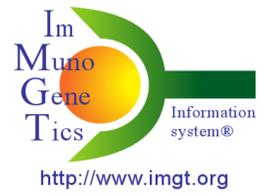


# Capra hircus and Ovis aries IGK loci: simultaneous annotation in IMGT®

Viviane Nguéfac Ngoune\*, Morgane Bertignac\*, Joumana Jabado-Michaloud, Géraldine Folch, Sofia Kossida, Marie-Paule Lefranc

\*Equal contribution

IMGT®, the international ImMunoGeneTics information system®, Laboratoire d'ImmunoGénétique Moléculaire (LIGM), Institut de Génétique Humaine (IGH), UMR 9002 CNRS-UM, Université de Montpellier (UM), Montpellier (France)



http://www.imgt.org



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 immunoglobulin (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.

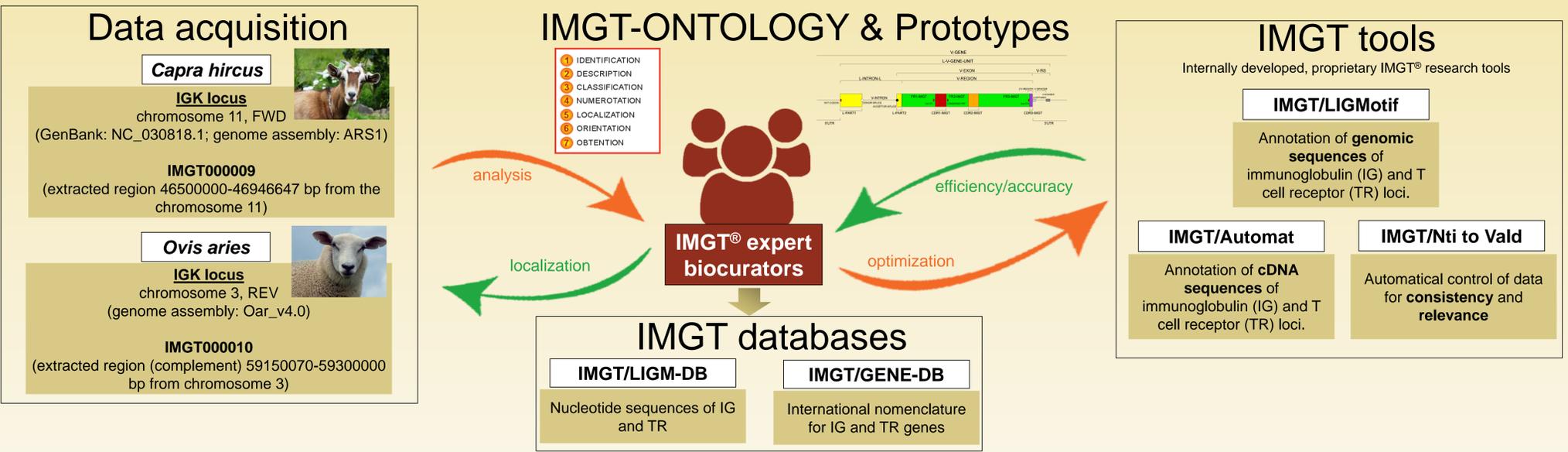
The genome of the vertebrates with jaws (*Gnathostomata*), which appeared in the evolution about 450 million years ago, includes the IG, TR and MH genes characteristic of the adaptive immune repertoires [2]. In humans and other mammals, there are seven main loci for IG and TR: three for IG (IGH, IGK 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.

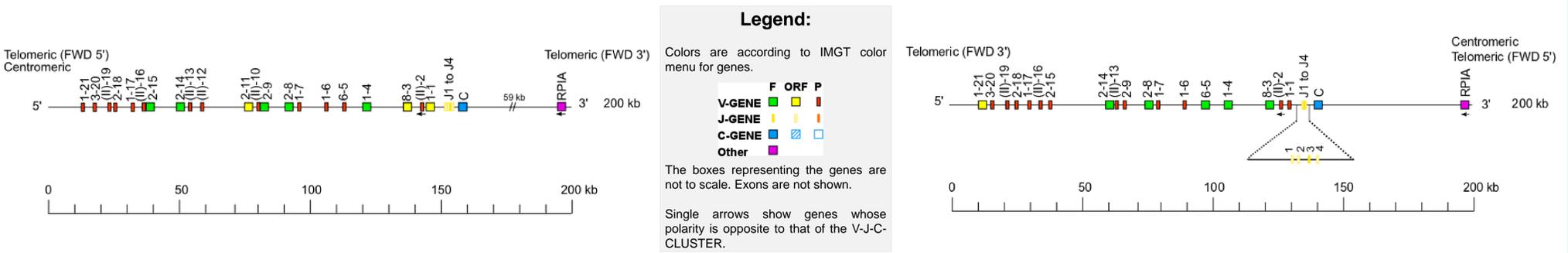
The IG loci have so far been explored in only a limited number of species. Indeed these loci are difficult to annotate owing to their multigene organization of highly similar genes and their biocuration requires a reliable and high quality locus assembly. The genomes of different ruminant species start becoming available, and among them, those of the domestic goat (*Capra hircus*) [3] and of the sheep (*Ovis aries*).

[1] Lefranc M.-P. et al., Nucl. Acids Res. 43:D413-422 (2015) PMID: 25378316 [2] Lefranc M.-P. Front. Immunol. 5:22. (2014) PMID: 24600447 [3] Schwartz J.C. et al., Immunogenetics 70(5):317-326 (2018) PMID : 29063126

## IMGT® Capra hircus and Ovis aries IGK loci annotation pipeline



## IGK locus representation of Capra hircus and Ovis aries: a unique and common nomenclature



The *Capra hircus* (goat) IGK locus on chromosome 11 in forward (FWD) orientation spans 160 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).

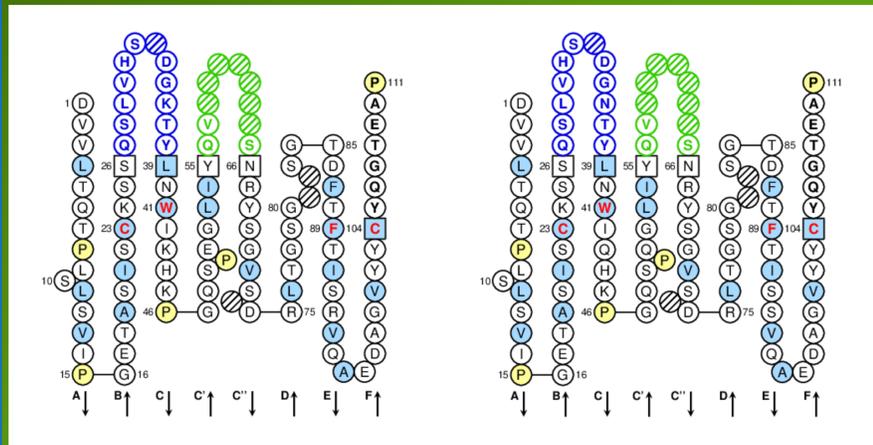
The *Ovis aries* (sheep) IGK locus on chromosome 3 in reverse (REV) orientation spans 150 kilobases and consists of 18 IGKV genes (5 genes are functional, 1 gene is ORF and 12 genes are pseudogenes), 4 IGKJ genes (1 gene is functional and 3 genes are ORF) and 1 IGKC gene (functional).

## Protein display of IGKV genes: Capra hircus & Ovis aries

Species	Gene	Allele	AccNum	Domain label	Functionality	FR1-IMGT (1-26)						CDR1-IMGT (27-38)				FR2-IMGT (39-55)				CDR2-IMGT (56-65)				FR3-IMGT (66-104)				CDR3-IMGT (105-117)					
						A (1-15)	B (16-26)	C (27-38)	C' (39-46)	C'' (47-55)	C''' (56-65)	C'''' (66-74)	D (75-84)	E (85-96)	F (97-104)	1 (1-15)	2 (16-26)	3 (27-38)	4 (39-46)	5 (47-55)	6 (56-65)	7 (66-74)	8 (75-84)	9 (85-96)	10 (97-104)	11 (105-117)	12 (105-117)	13 (105-117)	14 (105-117)				
Capra	IGKV1-1	IGKV1-1*01	IMGT000009	V-KAPPA	ORF	1	10	15	16	23	26	27	38	39	41	46	47	55	56	65	66	74	75	84	85	89	96	97	104	105	111	112	
Ovis	IGKV1-1	IGKV1-1*01	IMGT000010	V-KAPPA	P	DIQVTQSPSSLSACL	GRVSVITQAS	QNI	.....DTK	LAIWQKPK	RKAPKLLTY	AV	.....S	RSPSHFP	.S	QFSGSG	.FG	IDFTLTISLSKA	DDIATYYC	QQQHGLP													
Capra	IGKV1-6	IGKV1-6*01	IMGT000009	V-KAPPA	P	DIQVT*SPSVLSASL	GRVSVITQAS	QSV	.....SKH	LAIWQKPK	QKAPKLLTY	VA	.....S	NIQTGVS	.S	RFSGSS	.SG	TDFTLWISSLEA	EDVATYYC	LQVDSTP													
Ovis	IGKV1-6	IGKV1-6*01	IMGT000010	V-KAPPA	P	DIQVT*SPSVLSASL	GRVSVITQAS	QSV	.....SKH	LAIWQKPK	QKAPKLLTY	VA	.....S	NIQTGVS	.S	RFSGSS	.SG	TDFTLWISSLEA	EDVATYYC	LQVDSTP													
Capra	IGKV1-7	IGKV1-7*01	IMGT000009	V-KAPPA	P	GLQMTQSPS	LSASL	GRVVAICRAS	*SI	.....SKH	LDRYQKKS	G*VPKLLVY	AA	.....S	SLGTGAP	.S	RFSGSS	.SG	ADFTLTISLSLEA	EDAATYYM	LQTDSTP												
Ovis	IGKV1-7	IGKV1-7*01	IMGT000010	V-KAPPA	P	DLQMTQSPS	LSASL	GRVVAICRAS	QSI	.....SKH	LDRYQKKS	G*VPKLLVY	AA	.....S	SLGTGAP	.S	RFSGSS	.SG	ADFTLTISLSLEA	EDAATYYM	LQTDSTP												
Capra	IGKV2-8	IGKV2-8*01	IMGT000009	V-KAPPA	F	DVVL	TQTLRSLVTP	GETASISKSS	QSLVHS	DGNTY	LHWYQKPK	QSPQGLTY	QV	.....S	NRYSQVS	.D	RLTGSQ	.SG	TDFTLWISSRVA	EDAGVYYC	YQTEAP												
Ovis	IGKV2-8	IGKV2-8*01	IMGT000010	V-KAPPA	F	DVVL	TQTLRSLVTP	GETASISKSS	QSLVHS	DGNTY	LHWYQKPK	QSPQGLTY	QV	.....S	NRYSQVS	.D	RLTGSQ	.SG	TDFTLWISSRVA	EDAGVYYC	YQTEAP												
Capra	IGKV2-9	IGKV2-9*01	IMGT000009	V-KAPPA	F	AMVQ	TQTLRSLVTP	GETASISKSS	QSLVHS	DGNTY	LHWYQKPK	QSPQGLTY	QV	.....S	NRYSQVS	.D	RFTSSG	.SG	ADFTLTISLSRVA	EDARVYYC	QSLQAP												
Ovis	IGKV2-9	IGKV2-9*01	IMGT000010	V-KAPPA	F	ALVT	TQTLRSLVTP	GETASISKSS	QSLVHS	DGNTY	LHWYQKPK	QSPQGLTY	QV	.....S	NRYSQVS	.D	RFTSSG	.SG	ADFTLTISLSRVA	EDARVYYC	QSLQAP												
Capra	IGKV2-14	IGKV2-14*01	IMGT000009	V-KAPPA	F	DVVL	TQTLRSLVTP	GETASISKSS	QSLVHS	DGNTY	LHWYQKPK	QSPQGLTY	QV	.....S	NRYSQVS	.D	RFTSSG	.SG	ADFTLTISLSRVA	EDAGVYYC	QSLQAP												
Ovis	IGKV2-14	IGKV2-14*01	IMGT000010	V-KAPPA	F	DVVL	TQTLRSLVTP	GETASISKSS	QSLVHS	DGNTY	LHWYQKPK	QSPQGLTY	QV	.....S	NRYSQVS	.D	RFTSSG	.SG	ADFTLTISLSRVA	EDAGVYYC	QSLQAP												
Capra	IGKV2-15	IGKV2-15*01	IMGT000009	V-KAPPA	F	AVL	TQTLRSLVTP	GETASISKSS	QSLVHS	DGNTY	LHWYQKPK	QSPQGLTY	QV	.....S	NRYSQVS	.D	RFTSSG	.SG	ADFTLTISLSRVA	EDAGVYYC	QSLQAP												
Ovis	IGKV2-15	IGKV2-15*01	IMGT000010	V-KAPPA	F	AVL	TQTLRSLVTP	GETASISKSS	QSLVHS	DGNTY	LHWYQKPK	QSPQGLTY	QV	.....S	NRYSQVS	.D	RFTSSG	.SG	ADFTLTISLSRVA	EDAGVYYC	QSLQAP												
Capra	IGKV3-3	IGKV3-3*01	IMGT000009	V-KAPPA	ORF	EAVL	YQTPAVIAPL	GESISITCRAL	QSI	.....SDY	LSHWKQPK	QKAPKLLTY	DA	.....S	NRYSQVS	.D	RFTATQ	.SG	TDFTLWISSRVA	DDAARYYC	QSYVFPF												
Ovis	IGKV3-3	IGKV3-3*01	IMGT000010	V-KAPPA	F	EAVL	YQTPAVIAPL	GESISITCRAL	QSI	.....SDY	LSHWKQPK	QKAPKLLTY	DA	.....S	NRYSQVS	.D	RFTATQ	.SG	TDFTLWISSRVA	DDAARYYC	QSYVFPF												

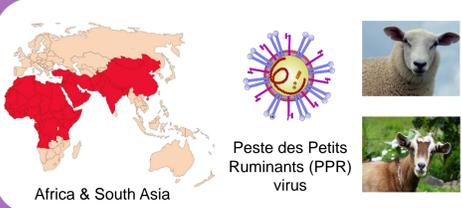
Protein display of different parts of the V-REGION shows for 3 selected subgroup, the strong similarity of amino acids of both species, designated in this study as Caphir and Oviari (using the 6-letter code IMGT taxon abbreviation). Three IGKV1 genes, Caphir IGKV1-1, IGKV1-6, and IGKV1-7 and Oviari IGKV1-1, 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.

## IMGT Collier de Perles: Caphir IGKV2-8 Oviari IGKV2-8



Hydrophobic amino acids and tryptophan (W): blue. Proline (P): yellow. The CDR-IMGT are limited by amino acids shown in squares, which belong to the neighbouring FR-IMGT. Arrows indicate the direction of the beta sheets and their different designations in 3D structures. CDR1-IMGT (blue), CDR2-IMGT (green), CDR3-IMGT (greenblue). C at position 23 (1st-CYS), W at position 41 (CONSERVED-TRP), letter of the conserved hydrophobic amino acid at position 89, and C at position 104 (2nd-CYS) are in red. Missing positions: hatched circles or squares

## Interest of this study



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 farmers 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 adaptive immune responses in these species would be beneficial for vaccine and immune reagent development.

## External links

