toxins, as well as to toxicity and efficacy of certain drugs.^{19–21} The malfunction of this gene family accounts for many human diseases, including cancers and pulmonary asbestosis.^{18,22} The gene family is a promising candidate for studies of the genetic variance and tissue-specific

expression profile. Although most of the genes were well characterized, the variations in introns had not yet been surveyed sufficiently. Recent reports have shown the importance of intronic conserved elements.^{23,24}

In this study, we performed bioinformatic analyses aiming to study the origin of the intronic LIR of GSTM5 gene and cast light on its significance in expression variance of GSTM genes. We collected the LIRs (no matter full- or half-sized) and their flanking sequences in seven mammalian genomes. One copy of the full-sized LIR was found in the genomes of rhesus monkey, orangutan, chimpanzee and human. By contrast, all the collected sequences in marmoset, mouse and dog genomes were homologous to one arm of the LIR. Phylogenetic relationship between the collected arms and multiple alignment of their flanking sequences showed that the left arm was derived from exon 5 of GSTM4 and the right arm was highly similar to exon 5 of GSTM1. The LIR is actually within inverted exons that were firstly formed in GSTM1 genes and acquired by the fifth intron of the GSTM5 gene later. The LIR in primates is probably under positive selection and therefore of potential importance in regulating the expression of GSTM gene family.

MATERIALS AND METHODS

We obtained full-length sequences of the genes of GSTM family from the NCBI (human build 36.3; <http://www.ncbi.nlm.nih.gov>). One LIR found in an intron of GSTM5 was identified with our program described elsewhere.²⁵ The sequences in high homology with the LIR were BLAT searched across the other mammalian genomes in the UCSC browser (<http://genome.ucsc.edu>). The species and its genome version for the searching are human genome NCBI

Table 1 Position of LIRs or their arms

Position	Start	End	Length (bp)	Similarity (%)	Gene
human_chr1	110058027	110058097	71	100	GSTM5 intron 5
human_chr1	110057815	110057855	41	100	GSTM5 exon 5
human_chr1	110001926	110001966	41	100	GSTM4 exon 5
human_chr1	110033379	110033419	41	100	GSTM1 exon 5
human_chr3	12274645	12274685	41	97.6	—
human_chr6	111475109	111475149	41	88	—
chimp_chr1	128075823	128075893	71	100	GSTM5 intron 5
chimp_chr1	128076403	128076443	41	100	GSTM5 exon 5
chimp_chr1	128100420	128100460	41	100	GSTM1 intron 5
chimp_chr1	128102282	128102322	41	97.6	GSTM1 exon 5
chimp_chr1	111249896	111249936	41	97.6	GSTM4 exon 5
chimp_chr3	12579979	12580019	41	97.6	—
Orangutan_chr1	118559827	118559886	71	98.4	GSTM5 intron 5
Orangutan_chr1	118560058	118560098	41	97.6	GSTM5 exon 5
Orangutan_chr1	118609049	118609089	41	100	GSTM4 exon 5
Orangutan_chr1	118578793	118578833	41	100	GSTM1 exon 5
Orangutan_chr3	57692771	57692811	41	97.6	—
Rhesus_chr1	112765740	112765810	71	91.6	GSTM5 intron 5
Rhesus_chr1	112765528	112765568	41	97.6	GSTM5 exon 5
Rhesus_chr1	112719724	112719764	41	100	GSTM4 exon 5
Rhesus_chr1	112750171	112750211	41	92.7	GSTM1 exon 5
Rhesus_chr2	48805562	48805602	41	100	—
Marmoset_Contig5186	67526	67566	41	100	GSTM*
Marmoset_Contig6612	10967	11007	41	100	GSTM*
Marmoset_Contig6612	39657	39697	41	100	GSTM*
Mouse_chr3	107788018	107788058	41	95.2	GSTM2 exon 5
Mouse_chr3	107833684	107833724	41	95.2	—
Mouse_chr18	31979864	31979904	41	95.2	WDR33 intron 1
Mouse_chr3	107846293	107846325	33	97	GSTM4 exon 5
Dog_chr6	45266185	45266219	35	100	GSTM* exon
Dog_chr5	50897046	50897080	35	97.2	GSTM* exon

Abbreviation: LIR, long inverted repeat.

The 71-bp sequences are intronic LIRs of GSTM5 in primate genomes, and the 41-bp sequences are the LIRs lacking one arm. Those shorter than 41 bp contain just one arm of the LIR. The similarities were obtained by BLAT search using the LIR in human GSTM5 gene as a reference. *GSTM genes that have not been fully characterized at present.

Build 36.1, chimpanzee genome panTro2, orangutan genome ponAbe2, rhesus monkey genome rheMac2, marmoset genome callJac1, mouse genome NCBI Build 37 and dog genome canFam2. Table 1 shows the positions and genes in which the homologous sequences are located. The flanking 100-bp sequences were also collected for multiple alignment with ClustalW,²⁶ followed by manual adjustment. As one arm of the LIR was homologous to exons of the GSTM genes, we performed a phylogenetic analysis to show the relationship between the arms and the exons. The arms were extended at flanking regions to find their corresponding exons. A total of about 82 sites in exon 5 of GSTM genes and other homologous fragments were obtained from multiple alignment and then used for phylogenetic analysis. Reconstruction of maximum-likelihood phylogeny was performed by the dnaml in Phylip package 3.6,²⁷ and an unrooted tree was drawn in MEGA3.0.²⁸

RESULTS

We found an LIR in the fifth intron of the human GSTM5 gene. The 71-bp LIR (G+C%=50%) is composed of 31-bp arms and a 9-bp internal spacer. The arms are highly complementary and thus tend to form a strong stem-loop structure with only one mispair (Figure 1). Also, in GSTM5 gene, we identified LIRs in other primates, including chimpanzee, orangutan and rhesus monkey. The stem-loop structures are highly similar to those in the human GSTM5 gene, although there are five mispairs on the stem for the rhesus LIR and one insertion for the orangutan LIR (Figure 1). They are all in the same intron of GSTM5 gene. Moreover, we also found fragments that are

homologous to one arm of the human LIR, largely in exons of GSTM genes and also in introns of chimpanzee GSTM1 and other unknown genes (Table 1). The LIRs seem to be unique to primates because we could not find them in genomes of non-primate mammals, such as marmoset, mouse and dog.

As the LIRs are homologous to a part of exon 5 of GSTM genes, we studied the relationship between the LIRs and the exons. The length of exon 5 is 93 bp, the same for all the primate GSTM genes. Within GSTM5, exon 5 of GSTM5 is 70-bp upstream of the LIR (Figure 2). The combination of one arm (31 bp) and the internal spacer (10 bp) is highly homologous to the last 41 bp of the fifth exons of the GSTM1, GSTM4 and GSTM5 genes. To verify that the arms of the LIR are within the earlier exons of GSTM genes, we obtained the flanking sequences of the LIR and compared them with the fifth exons of GSTM genes. Results show that the flanking sequences are indeed homologous to a part of exon 5 of a certain GSTM gene. Primarily, we concluded that the LIRs are actually two exons in different orientations (inverted exons). As the LIR was only exhibited in primates, it was perhaps a result of genomic insertion. We aligned the regions in which the LIR was located between human GSTM1 and GSTM5. The result clearly showed that the inverted exons are within a large inserted fragment at the downstream of GSTM5 exon 5. We then named the exon in which the left arm resides as the first inserted exon, and the one in which the right arm resides as the second one.

Figure 1 Stem-loop structures of LIRs in primates. The stem-loop structures formed by LIRs were all located in primate GSTM5 genes. The positions of the LIRs are referred in Table 1.

Figure 2 Schematics of GSTM5 and its LIR. The figure shows the LIR on the human GSTM5 gene. The location of the GSTM gene family on chromosome 1 is given, and the LIR is within the fifth intron of the GSTM5 gene. The two blank frames represent the inverted exons from GSTM4 and GSTM1, respectively. The arrows indicate the length of the arms of the LIR and transcription direction before their insertion. The distance between GSTM5 exon 5 and the LIR is 70bp.

To know the origin of the inverted exons, we then reconstructed a phylogenetic tree using the real and inserted exons. Our unrooted maximum-likelihood tree shows that the first inserted exon and

GSTM4 exon 5 are within the same clade, and the second inserted exon is clustered with GSTM1 exon 5 (Figure 3). This is the fact for all the primates and their inserted exons. Given the short distance, the exon 5 of GSTM5, interestingly, does not have the closest relationship with the inverted exons. The relationship, on the basis of phylogenetic analysis, was further supported by sequence homology between the downstream regions of the real and inserted exons (the upstream of the LIR is the downstream of the first inserted exon as shown in Figure 2). We found that more than 49bp downstream positions were highly homologous between GSTM1 exon 5 and the second inserted exon for all the primates (Figure 4). In humans, the alignment of the downstream sequences was largely maintained in more than 1900bp. The homology between GSTM4 exon 5 and the first inserted exon was clearly observed at 32 downstream positions. The homology disappears within 10 positions when the inserted exons were compared with GSTM5 exon 5 (Figure 4). It is worth remembering that the distance between GSTM5 exon 5 and the first inserted exon is 70bp. After removal of the above homologous positions, the remaining fragment is about 30bp. We found that it was a foreign

Figure 3 Phylogenetic classification of GSTM5 exon 5 and their homologous sequences. The phylogenetic unrooted tree was reconstructed using the maximum-likelihood algorithm. The tree is a bootstrap consensus tree based on 1000 replicates. The scale bar corresponds to 0.05 nucleotide substitution per site. The LIR sequences were split into left arm (reversely complementary and terminated with an 'L' in names) and right arm (terminated with an 'R' in names). The inverted exons covering the arms were used to reconstruct a phylogenetic tree with other homologous exonic and intronic sequences. The positions of the sequences are referred in Table 1. The abbreviated species names are hu for human; ch for chimpanzee; or for orangutan; rh for rhesus; ma for marmoset; mo for mouse; and do for dog.

fragment with an unknown source because homologous fragments were not found in other GSTM genes and even in the whole primate genomes.

As the two inserted exons are derived from different GSTM genes, they are not completely complementary. The conserved parts (41 bp) of the exons from GSTM1 and GSTM4 were used to form the LIR.

The rest of the regions in less homology were not used to form a larger LIR under our stringent criteria. If we had used relaxed criteria for LIR search, the LIR in GSTM5 intron 5 could have covered the whole inserted exons. The two exons actually overlap 9 bp at 5'-TCCTCTTCT-3', which is the internal spacer of the LIR. The overlapping part belongs to the first inserted exon, and thus the

Figure 4 Alignment of flanking sequences. We obtained upstream sequences of the left arm of the LIR and downstream sequences of the right arm. As the left arm is homologous to the antisense sequences of exon 5 of GSTM genes, we collected downstream sequences of real exons. The shaded sequences, if not in the LIR and intron, are a part of exon 5 of the GSTM genes, but on the antisense strand. The abbreviations of the species names are referred in Figure 3. The LIR in GSTM5 intron 5 is indicated as GSTM5_i5; exon 5 of GSTM genes is indicated as e5. The left and right arms are denoted by L and R at the end of the names.

second inserted exon lost its first 9 bp. In addition, the original upstream regions of both of the inserted exons are absent.

DISCUSSION

In this study, we discovered an intronic LIR in GSTM5 gene. Further analyses showed that the LIR was within inverted GSTM exons. The presence of the inverted exons is not likely a sort of genomic polymorphism shown in a small population of humans because they present in all the primate genomes studied. The inverted exons in the intron will probably not be spliced into the mRNAs for translation because (1) they tend to form a strong stem-loop structure, and (2) the upstream splicing site disappeared and the first 9 bp of the second inserted exon was occupied by the internal spacer.

The characteristics of the inverted exons and its flanking sequences allow us to postulate that the inverted exons were initially formed in GSTM1 and transferred to GSTM5 later. The whole process had been completed in the genome of the common ancestor of all primates. The formation of the inverted exons in GSTM1 is convincingly supported by the sequence homology between the downstream sequences of GSTM1 exon 5 and GSTM5 inverted exons. If it were formed in GSTM4, a small fragment belonging to the downstream of GSTM4 exon 5 would probably be exhibited downstream of the GSTM5 inverted exons. Moreover, the downstream alignment between the first inserted exon and GSTM4 exon 5 was broken at 32 bp (Figure 4). Together with the foreign fragment at the further downstream region, we believe that the inverted exons were not directly derived from GSTM4. The foreign fragment was thought to present in the GSTM1 gene of the common ancestor, and was taken into GSTM5 along with the transmission of the inverted exons. Under this assumption, the insertion site of the inverted exons would be at several nucleotides downstream of GSTM5 exon 5.

The mechanism accounting for the formation of the inverted exons is unknown at present. The abnormal recombination between GSTM1 exon 5 and GSTM4 exon 5 is probably responsible for introducing a reverse duplication of GSTM4 exon 5 into GSTM1 gene. In the process, the first 9 bp and the splicing site of GSTM5 exon 5 were

destroyed and thus the GSTM5 gene would malfunction theoretically. Along with the exon, a fragment downstream of the GSTM4 exon 5 was also inserted into GSTM1, which rationalizes the sequence homology that we observed downstream of the first inserted exon. The inverted exons on GSTM1 would form a large stem-loop structure with high potential. Owing to probable malfunction, this copy of GSTM1 would be under relaxed natural selection. The unstable inverted exons were easily knocked out and could be inserted elsewhere. As we exhibited in this study, the inverted exons were acquired by GSTM5. The acquisition was supposed to have evolutionary significance in primate speciation because it was maintained in primate genomes in spite of its unstable nature. Interestingly, the stem structure turns out to be more stable in humans than in rhesus monkeys (Figure 1). This is regarded as a process of compensating mutations, inferring the functional role of the stable stem-loop structure.

On the basis of the above results, we propose potential approaches through which the inverted exons and the LIR are involved in regulating the expression of GSTM gene family in more precision. RNA interference is a novel gene-regulation mechanism discovered in 1998.^{29,30} In the process, mRNA is bound by siRNA in length of 21–23 bp that is antisense to a part of the mRNA sequence. The binding of the siRNA triggers several steps for the digestion of mRNA and silences the expression of the target gene.^{6,7} The siRNA in this process comes from a stem-loop RNA produced by a microRNA gene. Some of the microRNA genes are located in introns and are associated with their own promoters.³¹ Also, researchers design to manually insert small IRs into introns. After the splicing process, the intronic IRs are transported into the cytoplasm and then cut into siRNA to silence the expression of the target gene.³² Exons and promoters are generally selected as target sites for binding of the siRNA. The GSTM5 gene in this study has a natural intronic IR potentially capable of regulating the expression of the GSTM gene family, because one arm of the LIR has a matching target at the exons of the GSTM genes. We used siRNA selection program to predict candidate siRNA on the LIR (Figure 5).³³ There are perhaps more candidates if the inverted exons were taken

sense 5'		GUUCUCCAAAAUGUCCACA	UU
mRNA:	TG	GTTCTCCAAAATGTCCACA	CG
antisense 3'	UU	CAAGAGGUUUUACAGGUGU	

Figure 5 Predicted siRNA sequence using mRNA of LIR in intron 5 of GSTM5. The siRNA sequence was predicted using WI siRNA selection server (<http://jura.wi.mit.edu/bioc/siRNAext/>).

into consideration. The hypothetical siRNA obtained from the intronic LIR will find the target on mature mRNAs of GSTM1, GSTM4 and GSTM5 genes, and finally result in the digestion of mRNAs. All the steps form a negative feedback loop for regulating the expression of GSTM genes. The transcription and subsequent mRNA splicing of GSTM5 gene accumulate siRNAs that are able to, in turn, silence the further expression of the GSTM5 gene, as well as GSTM1 and GSTM4 genes. The expression level of the GSTM5 gene is determinant to suppression degree of the three GSTM genes, serving as a 'buffer-like' mechanism controlling GSTM gene expression. If this hypothesis is verified by experiments, the study will help to expand our knowledge about RNA interference and understand a primate-specific mechanism for gene regulation in a more precise style.

The inverted exons in GSTM5 gene may also affect the usage of exon 5. The first inserted exon is close and similar to exon 5 as shown in this study. On account of the head-to-head arrangement, they can form a stem-loop structure as well. At the mRNA level, this will introduce complexity in splicing and difficulty in the inclusion of exon 5. A probable case is EST BM805585 obtained from human hippocampus tissue. The second inserted exon was used as an additional exon relative to other splicing variants. Finally, further experimental studies are required to evaluate the biological significance of the LIR and inverted exons discovered in this bioinformatic study.

- 1 Tanaka, H., Tapscott, S. J., Trask, B. J. & Yao, M.-C. Short inverted repeats initiate gene amplification through the formation of a large DNA palindrome in mammalian cells. *Proc. Natl. Acad. Sci. USA* **99**, 8772–8777 (2002).
- 2 Lin, C.-T., Lin, W.-H., Lyu, Y. L. & Whang-Peng, J. Inverted repeats as genetic elements for promoting DNA inverted duplication: implications in gene amplification. *Nucleic Acids Res.* **29**, 3529–3538 (2001).
- 3 Lobachev, K. S., Shor, B. M., Tran, H. T., Taylor, W., Keen, J. D., Resnick, M. A. *et al.* Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*. *Genetics* **148**, 1507–1524 (1998).
- 4 Lobachev, K. S., Stenger, J. E., Kozyreva, O. G., Jurka, J., Gordenin, D. A. & Resnick, M. A. Inverted Alu repeats unstable in yeast are excluded from the human genome. *EMBO J.* **19**, 3822–3830 (2000).
- 5 Voineagu, I., Narayanan, V., Lobachev, K. S. & Mirkin, S. M. Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc. Natl. Acad. Sci. USA* **105**, 9936–9941 (2008).
- 6 Montgomery, M. K., Xu, S. & Fire, A. RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **95**, 15502–15507 (1998).
- 7 Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281 (2004).
- 8 Miyaso, H., Okumura, M., Kondo, S., Higashide, S., Miyajima, H. & Imaizumi, K. An intronic splicing enhancer element in survival motor neuron (SMN) pre-mRNA. *J. Biol. Chem.* **278**, 15825–15831 (2003).

- 9 Chen, Y. & Stephan, W. (2003) Compensatory evolution of a precursor messenger RNA secondary structure in the *Drosophila melanogaster* Adh gene. *Proc. Natl. Acad. Sci. USA* **100**, 11499–11504.
- 10 Wang, Y. & Leung, F. C. C. A study on genomic distribution and sequence features of human long inverted repeats species-specific intronic inverted repeats. *FEBS J* (in press) (2009).
- 11 Sheehan, D., Meade, G., Foley, V. M. & Dowd, C. A. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem. J.* **360**, 1–16 (2001).
- 12 Pearson, W. R., Vorachek, W. R., Xu, S. J., Berger, R., Hart, I., Vannais, D. *et al.* Identification of class-mu glutathione transferase genes GSTM1–GSTM5 on human chromosome 1p13. *Am. J. Hum. Genet.* **53**, 220–233 (1993).
- 13 Ross, V. L., Board, P. G. & Webb, G. C. Chromosomal mapping of the human mu class glutathione S-transferases to 1p13. *Genomics* **18**, 87–91 (1993).
- 14 Nebert, D. W. & Vasiliou, V. Analysis of the glutathione S-transferase (GST) gene family. *Hum. Genomics* **1**, 460–464 (2004).
- 15 McLellan, R. A., Oscarson, M., Alexandrie, A.-K., Seidegard, J., Evans, D. A., Rannug, A. *et al.* Characterization of a human glutathione S-transferase cluster containing a duplicated GSTM1 gene that causes ultrarapid enzyme activity. *Mol. Pharmacol.* **52**, 958–965 (1997).
- 16 Katoh, T., Ymano, Y., Tsuji, M. & Watanabe, M. Genetic polymorphisms of human cytosol glutathione S-transferases and prostate cancer. *Pharmacogenomics* **9**, 93–104 (2008).
- 17 Denson, J., Xi, Z., Wu, Y., Yang, W., Neale, G. & Zhang, J. Screening for inter-individual splicing differences in human GSTM4 and the discovery of a single nucleotide substitution related to the tandem skipping of two exons. *Gene* **379**, 148–155 (2006).
- 18 Xu, S., Wang, Y., Roe, B. & Pearson, W. R. Characterization of the human class mu glutathione S-transferase gene cluster and the GSTM1 deletion. *J. Biol. Chem.* **273**, 3517–3527 (1998).
- 19 Townsend, D. & Tew, K. Cancer drugs, genetic variation and the glutathione-S-transferase gene family. *Am. J. Pharmacogenomics* **3**, 157–172 (2003).
- 20 Rao, A. V. S. K. & Shaha, C. Multiple glutathione S-transferase isoforms are present on male germ cell plasma membrane. *FEBS Lett.* **507**, 174–180 (2001).
- 21 McIlwain, C. C., Townsend, D. M. & Tew, K. D. Glutathione S-transferase polymorphisms: cancer incidence and therapy. *Oncogene* **25**, 1639–1648 (2006).
- 22 Ezer, R., Alonso, M., Pereira, E., Kim, M., Allen, J. C., Miller, D. C. *et al.* Identification of glutathione S-transferase (GST) polymorphisms in brain tumors and association with susceptibility to pediatric astrocytomas. *J. Neurooncol.* **59**, 123–134 (2002).
- 23 Sironi, M., Menozzi, G., Comi, G. P., Cagliani, R., Bresolin, N. & Pozzoli, U. Analysis of intronic conserved elements indicates that functional complexity might represent a major source of negative selection on non-coding sequences. *Hum. Mol. Genet.* **14**, 2533–2546 (2005).
- 24 Ying, S. Y. & Lin, S. L. Intron-derived microRNAs—fine tuning of gene functions. *Gene* **342**, 25–28 (2004).
- 25 Wang, Y. & Leung, F. C. C. Long inverted repeats in eukaryotic genomes: recombination motifs determine genomic plasticity. *FEBS Lett.* **580**, 1277–1284 (2006).
- 26 Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680 (1994).
- 27 Felsenstein, J. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).
- 28 Kumar, S., Tamura, K. & Nei, M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**, 150–163 (2004).
- 29 Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. & Mello, C. C. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- 30 Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404**, 293–296 (2000).
- 31 Smith, N. A. & Singh, S. P. Total silencing by intron-spliced hairpin RNAs. *Nature* **407**, 319–320 (2000).
- 32 Lin, S. L., Chang, D., Wu, D. Y. & Ying, S. Y. A novel RNA splicing-mediated gene silencing mechanism potential for genome evolution. *Biochem. Biophys. Res. Commun.* **310**, 754–760 (2003).
- 33 Yuan, B., Latek, R., Hossbach, M., Tuschl, T. & Liewitter, F. siRNA Selection Server: an automated siRNA oligonucleotide prediction server. *Nucleic Acids Res.* **32**, 130–134 (2004).