Molecular Phylogenetics and Evolution xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Phylogenetic relationships among Perissodactyla: *Secretoglobin 1A1* gene duplication and triplication in the Equidae family

Olivier Côté [a], Laurent Viel [b], Dorothee Bienzle [a,*]

[a] Department of Pathobiology, University of Guelph, Guelph, Ontario, Canada
[b] Department of Clinical Studies, University of Guelph, Guelph, Ontario, Canada

ARTICLE INFO

Article history:
Received 11 June 2013
Revised 8 August 2013
Accepted 16 August 2013
Available online xxxx

Keywords:
CC10
Clara cell secretory protein
Equid
Donkey
Horse
Zebra

ABSTRACT

Secretoglobin family 1A member 1 (SCGB 1A1) is a small anti-inflammatory and immunomodulatory protein that is abundantly secreted in airway surface fluids. We recently reported the existence of three distinct *SCGB1A1* genes in the domestic horse genome as opposed to the single gene copy consensus present in other mammals. The origin of *SCGB1A1* gene triplication and the evolutionary relationship of the three genes amongst Equidae family members are unknown. For this study, *SCGB1A1* genomic data were collected from various *Equus* individuals including *E. caballus*, *E. przewalskii*, *E. asinus*, *E. grevyi*, *and E. quagga*. Three *SCGB1A1* genes in *E. przewalskii*, two *SCGB1A1* genes in *E. asinus*, and a single *SCGB1A1* gene in *E. grevyi and E. quagga* were identified. Sequence analysis revealed that the non-synonymous nucleotide substitutions between the different equid genes coded for 17 amino acid changes. Most of these changes localized to the SCGB 1A1 central cavity that binds hydrophobic ligands, suggesting that this area of SCGB 1A1 evolved to accommodate diverse molecular interactions. Three-dimensional modeling of the proteins revealed that the size of the SCGB 1A1 central cavity is larger than that of SCGB 1A1A. Altogether, these findings suggest that evolution of the *SCGB1A1* gene may parallel the separation of caballine and non-caballine species amongst Equidae, and may indicate an expansion of function for *SCGB1A1* gene products.

1. Introduction

Secretoglobin family 1A member 1 (SCGB 1A1) is a small protein predominantly secreted by airway and uterine epithelial cells, and known to have anti-inflammatory and immunomodulatory functions. In most mammals, the *SCGB1A1* gene spans between 2800 and 4300 nucleotides and includes three exons encoding a mature secreted protein of 70–75 amino acids. In contrast to the single gene consensus observed in other mammals, we recently reported the presence of multiple *SCGB1A1* genes in the *E. caballus* genome (Côté et al., 2012). The three genes are closely positioned within a 512 kb region of equine chromosome 12 and share a high level of sequence identity, including large segments of the 5′- and 3′-untranslated regions (UTR). These observations suggest that an ancestral *SCGB1A1* gene triplication event occurred in horses. Interestingly, the three genes appear to have evolved independently, and can be distinguished by a combination of three single nucleotide polymorphisms (SNPs) located in the second exon of each gene. These patterns are referred to as signature sequences, which are highly conserved within individuals from different breeds.

Domestic horses (*Equus caballus*) are members of the Perissodactyla order, which contains three families named the Tapiridae, the Rhinocerotidae and the Equidae. The latter family includes eight existing species of horses, zebras and asses, organized into caballine (domestic and Przewalski's horses) and non-caballine (zebras and asses) groups (Wilson and Reeder, 2005). The Przewalski's horse (*Equus przewalskii*) represents the closest wild living relative to the domestic horse known for (1) never having been successfully domesticated and (2) having been reintroduced to its native environment after near extinction. Thus, the Przewalski's horse has been commonly referred to as the domestic horse ancestor. However, the origin of these two lineages continues to be debated (Goto et al., 2011; Lau et al., 2009; Wade et al., 2009). More recently, a phylogenetic study using massively parallel sequencing technology suggested that Przewalski's horses retained a comparable predecessor genetic background and/or experienced a similar gene flow as the domestic horse (Goto et al., 2011), following a diversification episode that occurred around 250,000–431,000 years ago (Lau et al., 2009; Orlando et al., 2013), also see graphical abstract. Interestingly, despite a different karyotype, Przewalski's and domestic horses can interbreed and produce fertile offspring (Short et al., 1974), a

* Corresponding author. Address: Department of Pathobiology, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada. Fax: +1 519 824 5930.
E-mail addresses: dbienzle@uoguelph.ca, lviel@uoguelph.ca, ocote@uoguelph.ca (D. Bienzle).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.08.012

characteristic not retained upon breeding with non-caballine individuals.

Likewise, the phylogenetic positions occupied by the different members of the non-caballine class are unresolved. It is assumed that the common ancestor of donkeys (*Equus asinus*) and zebras (*Equus zebra*, *Equus quagga burchellii* [henceforth called *E. quagga*], and *Equus grevyi*) diverged from the caballine species approximately 2.0–2.4 million years ago (Oakenfull and Clegg, 1998). Within non-caballine individuals, sequence analysis of mitochondrial, coding, and non-coding nuclear DNA clearly demonstrated the divergence of the donkey from the zebras and Asiatic asses as the first event, but the correct position of the different species within these two latter clades remains obscure (Steiner et al., 2012). Continuing molecular studies to characterize the genetic divergence within the Equidae family will undoubtedly contribute to refine the phylogenetic organization established to date. Interestingly, evidence suggests gene translocations, tandem fusions, and multiple inversions comprise the most important mechanisms leading to the chromosomal differences observed in living equids (Yang et al., 2003).

Whole genome sequencing has not been undertaken for Equidae members other than the domestic horse, limiting the data available to investigate the evolutionary relationship between *SCGB1A1* genes in additional caballine and non-caballine species. The main objective of this study was to provide the first extensive molecular characterization of *SCGB1A1* in the Equidae family. Remarkably, the presence of multiple *SCGB1A1* genes was not a consistent finding throughout *Equus* species. Rather, results show evidence of three genes in *E. przewalskii*, two genes in *E. asinus*, and a single gene in *E. grevyi* and *E. quagga*. We discuss the structure of the various *SCGB1A1* isoforms that provide insight into the origin of *SCGB1A1* gene triplication.

2. Materials and methods

2.1. Genome data sources

SCGB1A1 nucleotide and amino acid sequences were obtained from the National Center for Biotechnology Information (NCBI) database as follow; *E. caballus SCGB1A1-B*, JQ906259.1; *E. caballus SCGB1A1-A*, JQ906260.1; *E. caballus SCGB1A1A*, JQ906261.1; *Canis familiaris*, XM_533268.2; *Sus scrofa*, XM_03122618.1; *Bos taurus*, NM_001076976.2; *Oryctolagus cuniculus*, NM_001082237.1; *Pan troglodytes*, XM_001152276.1 and XP_001152343.1; *Homo sapiens*, NM_003357.4 and NP_003348.1. Sequence alignments were generated using Geneious Pro software (Drummond et al., 2011) using the following parameters: Needleman-Wunsch alignment, Gap open penalty 12, Gap extension penalty 3, cost matrix 65%.

2.2. Characterization of SCGB1A1 sequences in Equidae family members

Blood samples, buccal swabs or paraffin-embedded tissues were available from nine adult *E. przewalskii*, five *E. asinus*, seven *E. grevyi*, and four *E. quagga* individuals (Toronto Zoo, ON; and archives of the Department of Pathobiology, University of Guelph, ON). Genomic DNA purification was performed according to the manufacturer's protocol (DNA Mini kit, Qiagen, Mississauga, ON). A 526 bp genomic region covering *SCGB1A1* signature sequences was amplified by PCR using the forward UGn-F (5′-GCT TCT GCA GRA ATC TGC CAG AG-3′) and reverse UGn-R (5′-CTA AGC ACA CAG TGG GCT CTY TRC-3′) primers. PCR amplifications were carried out in duplicate using a Platinum Taq polymerase PCR kit (Invitrogen, Mississauga, ON). Each reaction was performed in a final volume of 20 μL, including 2 μL of 10× PCR buffer, 0.2 mM dNTPs, 2 mM $MgSO_4$, 0.3 μM of each primer, 2 U of Platinum Taq, and 1 μL of template DNA (100 ng). Conditions for amplification were 1 min at 94 °C followed by 30 cycles of 94 °C for 30 s; 57 °C for 30 s; and 72 °C for 1 min, and a final extension at 72 °C for 7 min. PCR products were subjected to electrophoresis; a band of appropriate size was excised from the gel, purified and submitted for automated sequencing (Laboratory Services Division, Guelph, ON). Amplicons were analyzed in duplicate using reverse and forward primer sequencing strategies.

The *SCGB1A1* genomic region coding for the full-length mature secreted protein (including exons 2–3) was amplified by end-point limiting dilution (EPLD)-PCR in a total volume of 20 μL, as described before (Côté et al., 2012), using primers UGe-F (5′-GGG CCT GGA TTC ATG TGC TTT C-3′) and UGe-R (5′-GCA ACT GCA GCT CTT CAG CTT CT-3′). The sequences were deposited in GenBank under accession numbers: *E. przewalskii SCGB1A1P*, KC853012; *E. przewalskii SCGB1A1*, KC853013; *E. przewalskii SCGB1A1A*, KC853014; *E. asinus SCGB1A1*, KC853015; *E. asinus SCGB1A1A*, KC853016; *E. grevyi SCGB1A1*, KC853017; *E. quagga SCGB1A1*, KC853018. The predicted amino acid sequences were generated using Geneious Pro software, and the isoelectric point was calculated using the bisection method with the following values for amino acids: D = −3.9, E = −4.1, C = −8.5, Y = −10.1, H = 6.5, K = 10.8, R = 12.5.

Total RNA was isolated from frozen or paraffin-embedded lung or uterus tissues (RNeasy, Qiagen) according to the manufacturer's recommendations. RNA integrity was verified through capillary electrophoresis in a 2100 Bioanalyzer (Agilent Technologies) prior to analysis. Complementary DNA (cDNA) and EPLD-PCR assays were performed as described above, and PCR products were separated by electrophoresis, purified, sequenced, and identified via their *SCGB1A1* gene-specific signature sequences.

2.3. Phylogenetic analyses

Alignments were performed with Geneious Pro software using the following parameters: Cost matrix (65% similarity); Gap open penalty (12); Gap extension penalty (3); Alignment type (Global alignment with free end gaps). Evolutionary relationships were inferred using minimum evolution (Rzhetsky and Nei, 1992), neighbor-joining (Saitou and Nei, 1987), and maximum likelihood (Jukes and Cantor, 1969) methods. A bootstrap analysis with 500 replicates was performed to test the robustness of the phylogenetic tree for each method. Phylogenetic analyses were performed using MEGA version 5 (Tamura et al., 2011).

2.4. Three-dimensional modeling

The various equid SCGB 1A1 and SCGB 1A1A protein structures were built using the chain A of the oxidized P21 form of leporine uteroglobin at 1.64 Angstrom resolution (2UTG_A; NCBI) as template (Lambert et al., 2002). The resultant molecular coordinate files were imported into Swiss-PdbViewer software (SPDBV 4.1.0) for analysis.

3. Results and discussion

3.1. Origin of SCGB1A1 ancestral triplication

To evaluate the presence of multiple *SCGB1A1* genes in other Equidae family members, we performed PCR assays targeting the signature sequences identified in equine *SCGB1A1P* (C–G–A), *SCGB1A1* (A–A–A), and *SCGB1A1A* (A–G–G) genes. Primers were designed in regions displaying the highest sequence identity (100%) amongst the three equine genes to avoid gene-specific preference

Please cite this article in press as: Côté, O., et al. Phylogenetic relationships among Perissodactyla: *Secretoglobin 1A1* gene duplication and triplication in the Equidae family. Mol. Phylogenet. Evol. (2013), http://dx.doi.org/10.1016/j.ympev.2013.08.012