

Capra hircus and Ovis aries IGK loci simultaneous annotation in IMGT®

Viviane Nguefack Ngoune *[†] ¹, Morgane Bertignac * [‡] ¹, Géraldine Folch ¹, Joumana Michaloud ¹, Sofia Kossida[§] ¹, Marie-Paule Lefranc[¶] ¹

¹ IMGT®, the international ImMunoGeneTics information system®, Laboratoire d'ImmunoGénétique Moléculaire (LIGM) – Institut de Génétique Humaine (IGH) CNRS Université de Montpellier UMR 9002 – 141 rue de la Cardonille 34396 Montpellier Cedex 5, France

INTRODUCTION

IMGT®, the international ImMunoGeneTics information system®, <http://www.imgt.org> [1], is the global reference in immunogenetics and immunoinformatics [2], founded in 1989 by Marie-Paule Lefranc at Montpellier (Université de Montpellier and CNRS). IMGT® is a high-quality integrated knowledge resource specialized in the immunoglobulins (IG) or antibodies, T cell receptors (TR), major histocompatibility (MH) of human and other vertebrate species, and in the immunoglobulin superfamily (IgSF), MH superfamily (MhSF) and related proteins of the immune system (RPI) of vertebrates and invertebrates.

The genome of the vertebrates with jaws (*Gnathostomata*), which appeared in the evolution about 450 million years ago, includes the IG, TR and MH genes characteristic of the adaptive immune repertoires [2]. In humans and other mammals, there are seven main loci for IG and TR: three for IG (IGH, IGK and IGL) and four for TR (TRA, TRB, TRD and TRG). IMGT® genomic annotated data are classically displayed in IMGT Repertoire Web Resources (Locus description, Locus representation, Gene tables, Alignments of alleles).

The IG are B cell antigen receptors, expressed at the membrane of the B cells or secreted by plasma cells, and characterized by the huge diversity of their binding specificities. Classically IG proteins comprise two identical heavy chains (H) associated with two identical light chains (L) which belong, in higher vertebrates, to two chain types, kappa (IGK) or lambda (IGL). Several variable (V), diversity (D) (only present in the IGH locus), joining (J) and constant (C) genes compose the IGH, IGK, and IGL loci. The IG biosynthesis requires the recombination of these genes [2] and it is the result of these complex mechanisms of V-(D)-J rearrangements and junctional N-diversity and, for the IG, somatic mutations which creates the IG high diversity.

The IG loci have so far been explored in only a limited number of species. Indeed these loci are difficult to annotate owing to their multigene organization of highly similar genes and their biocuration requires a reliable and high quality locus assembly. The genomes of different ruminant species start becoming available, and among them, those of the domestic goat (*Capra*

*Speaker

[†]Corresponding author: viviane.nguefack-ngoune@igh.cnrs.fr

[‡]Corresponding author: morgane.bertignac@igh.cnrs.fr

[§]Corresponding author: sofia.kossida@igh.cnrs.fr

[¶]Corresponding author: marie-paule.lefranc@igh.cnrs.fr

hircus) and of the sheep (*Ovis aries*).

METHODOLOGY

The San Clemente Island goat (*Capra hircus*) genome has been sequenced (GenBank BioProject accession: PRJNA290100) and the information related to the identification of the goat IGK locus is available on public databases [3]. The goat IGK locus is situated on chromosome 11 in forward (FWD) orientation (GenBank: NC_030818.1; genome assembly: ARS1). An IMGT flat file was created, IMGT000009, which comprises the extracted region 46500000-46946647 bp from the chromosome 11.

In order to localize the IGK locus in sheep (*Ovis aries*), the sequence of the most 5' V gene and of the most 3' C gene from the goat locus were used to BLAST the sheep genome (genome assembly: Oar_v4.0). The sheep IGK locus is situated on chromosome 3 in reverse (REV) orientation. An IMGT flat file was created, IMGT000010, which comprises the extracted region (complement) 59150070-59300000 bp from chromosome 3 in order to place the sequence in the 5' to 3' orientation of the locus.

The annotation tool IMGT/LIGMotif [4] was used for the identification of the variable (V), joining (J) and constant (C) genes. The biocuration and description of the identified genes were performed using IMGT® tools (IMGT/Automat, IMGT/NTItoVALD). All the labels (V-REGION, J-REGION, C-REGION...) of each type of genes (V, J and C genes) were characterized. Subsequently, a multiple alignment and a phylogenetic tree of the V-REGION sequences of the goat and sheep IGKV genes and of a representative of each IGKV subgroup in human were performed.

RESULTS

The goat (*Capra hircus*) IGK locus on chromosome 11 (FWD) spans 447 kilobases and consists of 21 IGKV genes (5 genes are functional, 3 genes are ORF and 13 genes are pseudogenes), 4 IGKJ genes (1 gene is functional and 3 genes are ORF) and 1 IGKC gene (functional).

The sheep (*Ovis aries*) IGK locus on chromosome 3 (REV) spans 150 kilobases and consists of 18 IGKV genes (5 genes are functional, 1 gene is ORF and 12 genes are pseudogenes), 4 IGKJ genes (1 gene is functional and 3 genes are ORF) and 1 IGKC gene (functional).

As expected, in more than 85% (18/21) of the cases, the V-REGION of goat and sheep are similar to each other. The same subgroups are found in both species: IGKV1, IGKV2, IGKV3, IGKV6 and a new subgroup was introduced for both species: IGKV8, which until now did not exist in the other species studied within IMGT®. Interestingly, beyond this similarity of the subgroups, the IGKV gene positions of both species are the same. A unique and common nomenclature of those genes was proposed.

Some degenerated IGKV genes which could not be assigned to a given subgroup were assigned to the same clan, IGKV(II), because these genes were closely related (on a phylogenetic tree) to *Homo sapiens* IGKV2, IGKV3, IGKV4 and IGKV6 subgroup genes, <http://www.imgt.org/IMGTindex>. They include IGKV(II)-2, IGKV(II)-13, IGKV(II)-16 and IGKV(II)-19 for both species and IGKV(II)-10, IGKV(II)-12 only for goat.

The CDR-IMGT lengths [2], which structurally define the IG and TR V-REGION of the germline

genes (<http://www.imgt.org/IMGTScientificChart/Nomenclature/IMGT-FRCDRdefinition.html>) and are visible in the IMGT Colliers de Perles structure [5], were used to characterize all the functional V-GENE of each IGKV subgroup in both species, designated in this study as Caphir and Oviari (using the 6-letter code IMGT taxon abbreviation). Three IGKV1 genes, Caphir IGKV1-1, IGKV1-6, and IGKV1-7 and Oviari IGKV1-1, IGKV1-6, and IGKV1-7, have the same [6.3.7] CDR-IMGT lengths. The IGKV2-8, IGKV2-9, IGKV2-14 and IGKV2-15 genes in both species have also the same [11.3.7] CDR-IMGT lengths, even if the functionality differs from goat to sheep for the IGKV2-9 and IGKV2-15 (pseudogenes in the sheep).

CONCLUSIONS AND PERSPECTIVES

The IMGT-ONTOLOGY axiom and concepts of CLASSIFICATION provide the rules for the IG gene classification [2]. Although *Homo sapiens* genes and subgroups are commonly used in IMGT® for the biocuration of mammalian species, it is the first time that within IMGT® the same type of locus for two relatively close species of ruminants was studied simultaneously. Owing to the multigene family structure of the IG locus, and the post-speciation locus evolution, no attempt was done so far for identification of orthologous V genes between species. However, the goat and sheep species are sufficiently close to each other in evolution that, for the first time, a unique nomenclature for orthologous V genes could be proposed for the two species. This study was possible owing to the V-REGION high similarity of IGKV genes which, at the same time, occupy similar gene order in the IGK locus. An extension of that approach is currently evaluated and underway for the IGL locus of goat and sheep within IMGT®. Standardization of the gene nomenclature of orthologous genes of IG and TR multigene families between closely related species will accelerate the exploration of the antigen receptor loci from novel genomes in veterinary ruminant species.

Beyond research, goat and sheep small ruminant species have a highly economic interest because they contribute significantly to the nutrition and cash income for many farmer in developing world such as Africa and South Asia. unfortunately both species are regularly affected by different strains of Morbillivirus, which are extremely contagious and lethal. The characterization of the adaptative immune responses in these species would be beneficial for vaccine and immune reagent development. [6]

REFERENCES

Lefranc M-P, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse T, Hadi-Saljoqi S, Sasorith S, Lefranc G, Kossida S. IMGT®, the international ImmunoGeneTics information system® 25 years on. Nucl. Acids Res. 2015 Jan;43(Database issue):D413-22. doi: 10.1093/nar/gku1056. Epub 2014 Nov 5 Free Article. PMID: 25378316

Lefranc M-P. Immunoglobulin (IG) and T cell receptor genes (TR): IMGT® and the birth and rise of immunoinformatics. Front Immunol. 2014 Feb 05;5:22. doi: 10.3389/fimmu.2014.00022. Open access. PMID: 24600447

Schwartz JC, Philp RL, Bickhart DM, Smith TPL, Hammond JA. The antibody loci of the domestic goat (*Capra hircus*). Immunogenetics 2017 Oct 23., doi:10.1007/s00251-017-1033-3. PMID: 29063126

Lane, J., Duroux, P. and Lefranc, M.-P. From IMGT-ONTOLOGY to IMGT/LIGMotif: the IMGT® standardized approach for immunoglobulin and T cell receptor gene identification and description in large genomic sequences. *BMC Bioinformatics* 2010, 11:223; doi:10.1186/1471-2105-11-223. PMID: 20433708

Kaas, Q., Ehrenmann, F. and Lefranc, M.-P. IG, TR, MHC, IgSf and MhcSF: what do we learn from the IMGT Colliers de Perles? *Brief. Funct. Genomic Proteomic*, 2007, 6, 253-264. Epub 2008 Jan 21. doi:10.1093/bfgp/elm032. PMID: 18208865

Naveen Kumar, Sunil Maherchandani, Sudhir Kumar Kashyap, Shoor Vir Singh, Shalini Sharma, Kundan Kumar Chaubey, Hinh Ly. Peste Des Petits Ruminants Virus Infection of Small Ruminants: A Comprehensive Review. *Viruses*. 2014 Jun; 6(6): 2287–2327. doi: 10.3390/v6062287. PMID: 24915458

Keywords: IMGT, immunoinformatics, immunogenetics, immunoglobulin, T cell receptor, *Capra hircus*, *Ovis aries*