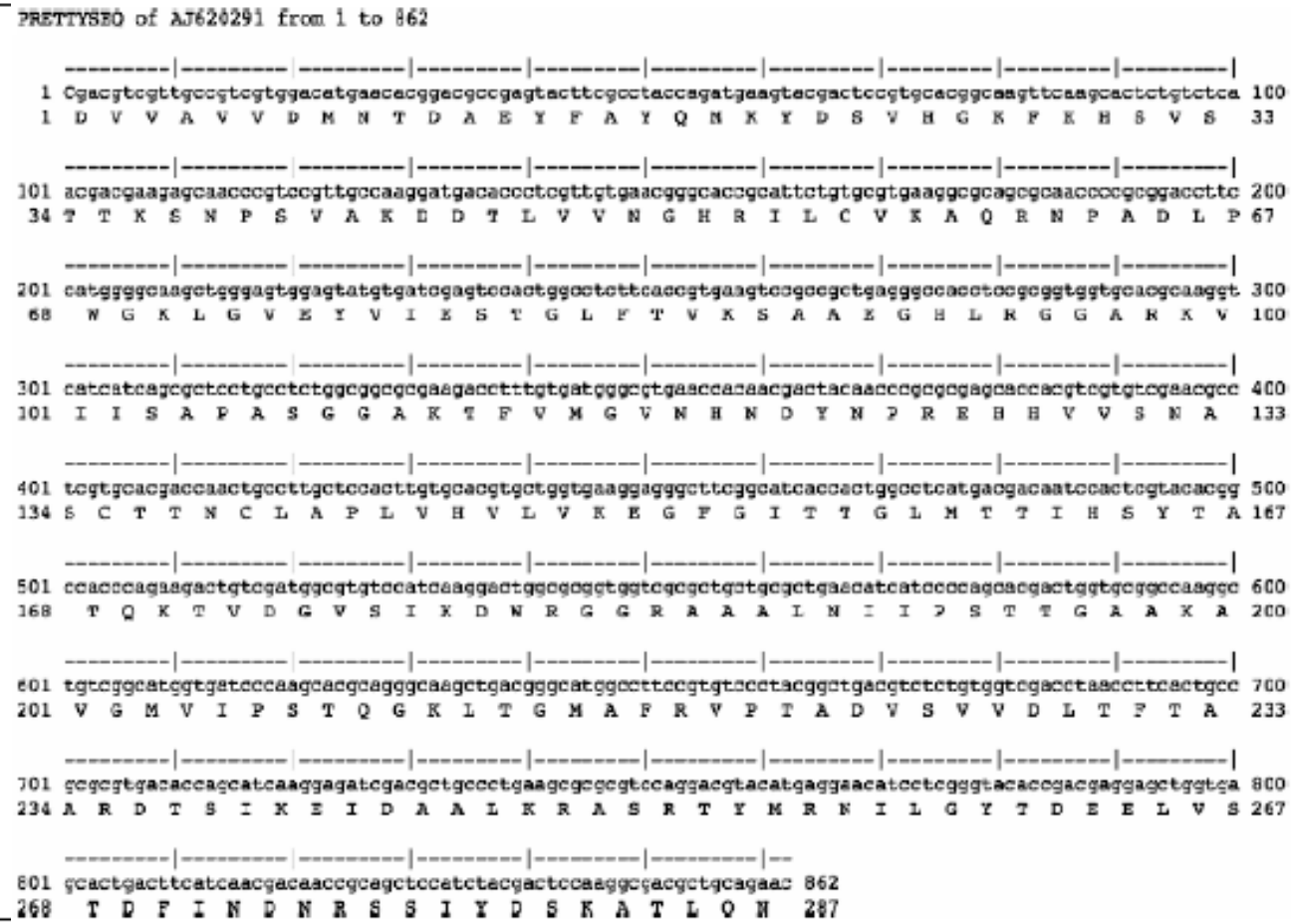
9. Protein dizilerinden filogenetik analiz

First Position	Second Position				Third Position
	U(T)	C	A	G	
U(T)	Phe	Ser	Tyr	Cys	U(T)
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U(T)
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U(T)
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U(T)
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G



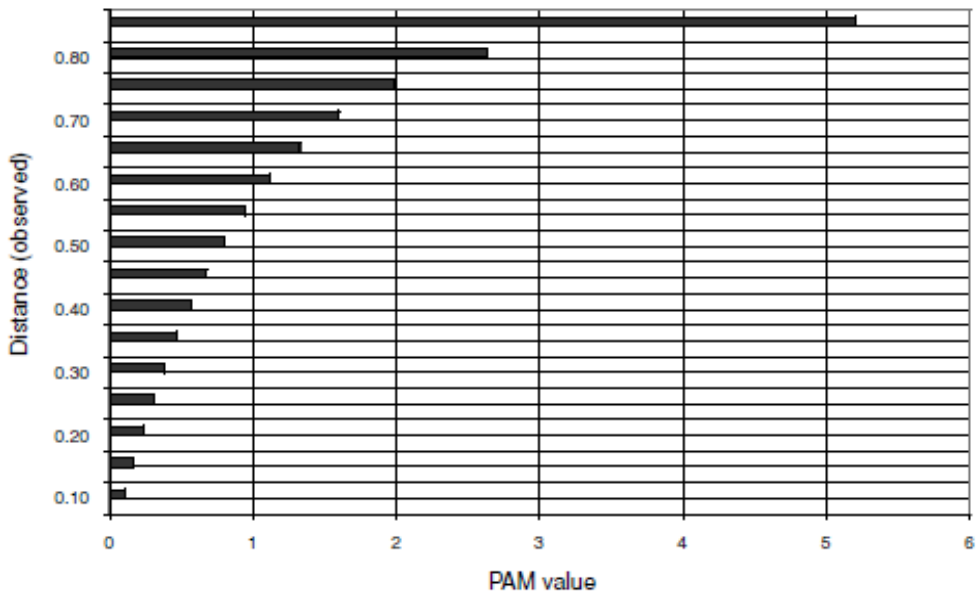
Figures and tables are taken from: Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press. For educational use only.

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

                                                    Score    E
Sequences producing significant alignments:         (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

                                                  Score    E
Sequences producing significant alignments:       (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
```

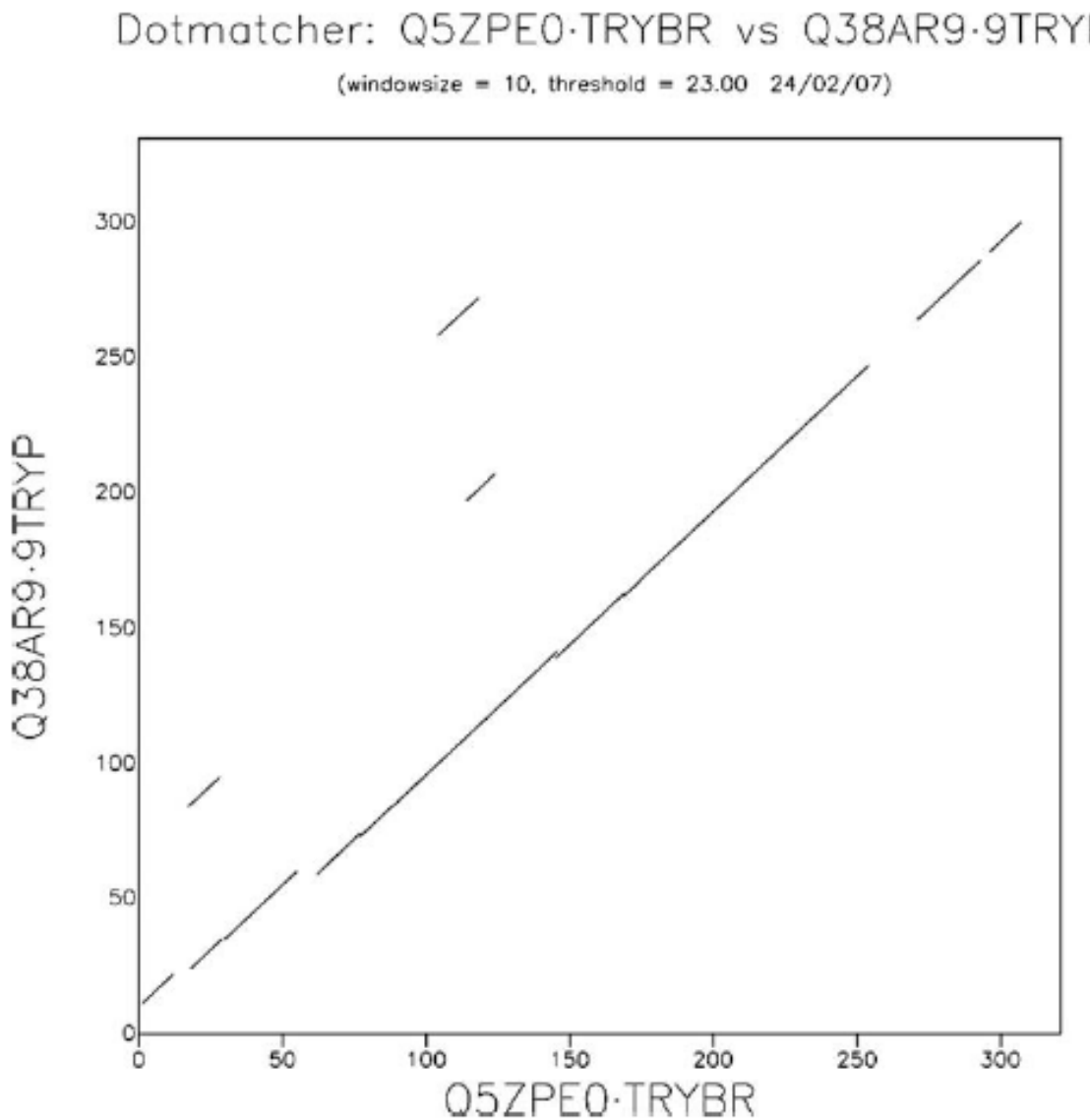
Figures and tables are taken from: Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press. For educational use only.

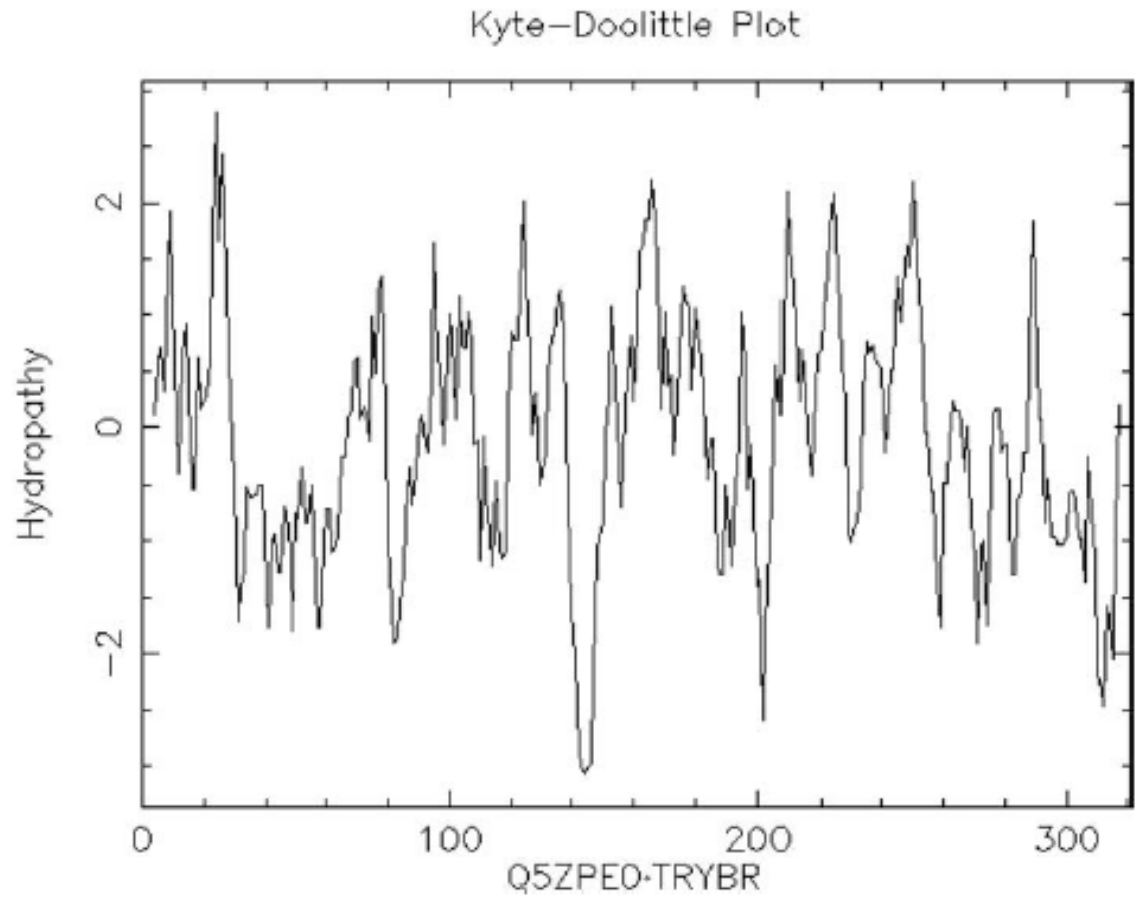


Observed percentage difference	Evolutionary distance in PAM
1	1
5	5
10	11
15	17
20	23
25	30
30	38
35	47
40	56
45	67
50	80
55	94
60	112
65	133
70	159
75	195
80	246
85	328 <- Twilight zone

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		

Figures and tables are taken from: Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press. For educational use only.





Figures and tables are taken from: Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press. For educational use only.

[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]
Taxonomy	
Protein	

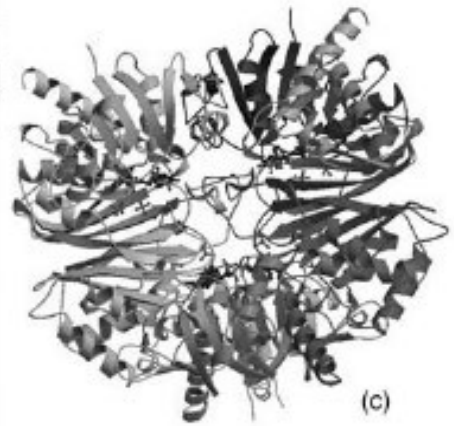
Sequence information

Length: **361 AA** [This is the length of the unprocessed precursor]

Molecular weight: **39033 Da** [This is the unprocessed precursor]

```
        10         20         30         40         50         60
MAPIKVGING FGRIGRMVFQ AICDQGLIGT EIDVVAVVDM STNAEYFAYQ MKHDTVHGRF

        70         80         90        100        110        120
KYTVEAVKSS PSVETADVLV VNGHRIKCVK AQRNPADLPW GKLGVDYVIE STGLFTDKLK

       130        140        150        160        170        180
AEGHIKGGAK KVVISAPASG GAKTIVMGVN QHEYSPASHH VVSNASCTTN CLAPIVHVLT

       190        200        210        220        230        240
KENFGIETGL MTTIHSYTAT QKTVDGVSLK DWRGGRAAAV NIIPSTTGAA KAVGMVIPST

       250        260        270        280        290        300
KGKLTGMSFR VPTPDVSVVD LTFRATRDTS IQEIDKAIKK AAQTYMKGIL GFTDEELVSA

       310        320        330        340        350        360
DFINDNRSSV YDSKATLQNN LPGEKRFFKV VSWYDNEWAY SHRVVDLVRY MAAKDAASSK

M
```

(b)



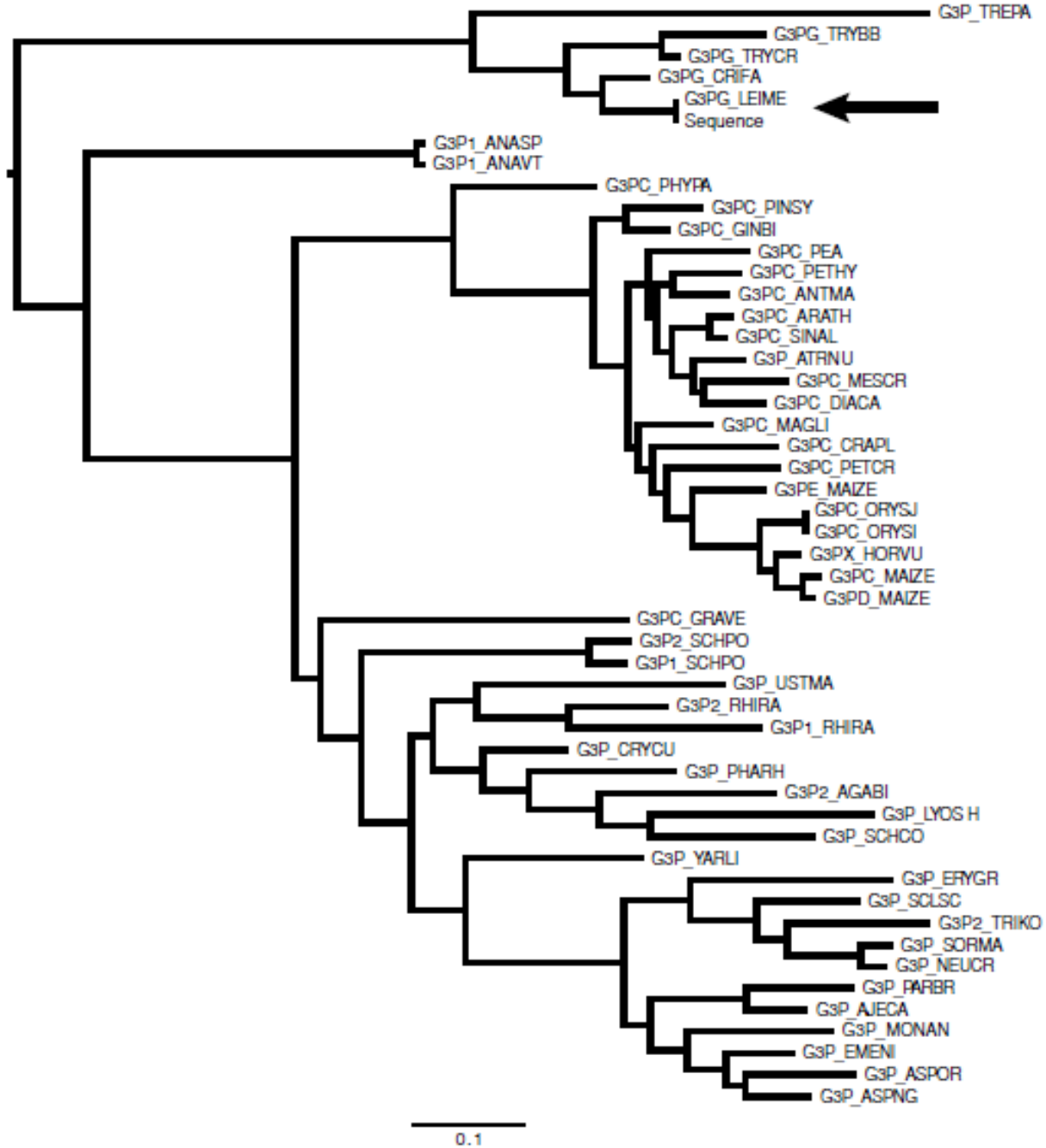
(c)

Q27890 in FASTA format

> X65226_1 L.mexicana gap gene for glycosomal glyceraldehyde-3-phosphate dehydrogenase

MAPIKVGINGFGRIGRMVFQAICDQGLIGTEIDVVAVVDMSTNAEYFAYQMKHDTVHGRPKYTV
EAVKSSPSVETADVLVVNGHRIKCVKAQRNPADLPWGKLGVDYVIESTGLFTDKLKAEGHIKGG
AKKVVISAPASGGAKTIVMGVNQHEYSPASHHVVSNASCTTNCLAPIVHVLTKENFGIETGLMTT
IHSYTATQKTVDGVSLKDWRGGRAAAVNIIPSTTGAAKAVGMVIPSTKGKLTGMSFRVPTPDVSV
VDLTFRATRDTSIQEIDKAIKKAAQTYMKGILGFTDEELVSADFINDNRSSVYDSKATLQNNLPG
EKRFFKVVSWYDNEWAYSHRVVDLVRYMAAKDAASSKM*

Figures and tables are taken from: Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press. For educational use only.

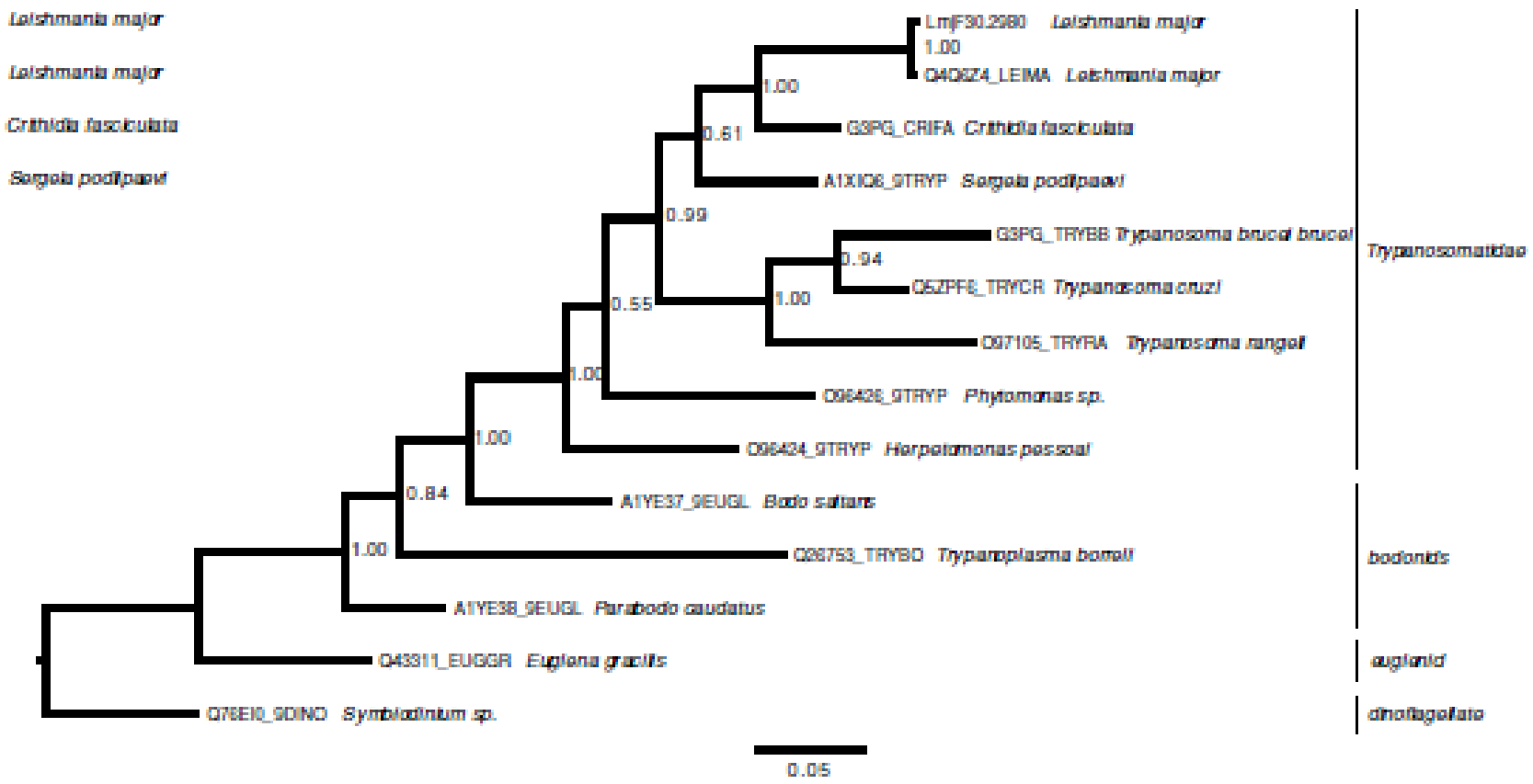


Accession Number	Organism	Entry name
Q4Q6Z4	*Leishmania major*	Q4Q6Z4_LEIMA
O96423	*Crithidia fasciculate*	G3PG_CRIFA
O96424	*Herpetomonas pessoai*	O96424_9TRYP
O96426	*Phytomonas* sp.	O96426_9TRYP
P22512	*Trypanosoma brucei brucei*	G3PG_TRYBB
A1YE37	*Bodo saltans* (bodonid)	A1YE37_9EUGL
A1YE38	*Parabodo caudatus* (bodonid)	A1YE38_9EUGL
O97105	*Trypanosoma rangeli*	O97105_TRYRA
Q5ZPF6	*Trypanosoma cruzi*	Q5ZPF6_TRYCR
Q43311	*Euglena gracilis*	Q43311_EUGGR
Q26753	*Trypanoplasma borreli* (bodonid)	Q26753_TRYBO
A1XIQ6	*Sergeia podlipaevi* (trypanosomatid)	A1XIQ6_9TRYP
Q76EI0	*Symbiodinium* sp. (dinoflagellate)	Q76EI0_9DINO

Figures and tables are taken from: Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press. For educational use only.

```
#NEXUS

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

  matrix
  LmjF30.2980    INGFGRIGRMVLQAICDQGLIGNEIDVVAVVD
  Q4Q6Z4_LEIMA   INGFGRIGRMVLQAICDQGLIGNEIDVVAVVD
  G3PG_CRIFA     INGFGRIGRMVFQSMCEDNVLGTELDVVAVVD
  O96424_9TRYP   INGFGRIGRMVFQAMCEQGVLGKDFDVVAVVD
  O96426_9TRYP   INGFGRIGRNVFQAICEGNHLGTDIDVVAVAD
  G3PG_TRYBB     INGFGRIGRMVFQALCDDGLLGNEIDVVAVVD
  A1YE37_9EUGL   INGFGRIGRMVFQAIADQGLLGKEIEVVAVVD
  A1YE38_9EUGL   INGFGRIGRMVFQAICDQGLLGTEIDLVAVVD
  O97105_TRYRA   INGFGRIGRMVFQAMCEADLLGTEIEVVAVVD
  Q5ZPF6_TRYCR   IRPFGRIGRMVFQALCEDGLLGTEIDVVAVVD
  Q43311_EUGGR   INGFGRIGRMVFQALCDQGLLGTTFDVVGVVD
  Q26753_TRYBO   INGFGRIGRMVLQAICDQGLLGTEIDVVAVVV
  A1XIQ6_9TRYP   XXXXXXXXXXXXXXXXXXXXXXXXXDVVAVVD
  Q76EI0_9DINO   INGFGRIGRMVFQAICDQNLLGTKLDVVGVVD

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

begin mrbayes;
  mcmcp filename=g3pdh.mb;
  prset aamodelpr=mixed;
  lset rates=equal;
  prset brlenspr=unconstrained:exponential(10.00);
  mcmcp ngen=200000 samplefreq=100 printfreq=100;
  mcmcp nruns=2 diagnfreq=1000 burninfrac=0.250000;
  mcmcp stoprule=yes stopval=0.004000;
  mcmcp 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;
```

Figures and tables are taken from: Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press. For educational use only.



Figures and tables are taken from: Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press. For educational use only.

10. Evrimsel model seçimi

Model	Free parameters	Base frequencies	Substitution code[a]	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 *et al.*, 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 *et al.*, 1990)

Figures and tables are taken from: Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press. For educational use only.

Model	Data type	Reference
JTT	nuclear	(Jones *et al.*, 1992)
MtREV	mitochondrial	(Adachi & Hasegawa, 1996)
MtMam	mammal mitochondria	(Cao *et al.*, 1998)
MtArt	arthropod mitochondria	(Abascal *et al.*, 2007)
Dayhoff	nuclear	(Dayhoff *et al.*, 1978)
WAG	nuclear	(Whelan & Goldman, 2001)
RtREV	retroviral polymerase	(Dimmic *et al.*, 2002)
CpREV	chloroplastic	(Adachi *et al.*, 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
```

Figures and tables are taken from: Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press. For educational use only.

Parameter[a]	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

[a] (I): averaged using only +I models; (G): averaged using only +G models; (IG): averaged using only +I+G models.

Figures and tables are taken from: Lemey, P., Salemi, M., & Vandamme, A. M. (Eds.). (2009). *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press. For educational use only.