Genomics 102 (2013) 301–309

Contents lists available at ScienceDirect

Genomics

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

Gene conversions are under purifying selection in the carcinoembryonic antigen immunoglobulin gene families of primates

Mouldi Zid, Guy Drouin *

Département de Biologie et Centre de Recherche Avancée en Génomique Environnementale, Université d'Ottawa, Ottawa, Ontario K1N 6N5, Canada

ARTICLE INFO

Article history:
Received 14 December 2012
Accepted 8 July 2013
Available online 15 July 2013

Keywords:
Carcinoembryonic antigen (CEA)
Gene conversion
Immunoglobulin
Primates
Human
Selective constraints
Purifying selection

ABSTRACT

The carcinoembryonic antigen (CEA) family contains a large number of glycoproteins belonging to the immunoglobulin superfamily. Here, we investigate whether the gene conversions occurring between primate CEA-related genes are adaptive. Our results show that primate CEA-related genes are subject to frequent and repeated gene conversion events. Furthermore, gene conversions occur most frequently between nearby genes sharing similar sequences, are not more frequent in Ig-like V-type 1 domains than in the Ig-like C2-type 1 domains and dN/dS ratio tests shown that both these domains evolve either neutrally or under purifying selection. Our results therefore suggest that CEA-related genes evolve under purifying selection and the frequent gene conversion events we observed likely represent selectively neutral events between genes having similar sequences and functions.

1. Introduction

The human carcinoembryonic antigen (CEA) multigene family is a subgroup of the immunoglobulin superfamily which includes 34 related genes and pseudogenes located within a 1.5 Mb region of chromosome 19 [1]. This family is involved in cell–cell adhesion and affects the growth and differentiation of various normal and pathogenic human tissues [2]. At present, the CEA family can be divided into three subgroups based on sequence similarity, developmental expression patterns, and their biological functions. The CEACAM (CEA-related cell adhesion molecule) subgroup contains twelve genes (*CEACAM1*, *CEACAM3* through *CEACAM8*, *CEACAM16*, and *CEACAM18* through *CEACAM21*), the PSG (pregnancy specific glycoprotein) subgroup consists of eleven closely related genes (*PSG1* through *PSG11*), and the third subgroup encloses eleven pseudogenes (*CEACAMP1* through *CEACAMP11*; [2]; Figs. 1–3). The extracellular regions of the CEACAM subgroup are heavily glycosylated and share 17–88% similarity in their protein sequences (Supplemental Table 1). The members of the PSG subgroup are more conserved; they share 79–91% similarity in their extracellular protein sequences (Supplemental Table 1). The common characteristic among CEACAM proteins is that they are made up of single 26–160 amino acids long Ig-like V-type 1 domains, followed by zero to six Ig-like C2-type domains of A and B subtypes in their extracellular region [3]. However, *CEACAM16* contains two Ig-like V-type domains at its NH_2-terminal (N1) and COOH-terminal (N2; [4]), whereas *CEACAM20* contains a truncated Ig-like V-type 1 domain ([5]; Fig. 1). Proteins encoded by members of the PSG subgroup have a structure similar to that encoded by the CEACAM members (Fig. 2). The extracellular domain of PSG proteins contains one Ig-like V-type domain (N), followed by three Ig-like C2-type domains of A1, A2 and B2 subtypes ([6]; Fig. 2). The majority of CEACAM members are anchored in the cell membrane, whereas *CEACAM16* and all of the PSGs appear to be secreted glycoproteins with no glycosyl phosphatidyl inositol (GPI) membrane anchor ([4,7]; Figs. 1 and 2).

Various CEA-family members exhibit diverse functions, including controlling tissue homeostasis, insulin metabolism, modulation of angiogenesis, tumor development, modulation of apoptosis and regulation of immune response during the period of pregnancy [2,8–11]. Interestingly, two members of this family, *CEACAM1* and *CEACAM3*, act as receptors for pathogenic bacteria and viruses. Both genes are expressed on the surface of immune cells such as granulocytes. Their role is to destroy the specific pathogens via their signaling motifs, yet diverse pathogenic bacteria, including *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Moraxella catarrhalis* and *Haemophilus influenzae*, as well as some strains of *Escherichia coli* utilize CEACAMs during the course of infection of humans [12].

The number of CEACAM and PSG genes varies between mammalian species. To date, humans and mice CEA gene family includes 34 and 32 members, which are localized on chromosome 19 and chromosome 7, respectively [1,13,14]. Thirteen genes were identified in the rat genome. There are 21 members of the CEA gene family in the chimpanzee.

Gene conversions are non-reciprocal homologous recombination events where part of a gene is replaced by the sequence from a related

* Corresponding author at: Département de Biologie, Université d'Ottawa, 30 Marie Curie, Ottawa, Ontario K1N 6N5, Canada. Fax: +1 613 562 5486.
E-mail address: gdrouin@science.uottawa.ca (G. Drouin).

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.07.003

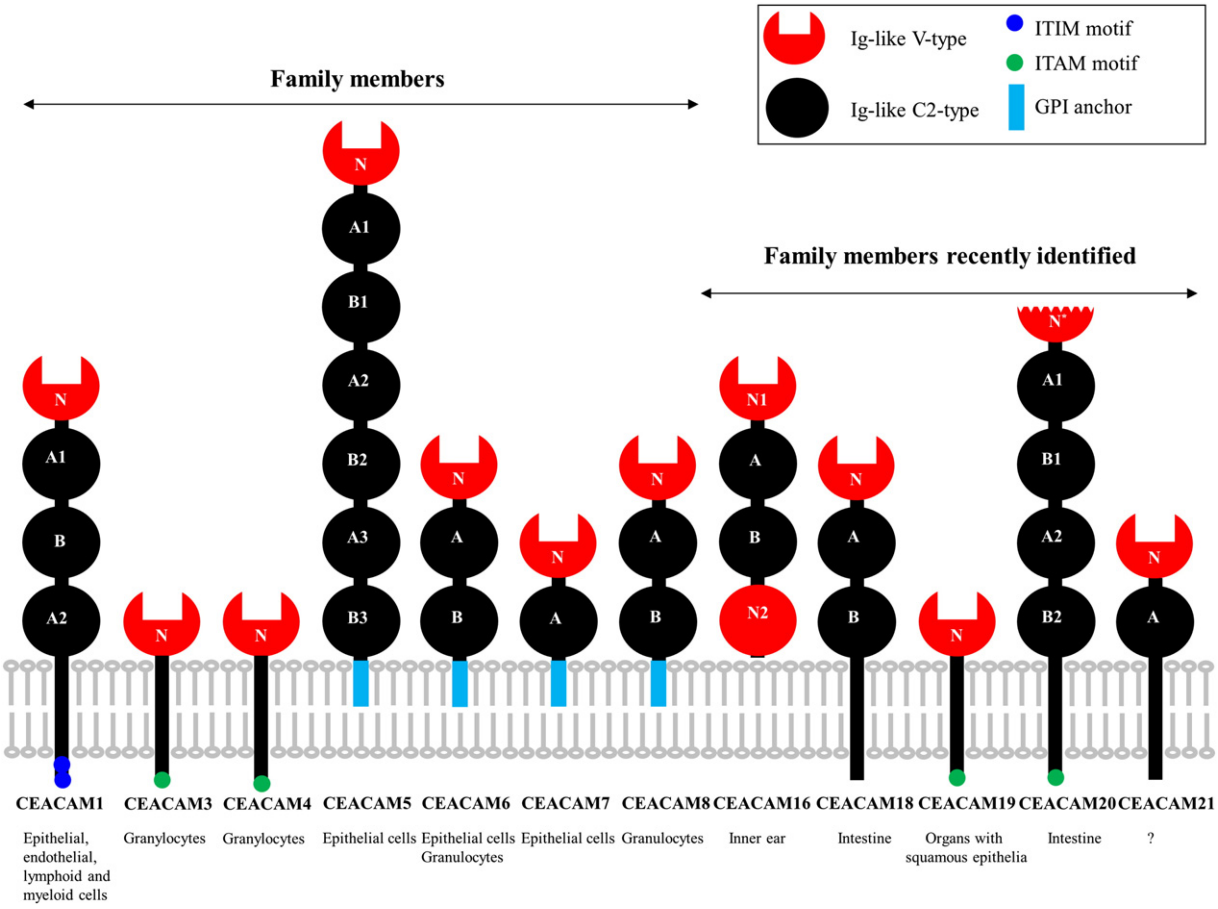

Fig. 1. Structure and tissue expression of human CEACAM proteins.
Adapted from http://www.carcinoembryonic-antigen.de/.

gene. Although the effect of most gene conversions is simply to increase the sequence similarity of related genes, they can also generate adaptive sequence diversity [15,16]. However, such exchanges can also be detrimental because they can lead to the elimination of an essential function [17,18]. The effect of gene conversions can therefore be neutral, adaptive or detrimental.

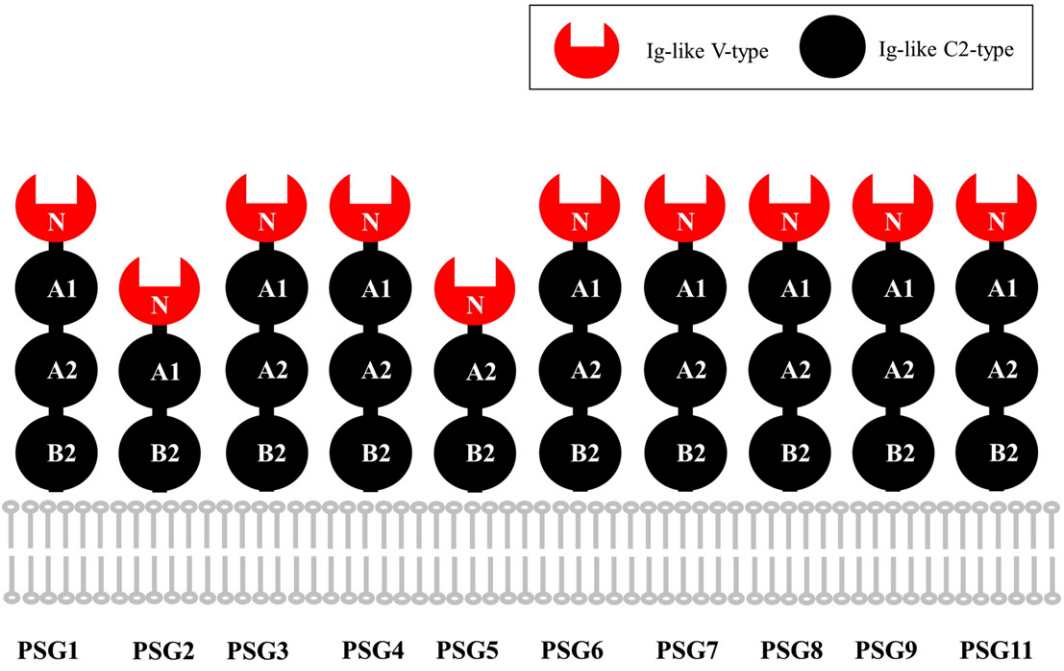

Fig. 2. Structure of human PSG proteins.
Adapted from http://www.carcinoembryonic-antigen.de/.