R4 regulators of G protein signaling (RGS) identify a conserved genomic region that contains MHC–related markers

Jaanus Suurväli[1], Jacques Robert[2], Pierre Boudinot[3] and Sirje Rüütel Boudinot[1]

[1]Department of Gene Technology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia

[2] Department of Microbiology and Immunology, University of Rochester Medical Center, Rochester, NY 14642

[3] INRA, Molecular Virology and Immunology, Domaine de Vilvert, 78350 Jouy en Josas, France

Corresponding author:

Jaanus Suurväli

Tallinn University of Technology

Department of Gene Technology

Akadeemia tee 15, 12618 Tallinn, Estonia

phone: +372 620 4447

fax: +372 620 4401

e-mail: jaanus.suurvali@ttu.ee

Online Resource 2.

This file contains the phylogenetic evidence supporting paralogy relationships between the *Homo sapiens* RGS1/RGS16 region gene ohnologs shown on Figure 2.

Online Resource 2.01 – Gives an overview of ohnologs for the genes in RGS1/RGS16 region. Includes paralogs in genomic locations termed by (Flajnik et al. 2012) as „minor paralogons“ (Flajnik MF, Tlapakova T, Criscitiello MF, Krylov V, Ohta Y (2012) Evolution of the B7 family: co-evolution of B7H6 and NKp30, identification of a new B7 family member, B7H7, and of B7's historical relationship with the MHC. Immunogenetics 64:571-90)

Online Resource 2.02 – 2.33

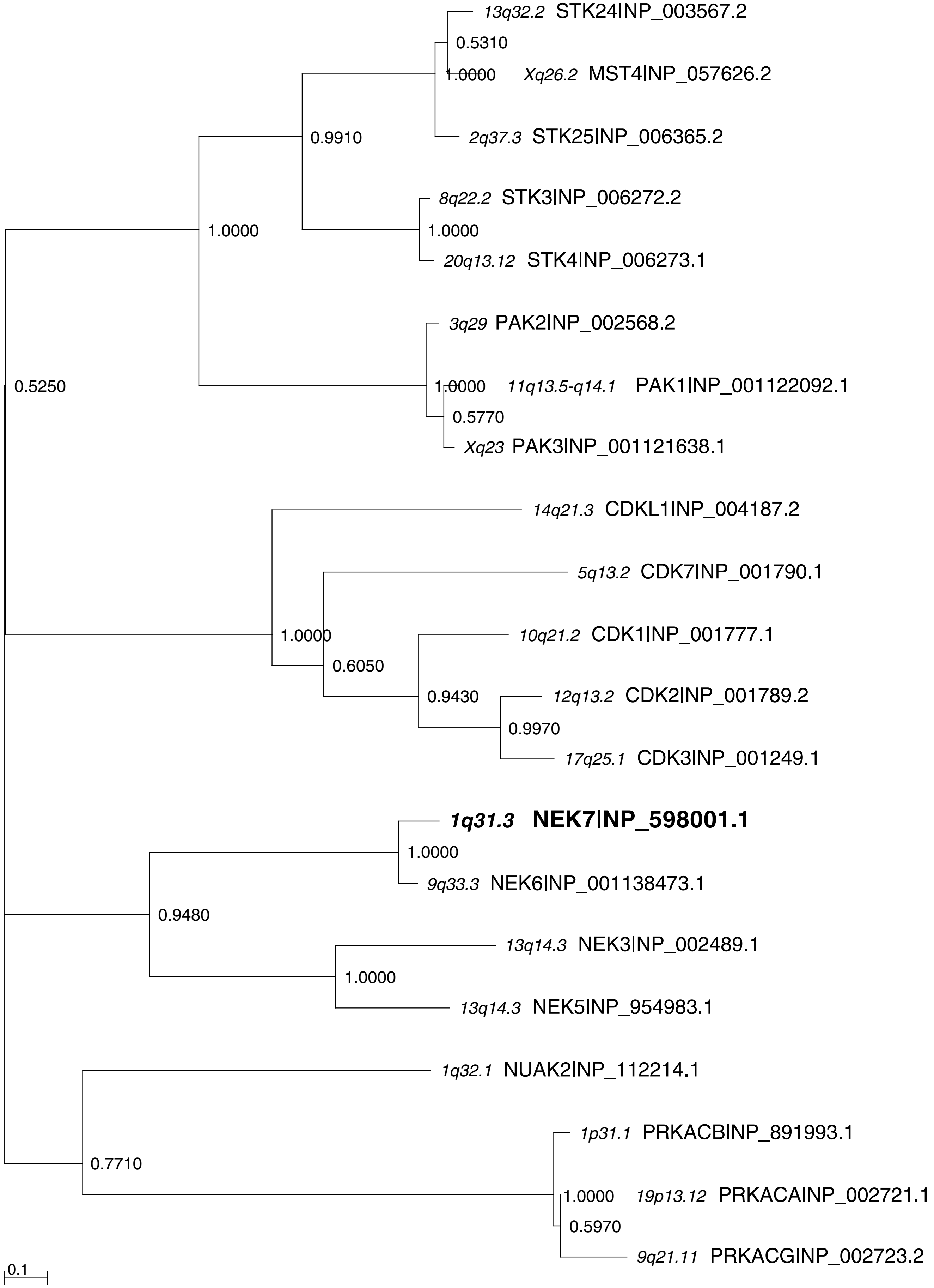
Phylogenetic evidence for both Figure 2 and the table in Online Resource 2.01. All phylogenys are Maximum Likelihood trees calculated with 1000 bootstrap replicates as described in Materials and Methods.

Online Resource 2.01. Genes of the human RGS1/RGS16 region have many paralogs on the „Major" and „Minor" MHC paralogous regions, as defined by (Flajnik et al. 2012).

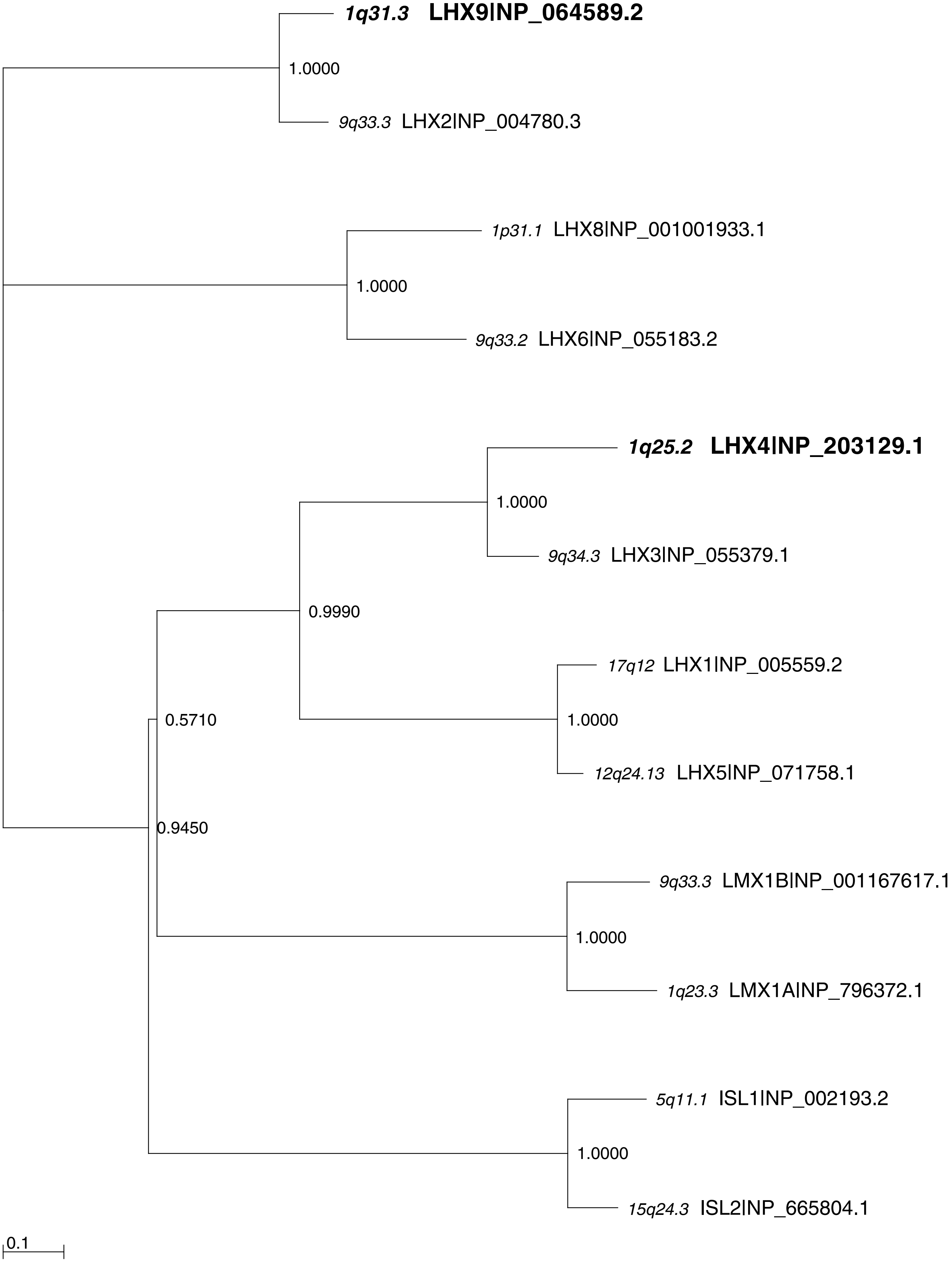
RGS1/RGS16 region on Chromosome 1		Paralogs on 6p21-p22 and 15q13-q26	Paralogs on 9p13-p24, 9q32- q34 and 5q11-q23		Paralogs on 19p13.1-p13.3		Paralogs on „Minor" MHC paralogous regions (Flajnik et al. 2012)		Phylogeny
NEK7	1q31.3		NEK6	9q33.3					Online Resource 2.02
LHX9	1q31.3		LHX2	9q33.3					Online Resource 2.03
C1orf53	1q31.3								No similar sequences found
DENND1B	1q31.3		DENND1A	9q33.3	DENND1C	19p13.3			Online Resource 2.04
CRB1	1q31.3		CRB2	9q33.3					Online Resource 2.05
ZBTB41	1q31.3								No similar sequences found
ASPM	1q31.3								No similar sequences found
F13B	1q31.3								Online Resource 2.06
CFHR5	1q31.3								Online Resource 2.06
CFHR2	1q31.3								Online Resource 2.06
CFHR4	1q31.3								Online Resource 2.06
CFHR1	1q31.3								Online Resource 2.06
CFHR3	1q31.3								Online Resource 2.06
CFH	1q31.3								Online Resource 2.06
KCNT2	1q31.3		KCNT1	9q34.3					Online Resource 2.07
B3GALT2	1q31.3								Online Resource 2.08
CDC73	1q31.2								No similar sequences found
GLRX2	1q31.2		GLRX	5q15			TXNRD3	3q21.3	Online Resource 2.09
							TXNRD1	12q23.3	
TROVE2	1q31.2								No similar sequences found
UCHL5	1q31.2								Online Resource 2.10
RGS2	1q31.2		RGS3	9q32					Online Resource 2.11
RGS13	1q31.2		RGS3	9q32					Online Resource 2.11
RGS1	1q31.2		RGS3	9q32					Online Resource 2.11
RGS21	1q31.2		RGS3	9q32					Online Resource 2.11
RGS18	1q31.2		RGS3	9q32					Online Resource 2.11
FAM5C	1q31.1		DBC1	9q33.1					Online Resource 2.12

RGS1/RGS16 region on Chromosome 1	Paralogs on 6p21-p22 and 15q13-q26	Paralogs on 9p13-p24, 9q32- q34 and 5q11-q23	Paralogs on 19p13.1-p13.3	Paralogs on „Minor“ MHC paralogous regions (Flajnik et al. 2012)	Phylogeny
PLA2G4A 1q31.1	PLA2G4F 15q15.1			PLA2G4C 19q13.33	Online Resource 2.13
	PLA2G4D 15q15.1				
	PLA2G4E 15q15.1				
	PLA2G4B 15q15.1				
PTGS2 1q31.1		PTGS1 9q33.2			Online Resource 2.14
PDC 1q31.1		PDCL 9q33.2			Online Resource 2.15
OCLM 1q31.1					No similar sequences found
C1orf27 1q31.1					No similar sequences found
TPR 1q31.1					Online Resource 2.16
PRG4 1q31.1					Online Resource 2.17
HMCN1 1q25.3		HMCN2 9q34.11			Only HMCN2 was found similar
IVNS1ABP 1q25.3					Online Resource 2.18
SWT1 1q25.3					No similar sequences found
TRMT1L 1q25.3			TRMT1 19p13.2		Only TRMT1 was found similar
RNF2 1q25.3	RING1 6p21.32				Only RING1 was found similar
FAM129A 1q25.3		FAM129B 9q34.11	FAM129C 19p13.11		Online Resource 2.19
EDEM3 1q25.3					Online Resource 2.20
C1orf21 1q25.3					No similar sequences found
TSEN15 1q25.3					No similar sequences found
GLT25D2 1q25.3		CERCAM 9q34.11	GLT25D1 19p13.11		Online Resource 2.21
APOBEC4 1q25.3				APOBEC1 12p13.31	Online Resource 2.22
				AICDA 12p13.31	
RGL1 1q25.3	RGL2 6p21.32	RALGDS 9q34.2	RGL3 19p13.2		Online Resource 2.23
ARPC5 1q25.3		ARPC5L 9q33.3			Only ARPC5L was found similar
NCF2 1q25.3		NOXA1 9q34.3			Only NOXA1 was found similar
SMG7 1q25.3					Online Resource 2.24
NMNAT2 1q25.3					Online Resource 2.25
LAMC2 1q25.3		LAMC3 9q34.12			Online Resource 2.26
LAMC1 1q25.3		LAMC3 9q34.12			Online Resource 2.26
SHCBP1L 1q25.3					Only SHCBP1 was found similar

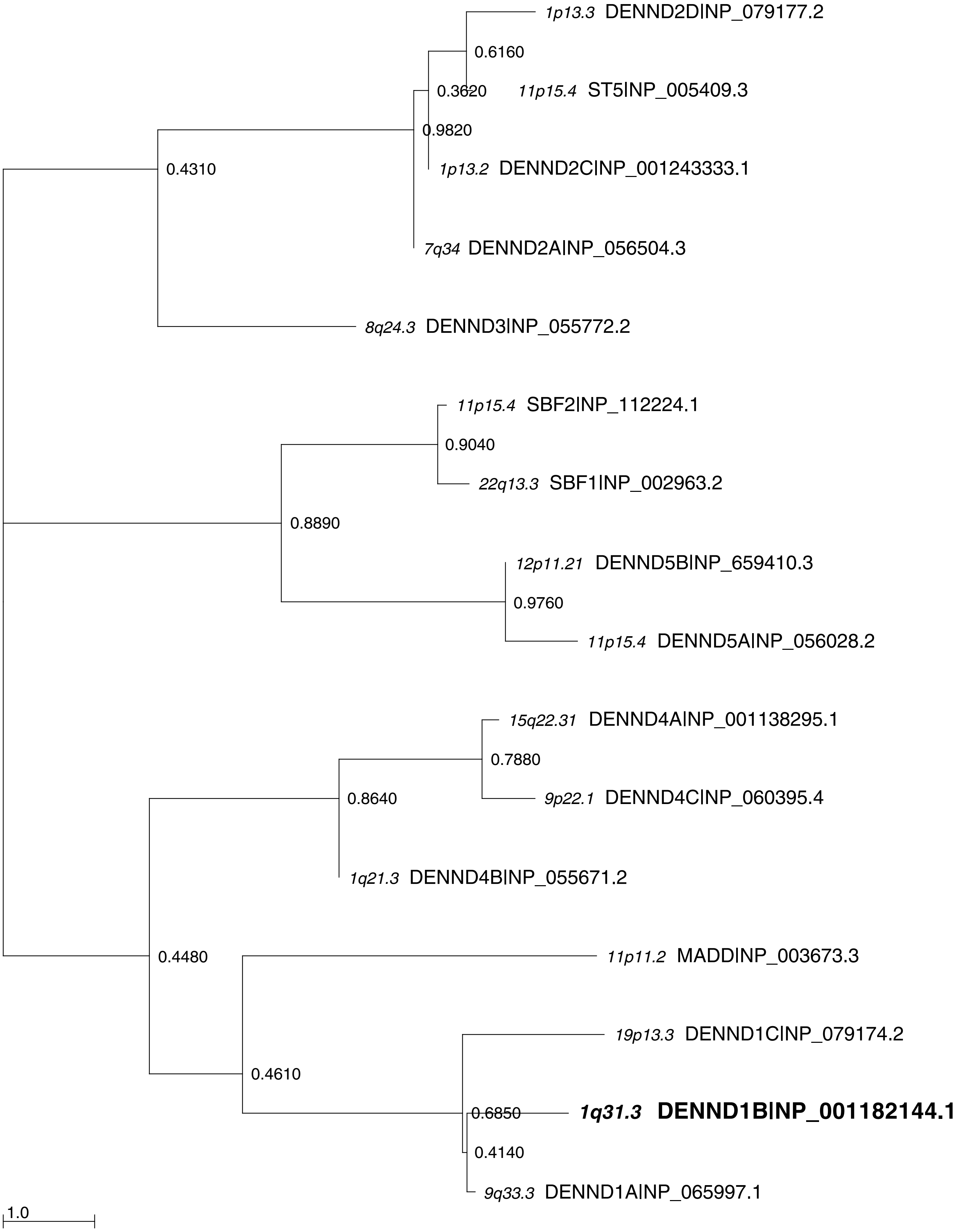
RGS1/RGS16 region on Chromosome 1	Paralogs on 6p21-p22 and 15q13-q26	Paralogs on 9p13-p24, 9q32- q34 and 5q11-q23	Paralogs on 19p13.1-p13.3	Paralogs on „Minor“ MHC paralogous regions (Flajnik et al. 2012)	Phylogeny
DHX9 1q25.3		YTHDC2 5q22.2			Online Resource 2.27
NPL 1q25.3					Only HOGA1 was found similar
RGS8 1q25.3		RGS3 9q32			Online Resource 2.11
RGS16 1q25.3		RGS3 9q32			Online Resource 2.11
RNASEL 1q25.3					No similar sequences found
RGSL1 1q25.3					Online Resource 2.11
TEDDM1 1q25.3				TMEM45A 3q12.2	Online Resource 2.28
				TMEM45B 11q24.3	
GLUL 1q25.3					Only LGSN was found similar
ZNF648 1q25.3					Reliable phylogeny could not be made
CACNA1E 1q25.3		CACNA1B 9q34.3	CACNA1A 19p13.2		Online Resource 2.29
IER5 1q25.3		IER5L 9q34.11	IER2 19p13.2		Online Resource 2.30
MR1 1q25.3	HLA-B 6p21.33			FCGRT 19q13.33	Online Resource 2.31
	HLA-C 6p21.33				
	HLA-E 6p21.33				
	HLA-A 6p22.1				
	HLA-F 6p22.1				
	HLA-G 6p22.1				
	HFE 6p22.2				
STX6 1q25.3			STX10 19p13.2		Online Resource 2.32
AL162431.1 1q25.3					No similar sequences found
KIAA1614 1q25.3					No similar sequences found
XPR1 1q25.3					No similar sequences found
ACBD6 1q25.3					No similar sequences found
LHX4 1q25.2		LHX3 9q34.3			Online Resource 2.03
QSOX1 1q25.2		QSOX2 9q34.3			Only QSOX2 was found similar
CEP350 1q25.2					No similar sequences found
TOR1AIP1 1q25.2					Only TOR1AIP2 was found similar
AL359853.2 1q25.2					No similar sequences found
TOR1AIP2 1q25.2					Only TOR1AIP1 was found similar
FAM163A 1q25.2		FAM163B 9q34.2			Online Resource 2.33



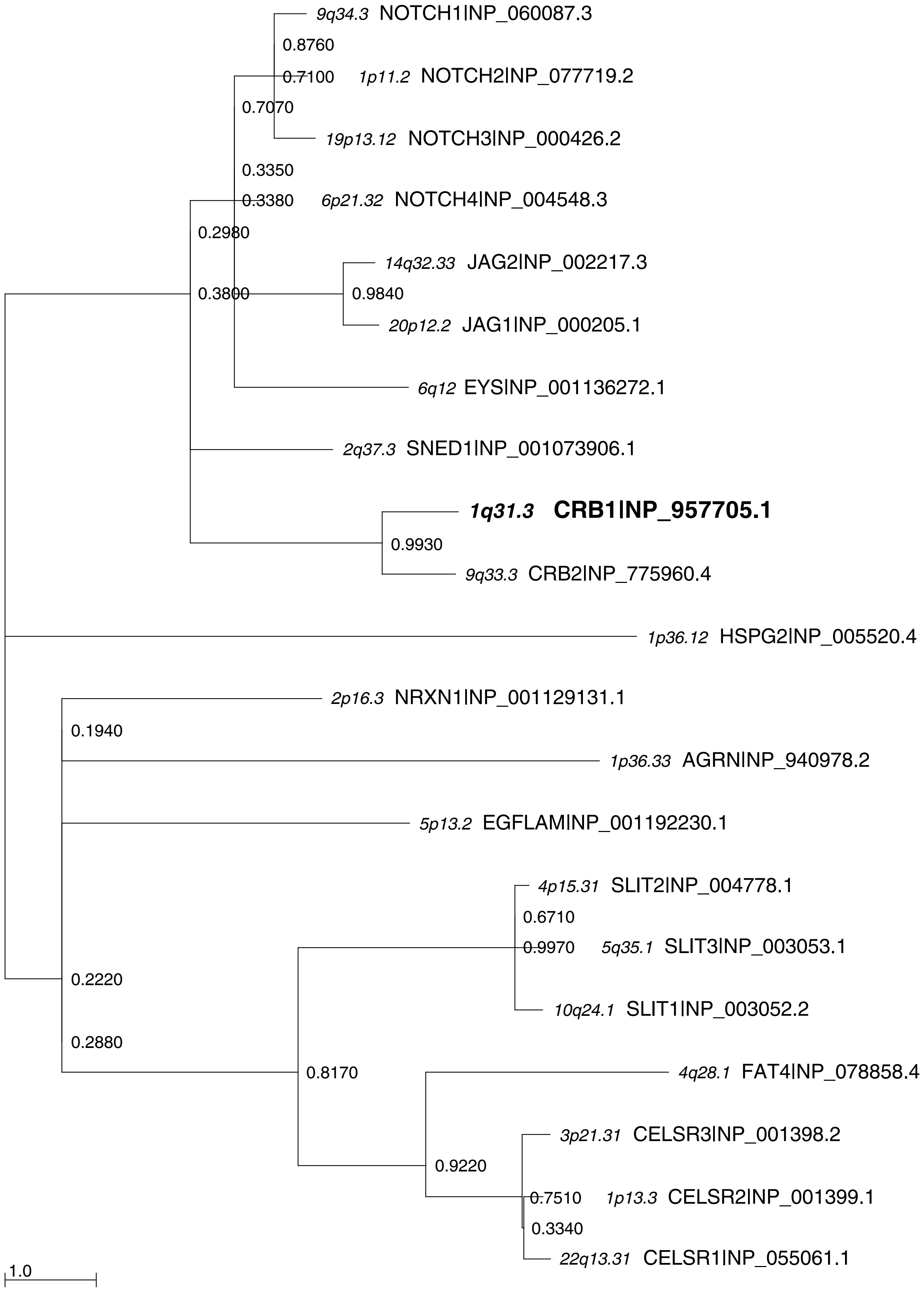
Online Resource 2.02. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein NEK7. This tree was constructed with 1000 bootstrap replicates from the protein sequences of NEK7 and 20 best hits obtained with the first iteration of DELTA BLAST against NCBI human reference sequences.



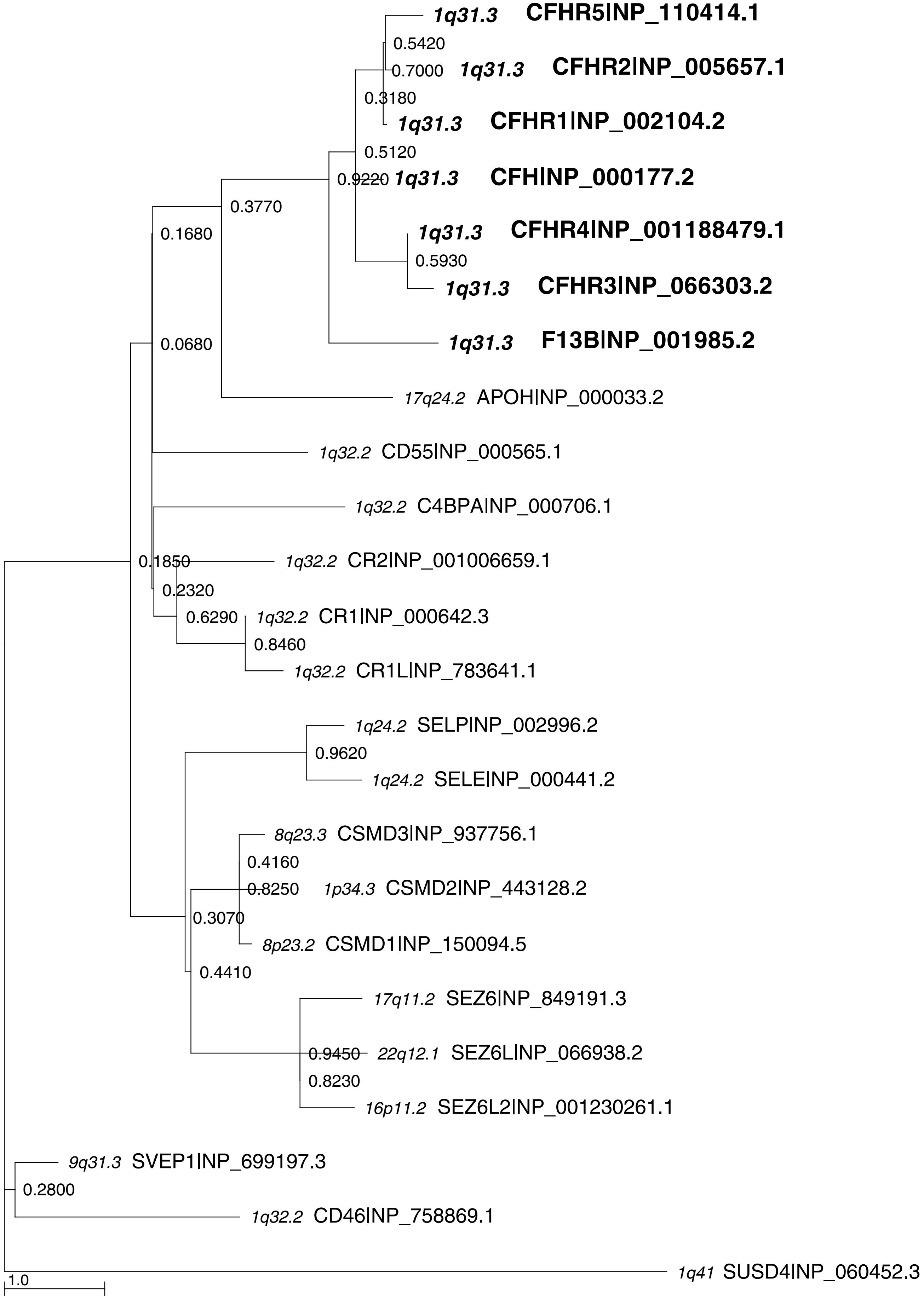
Online Resource 2.03. Maximum Likelihood tree showing the paralogues of *Homo sapiens* proteins LHX4 and LHX9. This tree was constructed with 1000 bootstrap replicates from the protein sequences of LHX4, LHX9 and all 10 results of DELTA BLAST's first iteration that had a similar domain structure.



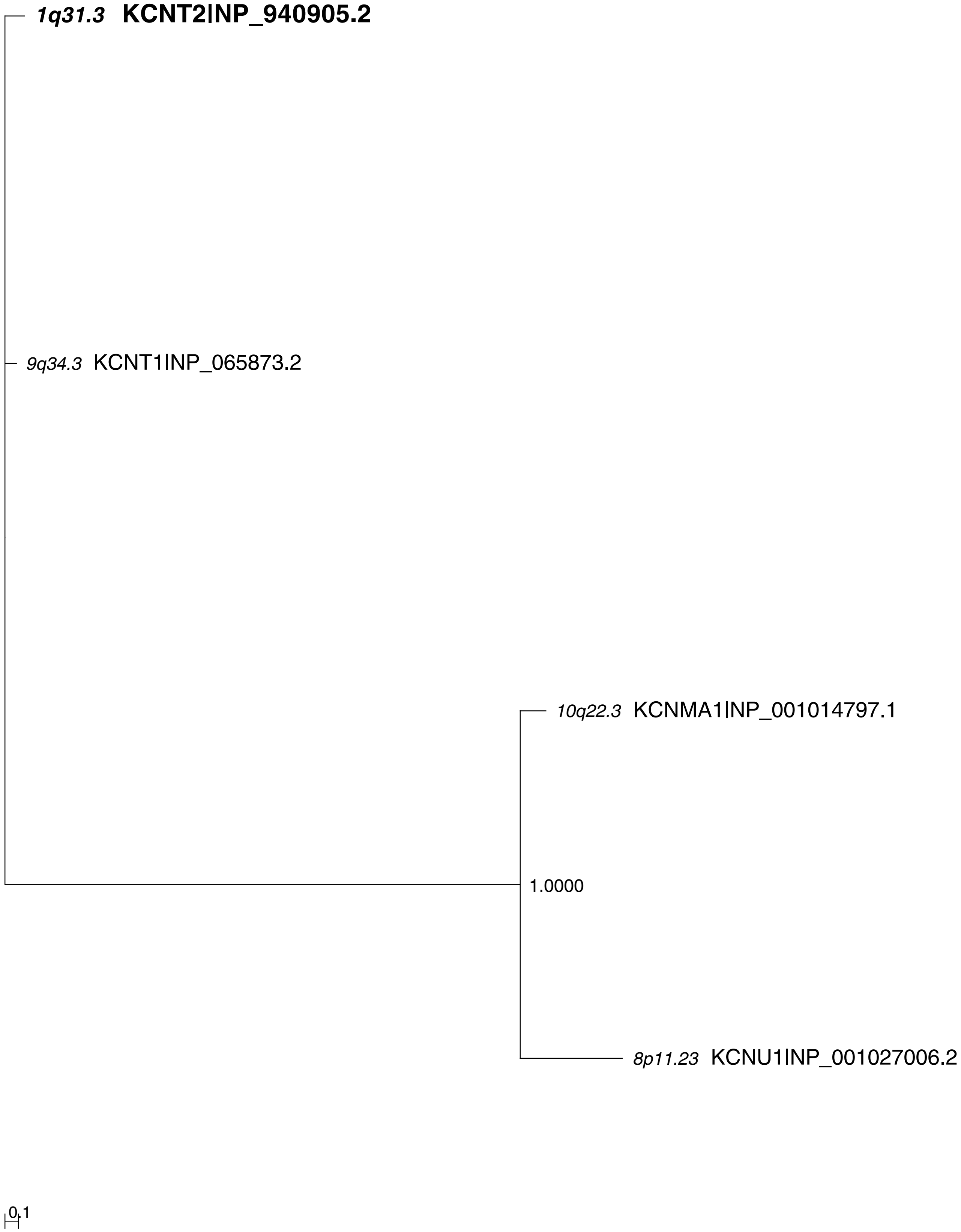
Online Resource 2.04. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein DENND1B. This tree was constructed with 1000 bootstrap replicates from the protein sequences of DENND1B and 15 results of DELTA BLAST's first iteration that had a similar domain structure.



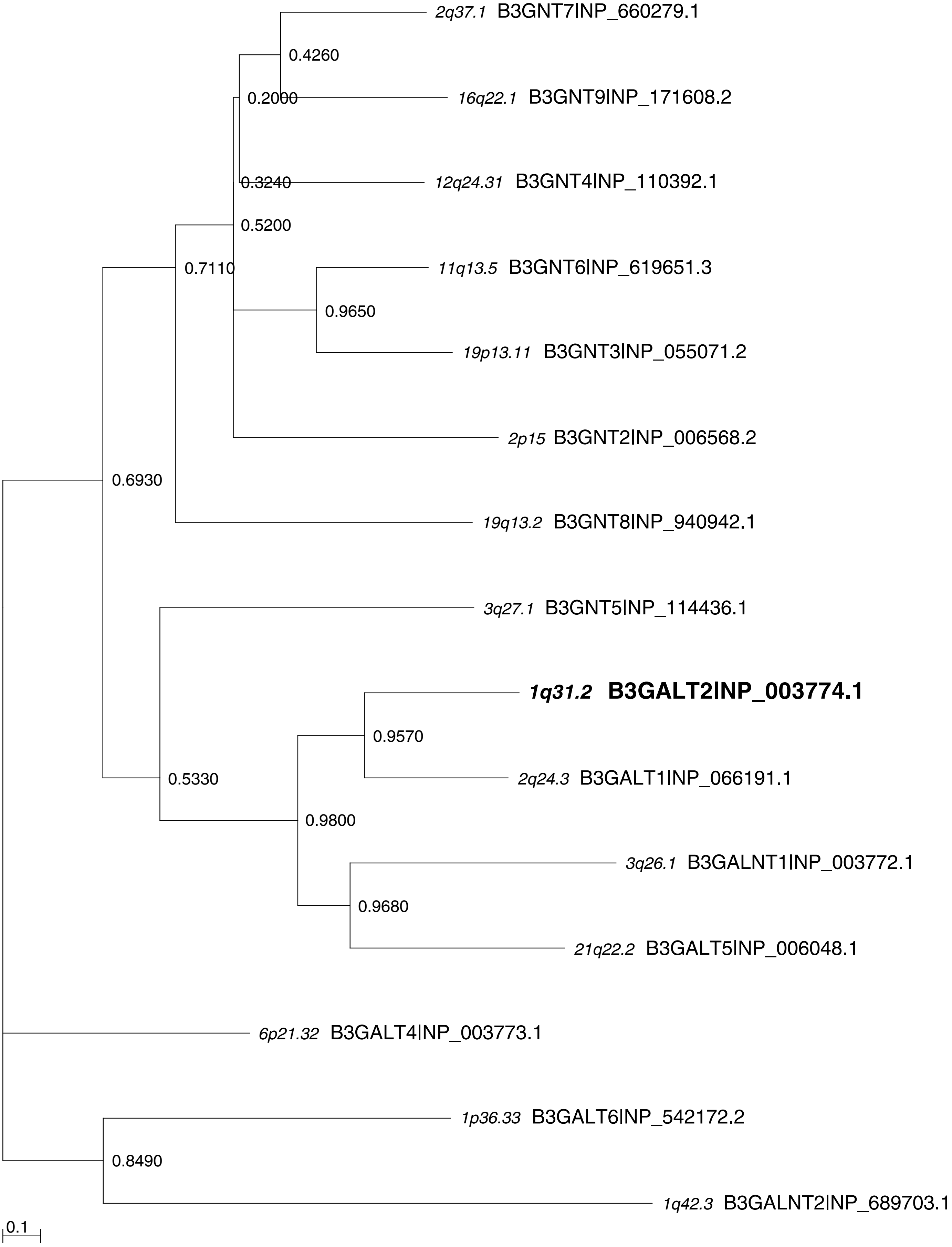
Online Resource 2.05. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein CRB1. This tree was constructed with 1000 bootstrap replicates from the protein sequences of CRB1 and 20 best hits obtained with the first iteration of DELTA BLAST against NCBI human reference sequences.



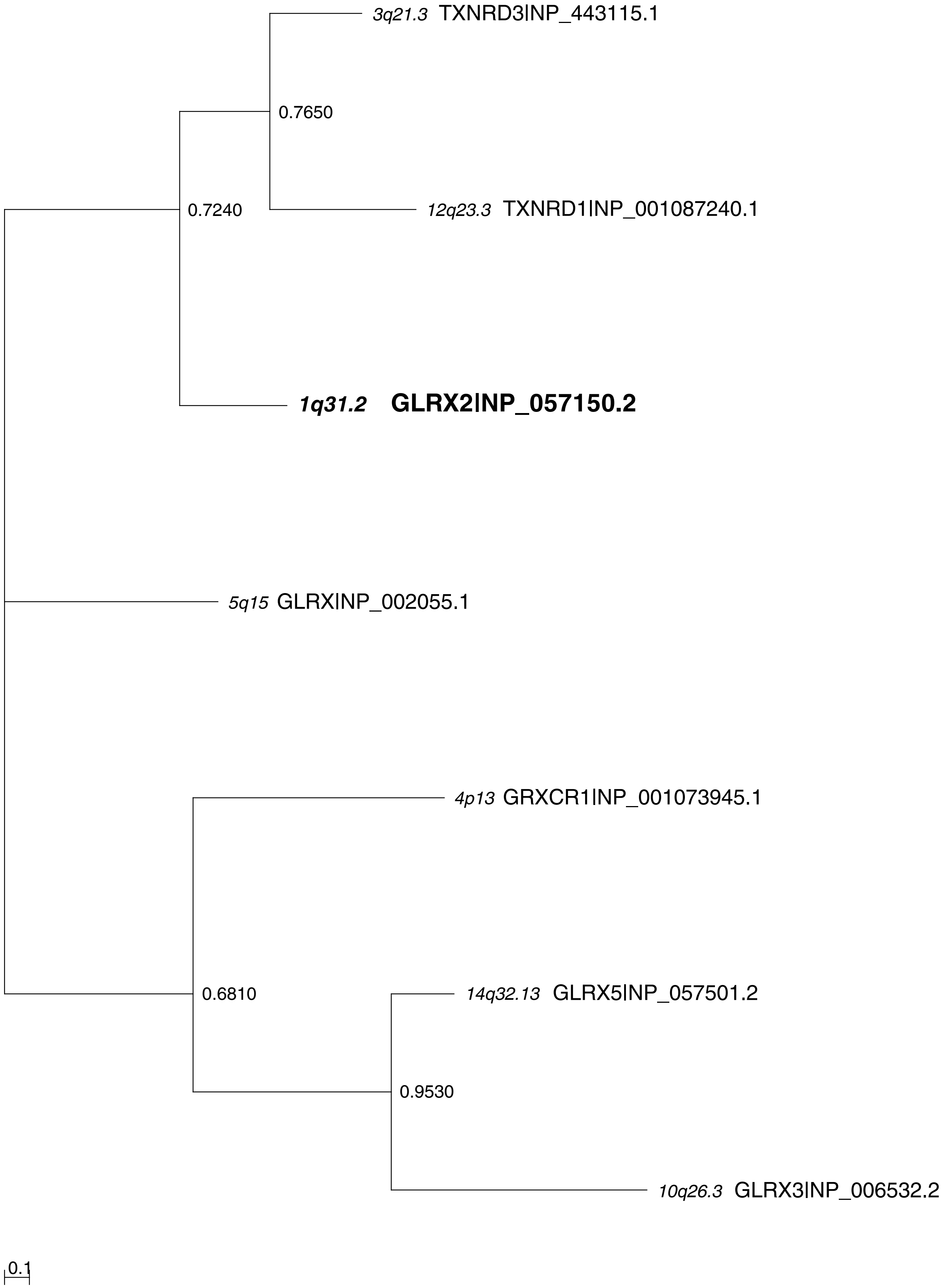
Online Resource 2.06. Maximum Likelihood tree showing the paralogues of *Homo sapiens* proteins F13B, CFH, CFHR1, CFHR2, CFHR3, CFHR4 and CFHR5. This tree was constructed with 1000 bootstrap replicates from the corresponding protein sequences and 15 additional hits obtained with the first iteration of DELTA BLAST for all of these 7 proteins.



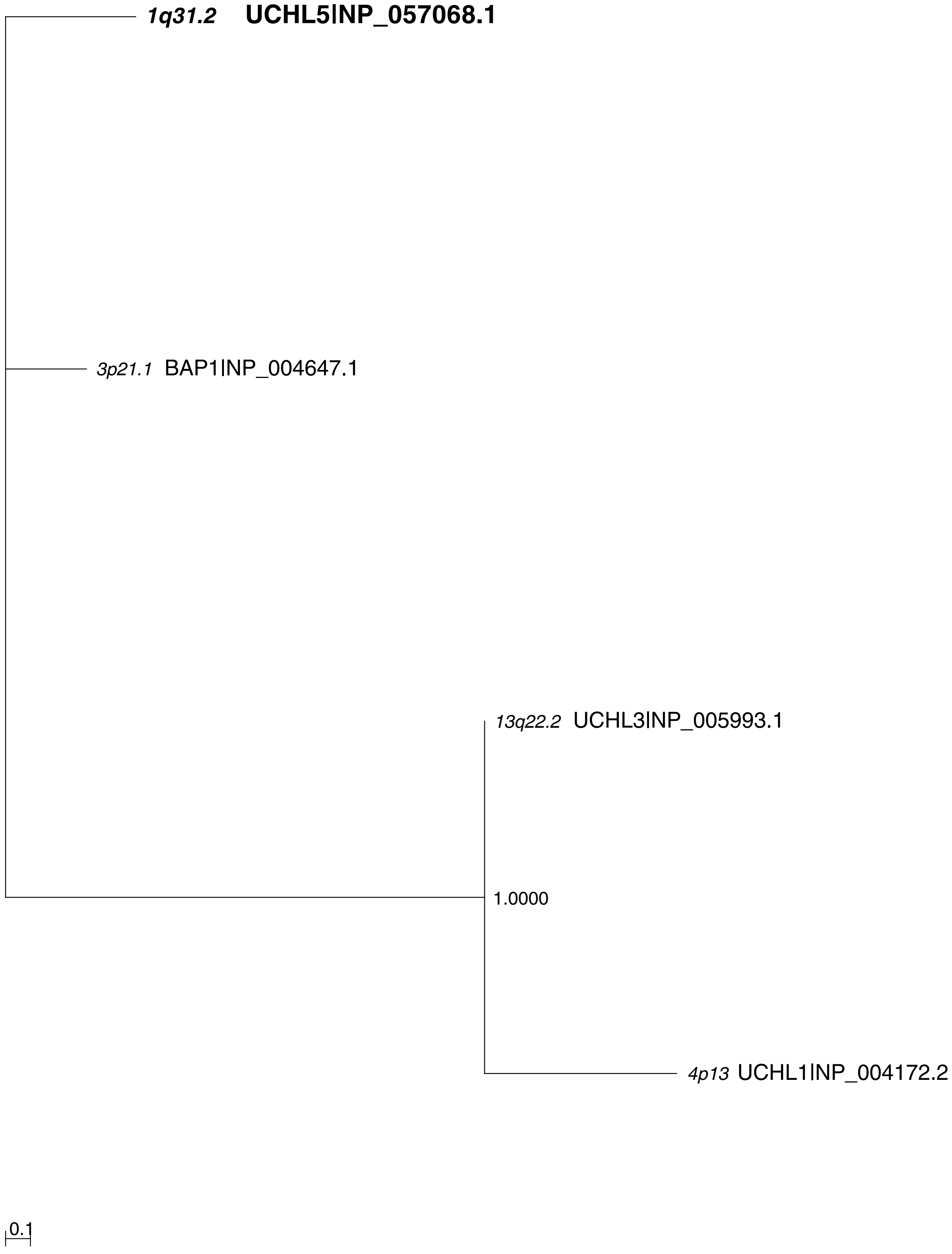
Online Resource 2.07. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein KCNT2. This tree was constructed with 1000 bootstrap replicates from the protein sequences of KCNT2 and all 3 hits of DELTA BLAST's first iteration that had a similar domain structure.



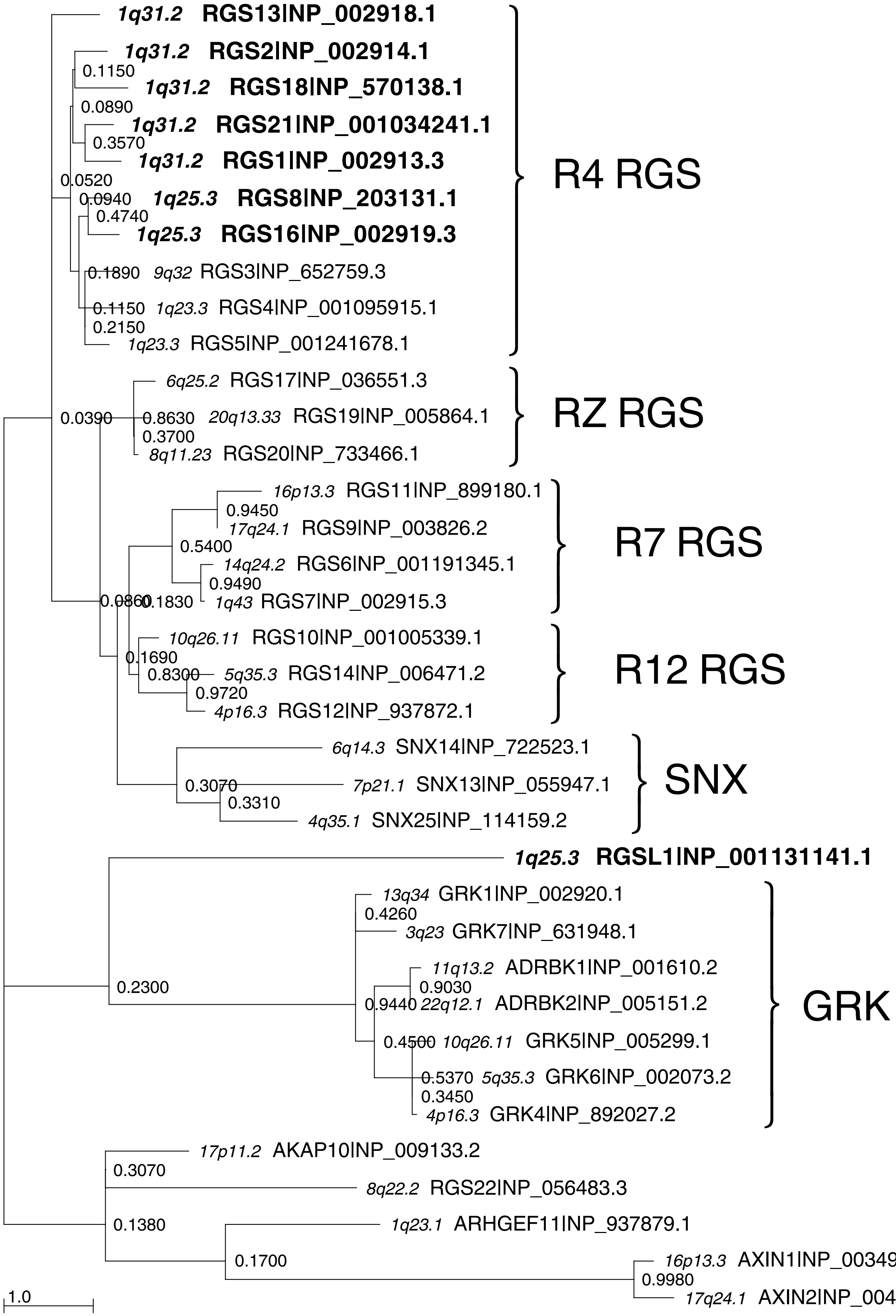
Online Resource 2.08. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein B3GALT2. This tree was constructed with 1000 bootstrap replicates from the protein sequences of B3GALT2 and all the hits of DELTA BLAST's first iteration with e values lower than 1e-20; all of which also have a galactosyltransferase domain.



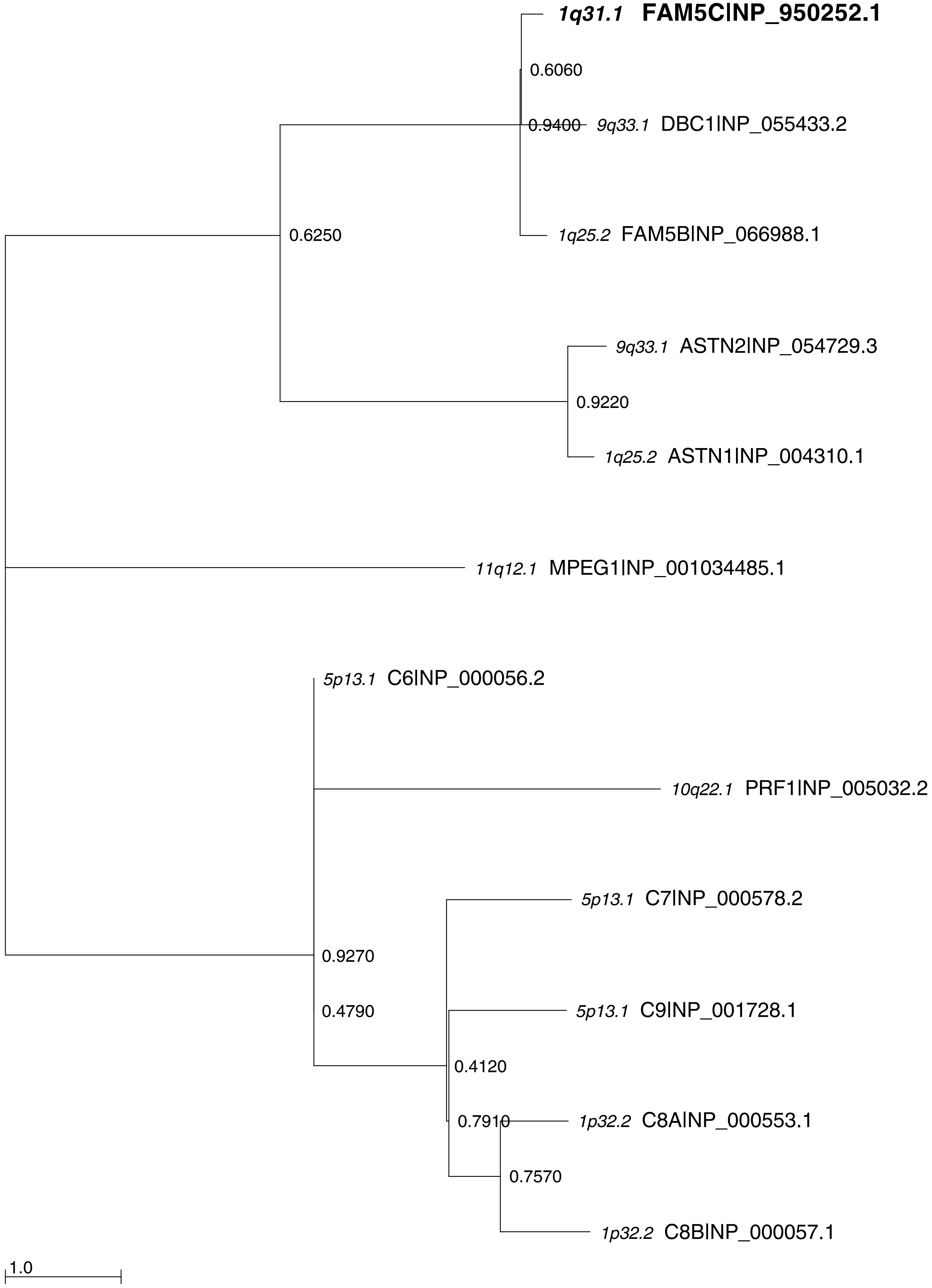
Online Resource 2.09. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein GLRX2. This tree was constructed with 1000 bootstrap replicates from the protein sequences of GLRX2 and DELTA BLAST hits with the lowest e values. The figure suggest GLRX not to be the closest paralogue of GLRX2. This is in contradiction with Ensembl Release 68, where it is listed as the only paralogue. For the purpose of the current study, it is considered one of the closer paralogues of GLRX2.



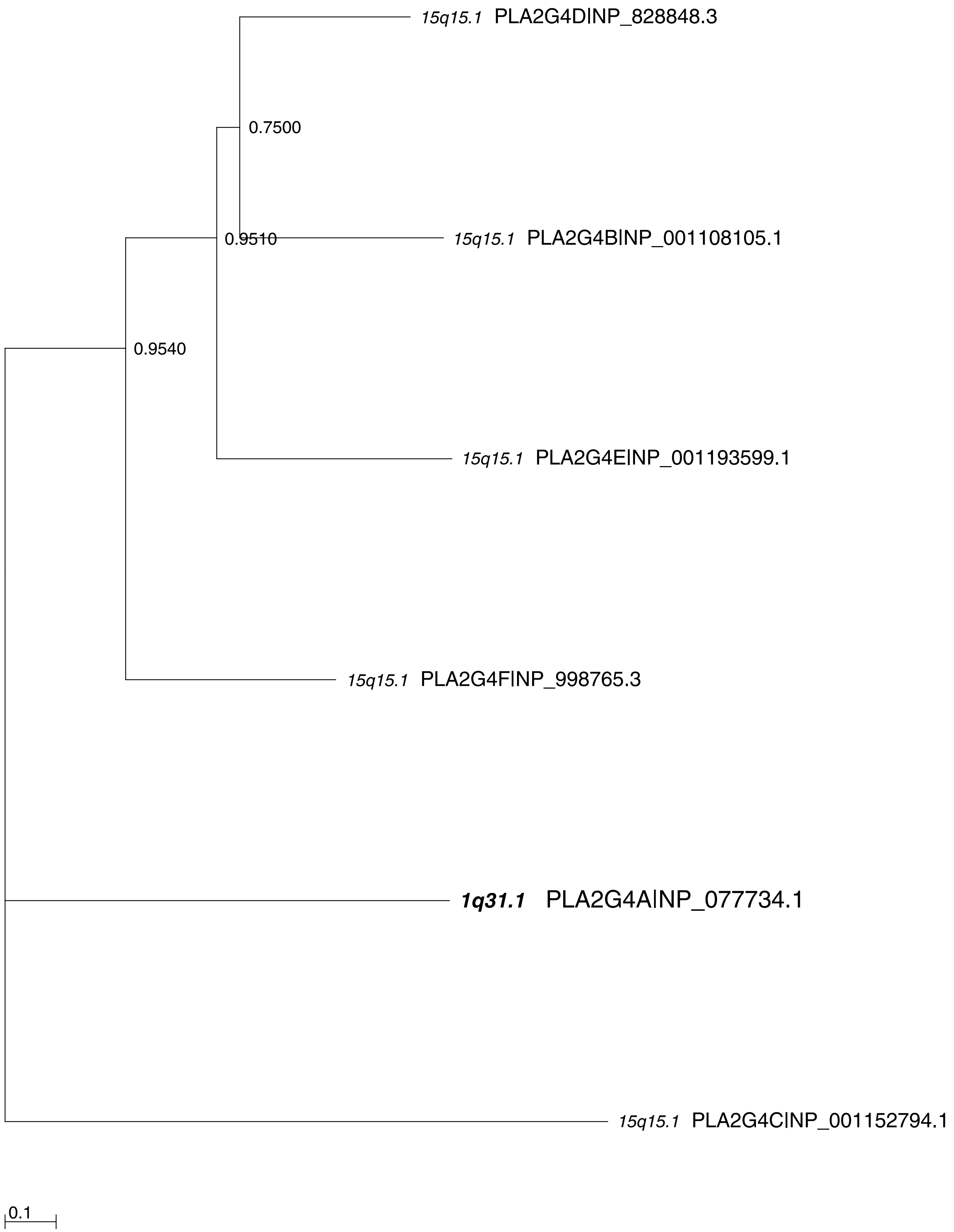
Online Resource 2.10. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein UCHL5. This tree was constructed with 1000 bootstrap replicates from the protein sequences of UCHL5 itself and the DELTA BLAST first iteration hits that were clearly more similar to UCHL5 than others.



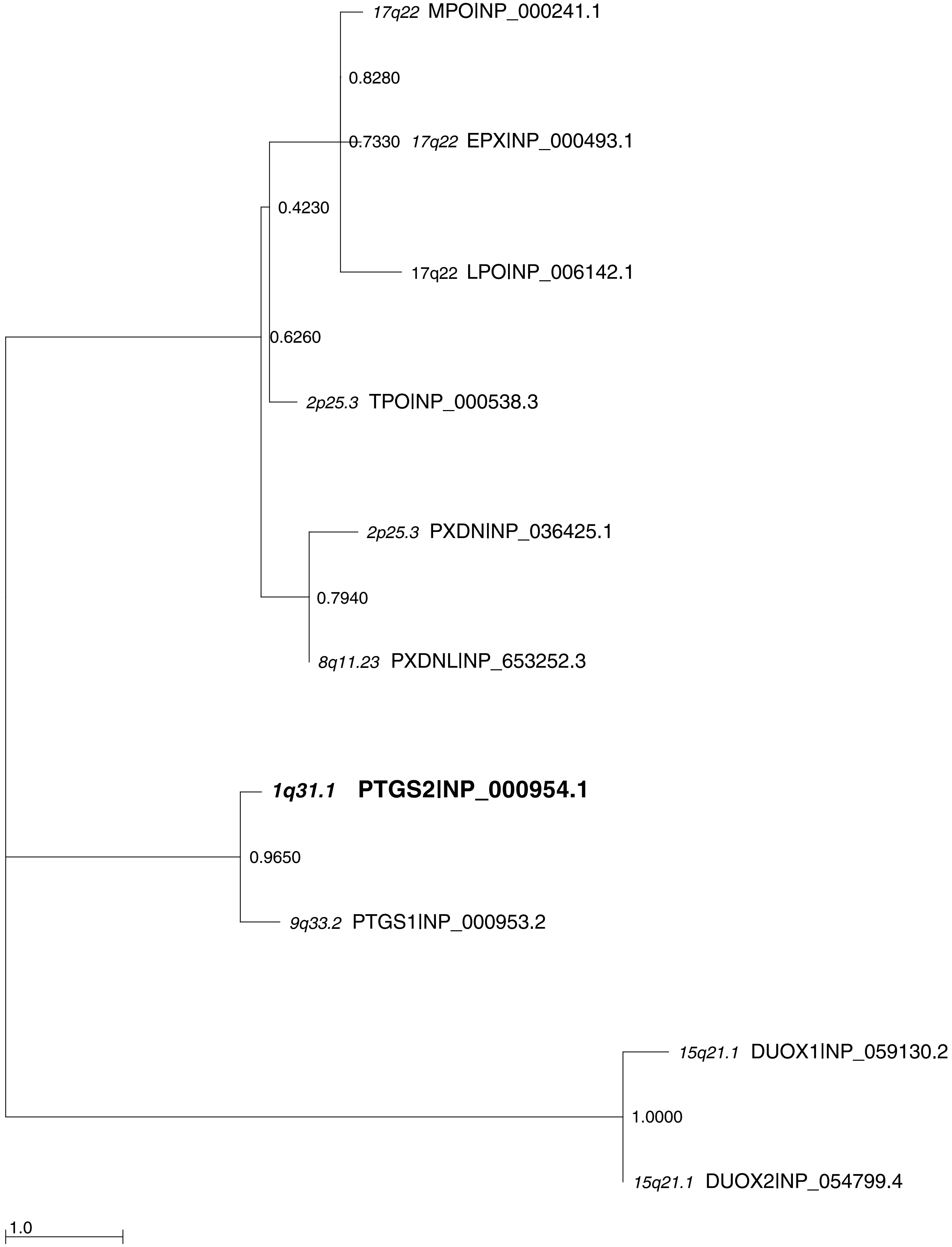
Online Resource 2.11. Maximum Likelihood tree showing the paralogues of *Homo sapiens* RGSL1 and the R4 RGS proteins in *RGS1/RGS16* region (RGS1, 2, 8, 13, 16, 18 and 21). This tree was constructed with 1000 bootstrap replicates from the protein sequences of the known human RGS-domain proteins using the first iteration of DELTA BLAST.



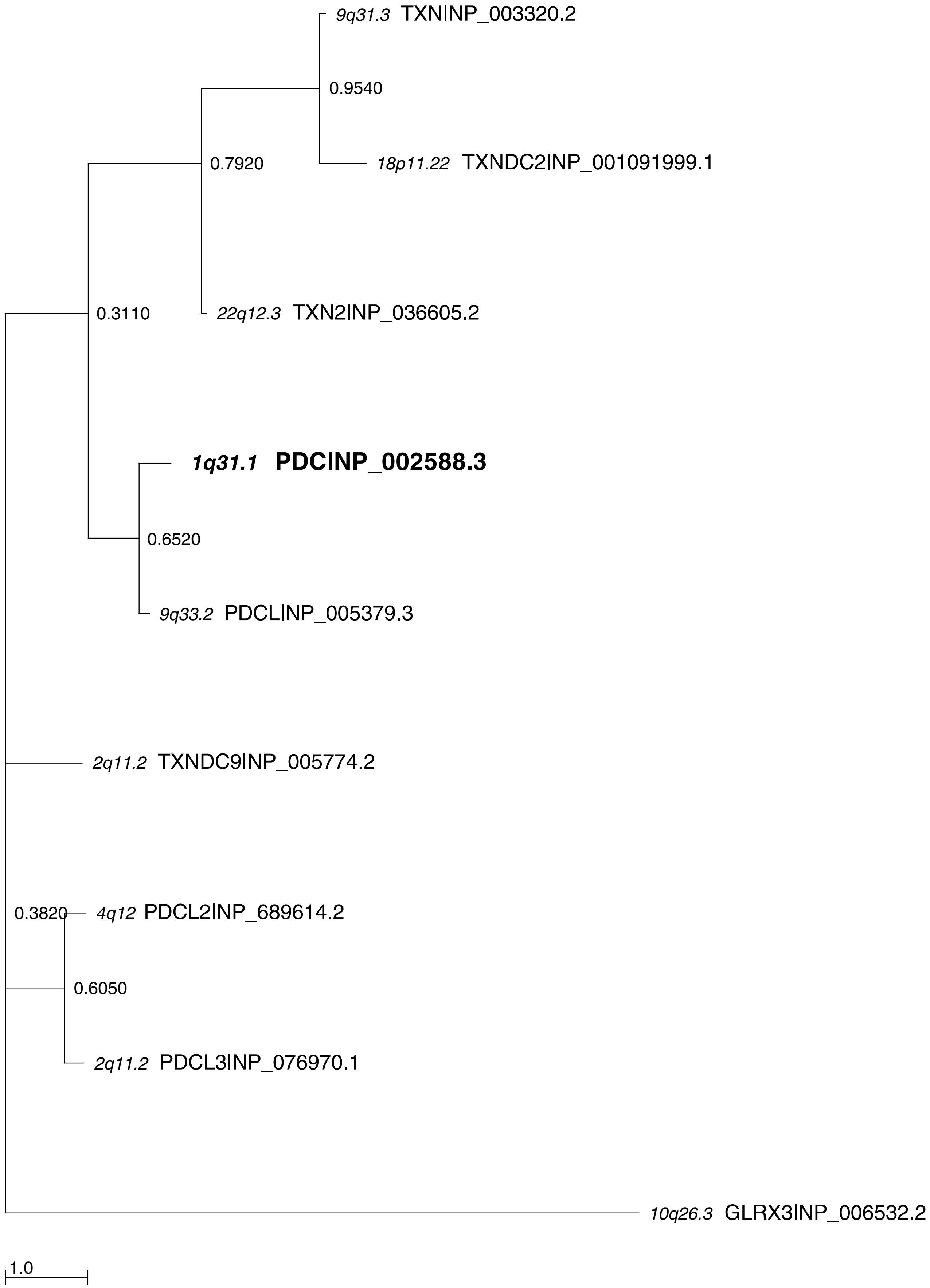
Online Resource 2.12. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein FAM5C. This tree was constructed with 1000 bootstrap replicates from the sequences of FAM5C and the best hits obtained with the first iteration of DELTA BLAST.



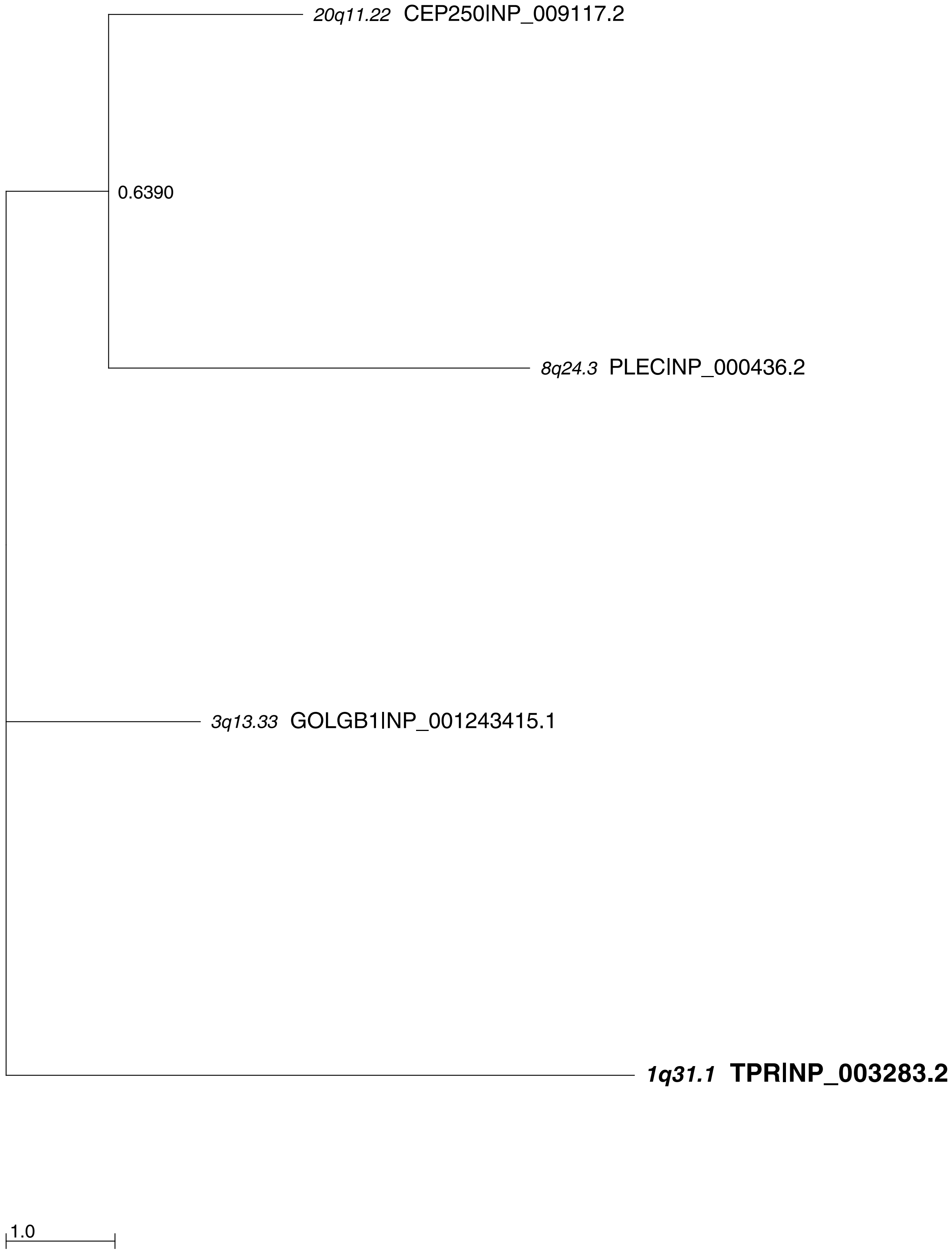
Online Resource 2.13. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein PLA2G4A. This tree was constructed with 1000 bootstrap replicates from the sequences of PLA2G4A itself and the best hits obtained with the first iteration of DELTA BLAST (PLA2G4C had an e-value on 9,00e-112 and coverage of 77%; the next hit had an e-value of 1e-28 and coverage of 16%).



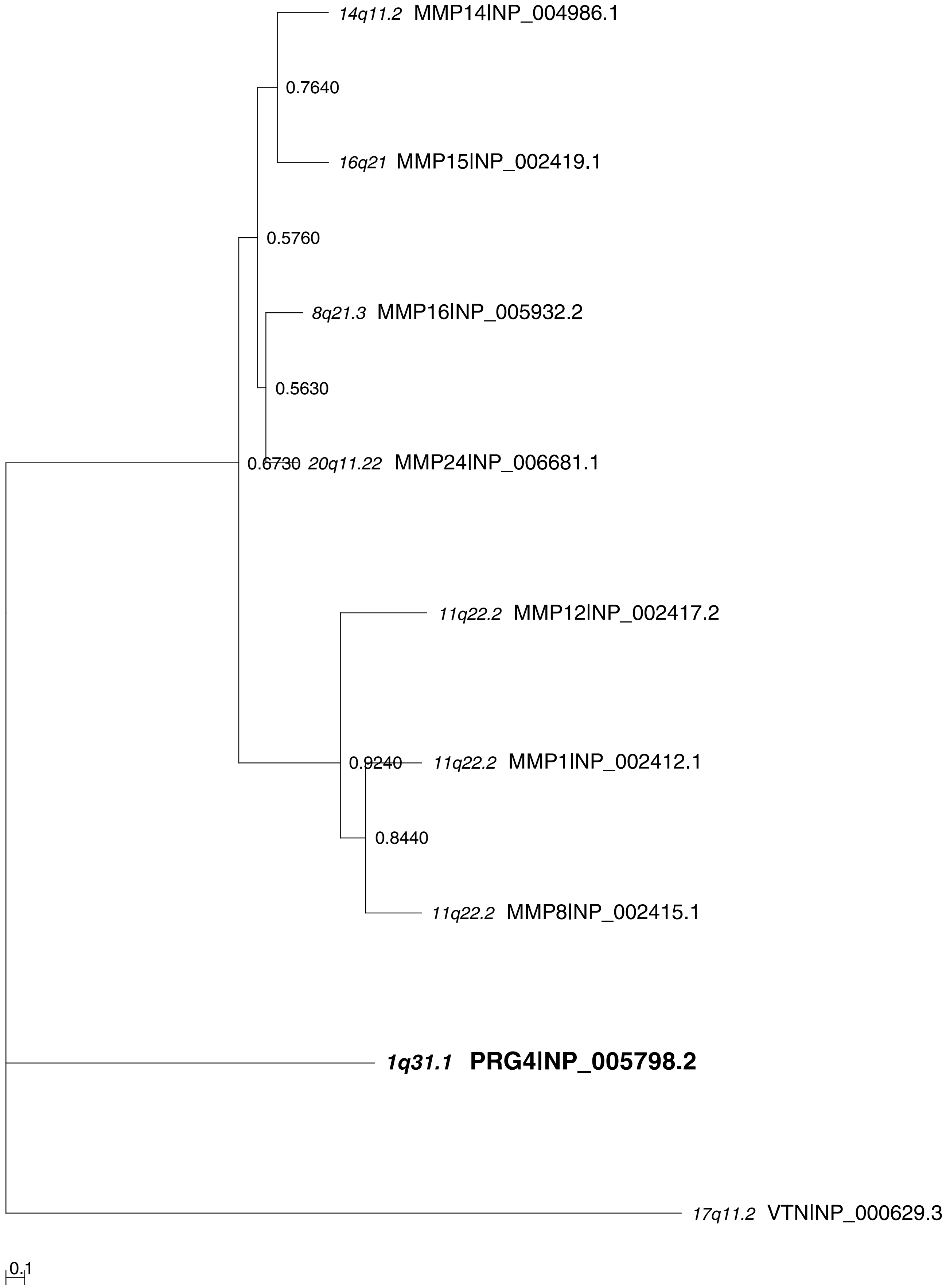
Online Resource 2.14. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein PTGS2. This tree was constructed with 1000 bootstrap replicates from the sequences of PTGS2 itself and the best hits obtained with the first iteration of DELTA BLAST that also had a peroxidase domain.



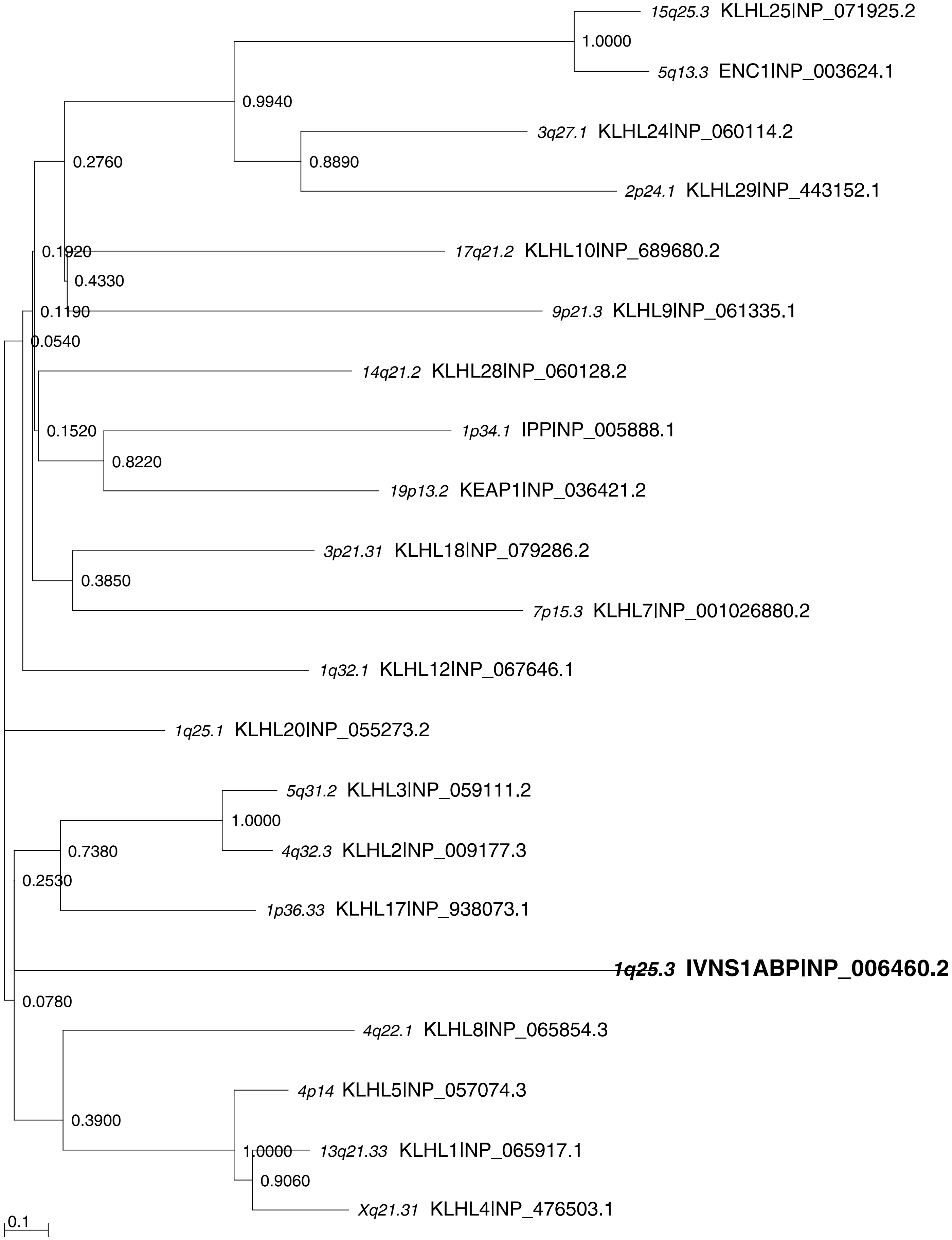
Online Resource 2.15. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein PDC. This tree was constructed with 1000 bootstrap replicates from the sequences of PDC itself and the best hits obtained with the first iteration of DELTA BLAST with a cutoff of 1e-20. However, based on domain structure alone only PDCL, PDCL1, PDCL2, PDCL3 are paralogs of PDC.



Online Resource 2.16. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein TPR. This tree was constructed with 1000 bootstrap replicates from the sequences of TPR itself and some of the best hits obtained with the first iteration of DELTA BLAST. However, even these sequences have very low identity to TPR itself (12%) and are thus likely not real paralogs.

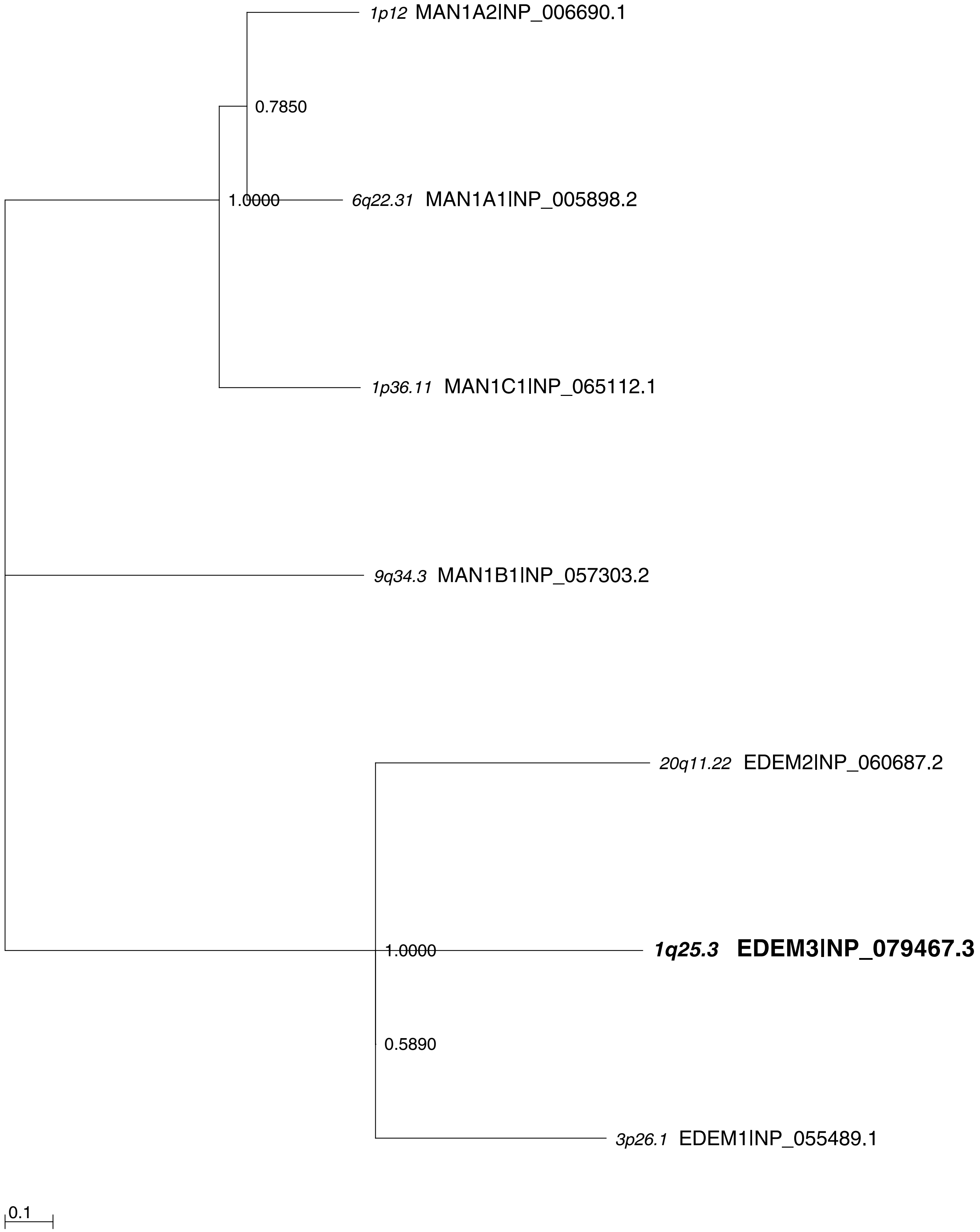


Online Resource 2.17. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein PRG4. This tree was constructed with 1000 bootstrap replicates from the sequences of PRG4 itself and the best hits obtained with the first iteration of DELTA BLAST. However, all except VTN have less than 30% identity and similarity, and even that protein is not very similar Phylogeny confirms this.

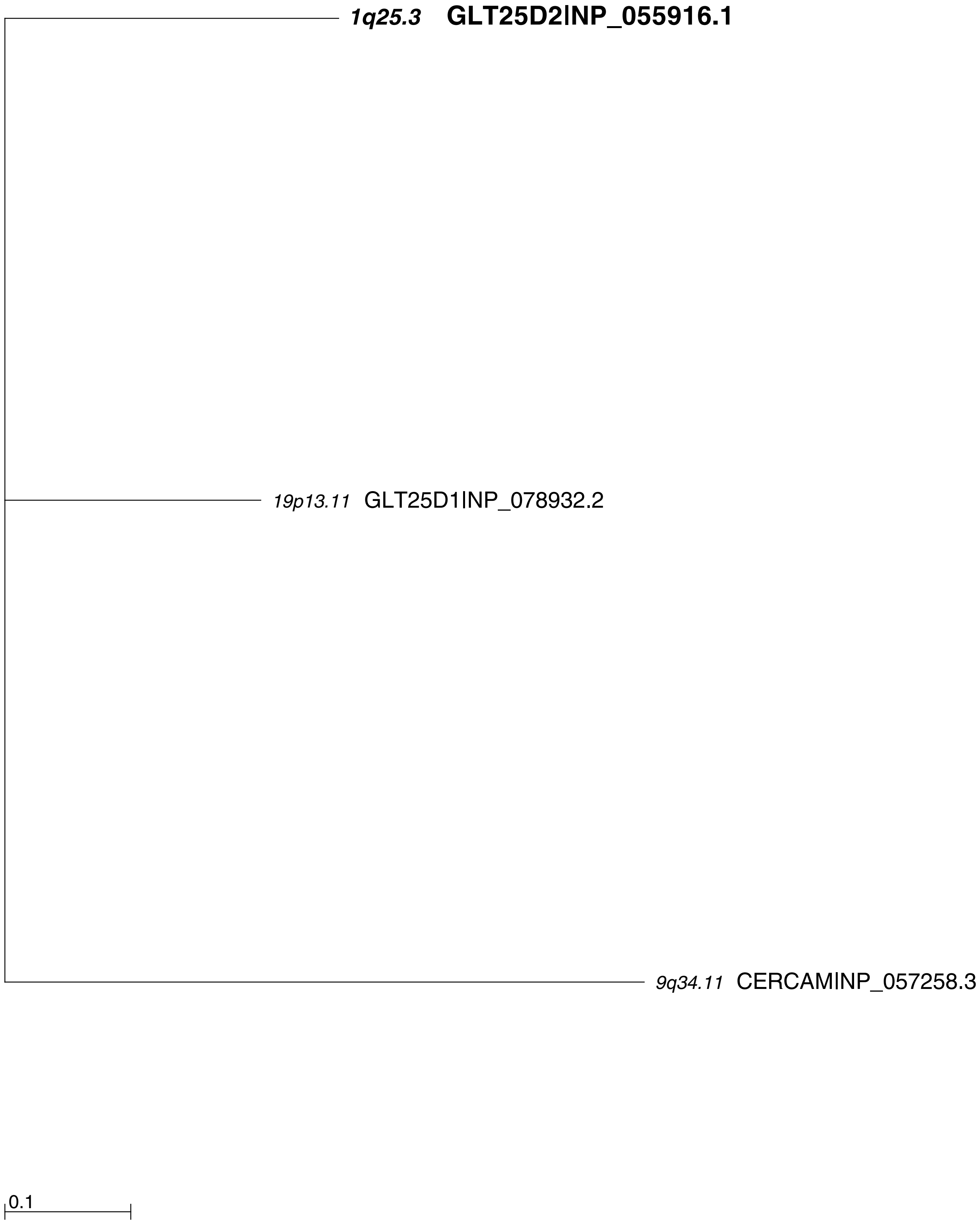


Online Resource 2.18. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein IVNS1ABP. This tree was constructed with 1000 bootstrap replicates from the protein sequences of IVNS1ABP itself and 20 best hits obtained with the first iteration of DELTA BLAST. IVNS1ABP consists mainly of kelch-like domains, which is a common feature of many proteins.

Online Resource 2.19. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein FAM129A. This tree was constructed from the protein sequences of FAM129A itself and the only two sequences with some identity and low e-values. Bootstrapping the phylogeny was not possible due to insufficient number of sequences.

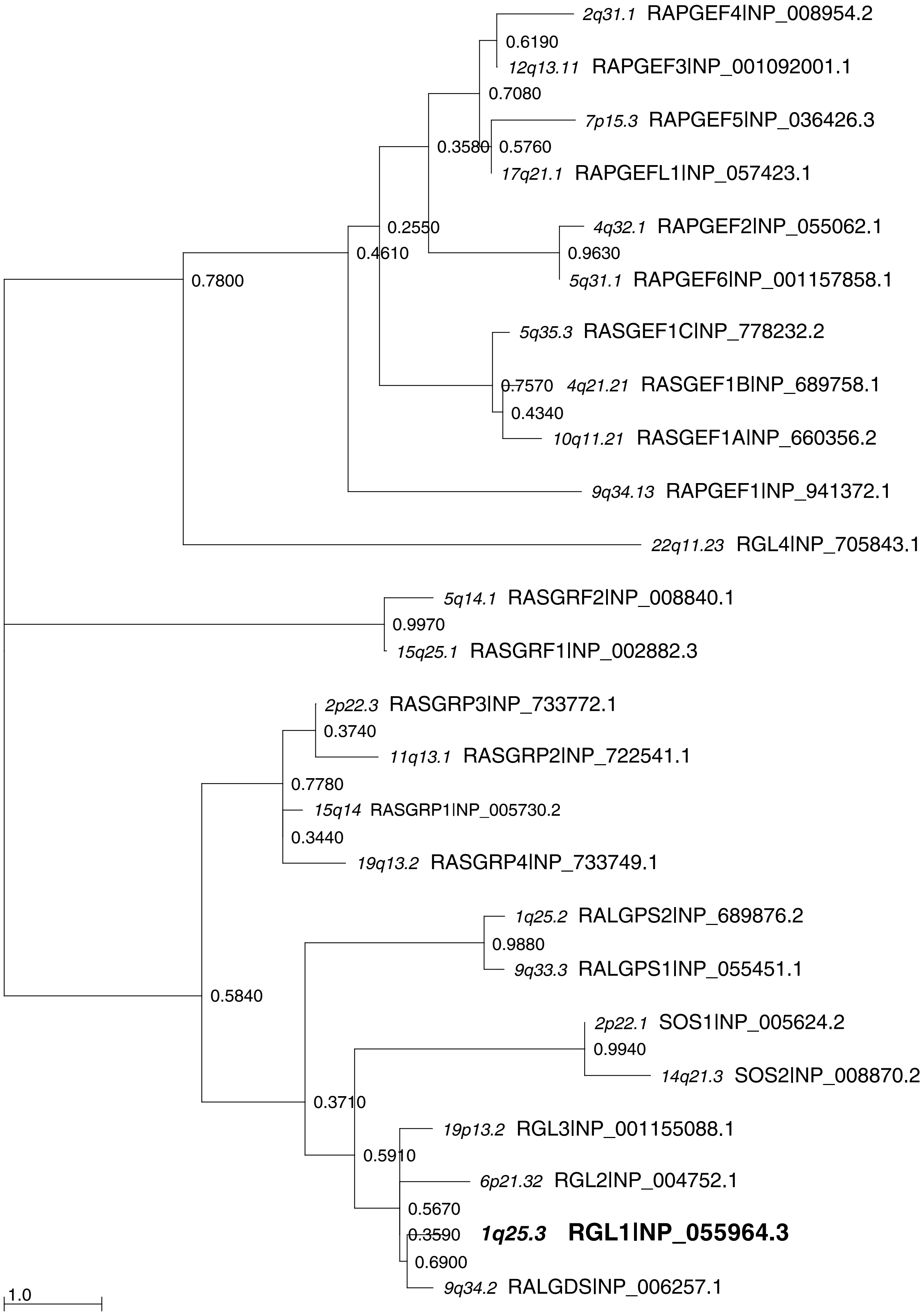


Online Resource 2.20. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein EDEM3. This tree was constructed with 1000 bootstrap replicates from the protein sequences of EDEM3 itself and the 6 best results of DELTA BLAST's first iteration against human reference sequences. All other hits had very low coverage.

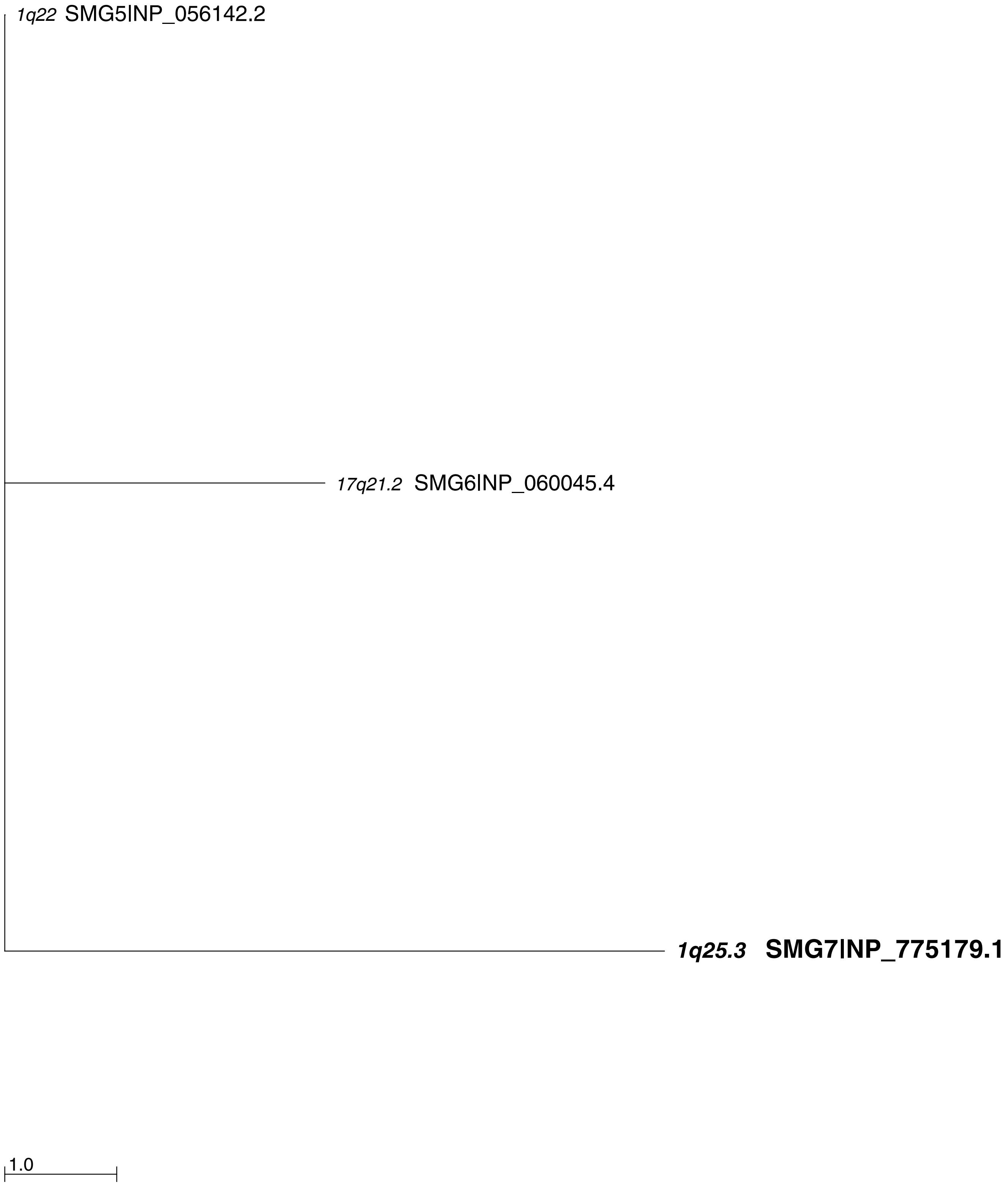


Online Resource 2.21. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein GLT25D2.This tree was constructed from the protein sequences of GLT25D2 itself and the only two hits of DELTA BLAST's first iteration against human reference sequences that had a similar domain structure. In addition, these were the only hits with low e values. Due to insufficient number of sequences, bootstrapping was not possible.

Online Resource 2.22. Maximum Likelihood tree showing the paralogues of *Homo sapiens* proteins APOBEC4. This tree was constructed with 1000 bootstrap replicates from the sequences of all the known human proteins with an APOBEC domain and an additional cytidine deaminase AICDA with the first iteration of DELTA BLAST.

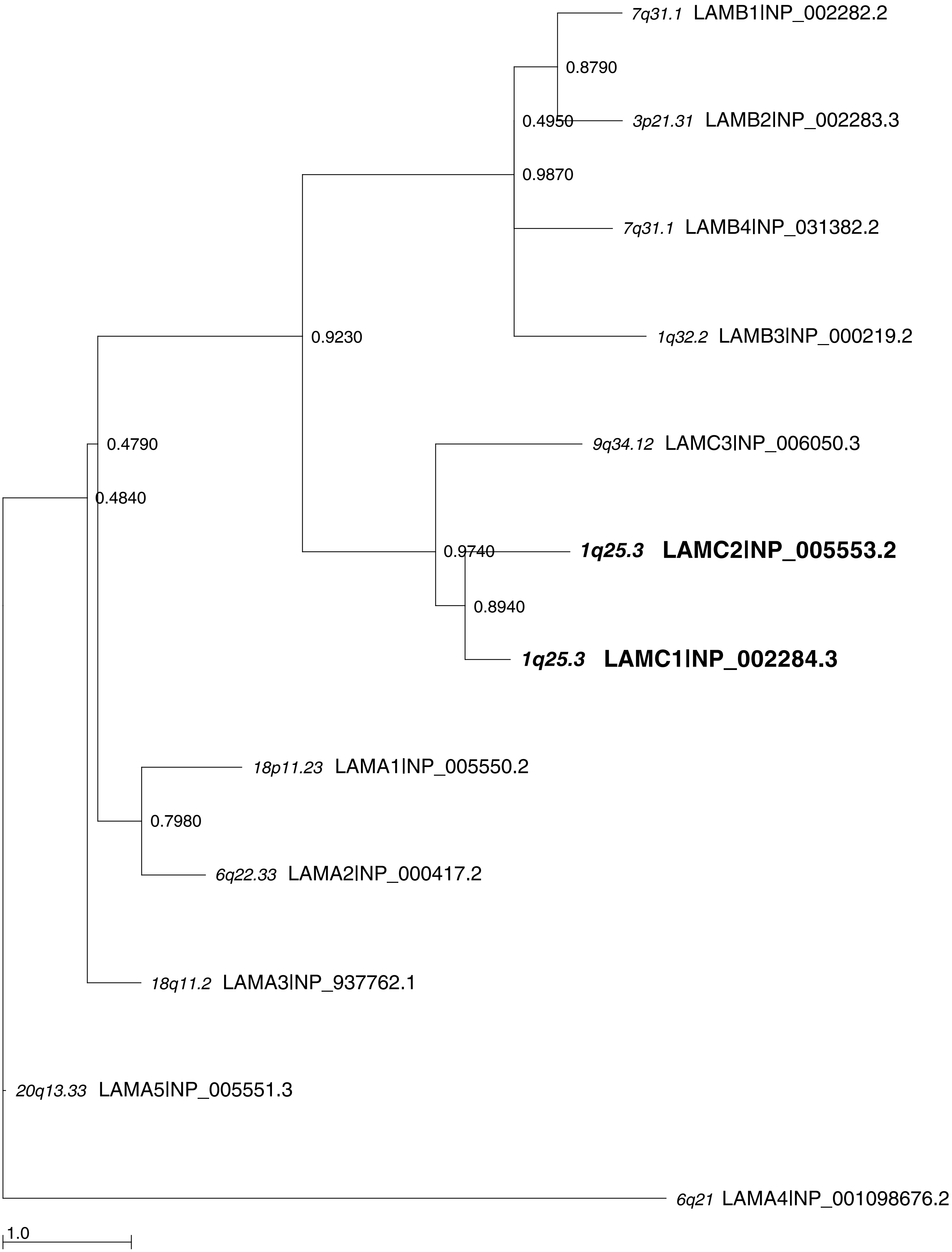


Online Resource 2.23. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein RGL1.This tree was constructed with 1000 bootstrap replicates from the protein sequences of RGL1 itself and 25 best hits obtained with the first iteration of DELTA BLAST. All of them had some domain similarities with RGL1; but only RALGDS, RGL2 and RGL3 had e values lower than 1e-140.

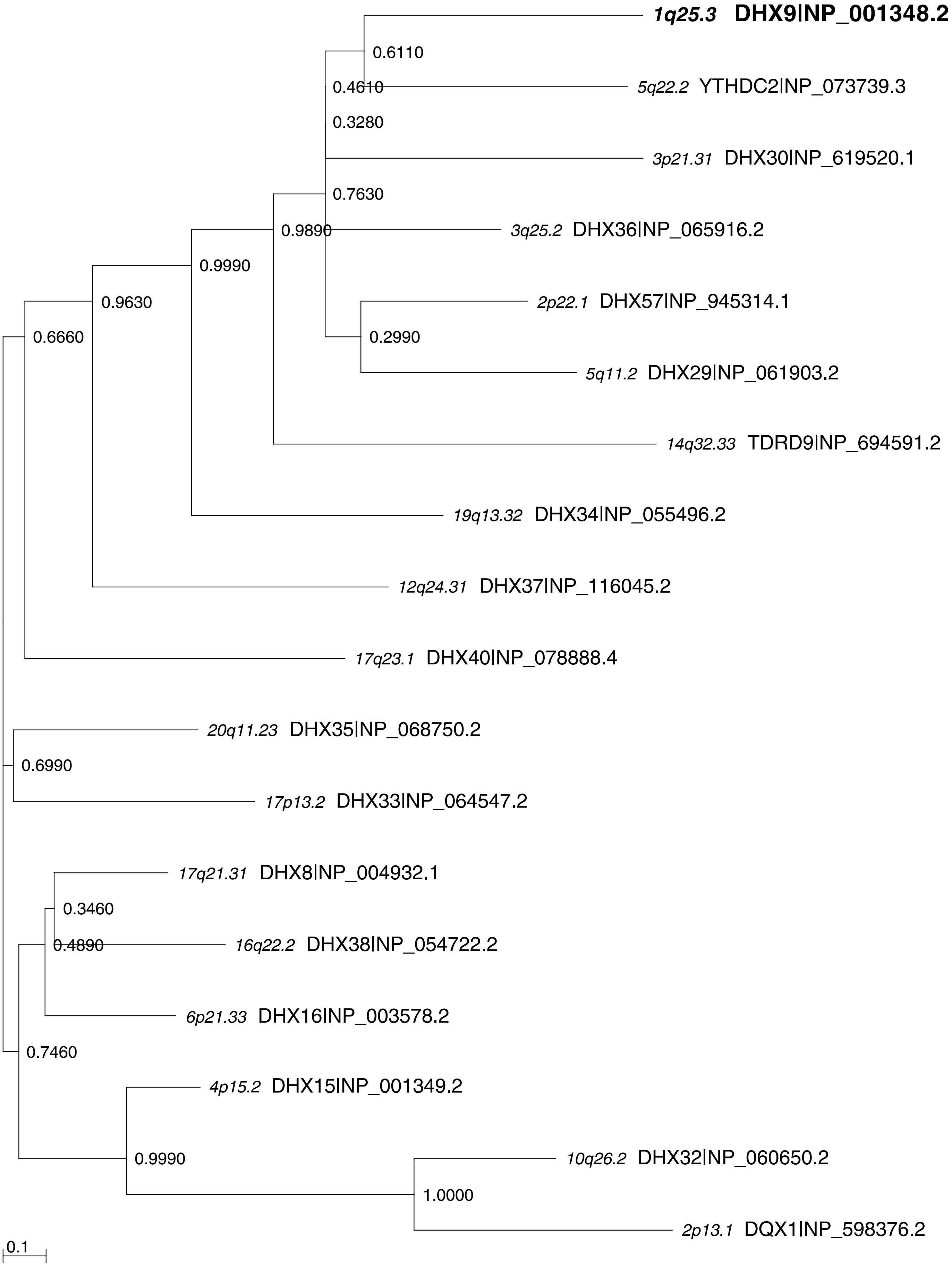


Online Resource 2.24. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein SMG7. This tree was constructed from the protein sequences of SMG7 itself and both proteins with some similar domains (first two DELTA BLAST first iteration hits when blasting against human reference sequences). Bootstrapping was not possible for this phylogeny due to insufficient number of sequences.

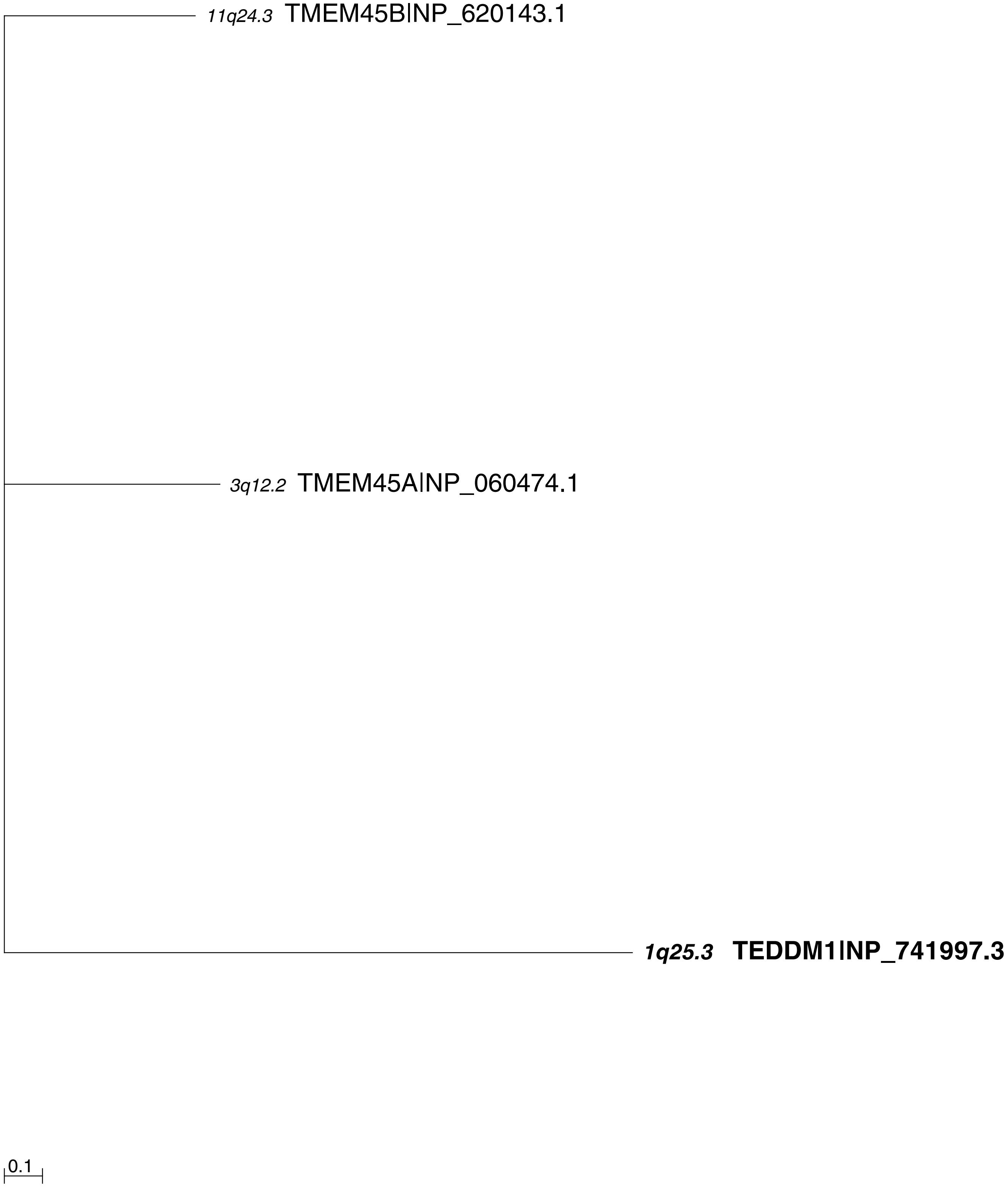
Online Resource 2.25. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein NMNAT2. This tree was constructed from the protein sequences of NMNAT2 itself and both proteins with some domain similarity (first two DELTA BLAST first iteration hits when blasting against human reference sequences. Bootstrapping was not possible due to insufficient number of sequences.



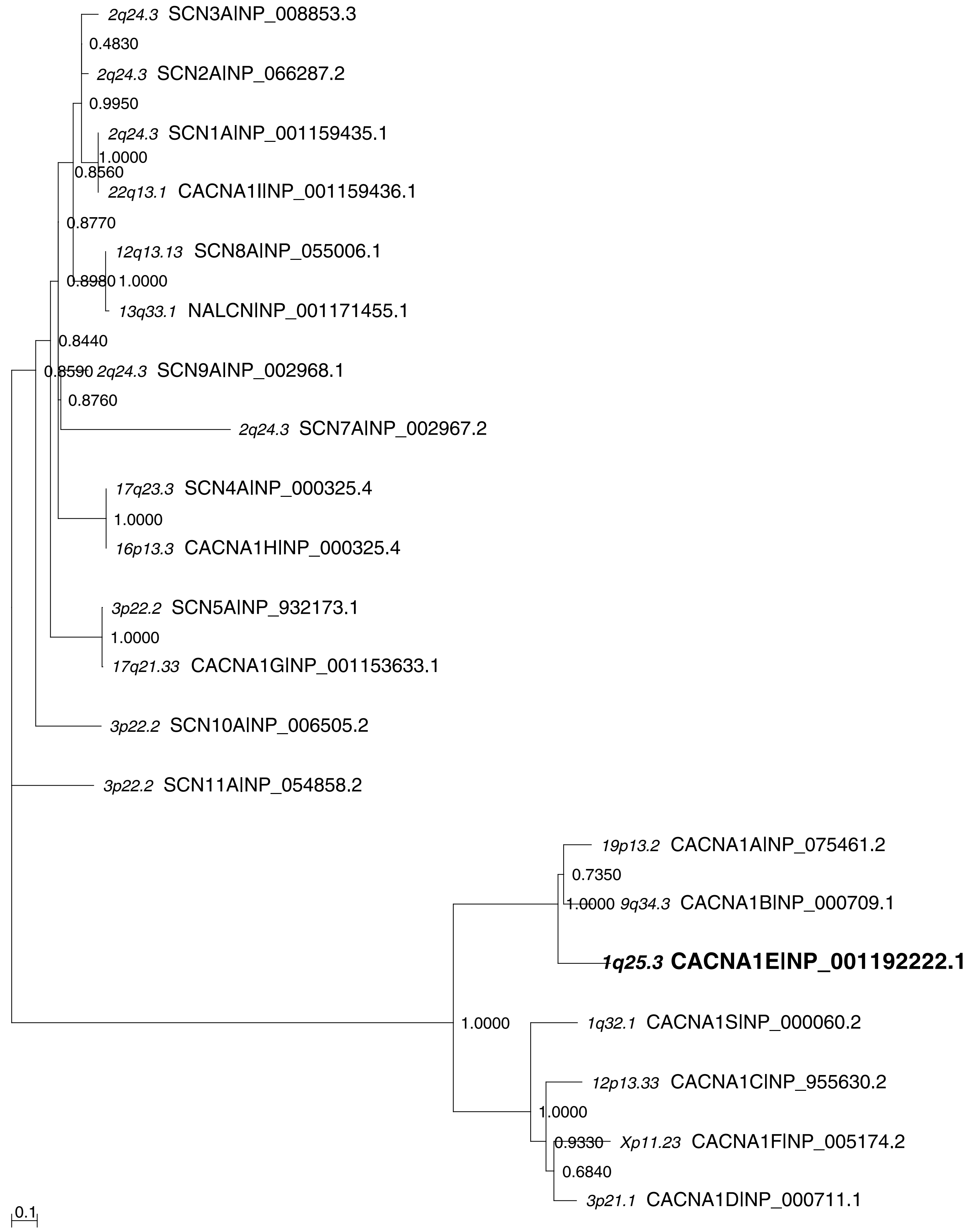
Online Resource 2.26. Maximum Likelihood tree showing the paralogues of *Homo sapiens* proteins LAMC1 and LAMC2. This tree was constructed with 1000 bootstrap replicates from the protein sequences of LAMC1, LAMC2 and 10 results of DELTA BLAST's first iteration that had a similar domain structure (laminin proteins). They also had the lowest e values.



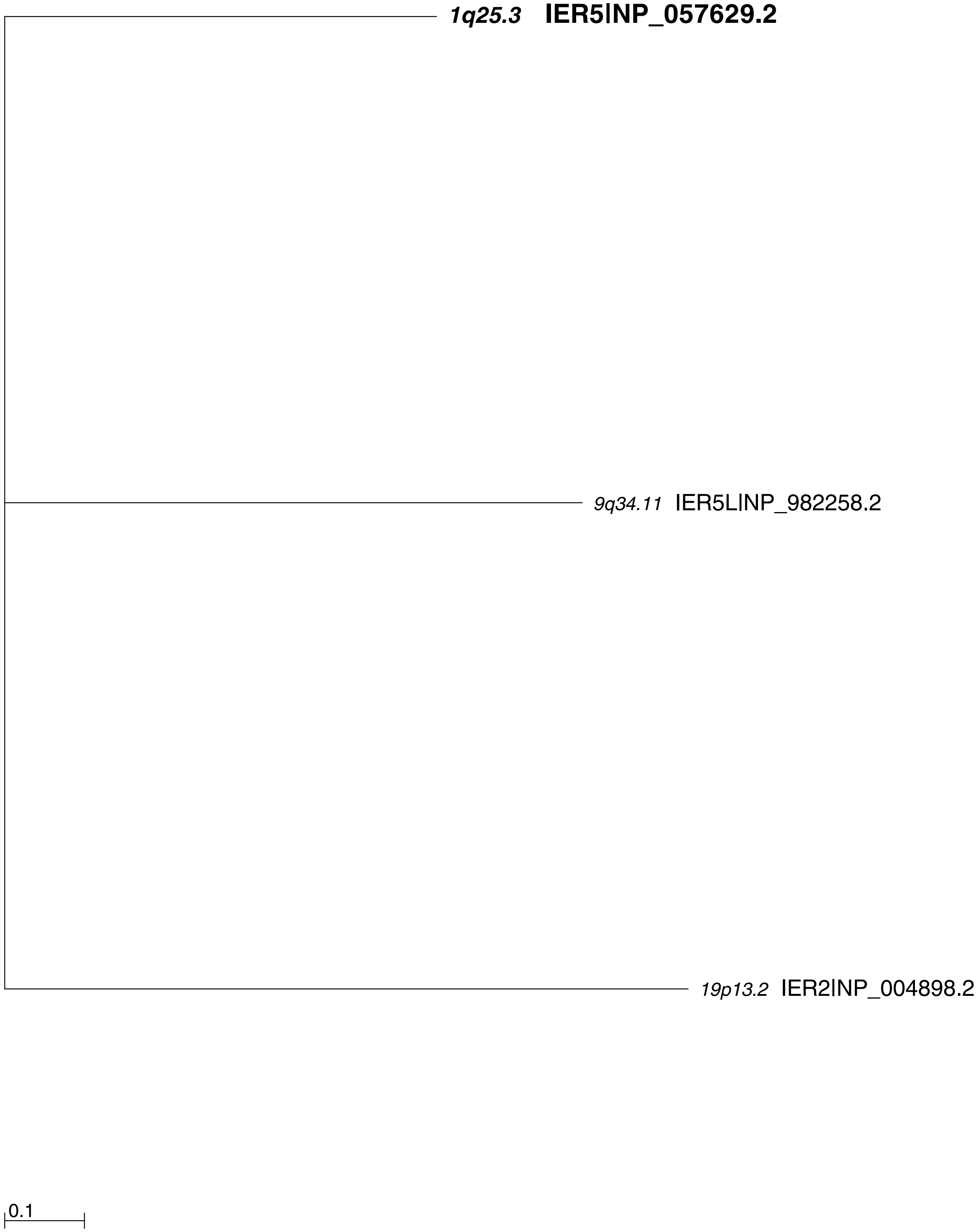
Online Resource 2.27. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein DHX9.This tree was constructed with 1000 bootstrap replicates from the protein sequences of DHX9 itself and all 17 results of DELTA BLAST's first iteration that had a similar domain structure.



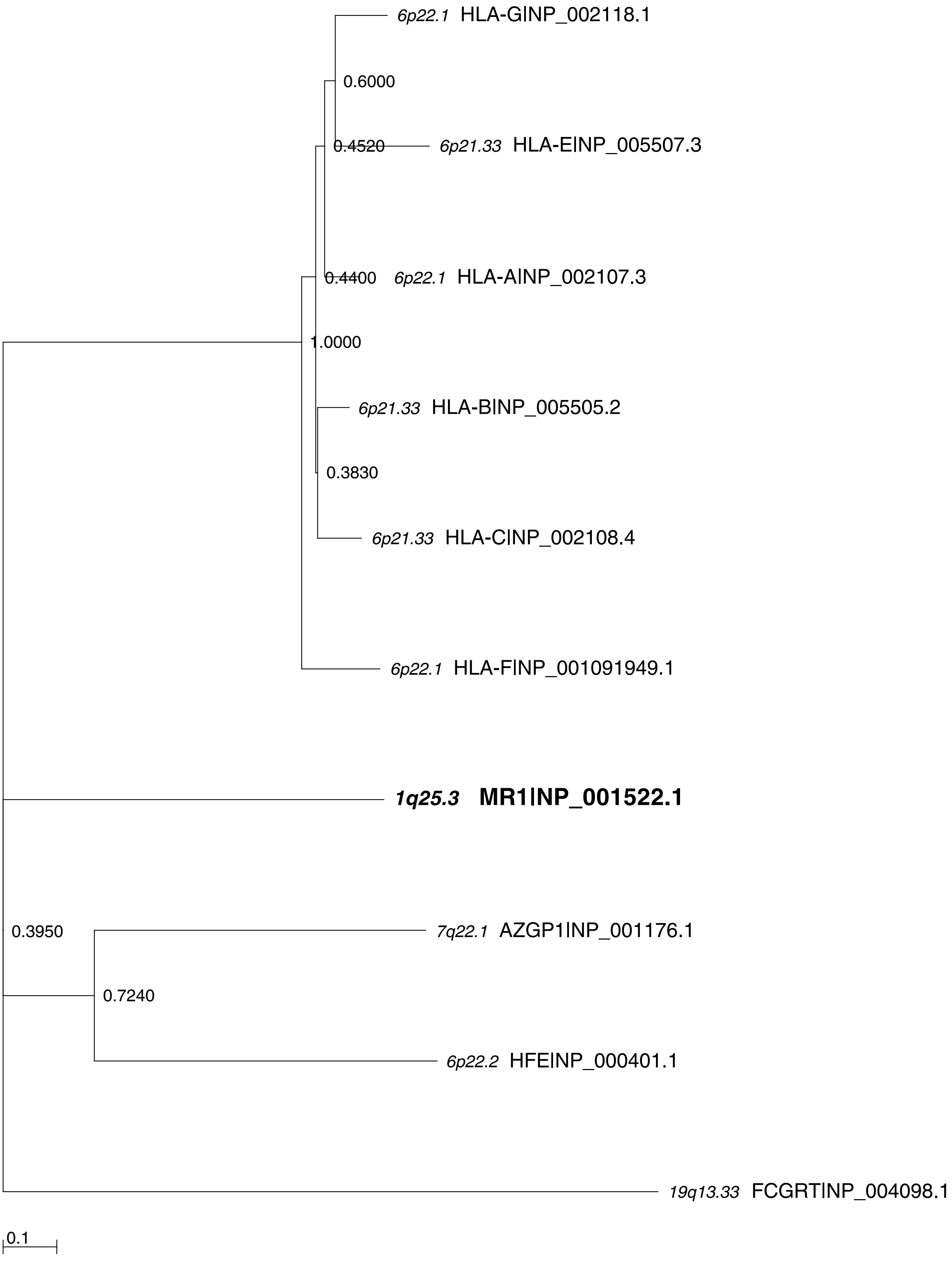
Online Resource 2.28. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein TEDDM1. This tree was constructed from the protein sequences of TEDDM1 itself and the only two hits of DELTA BLAST's first iteration against human reference sequences that had low e values. Due to insufficient number of sequences, bootstrapping was not possible.



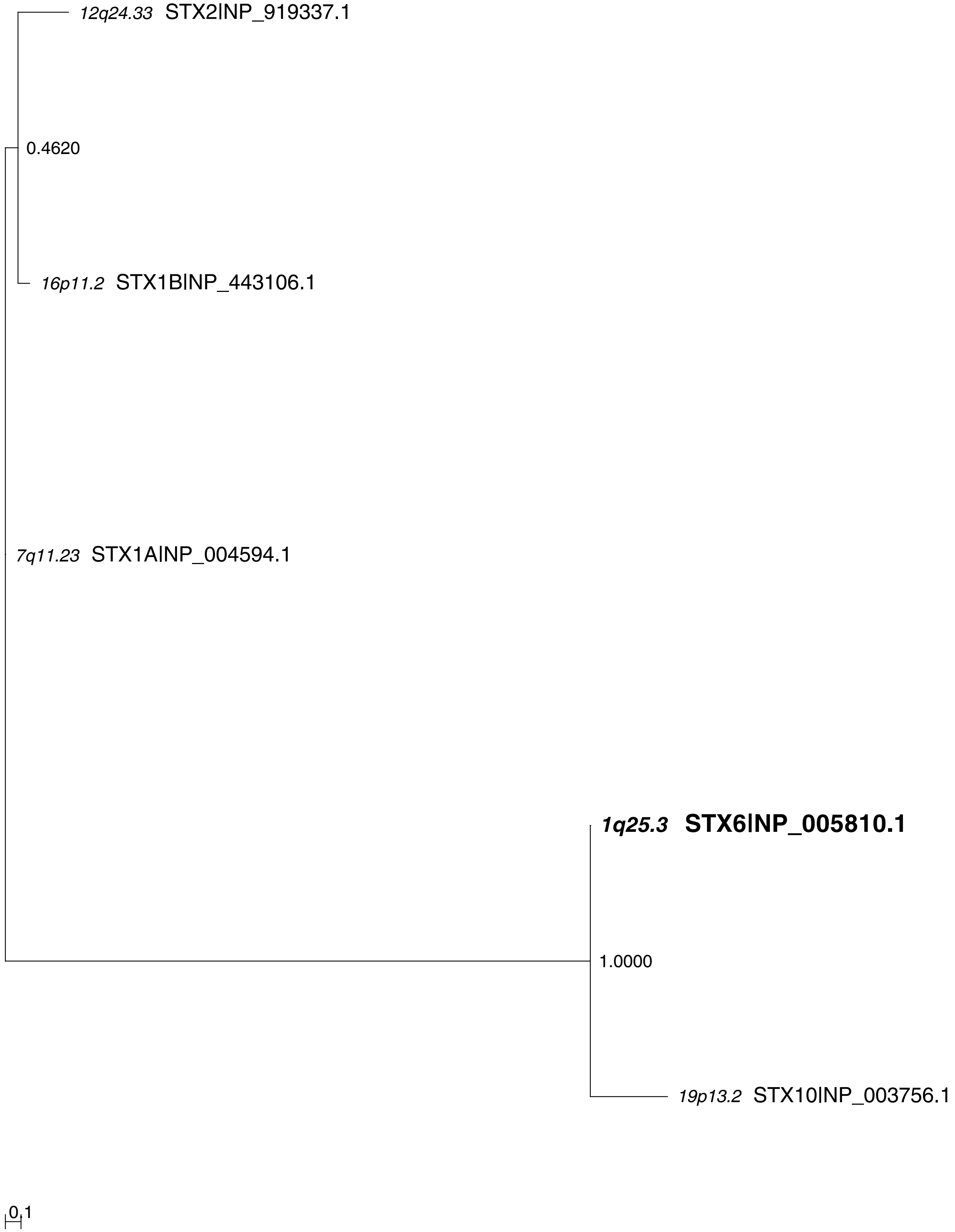
Online Resource 2.29. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein CACNA1E. This tree was constructed with 1000 bootstrap replicates from the protein sequences of CACNA1E itself and 20 best hits obtained with the first iteration of DELTA BLAST against NCBI human reference sequences.



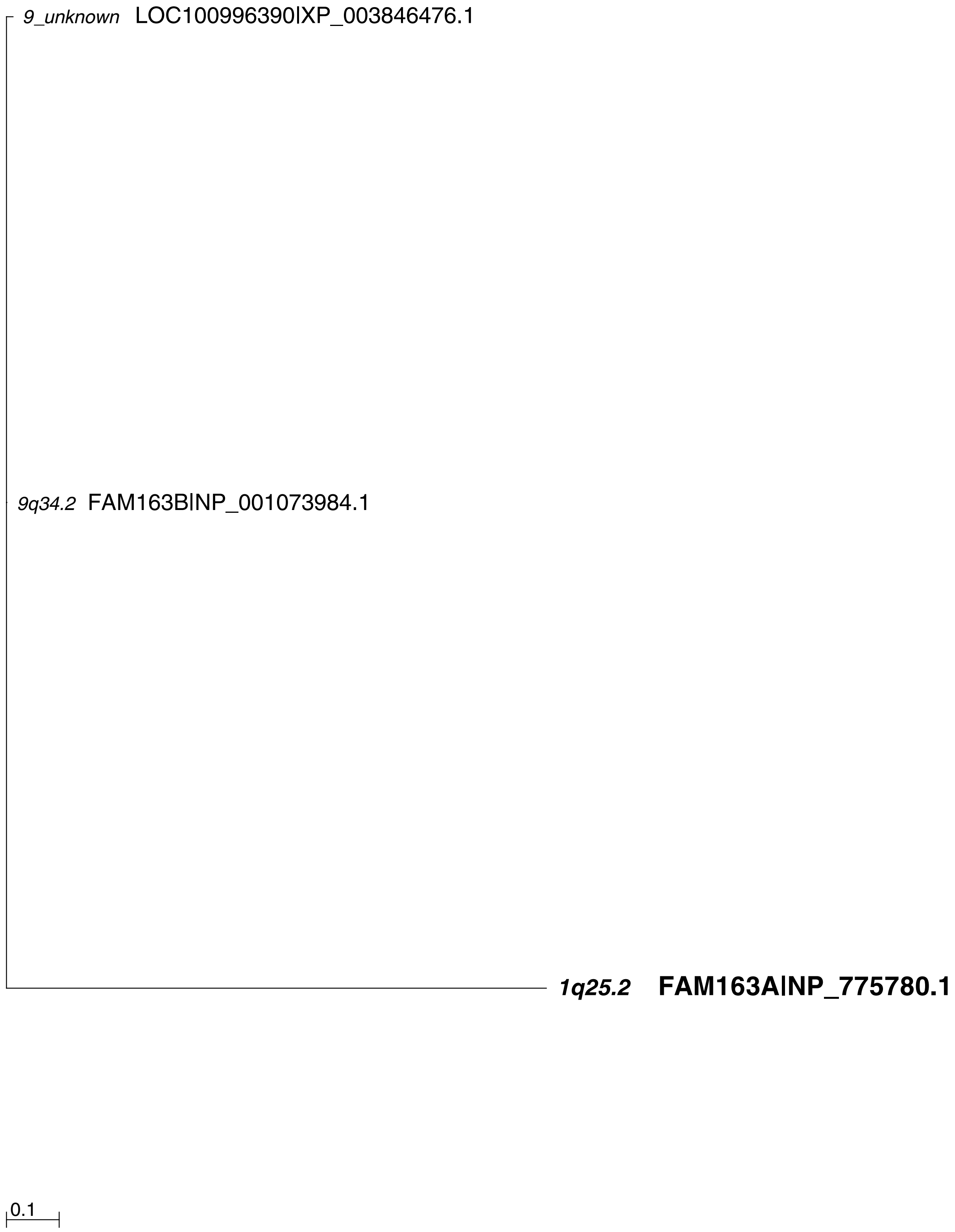
Online Resource 2.30. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein IER5. This tree was constructed from the protein sequences of IER5 itself and the only two hits of pblast against human reference sequences that had the characteristic IER domain. pblast was used instead of DELTA BLAST for this gene because DELTA BLAST failed to recognize the IER domain and therefore did not produce accurate results. Due to insufficient number of sequences, bootstrapping was not possible for this phylogenetic tree.



Online Resource 2.31. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein MR1.This tree was constructed with 1000 bootstrap replicates from the protein sequences of MR1 itself and 9 additional sequences amongst DELTA BLAST hits against NCBI human reference sequences. These sequences all had coverage > 90% and e values < 1e-90, being therefore clearly better matches than the other hits.



Online Resource 2.32. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein STX6. This tree was constructed with 1000 bootstrap replicates from the protein sequences of STX6 itself and all the hits of DELTA BLAST's first iteration against human reference sequences that had the syntaxin domain and e value < e-10.



Online Resource 2.33. Maximum Likelihood tree showing the paralogues of *Homo sapiens* protein FAM163A. This tree was constructed from the protein sequences of FAM163A itself and the only two hits of DELTA BLAST's first iteration against the only two human reference sequences that had e values e values < 1e-20 (next hit was e-04). Due to insufficient number of sequences, bootstrapping was not possible.