Immunology Letters 166 (2015) 79–86

Contents lists available at ScienceDirect

Immunology Letters

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

Immunoglobulin genomics in the prairie vole (*Microtus ochrogaster*)

Tong Qin[a,*], Huijing Zhao[b], Huabin Zhu[a], Dong Wang[a], Weihua Du[a], Haisheng Hao[a]

[a] Embryo Biotechnology and Reproduction Laboratory, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, PR China
[b] Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, PR China

ARTICLE INFO

Article history:
Received 13 April 2015
Received in revised form 15 May 2015
Accepted 3 June 2015
Available online 11 June 2015

Keywords:
Prairie vole
Immunoglobulin heavy chain
Immunoglobulin light chain

ABSTRACT

In science, the prairie voles are ideal models for studying the regulatory mechanisms of social behavior in humans. The utility of the prairie vole as a biology model can be further enhanced by characterization of the genes encoding components of the immune system. Here, we report the genomic organization of the prairie vole immunoglobulin heavy and light chain genes. The prairie vole IgH locus on chromosome 1 spans over 1600 kb, and consists of at least 79 V_H segments (28 potentially functional genes, 2 ORFs and 49 pseudogenes), 7 D_H segments, 4 J_H segments, four constant region genes (μ, γ, ε, and α), and two transmembrane regions of δ gene. The Igκ locus, found on three scaffolds (JH996430, JH996605 and JH996566), contains a totle of 124 $V_κ$ segments (47 potentially functional genes, 1 ORF and 76 pseudogenes), 5 $J_κ$ segments and a single $C_κ$ gene. Two different transcriptional orientations were determined for these $V_κ$ gene segments. In contrast, the Igλ locus on scaffold JH996473 and JH996489 includes 21 $V_λ$ gene segments (14 potentially functional genes, 1 ORF and 6 pseudogenes), all with the same transcriptional polarity as the downstream $J_λ$-$C_λ$ cluster. Phylogenetic analysis and sequence alignments suggested the prairie vole's large germline V_H, $V_κ$ and $V_λ$ gene segments appear to form limited gene families. Therefore, this species may generate antibody diversity via a gene conversion-like mechanism associated with its pseudogene reserves.

1. Introduction

The prairie vole (*Microtus ochrogaster*), is a small rodent found in central North America. This animal has been used in scientific experimentation since the 17th century. Unlike traditional rodent models, the prairie voles demonstrate features of social behavior similar to humans. Therefore, the prairie vole was a popular experimental animal for studying the role of social stress, and social isolation in particular, in the development of psychological and physiological disorders [1–4]. Social stressors have been consistently shown to negatively influence health, and a large body of literature has implicated changes in immune function as a link between social stressors and a wide range of disease states including cancers, cardiovascular diseases, diabetes, and infectious diseases [5–7]. The utility of the prairie vole as a biology model can be further enhanced by characterization of the genes encoding components of the immune system.

Immunoglobulins are the antigen-recognition molecules of B cells of jawed vertebrates, which usually consist of two identical heavy (H) and two identical light (L) chains. In some exceptional cases, such as shark IgNAR and selected subclasses of camelid IgGs, only heavy chains are used [8–10]. In the mammals studied so far, the locus of unique immunoglobulin heavy chain genes and loci of κ and λ light chain genes are commonly organized in a "translocon" pattern [11]. In the heavy chain locus, multiple variable (V_H), diversity (D_H), and joining (J_H) gene segments are followed by μ, δ, γ, ε and α gene segments [12]. In the κ light chain locus, multiple joining ($J_κ$) region gene segments are present within a cluster, followed by a single constant ($C_κ$) gene, whereas in the λ light chain locus, joining ($J_λ$) and constant ($C_λ$) genes occur as $J_λ$-$C_λ$ blocks, which usually have multiple copies [13].

The word "Prairie vole" is synonymous with scientific experimentation, but little is known about its Ig genes. We therefore used the recently available genome data of prairie vole provide as an opportunity to study the Ig genes of this species. Our study aimed to characterize the prairie vole IgH and IgL loci, in an effort to promote a better understanding of the immune system and evolutionary divergence of the Ig genes in placental mammals.

* Corresponding author at: Embryo Biotechnology and Reproduction Laboratory, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), No. 2 Yuanmingyuan West Road, Haidian District, Beijing 100193, PR China. Tel.: +86 10 62815892; fax: +86 10 62895971.
E-mail address: qintong@caas.cn (T. Qin).

http://dx.doi.org/10.1016/j.imlet.2015.06.001
0165-2478/

2. Materials and methods

2.1. Analysis of the prairie vole Ig genes

The genome sequence of the prairie vole (*M. ochrogaster*), provided by the Broad Institute, can be obtained from the Ensembl database (http://www.ensembl.org). Mouse immunoglobulin gene sequences were used as queries to search the prairie vole genome scaffolds that contained immunoglobulin genes. A conventional TBLASTN approach was used to identify constant region genes of the prairie vole immunoglobulins. FUZZNUC, an online software (http://embossgui.sourceforge.net/demo/fuzznuc.html), was used to find adjacent recombination signal sequences (RSSs) for the identification of variable, diversity, and joining gene segments. Five mismatched bases were allowed to cover all genes. The V gene domain (FRs or CDRs) was classified using IMGT numbering system [14]. Potentially functional-, ORF- and pseudo-V segments were all identified according to "Functionality" of IMGT [15].

2.2. Sequence analysis

EditSeq (DNAStar/Lasergene) was used for DNA and protein sequence editing. MegAlign (DNAStar/Lasergene) and Clustal X software program were used for sequence comparisons and identity calculations. The final alignment was made using BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html).

2.3. Phylogenetic analysis

Phylogenetic trees were made using MrBayes3.1.2 [16] and viewed in the TREEVIEW [17]. Multiple DNA sequence alignments for the tree construction were performed using *ClustalW*. Each V_H, V_κ and V_λ family is represented with one sequence per species chosen at random among the functional or ORF genes. The V_H, V_κ and V_λ sequences used in this study (except for prairie vole sequences) are presented in Tables S1–S3.

2.4. Definition of the V_H/V_L gene families

In mammals, germline V_H and V_L genes are categorized into different families according to their amino acid or nucleotide sequences similarity [18]. Sequences with greater than 75% similarity are general considered to belong to the same families, while those with less than 70% similarity are placed in different gene families, and those possessing between 70% and 75% similarity are inspected on a case-by-case basis [19]. We placed potentially functional V_H and V_L gene segments sharing more than 70% similarity into the same family.

3. Results

3.1. Prairie vole IgH locus

Analysis of the genomic sequence revealed that the prairie vole IgH locus is located on chromosome 1, and the entire IgH locus spans approximately 1600 kb. A single μ, one ε, one α gene, and three γ genes, as well as two transmembrane regions of δ gene could be identified. Upstream of the μ, there is approximately 59-kb DNA occupied by seven D_H and four J_H segments (Fig. 1A). A total of 79 V_H segments were identified (see Supplemental excel 1 for a summary of the prairie vole germline V_H segments). 28 of these appeared to be potentially functional because they contained leader exons (L), uninterrupted open reading frames (ORF), downstream RSS, and a V gene domain (framework regions and complementarity determining regions). The remaining 49 segments that contain either in-frame stop codons or are partial sequences were designated as pseudogenes. Given that gaps existed within the assembly, it is possible that as yet unidentified V_H genes are also present in the prairie vole genome. Except 20 pseudogenes that are too divergent or too short (truncated), all the V_H segments can be classified into four families (Table 1). While family 1 appears to be the largest V_H family consisting of 21 potentially functional V_H segments and 13 pseudogenes, family 4 contains only a single ORF member.

Seven D_H segments were identified in a region of 57 kb (some more might be missing due to sequence gaps). All the D_H segments were flanked at either side by RSS (nonamer and heptamer) with a 12-bp nucleotide spacer and existed within at least one alternative reading frame, suggesting that they are potentially functional. The length of these D_H segments ranges from 9 to 23 bp (Fig. 2A). J_H region contained four genes (designated J_H1–J_H4) spanning approximately 1.2 kb. Each J_H gene had an upstream RSS element with a 23 bp spacer, ORF, and a downstream RNA donor-splicing site at the 3′ end, suggesting that they are potentially functional (Fig. 2B).

3.2. Prairie vole Igκ locus

Using constant region sequence of the mouse Igκ (V01569) as a template, a BLAST search revealed one C_κ segment on scaffold JH996430 (Fig. 1B, top). Based on sequence alignment, the prairie vole C_κ segment shows a 74.5% sequence identity with the mouse C_κ segment (Fig. 3A). Upstream of the C_κ, five functional J_κs ($J_\kappa1$–$J_\kappa5$) are found (Fig. 3B). The prairie vole V_κ locus in this scaffold contains at least 33 potentially functional V_κ segments, one ORF and 49 pseudogenes. Whereas $V_\kappa1$, $V_\kappa16$, $V_\kappa18$, $V_\kappa25$, $V_\kappa30$–33 and 12 pseudogenes are oriented in the same transcriptional orientation as the J_κ-C_κ, the remaining V_κs, one ORF and 36 pseudogenes are arranged in an opposite orientation (Fig. 1B, *top*). Further searches for other V_κ-containing contigs uncovered scaffold JH996566 and JH996605 carried V_κ segments, which contained 7 potentially functional V_κ segments and 23 pseudogenes, 7 potentially functional V_κ segments and 4 pseudogenes, respectively (Fig. 1B, *middle and bottom*). In addition, V_κ segments located on scaffolds JH996566 and JH996605 also possess two different transcriptional directions. With the exception of 27 pseudogenes that are too divergent or truncated, all the V_κ segments can be classified into seven families (Table 2). Family 2 contains most members including 17 potentially functional V_κ segments and 20 pseudogenes. Family 1 includes 19 functional V_κ and 8 pseudogenes. The members of family 1 are most conserved in sequence identity, which share at least 81% identity at the nucleotide level. Families 3–5 contain four, three and two functional members, respectively. Families 6 and 7 both contain only one functional member (see Supplemental excel 2 for a summary of the prairie vole germline V_κ segments).

3.3. Prairie vole Igλ locus

A similar approach was used to identify the C_λ, and scaffolds JH996473 and JH996489 were determined to contain the prairie vole $C_\lambda1$ and $C_\lambda2$ segments, respectively (Fig. 1C). Each C_λ gene is preceded by one J_λ segment that is 5′ flanked by conserved RSS with a 12-bp nucleotide spacer (Fig. 4B). Two C_λ genes show approximately 42.9% amino acid identity, but the $C_\lambda1$ appears to be a pseudogene since it contains an in-frame stop codon (Fig. 4A). 14 potentially functional V_λ segments, one ORF and six pseudogene segments can be integrated into two families (Table 3). Family 1 contains all of members including 14 potentially functional V_λ segments, all of which