


RESEARCH ARTICLE

Open Access



# Genomic and expression analyses of *Tursiops truncatus* T cell receptor gamma (TRG) and alpha/delta (TRA/TRD) loci reveal a similar basic public $\gamma\delta$ repertoire in dolphin and human

Giovanna Linguiti<sup>1†</sup>, Rachele Antonacci<sup>1†</sup>, Gianluca Tasco<sup>2</sup>, Francesco Grande<sup>3</sup>, Rita Casadio<sup>2</sup>, Serafina Massari<sup>4</sup>, Vito Castelli<sup>1</sup>, Arianna Consiglio<sup>5</sup>, Marie-Paule Lefranc<sup>6</sup> and Salvatrice Ciccarese<sup>1\*</sup> 

## Abstract

**Background:** The bottlenose dolphin (*Tursiops truncatus*) is a mammal that belongs to the Cetartiodactyla and have lived in marine ecosystems for nearly 60 millions years. Despite its popularity, our knowledge about its adaptive immunity and evolution is very limited. Furthermore, nothing is known about the genomics and evolution of dolphin antigen receptor immunity.

**Results:** Here we report a evolutionary and expression study of *Tursiops truncatus* T cell receptor gamma (TRG) and alpha/delta (TRA/TRD) genes. We have identified *in silico* the TRG and TRA/TRD genes and analyzed the relevant mature transcripts in blood and in skin from four subjects.

The dolphin TRG locus is the smallest and simplest of all mammalian loci as yet studied. It shows a genomic organization comprising two variable (V1 and V2), three joining (J1, J2 and J3) and a single constant (C), genes. Despite the fragmented nature of the genome assemblies, we deduced the TRA/TRD locus organization, with the recent TRDV1 subgroup genes duplications, as it is expected in artiodactyls.

Expression analysis from blood of a subject allowed us to assign unambiguously eight TRAV genes to those annotated in the genomic sequence and to twelve new genes, belonging to five different subgroups. All transcripts were productive and no relevant biases towards TRAV-J rearrangements are observed.

Blood and skin from four unrelated subjects expression data provide evidence for an unusual ratio of productive/unproductive transcripts which arise from the TRG V-J gene rearrangement and for a “public” gamma delta TR repertoire. The productive cDNA sequences, shared both in the same and in different individuals, include biases of the TRGV1 and TRGJ2 genes.

The high frequency of TRGV1-J2/TRDV1- D1-J4 productive rearrangements in dolphins may represent an interesting oligo-clonal population comparable to that found in human with the TRGV9- JP/TRDV2-D-J T cells and in primates.

(Continued on next page)

\* Correspondence: salvatricemaria.ciccarese@uniba.it

†Equal contributors

<sup>1</sup>Department of Biology, University of Bari, via E. Orabona 4, 70125 Bari, Italy

Full list of author information is available at the end of the article



(Continued from previous page)

**Conclusions:** Although the features of the TRG and TRA/TRD loci organization reflect those of the so far examined artiodactyls, genomic results highlight in dolphin an unusually simple TRG locus. The cDNA analysis reveal productive TRA/TRD transcripts and unusual ratios of productive/unproductive TRG transcripts. Comparing multiple different individuals, evidence is found for a “public” gamma delta TCR repertoire thus suggesting that in dolphins as in human the gamma delta TCR repertoire is accompanied by selection for public gamma chain.

**Keywords:** T cell receptor, TRG locus, TRGV, TRGJ and TRGC genes, TRA/TRD locus, TRAV and TRDV genes, Dolphin genome, Expression analysis, IMGT

## Background

Bottlenose dolphin (*Tursiops truncatus*) and the other cetaceans represent the most successful mammalian colonization of the aquatic environment and have undergone a radical transformation from the original mammalian bodyplan. The discovery of two archaic whales with morphological homology between Cetacea and Artiodactyla brought conclusive anatomical support to clade Cetartiodactyla [1, 2]. Whales and hippos shared a common semi-aquatic ancestor that branched off from other artiodactyls around 60 million years ago [3–5]. One of the two branches would evolve into cetaceans, possibly beginning about 52 million years ago, with the protowhale *Pakicetus*, which underwent aquatic adaptation into the completely aquatic cetaceans [3]. So far nothing is known about the genomic organization of dolphin immunoglobulins (IG) and T cell receptor (TR) loci. The only studies of antigen receptors immunity revealed that IgG are present in whales [6, 7] and IGHG and IGHA genes have been described in the Atlantic bottlenose dolphin [8]. Within artiodactyls, the locus organization and expression of TRG and TRA/TRD genes have been characterized in ruminants; these species have been shown to possess a large TRG [9–11] and TRA/TRD [12–14] germline repertoire.

Here we present a evolutionary and expression analysis of *Tursiops truncatus* TRG and TRA/TRD genes. The surprising feature concerning TRG genes was, on the one hand, that the overall organization of the dolphin TRG locus resembles more the structure of a typical cassette of artiodactyls (IMGT<sup>®</sup>, the international ImMunoGeneTics information system<sup>®</sup>, <http://www.imgt.org> [15] > Locus representation: Sheep (*Ovis aries*) TRG1) than the structure typical of the human locus (IMGT<sup>®</sup> > Locus representation: Human (*Homo sapiens*) TRG). On the other hand, equally surprising was the finding of an unusual mechanism of biases in the V-J gene rearrangement usage, which is reminiscent of the most frequently used in the human peripheral  $\gamma\delta$  T cells repertoire of productively rearranged TRGV genes [16]. Despite the fragmented and incomplete nature of the assembly, we have obtained important information on the genomics and the evolution of the TRA/TRD dolphin potential repertoire and its relationship with the expressed chains. Furthermore, the structural 3D

visualization, computed by adopting a comparative procedure, using cDNA TRGV-J and TRDV-D-J rearranged amino acid sequences from a single individual, is consistent with the finding that the predicted  $\gamma\delta$  pairing, present both in the blood and in the skin, is shared among the organisms living in a controlled environment (kept under human care) as well as in those living in marine environment. This finding highlights in dolphin the existence of a basic “public”  $\gamma\delta$  repertoire of a given TR in a range of public T cell responses.

## Results

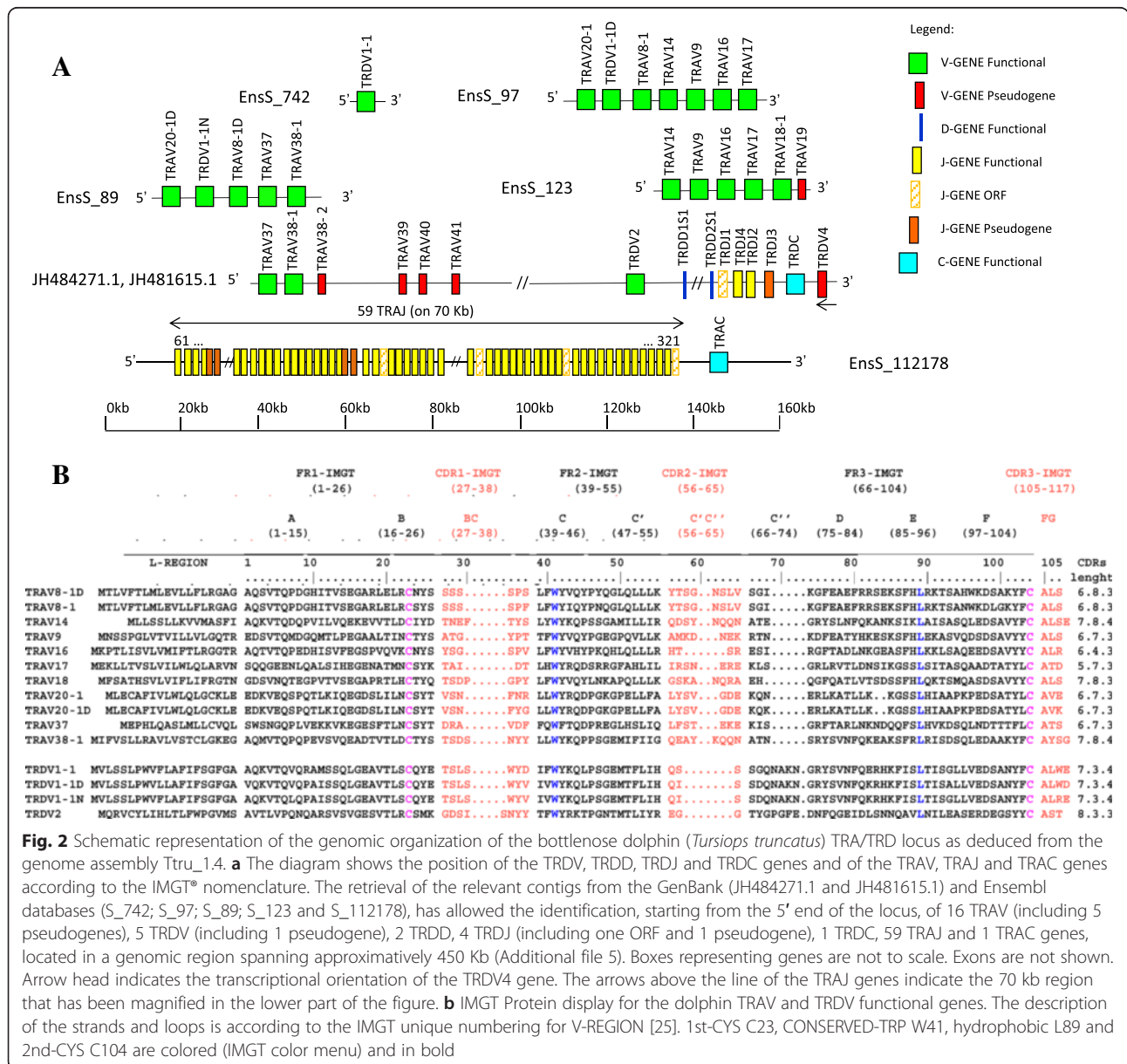
### Genomic arrangement and evolution of the dolphin TRG locus

The recent availability of a high quality draft sequence of the bottlenose dolphin (*Tursiops truncatus*) genome [17] (BioProject: PRJNA20367) allowed us to identify the dolphin TRG locus in two overlapping scaffolds (GEDI ID: JH473572.1; BCM-HGSC ID: contig 425448–578749) that provided a genomic sequence assembly of 188.414 kb (gaps included). In the dolphin, as in all mammalian species so far studied [18–20], the amphiphysin (AMPH) gene flanks the TRG locus at its 5' end and the related to steroido-genic acute regulatory protein D3-N-terminal like (STARD3NL) gene flanks the TRG locus at its 3' end. We annotated all the identified dolphin TRG genes using the human (GEDI ID: AF159056) and ovine (GEDI ID: DQ992075.1, DQ992074.1) TRG genomic sequences as a reference; the beginning and end of each coding exon were accurately identified by locating the splice sites and the flanking recombination signal (RS) sequences of the V and J genes (Fig. 1). According to our results, the dolphin locus is the simplest of the mammalian TRG loci identified to date (Additional file 1) [9, 10, 21–23]. It spans only 48 kb and its genes are arranged in a pattern comprising 2 TRGV, 3 TRGJ genes and a single TRGC (Additional file 2) gene. A closer inspection of the dolphin, human and sheep constant genes (Fig. 1c), reveals, that the dolphin TRGC (Additional file 2) gene possesses a single small exon 2 (EX2) which is more similar to the sheep TRGC5 EX2 than to the human TRGC1 EX2 (whereas in contrast the human TRGC2 gene has polymorphic duplicated or triplicated exons 2 [24]) (Fig. 1c). The dotplot matrix of



TRDV3 gene has yet been isolated in dolphin (Fig. 2a). This gene in dolphin, in contrast to all the so far analyzed species, is a pseudogene, since we have found both by *in silico* analysis and by PCR on genomic DNA, the presence of a stop codon (Additional file 6). Figure 2b shows the amino acid sequences of the dolphin TRA/TRD variable genes aligned according to the IMGT unique numbering for V domain [25]. Evolutionary analysis of dolphin, sheep, and human TRAV is shown in Additional file 7A. The tree shows that 10 dolphin subgroups form a monophyletic group with a corresponding human and sheep gene subgroup, consistent with the occurrence of distinct subgroups prior to the divergence of the three mammalian species. Three human TRAV

subgroup (TRAV19, TRAV20 and TRAV40) were found in dolphin, and not in sheep. In the TRA/TRD locus, the TRDJ, TRDD, and TRDC genes are followed finally by the TRAJ genes (Additional file 8 and Additional file 9), 61 in humans and 59 in dolphin, and by a single TRAC. Also in this case, the presence of several variable genes (TRDV1-1, TRDV1-1D and TRDV1-1 N) belonging to the TRDV1 subgroup, scattered in three different contigs, makes this portion of the dolphin locus more similar to the TRA/TRD locus of artiodactyls than to the human locus, as humans have only a single TRDV1 gene [16], while in cattle [13, 26], sheep [12] and other artiodactyls the TRDV1 subgroup is a large, multigene subgroup. In the phylogenetic tree, the membership of the



**Fig. 2** Schematic representation of the genomic organization of the bottlenose dolphin (*Tursiops truncatus*) TRA/TRD locus as deduced from the genome assembly Ttru\_1.4. **a** The diagram shows the position of the TRDV, TRDD, TRDJ and TRDC genes and of the TRAV, TRAJ and TRAC genes according to the IMGT® nomenclature. The retrieval of the relevant contigs from the GenBank (JH484271.1 and JH481615.1) and Ensembl databases (S\_742; S\_97; S\_89; S\_123 and S\_112178), has allowed the identification, starting from the 5' end of the locus, of 16 TRAV (including 5 pseudogenes), 5 TRDV (including 1 pseudogene), 2 TRDD, 4 TRDJ (including one ORF and 1 pseudogene), 1 TRDC, 59 TRAJ and 1 TRAC genes, located in a genomic region spanning approximately 450 Kb (Additional file 5). Boxes representing genes are not to scale. Exons are not shown. Arrow head indicates the transcriptional orientation of the TRDV4 gene. The arrows above the line of the TRAJ genes indicate the 70 kb region that has been magnified in the lower part of the figure. **b** IMGT Protein display for the dolphin TRAV and TRDV functional genes. The description of the strands and loops is according to the IMGT unique numbering for V-REGION [25]. 1st-CYS C23, CONSERVED-TRP W41, hydrophobic L89 and 2nd-CYS C104 are colored (IMGT color menu) and in bold



TRDV1 genes is supported by the monophyletic groupings, which are marked by 25 sheep, 6 dromedary and 3 dolphin members in contrast with the single human one (Additional file 7B).

#### 5' RACE PCR and RT-PCR on blood and skin RNA identified the dolphin TRG, TRA and TRD chains repertoire

Four types of 5' RACE and three types of RT-PCR (total of 6 and 6 experiments, respectively) on total RNA isolated from the peripheral blood of three unrelated adult animals (identified as M, K and L) and from the skin of animal identified by letter C (Tables 1 and 2) were carried out to investigate the dolphin TRG, TRA and TRD chains repertoire. We obtained a total of 105 unique (5RV and RTV) clonotypes of different length, each containing rearranged VJ-C (for TRG and TRA) and V-D-J-C (for TRDV) transcripts. A clonotype (AA) (AA for amino acid) is identified by a given rearranged V gene and allele, a given J gene and allele and a unique amino acid junction [27]. The V domains were checked for their typical features, i.e., the leader region and the five conserved amino acids (1st-CYS C23, CONSERVED-TRP W41, hydrophobic 89 (here, leucine L89), 2nd-CYS 104 and anchor 118 (J-PHE 118 or J-TRP) characteristic of a V-DOMAIN [25]. The functionality of each clonotype was determined based on the IMGT<sup>®</sup> criteria: transcripts were considered as productive if they had in-frame junctions and no stop codons, whereas transcripts were considered as unproductive if they had frameshifts

and/or stop codons. The junctions comprise the CDR3-IMGT and the two anchors C104 (2nd-CYS) and F118 (J-PHE for TRG, TRA and TRD) or W118 (for one TRA), identified in this study as TRAJ34) (Fig. 3 and Additional file 9). Fifty-nine TRG clonotypes were obtained from M, L, K and C, respectively. Twenty of the 59 TRG clonotypes contained out-of-frame cDNAs (Table 2). All the remaining 39 TRG clonotypes were productive, containing in-frame cDNA sequences and were submitted to and accepted by the ENA database in which they are identified by the HG328286 to HG328324 Accession numbers (Table 2 and Additional file 10). As all possible rearrangements between the two TRGV and the three TRGJ genes were found both in blood and in skin, it can be concluded that all dolphin TRG genes contribute to the formation of productive transcripts in all six TRGV-TRGJ combinations (Fig. 4).

To investigate the dolphin TRA chain repertoire, total RNA from the peripheral blood of a female dolphin (identified as L) was used as template in the single 5' RACE experiment (Tables 1 and 2). A total of 41 different TRA clonotypes were obtained and sequenced (Fig. 3). All sequences were productive (in-frame junction and no stop codon), and the leader region was of 17 to 20 amino acids depending on the V subgroup. The CDR1- and CDR2-IMGT lengths of the transcripts [6.4], [6.8], [7.8] corresponded to nine different TRAV subgroups and 29 different genes and were associated with diverse CDR3-IMGT of various length from 8 to 16 AA. In our cDNA collection, 8

**Table 1** List of primers used in 5' RACE, RT and genomic PCR

Locus	Primer	Genomic Orient. <sup>a</sup>	Sequence 5'-3'	Primer length	Location and sequence positions	Description
TRG	TC3L	REV	TGAGGAGGAGAAGGAGGT	18-mer	TRGC EX3 <sup>b</sup> 43364-43381	5'RACE, RT-PCR
	TC1L1	REV	GACGATACATACGAGTTCA	19-mer	TRGC EX1 <sup>b</sup> 37807-37825	dC-TAILED cDNA
	TC1L3	REV	AAGGCAAAGATGTGTTCCAG	20-mer	TRGC EX1 <sup>b</sup> 37635-37654	nested, RT-PCR
	TC1L2	REV	TGTTGCCATTCITTTCTTTCC	21-mer	TRGC EX1 <sup>b</sup> 37692-37712	nested, V1-V2 RT-PCR
	TV1LU	FWD	GCTCGCTCTGACAGTCCTT	19-mer	TRGV1 L-Part1 <sup>b</sup> 9989-10007	V1 RT-PCR, V1J2 genomic PCR
	TV7LU	FWD	GATCCTCTTCTCCTCCCTCTG	21-mer	TRGV2 L-Part1 <sup>b</sup> 21785-21805	V2 RT-PCR, V2J3 genomic PCR
	J2GL	REV	TGACGCTCTTGCCATGTGTT	20-mer	TRGJ2 <sup>b</sup> 31033-31052	V1J2 genomic PCR
TRA	J5BR	REV	CGGCGATGGGACAAAACCTG	20-mer	TRGJ3 <sup>b</sup> 34698-34717	V2J3 genomic PCR
	TA1C1L	REV	GAAGGTCTGGTTGAAGGTG	19-mer	TRAC EX1 <sup>c</sup> 86762-86780	5'RACE, RT-PCR
	TA1C2L	REV	TGCTCCGCATCCCAAATC	19-mer	TRAC EX1 <sup>c</sup> 86745-86763	dC-TAILED cDNA
TRD	TA1C3L	REV	TGCTGGATTGGGGCTTCT	19-mer	TRAC EX1 <sup>c</sup> 86574-86592	nested PCR
	TD1C1L	REV	AGAACTCCTTCACCAGAC	18-mer	TRDC EX1 <sup>d</sup> 85917-85934	5'RACE, RT-PCR
	TD1C2L	REV	CTTATAGTTACATCTTTGGG	20-mer	TRDC EX1 <sup>d</sup> 85938-85957	dC-tailed cDNA
	TD2CL	REV	CTGGAGTTTGAGTTTGATT	19-mer	TRDC EX2 <sup>d</sup> 86773-86791	nested PCR
	VD4U	FDW	GTGGAAGGTTTTGTGGGTCAGG	22-mer	TRDV4 EX <sup>d</sup> 91850-91871	V4 genomic PCR
VD4L	REV	TAACCAAGTGACCCAGATTT	20-mer	TRDV4 EX <sup>d</sup> 92062-92081	V4 genomic PCR	

<sup>a</sup>FWD: forward orientation, REV: reverse orientation (IMGT, Genomic orientation, <http://www.imgt.org/IMGTIndex/Orientation.php>)

<sup>b</sup>Acc. Number: JH473572.1

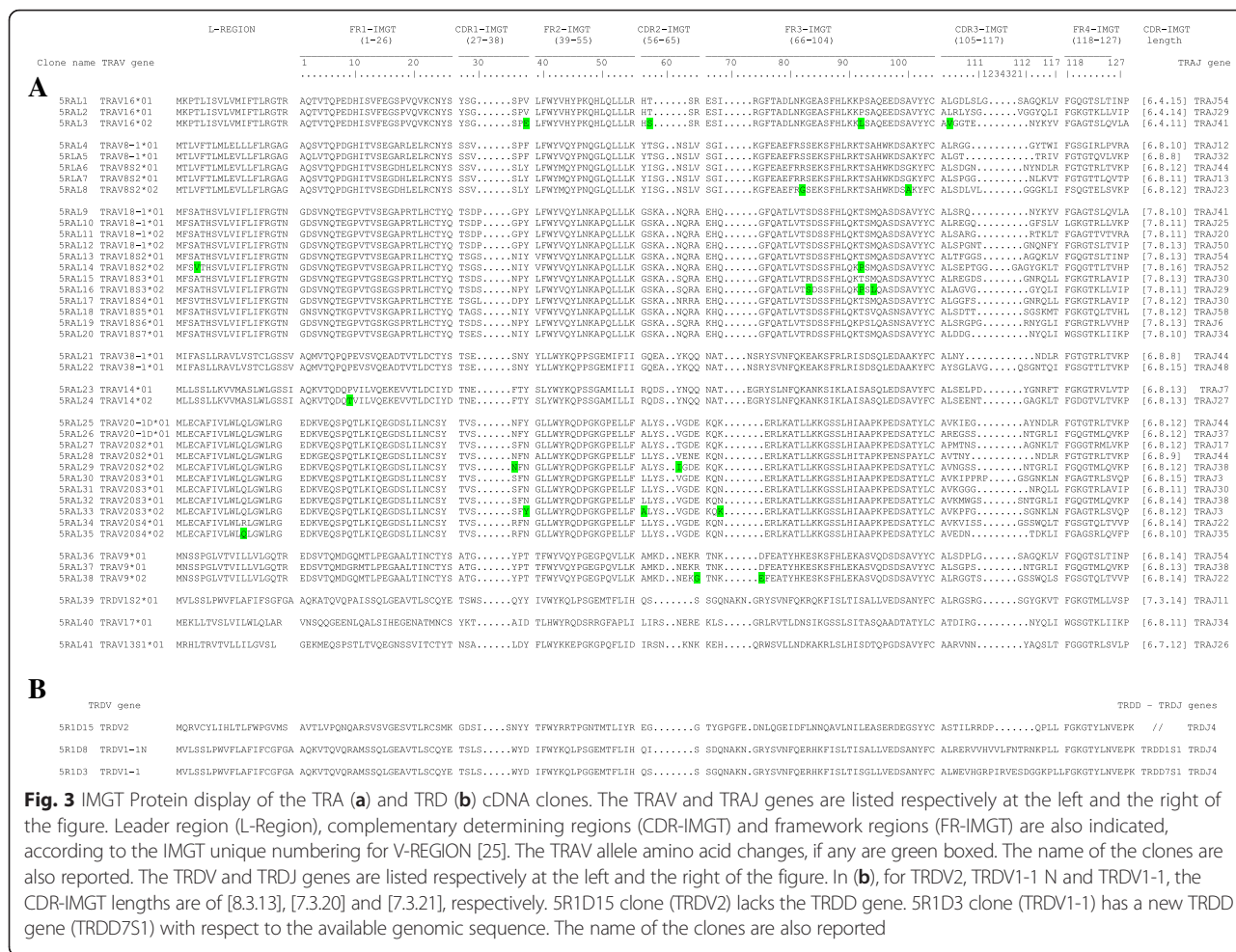
<sup>c</sup>Acc. Number: EnsS\_112178

<sup>d</sup>Acc. Number: JH481615.1

**Table 2** Summary of the different 5'RACE and RT-PCR experiments and the obtained rearrangement types

Locus	Animal tissue	Experiment	FWD primer name	REV primer name	Total number of non-redundant clonotypes	Number of non-redundant out-of-frame clonotypes	Number of non-redundant in-frame clonotypes	Non-redundant in-frame clonotypes by rearrangement type	GenBank (GEDI) accession numbers
TRG	Blood M	5'RACE	--	TC3L/TC1L1/TC1L2	7 <sup>a</sup>	(7) 4	3	1 TRGV1*01-TRGJ3*01	HG328298
								2 TRGV2*01-TRGJ1*01	HG328299/300
	Blood M	RT-PCR	TV1LU TV7LU	TC3L/TC1L2 TC3L/TC1L2	8	(8) 1	7	1 TRGV1*01-TRGJ2*01	HG328291
								4 TRGV2*01-TRGJ3*01	HG328292/93/94/95
	Blood L	5'RACE	--	TC3L/TC1L1/TC1L2	11 <sup>a</sup>	(11) 6	5	1 TRGV2*01-TRGJ1*01	HG328296
								1 TRGV2*01-TRGJ2*01	HG328297
Blood K	RT-PCR	TV1LU TV7LU	TC3L/TC1L2 TC3L/TC1L2	20	(20) 5	15	2 TRGV1*01-TRGJ2*01	HG328286/8	
							1 TRGV1*02-TRGJ2*01	HG328287	
							2 TRGV2*01-TRGJ3*01	HG328289/90	
							6 TRGV1*01-TRGJ2*01	HG328305/7/8/10/11/14	
Skin C	RT-PCR	TV1LU TV7LU	TC3L/TC1L3 TC3L/TC1L3	13	(13) 4	9	5 TRGV1*01-TRGJ3*01	HG328306/9/12/13/15	
							3 TRGV2*01-TRGJ3*01	HG328301/02/04	
							1 TRGV2*01-TRGJ1*01	HG328303	
							4 TRGV1*01-TRGJ2*01	HG328316/17/18/19	
							1 TRGV1*02-TRGJ2*01	HG328320	
TRD	Blood M	5'RACE		TD1C1L/TD1C2L/TD2CL	5 <sup>b</sup>	(5) 2	3	1 TRGV1*02-TRGJ1*01	HG328321
								3 TRGV2*01-TRGJ3*0	HG328322/3/4
								1 TRDV1*01-TRDJ4*01	LN610749
TRA	Blood L	5'RACE		TA1C1L TA1C2L TA1C3L	41	(41) 0	41	1 TRDV1*01-TRDJ4*01 <sup>c</sup>	LN610748
								1 TRDV2*01-TRDJ4*01	LN610747
								<sup>d</sup>	LN610706-LN610746

<sup>a</sup>One clonotype is an incomplete sequence; <sup>b</sup>One clonotype is a sterile germline transcript; <sup>c</sup>M TRDV1-1 N -TRDJ4 and TRDV1-1 -TRDJ4 rearrangements have one TRDD gene; <sup>d</sup>L TRA cDNA clonotypes by rearrangement type are displayed in Fig. 3



**Fig. 3** IMGT Protein display of the TRA (a) and TRD (b) cDNA clones. The TRAV and TRAJ genes are listed respectively at the left and the right of the figure. Leader region (L-Region), complementary determining regions (CDR-IMGT) and framework regions (FR-IMGT) are also indicated, according to the IMGT unique numbering for V-REGION [25]. The TRAV allele amino acid changes, if any are green boxed. The name of the clones are also reported. The TRDV and TRDJ genes are listed respectively at the left and the right of the figure. In (b), for TRDV2, TRDV1-1 N and TRDV1-1, the CDR-IMGT lengths are of [8.3.13], [7.3.20] and [7.3.21], respectively. 5R1D15 clone (TRDV2) lacks the TRDD gene. 5R1D3 clone (TRDV1-1) has a new TRDD gene (TRDD7S1) with respect to the available genomic sequence. The name of the clones are also reported

TRAV genes (TRAV16, TRAV8-1, TRAV18-1, TRAV38-1, TRAV14, TRAV20-1D, TRAV9, TRAV17) were assigned unambiguously to genes annotated in the genomic sequence (Fig. 2), while 12 could be assigned to new genes, belonging to five different subgroups. One gene belongs to a new subgroup, TRAV13 (TRAV13S1), not yet identified in the dolphin genomic sequence. This one gene belongs to subgroup TRDV1 (TRDV1S2), as shown by its CDR1- and CDR2-IMGT lengths [7.3.], and demonstrates that dolphin TRDV genes can, as in other species, participate to the synthesis of TRA chains by rearranging to a TRAJ gene (here, TRAJ11) [16]. Among the other new genes, six belong to subgroup TRAV45 (TRAV18S2, TRAV18S3, TRAV18S4, TRAV18S5, TRAV18S6, TRAV18S7), three belong to subgroup TRAV20 (TRAV20S2, TRAV20S3, TRAV20S4) and one to subgroup 42 (TRAV8S2). As these last subgroups have several members, an IMGT approved provisional nomenclature was assigned (with the letter S), allowing these genes to be entered in IMGT/GENE-DB and IMGT® tools (IMGT/V-QUEST and IMGT/HighV-QUEST) [15] while waiting for the identification and location of these genes in the reference genomic sequence. Three 5' RACE

experiments on total RNA isolated from the peripheral blood of two unrelated adult animals (identified as M and L) (Tables 1 and 2) were carried out to investigate the dolphin TRD chain repertoire; only one of these three PCR amplifications produced 3 in-frame, 1 out-of-frame and 1 sterile germline, clonotypes from the animal identified as M (Fig. 3b).

**Potential TRGV domain repertoire of productive and unproductive transcripts**

Analyzing the TRG in-frame transcripts it is noteworthy that 5 TRG clones were found identical in two or even three different individuals L and C (5RV1L2\*/C2/C5), M, C and K (RTV1M1/C1/K2/K3), K and M (RTV1K7/5RV1M1), and M and C (RTV2M4/C7 and RTV2M5/C8) (Additional file 10). This observation was rather intriguing as they represented together 14/39 in-frame sequences whereas in contrast each out-of-frame clonotype was found in a single individual. These shared clonotypes result from V1-J2 rearrangements in L and C (CDR-IMGT lengths [8.7.13]) and in M, C and K (CDR-IMGT lengths [8.7.14]), from V1-J3 rearrangements in K and M (CDR-IMGT

**A**

CLONE NAME	TRGV -J	- ntV	CDR3 SEQUENCE N REGION		- ntJ	Productive Rearrangement	CDR3 length nt/AA
			<b>C104</b>	<b>F118</b>			
5RV1L1	V1 - J2	1	TGTGCCCTCTGGGAAGT.	c . . . . . ACACATGGCAAGAGCGTCAAAGTGT <b>TT</b>	5	ALWEV // THGKSVKV	39/13
<u>5RV1L2*/C2/C5</u>		3	TGTGCCCTCTGGGA...	a . . . . . GGCTAACACATGGCAAGAGCGTCAAAGTGT <b>TT</b>	0	ALWE // RLTHGKSVKV	42/14
5RV1L3		8	TGTGCCCTCT.....	ttacaccgg . . . . . TAACACATGGCAAGAGCGTCAAAGTGT <b>TT</b>	3	AL/FTP// VTHGKSVKV	42/14
<u>RTV1M1/C1/K2</u>		11	TGTGCC.....	gagacac .GCTAACACATGGCAAGAGCGTCAAAGTGT <b>TT</b>	1	A //RTD// LTHGKSVKV	39/13
RTV1C3*/C4		8	TGTGCCCTCT.....	ttacacc .GCTAACACATGGCAAGAGCGTCAAAGTGT <b>TT</b>	1	AL/FTP// LTHGKSVKV	42/14
<u>RTV1K1/K4/K6</u>		3	TGTGCCCTCTGGGA...	ca . . . . . AACACATGGCAAGAGCGTCAAAGTGT <b>TT</b>	4	ALWE //Q// THGKSVKV	39/13
RTV1K3		11	TGTGCC.....	gagacac .GCTAACACATGGCAAGAGCGTCAAAGTGT <b>TT</b>	1	A //RDT// LTHGKSVKV	39/13
RTV1K5		4	TGTGCCCTCTGGGA.....	ggatgacatacaga.....GGCAAGAGCGTCAAAGTGT <b>TT</b>	11	ALWE//DDIR// GKSVKV	42/14
RTV1C6	V1 - J1	8	TGTGCCCTGT.....	cctcgg . . . . . GGTCAGGTTGGATCAAGATAT <b>TT</b>	3	AL // SS// GSGWIKI	33/11
<u>RTV1K7</u>	V1 - J3	0	TGTGCCCTCTGGGAGGTG	caactatcggggctg . . . . . TACATAAAAAAT <b>TTC</b>	7	ALWEV// OLSGS// YIKI	42/14
5RV1M1		0	TGTGCCCTCTGGGAAGTG	caactatcggggctg . . . . . TACATAAAAAAT <b>TTC</b>	7	ALWEV// OLSGS// YIKI	42/14
RTV1K8/K11/K12		4	TGTGCCCTCTGGGA.....	ggga . . . . . TACATAAAAAAT <b>TTC</b>	7	ALWE // G // YIKI	27/9
RTV1K9		11	TGTGCC.....	cgtc GAGTTATTACATAAAAAAT <b>TTC</b>	0	A // PS// SYYIKI	27/9
<u>RTV2M4/C7</u>	V2 - J3	2	TGTGCCTGGTGGGTT..	cttaaaggagga .AGTTATTACATAAAAAAT <b>TTC</b>	1	AWWV//LKGK// SYYIKI	42/14
5RV2L7		7	TGTGCCTGGT.....	acgaagtgg . . . . . AAAT <b>TTC</b>	14	AW // YEV // EI	21/7
<u>RTV2M5/C8</u>		1	TGTGCCTGGTGGGTTT.	cctctagttc GAGTTATTACATAAAAAAT <b>TTC</b>	0	AWWV//SSSS// SYYIKI	42/14
RTV2C12		8	TGTGCCTGG.....	gaatca . . . . . TATTACATAAAAAAT <b>TTC</b>	4	AW // ES // YYIKI	27/9
RTV2M6		3	TGTGCCTGGTGGGT...	aaggactgggcccacg . . . . . TACATAAAAAAT <b>TTC</b>	7	AWWV//RSGPT// YYIKI	39/13
RTV2M9		3	TGTGCCTGGTGGGT...	cggaag . . . . . TATTACATAAAAAAT <b>TTC</b>	4	AWWV//GK// YYIKI	33/11
RTV2K13		3	TGTGCCTGGTGGGT...	cat . . . . . TTACATAAAAAAT <b>TTC</b>	6	AWWV // I // YIKI	27/9
RTV2K15		4	TGTGCCTGGTGGG...	acc . . . . . GAGTTATTACATAAAAAAT <b>TTC</b>	0	AWW // DP // SYYIKI	33/11
RTV2K21		1	TGTGCCTGGTGGGTTT.	cgttg . . . . . ACATAAAAAAT <b>TTC</b>	8	AWWV// SVD // YKI	30/10
5RV2M4	V2 - J1	6	TGTGCCTGGTGGT.....	ctcat ATAGGTCAGGTTGGATCAAGATAT <b>TT</b>	0	AWC // S // YRSGWIKI	36/12
RTV2M10		8	TGTGCCTGG.....	cggaacttaatgcccgt ATAGGTCAGGTTGGATCAAGATAT <b>TT</b>	0	AW//RNLM//YRSGWIKI	45/15
5RV2M7		1	TGTGCCTGGTGGGTTT.	cggaagggc ATAGGTCAGGTTGGATCAAGATAT <b>TT</b>	0	AWWV//SEG//HRSGWIKI	45/15
RTV2K20		12	TGTGC.....	agcccctt ATAGGTCAGGTTGGATCAAGATAT <b>TT</b>	0	A // AL // YRSGWIKI	33/11
RTV2M11	V2 - J2	0	TGTGCCTGGTGGGTTTA	c . . . . . CTAAACACATGGCAAGAGCGTCAAAGTGT <b>TT</b>	2	AWWV//L//THGKSVKV	42/14

**B**

CLONE NAME	TRGV -J	- ntV	CDR3 SEQUENCE N REGION		- ntJ	Unproductive Rearrangement	CDR3 length nt
			<b>C104</b>				
RTV1M3	V1 - J2	0	TGTGCCCTCTGGGAAGTG	ctat . . . . . ACATGGCAAGAGCGTCAAAGTGT <b>TT</b>	7	ALWEV//LYMARASKC.	41
5RV2L8	V2 - J3	0	TGTGCCTGGTGGGTTTA	tgta . . . . . TTACATAAAAAAT <b>TTC</b>	6	AWWVY//VLT*KF	32
5RV2L4		17	.....	cctgg . . . . . TATTACATAAAAAAT <b>TTC</b>	4	//PGIT*KF	20
5RV2L5		11	TGTGCC.....	gactatca .AGTTATTACATAAAAAAT <b>TTC</b>	1	A//DYQVIT*KF	29
5RV2L9		16	.....A	aatctgttt . . . . .	22	//KSV(GDG)	10
RTV2C9		8	TGTGCCTGG.....	cggtcgggttaagacccttcgg . . . . . CATAAAAAT <b>TTC</b>	9	AW//RSQ*DPFGIKIF	39
RTV2C11		2	TGTGCCTGGTGGGTT..	aagg GAGTTATTACATAAAAAAT <b>TTC</b>	0	AWWV//KGVIT*KF	35
RTV2M7		0	TGTGCCTGGTGGGTTTA	gag . . . . .	19	AWWV*	17
5RV2M6		2	TGTGCCTGGTGGGTT..	gaatgtaa . .AGTTATTACATAAAAAAT <b>TTC</b>	1	AWWV//ECKVIT*KF	38
5RV2M2		0	TGTGCCTGGTGGGTTTA	ct GAGTTATTACATAAAAAAT <b>TTC</b>	0	AWWVY//VIT*KF	35
5RV2M3		7	TGTGCCTGG.....	acgac . . . . . ATTACATAAAAAAT <b>TTC</b>	5	AW//YDIT*KF	27
RTV2K14		2	TGTGCCTGGTGGGTT..	aggg . . . . . TATTACATAAAAAAT <b>TTC</b>	4	AWWV//RVLHKNF.	31
5RV2L10	V2 - J1	3	TGTGCCTGGTGGGT...	ccttatgcgggatt ATAGGTCAGGTTGGATCAAGATAT <b>TT</b>	0	AWWV//LMPGFIGVGSY.	51
5RV2L11		4	TGTGCCTGGTGGG...	gcagctcctat ATAGGTCAGGTTGGATCAAGATAT <b>TT</b>	0	AWW//GSSYIGVGSRY.	44
RTV2C13		0	TGTGCCTGGTGGGTTTA	tgctcggg .TAGGTCAGGTTGGATCAAGATAT <b>TT</b>	1	AWWVY//VG*VRLDQDI	43
RTV2C10		0	TGTGCCTGGTGGGTTTA	tagaatttctt ATAGGTCAGGTTGGATCAAGATAT <b>TT</b>	0	AWWVY//RIFL*VRLDQDI	49
5RV2M5		0	TGTGCCTGGTGGGTTTA	cggcacataaaaattg .TAGGTCAGGTTGGATCAAGATAT <b>TT</b>	1	AWWVY//GDIKIVGVGSY.	53
RTV2K16		1	TGTGCCTGGTGGGTTT.	ttgat ATAGGTCAGGTTGGATCAAGATAT <b>TT</b>	0	AWWV//FDIGVGSRY.	41
RTV2K17		4	TGTGCCTGGTGGG...	gcttt ATAGGTCAGGTTGGATCAAGATAT <b>TT</b>	0	AWW//GFIGVGSRY.	38
RTV2K18		2	TGTGCCTGGTGGGTT..	aagg ATAGGTCAGGTTGGATCAAGATAT <b>TT</b>	0	AWWV//KG*VRLDQDI	40

**Fig. 4** CDR3-IMGT nucleotide sequences retrieved from the cDNA clones with productive (a) and unproductive (b) rearrangements. Nucleotide sequences are shown from codon 104 (2nd-CYS) to codon 118 (J-PHE) (a and b). N-nucleotides added by the deoxynucleotidyltransferase terminal (DNNT, TdT) are indicated in lower cases. Numbers in the left and right columns indicate the number of nucleotides that are trimmed from the 3'V-REGION and 5' J-REGION, respectively; the germline region of the TRGV and TRGJ genes coincides with 0 in the nt V and nt J columns, respectively. In A, clonotypes with the same CDR3-IMGT nucleotide sequence deriving from two or more animals (letters M, L, K and C) are underlined. A shared clonotype (AA) between individuals has per definition a given V and J gene and allele and a given AA sequence for the junction. Three individuals (M, K and C) share the same CDR3 (AA) sequence with the V1-J2 rearrangement (RTV1M1/C1/K2/K3) however the junction in K3 differs from the junction in the other shared clonotypes by a nucleotide difference in the 5'J-REGION which may represent an allele of the TRGJ2 gene. Similarly, for the K and M shared clonotypes with a V1-J3 rearrangement (RTV1K7/5RV1M1), the junction in M1 differs from the junction in K7 by a nucleotide in the 3'V-REGION, which may represent an allele of the TRGV1 gene. This has been described as "convergent recombination" in which a given "public" TR amino acid sequence may be encoded by different nucleotide sequences both within the same and in different individuals [28]. In B, unproductive rearrangements (non-redundant out-of-frame clonotypes column of Table 2) for the presence of a stop codon (\*) and for frameshifts in the CDR3, are indicated

lengths [8.7.14]) and from V2-J3 rearrangements in M and C (CDR-IMGT lengths [8.6.14] (Fig. 4 and Additional file 10). This description of shared T cell clonotypes correspond to what is known in the literature as "public T cell response" in which T cells bearing identical TR may respond to the same antigenic epitope in different individuals [28].

Although the number of the germline TRG genes is low, which implies a reduced potential in the V-J recombination, a sufficient diversity and variability of the TR gamma transcripts seems to be guaranteed in the dolphin by the classical process of CDR3 diversity formation during somatic rearrangement [16]. Indeed, the creation of the CDR3



diversity results from the trimming of the 3'-V-REGION (up to 12 nucleotides (nt) for the in-frame junctions, up to 17 for the out-of-frame junctions), from the trimming of the 5'-J-REGION (up to 14 nt for the in-frame junctions, up to 22 for the out-of-frame junctions), and from the addition at random of the N nucleotides creating the N-REGION (up to 16 nt for the in-frame junctions, up to 23 for the out-of-frame junctions) (Fig. 4). This junction diversity is due to the activity of the terminal deoxynucleotidyl transferase (TdT) encoded by DNNT. The gene (NCBI ID: 101323636) has been identified in the bottle nosed dolphin genome and its amino acid sequence is 84 % identical to the human DNNT. The graphical representation of the number of in-frame versus out-of-frame sequences obtained for the 6 possible TRG rearrangements V1-J1, V1-J2, V1-J3, V2-J1, V2-J2 and V2-J3 display striking differences (Additional file 11). Both tests (Chi-squared *p*-value confirmed with Fisher's *p*-value) reject the null hypothesis for V1-J2 and V2-J1 (Additional file 12). This result confirms what was noticed at first sight and it follows that V1-J2 gene rearrangements were dominant among

the in-frame transcripts and were rare among the out-of-frame transcripts.

To investigate the high frequency of the out-of-frame rearranged V2-J3 cDNA, genomic PCR was carried out on DNA from blood of the animal identified by L. This choice was motivated both by the ratio of 1 in-frame (5RV2L7) on 4 out-of-frame (5RV2L8, L4, L5 and L9) clonotypes with the rearranged V2-J3 and by the highest number clones (5RV1L1, L2 and L3) with the rearranged V1-J2 (Fig. 4a). The frequency of the out-of-frame V2-J3 genomic rearrangements (Fig. 5a) is in agreement with that of all the respective rearranged cDNA clonotypes (Fig. 4); stop codons in the CDR3 seem to be generated in the unproductive V2-J3 rearrangements during the somatic recombination. Furthermore, genomic V1-J2 clonotypes, obtained by PCR performed on the same animal, demonstrate that they are all productive with two cases of sharing of the CDR3, i.e. V1J2L3/6/10/19 and V1J2L9/18 with 5RV1L2/C2/C5 and RTV1K1/K4/K6 cDNA clonotypes, respectively (Fig. 5c and Additional file 11).

**A**

CLONE NAME	Rearranged TRGV-J	-nt V	Nucleotide (nt) CDR3-IMGT sequence			-nt J	Amino acid (AA) CDR3-IMGT sequence	CDR3 length nt/AA
			C104	F118				
V2J3L7	V2-J3	5	TGTGCCTGGTGGG....	ccaacta	G.....	21	A//WWAN*	21
V2J3L21		2	TGTGCCTGGTGGGTTT..	tgggg	..GTTATTACATAAAAAATTTTC	2	A//WWWLGVIT*K	41
V2J3L35		4	TGTGCCTGGTGGGT....	cctgttag	.....TTACATAAAAAATTTTC	6	A//WWWLLVT*K	38
V2J3L36		4	TGTGCCTGGTGGGT....	gcta	.....TC	20	A//WWWLS	20
V2J3L39		0	TGTGCCTGGTGGGTTTAT	acgacagtcctgccc	.....ATTACATAAAAAATTTTC	5	A//WWWYTTIVLPLHKN	49
V2J3L6		1	TGTGCCTGGTGGGTTTA.	acctc	..GTTATTACATAAAAAATTTTC	2	A//WWW*PRYYIKI	42
V2J3L12		3	TGTGCCTGGTGGGTT...	atgggtct	.....ATTTTC	16	A//WWWML	29
V2J3L19		10	TGTGCCTG.....	tc	.....TACATAAAAAATTTTC	7	A//CLST*K	25
V2J3L28		11	TGTGCCT.....	ctagaggtgggatcaaa	GAGTTATTACATAAAAAATTTTC	0	A//SRGGIKELLHKN	46
V2J3L41		0	TGTGCCTGGTGGGTTTAT	gtegaa	GAGTTATTACATAAAAAATTTTC	0	A//WWWYVEELLHKN	46
V2J3L44		3	TGTGCCTGGTGGGTT...	gtttcct	..AGTTATTACATAAAAAATTTTC	1	A//WWWV*LLHKN	43
V2J3L45		6	TGTGCCTGGTGG.....	cccacaacctct	.....CATAAAAATTTTC	9	A//WWPTSHKN	37
V2J3L49		2	TGTGCCTGGTGGGTTT..		GAGTTATTACATAAAAAATTTTC	0	A//WWW*VIT*K	38
V2J3L48		1	TGTGCCTGGTGGGTTTA.	tggg	GAGTTATTACATAAAAAATTTTC	0	PGGFMGSYYIKI	43
V2J3L54		9	TGTGCCTGG.....	tgqccctc	GAGTTATTACATAAAAAATTTTC	0	A//WWPRVIT*K	38
V2J3L56		5	TGTGCCTGGTGGG.....	gggtctcgagacgc	..GTTATTACATAAAAAATTTTC	2	A//WWWVSRRVIT*K	47
V2J3L46		3	TGTGCCTGGTGGGTT...	ccga	..AGTTATTACATAAAAAATTTTC	1	A//WWWPKLLHKN	40
V2J3L47		5	TGTGCCTGGTGGG.....	ga	GAGTTATTACATAAAAAATTTTC	0	A//WWGELLHKN	37

**B**

V2J3L3	V2-J3	3	TGTGCCTGGTGGGTT...	aaagtt	..AGTTATTACATAAAAAATTTTC	1	A//WWWVSYYYIKI	42/12
V2J3L15		5	TGTGCCTGGTGGG.....	gggtca	.....ATTACATAAAAAATTTTC	5	A//WWGVNYIKI	36/10
V2J3L50		0	TGTGCCTGGTGGGTTTAT	gtgaaatc	GAGTTATTACATAAAAAATTTTC	0	A//WWWVYKSSYYIKI	48/14
V2J3L51		0	TGTGCCTGGTGGGTTTAT	gtcctagggaa	..TTATTACATAAAAAATTTTC	3	A//WWWVYVLYGNYIKI	48/14
V2J3L53		1	TGTGCCTGGTGGGTTTAT	gaaggg	...TATTACATAAAAAATTTTC	4	A//WWWYEGYYIKI	42/12

**C**

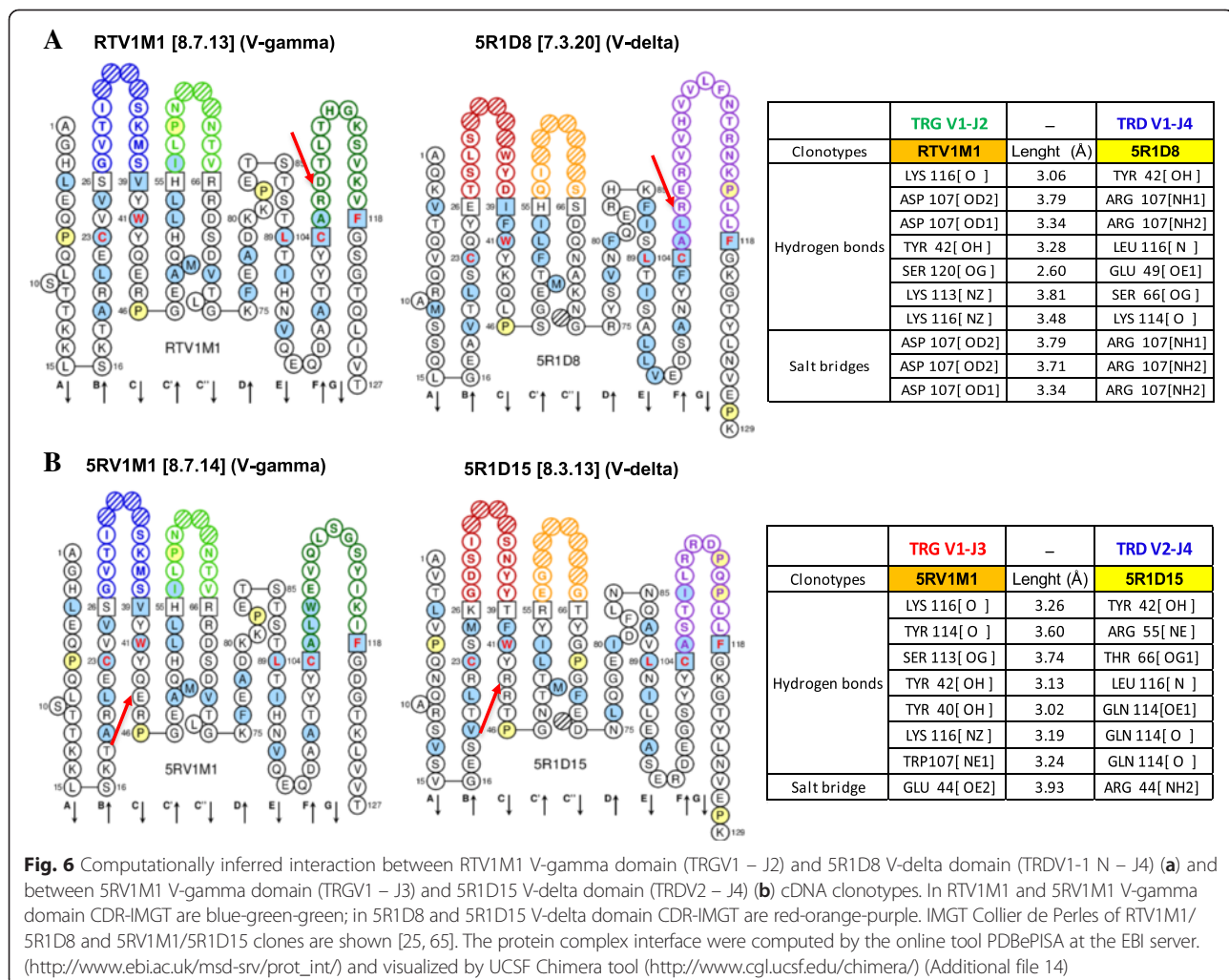
CLONE NAME	Rearranged TRGV-J	-nt V	Nucleotide (nt) CDR3-IMGT Sequence			-nt J	Amino acid (AA) CDR3-IMGT Sequence	CDR3 Length nt
			C104					
V1J2L5/13/22	V1-J2	2	TGTGCCCTCTGGGAAG..	cctctt	..TAACACATGGCAAGAGCGTCA	3	ALWE//ASVTHGKSV	43/13
V1J2L8/2/1/7		5	TGTGCCCTCTGGG....	tctctg	GGCTAACACATGGCAAGAGCGTCA	0	ALW//VLGLTHGKSV	43/13
V1J2L3/6/10/19		3	TGTGCCCTCTGGGAA...	a	GGCTAACACATGGCAAGAGCGTCA	0	<u>ALWE//RLTHGKSV</u>	40/12
V1J2L4/20		5	TGTGCCCTCTGGG....	cgatcc	GGCTAACACATGGCAAGAGCGTCA	0	ALW//AIRLTHGKSV	43/13
V1J2L9/18		3	TGTGCCCTCTGGGAA...	ca	...AACACATGGCAAGAGCGTCA	4	<u>ALWE//QTHGKSV</u>	37/11
V1J2L16		0	TGTGCCCTCTGGGAAGTG	accgccga	..CTAACACATGGCAAGAGCGTCA	2	ALWE//VTPRLTHGKSV	49/15
V1J2L17/28		2	TGTGCCCTCTGGGAAG..	ccgc	..GCTAACACATGGCAAGAGCGTCA	1	ALWE//AALTHGKSV	43/13
V1J2L31		4	TGTGCCCTCTGGG....	cggcc	GGCTAACACATGGCAAGAGCGTCA	0	ALW//GGRLLTHGKSV	43/13
V1J2L38		2	TGTGCCCTCTGGGAAG..	cctctg	..TAACACATGGCAAGAGCGTCA	3	ALWE//ASVTHGKSV	43/13

**Fig. 5** CDR3 nucleotide sequences retrieved from genomic rearranged clones. Nucleotide sequences are shown from codon 104 (2nd-CYS) to codon 118 (J-PHE) (a and b) and from codon 104 (2nd-CYS) to codon 115 (c). They are grouped on the basis of their rearrangement (a) and (b) TRGV2-TRGJ3 or (c) TRGV1-TRGJ2. N-nucleotides added by TdT are indicated in lower cases. Numbers in the left and right columns indicate the number of nucleotides that are trimmed from the 3'-V-REGION and 5'-J-REGION, respectively. The germline region of the TRGV and TRGJ genes coincides with 0 in the nt V and nt J columns, respectively. Clones with the same CDR3-IMGT nucleotide sequence deriving from two or more animals (L, K and C) are underlined (see also Fig. 4). In (a), unproductive rearrangements for the presence of a stop codon (\*) and for frameshifts in the CDR3, are indicated

### Computational analyses predict the pairing of the TRGV1-J2 and of the TRDV1-D1-J4 variable domains

To calculate the most likely computationally inferred interactions between the putative  $\gamma\delta$  pairing, we analysed the amino acid sequences of the three types of rearranged TRD cDNAs (originated by TRD V1-1 N-D1S1-J4 (5R1D8), V1-1-D7S1-J4 (5R1D3) and V2-J4 (5R1D15) rearrangements, respectively) found in the peripheral blood of the single animal identified in our study with the letter M (Table 2, Fig. 3b), and of three relevant TRG cDNAs (originated by TRG V1-J2, V1-J3 and V2-J3 rearrangements) among the six found in the peripheral blood of the same animal (clones identified by the letter M in Fig. 4a). The comparative inferred interactions of the TRGV1/TRDV1 and TRGV1/TRDV2 V domains, obtained using the AA sequences of the cDNA RTV1M1 (TRGV1-J2) and 5R1D8 (TRDV1-1 N-J4) clonotypes and the sequences of the 5RV1M1 (TRGV1-J3) and 5R1D15 (TRDV2-J4) clonotypes (Fig. 6, Additional file 13), respectively, were computed using as templates the human

$\gamma\delta$  T cell receptor chains (PDB and IMGT/3Dstructure-DB code: 1hxm) [29, 30]. We point out that, according to visualization of the RTV1M1/5R1D8 complex shown in Additional file 14A, the aspartic acid in the CDR3 position 107 (see IMGT Collier de Perles of the RTV1M1 clone) of the V-gamma domain, deriving from the addition of gagac nucleotides during the TRG V1-J2 recombination process, is predicted to be significantly involved in the formation of three possible salt bridge(s) and of two of the seven calculated hydrogen bonds with the arginine in position 107 of the 5R1D8 clonotype (V-delta domain) (red arrows in Fig. 6a). The arginine in position 107 derives from the TRDV1-1 N germline sequence in the CDR3 of the TRD V1-J4 rearrangement ('Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute ([http://www.ebi.ac.uk/pdbe/prot\\_int/pistart.html](http://www.ebi.ac.uk/pdbe/prot_int/pistart.html)) [31]. The computationally inferred interaction between the glutamic acid in position 44 in the FR2 of the TRG V1-J3 rearranged cDNA 5RV1M1 clonotype (V-gamma domain) and the arginine in position 44



in the FR2 of the TRD V2-J4 rearranged cDNA 5R1D15 clonotype (V-delta domain) is noteworthy because they are both involved in a possible salt bridge (red arrows in Fig. 6b). The Gln Q114 of the 5R1D15 clonotype is just as important because it is involved in three possible hydrogen bonds with Tyr Y40, Trp W107 and Lys K116 of the 5RV1M1 clone, respectively. In conclusion, we suggest that the RTV1M1/5R1D8 pairing is the most likely to form and it is the most stable (because of its ability to maintain the V-gamma/V-delta domain interactions better) (Fig. 6 and Additional file 13). This consideration, seems to be in compliance with the fact that the TRG V1-J2 rearrangement, found both in the peripheral blood and in the skin, is not only the most frequent among the six possible rearrangements, but it is shared among the organisms living in the same controlled environment (see animals identified by letters K, L and M) as well as in those living in marine environment (see animal identified by letter C) (Fig. 4).

## Discussion

In this study we report an extensive analysis of the genomic organization and expression of the TRG and TRA/TRD genes in dolphin. According to comparative analyses, dolphin TRG locus is the simplest and the smallest among the mammalian TRG loci identified to date [19, 20] and its organization is reminiscent of the structure of a typical single cassette of artiodactyls [9, 15] with a small number of genes, i.e. two TRGV, three TRGJ and one TRGC (Fig. 1).

The analysis of dolphin TRA/TRD locus confirmed that TRD genes are clustered within the TRA locus and that genes belonging to the TRDV1 subgroup are distributed among the TRAV genes as it is commonly expected in artiodactyls TRA/TRD locus [13, 14, 26, 32]. A total of 16 TRAV and 5 TRDV genes have been identified (Fig. 2). By the criterion that gene sequences having 75 % or greater nucleotide identity belong to the same subgroup, the TRAV and the TRDV genes belong to 13 and to three subgroups, respectively (Additional file 7). The sheep TRDV1 subgroup has been estimated to contain at least 40 genes [12], while only 25 TRDV1 genes have been identified in the genomic assembly [14]. The phylogenetic analysis assigns the membership of the dolphin TRDV1 genes due to the monophyletic groupings marked by 25 sheep, 6 dromedary and 3 dolphin members in contrast with the single human one (Additional file 7B).

Dolphin TR alpha chain expression analysis allowed us to identify new TRAV genes, with respect to the available genomic sequence. Furthermore a bias towards rearrangements containing TRA genes belonging to the TRAV18 (12/40 cDNA) and TRAV20 (11/40 cDNA) gene subgroups, was observed (Fig. 3a). On the contrary, the usage of the 61 TRAJ genes is generally random with a slight increase in usage of TRAJ (Fig. 2) between 54 and 22 (31 of

50 functional rearrangements) (Additional file 9); this finding being consistent with the widely accepted view that TRAV-TRAJ recombination proceeds in a coordinated, sequential manner from proximal to progressively more distal TRAV and TRAJ genes [33, 34].

Dolphin TR gamma chain expression analysis demonstrated that the two TRGV and three TRGJ were used in every possible combination, although a bias towards some transcripts (TRGV1-TRGJ2 and TRGV2-TRGJ3) was noted. Furthermore, about half the transcripts using TRGV2 were unproductive due to the presence of stop codons in CDR3. The percentage values of the productive/unproductive rearrangements are similar for both cDNA (Fig. 4) and genomic clones (Fig. 5), in contrast with what is usually observed (percentage of unproductive rearrangements lower in cDNA, due to nonsense-mediated decay of RNA).

In a previous work [35], it was reported that biased V-J gene rearrangement contributes to the regulation of the mature TRG repertoire. The biases in a given TR repertoire can stem from properties of the gene rearrangement process, as well as from thymic selection and the expansion of T cell clones. In the present work, we can make the following considerations: i) it seems to be a double preferentiality and that of the gene TRGV1 with respect to the gene TRGV2 as well as of the TRG V1-J2 rearrangement with respect to the five others (Fig. 4), the latter being supported given the comparison between the frequency of the in-frame and out-of-frame rearrangements both in cDNA and in genomic DNA (Additional file 11); ii) the fact that unrelated subjects show not only a biased usage of V-J genes, but also a biased number of nucleotides inserted/deleted at junction regions (Fig. 4 and Fig. 5c), could be explained by the presence of common antigens which can stimulate and expand T cells with a particular type of gamma chain, suggesting the existence of a basic “public” repertoire of a given TR in a range of public T cell responses; iii) finally we propose that the occurrence of clonotypes shared by different individuals who live both in marine and in artificial marine “habitat”, described as “convergent recombination” [28], could be strictly related to the biased V-J recombinational event.

The mechanisms that determine biases in genes use remain unclear. In a recent paper [36] a physical model of chromatin conformation at the TRB D-J genomic locus explains more than 80 % of the biases in TRBJ use that was measured in murine T cells. As a consequence of these structural and other biases, TR sequences are produced with different a priori frequencies, thus affecting their probability of becoming public TR that are shared among individuals. In dolphin, we could explain the abundance of TRGV1-J2 repertoire among individuals hypothesizing that this combination could be produced by the rearrangement process with different a priori probabilities because an



expanded role of chromatin conformation in TRGV-J rearrangement, which controls both the gene accessibility and the precise determination of gene use.

An evolutionary correlation between the dolphin TRGV1 and the human TRGV9 (Additional file 4A) genes and the dolphin TRGJ2 and the human TRGJP (Additional file 15) genes seems to exist, as in these two species the same mechanism pushes to an accurate determination of the J gene usage. In fact, dolphin TRGJ2 (this work) and human TRGJP, are the most frequently used J genes in the peripheral  $\gamma\delta$  T cells [16] and occupy an intermediate position with respect to the other two J genes. At present we have knowledge of the position of the genes on the physical map for human (IMGT), dromedary [23], dolphin (this work) and sheep [9] (Additional file 1) and cattle [10] TRG loci.

It is admitted that the expressed  $\gamma\delta$  T cell repertoire partly depends upon preferentially rearranged TRGV-J gene combinations, indeed in human the gamma delta TCR repertoire is accompanied by selection for public gamma chain sequences such that many unrelated individuals overlap extensive in their circulating repertoire [37]. As a consequence, the high frequency of TRGV1-J2/TRDV1-D1-J4 productive rearrangements in dolphins may represent a situation of oligoclonality comparable to that found in human with TRGV9-JP/TRDV2-D-J T cells, and in primates.

The similarity in dolphin and human of a basic public  $\gamma\delta$  repertoire, seems to be correlated with other recent findings. McGowen discovered several genes, potentially under positive selection in the dolphin lineage, associated with the nervous system, including those related to human intellectual disabilities, synaptic plasticity and sleep [38]. Moreover bottlenose dolphins are the only animals with man and apes, to be able to recognize themselves when confronted with a mirror [39], and have demonstrated the numerical skills [40]. While here, in the present work, the functional convergence of  $\gamma\delta$  domains is suggested among mammals, recently it was proposed similarity of dual-function TRA and TRD genes in jawed vertebrates and in the VLRA and VLRC genes in jawless vertebrates and their differential expression in two major T cell lineages [41–43]. Therefore comparative immunobiology of different vertebrate lineages may reveal heretofore unrealized features.

## Conclusions

The present study identifies the genomic organization and the gene content of the TRG and the TRA/TRD loci in the high quality draft sequence of the bottlenose dolphin (*Tursiops truncatus*) genome. The genomic structure of the smallest TRG locus thus described in mammals, includes two TRGV, three TRGV and only one TRGC genes. Through phylogenetic and expression analyses, 8 TRAV were assigned unambiguously to genes annotated in the TRA/TRD locus genomic sequence, while 12 TRAV could

be assigned to new genes, belonging to five different subgroups. The presence of several variable genes belonging to the TRDV1 subgroup, makes the TRA/TRD dolphin locus more similar to the TRA/TRD locus of artiodactyls than to the human locus.

By comparing multiple different individuals, we provide evidence of an unusual ratio of productive/unproductive TRG transcripts and of a bias towards TRGV1-TRGJ2 rearrangements, which were dominant among the in-frame transcripts and were rare among the out-of-frame transcripts. Moreover, the cDNA analysis revealed sharing of in-frame TRG sequences within the same and in different individuals living in a controlled environment as well as in marine environment, suggesting expansion of “public” TCR by a common antigen. The selection for public gamma chain and the high frequency of TRGV1-J2/TRDV1-D1-J4 productive rearrangements in dolphins may represent a situation comparable to that found in human with TRGV9-JP/TRDV2-D-J T cells.

## Methods

### Genome and sequence analysis

The bottlenose dolphin (*Tursiops truncatus*) genome is being sequenced at ~2X coverage (BioProject: PRJNA20367) by the Human Genome Sequencing Center at the Baylor College of Medicine and the Broad Institute using a whole genome shotgun sequencing strategy [17]. In 2008, Ensembl released the first low-coverage 2.59x assembly of the dolphin (turTru1). We employed these genome assemblies using BLAST algorithm to identify the TRG and TRA/TRD loci in this species.

For the TRG locus, two overlapping scaffolds were retrieved (GEDI ID: JH473572.1; BCM-HGSC ID: contig 425448–578749), respectively of 96017 and 284974 bp (gaps included). A sequence of 188414 bp was analysed. Amphiphysin (AMPH) and related to steroidogenic acute regulatory protein D3-N-terminal like (STARD3NL), flanking TRG locus at 5' and 3' ends, respectively, were included in the analysis. They were predicted to be functional in dolphin (GenBank ID: XM\_004317564.1; Ensembl ID: ENSTTRT00000004099). The TRG genes were identified using both our dolphin cDNA collection (this work) and the corresponding human (GEDI ID: AF159056) and sheep (GEDI ID: DQ992075.1, DQ992074.1) genomic sequences. Locations of the TRG genes are provided in Fig. 1b.

For the TRA/TRD locus, we retrieved a sequence of 482052 bp from two GenBank scaffolds, JH484271.1 and JH481615.1, and five EMBL-EBI scaffolds, Ens\_742, Ens\_97, Ens\_89, Ens\_123 and Ens\_112178. Scaffold Ens\_97 and Ens\_123 overlap for about 16,7 Kb, including TRA14/DV4, TRA9, TRA16 and TRA17 genes, while scaffold Ens\_89 and JH484271.1 overlap for about 10 Kb, a region that includes two genes, TRAV1S1 and TRAV38.1. The TRA/TRD genes were identified using



the corresponding human (GEDI ID: AE000521.1) genomic sequences. Sequences of all TRA/TRD genes are in Additional file 5. Computational analysis of the dolphin TR loci was conducted using the following programs: RepeatMasker for the identification of genome-wide repeats and low complexity regions [44] (RepeatMasker <http://www.repeatmasker.org>) and Pipmaker [45] (<http://www.pipmaker.bx.psu.edu/pipmaker/>) for the alignment of the dolphin sequence with the human counterpart. ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) and IMGT/V-QUEST ([http://www.imgt.org/IMGT\\_vquest/share/textes/](http://www.imgt.org/IMGT_vquest/share/textes/)) tools allowed the identification and characterization of the TR genes.

### Phylogenetic analyses

The TRGV, TRGJ, TRAV and TRDV genes used for the phylogenetic analyses were retrieved from IMGT/LIGM-DB and GenBank databases with the following accession numbers: AF159056 (human TRG locus), DQ992075 (sheep TRG1 locus), DQ992074 (sheep TRG2 locus), JN165102 (dromedary TRGV1), JN172913 (dromedary TRGV1), AE000521.1 (human TRA/TRD locus); sheep TRA/TRD accession numbers [14] and FN298219-FN298227 (dromedary TRD genes) [46]. Multiple alignments of the sequences under analysis were carried out with the MUSCLE program [47]. Phylogenetic analyses were performed using MEGA version 6.06 [48] and the bootstrap consensus tree inferred from 1000 replications using the Neighbor-Joining method [49, 50].

### Animals (source of tissue)

Blood samples were provided by Zoomarine Italia S.p.A. (Rome, Italy) and were collected from three dolphins, two males (Marco and King) and one female (Leah). The three individuals were born and kept under human care and are unrelated. In particular, Marco was born in the dolphinarium in Bruges (Belgium) and Tex, Marco's father, is from the United States (Texas, Gulf of Mexico). King was born in the dolphinarium in Albufeira (Portugal), and Sam, King's father had Cuban origins. Leah was born in the dolphinarium in Benidorm (Spain) and Eduardo, Leah's father has Cuban origins. The identifying letters are M, K and L, respectively. The Bank for the Tissues of Mediterranean Marine Mammals (Padua, Italy) provided us a sample of skin (epidermis plus dermis) belonging to a wild dolphin, that was found beached in the Northern Adriatic Sea; for this animal the identified letter is C.

### 5' RACE and RT-PCR

Four types of 5' RACE and three types of RT-PCR (total of six and six experiments, respectively) on total RNA from the peripheral blood of three unrelated adult animals (identified as M, K and L) and from the skin of animal (identified as C) (Table 1 and 2) were carried out to

investigate the dolphin TRG, TRA and TRD chains repertoire. Two 5' RACE experiments from the peripheral blood of the animals (identified as M and L) and three types of RT-PCR, two from blood (K and M) and one from skin (C), were carried out to investigate the dolphin TRG chain repertoire. A single 5' RACE experiment from the peripheral blood of the animal identified as L was carried out to investigate the dolphin TRA chain repertoire. Three 5' RACE experiments from the peripheral blood of the animals (identified as M and L) were carried out to investigate the dolphin TRD chain repertoire.

Total RNA was isolated from peripheral blood leukocytes (PBL) or skin using the Trizol method according to the manufacturer's protocol (Invitrogen, Carlsbad, CA), and integrity of RNA was verified on a 1 % agarose gel. About 5 µg of total RNA were reverse transcribed with Superscript II (Invitrogen, Carlsbad, CA) by using specific primers (Table 1), designed on the sequences of the first exon for each dolphin TR constant gene sequence (TC3L for gamma chain, TA1C1L for alpha chain and TD2CL for delta chain). After linking a poly-C tail at the 5' end of the cDNAs, the cDNAs was performed with Platinum Taq Polymerase (Invitrogen) by using specific primers as lower primers, TC1L1 for gamma chain, TA1C2L for alpha chain and TD1C2L for delta chain (Table 1) and an anchor oligonucleotide as upper primer (AAP) provided from the supplier (Invitrogen). PCR conditions were the following: one cycle at 94 °C for 1 min; 35 cycles at 94 °C for 30 s, 58 °C for 45 s, 72 °C for 1 min; a final cycle of 30 min at 72 °C. The products were then amplified in a subsequent nested PCR experiment by using specific lower primers, TC1L2 for gamma chain, TA1C3L for alpha chain and TD1C1L for delta chain (Table 1) and AUAP oligonucleotide as upper primer, provided from the supplier (Invitrogen). Nested PCR conditions were the following: one cycle at 94 °C for 1 min; 30 cycles at 94 °C for 30 s, 58 °C for 35 s, 72 °C for 30 s; a final cycle of 30 min at 72 °C. RT-PCR experiments were carried out amplifying rearranged transcripts containing TRGV1 and TRGV2 genes. Upper primers containing TRGV1 (TV1LU) and TRGV2 (TV7LU) sequences, and lower primer containing the I exon of TRGC (TC1L2) sequence were used on sscDNA (Table 1 and 2). RT-PCR conditions were: one cycle at 94 °C for 2,30 min; 35 cycles at 94 °C for 30 s, 58 °C for 40 s, 72 °C for 40 s; a final cycle of 30 min at 72 °C. The RT-PCR and RACE products were then gel-purified and cloned using StrataClone PCR Cloning Kit (Statagene). Random selected positive clones for each cloning were sequenced by a commercial service. cDNA sequence data were processed and analyzed using the Blast program (<http://www.blast.ncbi.nlm.nih.gov/Blast.cgi>), Clustal W2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) and IMGT\_ tools (IMGT/V-QUEST) [51, 52] with integrated IMGT/JunctionAnalysis tools [53, 54] and the IMGT unique numbering for V domain [25] (<http://www.imgt.org/>).

### Genomic DNA isolation and PCR

Genomic DNA was extracted from whole blood of a female subject (animal identifier letter L), with a salting-out method [55] with two modifications. First, whole blood was mixed with erythrocyte lysis buffer (155 mM  $\text{NH}_4\text{Cl}$ , 10 mM  $\text{KHCO}_3$ , 1 mM EDTA, pH 7.4) before the harvested white cell pellet was mixed with nucleus lysis buffer as described [55]. Second, incubation with proteinase K was carried out for 2 h at 56 °C, instead of overnight at 37 °C. The quality of the genomic DNA was evaluated by agarose gel electrophoresis and concentration determined by 260 nm absorbance measurements. Genomic PCR was performed with 50 ng to 100 ng of genomic DNA as template using specific upper primers (TV1L1 and TV7LU) designed on the two TRGV (TRGV1 and TRGV2) gene sequences in combination with two lower primers (J2GL and J5BR) designed on the two TRGJ (TRGJ2 and TRGJ3) gene sequences (Table 1). Two genomic PCR were performed to amplify TRGV1-TRGJ2 and TRGV2-TRGJ3 rearrangement combinations, respectively. High-fidelity polymerase was used to minimize possible PCR errors. PCR were performed following the manufacturer's instruction for the DNA polymerase (Platinum<sup>®</sup>Taq DNA Polymerase, Life Technologies). V1-J2 genomic PCR conditions were the following: one cycle at 94 °C for 3 min; 35 cycles at 94 °C for 30 s, 62 °C for 30 s, 72 °C for 30 s; a final cycle of 30 min at 72 °C. V2-J3 genomic PCR conditions were the following: one cycle at 94 °C for 3 min; 35 cycles at 94 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s; a final cycle of 30 min at 72 °C. The obtained fragments were agarose gel purified, cloned using StrataClone PCR Cloning Kit (Stratagene) and sequenced by a commercial service. Two genomic PCR were performed to amplify TRDV4 gene using a pair of primers designed based on the relative V-exon sequence (Table 1). PCR was performed following the manufacturer's instruction for the MyTaq<sup>™</sup> HS DNA Polymerase, (Bioline). The following settings were used: 94 °C for 2 min, followed by 30 cycles each comprising a denaturation step at 94 °C for 30 s, an annealing step of 30 s at 55 °C (according to the melting temperature of the primers), an extension step at 72 °C for 30 s, and a final extension period of 7 min at 72 °C.

### Statistical analysis

Statistical analyses were performed using 2 × 2 contingency tables. All the *p*-values shown in the Results were obtained using the Chi-squared test, considering as statistically significant a *p*-value <0.05. Fisher's Exact test was used to confirm the significance of the Chi-squared test when the counts of observed samples had values <5. When performing multiple comparisons among in-frame and out-of-frame TRG cDNA (Additional file 12), the Chi-squared test *p*-values were

adjusted using Benjamini–Hochberg false discovery rate [56]. All the analyses were performed using the R software environment for statistical computing (<https://www.r-project.org/>).

### Global alignments in protein secondary structure prediction and 3D visualization

Global alignment of the target and template sequences was performed with ClustalW (<http://www.ebi.ac.uk/clustalw/index.html>) [57]. Furthermore, when necessary, alignment was manually adjusted after predicting the secondary structure of the target and aligned to that of the template as derived with the DSSP program [58]. The secondary structure prediction was computed with SECPRED ([http://gpcr.biocomp.unibo.it/cgi/predictors/s/pred\\_seccgi.cgi](http://gpcr.biocomp.unibo.it/cgi/predictors/s/pred_seccgi.cgi)) and PSIPRED [59] and the target/template alignments were computed with YAP (<http://gpcr.biocomp.unibo.it/cgi/predictors/alignss/alignss.cgi>), that allows to align both primary and secondary structure at the same time. The template was selected from the Protein Data Bank (PDB) on the basis of sequence/function similarity with the target sequence and was the human  $\gamma\delta$  T cell receptor solved with an atomic resolution of 3 Å (PDB code and IMGT/3Dstructure-DB: 1hxm) [29, 30]. Target/ template alignments were then fed into Modeller version 9.8 [60]. For a given alignment, 50 3D models were routinely built and, then, evaluated and validated with the PROCHECK [61] and PROSA2003 [62] suites of programs. Models with the best stereochemical and energetics features were retained. 3D visualization (Additional file 14) of the RTV1M1/5R1D8 and of 5RV1M1/5R1D15 clones was computed, adopting as template the human  $\gamma\delta$  T cell receptor. The solvent accessibility was computed with DSSP program [58]. The protein complex interface were computed by the online tool PDBePISA at the EBI server ([http://www.ebi.ac.uk/msd-srv/prot\\_int/](http://www.ebi.ac.uk/msd-srv/prot_int/)) and visualized by UCSF Chimera tool (<http://www.cgl.ucsf.edu/chimera/>). The IMGT Collier de Perles of RTV1M1, 5R1D8, 5RV1M1 and 5R1D15 cDNA clonotypes were obtained using the IMGT/Collier-de-Perles tool (<http://www.imgt.org>) [63], starting from amino acid sequences.

### Additional files

**Additional file 1:** Schematic representation of the genomic organization of human, sheep, dromedary and dolphin TRG loci. The diagram shows the position of all V, J, and C TRG genes according to IMGT nomenclature (<http://www.imgt.org>). Boxes representing genes are not to scale. Exons are not shown. (PPT 146 kb)

**Additional file 2:** Nucleotide sequences of the dolphin TRGV (A), TRGJ (B) and TRGC (C) genes, as deduced from the genome assembly Ttru\_1.4. (DOC 19 kb)

**Additional file 3:** Dotplot matrix of dolphin/sheep (A) and of dolphin/human (B) TRG loci genomic comparison. Using the PipMaker program dolphin TRG has been plotted against sheep TRG1 (A) and dolphin TRG locus

has been plotted against human (B). The transcriptional orientation of each gene is indicated by arrows and arrowheads. Dolphin TRGV1 and TRGV2 genes were classified as orthologues to their corresponding human TRGV9 gene and sheep TRGV11-1 (a pseudogene) and human TRGV11 (an ORF) and sheep TRGV7 gene, respectively (red boxes). The correspondence is due to the highest nucleotide identity (see also Additional file 4A). (PPT 363 kb)

**Additional file 4:** The NJ tree inferred from the dolphin, sheep, dromedary and human TRGV (A) and TRGC (B) gene sequences. The evolutionary analysis was conducted in MEGA6.06 [48]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches [49]. The trees are drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. The evolutionary distances were computed using the *p*-distance method [50] and are in the units of the number of base differences per site. (PDF 67 kb)

**Additional file 5:** Description of the TRA/TRD genes in the dolphin genome assembly. The position of all genes and their classification and functionality are reported. (DOCX 149 kb)

**Additional file 6:** Nucleotide (A) and deduced amino acid (B) sequences of the dolphin TRDV4 gene. TRDV4 indicates the assembly gene; G1TRDV4 and G2TRDV4 genes refer to two different individual genomic sequences. The underlined A in G2 TRDV4 may represent an allele of the TRDV4 gene. (DOCX 18 kb)

**Additional file 7:** The NJ tree inferred from the dolphin, sheep, and human TRAV (A) and from the dolphin, sheep, dromedary and human TRDV (B) gene sequences. The evolutionary analysis was conducted in MEGA6.06 [48]. The percentage of replicate trees in which the associated taxa together in the bootstrap test (1,000 replicates) is shown next to the branches [49]. The trees are drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. The evolutionary distances were computed using the *p*-distance method [50] and are in the units of the number of base differences per site. The functionality of all genes is also indicated. (A) Subgroups TRAV45 and TRAV42 are officially adopted for dolphin: these genes are related to the TRAV18 and TRAV8 subgroups, respectively. (B) Dolphin TRDV1, TRDV2 and TRDV4 gene subgroups are indicated by A, B and C, respectively. (PPT 244 kb)

**Additional file 8:** Nucleotide and deduced amino acid sequences of the dolphin TRDJ (A) and TRDD (B) genes. The consensus sequences of the heptamer and nonamer [64] are provided at the top of the figure and underlined. The numbering adopted for the gene classification is reported on the left of each gene. The donor splice site for each TRDJ is shown. The canonical FGXG amino acid motifs are underlined. (DOC 16 kb)

**Additional file 9:** Nucleotide and deduced amino acid sequences of the dolphin TRAJ genes. The consensus sequence of the heptamer and nonamer [64] are provided at the top of the figure and are underlined. The numbering adopted for the gene classification, is reported on the left of each gene. The donor splice site for each TRAJ is shown. The canonical FGXG amino acid motifs are underlined. (DOCX 53 kb)

**Additional file 10:** IMGT Protein display of the TRG cDNA clones. The TRGV and TRGJ genes are listed respectively at the left and the right of the figure. Leader region (L-Region), complementarity determining regions (CDR-IMGT) and framework regions (FR-IMGT) are also indicated, according to the IMGT unique numbering for V-REGION [25]. The name of the clones are also reported. Shared sequences both within the same and in different individuals are in bold. The TRGV allele amino acid changes, if any, are green boxed. (DOC 60 kb)

**Additional file 11:** Percentages of in-frame and out-of-frame rearranged TRG V-J cDNA (A) and percentages of in-frame and out-of-frame rearranged TRG V-J genomic DNA (B). L is animal identifier letter. Blue areas of the bars indicate in-frame while the red areas of the bars indicate out-of-frame TRG V-J rearrangements (Fig. 4 and Additional file 10). (PPTX 37 kb)

**Additional file 12:** Summary of the results of the statistical test: Chi-squared *p*-value is confirmed with Fisher's *p*-value. We assume as

'null hypothesis' that in-frame and out-of-frame cDNAs are produced with the same probability and that there is no significant difference among in-frame and out-of frame occurrences. Thus, the null hypothesis is that the rate of in-frame and out-of frame cDNAs is proportional to the totals (or, to better say, to the remaining counts), for each category. The hypothesis has been tested using Chi-squared *p*-value and (due to the low counts) confirmed with Fisher's *p*-value. (DOC 31 kb)

**Additional file 13:** Overview of the analysis of the putative gd domains conducted with the software PDBePISA (M&M) (<http://www.ebi.ac.uk/pdbe/pisa/>). For each paired domain, 20 models were generated and after validation a representative was chosen. The columns represent, respectively: the number of H bond, the name and the position of the amino acid and of the atom involved in the H bond for delta domain; the 3x indicates that the amino acid is found in the CDR3 (IMGT\_Collier de Perles) [65]. The length of the hydrogen bond expressed in angstrom, the name and position of amino acids, numeration and the atom involved in the hydrogen bond of the gamma domain, follow respectively. The 3x at the end, indicates that the amino acid is found in the CDR3 of the gamma domain. The positions highlighted in yellow indicate the salt bridge (s). (PPTX 83 kb)

**Additional file 14:** Visualization of computationally inferred interaction between V-gamma and V-delta domain cDNA clonotypes. In RTV1M1 and 5RV1M1 V-gamma domain CDR-IMGT are blue-green-green (FR in orange); in 5R1D8 and 5R1D15 V-delta domain CDR-IMGT are red-pink-violet (FR in yellow). The protein complex interface were computed by the online tool PDBePISA at the EBI server. ([http://www.ebi.ac.uk/msd/srv/prot\\_int/](http://www.ebi.ac.uk/msd/srv/prot_int/)) and visualized by UCSF Chimera tool (<http://www.cgl.ucsf.edu/chimera/>). (PPTX 762 kb)

**Additional file 15:** The NJ tree inferred from the dolphin, sheep, dromedary and human TRGJ gene sequences. The evolutionary analysis was conducted in MEGA6.06 [48]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches [49]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. The evolutionary distances were computed using the *p*-distance method [50] and are in the units of the number of base differences per site. The functionality of all genes is also indicated. In the three, a clear cut subdivision of J sequences into two main sets is evident: set I (C-proximal) and set II (C-distal); genes of set III, have in the physical map an intermediate position with respect to J genes of the other two sets (Additional file 1). (PPT 159 kb)

#### Abbreviations

CDR, complementarity determining region; FR, framework region; IG, immunoglobulins; T cell receptor gamma locus; TR, T cell receptor; TRA/TRD locus, T cell receptor alpha/delta locus; TRG locus, TRGC, T cell receptor gamma constant; TRGJ, T cell receptor gamma joining; TRGV, T cell receptor gamma variable. All TR genes (functional, ORF, pseudogenes) reported here have been approved by the IMGT/WHO- IUIS nomenclature committee and their designations are in accord with the IMGT nomenclature for human (IMGT®, the international ImMunoGeneTics information system®, <http://www.imgt.org>)

#### Acknowledgments

The financial support of the University of Bari and of the University of Salento is gratefully acknowledged. A.C. is supported by Progetto MICROMAP PON01\_02589". Thanks are due to COST Action TD1101 and COST Action BM1405 delivered to R.C. We thank Stefania Brandini and Angela Pala for technical assistance in cDNA cloning experiments and Prof. P. Barsanti for critically reading of the manuscript. We thank all the staff of Zoomarine Italy. We are especially thankful to all the trainers who dedicate themselves every day to dolphins. Through their work, they make possible the study of medical behaviours. We thank the members of the IMGT® team for their expertise and constant motivation.

#### Funding

Università degli Studi di Bari Aldo Moro (ex 60 %).



**Authors' contributions**

GL, RA, and SC designed research; GL, GT, VC, and AC performed research; FG contributed new reagents/analytic tools; RA, RC, SM, M-PL and SC analyzed data; M-PL improved the manuscript; GL and SC wrote the paper. All authors have read and approved the final manuscript.

**Competing interests**

The author declares that he/she has no competing interests.

**Consent for publication**

Not applicable.

**Ethics approval and consent to participate**

Zoomarine is a seaside water park. The availability of blood of this study was a byproduct of standard health checks. The blood samples were taken from the caudal vein on the ventral surface of the caudal fin. Training of medical behaviours brings the animals to collaborate completely, assuming and maintaining the positions that allow the veterinarian, constantly assisted by the trainers, to perform the necessary procedures to monitor their welfare. Zoomarine Italia SpA, via Casablanca 61, 00071 Pomezia (RM), Italy <https://www.zoomarine.it/>.

**Data deposition**

TR cDNA and genomic sequences were submitted to the GenBank database. TRG cDNAs are under accession numbers: JF755948-JF755968 and JN011996; TRG genomic sequences are under accession numbers: LN886662 - LN886693. TRA cDNAs are under accession numbers: LN610706 - LN610746; TRD cDNAs are under accession numbers: LN610747 - LN610749. Phylogenetic trees presented in the Additional file 4A and B, in the Additional file 7A and B and in the Additional file 15 were deposited in TreeBase (<http://purl.org/phylo/treebase/phyloids/study/TB2S19396>).

**Author details**

<sup>1</sup>Department of Biology, University of Bari, via E. Orabona 4, 70125 Bari, Italy. <sup>2</sup>Biocomputing Group, CIRI-Health Science and Technologies/Department of Biology, University of Bologna, via Selmi 3, 40126 Bologna, Italy. <sup>3</sup>Zoomarine Italia SpA, via Casablanca 61, 00071 Pomezia, RM, Italy. <sup>4</sup>Department of Biological and Environmental Science e Technologies, University of Salento, via per Monteroni, 73100 Lecce, Italy. <sup>5</sup>CNR, Institute for Biomedical Technologies of Bari, via Amendola, 70125 Bari, Italy. <sup>6</sup>IMGT®, the international ImMunoGeneTics information system®, Laboratoire d'Immunogénétique Moléculaire, Institut de Génétique Humaine, UPR CNRS 1142, University of Montpellier, 34396 Montpellier Cedex 5, France.

Received: 29 January 2016 Accepted: 15 June 2016

Published online: 15 August 2016

**References**

- Gingerich PD, Haq M, Zalmout IS, Khan IH, Malkani MS. Origin of whales from early artiodactyls: hands and feet of Eocene Protocetidae from Pakistan. *Science*. 2001;293(5538):2239–42.
- Price SA, Bininda-Emonds OR, Gittleman JL. A complete phylogeny of the whales, dolphins and even-toed hoofed mammals (Cetartiodactyla). *Biol Rev Camb Philos Soc*. 2005;80(3):445–73.
- Boisserie JR, Lihoreau F, Brunet M. The position of Hippopotamidae within Cetartiodactyla. *Proc Natl Acad Sci U S A*. 2005;102(5):1537–41.
- Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. The delayed rise of present-day mammals. *Nature*. 2007;446(7135):507–12.
- Geisler JH, Theodor JM. Hippopotamus and whale phylogeny. *Nature*. 2009;458(7236):E1–4. discussion E5.
- Andresdottir V, Magnadottir B, Andresson OS, Petursson G. Subclasses of IgG from whales. *Dev Comp Immunol*. 1987;11(4):801–6.
- Mancia A, Romano TA, Gefroh HA, Chapman RW, Middleton DL, Warr GW, Lundqvist ML. The Immunoglobulin G Heavy Chain (IGHG) genes of the Atlantic bottlenose dolphin, *Tursiops truncatus*. *Comp Biochem Physiol Part B, Biochem mol Biol*. 2006;144(1):38–46.
- Mancia A, Romano TA, Gefroh HA, Chapman RW, Middleton DL, Warr GW, Lundqvist ML. Characterization of the immunoglobulin A heavy chain gene of the Atlantic bottlenose dolphin (*Tursiops truncatus*). *Vet Immunol Immunopathol*. 2007;118(3–4):304–9.
- Vaccarelli G, Miccoli MC, Antonacci R, Pesole G, Ciccarese S. Genomic organization and recombinational unit duplication-driven evolution of ovine and bovine T cell receptor gamma loci. *BMC Genomics*. 2008;9:81.
- Conrad ML, Mawer MA, Lefranc MP, McKinnell L, Whitehead J, Davis SK, Pettman R, Koop BF. The genomic sequence of the bovine T cell receptor gamma TRG loci and localization of the TRGC5 cassette. *Vet Immunol Immunopathol*. 2007;115(3–4):346–56.
- Ciccarese S, Vaccarelli G, Lefranc MP, Tasco G, Consiglio A, Casadio R, Linguiti G, Antonacci R. Characteristics of the somatic hypermutation in the Camelus dromedarius T cell receptor gamma (TRG) and delta (TRD) variable domains. *Dev Comp Immunol*. 2014;46(2):300–13.
- Antonacci R, Lanave C, Del Faro L, Vaccarelli G, Ciccarese S, Massari S. Artiodactyl emergence is accompanied by the birth of an extensive pool of diverse germline TRDV1 genes. *Immunogenetics*. 2005;57(3–4):254–66.
- Connelley TK, Degnan K, Longhi CW, Morrison WI. Genomic analysis offers insights into the evolution of the bovine TRA/TRD locus. *BMC Genomics*. 2014;15:994.
- Piccinni B, Massari S, Caputi Jambrenghi A, Giannico F, Lefranc MP, Ciccarese S, Antonacci R. Sheep (*Ovis aries*) T cell receptor alpha (TRA) and delta (TRD) genes and genomic organization of the TRA/TRD locus. *BMC Genomics*. 2015;16:709.
- Lefranc MP, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse T, et al. IMGT(R), the international ImMunoGeneTics information system(R) 25 years on. *Nucleic Acids Res*. 2015;43(Database issue):D413–22.
- Lefranc MP, Lefranc G. The T cell receptor fact book. 2001.
- Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. 2011;478(7370):476–82.
- Glusman G, Rowen L, Lee I, Boysen C, Roach JC, Smit AF, Wang K, Koop BF, Hood L. Comparative genomics of the human and mouse T cell receptor loci. *Immunity*. 2001;15(3):337–49.
- Massari S, Bellahcene F, Vaccarelli G, Carelli G, Mineccia M, Lefranc MP, Antonacci R, Ciccarese S. The deduced structure of the T cell receptor gamma locus in *Canis lupus familiaris*. *Mol Immunol*. 2009;46(13):2728–36.
- Massari S, Ciccarese S, Antonacci R. Structural and comparative analysis of the T cell receptor gamma (TRG) locus in *Oryctolagus cuniculus*. *Immunogenetics*. 2012;64(10):773–9.
- Herzig C, Blumerman S, Lefranc MP, Baldwin C. Bovine T cell receptor gamma variable and constant genes: combinatorial usage by circulating gammadelta T cells. *Immunogenetics*. 2006;58(2–3):138–51.
- Antonacci R, Vaccarelli G, Di Meo GP, Piccinni B, Miccoli MC, Cribiu EP, Perucatti A, Iannuzzi L, Ciccarese S. Molecular in situ hybridization analysis of sheep and goat BAC clones identifies the transcriptional orientation of T cell receptor gamma genes on chromosome 4 in bovids. *Vet Res Commun*. 2007;31(8):977–83.
- Vaccarelli G, Antonacci R, Tasco G, Yang F, Giordano L, El Ashmaoui HM, Hassanane MS, Massari S, Casadio R, Ciccarese S. Generation of diversity by somatic mutation in the Camelus dromedarius T-cell receptor gamma variable domains. *Eur J Immunol*. 2012;42(12):3416–28.
- Buresi C, Ghanem N, Huck S, Lefranc G, Lefranc MP. Exon Duplication and Triplication in the Human T-Cell Receptor-Gamma Constant Region Genes. *Cytogenet Cell Genet*. 1989;51(1–4):973.
- Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol*. 2003;27(1):55–77.
- Herzig CTA, Lefranc MP, Baldwin CL. Annotation and classification of the bovine T cell receptor delta genes. *BMC Genomics*. 2010;11:100.
- Li S, Lefranc MP, Miles JJ, Alamyar E, Giudicelli V, Duroux P, Freeman JD, Corbin VD, Scheerlinck JP, Frohman MA, et al. IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. *Nat Commun*. 2013;4:2333.
- Venturi V, Kedzierska K, Tanaka MM, Turner SJ, Doherty PC, Davenport MP. Method for assessing the similarity between subsets of the T cell receptor repertoire. *J Immunol Methods*. 2008;329(1–2):67–80.
- Xu B, Pizarro JC, Holmes MA, McBeth C, Groh V, Spiess T, Strong RK. Crystal structure of a gammadelta T-cell receptor specific for the human MHC class I homolog MICA. *Proc Natl Acad Sci U S A*. 2011;108(6):2414–9.



30. Ehrenmann F, Kaas Q, Lefranc MP. IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF. *Nucleic Acids Res.* 2010;38:D301–7.
31. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* 2007;372(3):774–97.
32. Bovine Genome S, Analysis C, Elisk CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, Weinstock GM, Adelson DL, Eichler EE, et al. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science.* 2009;324(5926):522–8.
33. Jouvin-Marche E, Fuschioti P, Marche PN. Dynamic aspects of TCRalpha gene recombination: qualitative and quantitative assessments of the TCRalpha chain repertoire in man and mouse. *Adv Exp Med Biol.* 2009;650:82–92.
34. Krangel MS. Mechanics of T cell receptor gene rearrangement. *Curr Opin Immunol.* 2009;21(2):133–9.
35. Kohsaka H, Chen PP, Taniguchi A, Ollier WE, Carson DA. Regulation of the mature human T cell receptor gamma repertoire by biased V-J gene rearrangement. *J Clin Invest.* 1993;91(1):171–8.
36. Ndifon W, Gal H, Shifrut E, Aharoni R, Yissachar N, Waysbort N, Reich-Zeliger S, Arnon R, Friedman N. Chromatin conformation governs T-cell receptor Jbeta gene segment usage. *Proc Natl Acad Sci U S A.* 2012;109(39):15865–70.
37. Pauza CD, Cairo C. Evolution and function of TCR Vgamma9 chain repertoire: It's good to be public. *Cell Immunol.* 2015;296:22–30.
38. McGowen MR, Grossman LI, Wildman DE. Dolphin genome provides evidence for adaptive evolution of nervous system genes and a molecular rate slowdown. *Proc Biol Sci / Royal Soc.* 2012;279(1743):3643–51.
39. Reiss D, Marino L. Mirror self-recognition in the bottlenose dolphin: a case of cognitive convergence. *Proc Natl Acad Sci U S A.* 2001;98(10):5937–42.
40. Jaakkola K, Fellner W, Erb L, Rodriguez M, Guarino E. Understanding of the concept of numerically "less" by bottlenose dolphins (*Tursiops truncatus*). *J Comp Psychol.* 2005;119(3):296–303.
41. Das S, Li J, Hirano M, Sutoh Y, Herrin BR, Cooper MD. Evolution of two prototypic T cell lineages. *Cell Immunol.* 2015;296(1):87–94.
42. Hirano M, Das S, Guo P, Cooper MD. The evolution of adaptive immunity in vertebrates. *Adv Immunol.* 2011;109:125–57.
43. Krangel MS, Carabana J, Abbarategui I, Schlimgen R, Hawwari A. Enforcing order within a complex locus: current perspectives on the control of V(D)J recombination at the murine T-cell receptor alpha/delta locus. *Immunol Rev.* 2004;200:224–32.
44. Smit AFA, Hubley R & Green P: RepeatMasker open-4.0. 2013–2015 <http://www.repeatmasker.org>.
45. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res.* 2000;10(4):577–86.
46. Antonacci R, Mineccia M, Lefranc MP, Ashmaoui HM, Lanave C, Piccinni B, Pesole G, Hassanane MS, Massari S, Ciccarese S. Expression and genomic analyses of Camelus dromedarius T cell receptor delta (TRD) genes reveal a variable domain repertoire enlargement due to CDR3 diversification and somatic mutation. *Mol Immunol.* 2011;48(12–13):1384–96.
47. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7.
48. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011;28(10):2731–9.
49. Felsenstein J. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution.* 1985;39(4):783–91.
50. Nei M, Kumar S. *Molecular Evolution and Phylogenetics.* 2000.
51. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res.* 2008;36(Web Server issue):W503–8.
52. Giudicelli V, Brochet X, Lefranc MP. IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. *Cold Spring Harbor protocols.* 2011;2011(6):695–715.
53. Yousfi Monod M, Giudicelli V, Chaume D, Lefranc MP. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONS. *Bioinformatics.* 2004;20 Suppl 1:i379–85.
54. Giudicelli V, Lefranc MP. IMGT/junctionanalysis: IMGT standardized analysis of the V-J and V-D-J junctions of the rearranged immunoglobulins (IG) and T cell receptors (TR). *Cold Spring Harbor protocols.* 2011;2011(6):716–25.
55. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16(3):1215.
56. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol.* 1995;57(1):289–300.
57. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947–8.
58. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 1983;22(12):2577–637.
59. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* 1999;292(2):195–202.
60. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al.]* 2006, Chapter 5:Unit 5 6.
61. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993;26(2):283–91.
62. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins.* 1993;17(4):355–62.
63. Lefranc MP, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, Wu Y, Gemrot E, Brochet X, Lane J, et al. IMGT, the international Immunogenetics information system. *Nucleic Acids Res.* 2009; 37(Database issue):D1006–12.
64. Hesse JE, Lieber MR, Mizuuchi K, Gellert M. V(D)J recombination: a functional definition of the joining signals. *Genes Dev.* 1989;3(7):1053–61.
65. Lefranc MP. IMGT Collier de Perles for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. *Cold Spring Harbor protocols.* 2011;2011(6):643–51.
66. Lefranc MP. WHO-IUIS Nomenclature Subcommittee for immunoglobulins and T cell receptors report. *Immunogenetics.* 2007;59(12):899–902.
67. Lefranc MP. Antibody nomenclature: from IMGT-ONTOLOGY to INN definition. *mAbs.* 2011;3(1):1–2.
68. Lefranc MP, Pommie C, Kaas Q, Duprat E, Bosc N, Guiraudou D, Jean C, Ruiz M, Da Piedade I, Rouard M, et al. IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev Comp Immunol.* 2005;29(3):185–203.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

