

9. Protein  
dizilerinden  
filogenetik analiz

First Position	Second Position				Third Position	PRETTYSED of AJ620291 from 1 to 862
	U(T)	C	A	G		1 Cgacgtcgttgcctgctggacatgaaccagcgcgcgcgactctcgcctaccgatgaagtcgactccgtgcacggcaagttcaagcactctgtctca 100 1 D V V A V V D M N T D A E Y F A Y Q N K Y D S V H G K F E H S V S 33
U(T)	Phe	Ser	Tyr	Cys	U(T)	101 acgacgaagagcaaccogtccgttgcgaagatgacaccctcgttctgaacggcaccgcattctgtgctgaaggcgcagcgcaaccocgcggaccttc 200 34 T T K S N P S V A K E D T L V V N G H R I L C V K A Q R N P A D L P 67
	Phe	Ser	Tyr	Cys	C	
	Leu	Ser	STOP	STOP	A	
	Leu	Ser	STOP	Trp	G	201 catggggcaagctgggagtgagatgctgatcagctccactggcctctcaccgtgaagtcggcgcctgaggccaccctccgggtggtgcacgcaagct 300 68 W G K L G V E Y V I E S T G L F T V K S A A E G H L R G G A R K V 100
C	Leu	Pro	His	Arg	U(T)	301 catcctcagcctcctcctcctggcggcgcgaagccttctgtgatggcctgaaccacaacgactacaaccocgcggagcaccactcgtgtcgaacgcc 400 101 I I S A P A S G G A K T F V M G V N H N D Y N P R E H H V V S N A 133
	Leu	Pro	His	Arg	C	
	Leu	Pro	Gln	Arg	A	
	Leu	Pro	Gln	Arg	G	401 tctgacagaccaactgccttgcctccacttctgcaactgctggggaaggagggtctcggcatcaccactggcctcatgacgacaatccactcgtacacgg 500 134 S C T T N C L A P L V H V L V K E G F G I T T G L M T T I H S Y T A 167
A	Ile	Thr	Asn	Ser	U(T)	501 ccaccagaaagactgtcgtatggcgtgtcactcaagcactggcgcgctggcgcgctgctgctgaadacatcccagcaagactggtcggcccaagcc 600 168 T Q K T V D G V S I X D N R G G R A A A I N I I P S T T G A A K A 200
	Ile	Thr	Asn	Ser	C	
	Ile	Thr	Lys	Arg	A	
	Met	Thr	Lys	Arg	G	601 tctcggcctcgtgctcccaagcagcgggcaagctgacgggcctggcctccgtgctcactcggctgactctctgtggctcagcctaaccttcactgcc 700 201 V G M V I P S T Q G K L T G M A F R V P T A D V S V V D L T F T A 233
G	Val	Ala	Asp	Gly	U(T)	701 gcgctgacaccagcatcaaggagatcgacgctgccctgaagcgcgcgctccaggacgtacatgaggaacatcctcgggtacacogacgaggagctggtca 800 234 A R D T S I K E I D A A L K R A S R T Y M R N I L G Y T D E E L V S 267
	Val	Ala	Asp	Gly	C	
	Val	Ala	Glu	Gly	A	
	Val	Ala	Glu	Gly	G	801 gcactgacttcatacaacgacaaccgcagctccactctcagactccaaggcagcgtgcagaac 862 268 T D F I N D N R S S I Y D S K A T L Q N 287

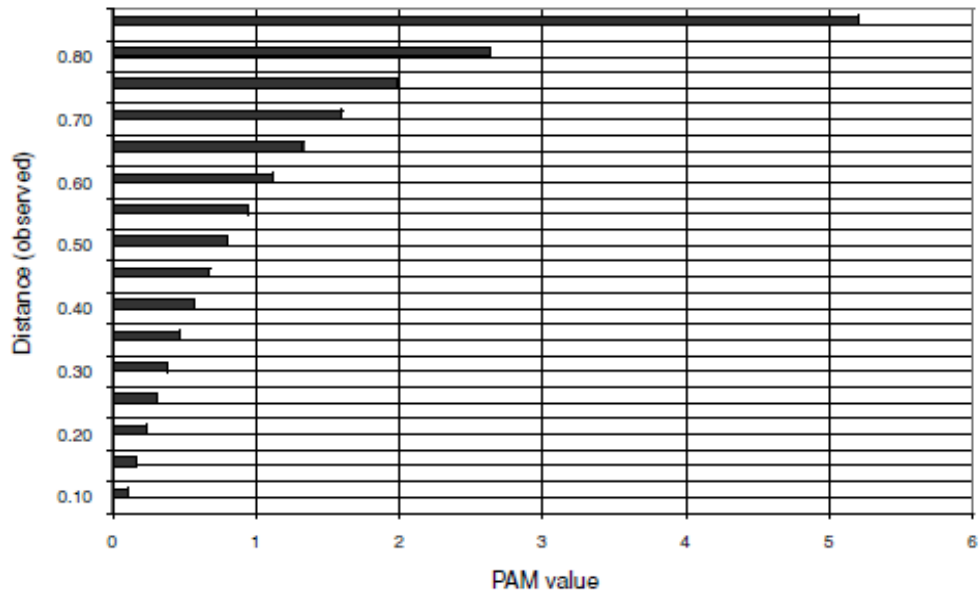
Alanine	Ala	A
Cysteine	Cys	C
Aspartic Acid	Asp	D
Glutamic Acid	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	Y

-----  
Query= LmjF30.2080 |||hypothetical protein, conserved|Leishmania major  
(1656 letters)  
Database: EMBL (release + updates)  
66277815 sequences; 69403745503 total letters

Sequences producing significant alignments:	Score	E
	(bits)	Value
em:CT005267[Leishmania major] Leishmani ...	3283	0.0e+00
em:CV663049[Leishmania donovani chagas ...	430	2.5e-116
em:DX961126[Trypanosoma cruzi] CHORI10 ...	54	6.1e-03
em:DX948996[Trypanosoma cruzi] CHORI10 ...	48	3.8e-01
em:AL939105[Streptomyces coelicolor] St ...	46	1.5e+00
em:BX628111[Anopheles gambiae (African ...	46	1.5e+00
em:AV628110[Chlamydomonas reinhardtii] ...	46	1.5e+00
em:AY686591[Escherichia coli] Escherich ...	46	1.5e+00
em:AI226793[Mus musculus (house mouse) ...	46	1.5e+00
em:BM016710[Homo sapiens (human)] 6036 ...	46	1.5e+00
em:CR529048[Anopheles gambiae (African ...	46	1.5e+00
em:CP000386[Rubrobacter xylanophilus DS ...	46	1.5e+00
em:AZ220839[Trypanosoma brucei] Sheare ...	46	1.5e+00
em:AC074258[Trypanosoma brucei] Trypano ...	46	1.5e+00
em:BC032244[Homo sapiens (human)] Homo ...	44	5.9e+00

Query= LmjF30.2080 |||hypothetical protein, conserved|Leishmania major  
(551 letters)  
Database: UniProt  
4103579 sequences; 1350324263 total letters

Sequences producing significant alignments:	Score	E
	(bits)	Value
trembl:Q4Q785_LEIMA Hypothetical protein...	1021	1.1e-296
trembl:Q4DCK6_TRYCR Hypothetical protein...	473	1.2e-131
trembl:Q584X8_9TRYP Hypothetical protein...	469	1.3e-130
trembl:Q4DB07_TRYCR Hypothetical protein...	466	1.1e-129
trembl:Q0ID02_SYNS3 R3H domain protein ...	203	2.2e-50
trembl:Q31QL2_SYNP7 Single-stranded nuc ...	191	8.7e-47
trembl:Q3AUV7_SYNS9 ATPase. ...	189	4.3e-46
trembl:Q2JIJ9_SYNJB R3H domain protein ...	189	4.3e-46
trembl:A0H4H4_9CHLR Single-stranded nuc ...	189	4.3e-46
trembl:Q05QB0_9SYNE Hypothetical protei ...	188	5.7e-46
trembl:Q3J6X9_NITOC Single-stranded nuc ...	188	5.7e-46

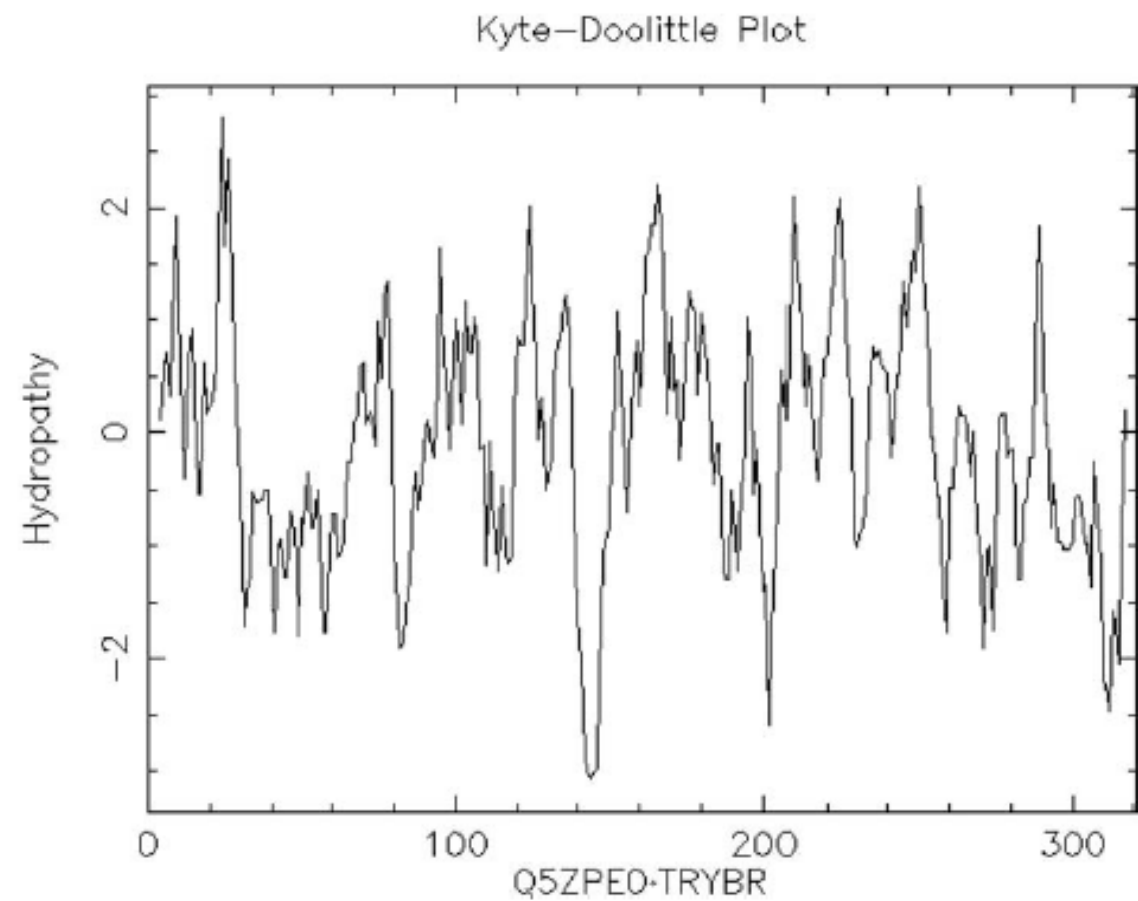
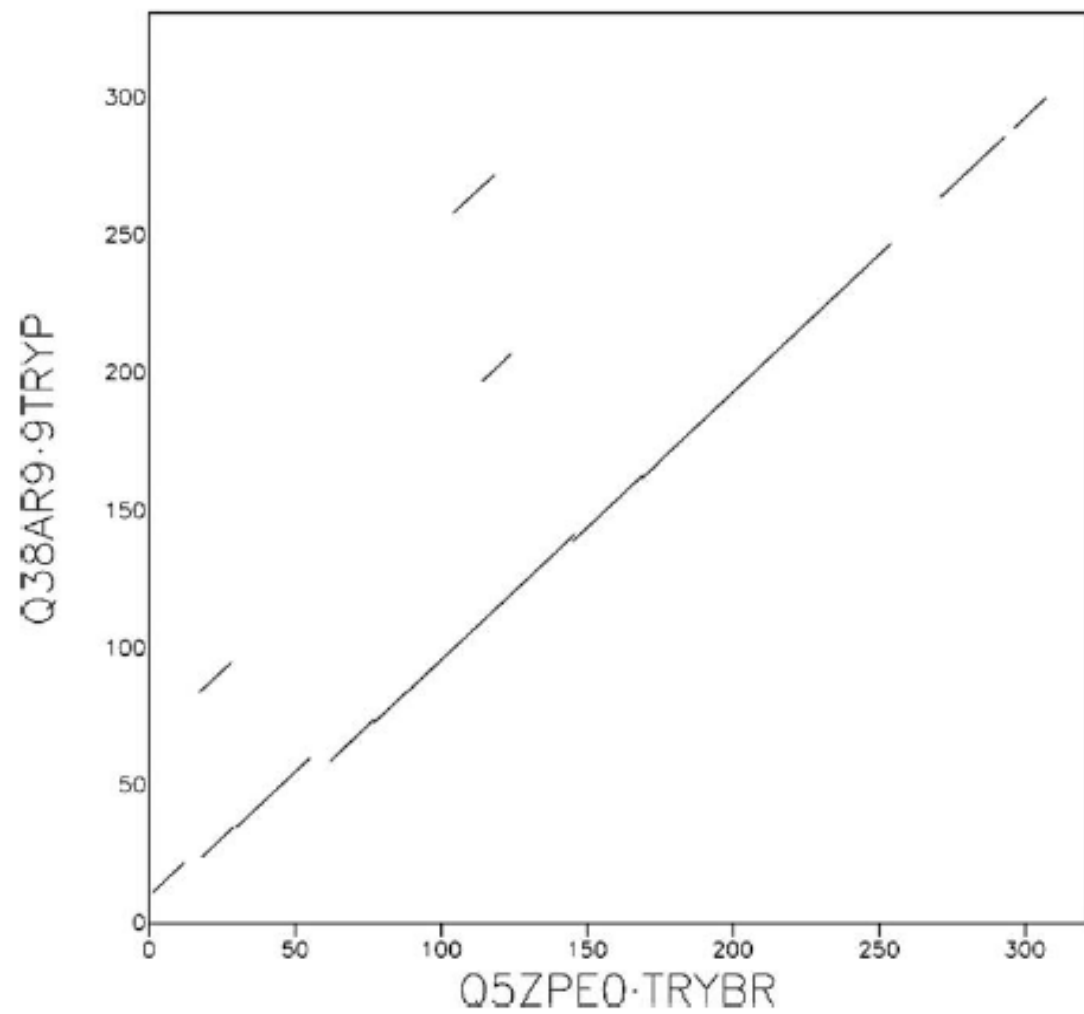


Observed percentage difference	Evolutionary distance in PAM
1	1
5	5
10	11
15	17
20	23
25	30
30	38
35	47
40	56
45	67
50	80
55	94
60	112
65	133
70	159
75	195
80	246
85	328 <- Twilight zone

Type of protein	Rate of Change (PAMs/100 myrs)	Theoretical Lookback Time (myrs)
Pseudogenes	400	45
Fibrinopeptides	90	200
Ig lambda chain C region	27	670
Somatotropin	25	800
Ribonucleases	21	850
Haemoglobin alpha chain	12	1500
Acid proteases	8	2300
Cytochrome c	4	5000
Adenylate kinase	3.2	6000
Glyceraldehyde-P dehydrogenase	2	9000
Glutamate dehydrogenase	0.9	18000
Useful lookback time = 360 PAMs		

# Dotmatcher: Q5ZPE0·TRYBR vs Q38AR9·9TRYF

(window size = 10, threshold = 23.00 24/02/07)



[Entry info] [Name and origin] [References] [Comments] [Cross-references] [Keywords] [Features] [Sequence]

Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.

Entry information	
Entry name	G3PG_LEIME
Primary accession number	Q27890
Secondary accession numbers	None
Integrated into Swiss-Prot on	July 15, 1998
Sequence was last modified on	January 23, 2007 (Sequence version 3)
Annotations were last modified on	July 24, 2007 (Entry version 56)

Name and origin of the protein	
Protein name	Glyceraldehyde-3-phosphate dehydrogenase, glycosomal
Synonyms	EC 1.2.1.12 GAPDH
Gene name	Name: GAPG
From	Leishmania mexicana (TaxID: 5665)

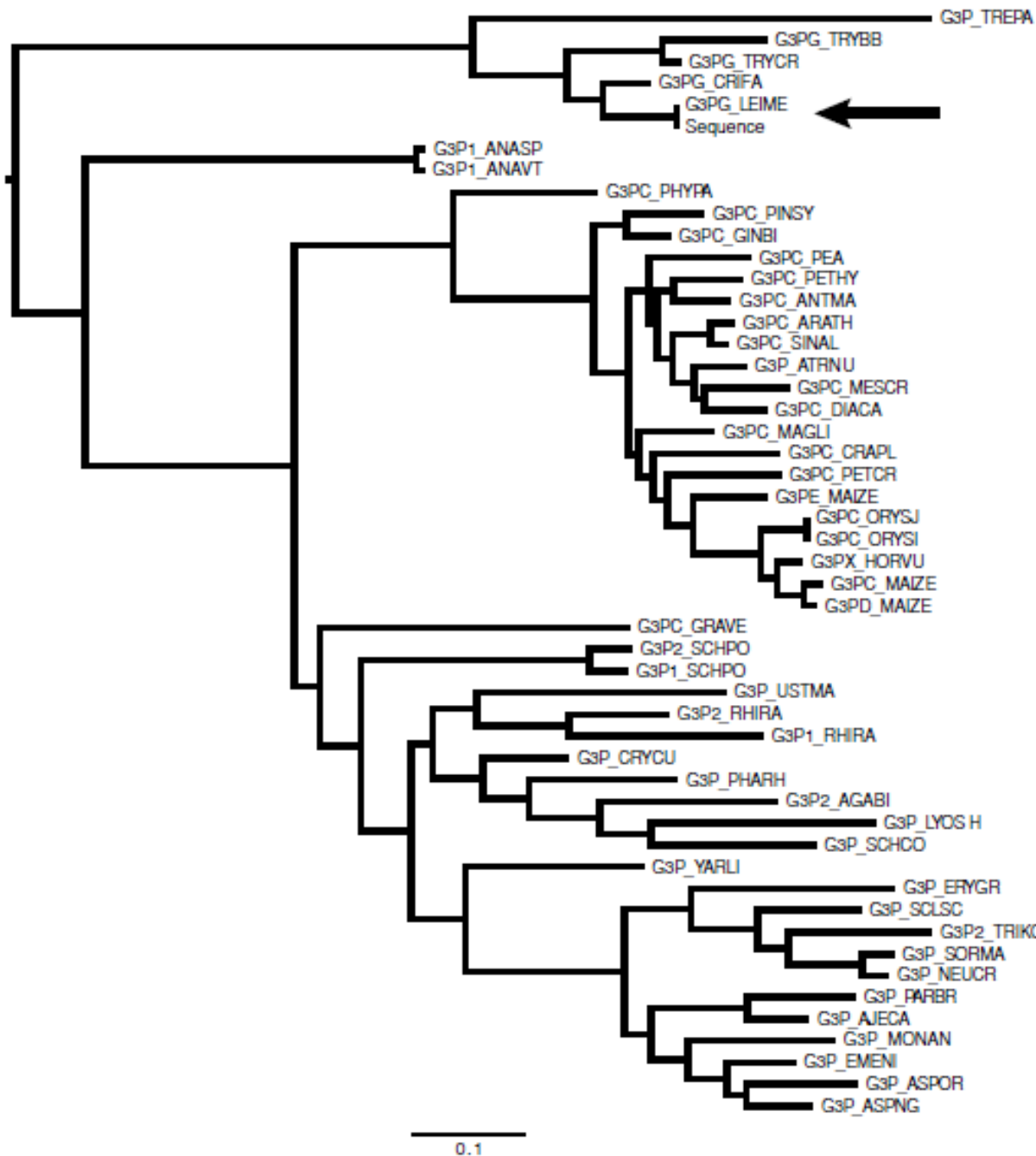
Taxonomic classification					
Sequence information					
Protein Length: 361 AA [This is the length of the unprocessed precursor]	Molecular weight: 39033 Da [This unprocessed precursor]				
10	20	30	40	50	60
MAPIKVGIHQ	YERIGRNVFD	AICDGLIGT	SIDVYVYVSH	STNAEYFATQ	HEEDTVIGRP
70	80	90	100	110	120
KYFVEAVKGG	PSVETADVLY	VNGIRIKDYK	AGRNPADLPH	GELGVGVVIE	STGLPTDLEK
130	140	150	160	170	180
AEGHKKDQAK	KYVISAPASG	GAKTIYNGVH	QREYSASRH	VVENASCTYH	CLAPIVYVLT
190	200	210	220	230	240
KENPDIRTGL	HTTINEVTAT	QRTVDGVSLK	DNROGAAAV	HTTPTTQAA	KAVGMPIPT
250	260	270	280	290	300
EKGLTGHSFR	VPTFQGVVVD	LTFRATRDTS	IQEISKATEK	AAQTYMKGIL	GFTDSELVSA
310	320	330	340	350	360
DFINDNRSQV	YDKATLQNS	LPSKERFFEV	VSMYDHSMAV	SHYVVLVRY	NAARDASSE



(c)

> X65226\_1 L.mexicana gap gene for glycosomal glyceraldehyde-3-phosphate dehydrogenase  
 MAPIKVGINGFGRIGRMVFQAI CDQGLIGTEIDVAVVDMSTNAEYFAYQMKHDTVHGRPKYTV  
 EAVKSSPSVETADVLLVNGHRIKCVKAQRNPADLPWGKLGVDYVIESTGLFTDKLKAEGHIKGG  
 AKKVVISAPASGGAKTIYVMGVNQHEYS PASHHVSNASCCTTNCLAPIVHVLTKENFGIETGLMTT  
 IHSYTATQKTVDGVSLKDWRGGRAAAVNIIPSTTGAAKAVGMVIPSTKGKLTGMSFRVPTPDVSV  
 VDLTFRATRDTSIQEIDKAIKAAQTYMKGILGFTDEELVSADFFINDNRSSVYDSKATLQNNLPG  
 EKRFKVVSWYDNEWAYSHRVVDLVRYMAAKDAASSKM\*

Q27890 in FASTA format



Accession Number

Q4Q6Z4  
O96423  
O96424  
O96426  
P22512  
A1YE37  
A1YE38  
O97105  
Q5ZPF6  
Q43311  
Q26753  
A1XIQ6  
Q76E10

Organism

*Leishmania major*  
*Crithidia fasciculata*  
*Herpetomonas pessoai*  
*Phytomonas* sp.  
*Trypanosoma brucei brucei*  
*Bodo saltans* (bodonid)  
*Parabodo caudatus* (bodonid)  
*Trypanosoma rangeli*  
*Trypanosoma cruzi*  
*Euglena gracilis*  
*Trypanoplasma borreli* (bodonid)  
*Sergeia podlipaevi* (trypanosomatid)  
*Symbiodinium* sp. (dinoflagellate)

Entry name

Q4Q6Z4\_LEIMA  
G3PG\_CRIFA  
O96424\_9TRYP  
O96426\_9TRYP  
G3PG\_TRYBB  
A1YE37\_9EUGL  
A1YE38\_9EUGL  
O97105\_TRYRA  
Q5ZPF6\_TRYCR  
Q43311\_EUGGR  
Q26753\_TRYBO  
A1XIQ6\_9TRYP  
Q76E10\_9DINO

```

#NEXUS

begin data;
  dimensions ntax = 14 nchar = 313;
  format interleave datatype = protein missing = X gap = -;

  matrix
LmjF30.2980   INGFGRI GRMVLQAICDQGLIGNEIDVVAVVD
Q4Q6Z4_LEIMA  INGFGRI GRMVLQAICDQGLIGNEIDVVAVVD
G3PG_CRIFA    INGFGRI GRMVFQSMCEDNVLGTELDVVAVVD
O96424_9TRYP  INGFGRI GRMVFQAMCEQGVLGKDFDVA VVD
O96426_9TRYP  INGFGRI GRNVFQAICEGNHLGTDIDVAVAD
G3PG_TRYBB    INGFGRI GRMVFQALCDDGLLGNEIDVVAVVD
A1YE37_9EUGL  INGFGRI GRMVFQAIADQGLLGKEIEVVAVVD
A1YE38_9EUGL  INGFGRI GRMVFQAICDQGLLGTEIDLVA VVD
O97105_TRYRA  INGFGRI GRMVFQAMCEADLLGTEIEVVAVVD
Q5ZPF6_TRYCR  IRPFGR I GRMVFQALCEDGLLGTEIDVVA VVD
Q43311_EUGGR  INGFGRI GRMVFQALCDQGLLGTTFDVVG VVD
Q26753_TRYBO  INGFGRI GRMVLQAICDQGLLGTEIDVVA VVV

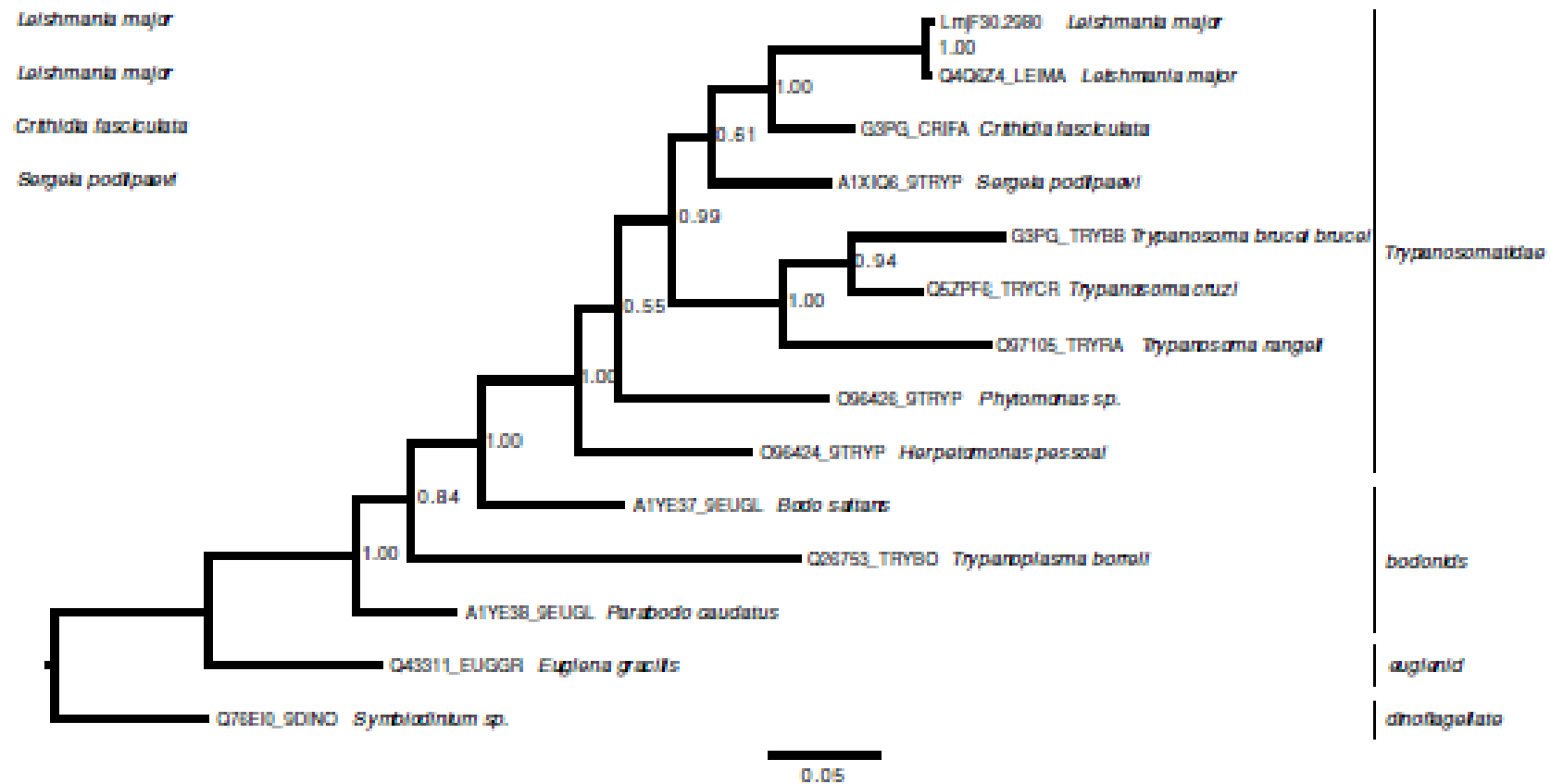
A1XIQ6_9TRYP  XXXXXXXXXXXXXXXXXXXXXXXXXXXX DVVAVVD
Q76EI0_9DINO  INGFGRI GRMVFQAICDQNLGTL DVVAVVD

[only a part of the alignment is shown for clarity]
;
end;

begin mrbayes;
  mcmc filename=g3pdh.mb;
  prset aamodelpr=mixed;
  lset rates=equal;
  prset brenspr=unconstrained:exponential(10.00);
  mcmc ngen=200000 samplefreq=100 printfreq=100;
  mcmc nruns=2 diagnfreq=1000 burnfrac=0.250000;
  mcmc stoprule=yes stopval=0.004000;
  mcmc nchains=4 swapfreq=1 temp=0.200000;
[ EDIT FILE AND REMOVE OUTCOMMENT BRACKETS AS NEEDED ]
[ mcmc; ]
[ sump burnin=XXX; ]
[ sumt contype=halfcompat burnin=XXX; ]
end;

```





# 10. Evrimsel model seçimi

Model	Free parameters	Base frequencies	Substitution code <sup>a</sup>	Reference
JC	0	equal	000000	(Jukes & Cantor, 1969)
F81	3	unequal	000000	(Felsenstein, 1981)
K80	1	equal	010010	(Kimura, 1980)
HKY	4	unequal	010010	(Hasegawa <i>et al.</i> , 1985)
TrNef	2	equal	010020	(Tamura & Nei, 1993)
TrN	5	unequal	010020	(Tamura & Nei, 1993)
K81	2	equal	012210	(Kimura, 1981)
K81uf	5	unequal	012210	(Kimura, 1981)
TIMef	3	equal	012230	(Posada, 2003)
TIM	6	unequal	012230	(Posada, 2003)
TVMef	4	equal	012314	(Posada, 2003)
TVM	7	unequal	012314	(Posada, 2003)
SYM	5	equal	012345	(Zharkikh, 1994)
GTR	8	unequal	012345	(Rodríguez <i>et al.</i> , 1990)



**ModelTest Server 1.0**  
 running ModelTest 3.8

[ModelTest Home](#)

---

**Input file**

Select likelihood scores file

**Likelihood Ratio Tests Options**

Enter confidence level for the LRTs

**Information Criterion options**

Model Selection Criterion

Enter sample size (for AICc and BIC)

Count branch lengths as parameters  Ignore them

Enter number of taxa (if count branch lengths)

Enter averaging confidence interval (0.01 - 1.00)

**Analysis**

Name this analysis

Contact: [posada@berkeley.edu](mailto:posada@berkeley.edu)  
 Citation:  
 - Posada D. 2005. ModelTest server: a web-based tool for the statistical selection of models of nucleotide substitution online. *Nucleic Acids Research* 33: W700-W703.  
 - Posada D and Crandall KA. 2002. Modeltest: testing the model of DNA substitution. *Bioinformatics* 18 (5): 817-818.

You are visitor number 3025 since January 1, 2006  
 This document last modified Wednesday December 20, 2006

Model	Data type	Reference
JTT	nuclear	(Jones <i>et al.</i> , 1992)
MtREV	mitochondrial	(Adachi & Hasegawa, 1996)
MtMam	mammal mitochondria	(Cao <i>et al.</i> , 1998)
MtArt	arthropod mitochondria	(Abascal <i>et al.</i> , 2007)
Dayhoff	nuclear	(Dayhoff <i>et al.</i> , 1978)
WAG	nuclear	(Whelan & Goldman, 2001)
RtREV	retroviral polymerase	(Dimmic <i>et al.</i> , 2002)
CpREV	chloroplastic	(Adachi <i>et al.</i> , 2000)
Blosum62	nuclear	(Henikoff & Henikoff, 1992)
VT	nuclear	(Muller & Vingron, 2000)

-----  
 \* AKAIKE INFORMATION CRITERION (AIC) \*  
 \*-----

Model selected: GTR+I+G  
 -lnL = 21148.9277  
 K = 41  
 AIC = 42379.8555  
 Base frequencies:  
 freqA = 0.3771  
 freqC = 0.2255  
 freqG = 0.1759  
 freqT = 0.2214  
 Substitution model:  
 Rate matrix  
 R(a) [A-C] = 3.4141  
 R(b) [A-G] = 5.0976  
 R(c) [A-T] = 3.5059  
 R(d) [C-G] = 0.4437  
 R(e) [C-T] = 14.9945  
 R(f) [G-T] = 1.0000  
 Among-site rate variation  
 Proportion of invariable sites (I) = 0.1595  
 Variable sites (G)  
 Gamma distribution shape parameter = 0.7279

Parameter <sup>a</sup>	Importance	Model-averaged estimates
fA	1.0000	0.3771
fC	1.0000	0.2255
fG	1.0000	0.1759
fT	1.0000	0.2214
TiTv	0.0000	–
rAC	1.0000	3.4152
rAG	1.0000	5.0991
rAT	1.0000	3.5073
rCG	1.0000	0.4440
rCT	1.0000	15.0010
pinv(I)	0.0000	–
alpha(G)	0.0022	0.4738
pinv(IG)	0.9978	0.1595
alpha(IG)	0.9978	0.7280

<sup>a</sup> (I): averaged using only +I models; (G): averaged using only +G models; (IG): averaged using only +I+G models.