Infection, Genetics and Evolution 31 (2015) 277–283

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Adaptive amino acid composition in collagens of parasitic nematodes

Austin L. Hughes *

Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA

ARTICLE INFO

Article history:
Received 2 February 2015
Received in revised form 4 February 2015
Accepted 4 February 2015
Available online 11 February 2015

Keywords:
Amino acid composition
Collagen
Immunogenicity
Parasitic nematodes

ABSTRACT

Amino acid composition was analyzed in the glycine-rich repeat region of 306 collagens belonging to three major families of collagens from both parasitic and free-living nematodes. The collagens of parasitic species showed a tendency toward decreased usage of the hydrophilic residues A, D, and Q and increased usage of the hydrophobic resides I, L, and M; and this trend was seen in parasitic species of both the order Rhabdita and the order Spirurida. The amino acid composition of collagens of parasitic Rhabdita thus tended to resemble those of Spirurida more than that of free-living Rhabdita, suggesting an association between amino acid composition and a parasitic lifestyle. Computer predictions suggested that the more hydrophobic amino acid composition was associated with a reduction of the propensity towards B-cell epitope formation, suggesting that evasion of host immune responses may be a major selective factor responsible for the parasite-specific trend in collagen amino acid composition.

1. Introduction

Parasitic nematodes exact a substantial burden on the health of humans and their domestic animals, leading to a search for improved treatments, including vaccines (Hoffmann et al., 2014). Development of effective vaccines has been hampered by an incomplete understanding of host immune responses to parasitic nematodes and other metazoan animal parasites (Maizels et al., 1999). The immune response to nematodes is known to be complex, both because of the large number of proteins expressed by nematodes and because of these parasites' ability to manipulate host immune responses (Finkelman et al., 1997; Maizels et al., 2001; Knox et al., 2003; Maizels and Yazdanbakhsh, 2003; Meeusen and Piedrafita, 2003; Kreider et al., 2007; Patel et al., 2009).

Because of its exposure to host defenses, the surface of the nematode body, including both the surface coat and the underlying cuticle, has been an area of interest for research on host immune responses to nematodes (Selkirk and Blaxter, 1990; Fetterer and Rhoads, 1993; Maizels et al., 2006). The nematode surface coat is rich in glycoproteins and negatively charged; in nematodes parasitic on vertebrates, the surface coat plays a role in immune evasion because it can be shed and thereby functions to divert the host antibody defenses away from the worm itself (Blaxter et al., 1992; Page and Johnstone, 2007). The cuticle itself is composed largely of cross-linked collagens, along with lipids and additional proteins including cuticulins (Cox, 1992; Fetterer and Rhoads, 1993; Page and Johnstone, 2007). Although collagens are not generally exposed to the host immune system because of the presence of the surface coat, antibodies against cuticle collagens are produced, implying an immune response to the cuticle in the brief periods when it is exposed, such as molting (Selkirk et al., 1989).

If host immune responses to the cuticle collagens of parasitic nematodes have been a constant feature of the evolutionary history of these parasites, it might be predicted that the amino acid composition of these collagens has evolved unique features related to evasion of host immune recognition. Here I test this hypothesis using data on 306 nematode collagens belonging to three major families. These analyses take advantage of the availability of complete or nearly complete genomic sequences of several free-living and parasitic nematodes, enabling comparisons of related collagens between free-living and parasitic species. Moreover, extensive collagen sequence data are available from free-living and parasitic species in the order Rhabdita, as well as parasitic species in the order Spirurida, making it possible to discriminate parasite-specific trends in amino acid composition from those based on evolutionary relationship.

In the model free-living nematode *Caenorhabditis elegans*, over 150 collagen genes have been identified, all of which are characterized by G-X-X repeats (glycine followed by two other residues; Fig. 1; Johnstone, 2000). Typically, there are two repeat domains, interrupted by a cysteine-containing region (Fig. 1). Outside the G-X-X repeats, there is little overall evidence of sequence

* Address: Department of Biological Sciences, University of South Carolina, Coker Life Sciences Bldg., 700 Sumter St., Columbia, SC 29208, USA. Tel.: +1 803 777 9186; fax: +1 803 777 4002.
E-mail address: austin@biol.sc.edu

http://dx.doi.org/10.1016/j.meegid.2015.02.002
1567-1348/

Fig. 1. Alignment of G-X-X repeat regions from examples of collagen sequences of *Caenorhabditis elegans* and *Necator americanus* belonging to (A) SQT-1 family; (B) SQT-2 family; and (C) SQT-3 family. Glycine residues in G-X-X repeats are indicated in *boldface*. Residues C-terminal motifs conserved in all members of a given family are *underlined*.

homology among all collagens encoded by *C. elegans*. However, sequences flanking the G-X-X repeats can be used to identify major families of *C. elegans* collagens (Johnstone, 2000). The present analysis focuses on three such families, identified by conserved amino acid sequence motifs located C-terminal to the G-X-X repeat region (Fig. 1).

2. Methods

2.1. Sequences analyzed

Amino acid sequences homologous to *C. elegans* SQT-1, SQT-2, and SQT-3 were obtained by BLASTP homology search from the NCBI database. The analyses reported here were based on sequences from the following seven nematode species: (1) *C. elegans* (order Rhabdita, family Rhabditidae); (2) *Caenorhabditis remanei* (order Rhabdita, family Rhabditidae); (3) *Ancylostoma ceylanicum* (order Rhabdita, family Ancylostomatidae); (4) *Necator americanus* (order Rhabdita, family Ancylostomatidae); (5) *Haemonchus contortus* (order Rhabdita, family Haemonchidae); (6) *Brugia malayi* (order Spirurida, family Onchocercidae); and (7) *Loa loa* (order Spirurida, family Onchocercidae). The phylogenetic relationships among these seven species are shown in Fig. 2 (Blaxter et al., 1998; van Megen et al., 2009). Other than the two species of *Caenorhabditis*, all species included are parasitic; thus the analyses included three intestinal parasites from the same order as *Caenorhabditis* (Rhabdita), along with two filarial parasites belonging to the order Spirurida. The SQT-1, SQT-2, and SQT-3 families were used in these analyses because they included numerous sequences representing free-living Rhabdita, parasitic Rhabdita, and Spirurida.

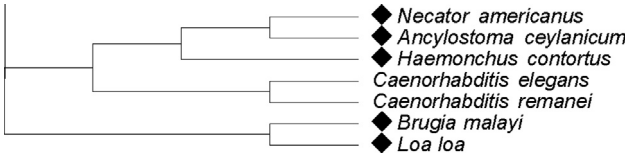

Fig. 2. Phylogenetic relationships of species used in analyses. Parasitic species are indicated by *solid diamonds*.

Amino acid sequences were aligned by the ClustalW algorithm in MEGA 6 (Tamura et al., 2011; Supplementary Figure S1). Sequence analyses were focused on the glycine-rich collagen repeat region, which includes the repeated G-X-X motif, occasionally interrupted by other short sequences (Fig. 1). Three collagen families were included in analyses, each defined by homology to *C. elegans* SQT-1, SQT-2, or SQT-3 (Fig. 1). All sequences included for analyses were characterized by a conserved cysteine at the N-terminus of the G-X-X repeat region and by 100% conservation of one of three family-specific sequence motifs located at the C-terminal end of the repeat region (Fig. 1). The SQT-1, SQT-2, and SQT-3 families appeared to correspond, respectively, to Group 1 (including Group 1a), Group 2, and Group 3 of *C. elegans* collagens as defined by Johnstone (2000). A total of 109 sequences from the SQT-1 family were included; 59 from the SQT-2 family; and 138 from the SQT-3 family (Supplementary Figure S1).

2.2. Statistical methods

In phylogenetic analyses of a set of sequences, any site at which the alignment postulated a gap in any sequence was excluded from