

**BAST 数据库相似性搜索**姓名 陈耿佳 学号 1301214752 组号 G01C

1. 以人血红蛋白 beta 亚基 (HBB\_HUMAN) 为检测序列, 搜索 Swiss-Prot 数据库中所有哺乳动物的血红蛋白, 找出前 10 个高分匹配, 列表说明搜索结果。

程序: blastp 参数: 默认 数据库: Swiss-Prot Organism: mammals

No.	Accession	Name	Species	Total score	Query cover	E value	Ident
1	P68871.2	HBB_HUMAN	Homo sapiens	301	100%	5e-105	100%
2	P02024.2	HBB_GORGO	Gorilla gorilla gorilla	300	100%	2e-104	99%
3	P02025.1	HBB_HYLLA	Hylobates lar	294	99%	3e-102	99%
4	P02032.1	HBB_SEMEN	Semnopithecus entellus	291	99%	2e-101	97%
5	P19885.2	HBB_COLPO	Colobus polykomos	291	100%	4e-101	96%
6	Q6WN22.3	HBB_ATEPA	Ateles paniscus	290	100%	1e-100	96%
7	P68232.2	HBB_ATEGE	Ateles geoffroyi	289	100%	2e-100	96%
8	P68222.2	HBB_MACFU	Macaca fuscata fuscata	289	100%	2e-100	95%
9	Q6WN27.3	HBB_PITPI	Pithecia pithecia	289	100%	3e-100	96%
10	Q6WN25.3	HBB_LAGLA	Lagothrix lagotricha	289	100%	3e-100	96%

2. 以大鼠血红蛋白 alpha 亚基(HBA\_RAT)为检测序列, 用 BlastP、PSI-Blast 和 DELAT-Blast 分别搜索 RefSeq 数据库中珠蛋白家族成员 (注意选择适当的计分矩阵、设置适当的 E 值)。

检测序列: NP\_001007723.1

数据库: Reference proteins (refseq\_protein)

Organism: Rattus

Exclude: Exclude Models (XM/XP), Exclude Uncultured/environmental sample sequences

Entrez Query: globin

Algorithm parameters: 默认

Blastp 搜索结果:

No.	Accession	Description	Total score	Query cover	E value	Ident
1	NP_001007723.1	hemoglobin, alpha 2	292	100%	2e-105	100%
2	NP_001013875.1	globin, alpha	228	100%	4e-80	77%
3	NP_001166316.1	hemoglobin, zeta	160	100%	2e-53	51%
4	NP_001106694.1	hemoglobin subunit beta-2	103	97%	3e-31	39%
5	NP_150237.1	hemoglobin subunit beta-1	98.6	97%	2e-29	39%
6	NP_742090.1	hemoglobin, gamma A	89	97%	8e-26	33%
7	NP_001019976.1	hemoglobin subunit epsilon	89	97%	9e-26	34%
8	NP_570100.1	cytoglobin	54.7	96%	3e-13	25%
9	NP_067599.1	myoglobin	49.7	100%	1e-11	25%
10	NP_203523.2	neuroglobin	32.7	92%	2e-05	26%

## PSI-Blast 搜索结果 (iteration 2):

No.	Accession	Description	Total score	Query cover	E value	Ident
1	NP_001166316.1	hemoglobin, zeta	187	100%	5e-64	51%
2	NP_001013875.1	globin, alpha	186	100%	1e-63	77%
3	NP_037228.1	hemoglobin subunit alpha-1/2	180	100%	2e-61	99%
4	NP_001007723.1	hemoglobin, alpha 2	180	100%	3e-61	100%
5	NP_742090.1	hemoglobin, gamma A	179	97%	1e-60	32%
6	NP_150237.1	hemoglobin subunit beta-1	177	97%	3e-60	36%
7	NP_001008890.1	hemoglobin, epsilon 1	177	97%	6e-60	37%
8	NP_001106694.1	hemoglobin subunit beta-2	176	97%	8e-60	37%
9	NP_001019976.1	hemoglobin subunit epsilon	176	97%	1e-59	33%
10	NP_942071.1	hemoglobin, beta adult major chain	173	97%	1e-58	35%
11	NP_067599.1	myoglobin	157	100%	5e-52	25%
12	NP_570100.1	cytoglobin	156	96%	3e-51	25%
13	NP_203523.2	neuroglobin	129	95%	4e-41	24%

## DELTA-Blast 搜索结果:

No.	Accession	Description	Total score	Query cover	E value	Ident
1	NP_001166316.1	hemoglobin, zeta	155	100%	9e-52	51%
2	NP_001013875.1	globin, alpha	149	100%	3e-49	77%
3	NP_001008890.1	hemoglobin, epsilon 1	144	97%	4e-47	37%
4	NP_037228.1	hemoglobin subunit alpha-1/2	141	100%	5e-46	99%
5	NP_001007723.1	hemoglobin, alpha 2	140	100%	1e-45	100%
6	NP_001019976.1	hemoglobin subunit epsilon	136	97%	5e-44	33%
7	NP_742090.1	hemoglobin, gamma A	134	97%	2e-43	32%
8	NP_001106694.1	hemoglobin subunit beta-2	132	97%	1e-42	37%
9	NP_150237.1	hemoglobin subunit beta-1	132	97%	1e-42	36%
10	NP_942071.1	hemoglobin, beta adult major chain	132	97%	2e-42	35%
11	NP_570100.1	cytoglobin	122	96%	3e-38	25%
12	NP_067599.1	myoglobin	104	100%	1e-31	25%
13	NP_203523.2	neuroglobin	102	92%	7e-31	24%

BlastP 是简单的 blast 算法, 以序列相似性为依据, 可能会遗漏远源同源序列。PSI-Blast 可以在当前 blast 结果的基础上分析序列特征如保守位点等, 构建位点特异性计分矩阵, 重新进行 blast, 即使是序列差异较大的远源同源序列也可以搜索到。DELTA-Blast 是基于保守结构域数据库构建位点特异性计分矩阵进行 blast, 也可以搜索到远源同源序列。

- 以大鼠血红蛋白 alpha 亚基 (HBA\_RAT) 为检测序列, 用 tBlastN 搜索 RefSeq 数据库中珠蛋白家族 mRNA 序列。参考 Hardison 论文和大鼠基因组数据库 RGD 中相关信息, 下载非预测序列, 提取其编码区序列, 进行多序列比对, 并构建系统发育树, 分析结果。

检测序列: NP\_001007723.1

数据库: Reference RNA sequences (refseq\_rna)

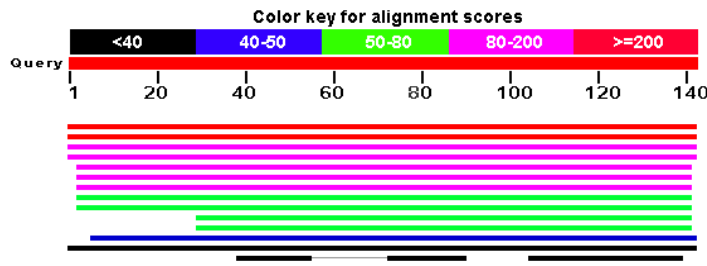
Organism: Rattus

Exclude: Exclude Models (XM/XP), Exclude Uncultured/environmental sample sequences

Entrez Query: globin

Algorithm parameters: 默认

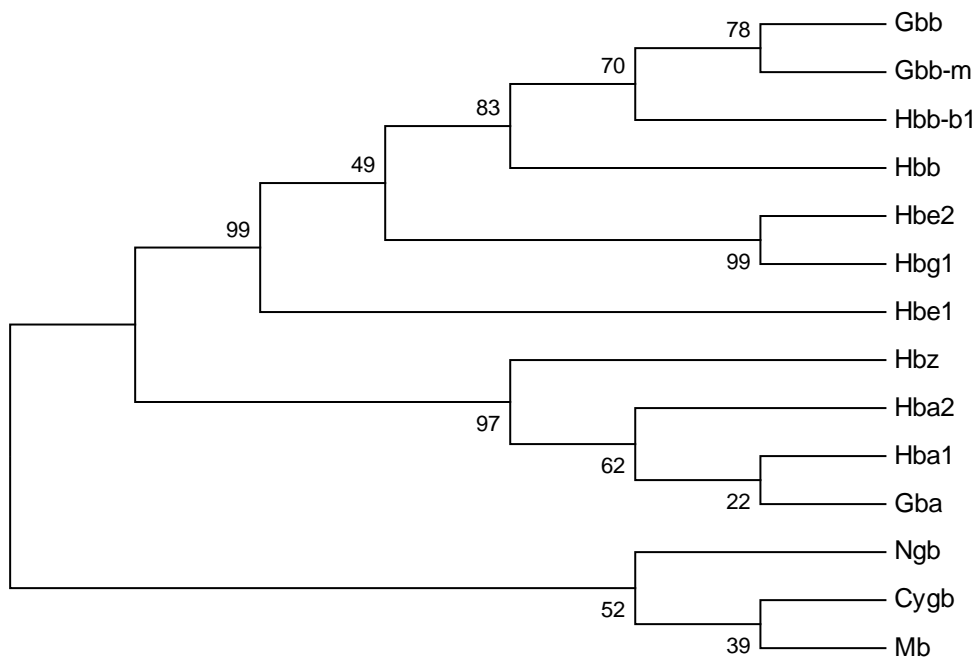
Blast 结果图形显示如下：



去掉最后一个结果，得 14 个序列如下表所示：

No	Accession	Name	Description	Total score	Query cover	E value	Ident
1	NM_001007722.1	Hba2	hemoglobin, alpha 2	233	100%	2e-81	82%
2	NM_013096.1	Hba1	hemoglobin, alpha 1	230	100%	3e-80	82%
3	NM_001013853.1	Gba	globin, alpha	191	100%	1e-65	66%
4	NM_001172845.1	Hbz	hemoglobin, zeta	137	100%	5e-44	46%
5	NM_001008890.1	Hbe1	hemoglobin, epsilon 1	81.3	97%	8e-23	34%
6	NM_001111269.1	Gbb	beta-globin	82	97%	8e-23	34%
7	NM_001113223.1	Gbb-m	beta globin minor gene	82	97%	1e-22	34%
8	NM_198776.1	Hbb-b1	hemoglobin, beta adult major chain	78.2	97%	2e-21	31%
9	NM_033234.1	Hbb	hemoglobin, beta (Hbb)	77.4	97%	4e-21	33%
10	NM_001024805.1	Hbe2	hemoglobin, epsilon 2 (Hbe2)	70.9	78%	4e-19	31%
11	NM_172093.1	Hbg1	hemoglobin, gamma A (Hbg1)	68.9	78%	2e-18	31%
12	NM_130744.2	Cygb	cytoglobin (Cygb)	40.8	96%	6e-08	22%
13	NM_021588.2	Mb	myoglobin (Mb)	31.6	100%	8e-05	21%
14	NM_033359.3	Ngb	neuroglobin (Ngb)	17.7	24%	3.5	32%

下载上述 14 个序列，在 WebLab 上用 coderet 程序提取其编码区序列，在 MEGA 6.0 软件上进行多序列比对，构建系统发育树，采用 Neighbor-Joining 算法，用 Bootstrap 法进行评估（500 次重复），所构建系统发育树如下图所示：



从系统发育树可以看到，Hemoglobin 各成员之间亲缘关系较近，分歧时间较近；而 Hemoglobin 各成员与 cytoglobin、myoglobin、neuroglobin 之间亲缘关系较远，分歧时间较早。

4. 搜索 RefSeq 数据库中人、小鼠和大鼠三个物种珠蛋白家族 mRNA 序列，下载非预测序列，提取其编码区序列，进行多序列比对，并构建系统发育树，分析结果。

程序: tblastn

检测序列: NP\_000508.1 (HBA\_HUMAN)

数据库: Reference RNA sequences (refseq\_rna)

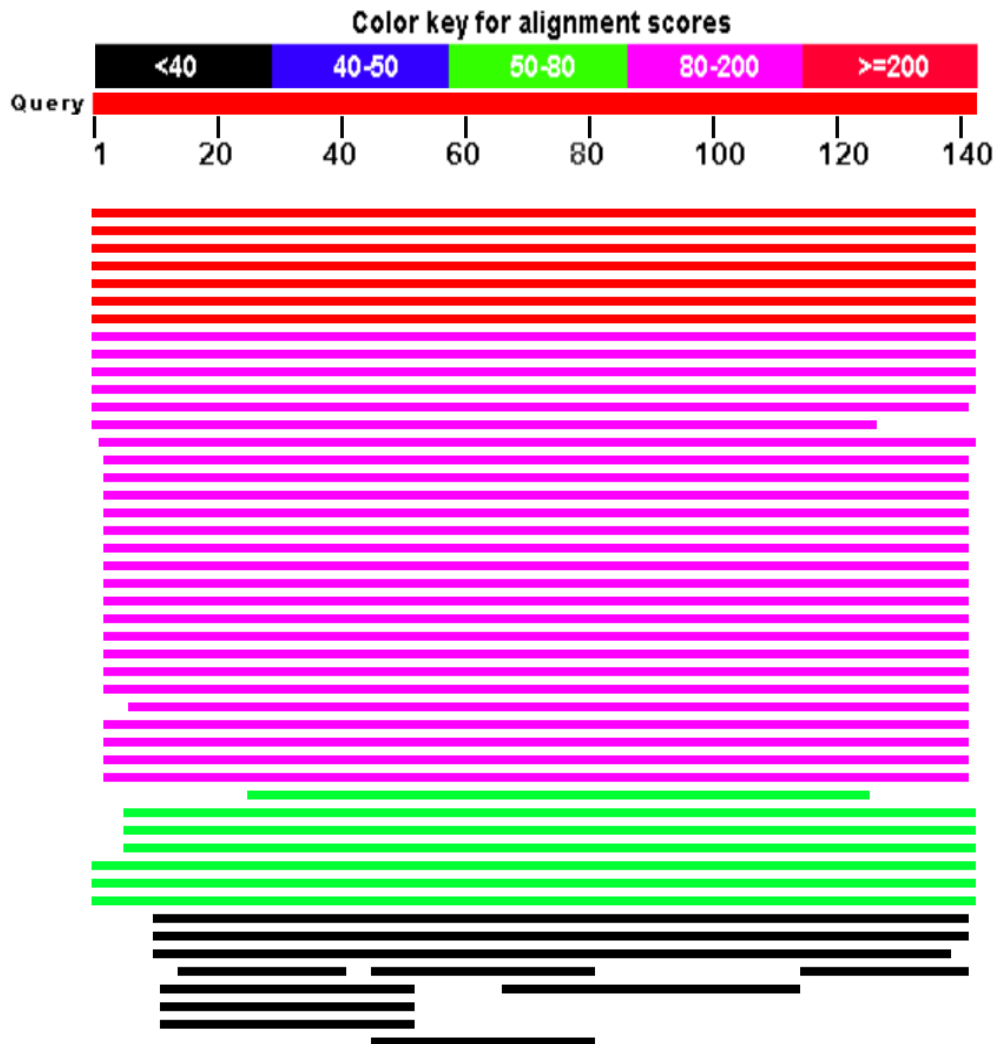
Organism: Rattus, Human, Mouse

Exclude: Exclude Models (XM/XP), Exclude Uncultured/environmental sample sequences

Entrez Query: globin

Algorithm parameters: 默认

Blast 结果图形显示如下:

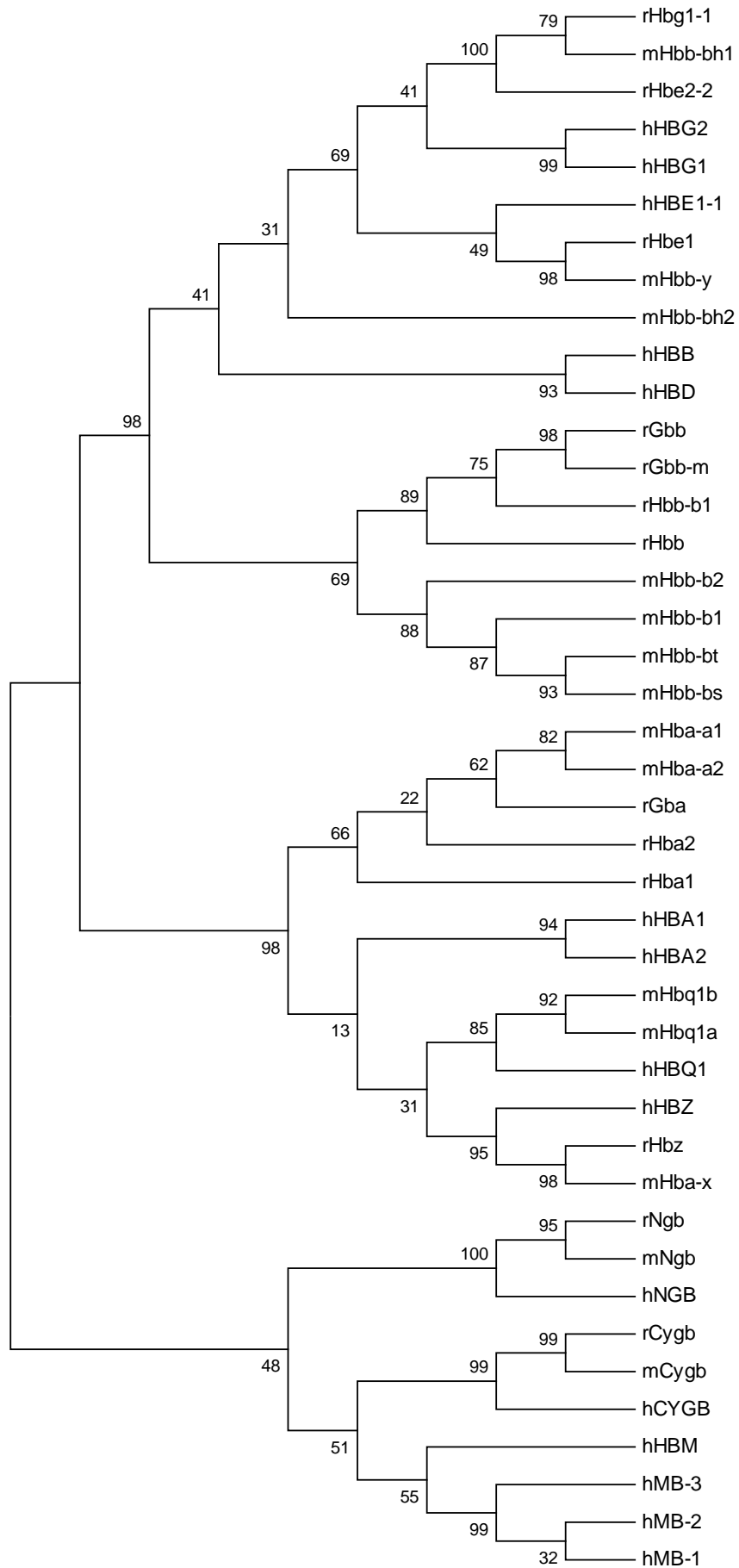


去掉其中非珠蛋白家族的结果，得 42 个序列如下表所示:

No	Accession	Name	Description	Species *	Total score	E value
1	NM_000558.3	hHBA1	hemoglobin, alpha 1	H	286	1e-100
2	NM_000517.4	hHBA2	hemoglobin, alpha 2	H	286	2e-100
3	NM_008218.2	mHba-a1	hemoglobin alpha, adult chain 1	M	255	2e-88
4	NM_001083955.1	mHba-a2	hemoglobin alpha, adult chain 2	M	255	2e-88
5	NM_001013853.1	rGba	globin, alpha (LOC287167)	R	223	1e-76
6	NM_001007722.1	rHba2	hemoglobin, alpha 2	R	211	2e-71
7	NM_013096.1	rHba1	hemoglobin, alpha 1	R	209	2e-70
8	NM_001033981.3	mHbq1b	hemoglobin, theta 1B	M	185	3e-60
9	NM_005331.4	hHBQ1	hemoglobin, theta 1	H	181	5e-59
10	NM_001172845.1	rHbz	hemoglobin, zeta	R	176	3e-57
11	NM_005332.2	hHBZ	hemoglobin, zeta	H	175	6e-57
12	NM_010405.4	mHba-x	hemoglobin X, alpha-like embryonic chain	M	171	4e-55
13	NM_175000.2	mHbq1a	hemoglobin, theta 1A	M	152	6e-48
14	NM_001003938.3	hHBM	hemoglobin, mu	H	134	2e-41
15	NM_001111269.1	rGbb	beta-globin (LOC689064)	R	116	6e-34
16	NM_001113223.1	rGbb-m	beta globin minor gene (LOC100134871)	R	115	7e-34
17	NM_000518.4	hHBB	hemoglobin, beta	H	115	1e-33
18	NM_001278161.1	mHbb-b1	hemoglobin, beta adult major chain	M	114	2e-33
19	NM_000519.3	hHBD	hemoglobin, delta	H	115	3e-33
20	NM_000184.2	hHBG2	hemoglobin, gamma G	H	113	3e-33
21	NM_000559.2	hHBG1	hemoglobin, gamma A	H	112	1e-32
22	NM_033234.1	rHbb	hemoglobin, beta	R	110	7e-32
23	NM_198776.1	rHbb-b1	hemoglobin, beta adult major chain	R	110	8e-32
24	NM_001008890.1	rHbe1	hemoglobin, epsilon 1	R	108	1e-31
25	NM_016956.3	mHbb-b2	hemoglobin, beta adult minor chain	M	108	4e-31
26	NM_008221.4	mHbb-y	hemoglobin Y, beta-like embryonic chain	M	107	2e-30
27	NM_008220.5	mHbb-bt	hemoglobin, beta adult t chain	M	106	3e-30
28	NM_001201391.1	mHbb-bs	hemoglobin, beta adult s chain	M	106	4e-30
29	NM_005330.3	hHBE1-1	hemoglobin, epsilon 1	H	102	4e-28
30	NM_172093.1	rHbg1-1	hemoglobin, gamma A	R	98.2	1e-27
31	NM_001024805.1	rHbe2-2	hemoglobin, epsilon 2	R	96.3	6e-27
32	NM_008219.3	mHbb-bh1	hemoglobin Z, beta-like embryonic chain	M	95.1	4e-26
33	NM_001127686.1	mHbb-bh2	hemoglobin beta, bh2	M	85.5	2e-22
34	NM_134268.4	hCYGB	cytoglobin	H	68.2	3e-15
35	NM_130744.2	rCygb	cytoglobin	R	63.2	1e-13
36	NM_030206.4	mCygb	cytoglobin	M	61.6	4e-13
37	NM_203378.1	hMB-3	myoglobin, transcript variant 3	H	51.2	6e-10
38	NM_203377.1	hMB-2	myoglobin, transcript variant 2	H	51.2	7e-10
39	NM_005368.2	hMB-1	myoglobin, transcript variant 1	H	51.2	7e-10
40	NM_033359.3	rNgb	neuroglobin	R	30.8	0.006
41	NM_021257.3	hNGB	neuroglobin	H	30.4	0.008
42	NM_022414.2	mNgb	neuroglobin	M	30.4	0.009

\*H = Homo sapiens, M = Mus musculus, R = Rattus norvegicus

下载上述 42 个序列，在 WebLab 上用 coderet 程序提取其编码区序列，在 MEGA 6.0 软件上进行多序列比对，构建系统发育树，采用 Neighbor-Joining 算法，用 Bootstrap 法进行评估（500 次重复），所构建系统发育树如下图所示：



从上面系统发育树可以看到，在各个大分支中三个物种的成员均有存在，表明珠蛋白各大分支的分化是在三个物种分歧之前就已经发生，即 Hemoglobin 各成员与 cytoglobin、myoglobin、neuroglobin 之间的分歧在三个物种的分歧之前。Hemoglobin 的主要分支  $\alpha$ 、 $\beta$  链的分歧也在三个物种分歧之前发生。而人 HBA1、HBA2 的分歧则是在三个物种分歧之前，在人的进化分支历程中经由复制产生。

5. 以你课题相关的蛋白质为检测序列，搜索 Swiss-Prot 或 RefSeq 数据库，分析搜索结果，阅读相关文献，找出同一物种中可能的旁系同源序列和近缘物种中可能的直系同源序列。

实验室主要关注的蛋白质是 MG53 (Mitsugumin-53)，也称为 TRIM72 (Tripartite motif-containing protein 72)。

#### 1) 直系同源序列

检测序列: NP\_001008275.2 (TRIM72\_HUMAN)

数据库: UniprotKB/Swiss-Prot (Swissprot)

Oganism: 不限

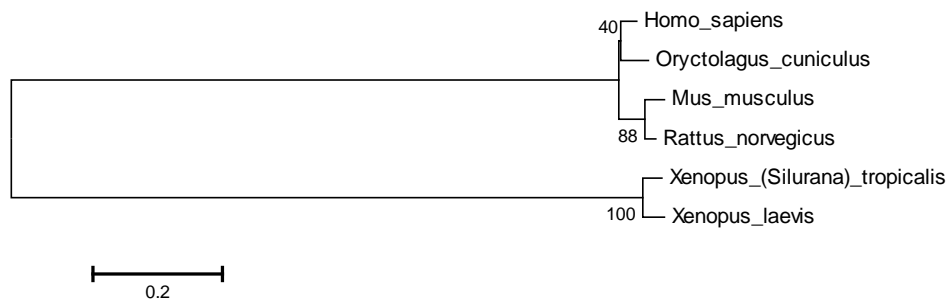
算法: DELTA-Blast

参数: 默认

搜索结果: 前 6 个高分匹配选项即为 MG53 在各个物种中的同源基因，列表如下:

No	Accession	Name	Species	Total score	Query cover	E value	Ident
1	Q6ZMU5.2	TRIM72	Homo sapiens	684	100%	0	100%
2	Q1XH17.1	TRIM72	Mus musculus	677	100%	0	91%
3	A0JPQ4.1	TRIM72	Rattus norvegicus	673	100%	0	91%
4	Q1XH18.1	TRIM72	Oryctolagus cuniculus	664	100%	0	94%
5	Q640S6.1	TRIM72	Xenopus (Silurana) tropicalis	545	98%	0	58%
6	Q6PGR9.1	TRIM72	Xenopus laevis	543	98%	0	57%

下载上述 6 个分子的氨基酸序列，在 MEGA 6.0 软件上进行多序列比对，构建系统发育树，采用 Neighbor-Joining 算法，用 Bootstrap 法进行评估 (500 次重复)，所构建系统发育树如下图所示:



可见 MG53 在进化过程中较为保守，其在不同物种中的分歧时间与物种进化历程一致。

## 2) 旁系同源序列

检测序列: NP\_001008275.2 (TRIM72\_HUMAN)

数据库: UniprotKB/Swiss-Prot (Swissprot)

Oganism: human

Exclude: Exclude Models (XM/XP), Exclude Uncultured/environmental sample sequences

Entrez Query: TRIM

参数: 默认

用 Blastp 得到 38 个匹配结果, 均为 TRIM 家族成员;

用 PSI-Blast (iteration 2) 得到 38 个匹配结果, 均为 TRIM 家族成员;

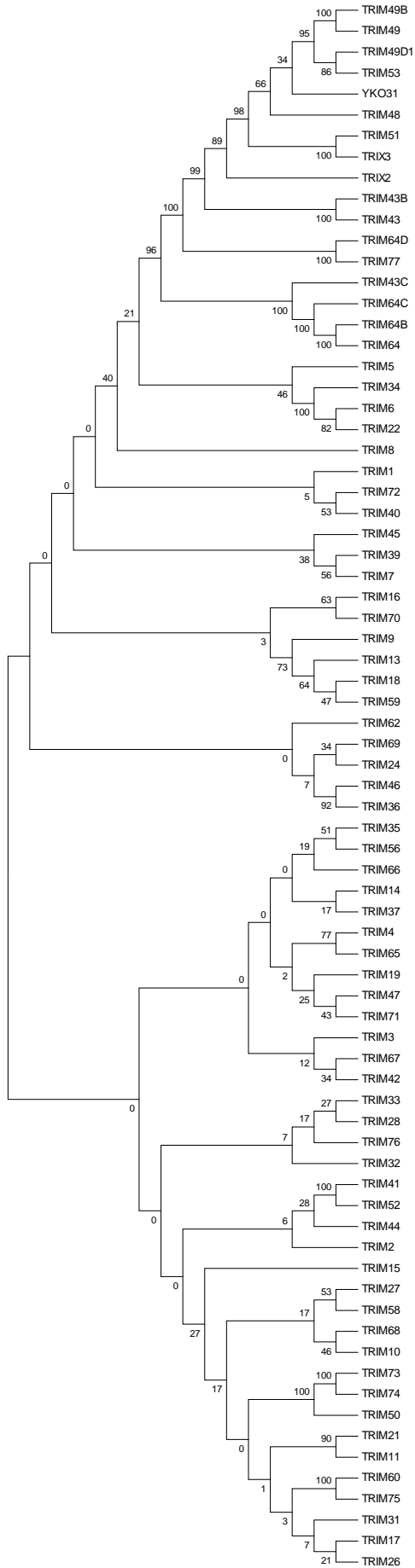
用 DELTA-Blast 得到 75 个匹配结果, 均为 TRIM 家族成员, 列表如下:

No	Accession	Name	Total score	Query cover	E value	Ident
1	Q6ZMU5.2	TRIM72	684	100%	0	100%
2	Q9HCM9.2	TRIM39	387	98%	4e-131	28%
3	P19474.1	TRIM21	380	97%	9e-129	28%
4	Q9BVG3.1	TRIM62	361	97%	2e-121	26%
5	Q86XT4.1	TRIM50	354	96%	9e-119	33%
6	P14373.1	TRIM27	346	98%	3e-115	28%
7	Q8NG06.2	TRIM58	343	98%	3e-114	30%
8	Q6AZZ1.1	TRIM68	339	97%	7e-113	25%
9	Q86WT6.2	TRIM69	338	97%	3e-112	28%
10	Q495X7.2	TRIM60	330	98%	2e-109	26%
11	Q96F44.2	TRIM11	327	99%	2e-108	28%
12	Q9C029.2	TRIM7	327	93%	6e-108	28%
13	A6NK02.2	TRIM75	311	98%	4e-102	24%
14	Q9UDY6.3	TRIM10	303	94%	7e-99	26%
15	Q9C030.1	TRIM6	297	97%	1e-96	25%
16	Q9UPQ4.2	TRIM35	281	96%	2e-90	26%
17	Q9C019.1	TRIM15	276	96%	6e-89	24%
18	Q8WV44.3	TRIM41	336	92%	1e-88	25%
19	Q9Y577.1	TRIM17	275	95%	4e-88	26%
20	Q9C037.2	TRIM4	270	93%	3e-86	26%
21	Q9C035.1	TRIM5	265	97%	4e-84	23%
22	Q8IYM9.1	TRIM22	258	97%	2e-81	24%
23	Q9BYJ4.2	TRIM34	232	97%	7e-72	21%
24	A6NI03.3	TRIM64B	230	94%	1e-71	23%
25	A6NGJ6.4	TRIM64	227	94%	3e-70	23%
26	A6NLI5.3	TRIM64C	221	94%	3e-68	22%
27	O95361.3	TRIM16	209	89%	2e-62	19%
28	Q12899.1	TRIM26	311	94%	2e-61	26%
29	A8MWY9.2	TRIM64D	201	93%	2e-60	21%
30	A6NCK2.3	TRIM43B	200	94%	3e-60	24%
31	Q96BQ3.1	TRIM43	199	94%	1e-59	24%
32	Q9BSJ1.2	TRIM51	194	93%	6e-58	23%
33	A6NMD1.2	TRIM43C	191	94%	5e-57	23%



No	Accession	Name	Total score	Query cover	E value	Ident
34	Q6PJ69.3	TRIM65	188	97%	3e-55	21%
35	A6NDI0.1	TRIM49B	180	93%	7e-53	25%
36	P0CI25.1	TRIM49	179	93%	1e-52	25%
37	C9J1S8.1	TRIM49D1	179	94%	2e-52	25%
38	A6NDQ2.2	TRIX3	178	93%	4e-52	21%
39	A6NCD1.4	TRIX2	176	93%	2e-51	22%
40	Q14142.2	TRIM14	206	55%	4e-50	24%
41	Q309B1.3	TRIM70	168	35%	2e-49	22%
42	Q3C1W6.2	YKO31	168	95%	1e-48	23%
43	I1YAP6.2	TRIM77	166	93%	7e-48	21%
44	O15344.1	TRIM18	233	67%	2e-42	25%
45	Q9BZY9.2	TRIM31	137	53%	5e-38	24%
46	Q86UV7.1	TRIM73	128	50%	4e-36	28%
47	Q86UV6.1	TRIM74	127	50%	9e-36	28%
48	O60858.2	TRIM13	130	53%	2e-35	24%
49	Q9UJV3.3	TRIM1	293	71%	1e-34	23%
50	O94972.2	TRIM37	122	53%	1e-31	24%
51	O75382.2	TRIM3	117	44%	3e-30	22%
52	Q8N3K9.3	TRIM76	186	69%	2e-29	25%
53	Q9C040.1	TRIM2	114	33%	2e-29	22%
54	Q8IWR1.1	TRIM59	111	27%	4e-29	30%
55	Q9C026.1	TRIM9	272	90%	2e-27	18%
56	Q6P9F5.3	TRIM40	102	53%	7e-27	25%
57	Q6ZTA4.3	TRIM67	253	93%	1e-25	18%
58	Q13049.2	TRIM32	102	45%	3e-25	25%
59	A6NNF0.3	TRIM53	94.4	51%	2e-23	28%
60	Q96LD4.2	TRIM47	231	84%	3e-22	17%
61	Q8IWZ4.2	TRIM48	87	31%	9e-22	30%
62	Q9BZR9.2	TRIM8	86.3	36%	3e-20	24%
63	Q9BRZ2.3	TRIM56	81.6	43%	1e-18	23%
64	Q96A61.1	TRIM52	137	23%	2e-18	32%
65	Q9UPN9.3	TRIM33	190	52%	2e-18	22%
66	Q13263.5	TRIM28	117	46%	3e-18	19%
67	P29590.3	TRIM19	80.1	24%	4e-18	25%
68	O15016.4	TRIM66	143	52%	7e-16	21%
69	Q7Z4K8.2	TRIM46	179	87%	2e-15	23%
70	O15164.3	TRIM24	170	55%	6e-15	19%
71	Q9NQ86.2	TRIM36	179	53%	7e-15	15%
72	Q8IWZ5.2	TRIM42	125	29%	5e-13	20%
73	Q9H8W5.2	TRIM45	124	50%	7e-13	19%
74	Q96DX7.1	TRIM44	80.9	11%	4e-11	38%
75	Q2Q1W2.1	TRIM71	93.6	38%	1e-09	24%

下载上述 75 个分子的氨基酸序列，在 MEGA 6.0 软件上进行多序列比对，构建系统发育树，采用 Neighbor-Joining 算法，用 Bootstrap 法进行评估（500 次重复），所构建系统发育树如下图所示：



## 3) 旁系同源序列——tBlastN

检测序列: NP\_001008275.2 (TRIM72\_HUMAN)

数据库: Reference RNA sequences (refseq\_rna)

Oganism: human

Exclude: Exclude Models (XM/XP), Exclude Uncultured/environmental sample sequences

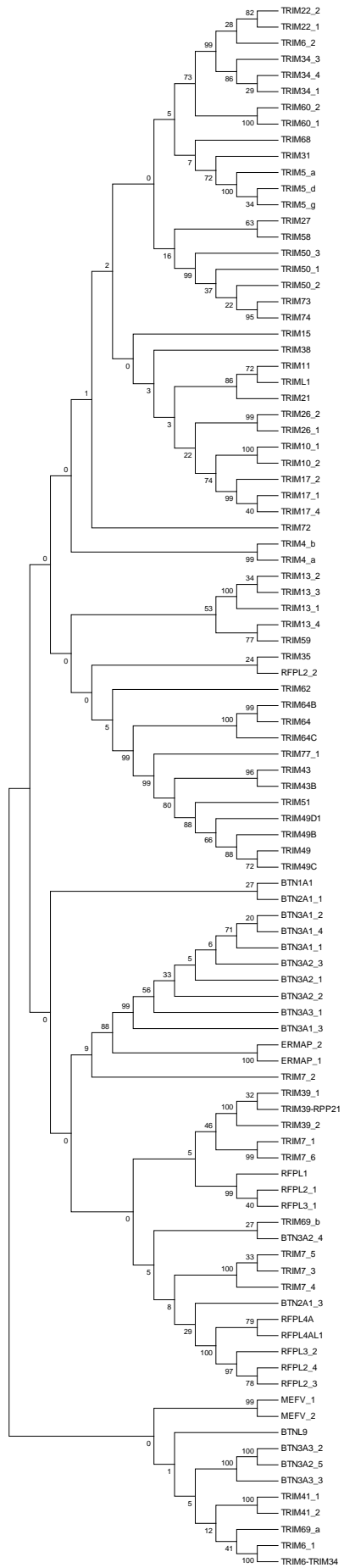
参数: 默认

得到 100 个匹配结果 (最大匹配数), 去掉非编码 RNA (假基因、反义链等), 得 97 个结果, 列表如下:

No	Accession	Name	Total score	Query cover	E value	Ident
1	NM_001008274.3	TRIM72	901	100%	0	95%
2	NM_001281451.1	TRIM50_1	240	96%	3e-70	31%
3	NM_178125.3	TRIM50_2	240	96%	5e-70	31%
4	NM_001281450.1	TRIM50_3	235	96%	3e-68	31%
5	NM_018207.2	TRIM62	171	97%	1e-44	25%
6	NM_003141.3	TRIM21	167	98%	2e-44	27%
7	NM_203293.2	TRIM7_1	169	93%	3e-44	28%
8	NM_182985.3	TRIM69_a	158	97%	2e-41	27%
9	NM_152620.2	TRIM60_2	157	98%	7e-41	24%
10	NM_001258025.1	TRIM60_1	157	98%	9e-41	24%
11	NM_006510.4	TRIM27	157	98%	3e-40	27%
12	NM_015431.3	TRIM58	157	97%	8e-40	29%
13	NM_172016.2	TRIM39_2	156	98%	9e-40	25%
14	NM_021253.3	TRIM39_1	154	98%	4e-39	25%
15	NM_018073.6	TRIM68	147	97%	1e-36	25%
16	NM_001003818.2	TRIM6_1	142	94%	5e-35	25%
17	NM_058166.4	TRIM6_2	142	94%	5e-35	25%
18	NM_145214.2	TRIM11	139	92%	4e-34	28%
19	NM_006778.3	TRIM10_1	134	94%	2e-32	27%
20	NM_033034.2	TRIM5_a	133	94%	3e-32	24%
21	NM_001024940.2	TRIM17_2	131	93%	4e-32	27%
22	NM_016102.3	TRIM17_1	131	93%	5e-32	27%
23	NM_001199573.1	TRIM22_2	130	93%	3e-31	25%
24	NM_006074.4	TRIM22_1	129	93%	4e-31	25%
25	NM_033091.2	TRIM4_b	127	93%	2e-30	27%
26	NM_006355.4	TRIM38	127	93%	4e-30	24%
27	NM_033017.3	TRIM4_a	124	93%	3e-29	26%
28	NM_178556.3	TRIML1	118	96%	5e-28	25%
29	NM_001164397.1	TRIM64B	116	93%	2e-27	24%
30	NM_001136486.1	TRIM64	115	93%	4e-27	24%
31	NM_171982.3	TRIM35	116	93%	1e-26	26%
32	NM_018538.3	ERMAP_2	114	31%	2e-26	36%
33	NM_001017922.1	ERMAP_1	114	31%	2e-26	36%
34	NM_080745.3	TRIM69_b	110	42%	2e-25	31%
35	NM_001206631.1	TRIM64C	109	93%	2e-25	23%
36	NM_020358.2	TRIM49	110	92%	5e-25	27%
37	NM_001206627.1	TRIM49D1	109	94%	6e-25	25%
38	NM_001195234.1	TRIM49C	109	92%	9e-25	27%
39	NM_001206626.1	TRIM49B	107	94%	1e-24	26%
40	NM_138800.1	TRIM43	108	93%	2e-24	25%
41	NM_001164464.1	TRIM43B	107	93%	3e-24	24%
42	NM_033549.4	TRIM41_1	149	91%	6e-24	25%
43	NM_052828.2	TRIM10_2	103	72%	1e-22	28%
44	NM_032681.3	TRIM51	96.7	93%	1e-20	22%
45	NM_203297.1	TRIM7_2	97.1	59%	2e-20	26%
46	NM_198924.3	TRIM73	95.5	50%	2e-20	28%

No	Accession	Name	Total score	Query cover	E value	Ident
47	NM_198853.1	TRIM74	95.5	49%	2e-20	28%
48	NM_203295.1	TRIM7_4	95.9	59%	4e-20	26%
49	NM_203294.1	TRIM7_5	95.5	59%	4e-20	26%
50	NM_203296.1	TRIM7_3	95.1	59%	6e-20	26%
51	NM_001199119.1	TRIM39-RPP21	91.7	72%	5e-19	24%
52	NM_033229.2	TRIM15	92.4	99%	5e-19	24%
53	NM_033093.2	TRIM5_d	92	25%	6e-19	34%
54	NM_033092.2	TRIM5_g	161	52%	1e-18	34%
55	NM_005798.4	TRIM13_1	90.5	53%	3e-18	25%
56	NM_052811.3	TRIM13_2	90.5	53%	3e-18	25%
57	NM_001007278.2	TRIM13_4	90.5	53%	3e-18	25%
58	NM_213590.2	TRIM13_3	90.5	53%	3e-18	25%
59	NM_001134855.1	TRIM17_4	87.4	38%	1e-17	31%
60	NM_001003819.3	TRIM6-TRIM34	238	75%	3e-17	25%
61	NM_197974.2	BTN3A3_2	86.7	27%	4e-17	31%
62	NM_006994.4	BTN3A3_1	86.7	27%	4e-17	31%
63	NM_001242803.1	BTN3A3_3	86.3	27%	4e-17	31%
64	NM_130390.2	TRIM34_3	83.6	24%	6e-17	32%
65	NM_033342.3	TRIM7_6	84.3	24%	9e-17	37%
66	NM_001003827.1	TRIM34_4	151	50%	2e-16	32%
67	NM_021616.5	TRIM34_1	151	50%	2e-16	32%
68	NM_152547.4	BTNL9	82.8	40%	5e-16	29%
69	NM_001145008.1	BTN3A1_3	82.4	27%	8e-16	30%
70	NM_007048.5	BTN3A1_1	82.4	27%	8e-16	30%
71	NM_021026.2	RFPL1	126	37%	1e-15	35%
72	NM_194441.2	BTN3A1_2	82	27%	1e-15	30%
73	NM_001145009.1	BTN3A1_4	82	27%	1e-15	30%
74	NM_001732.2	BTN1A1	79.3	34%	7e-15	33%
75	NM_001242783.1	TRIM26_2	122	43%	8e-15	34%
76	NM_003449.4	TRIM26_1	122	43%	8e-15	34%
77	NM_001145014.1	RFPL4A	117	48%	1e-14	31%
78	NM_000243.2	MEFV_1	114	55%	2e-14	28%
79	NM_007028.3	TRIM31	77.8	24%	2e-14	31%
80	NM_001277397.1	RFPL4AL1	115	48%	3e-14	30%
81	NM_173084.2	TRIM59	77.8	23%	3e-14	34%
82	NM_001198536.1	MEFV_2	113	55%	5e-14	28%
83	NM_201627.2	TRIM41_2	117	69%	6e-14	24%
84	NM_001197246.2	BTN3A2_2	75.5	27%	1e-13	29%
85	NM_001197248.2	BTN3A2_4	75.5	27%	1e-13	29%
86	NM_001197249.2	BTN3A2_5	75.5	27%	1e-13	29%
87	NM_007047.4	BTN3A2_1	75.5	27%	1e-13	29%
88	NM_001197247.2	BTN3A2_3	75.5	27%	1e-13	29%
89	NM_006605.3	RFPL2_1	119	37%	3e-13	34%
90	NM_006604.2	RFPL3_2	121	37%	4e-13	33%
91	NM_001159546.1	RFPL2_4	119	37%	4e-13	34%
92	NM_001098535.1	RFPL3_1	121	37%	4e-13	33%
93	NM_001159545.1	RFPL2_3	118	42%	5e-13	30%
94	NM_001098527.2	RFPL2_2	118	42%	8e-13	30%
95	NM_001197233.2	BTN2A1_3	72.4	27%	9e-13	30%
96	NM_007049.4	BTN2A1_1	72.4	27%	1e-12	30%
97	NM_001146162.1	TRIM77_1	71.6	94%	1e-12	21%

下载上述 97 个 RNA 序列, 在 WebLab 上用 coderet 程序提取其编码区序列, 在 MEGA 6.0 软件上进行多序列比对, 构建系统发育树, 采用 Neighbor-Joining 算法, 用 Bootstrap 法进行评估 (500 次重复), 所构建系统发育树如下图所示:



## 4) 直系/旁系同源序列——tBlastN

检测序列: NP\_001008275.2 (TRIM72\_HUMAN)

数据库: Reference RNA sequences (refseq\_rna)

Organism: human, mouse, rat

Exclude: Exclude Models (XM/XP), Exclude Uncultured/environmental sample sequences

参数: 默认

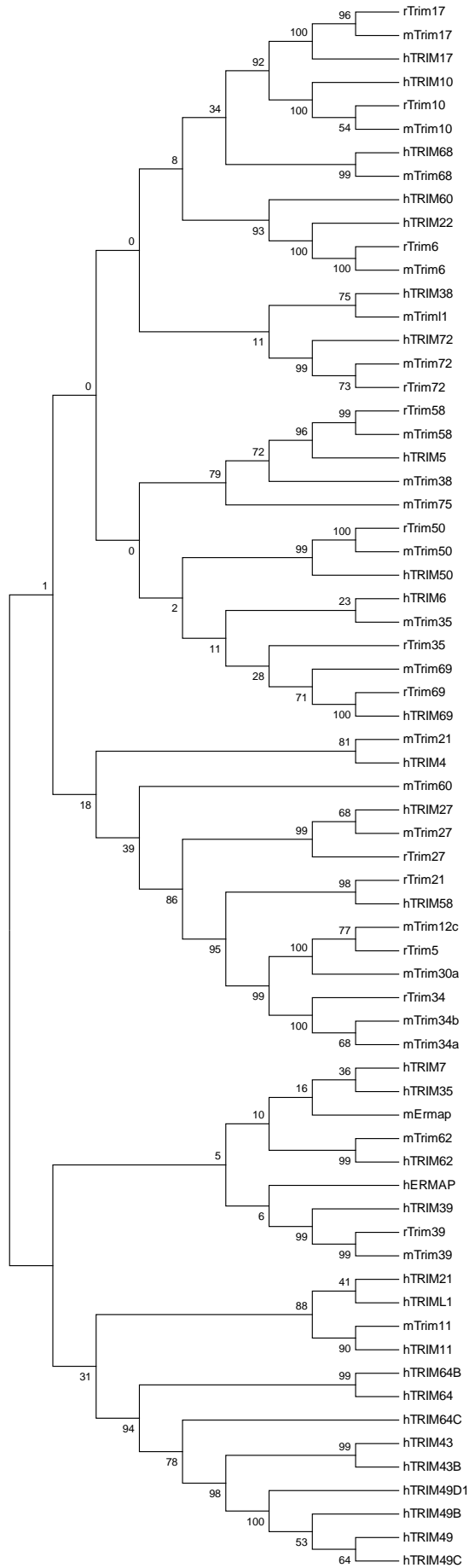
得到 100 个匹配结果 (最大匹配数), 去掉非编码 RNA (假基因、反义链等), 合并重复转录本 (取第一个), 取前 67 个匹配结果, 列表如下:

No	Accession	Name	Species *	Total score	Query cover	E value	Ident
1	NM_001008274.3	hTRIM72	H	901	100%	0	95%
2	NM_001079932.3	mTrim72	M	833	100%	0	86%
3	NM_001077675.1	rTrim72	R	829	100%	0	86%
4	NM_181080.1	rTrim50	R	238	97%	1e-70	31%
5	NM_178240.2	mTrim50	M	237	97%	1e-70	30%
6	NM_178125.3	hTRIM50	H	240	96%	9e-70	31%
7	NM_178110.2	mTrim62	M	171	97%	2e-44	24%
8	NM_018207.2	hTRIM62	H	171	97%	3e-44	25%
9	NM_003141.3	hTRIM21	H	167	98%	4e-44	27%
10	NM_203293.2	hTRIM7	H	169	93%	6e-44	28%
11	NM_001082572.1	rTrim21	R	159	96%	8e-42	26%
12	NM_001013160.1	rTrim69	R	159	97%	2e-41	25%
13	NM_001191641.1	rTrim58	R	157	98%	4e-41	28%
14	NM_182985.3	hTRIM69	H	158	97%	4e-41	27%
15	NM_001134974.1	rTrim27	R	158	98%	6e-41	27%
16	NM_080510.2	mTrim69	M	157	97%	6e-41	25%
17	NM_001258025.1	hTRIM60	H	157	98%	2e-40	24%
18	NM_006510.4	hTRIM27	H	157	98%	7e-40	27%
19	NM_009054.3	mTrim27	M	157	98%	1e-39	27%
20	NM_213562.1	rTrim39	R	156	98%	1e-39	25%
21	NM_015431.3	hTRIM58	H	157	97%	2e-39	29%
22	NM_024468.2	mTrim39	M	155	98%	4e-39	25%
23	NM_001039047.1	mTrim58	M	152	98%	8e-39	28%
24	NM_021253.3	hTRIM39	H	154	98%	8e-39	25%
25	NM_009277.4	mTrim21	M	152	96%	2e-38	25%
26	NM_153097.2	mTrim60	M	150	95%	2e-37	25%
27	NM_018073.6	hTRIM68	H	147	97%	2e-36	25%
28	NM_198012.2	mTrim68	M	144	98%	6e-36	24%
29	NM_022798.2	rTrim17	R	142	93%	9e-36	27%
30	NM_031172.2	mTrim17	M	140	93%	3e-35	27%
31	NM_001290988.1	mTrim11	M	141	92%	8e-35	29%
32	NM_001003818.2	hTRIM6	H	142	94%	1e-34	25%
33	NM_145214.2	hTRIM11	H	139	92%	7e-34	28%

No	Accession	Name	Species *	Total score	Query cover	E value	Ident
34	NM_009099.2	mTrim30a	M	137	98%	5e-33	24%
35	NM_001170461.1	rTrim6	R	135	94%	7e-33	24%
36	NM_006778.3	hTRIM10	H	134	94%	4e-32	27%
37	NM_001033429.2	mTrim75	M	133	98%	5e-32	26%
38	NM_001146007.1	mTrim12c	M	133	94%	5e-32	23%
39	NM_033034.2	hTRIM5	H	133	94%	6e-32	24%
40	NM_016102.3	hTRIM17	H	131	93%	1e-31	27%
41	NM_001009176.2	rTrim10	R	130	94%	2e-31	26%
42	NM_006074.4	hTRIM22	H	129	93%	8e-31	25%
43	NM_006355.4	hTRIM38	H	127	93%	8e-30	24%
44	NM_001013616.2	mTrim6	M	125	94%	2e-29	24%
45	NM_033017.3	hTRIM4	H	124	93%	6e-29	26%
46	NM_177742.4	mTrim11	M	119	96%	9e-28	25%
47	NM_178556.3	hTRIML1	H	118	96%	1e-27	25%
48	NM_001014023.1	rTrim5	R	119	97%	1e-27	22%
49	NM_001164397.1	hTRIM64B	H	116	93%	4e-27	24%
50	NM_001025142.2	rTrim35	R	117	94%	5e-27	26%
51	NM_029979.3	mTrim35	M	117	94%	7e-27	26%
52	NM_001136486.1	hTRIM64	H	115	93%	7e-27	24%
53	NM_171982.3	hTRIM35	H	116	93%	2e-26	26%
54	NM_001243916.1	mTrim34b	M	114	94%	2e-26	22%
55	NM_001017922.1	hERMAP	H	114	31%	5e-26	36%
56	NM_011280.2	mTrim10	M	112	94%	2e-25	24%
57	NM_013848.2	mErmap	M	112	31%	4e-25	35%
58	NM_001206631.1	hTRIM64C	H	109	93%	5e-25	23%
59	NM_020358.2	hTRIM49	H	110	92%	1e-24	27%
60	NM_001206627.1	hTRIM49D1	H	109	94%	1e-24	25%
61	NM_001195234.1	hTRIM49C	H	109	92%	2e-24	27%
62	NM_001206626.1	hTRIM49B	H	107	94%	2e-24	26%
63	NM_138800.1	hTRIM43	H	108	93%	4e-24	25%
64	NM_030684.3	mTrim34a	M	108	95%	4e-24	22%
65	NM_001164464.1	hTRIM43B	H	107	93%	5e-24	24%
66	NM_001029935.2	mTrim38	M	106	92%	8e-24	22%
67	NM_001276491.	rTrim34	R	107	93%	1e-23	23%

\*H = Homo sapiens, M = Mus musculus, R = Rattus norvegicus

下载上述 67 个序列，在 WebLab 上用 coderet 程序提取其编码区序列，在 MEGA 6.0 软件上进行多序列比对，构建系统发育树，采用 Neighbor-Joining 算法，用 Bootstrap 法进行评估（500 次重复），所构建系统发育树如下图所示：





**总结:**

- (1) MG53 从属于庞大的 TRIM 家族，具有众多旁系同源基因，其序列、结构具有不同程度的相似性。MG53 当前研究还较少，但其它 TRIM 家族成员有的已经有较充分的研究，它们的结构、功能具有一定的相似性，例如大部分 TRIM 家族成员都具有 RING finger / Zn finger domain、都具有 E3 泛素连接酶活性等。根据其它 TRIM 家族成员的已知功能、信号通路、相互作用分子，尤其是系统发育树中与 MG53 亲缘关系较近分子如 TRIM38 等，可以推测 MG53 的潜在功能、信号通路及互作分子。
- (2) TRIM 家族各成员分化时间较早，大部分在人、大鼠、小鼠分歧之前，部分为分歧之后重新分支形成。TRIM 家族成员在人和小鼠之间较为一致，大部分在人和小鼠上都存在同源基因。
- (3) tBlastN 将目标蛋白序列转译为多套 RNA 序列，再用这些 RNA 序列搜索 RNA 库，相比于简单的蛋白-蛋白 Blastp 而言，可以更精确地识别同源基因之间的微小差异，减少由于密码子简并性、重复突变等导致的匹配误差。同样，通过编码区序列进行系统发育树构建，相比于蛋白序列而言，能得到更加准确的结果。
- (4) Blast 和系统发育树构建都涉及较复杂的参数设置，根据实际研究需求，合理调整参数，可以更高效、更准确的获得结果。目前对这些参数的了解很有限，尚不能有效运用，需要继续深入学习。