Gene xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Short Communication

Molecular and functional evolution of human DHRS2 and DHRS4 duplicated genes

Franco Gabrielli [a], Sergio Tofanelli [b,*]

[a] Department of Experimental Pathology and Medical Biotechnology, University of Pisa, Pisa, Italy
[b] Department of Biology, University of Pisa, Pisa, Italy

ARTICLE INFO

Article history:
Accepted 5 September 2012
Available online xxxx

Keywords:
SDR enzymes
Duplicated genes
DHRS2
DHRS4
Molecular and functional evolution
Positive selection

ABSTRACT

Human DHRS2 and DHRS4 genes code for similar NADP-dependent short-chain carbonyl-reductase enzymes having different substrate specificity. Human DHRS2 and DHRS4 enzymes share several common sequence motives including residues responsible for coenzyme binding as well as for the intimate catalytic oxido-reductase mechanism, while their substrate-binding sequences have very low similarity. We found that DHRS2 and DHRS4 genes are syntenic outparalogues originated from a duplication of the DHRS4 gene that took place before the formation of the mammalian clade. DHRS2 gene evolved more rapidly and underwent positive selection on more sites than the DHRS4 gene. DHRS2 sites under positive selection were mainly located on the enzyme active site thus showing that substrate specificity drove the divergence from the DHRS4 enzyme. Rapid divergent evolution brought the human DHRS2 enzyme to have subcellular localization, synthesis regulation and specialized cellular functions very different from those of the human DHRS4 enzyme.

1. Introduction

DHRS2 (Gabrielli et al., 1995) and DHRS4 (Matsunaga et al., 2008) human genes code for two similar enzymes belonging to the short-chain dehydrogenase/reductase family (SDR family) (Bray et al., 2009; Jornvall et al., 1995; Tanaka et al., 2001). SDR enzymes are present in all life forms (Oppermann et al., 2001; Tanaka et al., 2001) and almost all are NAD/NADP dependent oxido-reductases active on a large and heterogeneous set of substrates including steroids, retinoids, prostaglandins, metabolites and xenobiotics (Bray et al., 2009; Oppermann, 2007; Oppermann et al., 2001; Tanaka et al., 2001; Wu et al., 2007). SDR enzymes exist in different non-allelic isoforms, which may share multiple overlapping substrate specificity and specific dehydrogenase or reductase catalytic activity (Bray et al., 2009; Oppermann, 2007; Oppermann et al., 2001; Tanaka et al., 2001; Wu et al., 2007). In addition, different isoforms may exhibit cell-specific distribution and specific intracellular localizations (Bray et al., 2009; Peltoketo et al., 1999; Wu et al., 2007). Although the amino acid identity values between different SDR enzymes can be as low as 25%, the SDR enzyme tridimensional structures are all superimposable (Bray et al., 2009; Tanaka et al., 2001; Wu et al., 2007).

SDR enzymes are characterized by several common distinct sequence motifs, which include the glycine consensus of NAD/NADP cofactor-binding domain, the catalytic amino acid as well as few highly conserved amino acids scattered along the enzyme sequences. The substrate sites of the SDR enzymes are all formed by amino acids localized at polypeptide C-terminal, but there are no common sequence features for substrate recognition (Tanaka et al., 2001; Wu et al., 2007). The sequence variability of SDR substrate-binding sites is associated to a large spectrum of substrates, among different SDR enzymes resulting in unique active sites and substrate specificities (Wu et al., 2007).

Human SDR enzymes are functionally active in the metabolism of signaling molecules as well as in intermediary and xenobiotic metabolism (Bray et al., 2009; Matsunaga et al., 2008; Oppermann, 2007; Peltoketo et al., 1999; Wu et al., 2007). In human peripheral tissues, SDR enzymes are active on steroid hormones, retinoids and prostaglandins, converting them to either active or inactive status (Oppermann, 2007; Oppermann et al., 2001; Peltoketo et al., 1999; Wu et al., 2007). By their action on regulatory molecules, SDR enzymes play important roles in the control of normal cellular functions (Oppermann et al., 2001; Peltoketo et al., 1999; Wu et al., 2007). SDR enzymes have been implicated in several human diseases (Oppermann et al., 2001; Peltoketo et al., 1999; Wu et al., 2007) and mutated SDR enzymes are responsible for a severe form of juvenile hypertension, male pseudohermaphroditism, Zellweger syndrome and night-blindness disorder (Oppermann et al., 2001; Peltoketo et al., 1999).

The human DHRS2 gene has been cloned and physically and cytogenetically mapped on chromosome 14q11.2 (Pellegrini et al., 2002). DHRS2 gene has two alternative promoter regions: a hepatocyte specific promoter is inducible by the histone deacetylase inhibitor sodium butyrate (Pellegrini et al., 2002) and a second upstream promoter is specifically active in human monocyte-derived dendritic cells (Heinz et al., 2002). DHRS2 gene is also activated by c-Myb (Rushton et al., 2003) and ETV5 (Monge et al., 2009) oncogenic transcription

* Corresponding author at: Department of Biology, University of Pisa, Via Derna 1, 56126 Pisa, Italy. Tel.: +39 050 2211352; fax: +39 050 2211475.
E-mail address: stofanelli@biologia.unipi.it (S. Tofanelli).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.09.013

factors. DHRS2 gene codes for Hep27, a cell cycle regulated (1,14 Donadel et al., 1991; Gabrielli et al., 1995) NADPH-dependent dicarbonyl reductase enzyme active on three toxic α-dicarbonyl-compounds (3,4-Hexanedione, 2,3-Heptanedione and 1-Phenyl-1,2-propanedione) (Shafqat et al., 2006). Hep27 is localized in the human cell nucleus (Donadel et al., 1991; Gabrielli et al., 1995; Monge et al., 2009; Pellegrini et al., 2002), the cytoplasm (Pellegrini et al., 2002) and the mitochondria (Deisenroth et al., 2010; Monge et al., 2009), and it is synthesized in several human normal tissues (Bhati et al., 2008; Heinz et al., 2002; Pellegrini et al., 2002; Shafqat et al., 2006). Hep27, which is highly concentrated in several types of cancer cells, is prognostic marker of prostate (Peterson et al., 2005) and bladder (Shedden et al., 2003) carcinomas and breast sporadic carcinoma (Bhattacharyya et al., 2003). Hep27 also has an important non-catalytic activity: it specifically binds to and inhibits Mdm2 protein, a negative regulator of p53 synthesis and stability. By this activity, nuclear Hep27, via p53, can play a regulatory role on cell cycle and apoptosis (Deisenroth et al., 2010; Monge et al., 2009; Thorner et al., 2010). The human DHRS4 gene is expressed in several human tissues and in several cultured cancer cells, and its expression is activated by ligands of the peroxisome-proliferator-activated receptor-α (Matsunaga et al., 2008). DHRS4 gene codes for a peroxisomal NADPH-dependent carbonyl-reductase active on a large heterogeneous set of substrates including α-dicarbonyl compounds, aldehydes, aromatic ketones and 3-keto-steroids (Endo et al., 2009; Matsunaga et al., 2008).

We established the molecular criteria to distinguish DHRS2 from DHRS4 genes in vertebrate and invertebrate species as well as between them and the other members of the SDR family in particular when the amino-acid sequence similarity with human DHRS4 are low (Supplementary paragraph). We demonstrated that human DHRS2 and DHRS4 genes are conserved syntenic outparalogues, originated by duplication of a vertebrate DHRS4 ancestor gene whose structural and sequence traces are present in insect, nematode and deuterostome species.

The knowledge of structural, molecular and cellular characteristics of DHRS2 and DHRS4 enzymes has allowed us to observe the evolution of DHRS2 and DHRS4 amino acid sequences associated with the evolution of their subcellular localizations and molecular activities. By analyzing the sites under positive control of mammalian DHRS2 and DHRS4 genes, we verified that the evolution of the DHRS2 enzyme mainly occurred in its active site. Indeed the evolution of human DHRS2 duplicated gene brought the coded enzyme to acquire an increased catalytic specificity by reduction of the substrate spectrum and the function to regulate cell cycle and apoptosis. The three steps of the DHRS2 gene evolution (duplication, intracellular relocalization of the coded protein and positive selection) confirms that duplication and the subsequent fast functional divergence of the duplicated genes is one of the most important mechanisms for the development of novel gene functions (Han et al., 2009; Taylor and Raes, 2004; Zhang, 2003).

2. Materials and methods

2.1. Sequence search

We have utilized computational programs and genomic, cDNA, EST and protein sequences obtained from public primary databases. Orthologs of human DHRS2 and DHRS4 genes were searched in fully sequenced and reassembled genomes of vertebrate and invertebrate species by BLAT at UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGatway). We have followed several methods (Tress et al., 2008) in order to determinate and validate DHRS2 and DHRS4 genes, duplicated genes and pseudogenes of vertebrate species as well as DHRS4 genes of invertebrate species. Initially, we used the amino acid sequence of human DHRS2 and DHRS4 enzymes as probe. By these analyses, we detected the nucleotide sequence, the exon–intron organization, the spliceosomal intron physical positions and phases of mammalian DHRS2 genes as well as vertebrate and invertebrates DHRS4 gene. Exon–intron gene organizations and intron physical positions and phases were manually checked verifying the exon–intron boundary sequences and the observance of the GT-AG/GC rule.

Spliceosomal intron physical positions and phases that resulted were very useful tools to define homology between genes and pseudogenes as they contain and retain useful information about gene homology, especially when genes belong to a large family and their sequences have experienced changes rendering difficult their alignment (Irimia and Roy, 2008; Tress et al., 2008).

An identical procedure was utilized to establish the nucleotide sequence of the DHRS2 and DHRS4 pseudogenes. The homology of pseudogenes with DHRS2 or DHRS4 genes was manually identified and the nucleotide sequence identity values between pseudogenes and genes calculated. Genes from different species were annotated as DHRS2 or DHRS4 if they had typical DHRS2 or DHRS4 exon–intron organization with exact intron positions and phases and if polypeptides, deducted from the DHRS2 and DHRS4 orf, included conserved amino acid sequence as well as single amino acid residues in specific sequence-positions that distinguish DHRS2 enzymes from DHRS4 ones. Genes and cDNA sequences identified by automated computational analysis from genomic sequences and already stored in public databases were annotated as DHRS2 and DHRS4 genes following the above indicated criteria. The program LALIGN was used to perform nucleotide and protein sequences alignments (Huang and Miller, 1991). Polypeptide consensus sequences were made with the ClustalW version 2.0 (Larkin et al., 2007).

2.2. Phylogenetic analyses

The alignment of cDNA sequences was performed with CLUSTALW version 2.0 (Larkin et al., 2007). Phylogenetic trees were constructed with the Neighbor Joining Algorithm (Saitou and Nei, 1987) based on the Kimura-2-parameter distance measure as implemented in the software package PHYLIP version 3.6 (Felsenstein, 2005). The data were also subjected to bootstrap analysis with 1000 replications in order to estimate the strength of support for each clade. As the topology is consistent with the most accepted phylogeny of vertebrate species and its clades have strong bootstrap support, we used the NJ tree as tree structure file to investigate patterns of selection with PAML (see Subsection 2.3).

2.3. Inference about selection

Effects of positive selection were detected by applying the maximum-likelihood method (Larkin et al., 2000; Nielsen and Yang, 1998) implemented in the program codeml of PAML version 4.4 (Yang, 1997). This method is based on the comparison of site-specific models with different assumptions regarding the ratio of non-synonymous to synonymous substitution rates ($\omega = d_N/d_S$) used as an estimate of the selective pressure across sites. We run codeml under a basic model with ω constant over all branches (M0), a nearly neutral model with $0<\omega<1$ estimated from the data and $\omega 1 = 1$ fixed (M1a), a model of variable ω with positive selection allowed (M2a), beta distributions of ω allowing (M8 and M8a) or not (M7) for positively selected sites. We applied likelihood ratio tests of positive selection by comparing a χ^2 distribution with the outputs generated under M1 against M2 ($df = 2$), M7 against M8 ($df = 2$) and M8a (the null model with fixed $\omega = 1$) against M8 ($df = 1$). We identified positively selected sites (PSS) at which non-synonymous substitutions d_N occurred at a higher rate than synonymous ones d_S with the Bayes Empirical Bayes (BEB) calculation of posterior probabilities for site classes under model M8 (Yang et al., 2005). We also estimated if the significance of posterior probabilities held true for model assumptions taking into consideration alternative equilibrium codon frequencies (equal nucleotide frequencies and averaged at third

Please cite this article as: Gabrielli, F., Tofanelli, S., Molecular and functional evolution of human DHRS2 and DHRS4 duplicated genes, Gene (2012), http://dx.doi.org/10.1016/j.gene.2012.09.013

codon positions) and alternative trees (the species tree and the gene tree).

As PAML does not have any proper methods to deal with alignment gaps, we limited the PSS analysis to 9 taxa (eutheria) and 251 codons at DHRS2, to 14 taxa (vertebrates) and 246 codons at DHRS4, and to 23 taxa and 245 codons at both genes. We also tested for evidence of positive selection along pre-specified foreground branches using the mixed branch-site model (model = 2; NS sites = 2) within codeml (Yang and Nielsen, 2002).

3. Results and discussion

3.1. Human DHRS2 and DHRS4 genes

The alignment of the nucleotide sequences of human DHRS2 and DHRS4 genes with those of their respective cloned cDNA showed that both genes have eight polypeptide-coding exons and an almost identical exon–intron organization. Five homologous exon polypeptides (exons 2, 3, 5, 6 and 7) have identical size and exons-1, 4 and 8 differ for 4, 1 and 1 amino acid residues, respectively (Fig. 1a). All human DHRS2 and DHRS4 spliceosomal introns are located in homologous positions and with identical phases, but the DHRS2-intron between 4th and 5th exon (Fig. 1b). This intron conserves its phase-2, but it is upstream slid causing loss of the glycine codon. The deletion of DHRS2 exon-4 last codon is an evolution step specific of primate species as DHRS2 genes of the other mammalian species include it (Supplementary Fig. 1).

DHRS2 gene, from the first ATG codon up to the first stop codon in frame, is about 2.4 times shorter than DHRS4 gene because of its shorter introns (Fig. 1c) and the DHRS2 introns have no significant or very low sequence identity with DHRS4 homologous introns. All intron splice-acceptor and splice-donor sequences of both human DHRS2 and DHRS4 genes follow the GT-AG rule except for the DHRS2 intron between exons 7 and 8 (Pellegrini et al., 2002), where the 5′-splice donor sequence starts with the rare GC dinucleotides (Thanaraj and Clark, 2001).

Human DHRS2 (GenBank ID: 10202) and DHRS4 (GenBank ID: 10901) genes are syntenic on Chr 14 and physically close. The DHRS4 gene is located about 315 kb downstream the DHRS2. A traditional pseudogene (DHRS4p1, GenBank ID: 317749) and a duplicated DHRS4 gene (DHRS4L1, GenBank ID: 72863) (Fig. 2b) are localized, respectively, about 20 kb and 47 kb downstream the DHRS4 gene. Analyzing the physical localization on Chr14, the exon–intron organization, intron phases and protein-coding sequences of human DHRS2 and DHRSL1 genes and DHRS4p1 pseudogenes with those of the DHRS4 gene, we could annotate human DHRS2 and DHRSL1 genes and DHRS4p1 pseudogenes as duplicate forms of the DHRS4 gene (Subsection 3.4). By the same analysis, we identified the vertebrate orthologues of human DHRS2 and DHRS4 genes (Section 3.3) as well as invertebrate orthologs of the human DHRS4 (Supplementary paragraph).

Human DHRS2 and DHRS4 enzymes share 59.2% identity (Supplementary Table 1), which is the highest value among members of the SDR family. Highly similar homologous sequence blocks are distributed along the orf nucleotide sequences: nucleotide sequences of the eight pairs of paralogous exons of human DHRS2 and DHRS4 genes share identity values ranging from 38.6% of exon-7 to 89.3% of exon-2 pairs, and sequence identity values of their coded polypeptides range from 15.8% to 84.7%, respectively (Supplementary Table 2). DHRS4L1 exons have sizes identical to homologous DHRS4 exons except for exon-3 that in DHRS4L1 is divided, by two extra introns, in three exons whose total length is 9 nucleotides longer than human DHRS4 exon-3 (Fig. 1c). Positions and phases of the other DHRS4L1 introns are identical to those of the DHRS4 gene. The mRNA deducted from the DHRSL1 gene does not include stop signals and it codes for a polypeptide 81.1% identical to the DHRS4 enzyme. However, there are no reported data showing that the DHRS4L1 gene is transcribed. The DHRS4p1 pseudogene has intron positions and phases identical to those of the DHRS4 gene with just an intron with sequence longer than that of the DHRS4 gene (Fig. 1c). DHRS4p1 exon-6 carries a frame shift mutation that generates a stop signal in the same exon-6. DHRS4p1 is transcribed and the polypeptide sequence deducted from the transcribed mRNA (GenBank ID: 317749) is 93.8% identical to the first-210 amino acids of the DHRS4 enzyme, but it lacks 46 amino acids at enzyme C-terminal.

3.2. Amino-acid sequence and molecular activities of human DHRS2 and DHRS4 enzymes

Recently, the crystal structure of human (PDB ID: 3o4r) and pig (Tanaka et al., 2008 and http://www.ebi.ac.uk/thornton-srv) DHRS4 enzymes has been established. On the enzyme primary structure, the residues responsible for the catalysis, the residues interacting with the NADP coenzyme molecule, the C-terminal polypeptide including the Peroxisomal Targeting Signal-1 tripeptide (PTS1) and several residues of the substrates-binding site are mapped. In the PDB ID: 3o4r, the secondary structural elements of the human DHRS4 enzyme that we have included in Fig. 1a are shown. Some residues, interacting with the substrate, were identified by site-specific mutagenesis and molecular modeling (Endo et al., 2009). Human DHRS2 and DHRS4 enzymes have in homologous positions the same catalytic residue triad (Ser, Tyr and Lys), almost identical amino acid residues involved in coenzyme NADP binding and the G/AxxxGxG consensus, which marks the NADP/NAD-binding domain called the Rossmann fold (Fig. 1a and Supplementary Table 3). Rossmann fold, typical of SDR enzymes, consists of six β-sheets and four α-helices supersecondary structure (El-Kabbani et al., 2005; Tanaka et al., 2001, 2008; Wu et al., 2007). The high similarity of DHRS2 and DHRS4 homologous functional amino acids is consistent with DHRS2 and DHRS4 identical enzymatic core activity for specifically and reversely associating the NADPH coenzyme and operating the reduction of carbonyl substrates. Such a fact strongly suggests that DHRS2 and DHRS4 homologous functional residues must be placed in almost identical geometrical positions, sustained by very similar tertiary, quaternary structures and inter-subunits interactions (Endo et al., 2009; Matsunaga et al., 2008; Tanaka et al., 2008).

Human DHRS4 C-terminal tail including PTS1 consensus is not conserved in human DHRS2 enzyme (Fig. 1a and Supplementary Table 3). This is in agreement with data obtained by biochemical and histochemical techniques showing that human DHRS4 is a peroxisomal enzyme (Matsunaga et al., 2008), while the human DHRS2 enzyme, which does not have a canonical PTS1, has been detected in the cytoplasm, mitochondria and nuclei (Deisenroth et al., 2010; Donadel et al., 1991; Gabrielli et al., 1995; Monge et al., 2009; Pellegrini et al., 2002). Human DHRS2 and DHRS4 enzymes show very low sequence similarity in the two homologous sequence blocks (DHRS2 Tyr176-Val184 and Lys224-Gln243; DHRS4: Phe173-Pro181 and Arg221-Arg240) inserted between highly conserved sequences (Fig. 1a and Supplementary Table 3). The two sequences are physically positioned in two different sides of the substrate-binding site of human and pig DHRS4 as well as of other SDR enzymes (El-Kabbani et al., 2005; Endo et al., 2009; Tanaka et al., 2001, 2008). The different residues of DHRS2 and DHRS4 substrate-binding sites are associated with different substrate specificity of the respective enzymes: human DHRS2 enzyme is specific for only three α-dicarbonyl compounds (one aromatic and two aliphatic small molecules), while human DHRS4 enzyme is active on aromatic and aliphatic α-dicarbonyl compounds as well as on a large group of aliphatic, aromatic and steroid carbonyl compounds (Endo et al., 2009; Matsunaga et al., 2008). The apparently kaleidoscopic variability of the C-terminal sequence (constituting the substrate-binding site) between different SDR enzymes is responsible of their unique substrate-binding properties (Wu et al., 2007) and recently it has been clearly demonstrated for human and pig DHRS4 enzymes (Endo et al., 2009) (see Subsection 3.5).

Please cite this article as: Gabrielli, F., Tofanelli, S., Molecular and functional evolution of human DHRS2 and DHRS4 duplicated genes, Gene (2012), http://dx.doi.org/10.1016/j.gene.2012.09.013

The three substrates (3,4-Hexanedione, 2,3-Heptanedione and 1-Phenyl-1,2-propanedione) of the DHRS2 enzyme are cytotoxic xenobiotics and their toxic effects include mitotic chromosome loss, inflammation and cell sclerotic complications. The three xenobiotics are generated during the course of metabolic process and oxidative stress, and are present in dietary constituents. According to the World Health Organization (WHO) and the International Program on Chemical Safety (IPCS), in Europe the estimated daily dietary intake of the three xenobiotics is altogether about 41 μg/day.

The reductive reaction of the DHRS2 enzyme can specifically quench the toxic action of the three xenobiotics by converting them to nontoxic or less toxic compounds (Oppermann, 2007; Shafqat et al., 2006). DHRS2 enzyme activity on only three α-carbonyl substrates is not a handicap for the cellular physiology, because the increased substrate specificity attributes to the human DHRS2 enzyme the specialized physiological function to selectively protect human cells from the toxic effects of the three α-dicarbonyl xenobiotics, even in the presence of other species of α-dicarbonyl compounds. An enzyme, such as DHRS4,

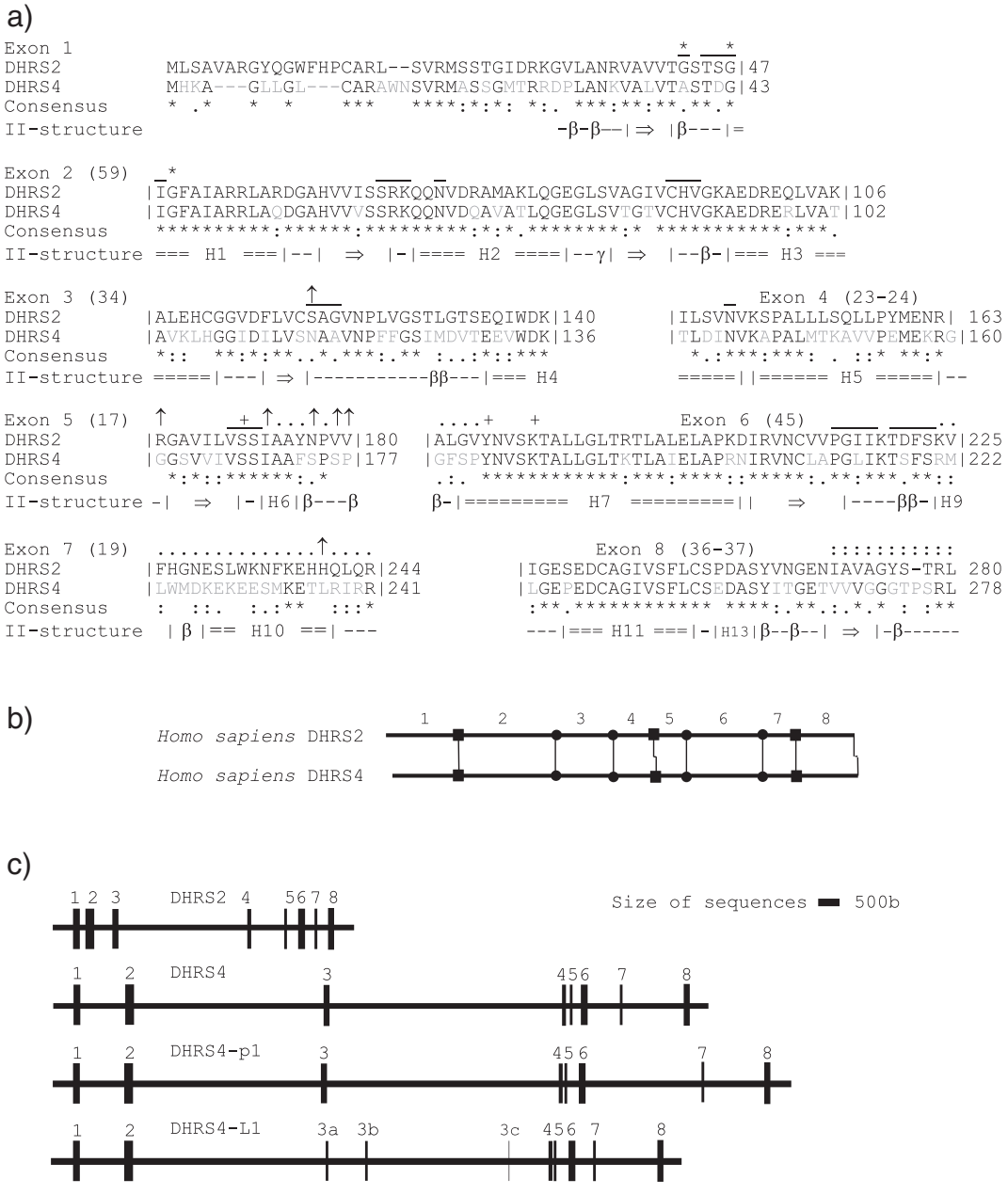

Fig. 1. a) Alignment of polypeptide sequences coded by human DHRS2 and DHRS4 homologous exons. In black and gray conserved and not conserved amino acids, respectively; dashed lines are introduced to optimize alignments; vertical bars, exon 3′- and 5′-ends. Numbers in parenthesis, exon sizes; numbers at the end of exon sequence, sequence position of exon last amino acid. Lines above sequence mark amino acid interacting with NADP coenzyme molecule; (*), amino acid belonging to G/AxxxGxG consensus; (+), catalytic amino acid; (…), amino acids of the enzyme active site; (:::), DHRS4 C-terminal sequence including PTS1. ↑, DHRS2 site under positive selection. Iso173, Asn177, Val179, Val180 and His239 of the enzyme active site coincide with sites under positive selection. Consensus symbols: (*), amino acid residues identical in all aligned sequences; (:), conserved substitutions; (.), semiconserved substitutions. Secondary structure symbols: H, helix; ⇒ strand; β, beta turn; γ, gamma turn (data from PDB ID: 3o4r of human DHRS4). b) Exon–intron structure of human DHRS2 and DHRS4 genes. ●, phase-0 intron (splicing between codons); ■, phase-2 intron (splicing between the second and the third nucleotide of the codon). c) Structure of human DHRS2 and DHRS4 genes, and DHRS4 pseudogenes. Vertical bars represent exons. Exon 1 starts from the first AUG triplets and exon 8 ends at the codon before the first stop codon of the open reading frame, respectively.

Please cite this article as: Gabrielli, F., Tofanelli, S., Molecular and functional evolution of human DHRS2 and DHRS4 duplicated genes, Gene (2012), http://dx.doi.org/10.1016/j.gene.2012.09.013

which is active on several different substrates (relative specificity), has its catalytic activity on a given substrate regulated by the relative amounts of others of its substrates present in the cell. A high concentration of a given substrate (e.g. ketosteroids) can saturate the DHRS4 enzyme making it almost inactive towards the other less concentrated substrates (e.g. α-dicarbonyl compounds). DHRS2 enzyme catalytic specificity and subcellular localizations become particularly important for the cell physiology when the DHRS4 enzymes as well as other α-dicarbonyl reductases (Oppermann, 2007), saturated by other types of substrates, are insufficiently active on the toxic substrates of the DHRS2 enzyme. It may explain the presence in the same cell of enzymes active on the same substrates (Oppermann, 2007).

The regulation of human DHRS2 and DHRS4 genes are different in the same cell type being in the same metabolic conditions (Bhati et

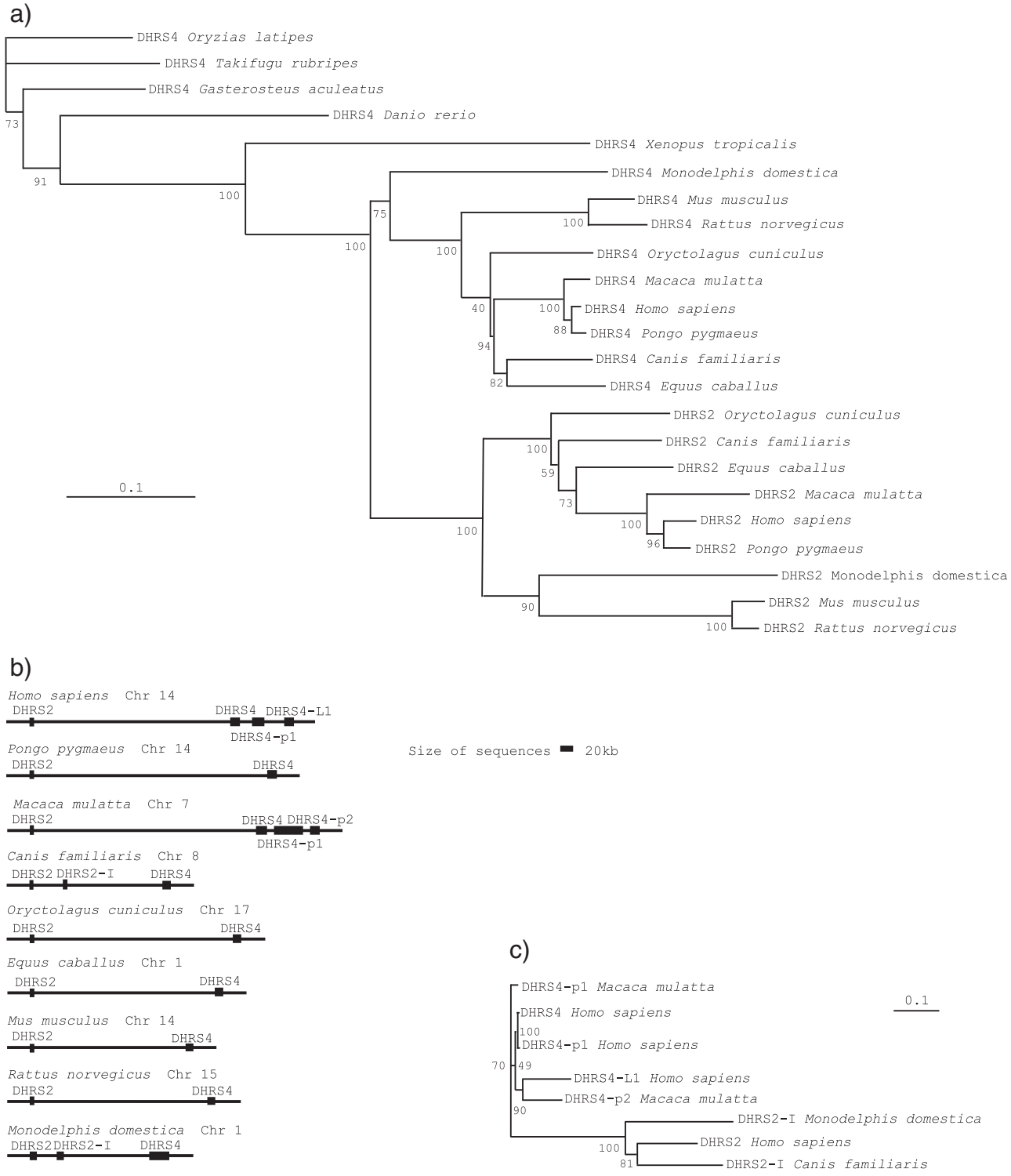

Fig. 2. a) Phylogram of vertebrate DHRS2 and DHRS4 genes. b) Physical loci of DHRS2 and DHRS4 genes and pseudogenes; black rectangles represent annotated genes and pseudogenes; L1 and I, paralogous gene; p, pseudogene. c) Phylogram of mammalian DHRS2 and DHRS4 genes, and pseudogenes. *Homo sapiens*: DHRS2, GenBank ID: 10202; DHRS4, GenBank ID: 10901; DHRS4L1, GenBank ID: 728635; DHRS4-p1, GenBank ID: 1317749; *Macaca mulatta* DHRS4-p1 manually deducted from the genome; DHRS4-p2, GenBank ID: 714661; *Canis familiaris* DHRS2-I, GenBank ID: 490617; *Monodelphis domestica* DHRS2-I, GenBank ID: 100030645.

Please cite this article as: Gabrielli, F., Tofanelli, S., Molecular and functional evolution of human DHRS2 and DHRS4 duplicated genes, Gene (2012), http://dx.doi.org/10.1016/j.gene.2012.09.013

al., 2008; Bhattacharyya et al., 2003; Heinz et al., 2002; Monge et al., 2009; Peterson et al., 2005; Thorner et al., 2010). In endometrial carcinoma cells, the oxidative stress upregulates the expression of the DHRS2 enzyme in order to cope with the stress-induced synthesis of α-dicarbonyl xenobiotics in mitochondria (Monge et al., 2009). Moreover, it has been recently shown that human DHRS2 gene has an important non-enzymatic activity (Deisenroth et al., 2010; Monge et al., 2009; Thorner et al., 2010). In osteosarcoma (Deisenroth et al., 2010) and in MCF7 breast cancer cultured cells (Thorner et al., 2010), Hep27 is part of the molecular pathway that regulates cell cycle and apoptosis. In this pathway, Hep27 (the DHRS2 enzyme) synthesis is up regulated by the oncogene Myc, the cytoplasmic Hep27 migrates into mitochondria where, after having proteolytically lost the N-terminal mitochondria targeting signal, migrates into nuclei and binds to the Mdm2 protein. Mdm2 has the function to inhibit p53 transcription and activate its degradation. Mdm2, when complexed with Hep27, loses its negative actions on p53 and by this mechanism, Hep27 nuclear concentration can control the onset of the cell cycle arrest and apoptosis (Deisenroth et al., 2010; Monge et al., 2009; Thorner et al., 2010).

3.3. Vertebrate DHRS2 and DHRS4 genes

Several mammalian species have two physically close syntenic genes, which have relative physical positions, exon–intron organizations and intron phases identical to those of either DHRS2 or DHRS4 human genes (Supplementary Figs. 1 and 2). Furthermore, they code for polypeptides having high sequence similarity with human DHRS2 or human DHRS4 enzymes (Supplementary Table 1).

In the great apes DHRS2 genes have several molecular characteristics identical to those of the human DHRS2 gene, which include: size of all homologous exons, in particular exon-4 which is 69 nucleotides long while DHRS2 exon-4 of the other mammalian species is 3 nucleotides longer (23 and 24 amino acids respectively, Supplementary Fig. 1); the rare GC dinucleotide at the 5′-splice-donor sequence of the intron located between exons 6 and 7 (DHRS2 genes of the other mammalian species have in homologous position GT dinucleotides); close size of orthologous introns (Supplementary Fig. 3). Furthermore, human and great apes DHRS2 coded enzymes share very high values of sequence identity (Supplementary Table 1). DHRS2 genes of primates and of the other mammalian species typically have identical positions and phases of orthologous introns with just an exception: the primate intron, located between exons 1 and 2, is upstream shifted by three bases (Fig. 1a and Supplementary Fig. 1). DHRS2 exon-8 of all mammalian species analyzed by us is 108 nucleotides long, while mammalian DHRS4 exon-8 is 3 nucleotides longer (36 and 37 amino acids respectively, Fig. 1a and Supplementary Figs. 1 and 2). DHRS2 and DHRS4 exon-8 sizes are diagnostic to distinguish between DHRS2 and DHRS4 genes.

Mammalian non-primate DHRS2 genes have some molecular characteristics of the vertebrate DHRS4 genes: all intron acceptor splicing-sites have GT signals and exon-4 is 72 nucleotides long. Single amino acid residues in homologous sequence positions respectively characterize mammalian enzymes encoded by DHRS2 or DHRS4 genes. Gly43, Cys119 and Leu157 of human DHRS2 enzymes are present in homologous positions in all mammalian DHRS2 enzymes we have analyzed (Supplementary Fig. 1). In vertebrate DHRS4 enzymes, these amino acids are replaced respectively by Ala39, Ser115 and Val153 (Supplementary Fig. 2).

The absence of duplicate forms of the DHRS4 gene in amphibians and fish, and the identical intron physical positions and phases as well as the identical sizes of most of homologous exons of mammalian DHRS2 and DHRS4 genes point out that the origin of the human DHRS2 gene has been from a syntenic duplication of a vertebrate DHRS4 ancestor, most probably reptilian or prototherian. In order to find out which of the two duplicated genes occupies the physical position of the DHRS4 gene of non-mammalian vertebrate genomes, we analyzed sequences at the 5′ and 3′ of the DHRS2 gene in mouse and opossum, and of the gene DHRS4 in mouse, opossum, amphibian and fish genomes, however we could not find any specific marker of the two genes in the genome of different species. DHRS2 and DHRS4 enzymes of mammalian species, as observed for the human enzymes, have identical catalytic residues and almost identical coenzyme-binding residues, while they do have different substrate-binding residues as well as the last 11 residues at C-terminal (Supplementary Table 3).

The phylogenetic tree shows that vertebrate DHRS2 and DHRS4 genes belong to two different evolutionary clusters with different evolutionary rates and that mammalian DHRS2 and DHRS4 genes have a common DHRS4 ancestor with amphibian, fish (Fig. 2a) and invertebrate species (Supplementary paragraph). The DHRS2/DHRS4 split node was supported by 100% of bootstrap replicates. The averaged branch lengths in the DHRS2 tree were 1.5 times the DHRS4 tree ones ($t_{16} = 60.76$, $p < 0.001$).

These data confirm that the DHRS2 gene had an evolution divergent from and faster than that of the DHRS4 gene. Interestingly, in each mammalian species sequence identity values of DHRS2 and DHRS4 enzymes remain about constant (Supplementary Table 1). By all the aforementioned analyses, we could annotate DHRS2 and DHRS4 genes of the analyzed vertebrate species.

3.4. Structure and chromosomal physical positions of vertebrate DHRS2 and DHRS4 genes and pseudogenes

Neither gains nor losses of introns took place during vertebrate DHRS2 and DHRS4 gene evolution, albeit almost all DHRS2 introns are shorter than DHRS4 paralogous ones (Supplementary Fig. 3). The different lengths of DHRS2 and DHRS4 homologous splicing introns suggest that the respective introns may have evolved under a different or differently regulated intron-size control mechanism(s).

Physical loci of vertebrate DHRS2 and DHRS4 genes are syntenic and quite close, with distances ranging from 200 kb to 400 kb. The DHRS2 and DHRS4 gene chromosomal regions may include, in between or downstream DHRS2 and DHRS4 paralogous genes or pseudogenes (*Homo sapiens*, *Macaca mulatta*, *Canis familiaris* and *Monodelphis domestica*) (Fig. 2b), or none of them (*Pongo pygmaeus*, *Equus caballus*, *Mus musculus* and *Rattus norvegicus*). Genomic sequences were annotated as DHRS2 or DHRS4 genes or pseudogenes by analyzing their chromosome physical positions (Fig. 2b), the positions and phases of their introns, the nucleotide sequence similarity, the size of homologous exons and the similarity of their coded polypeptide, with respect to those of the human DHRS2 and DHRS4 genes (Supplementary Figs. 1, 2 and Tables 1, 2, 5) and by a phylogenetic analysis (Figs. 2a, c). *Takifugu rubipres*, which like other fish does not have the DHRS2 gene, has two DHRS4 genes and a DHRS4 pseudogene. Polypeptides deducted from *T. rubipres* DHRS4 genes share 90.1% identity. DHRS2 and DHRS4 duplicated genes and pseudogenes are physically localized downstream of their respective mother genes (Fig. 2b).

3.5. Sequence similarity, subcellular localization and catalytic specificity of mammalian DHRS2 and DHRS4 enzymes

The alignment of human DHRS4 enzyme sequence with those of mammalian DHRS4 enzymes shows that the residues binding the NADP coenzyme, the catalytic residues and 11 residues of the C-terminal sequence, including the PTS1 tripeptide consensus, are highly conserved in mammals (Supplementary Table 4). The PTS1 consensus tripeptide of human peroxisomal enzymes is (S/A/C)(K/R/H)(L/M) (Lametschwandtner et al., 1998). The consensus of mammalian (human included) DHRS4 enzymes is restricted to S(R/H)L (Supplementary Table 4) and their peroxisomal localizations is proved by biochemical technologies (Endo et al., 2007; Lei et al., 2003; Matsunaga et al., 2008; Tanaka et al., 2008; Usami et al., 2003). Sequence variability mostly concerns amino acids of the substrate-binding site (Supplementary Table 4 and Supplementary Fig. 2) and suggests that

Please cite this article as: Gabrielli, F., Tofanelli, S., Molecular and functional evolution of human DHRS2 and DHRS4 duplicated genes, Gene (2012), http://dx.doi.org/10.1016/j.gene.2012.09.013

the DHRS4 enzyme from different species should have different substrate specificity. Substrate specificity and catalytic activity of human (Endo et al., 2009; Matsunaga et al., 2008), pig (Tanaka et al., 2008; Usami et al., 2003), rabbit (Usami et al., 2003), dog (Endo et al., 2007), and mouse (Lei et al., 2003) DHRS4 enzymes have been extensively studied and it has been shown that they are all carbonyl reductases, which utilize the NADP coenzyme, operate by identical catalytic residues and have multiple and overlapping substrate specificities. Their common substrates include α-dicarbonyl compounds, aromatic ketones, phenylquinones, aldehydes (retinal aldehyde included) and steroids (Endo et al., 2007, 2009; Lei et al., 2003; Matsunaga et al., 2008; Usami et al., 2003). However, the DHRS4 enzymes from the four aforementioned species have different values of affinity and/or catalytic efficiency towards the same substrate molecules (Endo et al., 2007, 2009; Matsunaga et al., 2008; Usami et al., 2003), which must be determined by different residues located in their respective substrate-binding sites. Elegantly, by multiple site-specific mutations, it has been shown that human and pig DHRS4 enzymes, which both accept α-ketosteroids as substrates, convert them into different stereospecific products: 3α-hydroxysteroids and 3β-hydroxysteroids, respectively (Endo et al., 2009). Two different residues (human Ser176 and Phe179, pig Phe177 and Leu180) (Supplementary Fig. 4a) are responsible for the different stereospecific reduction of 3-ketosteroids and for different substrate affinity of human and pig enzymes. The same authors also showed that a third residue (human Thr195, pig Asn196) of one subunit of the DHRS4 enzyme interacts with a residue (human Asn183, pig Asn184) of another subunit of the same tetramer. These inter-subunit interactions, modifying the orientation of the catalytic and substrate-binding residues, differently influence the kinetic constants of the human and pig DHRS4 enzymes towards the substrates: 3-ketosteroids and all-trans-retinal. The human DHRS4 enzyme is catalytically more active on 3-ketosteroids than on all-trans-retinal, while the pig enzyme does the opposite on the same substrates.

The evolution of mammalian DHRS2 enzymes is similar to that of DHRS4 enzymes but, as a whole, DHRS2 sequences are less conserved than DHRS4 homologous ones. DHRS2 coenzyme-binding and catalytic amino acids are highly conserved (Supplementary Table 4) and they have high similarity with DHRS4 homologous coenzyme-binding amino-acid sequences (Supplementary Table 3).

The mammalian DHRS2 C-terminal tail sequence has some conserved residues (Supplementary Table 4) but in each species, the consensus sequence is different from that of the DHRS4 C-terminal (Supplementary Table 3) suggesting that not all mammalian DHRS2 enzymes are peroxisomal. *P. pygmaeus* and *M. mulatta* C-terminals include the PTS1 tripeptide consensus, however, their DHRS2 enzymes should not be peroxisomal (Lametschwandtner et al., 1998) as their C-terminal sequence upstream of the PTS1 tripeptide is different from that of the DHRS4 enzymes (Supplementary Table 3), and in all mammalian species (*M. domestica* included) exon-8 lacks one amino acid residue of the PTS1 triplet or in a position close to PTS1. Tanaka et al. (2008) showed by site-specific mutagenesis that the pig DHRS4 enzyme, lacking a single amino acid in the Peroxisomal Targeting Signal is not able to migrate in the peroxisomes.

When combined, the above-mentioned data indicate that during the mammalian clade formation, the human DHRS2 gene, generated by DHRS4 gene duplication, started an independent evolution assuming new subcellular localization and new substrate specificity.

3.6. Sites under positive selection

Having found that the DHRS2 gene evolved faster than the DHRS4 gene we aimed at verifying whether the DHRS2 gene showed more sites under positive selection and whether they corresponded to codons of amino acids located in the substrate-binding site. At first, we considered the phylogeny obtained with paralogous genes in the same data set (Fig. 2a). When the hypothesis of positive selection pressure ($\omega > 1$) was tested for the DHRS2 + DHRS4 coding sequence, highly significant results were obtained for the Asn177 codon within exon 5 and a slightly lower statistical support was obtained for codon His239 within exon 7 (see Table 1 and Supplementary Table 6 for details). The comparison between a neutral model (M7) and a positive selection model (M8) yielded a likelihood test ratio of 7.86 ($p<0.05$) when performed for the whole cDNA tree and 12.47 ($p<0.001$) when analyzing the single exon. Under a branch-site model a higher rate of non-synonymous substitutions were observed after the split in the DHRS2 lineage (LRT = 411,12 $p<0.001$ for rate heterogeneity, LRT = 4.22 $p<0.05$ between selection and neutral $\omega = 1$ model). As a second step, we analyzed lineage-specific phylogenies. Sequence divergence among tree branches was similarly large at both cDNA regions to detect positive selection between paralogues with sufficient statistical power (d_S averaged over all branches = 0.97 and 1.36, respectively at DHRS2 and DHRS4). Pairwise substitution analyses among vertebrate DHRS genes showed an average global non-synonymous to synonymous substitution ratio ($\omega t = dN/dS$) of 0.213 at DHRS2 and 0.187 at DHRS4. Only values >1.0 are generally regarded as signals of positive selective pressure on the sequences analyzed. However, ω almost never reaches such values because selection generally acts with different rates along the branches or among sites of the same gene. Hence, a simple one-ratio model (M0) may mask evidence of positive selection that have affected a subset of lineages and/or genomic regions.

Consistency among more sophisticated analyses, namely significant LRTs applied separately to the eight exon sequences under both site and branch-site models, were used to detect PSS. Eight codons at the DHRS2 branch and one codon at the DHRS4 branch showed posterior probabilities > 0.70 (Table 1 and Fig. 1a). In particular, three DHRS2 codons for Ile173, Asn177 and Val179 at exon 5 and Ser174 at DHRS4 exon 5, the paralogous codon of Asn177, gave consistent signals of strong positive selection ($p > 0.95$ of $\omega > 1$ and $p < 0.05$). Eight DHRS2 codons under positive control and just one in the DHRS4 gene confirmed that the DHRS2 enzyme underwent much more adaptive changes than the DHRS4 enzyme. The localization in DHRS2 substrate-binding site of amino acids under positive selection is consistent with our hypothesis that DHRS2 gene neofunctionalization was under positive selection operated by DHRS2 enzyme substrate molecules.

As subcellular relocalization represents a common mechanism for the functional diversification of duplicated genes (Byun-McKay and Geeta, 2007; Marques et al., 2008), the independent evolution of the DHRS2 gene may have been triggered by minor changes in the C-terminal tail sequence, which relocalized within the cells the newly duplicated DHRS2 enzyme. In the new subcellular compartments (cytoplasm and mitochondria), the DHRS2 enzyme has very likely met different physical and chemical environments (e.g. different relative concentrations of carbonyl and α-dicarbonyl compounds). The DHRS2 enzyme, without having experienced any mutation in the active site, could have interacted mostly with α-dicarbonyl compounds, just because these compounds were abundant in the new cellular compartments. These compounds especially because of their cytotoxicity, could have been the selective instruments for the evolution of DHRS2 substrate-binding site. Thereafter, in the new cellular compartments, the DHRS2 enzyme has been subjected to positive selection in order to adjust to the new metabolic conditions. Substituting some amino acids of the active site, the DHRS2 substrate specificity has been restricted to a few α-dicarbonyl compounds that become independent of the presence of other species of carbonyl compounds. The capability of an enzyme to be active on multiple substrates (as that of the newly duplicated DHRS2 enzyme) likely acts as an important forge for a rapid functional divergence in such enzymes (Brock et al., 2008; Pichersky et al., 2006). Changing very few amino acids in an active site, which accepts molecular species having different structures and charges, can change its affinity towards some of the original substrates and trigger a new evolution of the active site. The enzyme can acquire new substrate specificities just by nullifying its affinity towards some of its original substrates and different specificities are expected

Please cite this article as: Gabrielli, F., Tofanelli, S., Molecular and functional evolution of human DHRS2 and DHRS4 duplicated genes, Gene (2012), http://dx.doi.org/10.1016/j.gene.2012.09.013

depending on the position in the active site of the substituted residues (Endo et al., 2009; Wu et al., 2007).

Several data about DHRS2 enzyme are in agreement with the aforementioned hypothesis. In *M. domestica* the C-terminal tail of the DHRS2 enzyme has low similarity with that of the DHRS4 enzyme C-terminal and it lacks the PTS1 three peptide consensus (Supplementary Table 3), suggesting that in Marsupialia the DHRS2 enzyme already had a subcellular localization different from that of the DHRS4 one and that its active site has been subjected to a fast and massive positive selection with respect to the DHRS4 enzyme (Table 1).

The presence of DHRS2 and DHRS4 enzymes in different subcellular compartments of the same cell sheds some light on how the DHRS2 and DHRS4 genes could have had different functional evolutions. The different metabolism induced in early terrestrial vertebrates by adaptive mechanisms to a non-aquatic lifestyle may have contributed to the evolution of the DHRS2 gene.

Our data about the DHRS2 and DHRS4 enzymes' evolution are in high agreement with those of Han and coworkers obtained in mammalian duplicate genes (Han et al., 2009). These authors demonstrated that high levels of positive selection are a common event in mammalian (human included) young duplicate genes and that the consequent amino acid substitutions are responsible for the maintenance of the duplicated copies. The mutation of the peroxisomal targeting signal brought the DHRS2 enzyme to new subcellular membrane-bound compartments and the maintenance in them has allowed the positive selection of the DHRS2 gene operated by the new molecular environments.

The proposed evolution of the human DHRS2 enzyme may throw some light about the high sequence variability of the substrate-binding site of the SDR superfamily enzymes. SDR enzymes may have a different substrate-binding sequence because they had a different evolution history of their catalytic site even when they are active on the same substrate/s.

We speculate that the capability of the DHRS2 enzyme to regulate cell cycle and apoptosis was acquired when the enzyme started to proteolytically loose the mitochondrial signal at the N-terminal (Deisenroth et al., 2010) and, consequently, the DHRS2 nuclear target signal (Gabrielli et al., 1995) become unmasked. Migration into nuclei opened the chance for the DHRS2 enzyme to bind to the nuclear Mdm2 protein and acquire the new important function to regulate cell cycle and apoptosis. In the nuclei, however, the DHRS2 enzyme may have started an adaptation to the new proteinaceous environment however without eliminating the DHRS2 catalytic activity. It suggests that the core enzyme structure, mostly constituted by the Rossmann fold, has not been changed as it still can associate with the NADPH and perform the reductive catalysis. In other words, the DHRS2 enzyme structure and its surface are necessary for both its molecular activities.

Literature (Byun-McKay and Geeta, 2007 and quoted references) and our data about DHRS2 suggest that genes coding a unique enzyme with different subcellular localizations in the same cell type can evolve in different unrelated functions.

In conclusion, single or very few mutations can cause the transfer of an enzyme in a new subcellular membrane-bound compartment in which, because the new chemical environment may be involved in a new evolutionary path and acquire new cellular functions. The necessity of the enzyme to adapt to the new metabolic environment can be the driving force of an accelerated evolution operated by positive selection. New mutations can cause the location of the neofunctionalized enzyme in a second membrane-bound compartment in which, maintaining its structural characteristics serendipitously acquires a new cellular function just by interacting with molecules specifically present in the new subcellular environment. Duplicated enzymes acting in different subcellular compartments, even if conserving the same primary structure, are able to perform different cellular functions.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2012.09.013.

Table 1
Posterior probabilities of ω>1 and hierarchical Likelihood Ratio Tests (BEB analysis) for detecting positively selected codon sites at DHRS2 and DHRS4 genes.

Codon	Probability ω>1[a]			LRT		
	By cDNA	By exon	M1a/M2a	M7/M8	M8a/M8	
DHRS2						
120 Ser	0.850	Exon 3	**0.999****	3.89	4.77	4.40
164 Arg		Exon 5	0.950	**11.12****	**10.94****	**10.40****
173 Ile	0.942		**0.990***			
177 Asn	0.830		**0.999****			
179 Val	**0.980***		**0.999****			
180 Val			0.845			
239 His	**0.957***	Exon 7	0.729	0.00	0.34	0.00
245 Ile		Exon 8	0.807	0.39	5.36	0.41
DHRS4						
174 Ser	0.883	Exon 5	0.945[b]	**6.25***	3.86	**3.54***
DHRS2+4						
177 Asn	**0.977***	Exon 5	**1.000****	**9.94****	**12.47****	**9.31****
239 His	0.872	Exon 7	[b]	**7.13***	0.00	0.28

First column on the left, amino acid positions in human DHRS2 and DHRS4 enzyme sequences as shown in Fig. 1. Human DHRS2 177 Cys and DHRS4 174Ser are located in homologous positions.

[a] Only p values>0.70 are reported.

[b] p>0.95 by a NEB analysis and/or under the M2a mode.

* p<0.05.

** p<0.01.

Acknowledgments

This study was supported by Ministero dell'Istruzione, dell'Università e della Ricerca (PRIN–COFIN grant). F.G. wishes to dedicate his contribution to his former supervisors Henry A. Lardy and Corrado Baglioni.

References

Bhati, R., et al., 2008. Molecular characterization of human breast tumor vascular cells. Am. J. Pathol. 172, 1381–1390.
Bhattacharyya, C., et al., 2003. Simultaneous classification and relevant feature identification in high-dimensional spaces: application to molecular profiling data. Signal Process. 83, 729–743.
Bray, J.E., Marsden, B.D., Oppermann, U., 2009. The human short-chain dehydrogenase/reductase (SDR) superfamily: a bioinformatics summary. Chem. Biol. Interact. 178, 99–109.
Brock, A., Brandt, W., Dräger, B., 2008. The functional divergence of short-chain dehydrogenases involved in tropinone reduction. Plant J. 54, 388–401.
Byun-McKay, S.A., Geeta, R., 2007. Protein subcellular relocalization: a new perspective on the origin of novel genes. Trends Ecol. Evol. 22, 338–344.
Deisenroth, C., Thorner, A.R., Enomoto, T., Perou, C.M., Zhang, Y., 2010. Mitochondrial HEP27 is a c-Myb target gene that inhibits Mdm2 and stabilizes p53. Mol. Cell. Biol. 30, 3981–3993.
Donadel, G., Garzelli, C., Frank, R., Gabrielli, F., 1991. Identification of a novel nuclear protein synthesized in growth-arrested human hepatoblastoma HepG2 cells. Eur. J. Biochem. 195, 723–729.
El-Kabbani, O., Carbone, V., Darmanin, C., Ishikura, S., Hara, A., 2005. Structure of the tetrameric form of human L-xylulose reductase: probing the inhibitor-binding site with molecular modeling and site-directed mutagenesis. Proteins 60, 424–432.
Endo, S., et al., 2007. Characterization of an oligomeric carbonyl reductase of dog liver: its identity with peroxisomal tetrameric carbonyl reductase. Biol. Pharm. Bull. 30, 787–1791.
Endo, S., 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.
Felsenstein, J., 2005. PHYLIP (Phylogeny Inference Package) version 3.6. . Distributed by the author Department of Genome Science, University of Washington, Seattle.
Gabrielli, F., Donadel, G., Bensi, G., Heguy, A., Melli, M., 1995. A nuclear protein, synthesized in growth-arrested human hepatoblastoma cells, is a novel member of the short-chain alcohol dehydrogenase family. Eur. J. Biochem. 232, 473–477.
Han, M.V., Demuth, J.P., McGrath, C.L., Casola, C., Hahn, M.W., 2009. Adaptive evolution of young gene duplicates in mammals. Genome Res. 19, 859–867.
Heinz, S., Krause, S.W., Gabrielli, F., Wagner, H.M., Andreesen, R., Rehli, M., 2002. Genomic organization of the human Hep27 gene: alternative promoter usage in HepG2 cells and monocyte-derived dendritic cells. Genomics 79, 608–615.
Huang, X., Miller, W., 1991. A time-efficient, linear-space local similarity algorithm. Adv. Appl. Math. 12, 337–357.
Irimia, M.I., Roy, S.W., 2008. Spliceosomal introns as tools for genomic and evolutionary analysis. Nucleic Acids Res. 36, 1703–1712.

Jornvall, H., et al., 1995. Short-chain dehydrogenases/reductases (SDR). Biochemistry 34, 6003–6013.
Lametschwandtner, G., Brocard, C., Fransen, M., Van Veldhoven, P., Berger, J., Hartig, A., 1998. The difference in recognition of terminal tripeptides as peroxisomal targeting signal 1 between yeast and human is due to different affinities of their receptor Pex5p to the cognate signal and to residues adjacent. J. Biol. Chem. 273, 33635–33643.
Larkin, M.A., et al., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.
Larkin, M.A., et al., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 24, 2947–2948.
Lei, Z., Chen, W., Zhang, M., Napoli, J.L., 2003. Reduction of all-trans-retinal in the mouse liver peroxisome fraction by the short-chain dehydrogenase/reductase RRD, induction by the PPARR ligand clofibrate. Biochemistry 42, 4190–4196.
Marques, A.C., Vinckenbosch, N., Brawand, D., Kaessmann, H., 2008. Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. Genome Biol. 9, R54.
Matsunaga, T., et al., 2008. Characterization of human DHRS4, an inducible short-chain dehydrogenase/reductase enzyme with 3b-hydroxysteroid-dehydrogenase activity. Arch. Biochem. Biophys. 477, 339–347.
Monge, M., et al., 2009. Proteomic approach to ETV5 during endometrial carcinoma invasion reveals a link to oxidative stress. Carcinogenesis 30, 1288–1297.
Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.
Oppermann, U., 2007. Carbonyl reductases: the complex relationships of mammalian carbonyl- and quinone-reducing enzymes and their role in physiology. Annu. Rev. Pharmacol. Toxicol. 47, 293–322.
Oppermann, U., Filling, C., Jornvall, H., 2001. Forms and functions of human SDR enzymes. Chem. Biol. Interact. 130–132, 699–705.
Pellegrini, S., et al., 2002. A human short-chain dehydrogenase/reductase gene: structure., chromosomal localization, tissue expression and subcellular localization of its product. Biochim. Biophys. Acta 1574 (3), 215–222.
Peltoketo, H., Luu-The, V., Simard, J., Adamski, J., 1999. 17b-hydroxysteroid dehydrogenase (HSD)/17-ketosteroid reductase (KSR) family; nomenclature and main characteristics of the 17HSD/KSR enzymes. J. Mol. Endocrinol. 23, 1–11.
Peterson, L.E., et al., 2005. Artificial neural network analysis of DNA microarray-based prostate cancer recurrence. Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, pp. 1–8 http://dx.doi.org/10.1109/CIBCB.2005.1594929.
Pichersky, E., Noel, J.P., Dudareva, N., 2006. Biosynthesis of plant volatiles: nature's diversity and ingenuity. Science 311, 808–811.
Rushton, J.J., Davis, L.M., Lei, W., Mo, X., Leutz, A., Ness, S.A., 2003. Distinct changes in gene expression induced by A-Myb., B-Myb and c-Myb proteins. Oncogene 22, 308–313.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Shafqat, N., et al., 2006. Hep27, a member of the short-chain dehydrogenase/reductase family, is an NADPH-dependent dicarbonyl reductase expressed in vascular endothelial tissue. Cell. Mol. Life Sci. 63, 1205–1213.
Shedden, K.A., et al., 2003. Accurate molecular classification of human cancers based on gene expression using a simple classifier with a pathological tree-based framework. Am. J. Pathol. 163, 1985–1995.
Tanaka, N., Nonaka, T., Nakamura, K.T., Hara, A., 2001. SDR: structure., mechanism of action, and substrate recognition. Curr. Org. Chem. 5, 89–111.
Tanaka, N., et al., 2008. Molecular basis for peroxisomal localization of tetrameric carbonyl reductase. Structure 16, 388–397.
Taylor, J.S., Raes, J., 2004. Duplication and divergence: the evolution of new genes and old ideas. Annu. Rev. Genet. 38, 615–643.
Thanaraj, T.A., Clark, F., 2001. Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. Nucleic Acids Res. 29, 2581–2593.
Thorner, A.R., Parker, J.S., Hoadley, K.A., Perou, C.M., 2010. Potential tumor suppressor role for the c-Myb oncogene in luminal breast cancer. PLoS One 5, e13073.
Tress, M.L., et al., 2008. Determination and validation of principal gene products. Bioinformatics 24, 11–17.
Usami, N., et al., 2003. Cloning, expression and tissue distribution of a tetrameric form of pig carbonyl reductase. Chem. Biol. Interact. 143–144, 353–361.
Wu, X., Lukacik, P., Kavanagh, K.L., Oppermann, U., 2007. SDR-type human hydroxysteroid dehydrogenases involved in steroid hormone activation. Mol. Cell. Endocrinol. 265–266, 71–76.
Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.
Yang, Z., Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. 19, 908–917.
Yang, Z., Wong, W.S., Nielsen, R., 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107–1118.
Zhang, J., 2003. Evolution by gene duplication: an update. Trends Ecol. Evol. 18, 292–298.

Please cite this article as: Gabrielli, F., Tofanelli, S., Molecular and functional evolution of human DHRS2 and DHRS4 duplicated genes, Gene (2012), http://dx.doi.org/10.1016/j.gene.2012.09.013