



Characterization of the ferret TRB locus guided by V, D, J, and C gene expression analysis

Bram Gerritsen^{1,2} · Aridaman Pandit³ · Fatiha Zaaoui-Boutahar⁴ · Mirjam C. G. N. van den Hout⁵ · Wilfred F. J. van IJcken⁵ · Rob J. de Boer¹ · Arno C. Andeweg⁴

Received: 17 October 2019 / Accepted: 18 October 2019 / Published online: 3 December 2019
© The Author(s) 2019

Abstract

The domestic ferret, *Mustela putorius furo*, is an important mammalian animal model to study human respiratory infection. However, insufficient genomic annotation hampers detailed studies of ferret T cell responses. In this study, we analyzed the published T cell receptor beta (TRB) locus and performed high-throughput sequencing (HTS) of peripheral blood of four healthy adult ferrets to identify expressed V, D, J, and C genes. The HTS data is used as a guide to manually curate the expressed V, D, J, and C genes. The ferret locus appears to be most similar to that of the dog. Like other mammalian TRB loci, the ferret TRB locus contains a library of variable genes located upstream of two D-J-C gene clusters, followed by a (in the ferret non-functional) V gene with an inverted transcriptional orientation. All TRB genes (expressed or not) reported here have been approved by the IMGT/WHO-IUIS nomenclature committee.

Keywords T cell receptor · TRB locus · Ferret (*Mustela putorius furo*) · IMGT · Comparative genomics · Immune repertoire sequencing

Introduction

The ferret (*Mustela putorius furo*) is an important mammalian model species to study human respiratory infections. Ferret infection models are well suited to study the pathogenicity and

transmissibility of Coronaviruses (SARS), Pneumoviridae (RSV) and Orthomyxoviruses that include human and avian influenza viruses (Enkirch and von Messling 2015; Oh and Hurt 2016). Ferrets are an attractive mammalian model species for these infections since ferrets and humans share similar lung physiology, and most notably a very similar (viral) receptor distribution throughout the respiratory tract (Belser et al. 2011; van Riel et al. 2006). A significant drawback of the ferret model is a lack of ferret specific reagents for detailed studies of the host immune response to these pathogens. Nevertheless, the use of the ferret model has increased over the years, and its usage, with the recent publication of the ferret (draft) genome (Peng et al. 2014), is likely to increase even further. Currently, little is known about the T cell receptor (TCR) repertoire of the ferret, limiting the options to monitor the immune response of ferrets to experimental infections with influenza virus and other pathogens.

TCRs mediate recognition of peptide antigens presented to T lymphocytes via the peptide-MHC complex (Davis and Bjorkman 1988). Conventional TCRs are $\alpha\beta$ or $\gamma\delta$ heterodimers that are formed by somatic rearrangement of Variable (V), Diversity (D), and Joining (J) gene segments for the β and δ chains, and V and J gene segments for the α and γ chains (Davis and Bjorkman 1988). Although the ratio between $\alpha\beta$ and $\gamma\delta$ T cell subsets is not known for the ferret, the $\alpha\beta$ T cells

This article is part of the Topical Collection on "Nomenclature, databases and bioinformatics in Immunogenetics"

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00251-019-01142-9>) contains supplementary material, which is available to authorized users.

✉ Arno C. Andeweg
a.andeweg@erasmusmc.nl

- Theoretical Biology and Bioinformatics, Utrecht University, Utrecht, the Netherlands
- Department of Pathology, Yale School of Medicine, New Haven, USA
- Rheumatology and Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands
- Department of Viroscience, Erasmus Medical Center, Rotterdam, The Netherlands
- Center for Biomics, Erasmus Medical Center, Rotterdam, The Netherlands

are much more common than $\gamma\delta$ T cells in both human and dog (Mineccia et al. 2012). The β chain (at least in humans) tends to interact more closely with the peptide antigen than the α chain (Glanville et al. 2017), making the TRB locus the most interesting first candidate to annotate in detail.

In this study, we annotate the expressed V, D, J, and C genes in the ferret TRB locus by combining genomic information from the locus with HTS of the ferret TRB repertoire. We find that the TRB locus of the ferret has a similar structure to that of other mammalian TRB loci, such as mouse and human (Glusman et al. 2001), bovine (Connelley et al. 2009), dog (Mineccia et al. 2012), and rabbit (Antonacci et al. 2014): a library of V genes, followed by two (or three in bovine) D-J-C clusters. Each cluster consists of one D gene, six or seven (six in ferret) J genes, and a single C gene. The D-J-C clusters are followed by a V gene with an inverted transcriptional orientation. We also performed a phylogenetic analysis, showing that the ferret V and J genes are indeed most closely related to those of the dog. The ferret locus is small like that of the dog, about 300 Kb, and has a (largely) conserved synteny with the dog TRB locus. Our annotation of the ferret TRB locus will enable detailed studies of T cell responses to support research on novel or improved antiviral strategies for influenza and other viral infections employing the ferret as a model organism. All TRB genes identified in our analysis (expressed or not) have been approved by the IMGT/WHO-IUIS nomenclature committee.

Materials and methods

Genome sequences

The ferret genomic scaffolds (GL896904.1 and GL897291.1) representing the TRB locus were retrieved from Genbank (ferret whole genome shotgun sequence Mus-PutFur 1.0, ref 5) guided by sequence homology with the dog TRB locus (chromosome:CanFam3.1:16:6706526:7027700:1). The blast algorithm (Altschul et al. 1990) and Mauve (Darling et al. 2004) software were applied to align the ferret genome scaffolds with the dog genome sequence.

Animals

Four surplus cryopreserved healthy control blood samples were obtained from an influenza vaccination-challenge study (Bodewes et al. 2010). The control blood samples originated from four 6 to 12 months old healthy outbred female ferrets.

RNA isolation and 5'-RACE

Peripheral blood mononuclear cells (PBMCs) were isolated from ferret blood using a standard Ficoll gradient separation protocol. Subsequently, total RNA was isolated and purified using the RNeasy Mini Kit (Qiagen, Hilden, Germany): 250 μ l of ethanol was added to the upper aqueous phase of the processed TRIzol samples and directly transferred to the RNeasy spin columns for purification. RNA concentrations and OD 260:280 nm ratios were measured with the NanoDrop^R ND-1000 UV-VIS spectrophotometer (NanoDrop Technologies, Wilmington, USA). TCR amplification was performed according to a protocol described by Mamedov et al. (Mamedov et al. 2013). Primer sequences are provided in Table 1. Various nested primer 2 primers were used since local sequence conservation is not yet established. Briefly, RNA obtained from unsorted PBMCs was reverse transcribed by RACE using a primer directed to the constant region. Twelve nucleotide long unique molecular identifiers (UMIs) were incorporated during cDNA synthesis (Kivioja et al. 2011). Subsequently, two-stage seminested and barcoded PCR amplification was performed including a size selection/agarose gel purification step after the first PCR (Mamedov et al. 2013).

TRB transcript sequence analysis

The resulting TCR amplicons were subjected to high-throughput sequencing according to the instructions of the manufacturers using the Ovation Low Complexity Sequencing System kit from NuGEN (San Carlos, CA, USA) and the Illumina MiSeq or the HiSeq2500 platforms both using indexed paired end 300 cycle runs. All sequence reads having the same UMI were collapsed into consensus sequences using the RTCR pipeline (Gerritsen et al. 2016). The BLAST+ (Camacho et al. 2009) and exonerate (Slater and Birney 2005) (version 2.2.0) software were used to align

Table 1 Primer sequences

Template switch primer	AAGCAGTGGTATCAACGCAG AGNNNNNNNNNNCTTGGGGG
FTRB-31	CAGTAGCTGGAGTCATTCT
Step-out primer 1	CACTCTATCCGACAAGCAGTGGTATC AACGCAG
FTRB-1	CACGTGGTCGGGGWAAAAGC
Step-out primer 2	NNNNCACTCTATCCGACAAGCAGT
FTRB-9F	TCTCTGCTTCTGACGGTTCAAAC
FTRB-9F1	GAGGTTTGACCTTTCCAGATTCTC
FTRB-9F2	GGAGTTTGACCTTTCCAGATTC
FTRB-9F3	CTGTGACCGTGGGAGGTTTG
FTRB-9F4	TGTGACCGTGGGAGGTTTGAC

the consensus sequences with the ferret genomic scaffolds to identify V, D, and J genes. After describing the V and J sequences in the ferret TRB locus, the RTCR pipeline was used to annotate the sequences, and perform additional error correction.

Results

Identification of expressed ferret TRB V, D, J, and C genes

Consensus sequences, formed by collapsing all sequence reads sharing the same UMI (here referred to as a “UMI-group”), were aligned against the ferret draft genome. Most consensus sequences (> 99% for primers 9F1-9F4, > 50% for primer 9F) of UMI-groups containing more than 1 sequence read targeted the region between the ferret MOXD2 (Ensembl ID ENSMPUG0000002940.1) and EPHB6 (ENSMPUG0000008478.1) genes. This result is in line with the genome location of the TRB regions of a diverse set of mammals, i.e., cattle, human, mouse, and dog, of which the TRB regions are also flanked by MOXD2 and EPHB6 (Antonacci et al. 2014). Consensus sequences from UMI-groups of size 2 and up were aligned against the ferret genome. Next, we searched in the TRB locus for expressed V, D, J, or C genes in the regions with over 50 coverage (see Fig. 1).

Using the HTS data, we were able to identify 23 ferret TRBV genes that had over 50 x coverage over

part of their region. Reads aligning to the same region were collapsed, and the resulting consensus sequence was used to identify the various features of a V-GENE, such as the location of the V-RS, SPLICE-DONOR, and SPLICE-ACCEPTOR sites. In cases where the consensus was unclear, either due to low coverage and/or poor agreement, we used the LIGMotif programme (Lane et al. 2010) in combination with manual curation to identify the various gene features (Fig. 2). In total, we identified 27 TRBV genes, of which IMGT classified 20 as functional (74%), 4 as pseudogenes (15%), and 3 as an ORF (see Figure S5 for genomic locations of V, D, J, and C genes).

Interestingly, in the functional V gene TRBV19, we discovered a non-functional splice variant (see Figure S1 for an example), leading to a frameshift and the loss of the FR1 and CDR1 regions of the V-EXON. This splice variant occurs on average in more than 25% of the sequence reads that align to TRBV19 in each of the four ferrets. To be able to detect this non-functional variant, sequence reads should extend nearly 200 bp into the V-EXON region. If reads do not extend this far into the V gene, this may lead to an overestimation of the expression level of V genes having such non-functional splice variants. To this end repertoire sequencing experiments should use reads that are long enough to detect the splice variants, or perform a bias correction based on dedicated experiments quantifying the proportion of non-functional splice variants in the ferret population.

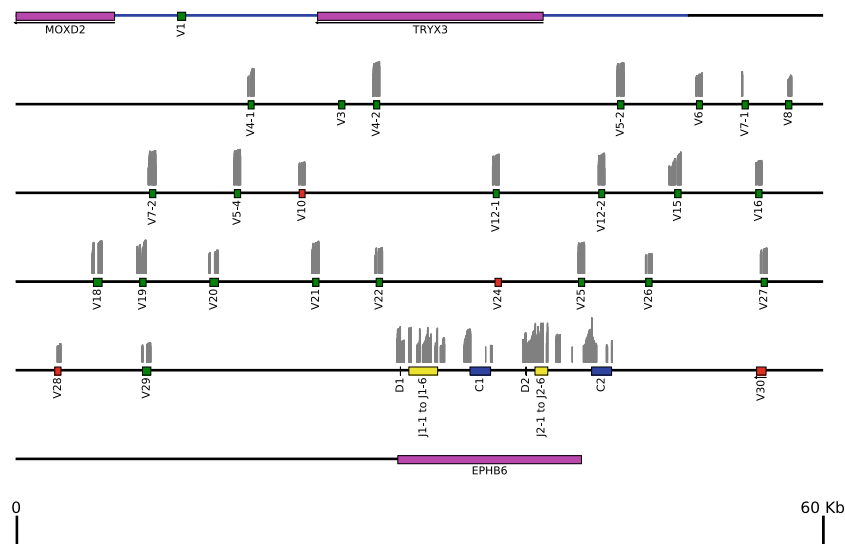


Fig. 1 Schematic representation of the genomic organization of V, D, J, and C genes in the ferret TRB locus. The TRB locus is covered by scaffolds GL896904.1 and GL897291.1, here shown as blue and black horizontal lines, respectively. Boxes are to scale and show the following regions for the various genes: V genes (functional or ORF, green; pseudogenes, red), from the 5' start of L-PART1 to the 3' end of V-REGION; for D genes (black), the D-region; for J genes (yellow), from

5' start of the J-REGION of TRBJ1-1 until 3' end of the J-REGION of TRBJ1-6 and similarly for TRBJ2; for C genes (blue), from 5' start of EX1 until 3' end of EX4. A few non-TRB genes are included: the canonical start and end genes of the TRB locus, MOXD2 and EPHB6, respectively, and TRYX3. Genomic coverage from the HTS (gray bars; bar width is 10 bp) is shown on a log10 scale, excluding bars that have less than 50x coverage on average.

gene name	L-REGION	FR1-IMGT (1-26)		CDR1-IMGT (27-38)			FR2-IMGT (39-55)		CDR2-IMGT (56-65)			FR3-IMGT (66-104)						CDR3-IMGT (105-115)	
		A (1-15)	B (16-26)	BC (27-38)	C (39-46)	C' (47-55)	C" (56-65)	C" (66-74)	D (75-84)	E (85-96)	F (97-104)	FG (105-115)	FG (105-115)						
		1	10 15 16 23 26	27 38	3941 46 47	55	56 65 66	74 75	80 84 85 89	96 97 104	105 111								
TRBV1	MGPPSLVWVWVATMSTGAT	ASLVEORPWWPVAC	GOAKTLHCILR	NSQ.....YWP	MSWYQDLS	QGLSLQVLT	LRN...PGD	EEVVSVP.G	ADYVRRVNS	TELRLHIVAMTQ	G...LYCTC	ST.....	[6.6.2]						
TRBV3	MGSTLFCMVLCLLGAAPL	DTAVSOTPKHLIAT	GNKXFLKCEKQ	LGH.....DT	MYWYQDLS	QKLLKXVFI	YNN...KDL	ILNETVP.R	CFLPESP.DK	TRLRHLHVFLET	GDTAVYFC	ASL.....	[5.6.4]						
TRBV4-1	MGCRLCQVWVLCMLVAVSTGN	.GVTOTPKHLVSGT	GRSVTLKCKQH	LGH.....NA	MYWYQSV	QKPKMLLA	YSY...KDL	LONETAS.S	RFLPDRS.NN	SOLDLOIKALEP	EDSALYLC	ASSK.....	[5.6.4]						
TRBV4-2	MGCRLPWWALCLLGAVSH	DPGVITPKHLVGT	TRSVTLKCKQH	LGH.....NA	MYWYQSV	QKPKMLLA	YNY...KDL	FENATAS.S	RFLPDRS.NN	SOLDLOIKALEP	EDSALYLC	ASSK.....	[5.6.4]						
TRBV5-2	MCSRLCCVALLCLLGGAPVA	.SOVTTPRHLIKTR	GQKVTLCFSV	SGH.....LS	VWYQDVL	GGPRFLIQ	YDQ...RKE	RDKGMP.E	RFSARQ.SN	SSSOLDLDLLE	EDSALYLC	ASSE.....	[5.6.4]						
TRBV5-4	MRRVLPFGSARGTQLCCVSLCLLGGAPVA	GAEVITPRHLIQQS	GGTAVTLCPV	SGH.....NS	VWYQDVL	GGPKFLIE	YFA...QAP	RDKGMP.E	RFSARQ.SN	YSSQLMLLSEP	GDSALYLC	ASSV.....	[5.6.4]						
TRBV6	MSVSRLLCCVFLQAGSL	NAGITQTPKFOVLGT	GNMTLKCHD	MNH.....NS	MYWYQDH	GHLRLHY	SVGPAPLSGA	SASGEVP.D	GYSASRP.NV	TIFLL.L.STAP	SQTSVYFC	ASSV.....	[5.10.4]						
TRBV7-1	MDPRLCQVTLCLLGGQHGAG	.PQVSOFSRHRTVER	GNMVLGCDPT	SGH.....AV	LYWYRIT	GKGFELVY	FQN...ADA	LKSGMTND	RLSARRS.EG	TNLSLQIQRAEQ	GDSAVHLC	ASSP.....	[5.6.4]						
TRBV7-2	MDPRLCQVTLCLLGGQHGAG	.PQVSOFSRHRTVER	GNMVLGCDPT	SGH.....TV	LYWYRIT	GKGFELVY	FQN...TDA	VKSGMTND	RLSARRS.KG	TNLSLQIQRAEQ	GDSAVHLC	ASSP.....	[5.6.4]						
TRBV9	MNLNLFVTLMLLGGAPM	EAEITQTPRYSIIQI	ETKXTECSDD	MNH.....FA	MFWYQDP	GQGLRLYY	SSD...SPS	TTKGDVA.E	GYSVSRK.EQ	KFSPLTESAST	NOTSVYLC	ASSE.....	[5.6.4]						
TRBV12-1	MGTLALCLLGLCVLWFEFT	DAGVITPRLHKVTKM	GOEVLTKCQPI	SGH.....EG	LFWYSQS	VQGPLLLIS	FNN...EAP	IDDSGMPKE	WFSAEIS.NK	SLSLTKIKSTEP	GDSATYLC	ASSI.....	[5.6.4]						
TRBV12-2	MATRLCCVALLCLLGREFT	DAGVITPRLHNVTKM	GOEVLTKCQPI	SGH.....AA	LYWYSQS	VQGPLLLIS	FNN...QAP	IDDSGMPKE	WFSAEIS.NK	SLSLTKIKSTEP	GDSATYLC	ASSV.....	[5.6.4]						
TRBV15	MAPELLHCVAFLLCAGLIM	GAMVIOSSRYVTVR	GKPVNLSGSDN	MNH.....DS	MYWYQRL	NDAKPLLY	YYD...KNF	NREAEIS.D	NFQASRP.NT	SFCSLNLSPGL	GDSVQVLC	ASSK.....	[5.6.4]						
TRBV16	MCVFCSLIVMLLNAGAL	HAEVITQPKHLKQK	GOKARHCVPTI	KGH.....SY	VWYQDVP	AKFKFFIS	FQN...DAI	FDKTGPK	RFSARQ.QN	ASCLEIKPTEL	QDSAVYFC	ASSE.....	[5.6.4]						
TRBV18	MCTRVLCVWVILLGAGLM	NAGVTONPRHLVIRT	GOEATLKCSE	KGH.....YV	MYWYQDP	GESLKFMY	LON...AQT	.DDSGMKE	REFSTNIS.KE	KFSPLTESAST	GDSAVYFC	ASSL.....	[5.7.4]						
TRBV19	MGNVICVALCLLRAGTT	DGGITQTPKYLREE	GREVTLKEQD	FNH.....DY	MYWYQDP	GQGPLRYY	SPL...KDD	VORGDP.E	GYGSRG.KK	TIFSLAVTSTRK	NHTALYLC	ASSR.....	[5.6.4]						
TRBV20	MLLLLLLGPSSGL	GARVSDPSAVICKS	GasVOTECRAV	DLO.....VT	VFGYRQLP	GQGLTLMAT	SNVG...SSA	THEPDFPKA	KPFTSPH.NL	IFFSLTITSASP	ADSSLYLC	GA.....	[6.7.2]						
TRBV21	MFRLLCCVALLFSWAGSL	DPWVITQPKYLKQK	GOKARHCVPTI	KGH.....SY	VFWYRKL	EEFKFLYY	LQD...KDI	VKLEGFDM	QFSAECP.KN	SPLTEINSTEP	GDSALYLC	ASSK.....	[5.6.4]						
TRBV22	MSMALCYVALLCLLGGTGP	THAEYQMPAFLLTR	AGR..AVTLCK	QNL.....RYNA	MYWYQDP	GQSLRLYY	STV...EKD	VORGDI.T	GYWSRE.EK	GLFPLTMLAHT	NOTGVLC	SQSA.....	[7.6.4]						
TRBV25	MVALLCVYVALLLEADLV	DAGITQTPRHCVSGT	GKIVLCESTQ	MCV.....EN	MYWYQDP	KALQALHY	SYG...VNN	TENEELS.S	GTSVRL.RK	EVFPLTESASP	GDSLYLC	ASSE.....	[5.6.4]						
TRBV26	MNRRLLCCVYIILSVGLGK	.AAVTFPKHRLVPG	GKELTLQCLQD	MNY.....VL	MYWYQDP	GFGLQLYY	STG...ARN	FEKGDVPG	YRVSRRD..L	QFSPLTESAST	NOTSVYLC	TSSE.....	[5.6.4]						
TRBV27	MDPRLCQVTLCLLGGQHGAG	EAEVITPRLHFTRK	GROLTYLCSDD	MDH.....DV	MFWYQDP	GGKLLYF	SRN...VKF	IEVGDV.P	GYSVLRK.EK	KDFPLTKSTGT	NOTSVYLC	ASSV.....	[5.6.4]						
TRBV29	MLPCLLLLGGQGVST	.LLSQEPRRDICOC	GTSMTCQCTD	TQV.....SL	MHWYRQL	GQSLTLTAT	ANQG...MEA	TYESGTFKE	KPFTSRP.TL	TFFSLVWMMFS	EDSSFYLC	SAE.....	[5.7.3]						

Fig. 2 Protein display of the ferret consensus TRBV genes, showing only genes classified by IMGT as functional, F, or open reading frame, ORF. Alignment of the V-REGIONS, performed using DomainGapAlign (Ehrenmann et al. 2010; Ehrenmann and Lefranc 2011), is displayed

Similarly to the V genes, we manually curated the D, J, and C genes guided by the HTS data (Figs. 3 and 4). The general structural organization of the ferret TRB locus (see Fig. 1) is similar to that of the dog (see Ref. (Mineccia et al. 2012) for the locus of *Canis lupus familiaris*). Like in the dog, the ferret TRB locus spans about 300 Kb, which is much less than the 650 Kb long locus of humans. The ferret TRB locus contains a region of V genes followed by two D-J-C clusters. Both D-J-C clusters span about 7 Kb and consist of one D gene, six J genes, followed by a C gene. About 11 Kb downstream of

according to IMGT unique numbering for V-REGION (Lefranc et al. 2003), and the amino acid length of CDR-IMGT is indicated in square brackets

the second D-J-C cluster, there is a TRBV gene (TRBV30) with an inverted transcriptional orientation. In the ferret, IMGT classified TRBV30 as a pseudogene because of a missing DONOR-SPLICE. Instead, the homologous TRBV30 in both human and dog is functional.

The ferret TRBD1 and TRBD2 genes are 12 bp and 15 bp long, respectively, and both genes are productively read in all 3 reading frames (Fig. 3a). The ferret J genes are between 44 and 53 bp long and conserve the FGXG motif, required for a functional J gene. The only exception is TRBJ1-4, classified

(a)			(b)				
gene name	5'D-NONAMER	5'D-SPACER	5'D-HEPTAMER	D-REGION	3'D-HEPTAMER	3'D-SPACER	3'D-NONAMER
TRBD1	<u>GGTTTTGT</u> gtttttgt	***** acaagaatgtaa	<u>CACTGTG</u> cattgtg	GGGACAGGGGGC G T G G G Q Q G D R G	<u>CACAGTG</u> cacggtg	***** attcaactctatgggaagctttt	<u>ACAAAAAC</u> acaaaaaacc
TRBD2	catttttgt	atcctgttgtaa	cattgtg	GGAGCTGGGGGGGG G A G G G E L G G S W G G	cacgatg	actcaggtagaggggtgctttt	acaaaaaacc
(b)							
gene name	J-NONAMER	J-SPACER	J-HEPTAMER	J-REGION	DONOR-SPLICE		
TRBJ1-1	<u>GGTTTTGT</u> gtttgttc	***** cctgcccaacat	<u>CACTGTG</u> cactgtg	TGAACACTGAACTTTTCTTTGGAGAAGGCCACCGGCTCACAGTTATAG N T E L F F G E G T R L T V I	gtaaga		
TRBJ1-2	catttgaga	gtggccgatgc	tgatgtg	CTATGATTTTACCTTCGGCCAGGGACCAAGCTGACGGTTGTCG Y D F T E G P G T K L T V V	gtaagg		
TRBJ1-3	ggttctgaa	gtggatctggga	ggctgtg	CTCTTGAGACACCCCTCCATTTTGGGAGGGGACCCGGCTCACTGTTGTAG S * D T L H F G E G T R L T V V	gtaagt		
TRBJ1-4	agttttcct	accgggctgcag	tgttgtg	TAACAATGAAAAACTGTATTTTGGCAAGTGAACCAAGCTGTCCTCGTGG N N E K L Y F A S G T K L S V L	gtaagt		
TRBJ1-5	gggtttgtc	acacctgcctg	tgctgtg	GAGCAACGACGACCAAGCTTTGGAGATGGTACTCAACTTTGGCTCTCG S N Q A Q Q H E F G D G T Q L L V L	gtaagg		
TRBJ1-6	ggttttacc	acggctgcctgc	agctgtg	TTCTATAACTCGCGCTCTACTTTGGGATGGGCACAGGCTCACGCTGACAG S Y N S P L Y F G M G T R L T V T	gtatgg		
TRBJ2-1	aaattcttg	gcagccctgcag	cactgtg	CTCATAACGAGGACAGCACTTTGGCCAGGGACCCGGCTCACAGTACTAG S Y S E Q H F G P G T R L T V L	gtaaga		
TRBJ2-2	ggttgtgtc	ctggctccaccag	agctgtg	CAAAACAGGACAGCTGACTTTGGGGCGGTTCCAAGCTGACGGTTCTGG N T G Q L Y F G A G S K L T V L	gtaagg		
TRBJ2-3	gattcttgt	gctgggctccg	ggcctgt	AGTGAGAAACTCAGTATTTGGCAGGGGACCCGGCTGACGGTGTAG S G E T Q Y F G R G T R L T V L	gtaagc		
TRBJ2-4	tgtttttgt	gctgagccagg	ggctgtg	AGCCAAAATACCCGACTCTTGGCGGGGACCCGGCTGACGGTTCTAG S Q N T Q Y F G A G T R L T V L	gtgagc		
TRBJ2-5	tctctcgga	gaggggcccgg	ggccctg	CTTTGACGCTCTGCCCTGTCTTGGGGCGGGACCCGGCTGACCGTCTGG F A A S A L S F G A G S R L T V L	gtgggt		
TRBJ2-6	ggttgtgac	gcgggtctgggc	ctctgtg	TTCTACGAGCAGTATTTGGCTCCGGCACAGGCTCACGCTCATAG S Y E Q Y F G S G T R L T V I	gtgaga		

Fig. 3 Nucleotide and deduced amino acid sequences of the ferret TRBD (a) and TRBJ (b) genes. TRBJ1-3 is classified as a pseudogene because of the presence of a stop codon (asterisk).

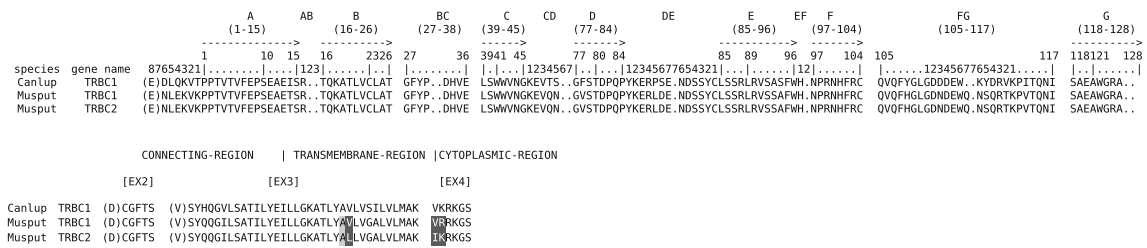


Fig. 4 Protein display of the ferret TRBC genes, including dog TRBC1, according to the IMGT unique numbering for C-DOMAIN (Lefranc et al. 2005). Nucleotide differences between TRBC1 and TRBC2 exons of the

as an ORF because it has a noncanonical J-MOTIF (“FASG”; Fig. 3b) identical to the homologous gene, TRBJ1-4, in the dog. Both TRBJ2-1 and TRBJ2-5 are also classified as an ORF due to having a noncanonical J-NONAMER. Although TRBJ1-3 is a pseudogene (because of a stop codon), it can still lead to functional transcripts by VDJ recombination, as it is used in about 3% of the TCRB clonotypes (Fig. 6b). Like other mammalian species such as human, mouse, dog, and rabbit (Antonacci et al. 2014), the ferret TRBC genes consist of 4 exons each (Fig. 4). The TRBC genes of the ferret are identical to each other for the first 2 exons (EX1 and EX2), and differ by only 2 nt in EX3 and also by 2 nt in EX4. The FG loop is one amino acid longer than the longest TRBC FG loop described by IMGT (Lefranc et al. 2005). We extended the numbering of the FG loop to accommodate the additional amino acid (Fig. 4). Both TRBC genes appear to be functional, having proper acceptor and donor splice sites for each exon, not containing any stop codon or frameshifts.

Phylogenetic analysis of the ferret TRBV and TRBJ genes

We aligned the ferret TRBV (V-REGION) and TRBJ (J-REGION) amino acid sequences to human and dog TRBV and TRBJ sequences, and constructed a phylogenetic tree using a maximum likelihood approach (Guindon and Gascuel 2003) (Fig. 5). We included both functional and non-functional ferret V and J genes. The ferret TRBV genes tend to cluster closer together with the dog than human TRBV genes. Most ferret TRBJ genes are more closely related to the dog TRBJ genes than to the human TRBJ genes, except for TRBJ1-5 and TRBJ2-6, which are closer to human TRBJ1-5 and TRBJ2-7, respectively.

Analysis of the ferret TRBV and TRBJ usage

After identifying the ferret TRBV and TRBJ genes, we used the RTCR pipeline to annotate and error correct the sequence reads. Overall, the TRBV and TRBJ usage is very similar among the four ferrets (Fig. 6). Interestingly, TRBV5-4, which contains a stop codon at the 3’ end of the CDR3, is the most common TRBV gene in the repertoire. Similarly, the

ferret are shown in light gray and dark gray, for synonymous and non-synonymous substitutions, respectively

pseudogene TRBJ1-3 also produces functional transcripts (i.e., about 3% of all transcripts; Fig. 6).

Discussion

We combined HTS and genome analysis to describe the (expressed) T cell receptor genes in the TRB locus of the ferret. The genomic organization of the ferret TRB locus is very similar to that described in other mammals such as human, mouse, dog, and rabbit (Mineccia et al. 2012; Antonacci et al. 2014): the locus is flanked by MOXD2 and EPHB6 at the 5’ and 3’ ends, respectively, and consists of a library of V genes followed by two D-J-C clusters, followed by a V gene, which is non-functional in the ferret, with an inverted transcriptional orientation. Thus, the ferret confirms the strong organizational conservation of mammalian TRB loci. The ferret and dog TRB loci are closely related, because the ferret and the dog are in the same mammalian order, the Carnivora. Like in the dog, the ferret TRB locus is relatively small (300 Kb) and both contain about 20 functional TRBV genes. The ferret also expresses TRBV and TRBJ genes that contain stop codons, which nonetheless lead to functional transcripts, because these stop codons are deleted during VDJ recombination.

The ferret TRB locus is represented by two scaffolds of the draft genome assembly of the ferret as indicated in Fig. 1. Since, one, the highly conserved synteny with the dog TRB locus (any genomic sub-region of the dog TRB locus has a ferret counterpart on either of the two contigs), and two, the facing ends of the two scaffolds (Fig. 1, top line) display considerable sequence homology (> 1400 nt, data not shown), all ferret TRB VDJ genes are most likely contained in the current genome build. It is to be expected that the two scaffolds will be connected in a future build of the ferret genome when additional sequence information is available to complete genome regions encompassing gene families like the TRB locus that are particularly difficult to assemble correctly.

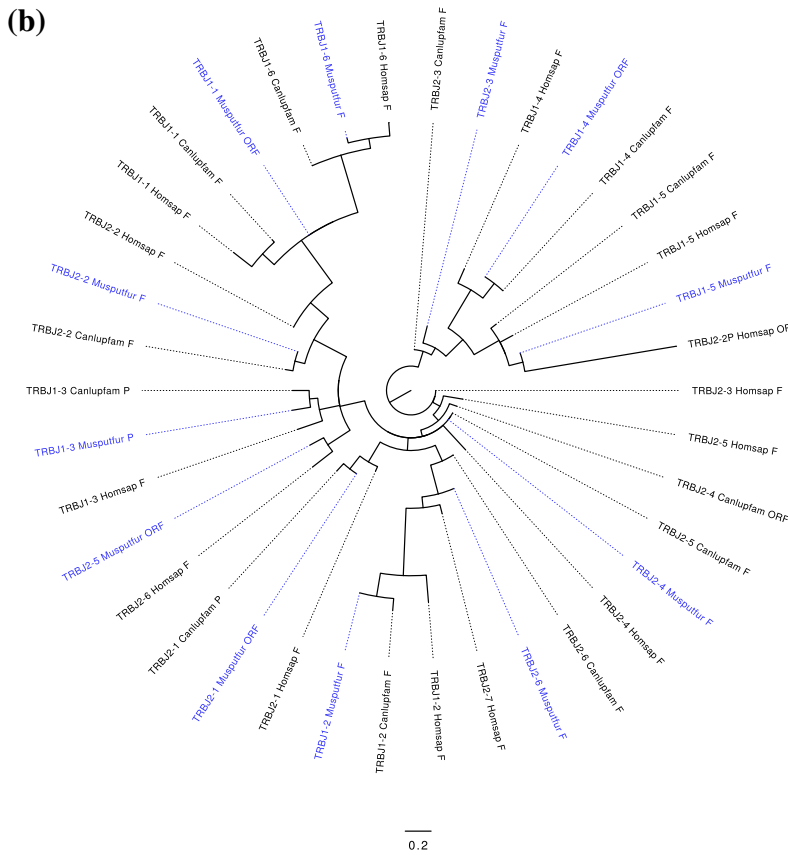
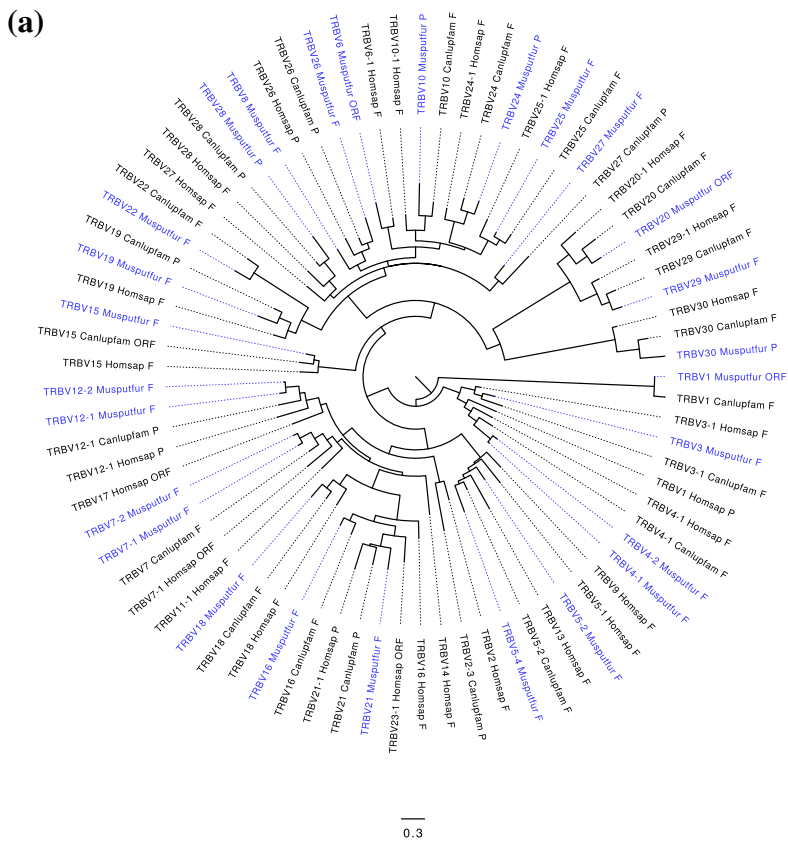


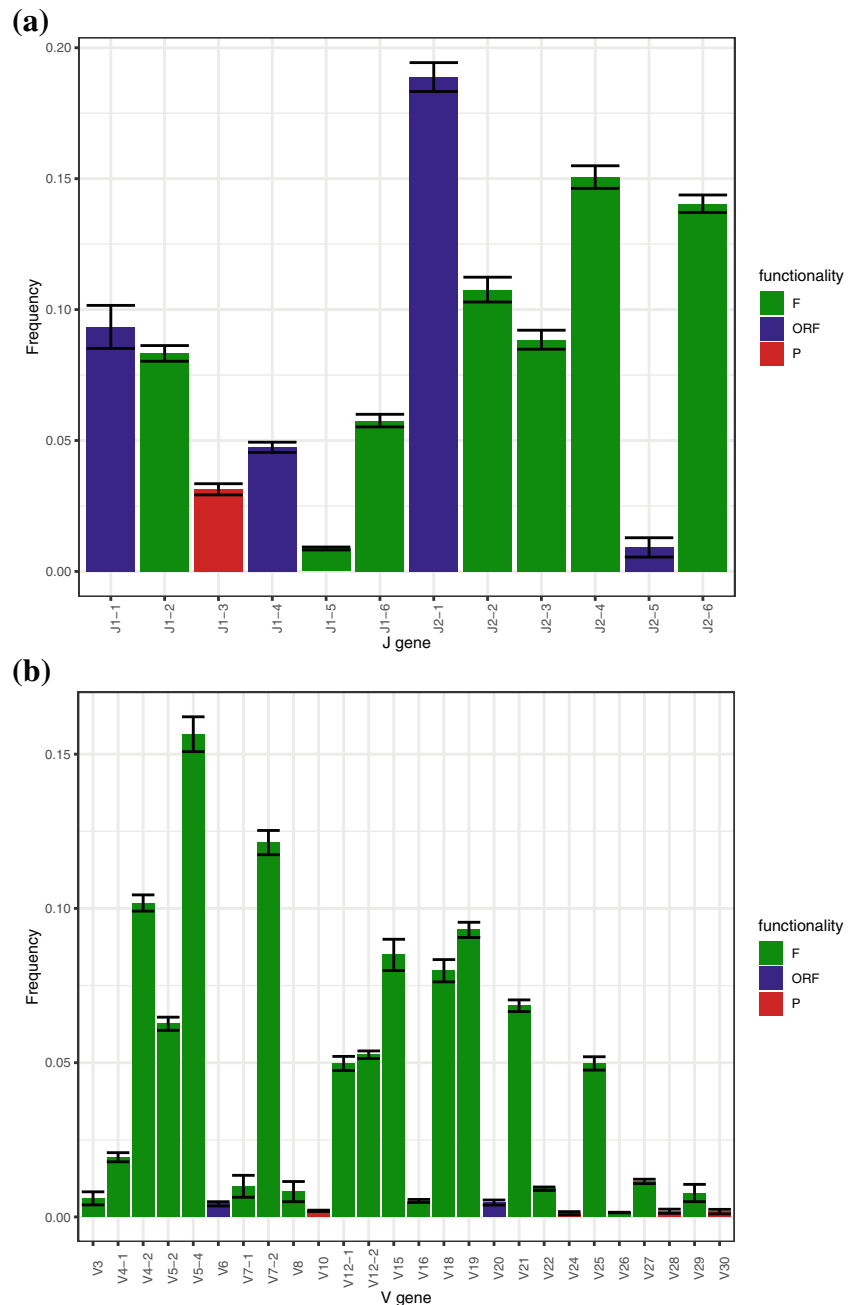
Fig. 5 Phylogenetic trees of the dog, human, and ferret TRBV (a) and TRBJ (b) genes. Unrooted phylogenies were inferred from V-REGION (a) and J-REGION (b) amino acid sequences using the PhyML programme (Guindon and Gascuel 2003). The phylogenetic trees were visualized using FigTree (Rambaut 2009). Ferret genes are indicated in blue, human and dog genes in black

As previously described (Mineccia et al. 2012), the CDR3 length distribution is highly conserved, which in this study is also confirmed as the ferret and the human have nearly identical CDR3 length distributions (Supplemental Figure S2). Despite the relatively low number of V genes in the TRB locus of the ferret, the ferret

repertoire is highly diverse as there is hardly any overlap in TCR β chains between the ferrets (Supplemental Figure S3 and S4). Within ferrets the TCR β chain overlap between samples is about 50%, probably reflecting the presence of memory clonotypes in the blood that have expanded due to antigen stimulation.

Although the ferret is an important animal model in research on respiratory infections, its adaptive immune responses were until now poorly characterized (Enkirch and von Messling 2015). Our characterization of expressed TRB genes in the ferret paves the way for detailed analysis of the cellular immune responses of ferrets in health and disease.

Fig. 6 TRBV (a) and TRBJ (b) gene usage in the ferret. Bars indicate average fraction of distinct TCR sequences encoding a particular gene. Error bars indicate SEM across 16 HTS datasets (4 ferrets total, of which 3 ferrets are represented by 5 repertoire sequencing runs each)



Funding information BG, AA, and FZB were supported by the VIRGO consortium funded by the Netherlands Genomics Initiative and the Dutch Government (FES0908). RdB and AP were supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement 317040 (QuanTI). AP was also supported by the Netherlands Organization for Scientific Research (NWO) VENI grant agreement 016.178.027.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990. ISSN 0022-2836) Basic local alignment search tool. *J Mol Biol* 215(3): 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Antonacci R, Giannico F, Ciccarese S, Mas-sari S (2014. ISSN 1432-1211) Genomic characteristics of the T cell receptor (TRB) locus in the rabbit (*Oryctolagus cuniculus*) revealed by comparative and phylogenetic analyses. *Immunogenetics* 66:255–266. <https://doi.org/10.1007/s00251-013-0754-1>
- Belser JA, Katz JM, Tumpey TM (2011. ISSN 1754-8411) The ferret as a model organism to study influenza a virus infection. *Dis Models Mech* 4:575–579. <https://doi.org/10.1242/dmm.007823>
- Bodewes R, Kreijtz JHCM, van Amerongen G, Geelhoed-Mieras MM, Verburgh RJ, Heldens JGM, Bedwell J, van den Brand JMA, Kuiken T, van Baalen CA, Fouchier RAM, Osterhaus ADME, Rimmelzwaan GF (2010. ISSN 1098-5514) A single immunization with CoVaccine HT-adjuvanted H5N1 influenza virus vaccine induces protective cellular and humoral immune responses in ferrets. *J Virol* 84:7943–7952. <https://doi.org/10.1128/JVI.00549-10>
- Camacho C, Coulouris G, Avagyan V, Ma N, Pa-padopoulos J, Bealer K, Madden TL (2009. ISSN 1471-6) BLAST+: architecture and applications. *BMC Bioinformatics* 10(1):421. <https://doi.org/10.1186/1471-2105-10-421>
- Connelley T, Aerts J, Law A, Morrison WI (2009) Genomic analysis reveals extensive gene duplication within the bovine TRB locus. *BMC genomics* 10(1):192
- Darling ACE, Mau B, Blattner FR, Perna NT (2004. ISSN 1088-9051) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14(7):1394–1403. <https://doi.org/10.1101/gr.2289704>
- Davis MM, Bjorkman PJ (1988) T-cell antigen receptor genes and T-cell recognition. *Nature* 334(6181):395–402
- Ehrenmann F, Lefranc M-P (2011. ISSN 1559-6095) IMGT/DomainGapAlign: IMGT standardized analysis of amino acid sequences of variable, constant, and groove domains (IG, TR, MH, IgSF, MhSF). *Cold Spring Harb Protoc* 2011:737–749. <https://doi.org/10.1101/pdb.prot5636>
- Ehrenmann F, Kaas Q, Lefranc M-P (2010. ISSN 1362-4962) 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* 38:D301–D307. <https://doi.org/10.1093/nar/gkp946>
- Enkirch T, von Messling V (2015. ISSN 1096-0341) Ferret models of viral pathogenesis. *Virology* 479-480:259–270. <https://doi.org/10.1016/j.virol.2015.03.017>
- Gerritsen B, Pandit A, Andeweg AC, de Boer RJ (2016. ISSN 1367-4811) RTCR: a pipeline for complete and accurate recovery of T cell repertoires from high throughput sequencing data. *Bioinformatics (Oxford, England)* 32:3098–3106. <https://doi.org/10.1093/bioinformatics/btw339>
- Glanville J, Huang H, Nau A, Hatton O, Wagar LE, Rubelt F, Ji X, Han A, Krams SM, Pettus C, Haas N, Lindestam Arlehamn CS, Sette A, Boyd SD, Scriba TJ, Martinez OM, Davis MM (2017. ISSN 1476-4687) Identifying specificity groups in the T cell receptor repertoire. *Nature* 547:94–98. <https://doi.org/10.1038/nature22976>
- Glusman G, Lee R, Lee I, Boyesen C, Roach JC, Smit A-i FA, Wang K, Koop BF, Hood L (2001) Comparative genomics of the human and mouse T cell receptor loci. *Immunity* 15(3):337–349
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52: 696–704 ISSN 1063-5157
- Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, Taipale J (2011. ISSN 1548-7105) Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods* 9:72–74. <https://doi.org/10.1038/nmeth.1778>
- Lane J, Duroux P, Lefranc M-P (2010. ISSN 1471-2105) From IMGT-ONTOLOGY to IMGT/LIGMotif: the IMGT standardized approach for immunoglobulin and T cell receptor gene identification and description in large genomic sequences. *BMC bioinformatics* 11:223. <https://doi.org/10.1186/1471-2105-11-223>
- Lefranc M-P, Pommier C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Developmental and comparative immunology* 27:55–77 ISSN 0145-305X
- Lefranc M-P, Pommier C, Kaas Q, Duprat E, Bosc N, Guiraudou D, Jean C, Ruiz M, Da Piédade I, Rouard M, Foulquier E, Thouvenin V, Lefranc G (2005. ISSN 0145-305X) IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev Comp Immunol* 29:185–203. <https://doi.org/10.1016/j.dci.2004.07.003>
- Mamedov IZ, Britanova OV, Zvyagin IV, Turchaninova MA, Bolotin DA, Putintseva EV, Lebedev YB, Chudakov DM (2013. ISSN 1664-3224) Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Frontiers in immunology* 4: 456. <https://doi.org/10.3389/fimmu.2013.00456>
- Mineccia M, Massari S, Linguiti G, Ceci L, Ci-ccarese S, Antonacci R (2012. ISSN 1879-0089) New insight into the genomic structure of dog T cell receptor beta (TRB) locus inferred from expression analysis. *Developmental and comparative immunology* 37:279–293. <https://doi.org/10.1016/j.dci.2012.03.010>
- Oh DY, Hurt AC (2016. ISSN 1664-302X) Using the ferret as an animal model for investigating influenza antiviral effectiveness. *Frontiers in microbiology* 7:80. <https://doi.org/10.3389/fmicb.2016.00080>
- Peng X, Alföldi J, Gori K, Eisfeld AJ, Tyler SR, Tisoncik-Go J-n, Brawand D, Law GL, Skunca N, Hatta M, Gasper DJ, Kelly SM, Chang J, Thomas MJ, John-son J, Berlin AM, Lara M, Russell P, Swofford R, Turner-Maier J, Young S, Hourlier T, Aken B, Searle S, Sun X, Yi Y, Suresh M, Tumpey TM, Siepel A, Wisely SM, Dessimoz C, Kawaoka Y, Birren BW, Lindblad-Toh K, Di Palma F, Engelhardt JF, Palermo RE, Katze MG (2014. ISSN 1546-1696) The draft genome sequence of the ferret (*Mustela putorius furo*) facilitates study of human respiratory disease. *Nature biotechnology* 32:1250–1255. <https://doi.org/10.1038/nbt.3079>
- Andrew Rambaut (2009), FigTree v1.4.4. <<http://tree.bio.ed.ac.uk/software/figtree/>>.
- Slater GSC, Birney E (2005. ISSN 1471-2105) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6(1):31. <https://doi.org/10.1186/1471-2105-6-31>
- van Riel D, Munster VJ, de Wit E, Rimmelzwaan GF, Fouchier RAM, Osterhaus ADME, Kuiken T (2006. ISSN 1095-9203) H5N1 virus attachment to lower respiratory tract. *Science (New York, N.Y.)* 312: 399. <https://doi.org/10.1126/science>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.