Functional consequences of new exon acquisition in mammalian chromodomain Y-like (CDYL) genes

Xin Li1,5*, Jing Liang2*, Haijing Yu3, Bing Su1,4, Chunjie Xiao3, Yongfeng Shang2 and Wen Wang1,4

1 CAS-Max Planck Junior Research Group, Key Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences (CAS), Kunming, Yunnan 650223, China
2 Department of Biochemistry and Molecular Biology, Peking University Health Science Center, Beijing 100083, China
3 Key Laboratory of Bioresources Conservation and Utilization and Human Genetics Center of Yunnan University, Kunming 650091, China
4 Kunming Primate Research Center, Chinese Academy of Sciences, Kunming, Yunnan 650223, China
5 Graduate School of Chinese Academy of Sciences, Beijing 100039, China

The origin of new exons is an important mechanism for proteome diversity. Here, we report the recurrent origination of new exons in mammalian chromodomain Y-like (CDYL) genes and the functional consequences associated with the acquisition of the new exons. Driven by positive selection, the newly evolved longer peptide exhibits weaker transcription repression activity and attenuates the repression activity of the old form, suggesting that the acquisition of the new exons is functionally significant.

Introduction
The discrepancy between the complexity of organisms and the relatively conserved gene numbers and alternative splicing levels [1–5] indicates that complexity in higher animals might require new mechanism(s) in addition to those found in less complex animals, such as the acquisition of new exons in orthologous genes that are present in different lineages. This idea was initially supported by the identification of many exons that are newly evolved in rodents [6,7] and was further strengthened by a recent report of many new exons in humans [8]. In addition, when analysing alternative splicing, Lee and co-workers [9,10] also noted the evolutionary significance of the gain and loss of exons. Using computational and experimental approaches, we characterized the new exons that have arisen in mammalian chromodomain Y-like (CDYL) genes (Box 1). We observed independent acquisition of new exons in mammalian CDYL genes. Our functional analysis revealed that the new exons have conferred new functions to their encoded product, which might have contributed to the evolution of these mammalian species, including humans.

Acquisition of new exons in mammalian CDYL genes
We compared the gene structures and cDNA sequences of all CDYL genes retrieved from GenBank and Ensembl and grouped these sequences into two types based on the promoters that are used to transcribe them (Figure 1). Type I CDYLs are similar to mouse and human CDYL genes in gene structure. This type includes human CDYL gene isoform b (CDYLb; GenBank accession number NM_170751) and the reported macaque CDYL transcription form (AY275460) (Figure 1a). By contrast, type II CDYLs, which include the human CDYL gene isoform a (CDYLa; NM_004824) and the mouse (NM_009881) and dog (GenBank XM_843680 and Ensembl ENSCAFG00000009417, respectively) CDYL genes, have several extra exons in the 5' region and use a different promoter that is upstream of the promoter of type I CDYL forms (Figure 1b). We also searched for CDYL genes in the cow, pig and opossum (see Figure S1 and Table S1 in the supplementary material online). For the cow, although only the type I CDYL transcript was identified using expressed sequence tag (EST) data, the existence of conserved genomic sequences homologous to the 5' region of the type II CDYL gene in other species implies that it also has a type II CDYL transcript. Indeed, EST data for another artiodactyl species, the pig, clearly show that the pig has both type I and type II CDYL transcripts, although the detailed gene structures could not be determined owing to a lack of genomic sequences in the CDYL region. For the opossum, we were able to analyze the CDYL gene structure only using genomic sequences because of a lack of related expression data. Our results suggested that the opossum only has a type I CDYL transcript because there are no encodable upstream sequences similar to the eutharian type II CDYL transcript. We further analysed the chicken CDYL gene and its cDNA sequences (GenBank XM_418960, Ensembl ENSGALG00000012808) and found that the chicken CDYL gene possesses characteristics of the type I transcript. Based on the above analyses, we propose that the type I transcript is the ancestral form of the CDYL gene, an argument that is supported by the fact that the structure of this transcript is also shared by CDYL2 (see Materials and methods in the supplementary material online for analyses).

Based on sequence comparisons, we propose that the CDYL gene (type I) obtained a new promoter upstream of the ancestral promoter, together with three new exons.
(exons 1, 2 and 3), and formed a new transcript (type II) from the common ancestor of Eutheria. Interestingly, after the acquisition of these three exons, a fourth new exon appeared in the CDYL gene of mouse, dog, and human; there is no similarity in either sequence or length in exon 4 among these species, suggesting that the recruitment of the new exon 4 occurred independently in each of the above three eutherian lineages.

In primates, the evolutionary process of the CDYL gene is even more complicated. The human type II transcript CDYLa has been reported previously [11], and we identified a new type II transcript, CDYLa1. As mentioned earlier, in these two type II transcripts, a newly inserted exon 4, of 83 nucleotides in length (purple box in Figure 1b), with no similarity to exons in the mouse and dog, was identified. The insertion of the new exon 4 caused the start codon of CDYLa1 to shift to exon 5, resulting in a coding sequence for a shorter peptide. We identified this form by reverse transcription–PCR in the rhesus monkey, gibbon, orangutan, chimpanzee and human (Figure 1b).

In addition, in the human, chimpanzee and gorilla, the new transcript, CDYLa, was created by shifting the start codon to a more upstream position in exon 2 and subsequent skipping of exon 3 to generate a longer peptide. In other words, from gorilla to human, a novel transcript was created through a combination of acquisition of a new exon 4, shifting of the start codon and skipping of exon 3 (Figure 2).

### Rapid evolution of new exons driven by positive selection

To identify the origin of the new exon 4 of CDYL genes in the human, mouse and dog, we performed a BLASTN search against the genomes of these species [12]. We found

---

**Box 1. The CDY-related gene family**

CDYL belongs to a multigene family called the CDY-related gene family. Previous studies have shown that members of the CDY-related gene family function in gene transcription regulation and have an important role in mammal spermatogenesis [17,18]. Members of this family contain two key functional domains: the chromodomain, which is often implicated in chromatin binding [20], and the enoyl-coenzyme A (CoA) hydratase–isomerase catalytic domain. In the mouse, the CDY-related gene family contains two autosomal members, CDYL and CDYL2. In humans, in addition to the autosomal CDYL and CDYL2 genes, the family consists of a group of closely related gene copies called chromodomain Y (CDY) genes, located on the Y chromosome. Comparative evolutionary studies suggested that one of the autosomal CDYL members arose from genomic duplication, whereas the human Y-linked CDY genes developed by retroposition of CDYL mRNA followed by several rounds of amplification [11,19]. In the mouse, each autosomal gene member produces a long, ubiquitously expressed transcript and a short, testis-specific transcript. Human autosomal CDYL and CDYL2 are ubiquitously expressed in multiple tissues, whereas the Y-linked copies, CDY genes, are expressed solely in the testis [11,19]. The different expression profiles indicate that the members of CDY-related gene family have evolved divergent functions. Therefore, mammalian, especially hominoid, CDYL genes can encode several functionally differentiated peptides by acquisition of new exons.

---

Figure 1. Gene structures of different mammalian CDYL genes. Black boxes, coding regions; white boxes, untranslated regions. The boxes representing exon 4s are in different colours to indicate that there is no homology among species. CDYL transcripts can be grouped into (a) type I and (b) type II, based on their gene structures. Type II CDYL genes use a different promoter, which is upstream of the promoter of type I CDYL forms, and the two types of CDYL genes have different N-terminal sequences.
that the exon 4s of mouse and dog CDYL genes both have sequences that are homologous exclusively to the corresponding intronic regions of the human CDYL gene. These results suggest that the new exon 4s evolved through an important mechanism of exon origination – exonization of unique intronic sequences [7,13].

To understand the evolutionary process of the new exons after their origination, we firstly calculated the number of nucleotide substitutions per site (d) in the new exons 2, 3 and 4, and in the old exons, between the mouse and the rat. The d value for the new exons between the two species (0.1298, Kimura 2-parameter model) is significantly higher than that for the old exons (0.0434, one-tailed Z test, P = 0.002), indicating that these new exons have experienced rapid evolution. Further analysis indicated that the ratio of non-synonymous to synonymous substitution rates (Ka/Ks) for the new exons is 4.55, which is significantly higher than 1 (Z test, P = 0.0190) [14,15] (see Table S2 in the supplementary material online). This result indicates that the rapid evolution of new exons in rodents was driven by positive selection rather than relaxation of functional constraints [16]. We also obtained a significantly higher d value (0.0219) for the coding region of exons 2 and 4 between human and chimpanzee than for the old exons (0.0037, one-tail Z test, P = 0.0506). Furthermore, all of the four nucleotide substitutions between human and chimpanzee are nonsynonymous (Figure S2 in the supplementary material online). Although it is impossible to conduct a statistical test, owing to the small sample size, this mutation pattern suggests that the new exons in hominoids have also experienced positive selection [7].

**Functional divergence due to acquisition of the new exon**

It has been shown that the mouse CDYL protein functions mainly as a transcriptional corepressor in somatic cells [17]. Although CDYL also shows histone acetyltransferase (HAT) activity in vitro, it was proposed that in vivo this HAT activity primarily functions in germline cells during the late stages of spermatid maturation [18]. In humans, it is likely that the testis-specific CDYL retains spermatogenesis-related functions, whereas the ubiquitously expressed CDYL preserves transcriptional repression activity in somatic cells.

Amino acid sequence alignments of human CDYLα and CDYLβ show that both proteins contain two basic domains – the chromodomain and the CoA pocket – and that the only difference between these proteins is that CDYLα has an extra 62 amino acids encoded by the new exons at the N-terminus. Mechanistically, a longer N-terminus upstream of the chromodomain might affect the interaction of CDYLα with other proteins or with chromatin. To gain insight into the functional significance associated with acquisition of the new exon, the transcriptional repression activities of human CDYLα and CDYLβ were investigated.
Further, we co-transfected HeLa cells with both CDYLa and CDYLb to investigate the functional relationship between CDYLa and CDYLb. To examine whether there is indeed a functional difference in gene repression ability between CDYLa and CDYLb, we transfected HeLa cells with Gal4DBD–CDYL constructs in different combinations to examine how the relative abundance of CDYLa might interfere with or enhance the transcription activity of CDYLb. As shown in Figure 3c, the repression of reporter gene expression by CDYLb was gradually alleviated with increasing amounts of CDYLa, suggesting that CDYLa interferes with the transcriptional repression activity of CDYLb.

CDYLb has been shown to be the major splicing form in most somatic cells [11,19]. Although the functional differentiation and interference between CDYLa and CDYLb need to be confirmed by in vivo experiments in the future, the above experiments nevertheless suggest that it is possible that the addition of CDYLa provides a finely tuned regulation of gene expression. At the genomic level, the CDYL gene maintains orthologous status across mammals. However, the existence of different transcripts of the CDYL gene, with different regulatory behaviours, expands the proteome and adds to the complexity of gene regulation in the gorilla, chimpanzee and human.

Acknowledgements
This work was supported by a CAS–Max Planck Society Fellowship, grants 30325016 and 30434000 to W.W.; 30621002, 30393113 and 30470912 to Y.S.) from the National Natural Science Foundation of China, and grants (863 Program: 2006AA02Z466 and 973 Program: 2005CB522404 to Y.S.) from the Ministry of Science and Technology of China.

Supplementary data
Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2007.06.001.

References