Published in final edited form as:

Comp Biochem Physiol Part D Genomics Proteomics. 2009 March ; 4(1): 54–65. doi:10.1016/j.cbd.2008.10.004.

Horse Carboxylesterases: Evidence for Six *CES1* and Four Families of *CES* Genes on Chromosome 3

Roger S Holmes^{1,2,3,4}, Laura A Cox^{1,2}, and John L VandeBerg^{1,2}

¹Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX, USA

²Southwest National Primate Research Center, Southwest Foundation for Biomedical Research, San Antonio, TX, USA

³School of Biomolecular and Physical Sciences, Griffith University, Nathan, QLD, Australia

Abstract

Carboxylesterases (CES) are responsible for the detoxification of a wide range of drugs and xenobiotics, and may contribute to cholesterol, fatty acid and lung surfactant metabolism. In this study, *in silico* methods were used to predict the amino acid sequences, secondary and tertiary structures, and gene locations for horse *CES* genes and encoded proteins, using data from the recently completed horse genome project. Evidence was obtained for six *CES1* genes closely localised on horse chromosome 3, for which the predicted *CES1* gene products are ≥74% identical. The horse genome also showed evidence for three other *CES* gene classes: *CES5*, located in tandem with the *CES1* gene cluster; and *CES2* and *CES3*, located more than 9 million base pairs downstream on chromosome 3. Horse *CES2*, *CES3* and *CES5* gene products shared 42–46% identity with each other, and with the *CES1* protein subunits. Sequence alignments of these enzymes demonstrated key enzyme and family specific CES protein sequences reported for human *CES1*, *CES2*, *CES3* and *CES5*. In addition, predicted secondary and tertiary structures for horse *CES1*, *CES2*, *CES3* and *CES5* subunits showed extensive conservation with human *CES1*. Phylogenetic analyses demonstrated the relationships and potential evolutionary origins of the horse CES sequences with previously reported sequences for human and other mammalian *CES* gene products. Several *CES1* gene duplication events have apparently occurred following the appearance of the ‘dawn’ horse ~ 55 million years ago.

Keywords

Horse; amino acid sequence; carboxylesterase; evolution; gene duplication

Introduction

Carboxylesterases (CES; E.C.3.1.1.1) catalyse hydrolytic and transesterification reactions using a broad range of substrates, including xenobiotics, anticancer pro-drugs, narcotics,

© 2008 Elsevier Inc. All rights reserved.

⁴Corresponding Author: Roger S Holmes, D.Sc., Department of Genetics, Southwest National Primate Research Center, Southwest Foundation for Biomedical Research, San Antonio, TX, 78227, USA, rholmes@sfbgenetics.org, Phone: 210-258-9687 ; Fax: 210-258-9600.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

clinical drugs and pro-herbicide esters (Redinbo & Potter 2005; Gershater *et al.* 2006; Satoh & Hosokawa, 2006). CES also detoxifies organophosphates, carbamate compounds and insecticides (Leinweber 1987), catalyses several cholesterol and fatty acid metabolic reactions (Hosokawa *et al.* 2007) and the conversion of alveolar surfactant in lung (Ruppert *et al.* 2006); and has been linked with the assembly of low density lipoprotein particles in liver (Wang *et al.* 2007).

Five families of mammalian CES have been reported (Holmes *et al.* 2008a) including CES1, the major liver enzyme (Shibata *et al.* 1993); CES2, the major intestinal enzyme (Schewer *et al.* 1997); CES3, expressed in liver, colon and brain (Sanghani *et al.* 2004); CES5, a major urinary protein of the domestic cat (Miyazaki *et al.* 2003; Holmes *et al.* 2008b); and CES6, a predicted CES-like enzyme in brain (Clark *et al.*, 2003). Three-dimensional structural analyses of human CES1 have clarified the structure-function relationships for this enzyme, and the identification of three ligand binding sites, including the promiscuous active site, 'side door' and 'Z-site', where substrates, fatty acids and cholesterol analogues respectively, are bound; and a 'product releasing' residue (Bencharit *et al.* 2003; 2006; Fleming *et al.* 2005).

Structures for several human and animal CES genes have been determined, including human (see Marsh *et al.* 2004) and rodent CES1 and CES2 'like' genes (see Hosokawa *et al.* 2007). Moreover, following the release of a number of mammalian genome sequences, predicted CES gene structures have been described for five classes of CES genes in several mammals and other animal species (Holmes *et al.* 2008a-d). Recently, the horse (*Equus caballus*) genome sequence has been reported (Horse Genome Project, 2008) enabling *in silico* interrogation and analyses of horse genes and proteins to be undertaken. This paper reports the predicted gene and amino acid sequences; predicted secondary and tertiary structures for multiple horse CES1 protein subunits and CES2, CES3 and CES5 protein subunits; and describes the structural, phylogenetic and evolutionary relationships for these enzymes.

Even though horse liver CES was one of the first enzymes subjected to large scale purification (Burch, 1955) and has been biochemically characterized (Stoops *et al.* 1975; Inkerman *et al.* 1975), there are no previous reports of protein and genomic structures and sequences for horse CES for any of the mammalian CES classes. CES has been extensively investigated in other mammals and shown to serve a range of metabolic (Satoh & Hosokawa 2006; Redinbo & Potter 2005) and biomedical roles (Pindel *et al.*, 1997; Xu *et al.*, 2002; Imai, 2006; Mutch *et al.*, 2007; Wang *et al.*, 2007).

Methods

In silico horse CES gene and protein identification

BLAST (Basic Local Alignment Search Tool) studies were undertaken using web tools from the National Center for Biotechnology Information (NCBI) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul *et al.* 1997). Protein BLAST analyses used the human CES1 amino acid sequence (see Table 1; Figure 1) to examine the non-redundant protein sequences database available for the horse genome (Horse Genome project, 2008) using the blastp algorithm. This procedure produced 25 BLAST 'hits' which were individually examined and retained in FASTA format, and a record kept of the sequences for predicted mRNAs and encoded forms of horse CES and related proteins. These records were derived from annotated genomic sequences using the gene prediction method: GNOMON and predicted sequences with high similarity scores for human CES1 were further examined. Nine CES-like reference sequences were obtained, including six predicted as being CES1-like, and three predicted as CES2-like, CES3-like and CES5-like (see Table 1).

BLAT (BLAST-Like Alignment Tool) *in silico* analyses were subsequently undertaken for each of the predicted horse CES amino acid sequences using the UC Santa Cruz web browser [<http://genome.ucsc.edu/cgi-bin/hgBlat>] (Kent *et al.* 2003) with the default settings to obtain the predicted locations for each of the horse CES genes, including predicted exon boundary locations and gene sizes. Sequences for other known human CES gene products, including CES2, CES3, CES5 and CES6 (Table 1), were also used in BLAST analyses to examine the horse non-redundant protein sequence database. With the exception of CES6, predicted mRNA and encoded protein sequences were obtained for each of the corresponding horse CES genes, and BLAT analyses revealed predicted gene locations and exon boundaries for each of the horse CES genes. BLAT analyses were also undertaken of the horse genome using the UC Santa Cruz web browser to obtain predicted nucleotide sequences for exons 13 and 14 and intron 13 for each of the six CES1-like genes (designated CES1.1; CES1.2; CES1.3; CES1.4; CES1.5; and CES1.6) using the derived amino acid sequences to interrogate the horse genome.

Predicted Structures and Properties for Horse CES Gene Products

Predicted secondary and tertiary structures for horse CES1 subunits (1.1-1.6), CES2, CES3 and CES5 were obtained using the PSIPRED v2.5 web site tools provided by Brunel University [<http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>] (McGuffin *et al.* 2000) and the SWISS MODEL web tools [<http://swissmodel.expasy.org/>], respectively (Kopp & Schwede 2004). The reported tertiary structure (2.0 Å resolution) for the human CES1 Coenzyme A complex (Benchari *et al.* 2003, 2006; Fleming *et al.* 2005) served as the reference for obtaining the predicted horse CES tertiary structures, with a modeling range of residues 21-551 for the horse CES1 subunits; residues 29-539 for horse CES2; residues 32-550 for horse CES3; and residues 31-540 for horse CES5. Theoretical isoelectric points and molecular weights for horse CES subunits were obtained using Expasy web tools (http://au.expasy.org/tools/pi_tool.html). SignalP 3.0 web tools were used to predict the presence and location of signal peptide cleavage sites (<http://www.cbs.dtu.dk/services/SignalP/>) for each of the predicted horse CES sequences (Emanuelsson *et al.* 2007).

Phylogenetic Studies and Sequence Divergence

Phylogenetic trees were constructed using an amino acid alignment from a ClustalW-derived alignment of CES protein sequences, obtained with default settings and corrected for multiple substitutions (Chenna *et al.* 2003; Larkin *et al.* 2007) [<http://www.ebi.ac.uk/clustalw/>]. An alignment score was calculated for each aligned sequence by first calculating a pairwise score for every pair of sequences aligned. Alignment ambiguous regions, including the amino and carboxyl termini, were excluded prior to phylogenetic analysis yielding alignments of 526 residues for comparisons of mammalian CES1 sequences, and alignments of 509 residues for comparisons of human and horse CES1, CES2, CES3 and CES5 sequences with *Xenopus laevis* and zebrafish (*Danio rerio*) CES, which served as outgroup sequences (Table 1). The extent of divergence for the mammalian CES1-like subunits, and the human and horse CES2, CES3 and CES5 subunits were determined using the SIM-Alignment tool for Protein Sequences [<http://au.expasy.org/tools/sim-prot.html>] (Schwede *et al.* 2003).

Results

Alignments of Human CES1 and Predicted Horse CES1 Amino Acid Sequences

The deduced amino acid sequences for six distinct subunits of horse CES1 (designated as CES1.1-CES1.6) are shown in Figure 1 together with the previously reported sequence for human CES1 (Shibata *et al.*, 1993). Alignments of the six horse CES1 subunits with human

CES1 showed between 69-81% sequence identities (Table 2). The predicted amino acid sequences for horse CES1 subunits were two (horse CES1.2-CES1.5) or five (horse CES1.1) residues shorter or seven residues longer (horse CES1.6) than that of human CES1 (567 residues) (Figure 1; Table 1). Of particular interest are key residues which have been previously shown to contribute to the catalytic, subcellular localization, oligomeric and regulatory functions for human CES1 (sequence numbers refer to human CES1) (Table 3). These included the catalytic triad for the active site, the microsomal targeting sequences, including the hydrophobic N-terminus signal peptide and the C-terminal endoplasmic reticulum (ER) retention sequence (His-Ile-Glu-Leu), disulfide bond forming residues (Cys95/Cys123 and Cys280/Cys291) and ligand binding sites, including the 'Z-site' (Gly356), the 'side door' (Val424-Met425-Phe426) and product releasing (Phe552) residues (see Table 3). Identical or conservatively substituted residues were observed for each of the six horse CES1 subunits for the key human CES1 residues previously described (see Table 3).

Other key CES1 sequences included two 'ion pairs' which have been reported to be responsible for subunit-subunit interaction, namely residues Lys78/Glu183 and Glu72/Arg193 (Bencharit *et al.* 2003; 2006; Fleming *et al.* 2005). Predicted horse CES1 subunit sequences for these sites showed that only one of the charge clamps was retained for all six horse CES1 subunits, namely Lys78 (or Arg)/Glu183 (Figure 1). A second 'ion pair' site for horse CES1.2 – CES1.6 subunits is unlikely to function since Arg186 for human CES1 has been replaced by a neutral amino acid (Pro193 for horse CES1.2-CES1.6), whereas the predicted horse CES1.1 subunit has retained Arg193 and may support the second 'ion pair' (Figure 1). The N-glycosylation site for human CES1 at Asn79-Ala80-Thr81 has been retained for all of the predicted horse CES1 sequences (CES1.1-CES1.6) (Figure 1). Additional potential N-glycosylation sites were observed for horse CES1-like subunits, including CES1.1 and CES1.4 (382Asn-383Ser-384Ser); CES1.1 (269Asn-270Phe-271Ser); CES1.2 and CES1.3 (382Asn-383Ser-384Ser); and CES1.4 (499Asn-500Phe-Ser501; and 516Asn-517Pro-518Ser) (Table 3).

Alignments of Horse CES2, CES3 and CES5 with Human CES1, CES2, CES3 and CES5 Amino Acid Sequences

The deduced amino acid sequences for horse CES2, CES3 and CES5 are shown in Figure 2 together with the previously reported sequences for human CES2 (Schewer *et al.*, 1997; Pindel *et al.*, 1997); human CES3 (Clark *et al.*, 2003; Sanghani *et al.*, 2004); human CES5 (Ota *et al.*, 2004); and human CES1 (Shibata *et al.*, 1993) (see Table 1). Interrogation of the horse genome with the human CES6 amino acid sequence (Clark *et al.*, 2003; Ota *et al.*, 2004) however failed to identify a corresponding horse CES6 homologue using the horse genome sequence (Horse Genome Project, 2008).

Alignments of predicted amino acid sequences for horse CES2, CES3 and CES5 with the corresponding human CES sequences confirmed several key CES residues discussed earlier for human CES1, including the active site 'triad', the hydrophobic N-terminus signal peptide and the disulfide bond forming residues; however, only horse and human CES2 and CES3 sequences contained the C-terminal endoplasmic reticulum retention sequences, namely HTEL (human and human CES2), QEDL (human CES3) and QEEL (horse CES3) (Table 3). Horse and human CES5 C-terminal sequences lacked the endoplasmic reticulum retention tetrapeptide sequences and showed high content of hydrophobic amino acids for the additional 12 amino acids concluding with an Ala-Pro sequence in each case (Figure 2). Horse and human CES2, CES3 and CES5 lacked the human CES1 N-glycosylation site at Asn79-Ala80-Thr81, but exhibited other potential N-glycosylation sites: horse CES2 (2 sites: Asn111-Gln112-Ser113 and Asn276-Leu277-Ser278); horse CES3 (4 sites at Asn212-Ile213-Thr214; Asn285-Ser286-Ser287; Asn433-Phe434-Ser435; and

530Asn-531Gln-532Ser); and horse CES5 (4 sites at Asn281-Ser282-Ser283; Asn363-Lys364-Ser365; Asn522-Lys523-Thr524 and Asn533-Val534-Ser535) (Table 4). 'Ion pair' residues reported for human CES1 were absent in the predicted horse CES2, CES3 and CES5 sequences (Fig 2). The 'Z-site' residue (Gly356 for human CES1) was retained for horse CES2, CES3 and CES5 sequences, whereas 'side-door' (Val422-Met423-Phe424 for human CES1) sequences have undergone conservative substitutions for horse CES3 (Ile425-Ile426-Leu427) and horse CES5 (Val419-Phe420-Phe421), but was reduced in length to two hydrophobic residues for horse CES2 (Ile423-Phe424). The 'product releasing' residue for human CES1 (Phe552) has undergone conservative substitutions for all of the predicted horse CES sequences (Leu for horse CES1, CES2 and CES5 sequences and Trp for horse CES3) (Figures 1 and 2; Table 3).

Predicted N-terminal signal cleavage sites for horse CES1 subunits were examined for the horse CES1-like subunits, with CES1.1-CES1.5 retaining the 18 residue sequence reported for human CES1 (von Heinje 1983), whereas a longer N-terminus was observed for horse CES1.6 which contained a predicted 25 residue signal cleavage sequence (Figure 1). Horse CES2, CES3 and CES5 subunits also showed predicted N-terminal signal cleavage site sequences of different lengths, with 26, 28 and 25 residues, respectively (Figure 2). Sequence identities for horse and human CES1, CES2, CES3 and CES5 sequences showed that respective CES subunits showed higher levels of identities in each case for enzymes of the same proposed class (68-81%) whereas CES subunits from different classes exhibited lower levels of sequence identity (41-49%) (Table 2).

Predicted Structures and Properties for Horse CES Gene Products

Predicted secondary structures for six horse CES1-like and horse CES2-, CES3- and CES5-like subunits were compared with the previously reported secondary structure for human CES1 (Bencharit et al., 2003; 2006) (Figures 1 and 2). Similar α -helix β -sheet structures were observed for all of the horse and human CES gene products examined. Consistent structures were predicted near key residues or functional domains including the α -helix within the N-terminal signal peptide; the β -sheet and α -helix structures near the active site Ser228 (human CES1) and 'Z-site' (Glu354/Gly356 respectively); the α -helices bordering the 'side door' site; and the α -helix containing the 'product releasing' residue (Phe551 for human CES1). In addition, two regions lacking helical or strand structures (residues 51-115 and 169-188 for human CES1) were predominantly retained for all forms of horse CES subunits examined which have been shown for human CES1 to contain 2 charge clamps sites: (Lys79/Glu183 and Glu73/Arg186); an N-glycosylation site at Asn79-Ala80-Thr81; a second potential N-glycosylation site for horse and human CES2 (Asn111-Gln112-Ser113 for horse CES2), and one of the disulfide bridges (87Cys/117Cys) reported for human CES1. Human and horse CES5 secondary structures, however, predicted an additional strand and helix at the hydrophobic C-termini, in each case.

Predicted 3-D structures for horse CES1, CES2, CES3 and CES5 showed a high degree of similarity with each other and with the reported structure for human CES1 (Bencharit et al., 2003; 2006) (Figure 3). The rainbow based color code (red for carboxyl-end and deep blue for amino terminus end) illustrated the high degree of conservation for predicted horse CES1, CES2, CES3 and CES5 secondary and tertiary structures despite having <49% sequence identities. Horse CES subunits exhibited similar theoretical pI values (5.5-6.5) as compared with human CES subunits (5.4-6.1), with the exception of the human CES6 subunit which exhibited a much higher theoretical value (9.4) (Table 1).

Predicted Gene Locations and Exonic Structures for Horse CES1, CES2, CES3 and CES5 Genes

Table 1 and Figure 4 summarize the predicted locations for horse *CES1*, *CES2*, *CES3* and *CES5* genes based upon BLAT interrogation of the horse genome (Horse Genome Consortium, 2008) using the derived sequences for horse CES1, CES2, CES3 and CES5 subunits and the UC Santa Cruz Web Browser (Kent et al., 2003). The predicted horse CES genes were located on chromosome 3 in two clusters, with the six *CES1* 'like' genes located in a tandem sequence between nucleotides 7,844,258 - 8,199,077, near to the predicted horse CES5 gene located between nucleotides 8,287,811-8,319,752 (Fig 4). The predicted horse *CES2* and *CES3* genes were found in a distant location on chromosome 3 between nucleotides 17,402,324-17,428,268. BLAT interrogations of the horse and human genomes with the corresponding CES sequences also demonstrated that the two CES gene clusters on the horse and human genomes were syntenic for chromosomes 3 and 16, respectively.

The horse *CES1* and *CES5* 'like' genes were transcribed on the negative strand whereas the *CES2* and *CES3* horse genes were transcribed on the positive strand. Figure 1 summarizes the predicted exonic start sites within each of the horse *CES1*-like genes, with the exception of the predicted horse *CES1.6* gene, having 14 exons, in identical or similar positions to those described for the human CES1 gene (Langmann et al. 1997). The predicted horse *CES1.6* gene contained 13 predicted exons and lacked a start site in exon 8 compared with the other predicted horse *CES1* genes. Horse *CES2* and *CES3* genes contained 12 and 13 predicted exons respectively (Figure 2), in similar positions to those observed for the human *CES2* (Tang et al. 2008) and *CES3* (Clark et al. 2003) genes. The horse *CES5* gene contained 13 predicted exons in identical positions to those reported for human *CES5* (Figure 2) (Ota et al., 2004).

Figure 5 shows the predicted nucleotide sequences for exon 13, intron 13 and exon 14 for horse *CES1.1*, *CES1.2*, *CES1.3*, *CES1.4*, *CES1.5* and *CES1.6* genes. Exons 13 and 14 showed distinct nucleotide sequences for each of the horse *CES1* like genes and were 90% identical, whereas the intron corresponding to intron 13 for horse *CES1.1* also showed distinct sequences for each of the horse *CES1* like genes, showing that 80% of the nucleotide residues were identical.

Phylogeny and Divergence of Horse and Other Mammalian CES Sequences

A phylogenetic tree (Figure 6) was calculated by the progressive alignment of human, baboon, mouse, cow and horse CES1 like CES amino acid sequences showed 2 major groups, namely the horse CES1 sequences and other mammalian CES1-like sequences. The six horse CES1-like subunits clustered together, with horse CES1.2 and CES1.3 showing a higher degree of relatedness than with the other forms, and with horse CES1.5 and CES1.6 subunits showing lower sequence identities with other horse CES1 sequences (Table 2). A second phylogenetic tree (Figure 7) was calculated from the progressive alignment of 8 mammalian CES subunits with xenopus and zebrafish CES subunits: human CES1 and horse CES1.1; as well as human and predicted horse subunits sequences for CES2, CES3 and CES5 (Table 1). The sequences for the mammalian CES subunits clustered into 4 discrete groups, namely horse and human CES1, CES2, CES3 and CES5. The xenopus and zebrafish CES sequences served as outgroups for the mammalian CES1, CES2, CES3 and CES5 sequences. Table 5 presents the calculated genetic distances for several mammalian *CES1*-like common ancestral genes, including human and baboon *CES1*-like genes, human, mouse and cow *CES1*-like genes and proposed common ancestors for six horse *CES1*-like genes.

Discussion

Sequencing of the genome of the domestic horse, *Equus caballus*, has assisted in identifying genomic features shared with other mammals and has provided a tool for researchers to better understand equine molecular evolution, genetics, biochemistry and disease (<http://www.broad.mit.edu/mammals/horse/>). The horse genome sequence will also assist trait mapping of specific diseases within modern breeds of horses and contribute to a better understanding of medical conditions such as influenza, arthritis and allergies that are shared by horses and humans (Ring *et al.* 1977; Livesay *et al.* 1987; Dolvik & Klemetsdal 1996). Many drugs are used to improve health and treat disease for both species and given the roles of CES in drug metabolism (Redinbo and Potter 2005), lung physiology (Ruppert *et al.* 2006) and xenobiotic, insecticide, lipid and cholesterol metabolism (see Imai 2006), studies of the genetics and biochemistry of horse CES will contribute to an improved understanding of the contribution of genetics to these important metabolic processes in the horse.

Horse liver carboxylesterase (CES) was the first mammalian CES to be subjected to large scale purification and biochemical analysis (Burch 1955). Subsequent studies examined the kinetic properties of horse liver CES using a range of substrates and reported an equivalent weight of 70,000 for this enzyme based on titration with *p*-nitrophenyl dimethylcarbamate (Stoops *et al.* 1975; Inkerman *et al.* 1975). There are however no reports of amino acid sequences for horse liver CES or for other CES gene family members. Limited genetic analyses of horse CES have been undertaken which have used electrophoretic polymorphisms of horse serum CES to map the responsible gene (*Es*) to chromosome 3 for which variants have been useful in examining the relatedness of various horse breeds (Andersson *et al.* 1983; Kelly *et al.* 2002).

Human CES1 has been extensively studied biochemically and the 3-D structure determined at high resolution (2.0Å) (Bencharit *et al.* 2003; 2006; Fleming *et al.* 2005). The enzyme is divided into three functional domains: the catalytic domain contains the active site 'triad' and the carbohydrate binding site; the $\alpha\beta$ domain provides the majority of the hydrophobic internal structure and assists in forming the trimeric subunit structure for this enzyme; and the regulatory domain which facilitates substrate binding, product release and the trimer-hexamer equilibrium. Several key amino acid residues or sequences have been strictly conserved among the seven predicted horse CES1 sequences (Figure 1) which correlated with CES functions (sequences quoted are for horse CES1.2): the active site 'triad' (Ser221, Glu354 and His468); Gly356 or the Z-site, which binds cholesterol-like compounds; Cys95/Cys123 and Cys280/Cys291, the sites for disulfide bond formation (Lockridge *et al.*, 1987); and two microsomal targeting sequences, including the hydrophobic N-terminus signal peptides for CES1 (residues 1-18) and the C-terminal endoplasmic reticulum (ER) retention sequences His-Val/Ile-Glu-Leu, which function in protein retrieval from the Golgi apparatus and in CES retention in the ER lumen (Munro & Pelham, 1987). Other conserved amino acid residues or sequences which are CES1 specific among the horse CES1 sequences examined, correlate with the functions reported for the human CES1 tertiary structure studies (Bencharit *et al.* 2003; 2006; Fleming *et al.* 2005). The *N*-glycosylation site reported for human CES1 (Asn79-Ala80-Thr81) (Kroetz *et al.* 1993) was retained for six of the horse CES1 sequences (either as Asn79-Ala80-Thr81 for horse CES1.1 to CES1.5, or as Asn79-Thr80-Thr80 for horse CES1.6), although other potential carbohydrate binding sites were observed for horse CES1.1, CES1.2, CES1.3 and CES1.4 (Table 3). Given the role of the *N*-glycosylation contributing to CES1 stability and maintaining catalytic efficiency (Kroetz *et al.*, 1993), it is likely that this property has been retained for all of the horse CES subunits. One of two 'ion pairs' that maintains the trimeric-hexameric subunit structures for human CES1 is retained by all horse CES1 subunits (78Lys/183Glu), whereas the second has been retained only for horse CES1.1, but has been lost as a result of a 186Arg → 186Pro

substitution for horse CES1.2-1.6 subunits, respectively. Given the reported oligomeric structure for horse liver CES (Inkerman *et al.* 1975), it is likely that a subunit-subunit binding site has been maintained for the horse CES1-like subunits.

Previous reports have shown that human and baboon CES2 behave as monomers (Pindel *et al.* 1997; Holmes *et al.* 2008c) explained by the absence of key residues supporting the two charge clamps previously reported for human CES1 (Fleming *et al.* 2005). The predicted horse CES2 sequence has undergone amino acid substitutions for those residues contributing to the human CES1 charge clamps: human CES1 Glu183 and Arg186 have been replaced for horse CES2 by amino acids that would not support charge clamp formation: Glu183 → Lys183; and Arg186 → Ala183 (Figure 2). It would appear then that the respective mammalian oligomeric and monomeric subunit structures for CES1 and CES2 have been retained for horse CES1 subunits and CES2, respectively, which is likely to have a major influence on the kinetics and biochemical roles for horse CES1 and CES2. Three dimensional studies have indicated that ligand binding to the human CES1 'Z-site' shifts the trimer-hexamer equilibrium towards the trimer facilitating substrate binding and enzyme catalysis (Redinbo & Potter 2005). This property is predicted to be shared by each of the horse CES1-like subunits while horse CES2 may serve a distinct set of roles as a monomeric enzyme in drug, lipid and cholesterol metabolism in the body.

Other key residues for human CES1 that have been conserved for the horse CES1 subunits include the 'Z-site' Gly355, also found in horse CES2, CES3 and CES7 sequences, and 'side door' residues Val422-Met423-Phe424. The CES1 'product releasing' residue (Phe552 for human CES1) has however undergone a conservative substitution for the horse CES1 subunits and for horse CES2 and CES5 (552Phe→552Leu), whereas horse CES3 has Tyr560 at this site. Human CES1 Met425 has been described as a key residue in regulating the release of fatty acids following the hydrolysis of cholesterol esters within the 'side door' of human CES1; Phe426 apparently serves as a 'switch'; and Phe552 acts as an aromatic releasing residue. Horse CES1 subunits have may retained residues which contribute to CES1 'like' properties reported for human CES1 within these key regions for this enzyme.

The presence of six *CES1* like genes on the horse genome and six predicted CES1 subunits is supported by several lines of evidence presented in this paper: the distinct locations for six *CES1*-like genes on chromosome 3 (Table 1; Figure 4); the identities in each case of the six CES1 amino acid sequences with the predictions derived from BLAT interrogations of the horse genome (Table 1; Figure 1); the distinct nucleotide sequences observed for horse *CES1*-like exons 13 and 14 and intron 13 for each of the predicted *CES1* like genes (Figure 5); the similarities observed in the number and positions for the intron-exon junctions within the predicted horse *CES1* like genes (Figure 1); and the distinct yet similar amino acid sequences observed for the six horse CES1-like subunits (Figure 1; Table 2). Multiple mammalian CES1 like genes have been previously reported, which are closely linked on chromosome 8 (Furihata *et al.* 2003); rat (Ghosh *et al.* 1995), which are located on chromosome 19 in this organism; and cow, where two *CES1*-like genes are located on chromosome 18 (Holmes *et al.* 2008d).

The deduced amino acid sequences for horse CES2, CES3 and CES5 showed a higher level of identity with the corresponding human enzymes and shared similar sequences for the N- and C-termini in each case (Figure 2; Table 3; Figure 7). Horse CES5 also exhibited the distinctive C-terminus for human CES5 which lacked the microsomal retention sequence reported for human CES1 (His564-Ile565-Glu566-Leu567). Domestic cat CES5 (also called cauxin for carboxylesterase-like urinary excreted protein or CES7) is designed for secretion from the epithelial cells of kidney distal tubules where it is apparently functions in regulating the production of a pheromone precursor (Miyazaki *et al.* 2003). This is also

shared by horse CES5 which shows a similar 12 amino acid C-terminus sequence to that of human and other mammalian CES5 subunits (Figure 2) (Holmes *et al.* 2008b).

Predicted secondary structures observed for horse CES1 protein subunits were similar to the reported secondary structure for human CES1 (Figure 1). Some differences were observed however, including an extension of a neighboring α -helix into the 'Z-site' for the horse CES1.2 and CES1.4-CES1.6 subunits; predicted changes in the secondary structures for the 'gate' region, with horse CES1.1, CES1.3 and CES1.4 subunits containing a β -strand in this region; and extensions of the α -helix near the C-terminus end for CES1.2-CES1.6. Horse and human subunits from all four CES families also showed extensive similarities in secondary structures although some predicted differences were observed, including: the 'Z-site' for horse CES2 and CES3, which contain an extended α -helix in this region; and the 'side door' region, for which helical (horse CES5) and strand structures (horse CES2) were observed. Both human and horse CES5 subunits contained a C-terminus helix structure not present in the other CES classes which may be of significance for the secretion role reported for this enzyme (Miyazaki *et al.*, 2003). Predicted tertiary structures for the horse CES1.1, CES2, CES3 and CES5 subunits were sufficiently similar (Figure 3) to be based on the same structure for human CES1-Coenzyme A complex reported by Bencharit and coworkers (2003;2006). It should be noted however that these predicted 3-D structures were based on incomplete sequences in each case (residues 21-551 for horse CES1.1; 29-539 for horse CES2; 32-550 for horse CES3; and 31-540 for horse CES5) which would exclude structures for the N- and C-termini regions for these enzymes.

The phylogenetic tree reported here for human and horse CES1 subunits (Figure 6) was calculated by the progressive alignment of six predicted horse CES1 like amino acid sequences with human, baboon, mouse and cow CES1 like sequences and showed a cluster into two main groups consistent with the horse and other mammalian origins for these enzymes. The tree indicates that a common ancestor for horse *CES1*-like genes (called CA3) postdated the common ancestor for horse and cow and that the gene duplication events forming the horse *CES1*-like genes occurred during horse evolution and likely to be subsequent to the appearance of the *Hyracotherium* (or 'dawn' horse) at ~ 52 MY ago (MacFadden 1987). The following horse CES1 gene duplication events are proposed: the ancestral mammalian *CES1* gene in the *Hyracotherium* ancestor forms the ancestral horse *CES1.5* and *CES1.6* genes; with subsequent gene duplication events generating the horse *CES1.4* gene, and finally the *CES1.2* and *CES1.3* genes. It is suggested that the multiple horse *CES1*-like genes have appeared in parallel with the evolution of the horse during successive epochs, which may include Eocene (34-55 MY ago), Oligocene (24-34 MY ago), Miocene (5-24 MY ago), and Pliocene and Pleistocene epochs (up to 5 MY ago) (MacFadden 1987). Estimated times for horse *CES1*-like gene duplication events are consistent with this proposal (Table 5). The tandem locations for the horse CES1 genes (Table 1; Figure 4) lend support to a mechanism for generating these genes by successive gene duplications through unequal crossover events, similar to that observed for hemoglobin and alpha-satellite genes (Alkan *et al.* 2004). Mammalian CES gene duplications have been previously reported in the mouse, for which CES1 and CES2 like genes exist as multiple copies closely located on chromosome 8 (Dolinsky *et al.*, 2001; Furihata *et al.* 2003; Hosokawa *et al.* 2007); opossum *CES2*-like genes, where three genes are located in tandem on chromosome one (Holmes *et al.*, 2008a); and bovine CES1-like genes, with two genes located together on chromosome 18 (Holmes *et al.* 2008d). A second phylogenetic tree was calculated using previously reported human and predicted horse CES1, CES2, CES3 and CES5 sequences with xenopus and zebrafish CES (Table 1; Figure 7). This is consistent with a recently published phylogenetic tree which proposed that the mammalian CES gene duplication events generating ancestral mammalian *CES1*, *CES2*, *CES3* and *CES5*

genes have predated the common ancestor for eutherian and marsupial mammals and occurred around 328-378 MY ago (Holmes *et al.* 2008a).

Mammalian *CES* genes encode enzymes of broad substrate specificity which are responsible for the detoxification and metabolism of a range of xenobiotics, narcotics and clinical drugs, and catalyze several cholesterol and lipid metabolic reactions (Dolinsky *et al.* 2001; Satoh and Hosokawa, 2006; Redinbo and Potter, 2005). More specific roles for mammalian CES include the activation of lung surfactant (Ruppert *et al.* 2006); detoxifying organophosphate and carbamate poisons (Satoh and Hosokawa, 2006); activating a number of prodrugs used in treating diseases such as influenza, cancer, asthma and high blood pressure (Tabata *et al.* 2004; Mutch *et al.* 2007; Tang *et al.*, 2008), and regulating the production of a pheromone precursor in cat kidney (Miyazaki *et al.* 2003). Mammalian liver is predominantly responsible for drug and xenobiotic clearance from the body with CES1 and CES2 (with CES1 > CES2) playing major roles, following absorption of drugs and xenobiotics into the circulation (Pindel *et al.* 1997; Imai, 2006). Mammalian intestine (with CES2 > CES1) is predominantly responsible for first pass clearance of several drugs and xenobiotics, with the activity occurring mostly in the ileum and jejunum and processed via CES2 (Imai 2006).

CES1 and CES2 also serve different roles in prodrug activation, as shown for the anti-cancer drug irinotecan (CPT-11) which is converted to its active form SN-38 predominantly by CES2 (Xu *et al.* 2002). In contrast with *CES1* and *CES2* genes, mammalian *CES5* is predominantly expressed in peripheral tissues, including brain, kidney, lung and testis (Thierry-Mieg and Thierry-Mieg, 2006), and is a secreted form of CES enzyme due to the absence of the microsomal targeting sequence found at the carboxy-terminus (Figure 2) (Miyazaki *et al.* 2003; 2006; Holmes *et al.* 2008b). Mammalian CES5 may serve in two major roles within mammalian fluids and peripheral tissues, including regulating the production of a pheromone precursor in urine (Miyazaki *et al.* 2003) and contributing to lipid and cholesterol transfer processes within male reproductive fluids (Ecroyd *et al.*, 2005). CES5 has also been identified in human brain (Ota *et al.* 2004) and may contribute to drug metabolism in the cerebrospinal fluid or other fluids of the brain. The metabolic roles for mammalian CES3 have not been extensively investigated however the enzyme is capable of activating prodrugs such as irinotecan (Sanghani *et al.* 2004) and is located in several tissues of the body, including the colon, placenta and neural tissues, such as the cerebellum and hippocampus (Thierry-Mieg and Thierry-Mieg, 2006). CES3 has retained the microsomal targeting sequence at the C-Terminus (QEDL and QEEL for human and horse CES3, respectively) and is therefore likely to be localized within the endoplasmic reticulum, assisting in drug metabolism in peripheral tissues of the body. Given the similarities in structure for horse CES family members, in comparison with the corresponding human enzymes, horse CES1, CES2, CES3 and CES5 may serve similar roles to those reported for other mammals. In addition, given the presence of six horse CES1 like subunits, these enzymes may serve more specialized roles in the horse, which remain to be determined.

In conclusion, horse CES1, CES2, CES3 and CES5 subunits have similar predicted amino acid sequences with the corresponding human enzymes, share key conserved sequences and structures that have been reported for human CES1 and have family specific sequences consistent with the oligomeric and monomeric subunit structures for CES1 and CES2, respectively. The horse genome also contains at least six *CES1*-like genes, which are located in tandem with the *CES5* gene, and with the more distantly located *CES2* and *CES3* genes on chromosome 3. Predicted secondary and tertiary structures for horse CES1, CES2, CES3 and CES5 showed a high degree of conservation with human CES1. Phylogeny studies using horse, human and other mammalian CES1 amino acid sequences indicated that six *CES1* like genes have appeared following the appearance of the 'dawn' horse common ancestor about 55 MY ago, and that horse *CES2*, *CES3* and *CES5* genes have evolved from

respective ancestral genes which appeared prior to the eutherian and marsupial common ancestor. Even though the metabolic roles for horse CES subunits remain to be determined, given the similarities in structures for the CES family members with those reported in human, it is proposed that horse CES1 and CES2 are predominantly responsible for drug clearance (in liver) and first pass metabolism (in intestine and lung); horse CES3 plays a role in drug metabolism in peripheral tissues such as colon and brain; and that horse CES5 catalyses lipid transfer and drug metabolism reactions within male reproductive and neural fluids of the body.

Acknowledgments

This project was supported by NIH Grants P01 HL028972 and P51 RR013986. In addition, this investigation was conducted in facilities constructed with support from Research Facilities Improvement Program Grant Numbers 1 C06 RR13556, 1 C06 RR15456, 1 C06 RR017515.

References

- Alkan C, Eichler EE, Bailey JA, Sahinalp SC, Tuzun E. The role of unequal crossover in alpha-satellite DNA evolution: a computational analysis. *J. Comput. Biol* 2004;11:933–944. [PubMed: 15700410]
- Altschul SF, Madden L, Schäffer AA, Zhang J, Zhang Z, Webb Miller W, Lipman DJ. Gapped ST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–3402. [PubMed: 9254694]
- Andersson L, Sandberg K, Adalsteinsson S, Gunnarsson E. Linkage of the equine serum esterase (Es) and mitochondrial glutamate oxaloacetate transaminase (GOTM) loci. A horse-mouse homology. *J. Heredity* 1983;74:361–364.
- Bencharit S, Edwards CC, Morton CL, Howard-Williams EL, Kuhn P, Potter PM, Redinbo MR. Multisite promiscuity in the processing of endogenous substrates by human carboxylesterase 1. *J. Mol. Biol* 2006;363:201–214. [PubMed: 16962139]
- Bencharit S, Morton CL, Xue Y, Potter PM, Redinbo MR. Structural basis of heroin and cocaine metabolism by a promiscuous human drug-processing enzyme. *Nature Struct. Biol* 2003;10:349–356. [PubMed: 12679808]
- Burch J. The purification and properties of horse liver esterase. *Biochem. J* 1954;58:415–26. [PubMed: 13208632]
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Thompson JD. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 2003;31:3497–3500. [PubMed: 12824352]
- Clark HF, Gurney AL, Abaya E, Baker K, Baldwin DT, Brush J, Chen J, Chow B, Chui C, Crowley C, Currell B, Deuel B, Dowd P, Eaton D, Foster JS, Grimaldi C, Gu Q, Hass PE, Heldens S, Huang A, Kim HS, Klimowski L, Jin Y, Johnson S, Lee J, Lewis L, Liao D, Mark MR, Robbie E, Sanchez C, Schoenfeld J, Seshagiri S, Simmons L, Singh J, Smith V, Stinson J, Vagts A, Vandlen RL, Watanabe C, Wieand D, Woods K, Xie M-H, Yansura DG, Yi G, Yuan J, Zhang M, Zhang Z, Goddard AD, Wood WI, Godowski PJ, Gray AM. The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. *Genome Res* 2003;13:226–2270.
- Cygler M, Schrag JD, Sussman JL, Harel M, Silman L, Gentry MK, Doctor BP. Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases and related proteins. *Protein Sci* 1993;2:366–382. [PubMed: 8453375]
- Dolinsky VW, Sipione S, Lehner R, Vance DE. The cloning and expression of murine triacylglycerol hydrolase cDNA and the structure of the corresponding gene. *Biochim. Biophys. Acta* 2001;1532:162–172. [PubMed: 11470237]
- Dolvik NI, Klemetsdal G. The effect of arthritis in the carpal joint on performance in Norwegian cold-blooded trotters. *Vet. Res. Commun* 1996;20:505–512. [PubMed: 8950831]
- Ecroyd H, Belghazi M, Dacheux J-L, Miyazaki M, Yamashita T, Gatti J-L. An epididymal form of cauxin, a carboxylesterase-like enzyme, is present and active in mammalian male reproductive fluids. *Biol. Reprod* 2006;74:439–447. [PubMed: 16251497]

- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* 2007;2:953–971.
- Fleming CD, Benchari S, Edwards CC, Hyatt JL, Trurkan L, Bai B, Fraga C, Morton CL, Howard-Williams EL, Potter PM, Redinbo MR. Structural insights into drug processing by human carboxylesterase 1: tamoxifen, Mevastatin, and inhibition by Benzil. *J. Mol. Biol* 2005;352:165–177. [PubMed: 16081098]
- Furihata T, Hosokawa M, Nakata F, Satoh T, Chiba K. Purification, molecular cloning and functional expression of inducible acylcarnitine hydrolase in C57BL/6J mouse belonging to the carboxylesterase gene family. *Arch. Biochem. Biophys* 2003;416:101–109. [PubMed: 12859986]
- Gershater M, Sharples K, Edwards R. Carboxylesterase activities toward pesticide esters in crops and weeds. *Phytochem* 2006;67:2561–7.
- Ghosh S. Cholesteryl ester hydrolase in human monocyte/macrophage: cloning, sequencing and expression of full-length cDNA. *Physiol. Genomics* 2000;2:1–8. [PubMed: 11015575]
- Ghosh S, Mallonee DH, Hylemon PB, Grogan WM. Molecular cloning and expression of rat hepatic neutral cholesteryl ester hydrolase. *Biochim. Biophys. Acta* 1995;1259:305–312. [PubMed: 8541339]
- Holmes RS, Chan J, Cox LA, Murphy WM, VandeBerg JL. Opossum carboxylesterases: sequences, phylogeny and evidence for CES duplication events predating the marsupial-eutherian common ancestor. *BMC Evol. Biol* 2008a;8:54. [PubMed: 18289373]
- Holmes RS, Cox LA, Vandeberg JL. Mammalian carboxylesterase 5: comparative biochemistry and genomics. *Comp. Biochem. Physiol. D* 2008b;3:195–204.
- Holmes RS, Glenn JP, VandeBerg JL, Cox LA. Baboon carboxylesterases 1 and 2: sequences, structures and phylogenetic relationships with human and other primate carboxylesterases. *J. Med. Primatol.* 2008c in press.
- Holmes RS, Cox LA, VandeBerg JA. Bovine carboxylesterases: evidence for two CES1 and five families of CES genes on chromosome 18. *Comp. Biochem. Physiol. D.* 2008d in press.
- Horse Genome Project. 2008. <http://www.uky.edu/Ag/Horsemap/>
- Hosokawa M, Furihata T, Yaginuma Y, Yamamoto N, Nao K, Ayako F, Yuko N, Testuo S, Kan C. Genomic structure and transcriptional regulation of the rat, mouse and human carboxylesterase genes. *Drug Metab. Revs* 2007;39:1–15. [PubMed: 17364878]
- Imai T. Human carboxylesterase isozymes: catalytic properties and rational drug design. *Drug Metab. Pharmacogen* 2006;21:173–185.
- Inkerman PA, Scott K, Runnegar MT, Hamilton SE, Bennett EA, Zerner B. Carboxylesterases (EC 3.1.1). Purification and titration of chicken, sheep and horse liver carboxylesterases. *Canad. J. Biochem* 1975;53:536–46. [PubMed: 1139397]
- Kelly L, Postiglioni A, de Andres DF. Genetic characterisation of the Uruguayan Creole horse and analysis of relationships among horse breeds. *Res. Vet. Science* 2002;72:69–73.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Hussler B. The human genome browser at UCSC. *Genome Res* 2003;12:994–1006.
- Kopp J, Schwede T. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res* 2004;32:D230–D234. [PubMed: 14681401]
- Kroetz DL, McBride OW, Gonzalez FJ. Glycosylation-dependent activity of Baculovirus-expressed human liver carboxylesterases: cDNA cloning and characterization of two highly similar enzyme forms. *Biochemistry* 1993;32:11606–11617. [PubMed: 8218228]
- Langmann T, Becker A, Aslanidis C, Notka F, Ulrich H, Schmidt G. Structural organization and characterization of the promoter region of a human carboxylesterase gene. *Biochim. Biophys. Acta* 1997;1350:65–74. [PubMed: 9003459]
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. ClustalW and ClustalX version 2. *Bioinformatics* 2007;23:2947–2948. [PubMed: 17846036]
- Leinweber FJ. Possible physiological roles of carboxyl ester hydrolases. *Drug Metab. Revs* 1987;18:379–439. [PubMed: 3286170]

- Livesay GJ, O'Neill T, Hannant D, Yadav MP, Mumford JA. The outbreak of equine influenza (H3N8) in the United Kingdom in 1989: diagnostic use of an antigen capture ELISA. *Vet. Rec* 1987;133:515–9. [PubMed: 8310627]
- Lockridge O, Adkins S, La Due BN. Location of disulfide bonds within the sequence of human serum cholinesterase. *J. Biol. Chem* 1987;262:12945–12952. [PubMed: 3115973]
- MacFadden BJ. Fossil horses from 'Eohippus' (Hyracotherium) to Equus: scaling, Cope's law, and the evolution of body size. *Paleobiology* 1987;12:355–69.
- Marsh S, Xiao M, Yu J, Ahluwalia R, Minton M, Freimuth RR, Kwok P-Y, McLeod HL. Pharmacogenomic assessment of carboxylesterases 1 and 2. *Genomics* 2004;84:661–8. [PubMed: 15475243]
- McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics* 2000;16:404–405. [PubMed: 10869041]
- Miyazaki M, Kamiie K, Soeta S, Taira H, Yamashita T. Molecular cloning and characterization of a novel carboxylesterase-like protein that is physiologically present at high concentrations in the urine of domestic cats (*Felis catus*). *Biochem. J* 2003;370:101–110. [PubMed: 12401131]
- Munro S, Pelham HR. A C-terminal signal prevents secretion of luminal ER proteins. *Cell* 1987;48:899–907. [PubMed: 3545499]
- Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. Molecular phylogenetics and the origins of placental mammals. *Nature* 2001;409:614–8. [PubMed: 11214319]
- Mutch E, Nave R, McCracken N, Zech K, Williams FM. The role of esterases in the metabolism of ciclesonide to desisobutryl-ciclesonide in human tissue. *Biochem. Pharmacol* 2007;73:1657–1664. [PubMed: 17331475]
- Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R, Wakamatsu A, Hayashi K, Sato H, Nagai K, Kimura K, Makita H, Sekine M, Obayashi M, Nishi T, Shibahara T, Tanaka T, Ishii S, Yamamoto J, Saito K, Kawai Y, Isono Y, Nakamura Y, Nagahari K, Murakami K, Yasuda T, Iwayanagi T, Wagatsuma M, Shiratori A, Sudo H, Hosoiri T, Kaku Y, Kodaira H, Kondo H, Sugawara M, Takahashi M, Kanda K, Yokoi T, Furuya T, Kikkawa E, Omura Y, Abe K, Kamihara K, Katsuta N, Sato K, Tanikawa M, Yamazaki M, Ninomiya K, Ishibashi T, Yamashita H, Murakawa K, Fujimori K, Tanai H, Kimata M, Watanabe M, Hiraoka S, Chiba Y, Ishida S, Ono Y, Takiguchi S, Watanabe S, Yosida M, Hotuta T, Kusano J, Kanehori K, Takahashi-Fujii A, Hara H, Tanase TO, Nomura Y, Togiya S, Komai F, Hara R, Takeuchi K, Arita M, Imose N, Musashino K, Yuuki H, Oshima A, Sasaki N, Aotsuka S, Yoshikawa Y, Matsunawa H, Ichihara T, Shiohata N, Sano S, Moriya S, Momiyama H, Satoh N, Takami S, Terashima Y, Suzuki O, Nakagawa S, Senoh A, Mizoguchi H, Goto Y, Shimizu F, Wakebe H, Hishigaki H, Watanabe T, Sugiyama A, Takemoto M, Kawakami B, Yamazaki M, Watanabe K, Kumagai A, Itakura S, Fukuzumi Y, Fujimori Y, Komiyama M, Tashiro H, Tanigami A, Fujiwara T, Ono T, Yamada K, Fujii Y, Ozaki K, Hirao M, Ohmori Y, Kawabata A, Hikiji T, Kobatake N, Inagaki H, Ikema Y, Okamoto S, Okitani R, Kawakami T, Noguchi S, Itoh T, Shigeta K, Senba T, Matsumura K, Nakajima Y, Mizuno T, Morinaga M, Sasaki M, Togashi T, Oyama M, Hata H, Watanabe M, Komatsu T, Mizushima-Sugano J, Satoh T, Shirai Y, Takahashi Y, Nakagawa K, Okumura K, Nagase T, Nomura N, Kikuchi H, Masuho Y, Yamashita R, Nakai K, Yada T, Nakamura Y, Ohara O, Isogai T, Sugano S. Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nature Genet* 2004;36:40–45. [PubMed: 14702039]
- Pindel EV, Kedishvili NY, Abraham TL, Brzezinski MR, Zhang A, Dean RA, Bosron WF. Purification and cloning of a broad substrate specificity human liver carboxylesterase that catalyzes the hydrolysis of cocaine and heroin. *J. Biol. Chem* 1997;272:14769–14775. [PubMed: 9169443]
- Raum R, Sterner KN, Noviello CM, Stewart C-B, Disotell TR. Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *J. Human Evol* 2005;48:237–57. [PubMed: 15737392]
- Redinbo MR, Potter PN. Mammalian carboxylesterases: from drug targets to protein therapeutics. *Drug Disc. Today* 2005;10:313–20.
- Ring J, Seifert J, Brendel W. High incidence of horse serum protein allergy in various autoimmune disorders. *J. Allergy Clin. Immunol* 1977;59:185–189. [PubMed: 320246]

- Ruppert C, Bagheri A, Markart P, Schmidt R, Seeger W, Gunther A. Liver carboxylesterase cleaves surfactant protein (SP-B) and promotes surfactant subtype conversion. *Biochem. Biophys. Res. Commun* 2006;348:1449–1454. [PubMed: 16919595]
- Sanghani SP, Quinney SK, Fredenberg TB, Davis WI, Murry DJ, Bosron WF. Hydrolysis of irinotecan and its oxidative metabolites, 7-ethyl-10-[4-N(5-aminopentanoic acid)-1-piperidino] carbonyloxycamptothecin and 7-ethyl-10-[4-(1-piperidino)-1 amino]-carbonyloxycamptothecin, by human carboxylesterases CES1A1, CES2, and a newly expressed carboxylesterase isoenzyme, CES3. *Drug Metab. Disp* 2004;32:505–511.
- Satoh T, Hosokawa M. Structure, function and regulation of carboxylesterases. *Chem.-Biol. Inters* 2006;162:195–211.
- Schewer H, Langmann T, Daig R, Becker A, Aslandis C, Schmitz G. Molecular cloning and characterization of a novel putative carboxylesterase, present in human intestine and liver. *Biochem. Biophys. Res. Commun* 1997;233:117–120.
- Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 2003;31:3381–3385. [PubMed: 12824332]
- Shibita F, Takagi Y, Kitajima M, Kuroda T, Omura T. Molecular cloning and characterization of a human carboxylesterase gene. *Genomics* 1993;17:76–82. [PubMed: 8406473]
- Stoops JK, Hamilton SE, Zerner B. Carboxylesterases (EC 3.1.1). A comparison of some kinetic properties of horse, sheep, chicken, pig, and ox liver carboxylesterases. *Can. J. Biochem* 1975;53:565–73. [PubMed: 237618]
- Tabata T, Katoh M, Tokudome S, Nakajima M, Yokoi T. Identification of the cytosolic carboxylesterase catalyzing the 5'-deoxy-5-fluorocytidine formation from capecitabine in human liver. *Drug Metab. Dispos* 2004;32:1103–1110. [PubMed: 15269188]
- Tang X, Wu H, Wu Z, Wang G, Wang Z, Zhu D. Carboxylesterase 2 is downregulated in colorectal cancer following progression of the disease. *Cancer Invest* 2008;26:178–181. [PubMed: 18259949]
- Thierry-Mieg D, Thierry-Mieg J. AceView: A comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* 2006;7:S12. <http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/index.html?human>. [PubMed: 16925834]
- von Heijne G. Patterns of amino acids near signal-sequence cleavage sites. *Eur. J. Biochem* 1983;133:17–21. [PubMed: 6852022]
- Wang H, Gilham D, Lehner R. Proteomic and lipid characterization of apo-lipoprotein B-free luminal lipid droplets from mouse liver microsomes: implications for very low density lipoprotein assembly. *J. Biol. Chem* 2007;282:33218–33226. [PubMed: 17848546]
- Xu G, Zhang W, Ma MK, McLeod HL. Human carboxylesterase 2 is commonly expressed in tumor tissue and is correlated with the activation of irinotecan. *Clin. Cancer Res* 2002;8:2605–2611. [PubMed: 12171891]

Figure 1. Amino Acid Sequence Alignments for Human and Horse CES1 Subunits

See Table 1 for sources of CES sequences; * shows identical residues for human (hs: *Homo sapiens*) and horse CES1 subunits; : 2 alternate residues; . 3 alternate residues. Residues involved in endoplasmic reticulum processing at N- (Signal peptide) and C- termini (MTS-microsomal (endoplasmic reticulum) targeting sequence) are in **red**; N-glycosylation residues at 79NAT (Human CES1) and potential N-glycosylation sites are in **blue**; active site triad residues are in pink. 'side door', 'product releasing' residues and cholesterol binding Gly residue (**Z site**) for human CES1 are in brown; Disulfide bond ---- Cys residues for human CES1 are shown as C (Lockridge *et al* 1987). Charge clamp residues identified for human CES1 are in pink; Helix (Human CES1 or predicted helix; Sheet (Human CES1) or predicted sheet. **Bold font underlined** shows residues at known or predicted exon junctions. Exons are numbered for human CES1.

Figure 2. Amino Acid Sequence Alignments for Human CES1 and for Human and Horse CES2, CES3 and CES5 Subunits

See Table 1 for sources of CES sequences; * shows identical residues for human (Hs for *Homo sapiens*) and horse (Ec for *Equus caballus*) CES1 subunits; : 2 alternate residues; . 3 alternate residues. Residues involved in endoplasmic reticulum processing at N- (Signal peptide) and C- termini (MTS-microsomal (endoplasmic reticulum) targeting sequence) are in **red**; N-glycosylation residues at 79NAT (Human CES1) and potential N-glycosylation sites are in **blue**; active site triad residues are in pink. 'side door', 'product releasing' residues and cholesterol binding Gly residue (**Z site**) for human CES1 are in brown; Disulfide bond ---- Cys residues for human CES1 are shown as C. Charge clamp residues identified for human CES1 are in pink; Helix (Human CES1 or predicted helix; Sheet (Human CES1) or predicted sheet. **Bold font underlined** shows residues at known or predicted exon junctions. Exons are numbered for human CES1.

Figure 3. Predicted Three Dimensional Structures for Horse CES1.1, CES2, CES3 and CES5 Subunits

Predicted horse CES1.1, CES2, CES3 and CES7 3-D structures were obtained using the SWISS MODEL web site <http://swissmodel.expasy.org/workspace/index.php?> and known amino acid sequences (see Table 1 for sources). The rainbow color code describes the 3-D structures from the N- (blue) to C-termini (red color). The structures are based on the known 3-D structure for human CES1 complexed with Coenzyme A (Bencharit et al., 2003) for residues 21-551 (CES1); 29-539 (CES2); residues 32-550 (CES3); and 31-540 (CES5).

Chromosome 3

22	79	32	32	31	32	90	32	9	10	Gene Size kb
CES1.1	CES1.2	CES1.3	CES1.4	CES1.5	CES1.6	CES1.7	CES5	CES2	CES3	Gene order
37	18	27	30	31	34	21	9,100	6		Distance kb

Figure 4. Locations of CES Genes on Horse Chromosome 3
Numbers refer to kilobases of DNA.

[illegible]

Figure 5. Nucleotide Sequence Alignments for Horse CES1 Genes: Predicted Exons 13 and 14; and Predicted Intron 13
Exon sequences are in blue. Intron sequences are in black.

Figure 6. Phylogenetic Tree of Mammalian CES1-like Sequences

The tree is labeled with the gene name, the species name and the calculated genetic distances between pairs of sequences. Note the separation into 2 clusters for mammalian CES1: six horse CES1 like genes; and other mammalian CES1 like genes, including human, baboon, mouse and cow. Common ancestral CES1-like genes are shown: CA1 (human and baboon CES1); CA2 (mammalian CES1-like genes other than horse CES1-like genes); CA3 (horse CES1-like genes); CA4 (horse CES1-like genes other than CES1.6); CA5 (horse CES1.2, CES1.3 and CES1.4-like genes); CA6 (horse CES1.2 and CES1.3-like genes). Phylogenetic trees were constructed using an amino acid alignment from a ClustalW-derived alignment of CES protein sequences, obtained with default settings and corrected for multiple substitutions (Chenna *et al* 2003; Larkin *et al.* 2007) [<http://www.ebi.ac.uk/clustalw/>]. Pairwise scores were calculated using the number of identities in the best alignment divided by the number of residues compared. These percent identity scores were then converted to genetic distances by dividing by 100 and subtracting from 1.0 to give the number of differences per site.

Figure 7. Phylogenetic Tree of Human and Horse CES1, CES2, CES3 and CES5 with Xenopus and Zebrafish CES Sequences

The tree is labeled with the gene name and the species name. Note the separation into 4 clusters for human and horse *CES* genes: *CES1*, *CES2*, *CES3* and *CES5*. The gene duplication events generating the four distinct gene families (*CES1*, *CES2*, *CES3* and *CES5*) have been previously estimated to have occurred ~ 328-378 million years ago (Holmes et al., 2008a). Phylogenetic trees were constructed using an amino acid alignment from a ClustalW-derived alignment of CES protein sequences, obtained with default settings and corrected for multiple substitutions (Chenna *et al* 2003; Larkin *et al.* 2007) [<http://www.ebi.ac.uk/clustalw/>].

Table 1

Horse, Other Mammalian, Xenopus and Zebrafish CES Genes

CES	<i>I</i> _{GenBank mRNA}	No. of Amino Acids	Chromosome Location	Strand	Gene	Size kbs	Predicted subunit pI	Predicted subunit MW
Gene	² NCBI Locus							
	³ UNIPROT ID							
Horse CES1.1	² XP1491160	565	3:7,903,234-7,982,438	Negative	22	5.5	5.5	61,486
Horse CES1.2	² XP1491576	565	3:8,000,851-8,032,477	Negative	79.2	5.6	5.6	61,556
Horse CES1.3	² XP1491752	565	3:8,048,349-8,079,872	Negative	31.6	5.5	5.5	61,935
Horse CES1.4	² XP1491878	565	3:8,106,399-8,137,176	Negative	31.5	5.6	5.6	61,742
Horse CES1.5	² XP1491878	565	3:8,167,845-8,199,077	Negative	30.8	6.5	6.5	61,037
Horse CES1.6	² XP1915508	566	3:8,233,589-8,266,613	Negative	31.9	5.5	5.5	64,050
Horse CES2	² XP19115822	559	3:17,402,324-17,411,849	Positive	8.9	5.5	5.5	61,670
Horse CES3	² XP1496251	571	3:17,417,834-17,428,268	Positive	10.4	6.2	6.2	62,719
Horse CES5	² XP1493477	575	3: 8,287,811-8,319,572	Negative	31.8	5.6	5.6	63,859
Human CES1 Baboon	¹ L07765 ³ p23141	567	16:54,394,266-54,424,4894	Negative	30	6.2	6.2	62,521
CES1	¹ FJ147178	567				6.1	6.1	62,483
Mouse CES1 Bovine	¹ Y12887 ³ Q8VCC2	565	8:95,826,807-95,861,053	Negative	34.2	5.6	5.6	62,680
CES1.1	¹ BC102781	565	18: 24,344,904-24,371,523	Negative	26.6	6.3	6.3	61,723
Human CES2	¹ BX538086 ³ O00748	559	16:65,527,040-65,535,426	Positive	8.4	5.7	5.7	61,807
Human CES3	¹ AK025389 ³ Q6UWW8	571	16:65,552,712-65,564,450	Positive	11.7	5.4	5.4	62,282
Human CES5	¹ AK056109 ³ Q6NT32	575	16:54,437,867-54,466,634	Negative	28.8	6	6	63,926
Human CES6	¹ FLJ37464 ³ Q5XG92	561	16:65,580,177-65,600,543	Positive	20.4	9.4	9.4	63,529
Zebrafish CES	¹ BC091470 ³ Q1LUZ9	548	18: 16,714,030 16,721,981	Negative	7.9	6	6	60,297

CES	¹ GenBank mRNA	No. of Amino Acids	Chromosome Location	Strand	Gene Size kbs	Predicted subunit pI	Predicted subunit MW
Gene	² NCBI Locus						
	³ UNIPROT ID						
Xenopus CES	¹ BC082503 ³ A1L2G7	557	scaffold170:71,809-89,385	Negative	17.6	5	61,707

¹ Locus derived from a BLAST of the NCBI data base (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>)

² GenBank ID (<http://www.ncbi.nlm.nih.gov/Genbank/>); and

³ UNIPROT ID of the CES protein using the SWISS-PROT Web Browser (<http://au.expasy.org/>). pI refers to theoretically determined isoelectric point. kbs refers to kilobases of DNA nucleotides.

Table 2
Percentage Identities for Horse and Other Mammalian CES Subunit Amino Acid Sequences

Mammal	CES Gene	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 11	Site 12	Site 13	Site 14	Site 15	No of sites
Human	CES1	79NAT															1
Horse	CES1.1	79NAT								382NSS							2
	CES1.2	79NAT								382NSS							2
	CES1.3	79NAT			108NIS												1
	CES1.4	79NAT											489NFS	506NPS			3
	CES1.5	79NAT															1
	CES1.6																0
Human	CES2			111NMT			276NLS										2
Horse	CES2			111NQS			276NLS					472NYS					3
Human	CES3		105NSS														1
Horse	CES3					212NIT		285NSS			423NFS				519NQS		4
Human	CES5							281NAS	363NKS						511NLT	522NMS	4
Horse	CES5							281NSS	363NKS						511NKT	522NVS	4

Table 3
Key Residues and Sequences for Human and Horse CES Subunits

CES Gene	human CES1	horse CES1.1	horse CES1.2	horse CES1.3	horse CES1.4	horse CES1.5	horse CES1.6	baboon CES1	mouse CES1
human CES1	100	80	80	81	77	80	69	94	73
horse CES1.1	80	100	95	91	85	85	74	80	73
horse CES1.2	80	95	100	93	85	86	74	74	73
horse CES1.3	81	91	93	100	85	88	75	81	74
horse CES1.4	77	85	85	85	100	85	75	78	74
horse CES1.5	80	85	86	88	85	100	76	81	73
horse CES1.6	69	74	74	75	75	76	100	71	67
baboon CES1	94	80	81	81	78	81	71	100	73
mouse CES1	73	73	73	74	73	73	67	73	100

human CES1	horse CES1.1	human CES2	horse CES2	human CES3	horse CES3	human CES5	horse CES5	
human CES1	100	80	45	44	42	43	42	43
horse CES1.1	80	100	45	44	44	43	42	45
human CES2	45	45	100	77	41	46	43	45
horse CES2	44	44	77	100	49	49	43	46
human CES3	42	44	41	49	100	75	41	43
horse CES3	43	43	46	49	75	100	41	44
human CES5	42	42	43	43	41	41	100	77
horse CES5	43	45	45	46	43	44	77	100

¹ N-signal peptides identified using SignalP 3.0 web based tools (Emanuelsson *et al* 2007);

² predicted 'ion pair' supporting CES oligomers (see Fleming *et al.* 2005) and predicted key Z-site, 'side door' and product releasing residues (see Bencharit *et al.* 2003; 2006);

³ based on Cygler *et al.* (1983); based on Munro & Pelham, 1983.

Table 4
Potential N-Glycosylation Sites for Horse and Human CES Subunits

CES1 Gene Common Ancestor	Genetic Distance	Common Ancestor MY Ago
CES1.1-CES1.2 [CA1]	0.022	16
CA1-CES1.3 [CA2]	0.034	25
CA2-CES1.5 [CA3]	0.064	47
CA3-CES1.4 [CA4]	0.072	52
CA4-CES1.6	0.129	94

¹ Identified N-glycosylation site for human CES1 (Kroetz *et al.* 1993). Standard amino acid abbreviations were used.