Experimental and Clinical Immunogenetics

Exp Clin Immunogenet 2000;17:107–114 Received: October 15, 1999

The Human T Cell Receptor Beta Diversity (TRBD) and Beta Joining (TRBJ) Genes

Géraldine Folch Marie-Paule Lefranc

Laboratoire d'ImmunoGénétique Moléculaire, CNRS, Université Montpellier II, Montpellier, France

Key Words

Human genes · IMGT · T cell receptor · Beta diversity genes · Beta joining genes

Abstract

'Human T cell Receptor Beta Diversity (TRBD) and Beta Joining (TRBJ) Genes', the 10th report of the 'IMGT Locus in Focus' section, comprises 6 tables: (1) 'Human germline TRBD genes'; (2) 'Human TRBD allele table'; (3) 'Nucleotide and protein displays of the human TRBD alleles (overview)'; (4) 'Human germline TRBJ genes'; (5) 'Human TRBJ allele table', and (6) 'Nucleotide and protein displays of the human TRBJ alleles (overview)'. These tables are available on the IMGT Marie-Paule page from **IMGT**, the international ImMunoGeneTics database **(http://imgt.cines.fr:8104)** created in 1989 by Marie-Paule Lefranc, Université Montpellier II, CNRS, France.

Introduction

'Human T cell Receptor Beta Diversity (TRBD) and Beta Joining (TRBJ) Genes' is the 10th report of the 'IMGT Locus in Focus' section launched in the April 1998 issue of *Experimental and Clinical Immunogenetics* [1]. We have previously reported the complete repertoire of the human germline IGH, IGK and IGL genes [2–6], and that of the germline human TRBV [7], TRAV [8] and TRAJ [9] genes. This 10th report on the human T cell receptor beta diversity and beta joining genes comprises 6 tables: (1) 'Human germline TRBD genes'; (2) 'Human TRBD allele table'; (3) 'Nucleotide and protein displays of the human TRBD alleles (overview)'; (4) 'Human germline TRBJ genes'; (5) 'Human TRBJ allele table' and (6) 'Nucleotide and protein displays of the human TRBJ alleles (overview)'. These tables are available on the IMGT Marie-Paule page from **IMGT**, the international ImMunoGeneTics database **(http://imgt.cines.fr:8104)** created in 1989 by Marie-Paule Lefranc, Université Montpellier II, CNRS, France [10, 11]. Descriptions of functionality (functional, open reading frame, pseudogene) and of mutations are according to the IMGT Scientific chart [10, 11], available on the IMGT Marie-Paule page.

KARGER

Fax + 41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0254–9670/00/0172–0107$17.50/0

Accessible online at:
www.karger.com/journals/eci

Prof. Marie-Paule Lefranc, Université Montpellier II
Lab. d'ImmunoGénétique Moléculaire, LIGM, UPR CNRS 1142, IGH
141, rue de la Cardonille, F–34396 Montpellier Cedex 5 (France)
Tel. +33 4 99 61 99 65, Fax +33 4 99 61 99 01
E-Mail lefranc@ligm.igh.cnrs.fr, IMGT:http://imgt.cines.fr.8104

Table 1. Human germline TRBD genes

Fct: FUNCTIONALITY
F: Functional
P: Pseudogene
ORF: Open Reading Frame
vg: Vestigial
R: Rearranged
T: Transcribed
Pr: Translated into protein

"**+**" or "**-**" indicates if the gene sequences have been found (+) or not been found (-) rearranged (R), transcribed (T), and/or translated into protein (Pr).

TRBD name	Fct	R	T	P	Reference sequences	Accession numbers	Sequences from the literature
TRBD1	F	+	+	+	**Dbeta1.1**	X00936 [1]	**Dbeta1.1**[K02545][1], **Dbeta1.1**[M14158][2], **Dbeta1.1**[L36092/U66061][4]
TRBD2	F	+	+	+	**Dbeta2.1**	X02987 [3]	
	F	+	+	+	**Dbeta2.1**	M14159 [2]	**Dbeta2.1**[L36092/U66061][4]

Note:

(1) The original L36092 sequence (684973 bp) has been split in EMBL into three sequences of 267156 bp (U66059), 215422 bp (U66060) and 232650 bp (U66061), L36092 has become secondary accession number of U66059, U66060 and U66061. In IMGT, the original sequence L36092 which is fully annotated has also been kept as primary accession number, in addition to 'U66059, U66060 and U66061.

References:

[1] Clark, S.P. et al., Nature, 311, 387-389 (1984).
[2] Toyonaga, B. et al., Proc. Natl. Acad. Sci. USA., 82, 8624-8628 (1985).
[3] Tunnacliffe, A. et al., Nucleic Acids Res., 13, 6651-6661 (1985).
[4] Rowen, L. et al., Science, 272, 1755-1762 (1996). See note (1)

Author: Géraldine Folch imgt@ligm.igh.cnrs.fr

Created: 23/07/97

Table 2: Human TRBD allele table

Fct: FUNCTIONALITY
F: Functional
P: Pseudogene
ORF: Open Reading Frame
vg: Vestigial

D-REGION alleles are only described at the nucleotide level since D-REGION can be used in the three reading frames.
The numbering starts with the first nucleotide downstream the 5'D-HEPTAMER.
The accession number of a reference sequence is given for each allele.

TRBD name	Fct	TRBD allele name	Accession number	Confirmed by genetics and/or data	Description of mutations
TRBD1	F	TRBD1*01	X00936	+	
TRBD2	F	TRBD2*01	X02987		g13 |
	F	TRBD2*02	M14159	+	g13>a |

Author: Géraldine Folch imgt@ligm.igh.cnrs.fr

Created: 20/01/98

Exp Clin Immunogenet 2000;17:107–114

Table 3: Nucleotide and protein displays of the human TRBD alleles (overview)

Only one sequence for each allele is shown. This set of sequences is part of the IMGT reference directory.
Other known sequences are shown in the individual Alignments of alleles, at http://imgt.cines.fr:8104.
When several alleles are shown, the nucleotide mutations are indicated in bold characters, and the amino acids changes in bold characters and underlined.
These polymorphic mutations are reported in Table 2 (Table of alleles).

		Direct 5'- 3' orientation	Inverted orientation
		G T G G	A P C P
		G Q G	P P V
		D R G	P L S
TRBD1	X00936,TRBD1*01	GGGACAGGGGGC	GCCCCCTGTCCC
		G T S G G	P P R * S
		G L A G G	P P A S P
		D * R G	P P L V
TRBD2	X02987,TRBD2*01	GGGACTAGCGGGGGGG	CCCCCCCGCTAGTCCC
		G T S G **R**	P **S** R * S
		G L A G G	P P A S P
		D * R **E**	**L** P L V
	M14159,TRBD2*02	GGGACTAGCGGG**A**GGG	CCC**T**CCCGCTAGTCCC

Author: Géraldine Folch imgt@ligm.igh.cnrs.fr
Created: 21/06/99

Exp Clin Immunogenet 2000;17:107–114

Table 4. Human germline TRBJ genes

Fct: FUNCTIONALITY **P**: Pseudogene **vg**: Vestigial **T**: Transcribed
F: Functional **ORF**: Open Reading Frame **R**: Rearranged **Pr**: Translated into protein

"**+**" or "**-**" indicates if the gene sequences have been found (+) or not been found (-) rearranged (R), transcribed (T), and/or translated into protein (Pr).
Arbitrarily that information is shown on the first line of each gene when the data have been confirmed by several studies.
Reference sequences in bold have been mapped: "mapped" refers to sequences which have been obtained from clones (phages, cosmids, YACs...) either by subcloning or PCR, and does not apply to sequences obtained directly from genomic DNA.
The original L36092 sequence (684973 bp) has been split in EMBL into three sequences of 267156 bp (U66059), 215422 bp (U66060) and 232650 bp (U66061), L36092 has become secondary accession number of U66059, U66060 and U66061.
In IMGT, the original sequence L36092 which is fully annotated has also been kept as primary accession number, in addition to U66059, U66060 and U66061.

TRBJ name	Fct	R	T	P	Reference sequences	Accession numbers	Sequences from the literature
TRBJ1-1	F	+	+	+	**Jbeta1.1**	X00936 [1]	**Jbeta1.1**[K02545][1], **Jbeta1.1**[M14158][3], **Jbeta1.1**[L36092/U66061][5]
TRBJ1-2	F	+	+	+	**Jbeta1.2**	X00936 [1]	**Jbeta1.2**[K02545][1], **Jbeta1.2**[M14158][3], **Jbeta1.2**[L36092/U66061][5]
TRBJ1-3	F	+	+	+	**Jbeta1.3**	M14158 [3]	**Jbeta1.3**[L36092/U66061][5]
TRBJ1-4	F	+	+	+	**Jbeta1.4**	M14158 [3]	**Jbeta1.4**[L36092/U66061][5]
TRBJ1-5	F	+	+	+	**Jbeta1.5**	M14158 [3]	**Jbeta1.5**[L36092/U66061][5]
TRBJ1-6	F	+	+	+	**Jbeta1.6**	M14158 [3]	
	F				**Jbeta1.6**	L36092/U66061 [5]	
TRBJ2-1	F	+	+	+	**Jbeta2.1**	X02987 [2]	**Jbeta2.1**[M14159][3], **Jbeta2.1**[M31346][4],**Jbeta2.1**[L36092/U66061][5]
TRBJ2-2	F	+	+	+	**Jbeta2.2**	X02987 [2]	**Jbeta2.2**[M14159][3], **Jbeta2.2**[L36092/U66061][5]
TRBJ2-2P	ORF(1)				**Jbeta2P**	X02987 [2]	**Jbeta2.2P**[M14159][3], **Jbeta2.2P**[L36092/U66061][5]
TRBJ2-3	F	+	+	+	**Jbeta2.3**	X02987 [2]	**Jbeta2.3**[M14159][3], **Jbeta2.3**[L36092/U66061][5]
TRBJ2-4	F	+	+	+	**Jbeta2.4**	X02987 [2]	**Jbeta2.4**[M14159][3], **Jbeta2.4**[L36092/U66061][5]
TRBJ2-5	F	+	+	+	**Jbeta2.5**	X02987 [2]	**Jbeta2.5**[M14159][3], **Jbeta2.5**[L36092/U66061][5]
TRBJ2-6	F	+	+	+	**Jbeta2.6**	X02987 [2]	**Jbeta2.6**[M14159][3], **Jbeta2.6**[L36092/U66061][5]
TRBJ2-7	F	+	+	+	**Jbeta2.7**	M14159 [3]	**Jbeta2.7**[L36092/U66061][5]
	ORF(2)				**Jbeta2.7**	X02987 [2]	

IMGT notes:

(1) This sequence has diverged considerably from that of other TRBJ2-2, and it has lost J-NONAMER [2], J-PHE replaced by Leu in J-SEGMENT.
(2) J-PHE replaced by Val in J-SEGMENT.

Author: Géraldine Folch imgt@ligm.igh.cnrs.fr
Created: 23/07/97

References:

[1] Clark, S.P. et al., Nature, 311, 387-389 (1984).
[2] Tunnacliffe, A. et al., Nucleic. Acids. Research, 13, 6651-6661 (1985).
[3] Toyonaga, B. et al., Proc. Natl. Acad. Sci. USA., 82, 8624-8628 (1985).
[4] Ikuta, K. et al., Nucleic Acids Res., 14, 4899-4909 (1986).
[5] Rowen, L. et al., Science, 272, 1755-1762 (1996).

Table 5. Human TRBJ allele table

Fct: FUNCTIONALITY
F: Functional
P: Pseudogene
ORF: Open Reading Frame
vg: Vestigial

IMGT numbering and description of alleles for germline J-REGIONs start with the first nucleotide of the first codon.
The accession number of a reference sequence is given for each allele.
The original L36092 sequence (684973 bp) has been split in EMBL into three sequences of 267156 bp (U66059), 215422 bp (U66060) and 232650 bp (U66061), L36092 has become secondary accession number of U66059, U66060 and U66061.
In IMGT, the original sequence L36092 which is fully annotated has also been kept as primary accession number, in addition to U66059, U66060 and U66061.

TRBJ name	Fct	TRBJ allele name	Accession number	Confirmed by genetics and/or data	Description of mutations
TRBJ1-1	F	TRBJ1-1*01	X00936	+	
TRBJ1-2	F	TRBJ1-2*01	X00936	+	
TRBJ1-3	F	TRBJ1-3*01	M14158	+	
TRBJ1-4	F	TRBJ1-4*01	M14158	+	
TRBJ1-5	F	TRBJ1-5*01	M14158	+	
TRBJ1-6	F	TRBJ1-6*01	M14158		t30 |
	F	TRBJ1-6*02	L36092/U66061	+	t30>c|
TRBJ2-1	F	TRBJ2-1*01	X02987	+	
TRBJ2-2	F	TRBJ2-2*01	X02987	+	
TRBJ2-2P	ORF	TRBJ2-2P*01	X02987	+	
TRBJ2-3	F	TRBJ2-3*01	X02987	+	
TRBJ2-4	F	TRBJ2-4*01	X02987	+	
TRBJ2-5	F	TRBJ2-5*01	X02987	+	
TRBJ2-6	F	TRBJ2-6*01	X02987	+	
TRBJ2-7	F	TRBJ2-7*01	M14159	+	t16 ,F6 |
	ORF	TRBJ2-7*02	X02987		t16>g,F6>V|

Author: Géraldine Folch imgt@ligm.igh.cnrs.fr
Created: 22/02/99

Table 6: Nucleotide and Protein displays of the human TRBJ alleles

Only one sequence for each allele is shown. This set of sequences is part of the IMGT reference directory.
Other known sequences are shown in the individual Alignments of alleles, at http://imgt.cines.fr:8104.
When several alleles are shown, the nucleotide mutations are indicated in bold characters, and the amino acid changes in bold characters and underlined.
These polymorphic mutations are reported in Table 5 (Table of alleles).
Dashes indicate identical nucleotides.

```
                                    N   T   E   A   F   F   G   Q   G   T   R   L   T   V   V
  X00936  ,TRBJ1-1*01             TG AAC ACT GAA GCT TTC TTT GGA CAA GGC ACC AGA CTC ACA GTT GTA G

                                    N   Y   G   Y   T   F   G   S   G   T   R   L   T   V   V
  X00936  ,TRBJ1-2*01             CT AAC TAT GGC TAC ACC TTC GGT TCG GGG ACC AGG TTA ACC GTT GTA G

                                    S   G   N   T   I   Y   F   G   E   G   S   W   L   T   V   V
  M14158  ,TRBJ1-3*01              C TCT GGA AAC ACC ATA TAT TTT GGA GAG GGA AGT TGG CTC ACT GTT GTA G

                                    T   N   E   K   L   F   F   G   S   G   T   Q   L   S   V   L
  M14158  ,TRBJ1-4*01             GA ACT AAT GAA AAA CTG TTT TTT GGC AGT GGA ACC CAG CTC TCT GTC TTG G

                                    S   N   Q   P   Q   H   F   G   D   G   T   R   L   S   I   L
  M14158  ,TRBJ1-5*01              T AGC AAT CAG CCC CAG CAT TTT GGT GAT GGG ACT CGA CTC TCC ATC CTA G

                                  S   Y   N   S   P   L   H   F   G   N   G   T   R   L   T   V   T
  M14158  ,TRBJ1-6*01           C TCC TAT AAT TCA CCC CTC CAC TTT GGG AAT GGG ACC AGG CTC ACT GTG ACA G

(1)L36092,U66061,TRBJ1-6*02     - --- --- --- --- --- --- --- --- --- ---C --- --- --- --- --- --- --- -

                                    S   Y   N   E   Q   F   F   G   P   G   T   R   L   T   V   L
  X02987  ,TRBJ2-1*01              C TCC TAC AAT GAG CAG TTC TTC GGG CCA GGG ACA CGG CTC ACC GTG CTA G

                                    N   T   G   E   L   F   F   G   E   G   S   R   L   T   V   L
  X02987  ,TRBJ2-2*01             CG AAC ACC GGG GAG CTG TTT TTT GGA GAA GGC TCT AGG CTG ACC GTA CTG G

                                  L   R   G   A   A   G   R   L   G   G   G   L   L   V   L
  X02987  ,TRBJ2-2P*01           CTG AGA GGC GCT GCT GGG CGT CTG GGC GGA GGA CTC CTG GTT CTG G
```

```
                                    S   T   D   T   Q   Y   F   G   P   G   T   R   L   T   V   L
  X02987   ,TRBJ2-3*01             AGC ACA GAT ACG CAG TAT TTT GGC CCA GGC ACC CGG CTG ACA GTG CTC G

                                    A   K   N   I   Q   Y   F   G   A   G   T   R   L   S   V   L
  X02987   ,TRBJ2-4*01           A GCC AAA AAC ATT CAG TAC TTC GGC GCC GGG ACC CGG CTC TCA GTG CTG G

                                        Q   E   T   Q   Y   F   G   P   G   T   R   L   L   V   L
  X02987   ,TRBJ2-5*01             AC CAA GAG ACC CAG TAC TTC GGG CCA GGC ACG CGG CTC CTG GTG CTC G

                                  S   G   A   N   V   L   T   F   G   A   G   S   R   L   T   V   L
  X02987   ,TRBJ2-6*01         C TCT GGG GCC AAC GTC CTG ACT TTC GGG GCC GGC AGC AGG CTG ACC GTG CTG G

                                     S   Y   E   Q   Y   F   G   P   G   T   R   L   T   V   T
  M14159   ,TRBJ2-7*01              C TCC TAC GAG CAG TAC TTC GGG CCG GGC ACC AGG CTC ACG GTC ACA G
                                                         V
  X02987   ,TRBJ2-7*02              - --- --- --- --- --- G-- --- --- --- --- --- --- --- --- --- -
```

IMGT note:

(1) The original L36092 sequence (684973 bp) has been split in EMBL into three sequences of 267156 bp (U66059), 215422 bp (U66060) and 232650 bp (U66061), L36092 has become secondary accession number of U66059, U66060 and U66061.
In IMGT, the original sequence L36092 which is fully annotated has also been kept as primary accession number, in addition to U66059, U66060 and U66061.

Author: Géraldine Folch imgt@ligm.igh.cnrs.fr
Created: 04/08/99

Acknowledgments

We thank Valérie Contet for editorial work, Véronique Giudicelli and Gérard Lefranc for helpful discussion. IMGT is funded by the European Union's BIOTECH programme (BIO4CT96-0037), Centre National de la Recherche Scientifique, and Ministère de l'Education Nationale, de la Recherche et de la Technologie. Subventions have been received from Association pour la Recherche sur le Cancer, Association de la Recherche sur la Polyarthrite, Fondation pour la Recherche Médicale, Ligue Nationale contre le Cancer and Région Languedoc-Roussillon.

References

1 Lefranc MP: IMGT Locus on focus. Exp Clin Immunogenet 1998;15: 1–7.
2 Pallarès N, Frippiat JP, Giudicelli V, Lefranc MP: The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. Exp Clin Immunogenet 1998;15:8–18.
3 Barbié V, Lefranc MP: The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. Exp Clin Immunogenet 1998;15:171–183.
4 Pallarès N, Lefebvre S, Contet V, Matsuda F, Lefranc MP: The human immunoglobulin heavy variable genes. Exp Clin Immunogenet 1999;16:36–60.
5 Ruiz M, Pallarès N, Contet V, Barbié V, Lefranc MP: The human immunoglobulin heavy diversity (IGHD) and joining (IGHJ) segments. Exp Clin Immunogenet 1999;16:173–184.
6 Scaviner D, Barbié V, Ruiz M, Lefranc MP: Protein displays of the human immunoglobulin heavy, kappa and lambda variable and joining regions. Exp Clin Immunogenet 1999;16:234–240.
7 Folch G, Lefranc MP: The human T cell receptor beta variable (TRBV) genes. Exp Clin Immunogenet 2000;17:42–54.
8 Scaviner D, Lefranc MP: The human T cell receptor alpha variable (TRAV) genes. Exp Clin Immunogenet 2000;17:83–96.
9 Scaviner D, Lefranc MP: The human T cell receptor alpha joining (TRAJ) genes. Exp Clin Immunogenet, in press.
10 Lefranc MP, Giudicelli V, Ginestoux C, Bodmer J, Müller W, Bontrop R, Lemaitre M, Malik A, Barbié V, Chaume D: IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 1999;27: 209–212.
11 Ruiz M, Giudicelli V, Ginestoux C, Stoehr P, Robinson J, Bodmer J, Marsh S, Bontrop R, Lemaitre M, Lefranc G, Chaume D, Lefranc MP: IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 2000;28:219–221.