Journées Ouvertes de Biologie Informatique et de Mathématiques (JOBIM)
Marseille, France, Juillet 3-6, 2018
Abstracts des Journées, Brun C and Ballester B (eds.), A583, P196, p. 600-603.

IMGT® genomic annotation of the dog (Canis lupus familiaris) seven immunoglobulin (IG) or antibody and T cell receptor (TR) loci

Imène Chentli [*†] [1], Perrine Pégorier [* ‡] [1], Saïda Saljoqi[§] [1], Géraldine Folch [1], Joumana Michaloud [1], Véronique Giudicelli [1], Patrice Duroux [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. Currently, there are 244 annotated vertebrate genomes including 112 from mammals at NCBI.

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). So far the number of species present in the IMGT Web Resources reaches 40, however only two species, *Homo sapiens* and *Mus musculus*, have been fully annotated for their seven antigen receptor *loci*. This biocuration has been perfomed manually and the standardized annotation has allowed data entry in IMGT® databases

*Speaker
†Corresponding author: imene.chentli@igh.cnrs.fr
‡Corresponding author: perrine.pegorier@igh.cnrs.fr
§Corresponding author: saida.saljoqi@igh.cnrs.fr
¶Corresponding author: sofia.kossida@igh.cnrs.fr
‖Corresponding author: marie-paule.lefranc@igh.cnrs.fr

and tools [2]. The dog (*Canis lupus familiaris*) represents the first species for which the seven *loci* are annotated simultaneously. The biocuration was performed on the *loci* extracted from genome assembly.

METHODOLOGY

The seven IG and TR *loci* of the dog genome were recently described [3]. *CanFam3.1* is the last assembly (March 2015) of a female boxer (GenBank assembly accession: GCA_000002285.2, GenBank BioProject accession: PRJNA13179). Each *locus* sequence was localized on the corresponding chromosome and extracted. As the *locus* orientation on a chromosome can be either forward (FWD) or reverse (REV), the REV *locus* sequences were placed in the 5' to 3' *locus* orientation. Each *locus* sequence was assigned to a unique IMGT® accession number (IGH: IMGT000001, IGK: IMGT000002, IGL: IMGT000003, TRA/TRD: IMGT000004, TRB: IMGT000005 and TRG: IMGT000006).

RESULTS

The dog **IGH** *locus*, on chromosome 8 (REV), spans 1425 kilobases (kb) and consists of 89 IGHV genes belonging to 4 IGHV subgroups (36 functional, 2 ORF and 51 pseudogenes), 6 IGHD genes (5 functional and 1 ORF), 6 IGHJ genes (5 functional and 1 ORF) and 5 IGHC genes (4 functional and 1 ORF). The dog **IGK** *locus*, on chromosome 17 (REV), spans 349 kilobases (kb) and consists of 22 IGKV genes belonging to 5 IGKV subgroups (13 functional, 1 ORF and 8 pseudogenes), 5 IGKJ genes (4 functional and 1 ORF) and 1 IGKC gene (functional). The dog **IGL** *locus*, on chromosome 26 (FWD), spans 2583 kilobases (kb) and is still in progress. In June 2018, 261 IGLV genes belonging to 7 IGLV subgroups (69 functional, 12 ORF and 180 pseudogenes), 9 IGLJ genes (functional) and 9 IGLC genes (functional) were identified.

The dog **TRA** *locus*, on chromosome 8 (FWD), spans 743 kilobases (kb) and consists of 56 TRAV genes (34 functional and 22 pseudogenes) belonging to 30 TRAV subgroups, 59 TRAJ genes (40 functional, 12 ORF and 7 pseudogenes) and 1 TRAC gene (functional). The dog **TRB** *locus*, on chromosome 16 (REV), spans 271 kilobases (kb) and consists of 36 TRBV genes (22 functional, 1 ORF and 13 pseudogenes) belonging to 25 TRBV subgroups, 2 TRBD genes (functional), 12 TRBJ genes (9 functional, 2 ORF and 1 pseudogene) and 2 TRBC genes (functional). Like in the human (*Homo sapiens*) TRB *locus*, the functional TRBV30 gene is in inverted orientation of transcription downstream of the TRBC2 gene and rearranges by a mechanism of inversion. The dog **TRD** *locus*, on chromosome 8 (FWD), spans 344 kilobases (kb) and is embedded in the TRA *locus* between the TRAV and the TRAJ genes. It consists of 5 TRDV genes (3 functional belonging to 3 subgroups; 1 ORF and 1 pseudogene interspersed among the TRAV genes), 2 TRDD genes (functional), 4 TRDJ genes (3 functional and 1 ORF), and 1 TRDC gene (functional). Like in the human TRD *locus*, the functional TRVD3 gene is in inverted orientation of transcription downstream of the TRDC gene and rearranges by a mechanism of inversion. The dog **TRG** *locus*, on chromosome 18 (FWD), spans 447 kilobases (kb) and consists of 16 TRGV genes (8 functional and 8 pseudogenes) belonging to 7 TRGV subgroups, 16 TRGJ genes (7 functional, 3 ORF and 6 pseudogenes) and 8 TRGC genes (6 functional, 1 ORF and 1 pseudogene).

CONCLUSION AND PERSPECTIVES

Information on the IMGT® gene and *locus* reference sequences is available in the classical IMGT Repertoire pages, to which were added two novel pages: Locus in genome assembly and Locus gene order (with links to Locus representation). The annotation of the seven dog *loci* gives access to the study and comparison of the expressed adaptive immune repertoires in veterinary,

normal and pathological situations, using IMGT® tools such as, for nucleotide sequence analysis, IMGT/V-QUEST [4], for high throughput next generation sequencing (NGS), IMGT/HighV-QUEST [5], and for domain amino acid sequence analysis, IMGT/DomainGapAlign [6]. The curated IG and TR dog genes and alleles have been entered in the IMGT/GENE-DB database [7] and the corresponding IMGT® reference directories [1] will be used for coherent gene and sequence annotations of IG and TR *loci* in other *Canidae*. Indeed, this simultaneous biocuration of the seven IG and TR *loci* in *Canis lupus familiaris* has extended, between *loci*, the IMGT® concepts of classification (nomenclature) and description (labels) for a standardized IMGT® genomic annotation at the chromosome and genome assembly levels.

Dogs are an excellent model for human disease. For example, the treatment of canine lymphoma has been predictive of the human response to that treatment [3]. Study of the antigen receptor and immune response in dogs offers a unique opportunity for potential applications in veterinary and human medicine.

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

Martin J, Ponstingl H, Lefranc M-P, Archer J, Sargan D, Bradley A. Comprehensive annotation and evolutionary insights into the canine (*Canis lupus familiaris*) antigen receptor loci. Immunogenetics. 2018 Apr; 70(4):223-236. doi: 10.1007/s00251-017-1028-0. Epub 2017 Sep 19. Free article. PMID: 28924718

Brochet X, Lefranc M-P, Giudicelli, V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucl. Acids Res, 36, W503-508 (2008); doi:10.1093/nar/gkn316. PMID: 18503082

Alamyar E, Giudicelli V, Shuo L, Duroux P, Lefranc M-P. IMGT/HighV-QUEST: the IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. Immunome Res. 2012, April 20;8:1:2. doi: 10.4172/1745-7580.1000056. PMID: 22647994

Ehrenmann F, Kaas Q, Lefranc M-P. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. Nucl. Acids Res., 38, D301-307 (2010). Epub 2009 Nov 9; doi:10.1093/nar/gkp946. PMID: 19900967

Giudicelli V, Chaume D, Lefranc M-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucl. Acids Res., 33, D256-D261 (2005). PMID: 15608191

Journées Ouvertes de Biologie Informatique et de Mathématiques (JOBIM)
Marseille, France, Juillet 3-6, 2018
Abstracts des Journées, Brun C and Ballester B (eds.), A583, P196, p. 600-603.

Keywords: IMGT, immunoinformatics, immunogenetics, immunoglobulin, T cell receptor, Canis lupus familiaris