人类白细胞抗原I型分子的功能和结构探究

Function and structure analysis of human leukocyte antigen class I

G12_13:王涛、吴而已、黄思源、贺淼清、刘潇阳、彭天博、田玉

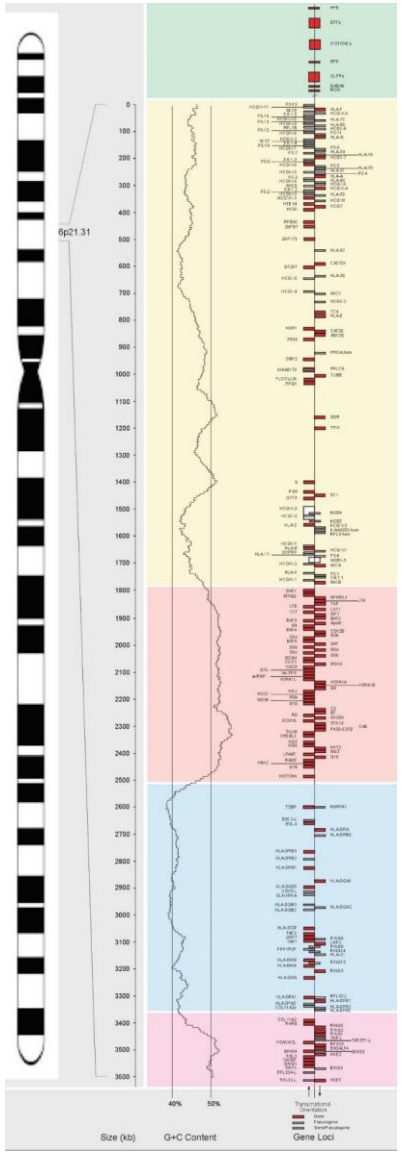
讲述人:吴而已

2020年1月4日

Contents

HLA

HLA (human leukocyte antigen)
人类白细胞抗原是人类的主要组织相容性复合体（MHC）的表达产物，在人体免疫过程中起到非常重要的调节作用。



Extended Class I

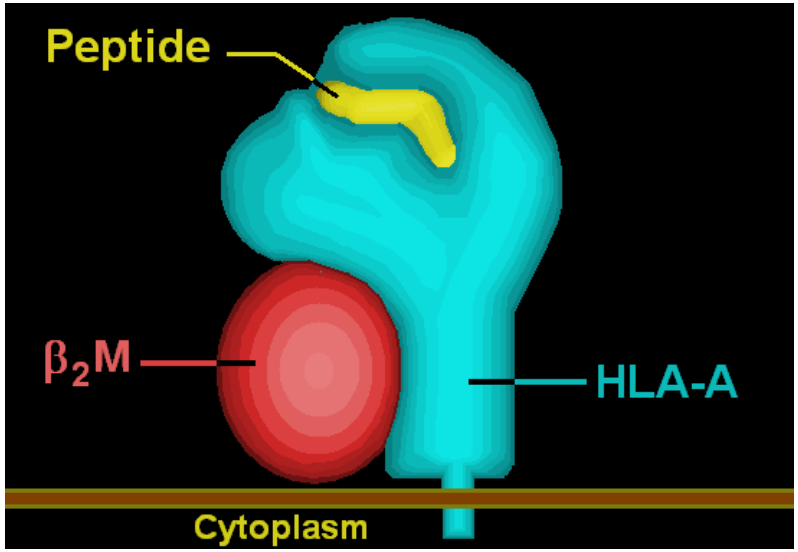
Classical Class I

Class III

Classical Class II

Extended Class II

HLA Class I



HLA class I是由α、β两个亚基形成的异二聚体，其中编码α链的基因主要有HLA-A、HLA-B、HLA-C三个，编码β亚基的B2M基因位于第15号染色体长臂。

Structure of the human class I histocompatibility antigen, HLA-A2

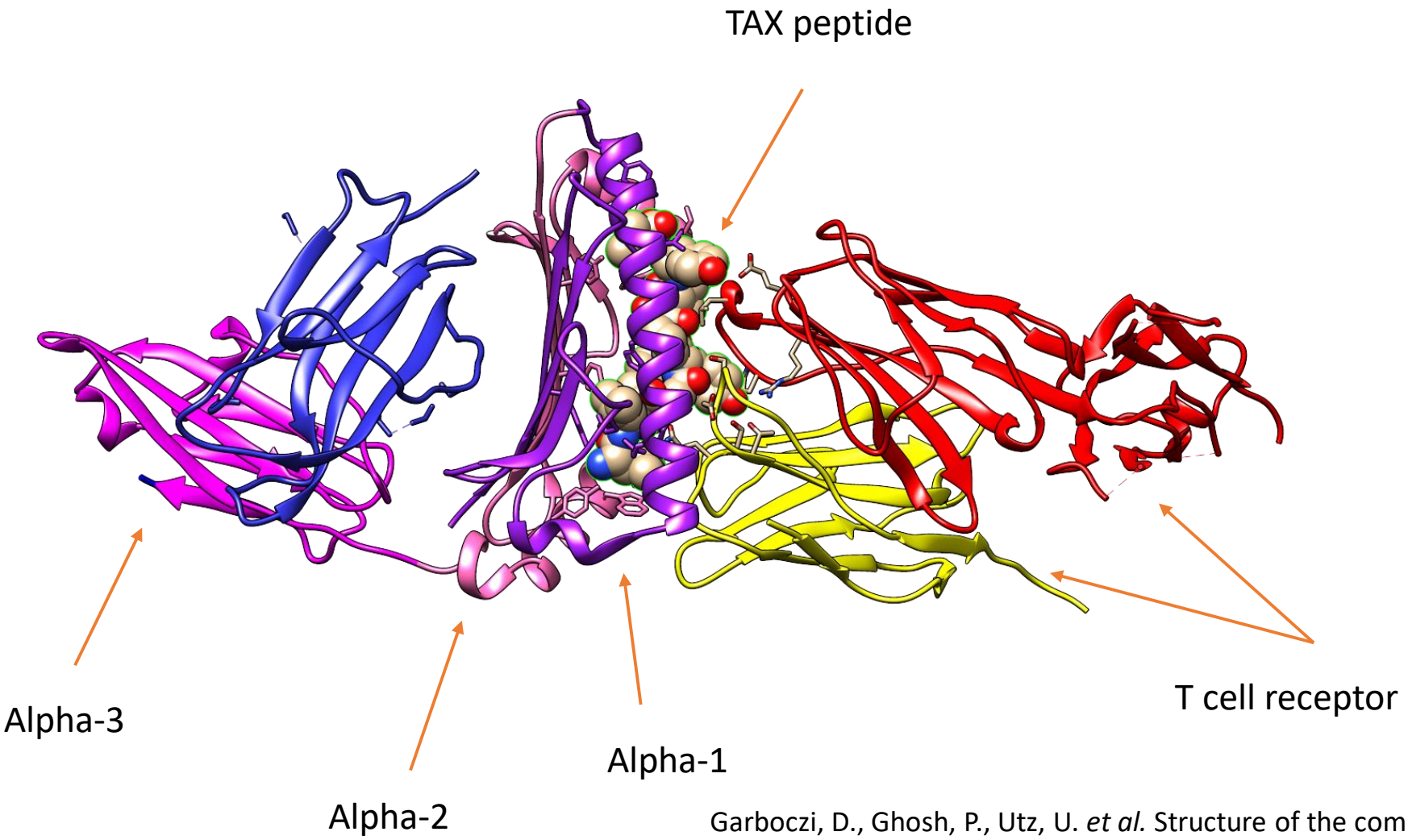
P. J. Bjorkman*†‡, M. A. Saper*, B. Samraoui*, W. S. Bennett*‡, J. L. Strominger* & D. C. Wiley*§

* Department of Biochemistry and Molecular Biology, Harvard University, Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA
† Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA
‡ Present addresses: Department of Medical Microbiology, Stanford University, Stanford, California 94305, USA (P.J.B.) and Max-Planck-Institute für Molekulare Genetik, Abteilung Wittmann, D-1000 Berlin 33, FDR (W.S.B.)
§ To whom correspondence should be addressed

HLA I型分子最早的结构解析完成于1987年，研究发现α亚基的顶部有一个凹槽，可以结合抗原，β亚基起到稳定结构的作用。

HLA I型分子α链的结构分析

1AO7示例



Garboczi, D., Ghosh, P., Utz, U. *et al.* Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384,** 134–141 (1996) doi:10.1038/384134a0

HLA Class I 功能

- beta-2-microglobulin binding
- CD8 receptor binding
- peptide antigen binding
- signaling receptor binding
- TAP binding
- TAP complex binding
- T cell receptor binding

基于Blast算法搜索HLA I型分子α链的相似序列

BLAST算法介绍

BLAST（Basic Local Alignment Search Tool）是一个用来比对生物序列的一级结构（如不同蛋白质的氨基酸序列或不同基因的DNA序列）的算法。已知一个包含若干序列的数据库，BLAST可以让研究者在其中寻找与其感兴趣的序列相同或类似的序列。

1) 移除Query序列中之低复杂度以及有串接重复现象的区域
2) 将Query序列中每k个字的组合做成一个表
3) 列出我们所关心的所有可能之字组
4) 将这些经筛选后剩下的高分字组组织成快速搜索树的结构
5) 对每个Query序列中的字组重复步骤1到4，并得到所有相应的高分字组
6) 扫描数据库中的序列，看看是否有跟剩下的高分字组完全匹配的情形
7) 将这些完全匹配的地方扩展为高分序对（high-scoring segment pair, HSP）
8) 将所有分数够高的HSP列出来
9) 评估这些留下来的HSP它们的分数是否具有意义
10) 将一个数据库序列中的多个HSP区域结合成一个更长的排比
11) 显示Query序列和所有之前找到可能相关的数据库序列之有间隙区域排比
12) 列出上一步骤中期望分数E低于我们所要求的门槛值之数据库序列

基于Blast算法搜索HLA I型分子α链的相似序列

搜索相似序列

从Uniprot数据库下载人类白细胞抗原I型分子（P04439）的蛋白质序列，长度为365，然后利用NCBI的blastp算法搜索人类白细胞抗原I型分子的相似序列。

Blastp是使用蛋白质序列与蛋白质数据库中的序列进行比较，可以寻找较远的关系；数据库选择UniProtKB/Swiss-Prot(swissprot) 数据库，其它参数选择默认。得到了100条显著的蛋白质序列，比对最大的得分为756分。

Sequences producing significant alignments Download Manage Columns Show 100

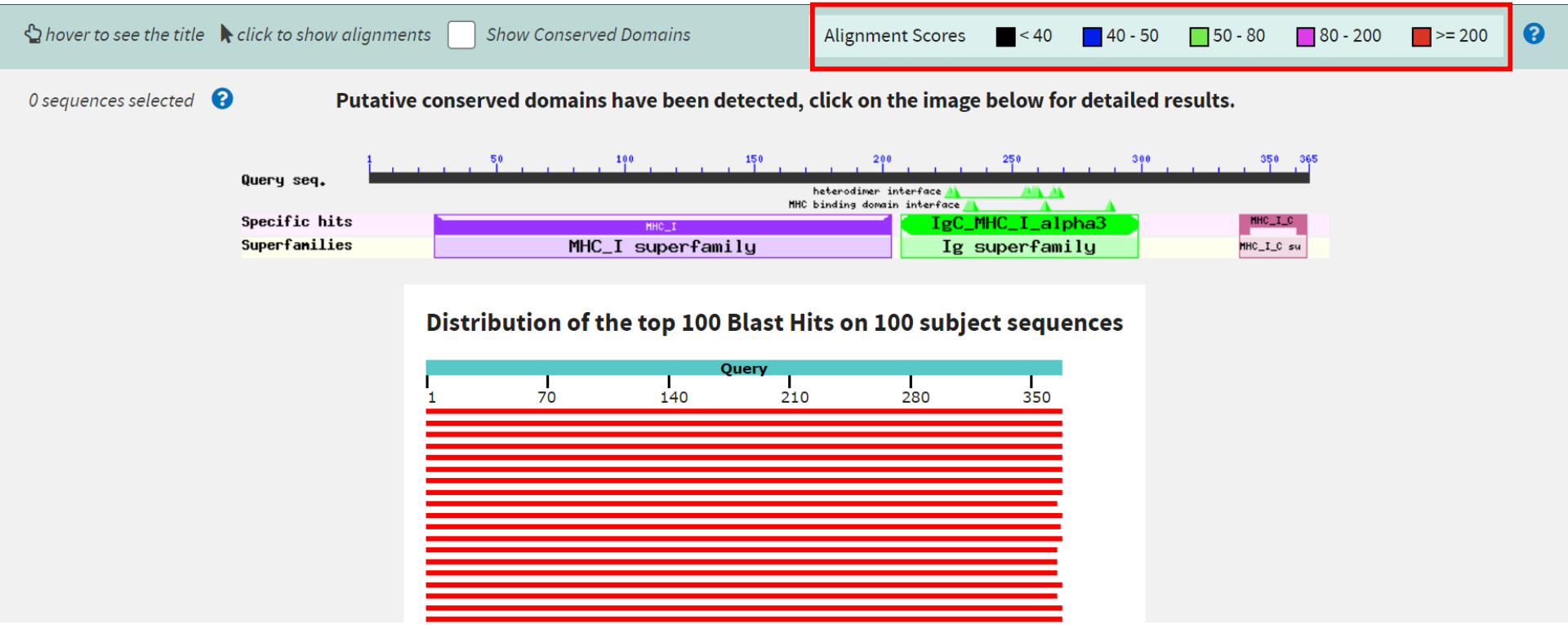
select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
RecName: Full=HLA class I histocompatibility antigen, A alpha chain; AltName: Full=Human leukocyte antigen A; Short=HLA-A; Flags: Precurs	756	756	100%	0.0	100.00%	P04439.2
RecName: Full=Patr class I histocompatibility antigen, A-2 alpha chain; AltName: Full=ChLa class I histocompatibility antigen, A-2 alpha chain;	734	734	100%	0.0	97.26%	P16209.1
RecName: Full=Patr class I histocompatibility antigen, A-108 alpha chain; AltName: Full=ChLa class I histocompatibility antigen, A-108 alpha ch	730	730	100%	0.0	96.71%	P13748.2
RecName: Full=Patr class I histocompatibility antigen, A-126 alpha chain; AltName: Full=ChLa class I histocompatibility antigen, A-126 alpha ch	728	728	100%	0.0	95.89%	P13749.1
RecName: Full=Class I histocompatibility antigen, Gogo-A*0401 alpha chain; Flags: Precursor [Gorilla gorilla gorilla]	689	689	100%	0.0	93.97%	P30377.1
RecName: Full=Class I histocompatibility antigen, Gogo-A*0101 alpha chain; Flags: Precursor [Gorilla gorilla gorilla]	684	684	100%	0.0	93.42%	P30375.1
RecName: Full=Popy Class I histocompatibility antigen, A-1 alpha chain; AltName: Full=Class I histocompatibility antigen, A-1 alpha chain; Flag	673	673	100%	0.0	91.51%	P16211.1
RecName: Full=Class I histocompatibility antigen, Gogo-A*0201 alpha chain; Flags: Precursor [Gorilla gorilla gorilla]	658	658	100%	0.0	92.88%	P30376.1
RecName: Full=Class I histocompatibility antigen, Gogo-OKO alpha chain; Flags: Precursor [Gorilla gorilla gorilla]	649	649	99%	0.0	89.23%	P30388.1
RecName: Full=Class I histocompatibility antigen, Gogo-A*0501 alpha chain; Flags: Precursor [Gorilla gorilla gorilla]	619	619	100%	0.0	88.49%	P30378.1
RecName: Full=Class I histocompatibility antigen, Gogo-C*0101/C*0102 alpha chain; Flags: Precursor [Gorilla gorilla gorilla]	612	612	99%	0.0	83.52%	P30383.1
RecName: Full=Saoe class I histocompatibility antigen, A alpha chain; AltName: Full=Class I histocompatibility antigen, A alpha chain; Flags: P	610	610	100%	0.0	85.75%	P30515.1
RecName: Full=Class I histocompatibility antigen, Gogo-B*0101 alpha chain; Flags: Precursor [Gorilla gorilla gorilla]	602	602	99%	0.0	85.91%	P30379.1
PUTATIVE PSEUDOGENE: RecName: Full=Putative HLA class I histocompatibility antigen, alpha chain H; AltName: Full=HLA-12.4; AltName: I	602	602	99%	0.0	85.64%	P01893.3
RecName: Full=Class I histocompatibility antigen, Gogo-B*0103 alpha chain; Flags: Precursor [Gorilla gorilla gorilla]	601	601	99%	0.0	85.91%	P30381.1
RecName: Full=Saoe class I histocompatibility antigen, C alpha chain; AltName: Full=Class I histocompatibility antigen, C alpha chain; Flags: P	600	600	100%	0.0	82.47%	P30517.1
RecName: Full=Class I histocompatibility antigen, Gogo-B*0102 alpha chain; Flags: Precursor [Gorilla gorilla gorilla]	600	600	99%	0.0	85.64%	P30380.1

基于Blast算法搜索HLA I型分子α链的相似序列

搜索相似序列

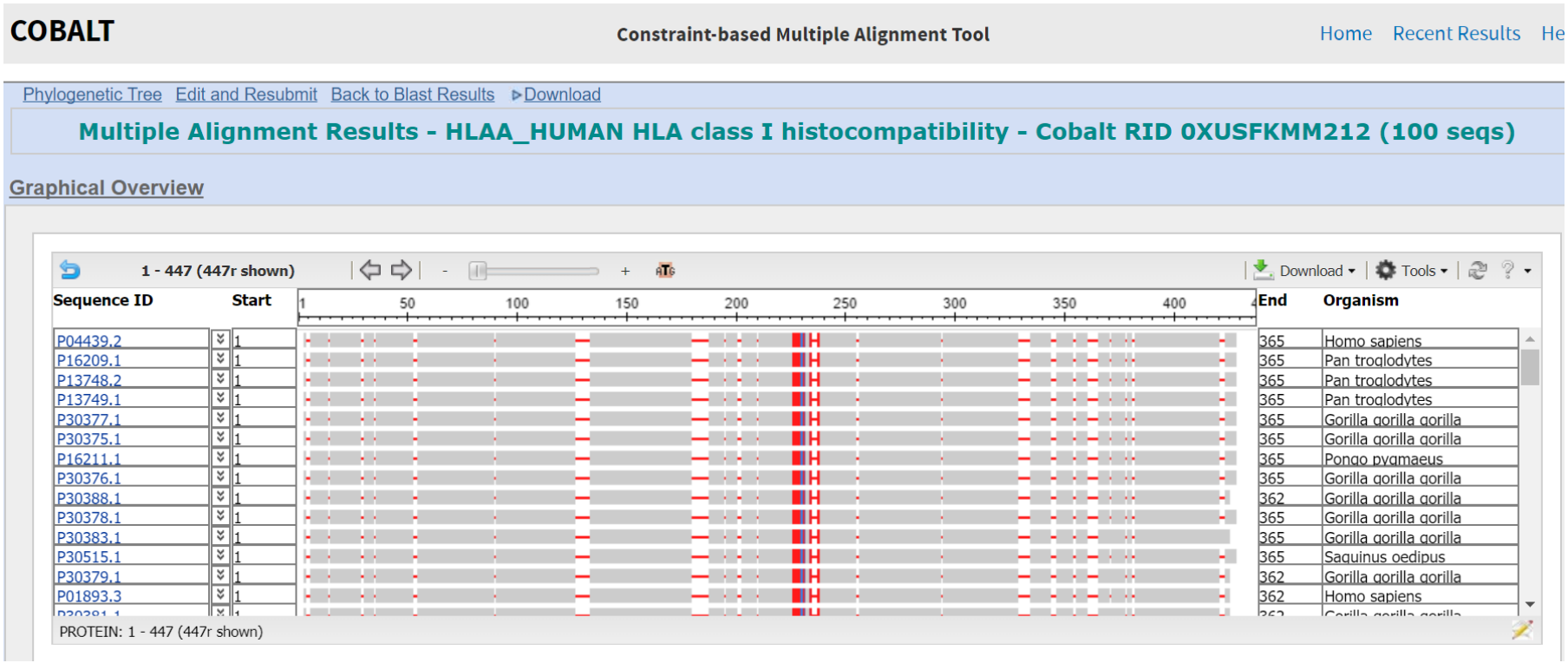
相似序列的graphic summary:



基于Blast算法搜索HLA I型分子α链的相似序列

搜索相似序列

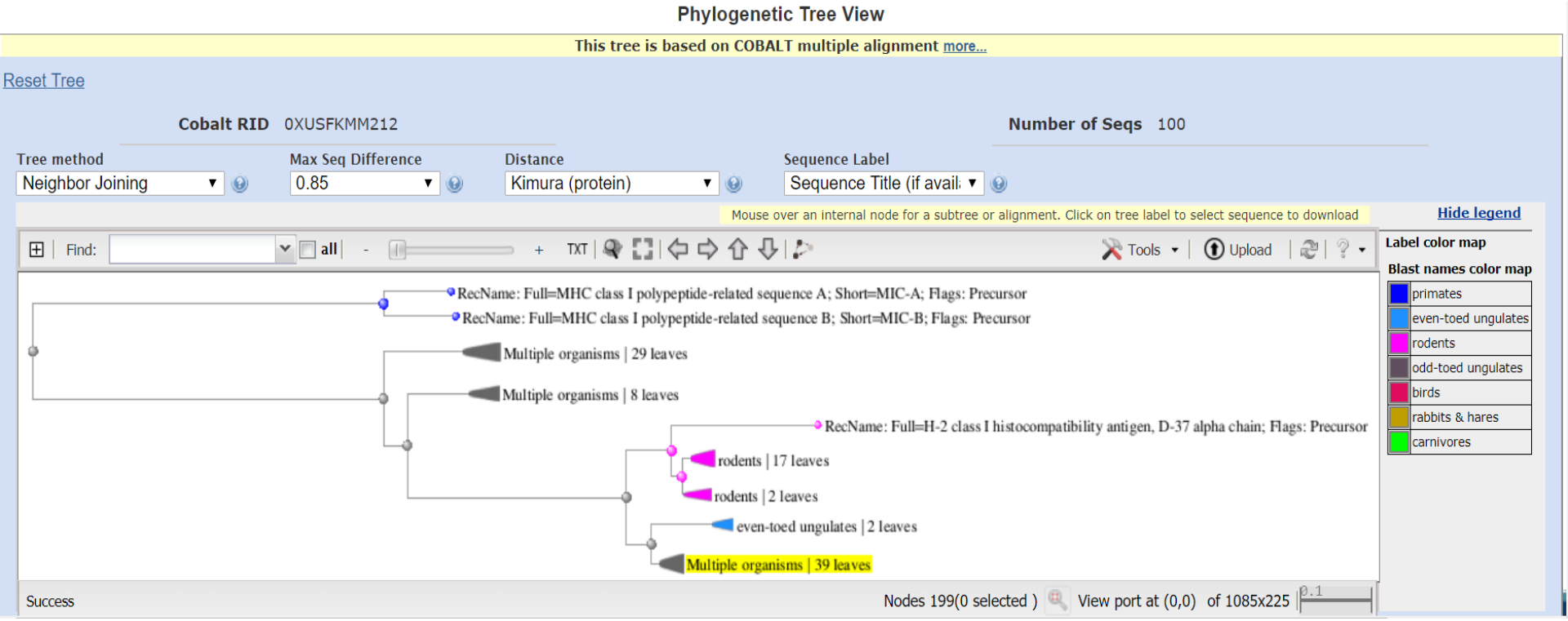
选中这100条条目，使用COBALT方法，进行多序列比对，其结果如下：



基于Blast算法搜索HLA I型分子α链的相似序列

搜索相似序列

基于COBALT 多序列比对结果，构建系统发生树，建树的方法为Neighbor Joining法，最大的序列相异性参数为0.85，计算距离方法选择Kimura(protein)，生成的系统发生树如下：



人类和大鼠、小鼠的HLA I型分子α链的序列比较

全局比对算法比对

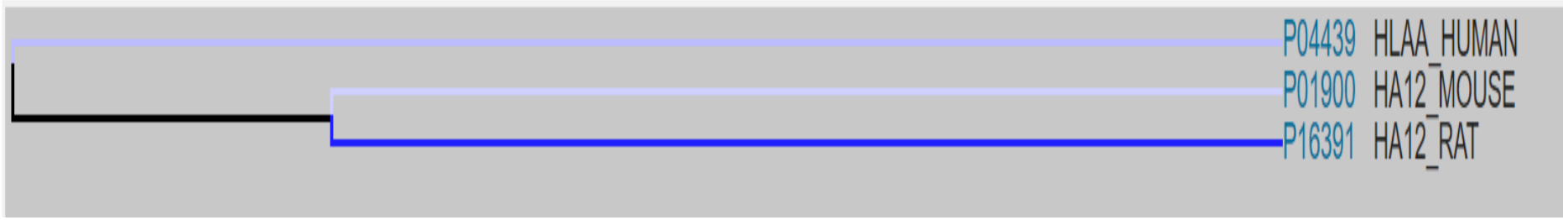
```
P04439 HLAA_HUMAN             1   MAVMAPRTLLLLLSGALALTQTWAGSHSMRYFFTSVSRPGRGEPRFIAVGYVDDTQFVRF        60
P01900 HA12_MOUSE             1   MGAMAPRTLLLLLAAALGPTQTRAGSHSLRYFVTAVSRPGFGEPRYMEVGYVDNTEFVRF        60
P16391 HA12_RAT               1   MEAMAPRTLLLLLAAALAPTQTRAGSHSLRYFYTAVSRPGLGEPRFIAVGYVDDTEFVRF        60
                                  * .**********:.**. *** *****:*** *:***** ****:: *****:*:****

P04439 HLAA_HUMAN            61   DSDAASQRMEPRAPWIEQEGPEYWDQETRNVKAQSQTDRVDLGTLRGYYNQSEAGSHTIQ        120
P01900 HA12_MOUSE            61   DSDAENPRYEPRARWIEQEGPEYWERETRRAKGNEQSFRVDLRTALRYYNQSAGGSHTLQ        120
P16391 HA12_RAT              61   DSDAENPRMEPRARWMEREGPEYWEQQTRIAKEWEQIYRVDLRTLRGYYNQSEGGSHTIQ        120
                                  **** . * **** *:*:******:::** .*  .*  **** *   ***** .****:*

P04439 HLAA_HUMAN           121   IMYGCDVGSDGRFLRGYRQDAYDGKDYIALNEDLRSWTAADMAAQITKRKWEAAHEAEQL        180
P01900 HA12_MOUSE           121   WMAGCDVESDGRLLRGYWQFAYDGCDYIALNEDLKTWTAADMAAQITRRKWEQAGAAERD        180
P16391 HA12_RAT             121   EMYGCDVGSDGSLLRGYRQDAYDGRDYIALNEDLKTWTAADFAAQITRNKWERARYAERL        180
                                   * **** *** :**** * **** *********::*****:*****:.*** *  **:

P04439 HLAA_HUMAN           181   RAYLDGTCVEWLRRYLENGKETLQRTDPPKTHMTHHPISDHEATLRCWALGFYPAEITLT        240
P01900 HA12_MOUSE           181   RAYLEGECVEWLRRYLKNGNATLLRTDPPKAHVTHHRRPEGDVTLRCWALGFYPADITLT        240
P16391 HA12_RAT             181   RAYLEGTCVEWLSRYLELGKETLLRSDPPEAHVTLHPRPEGDVTLRCWALGFYPADITLT        240
                                  ****:* ***** ***: *: ** *:***::*:* *   : :.************:****

P04439 HLAA_HUMAN           241   WQRDGEDQTQDTELVETRPAGDGTFQKWAAVVVPSGEEQRYTCHVQHEGLPKPLTLRWEL        300
P01900 HA12_MOUSE           241   WQLNGEELTQEMELVETRPAGDGTFQKWASVVVPLGKEQKYTCHVEHEGLPEPLTLRWGK        300
P16391 HA12_RAT             241   WQLNGEDLTQDMELVETRPAGDGTFQKWASVVVPLGKEQNYTCRVEHEGLPKPLSQRWEP        300
                                  ** :**: **: *****************:**** *:**.***:*:*****:**: **

P04439 HLAA_HUMAN           301   SSQPTIPI-----VGIIAGLVLLG-AVITGAVVAAVMWRRKSSDRKGGSYTQAASSDSAQ        354
P01900 HA12_MOUSE           301   EEPPSSTKTNT---VIIAVPVVLGAVVILGAVMAFVMKRRRNTGGKGGDYALAPGS---Q        354
P16391 HA12_RAT             301   SPSTDSNMETTVIYVILGAVAMIGAVAIIGAMVAVVRRRKRNTGGKGGDYAPAPGRDSSQ        360
                                  .             *:.  .::* ..* **::* *  *::.:. ***.*: * .    *

P04439 HLAA_HUMAN           355   GSDVSLTACKV                                                  365
P01900 HA12_MOUSE           355   SSDMSLPDCKV                                                  365
P16391 HA12_RAT             361   SSDVSLPDCKA                                                  371
                                  .**:**  **.
```

人类和大鼠、小鼠的HLA I型分子α链的序列比较

全局比对算法比对

生成的物种发生树:

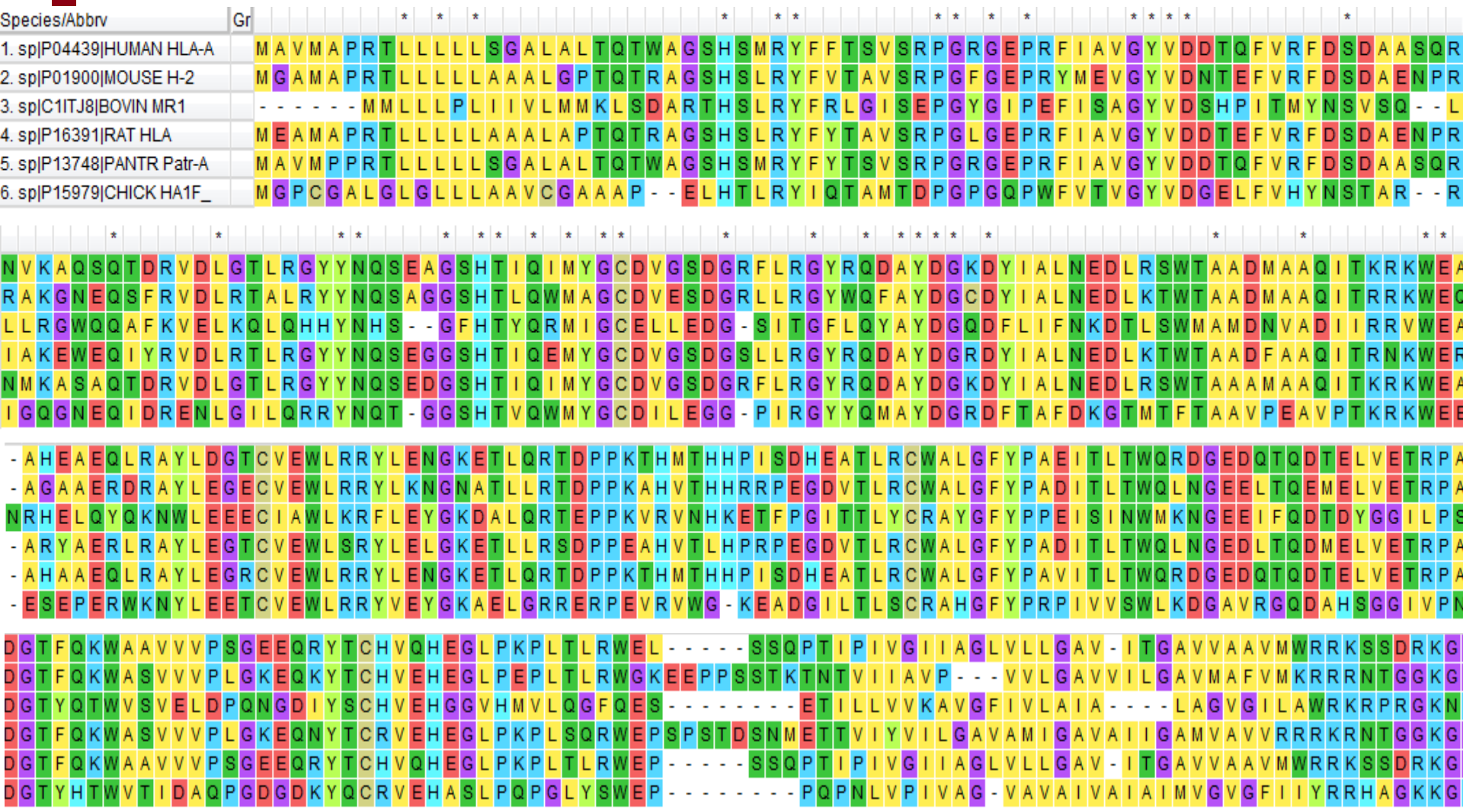
Tree



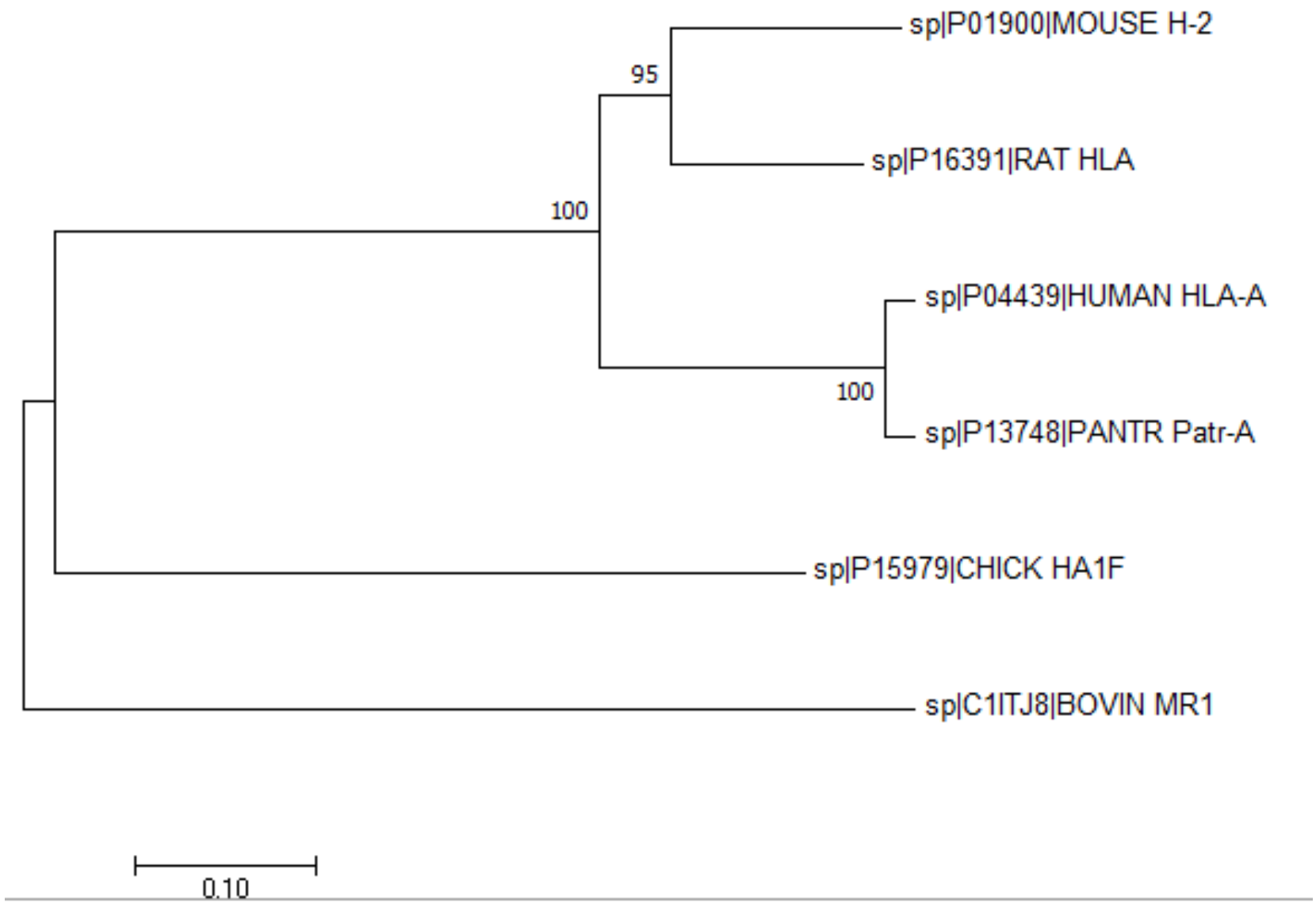
Result information

Date of job execution	2019-12-29
Job identifier	A201912295C475328CEF75220C360D524E9D456CE068CC6O (jobs are stored for 7 days)
Running time	16.5 seconds
Identical positions	224
Identity	60.377%
Similar positions	69
Program	clustalo
Default parameters	Default parameters: The default transition matrix is Gonnet, gap opening penalty is 6 bits, gap extension is 1 bit. Clustal-Omega uses the HHalign algorithm and its default settings as its core alignment engine. The algorithm is described in Söding, J. (2005) 'Protein homology detection by HMM–HMM comparison'. Bioinformatics 21, 951-960.

不同物种HLA I型分子序列比对



不同物种HLA I型分子进化树



HLA I型分子α链的结构分析

UniProt数据库中的结构域与区域划分

Molecule processing

Feature key	Position(s)	Description	Actions	Graphical view	Length
Signal peptide	1 – 24	1 Publication	Add BLAST		24
Chain (PRO_0000018815)	25 – 365	HLA class I histocompatibility antigen, A alpha chain	Add BLAST		341

Topology

Feature key	Position(s)	Description	Actions	Graphical view	Length
Topological domain	25 – 308	Extracellular Sequence analysis	Add BLAST		284
Transmembrane	309 – 332	Helical Sequence analysis	Add BLAST		24
Topological domain	333 – 365	Cytoplasmic Sequence analysis	Add BLAST		33

Domains and Repeats

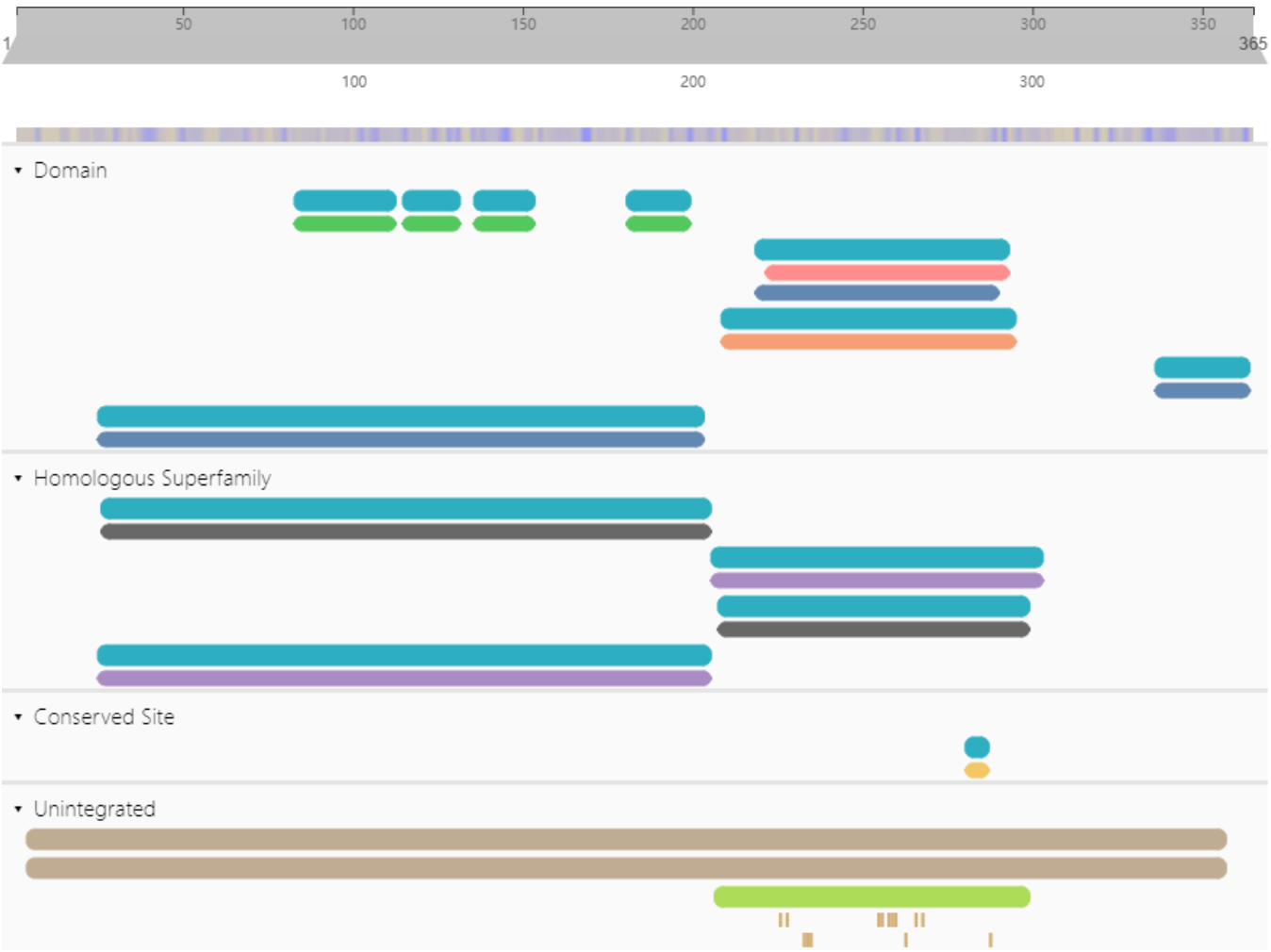
Feature key	Position(s)	Description	Actions	Graphical view	Length
Domain	209 – 295	Ig-like C1-type Sequence analysis	Add BLAST		87

Region

Feature key	Position(s)	Description	Actions	Graphical view	Length
Region	25 – 114	Alpha-1 Sequence analysis	Add BLAST		90
Region	115 – 206	Alpha-2 Sequence analysis	Add BLAST		92
Region	207 – 298	Alpha-3 Sequence analysis	Add BLAST		92
Region	299 – 308	Connecting peptide Sequence analysis			10

HLA I型分子α链的结构分析

InterPro数据库的结构域分析



参考文献

[1] Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. Structure of the human class I histocompatibility antigen, HLA-A2. Nature. 1987;329(6139):506-12

[2] Madden DR, Garboczi DN, Wiley DC. The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. Cell. 1993;75(4):693-708

[3] Salter RD, Norment AM, Chen BP, et al. Polymorphism in the alpha 3 domain of HLA-A molecules affects binding to CD8. *Nature*. 1989;338(6213):345-7

[4] Ding YH, Baker BM, Garboczi DN, Biddison WE, Wiley DC. Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical. *Immunity*. 1999;11(1):45-56

[5] Neisig A, Wubbolts R, Zang X, Melief C, Neefjes J. Allele-specific differences in the interaction of MHC class I molecules with transporters associated with antigen processing. *J Immunol*. 1996;156(9):3196-206