甘蓝cer1同源基因预测及分析

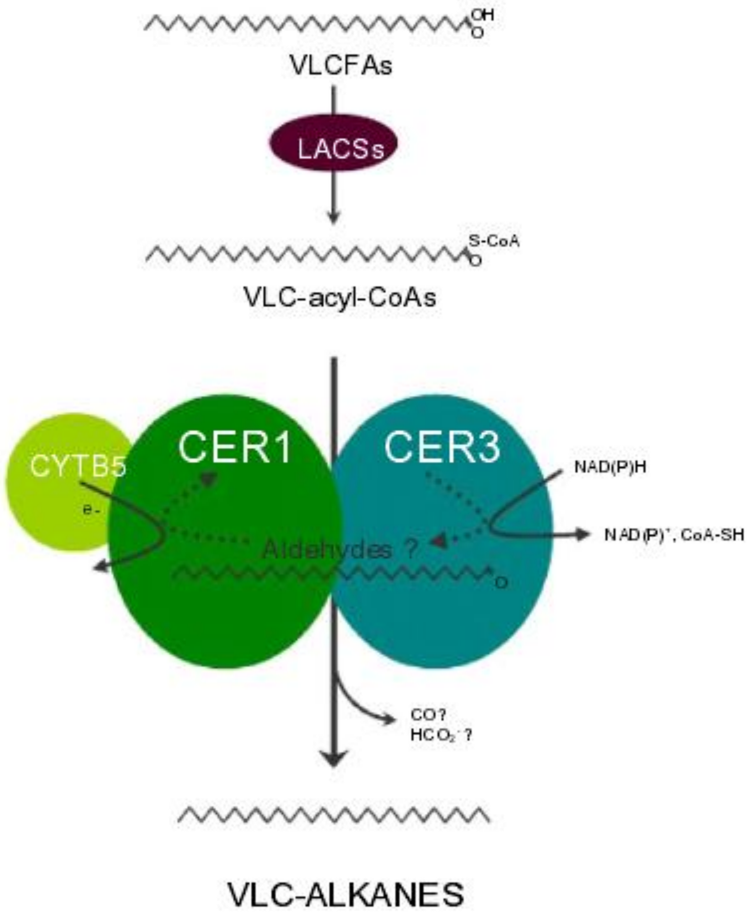
2013/1/20 唐俊 周慧 张薇 马振国

研究背景

- 植物外表面有一层疏水角质层,其最外层由表皮蜡质覆盖,对植物抵抗恶劣的外部环境具有十分重要意义
- 表皮蜡质主要由VLCFAs的衍生物组成，包括烷烃、伯醇、次级醇、醛和酮等.外表皮蜡质合成有2条途径:烷烃合成途径和醇合成途径,其中拟南芥中烷烃合成途径合成了80%左右的蜡质成分.
- 甘蓝为重要蔬菜,同为十字花科植物,其外表皮主要蜡质成分同为烷烃,且甘蓝基因组测序已经完成.

研究背景

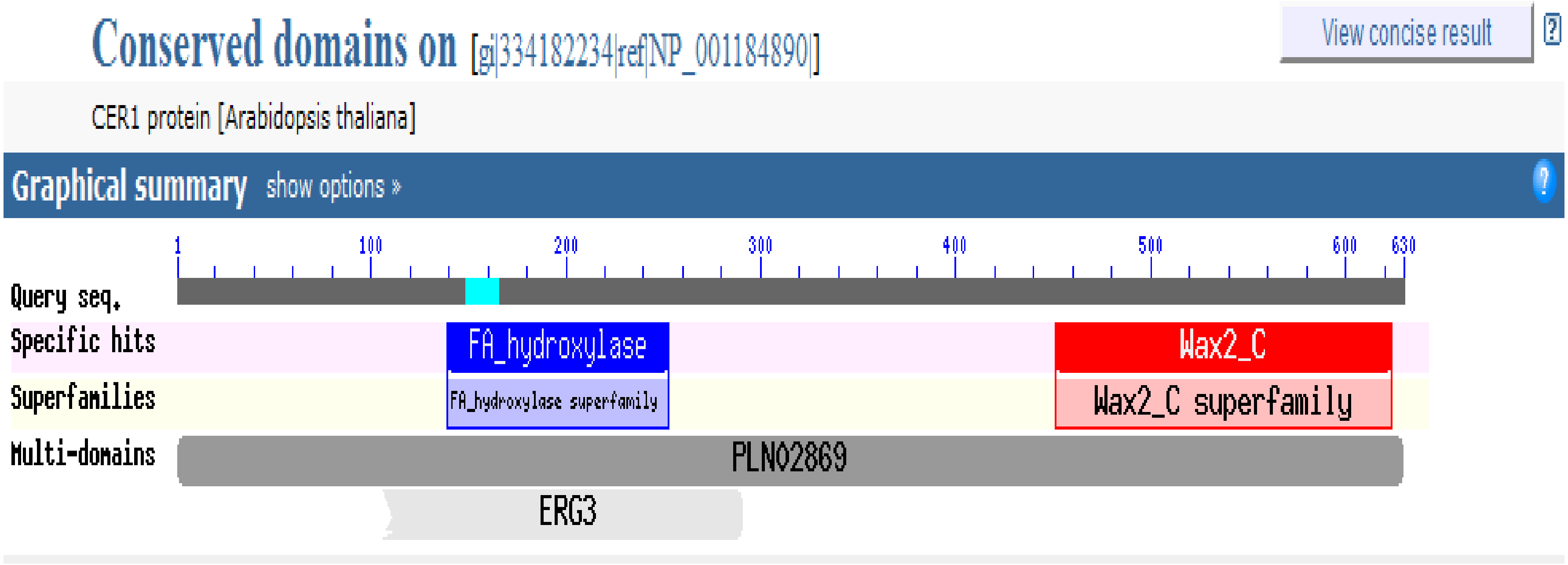
- 烷烃合成途径


OH
O
VLCFAs
LACSs
S-CoA
O
VLC-acyl-CoAs
CYTB5
e-
CER1
CER3
NAD(P)H
NAD(P)⁺, CoA-SH
Aldehydes ?
O
CO?
HCO₂⁻?
VLC-ALKANES

研究背景

- Cer1登陆信息NP_001184890 NM_001197961.1
- CER1蛋白含有630aa,含有4个保守序列:
- Wax2 C-terminal domain,此保守序列为451-624,与短链脱氢酶相似
- FA_hydroxylase, Fatty acid hydroxylase superfamily.序列为139-253,此超家族与植物表皮蜡质合成相关,含有2个HXHH拷贝,是完整的膜蛋白.
- PLN02869,fatty aldehyde decarbonylase,1-620,脂肪醛脱羧酶
- ERG3, Sterol desaturase序列为106-290,固醇脱氢酶,参与脂合成.

研究背景



Conserved domains on [gi|334182234|ref|NP_001184890|]

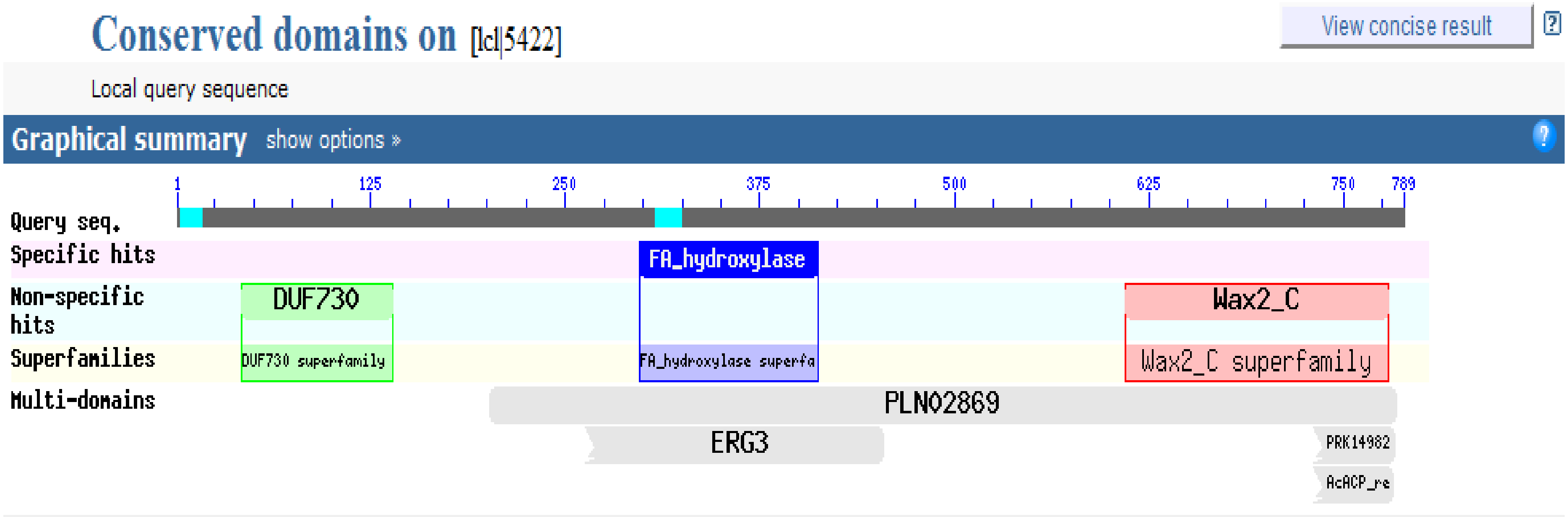
CER1 protein [Arabidopsis thaliana]

Graphical summary show options »

甘蓝基因组中cer1同源序列获得

- 利用拟南芥cer1cDNA序列对甘蓝基因组序列进行Blastn,分别在第8号染色体,第3号染色体和第4号染色体上找到同源序列,利用Cutseq脚本分别截取同源序列附近NDA序列。
- 利用Weblab工具augustus（v2.1）进行序列分析，并将候选基因预测的蛋白进行在NCBI上进行blast，获得候选3个候选基因，分别位于8号，3号，4号染色体上。

Cer1—8号染色体

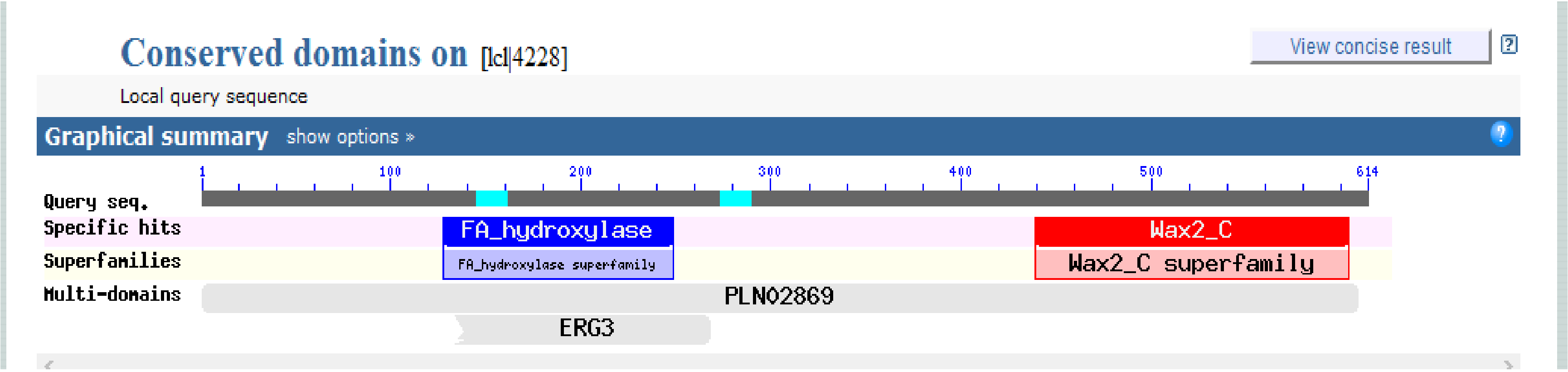


此序列包含789个aa，7个保守序列，除cer1含有的4个保守序列，其中DUF序列功能未知

PPK14982, acyl-ACP reductase脂酰COA还原酶序列731-783

AcACP_reductase[TIGR04058]:long-chain fatty acyl-ACP reductase (aldehyde-forming)，序列为731-782.都为参与蜡质合成酶。

cer1-3号染色体

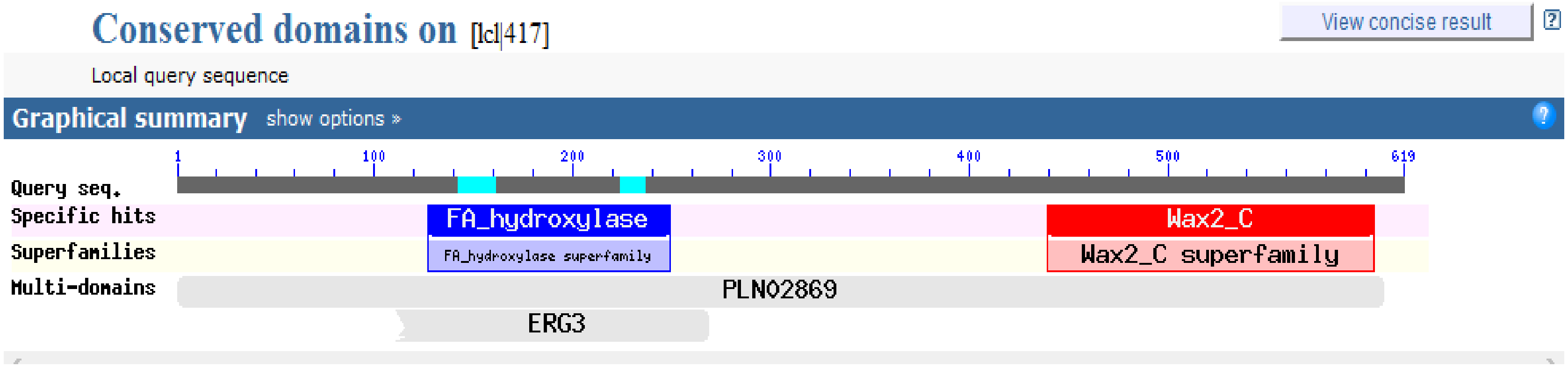


Conserved domains on [lcl|4228]

Local query sequence

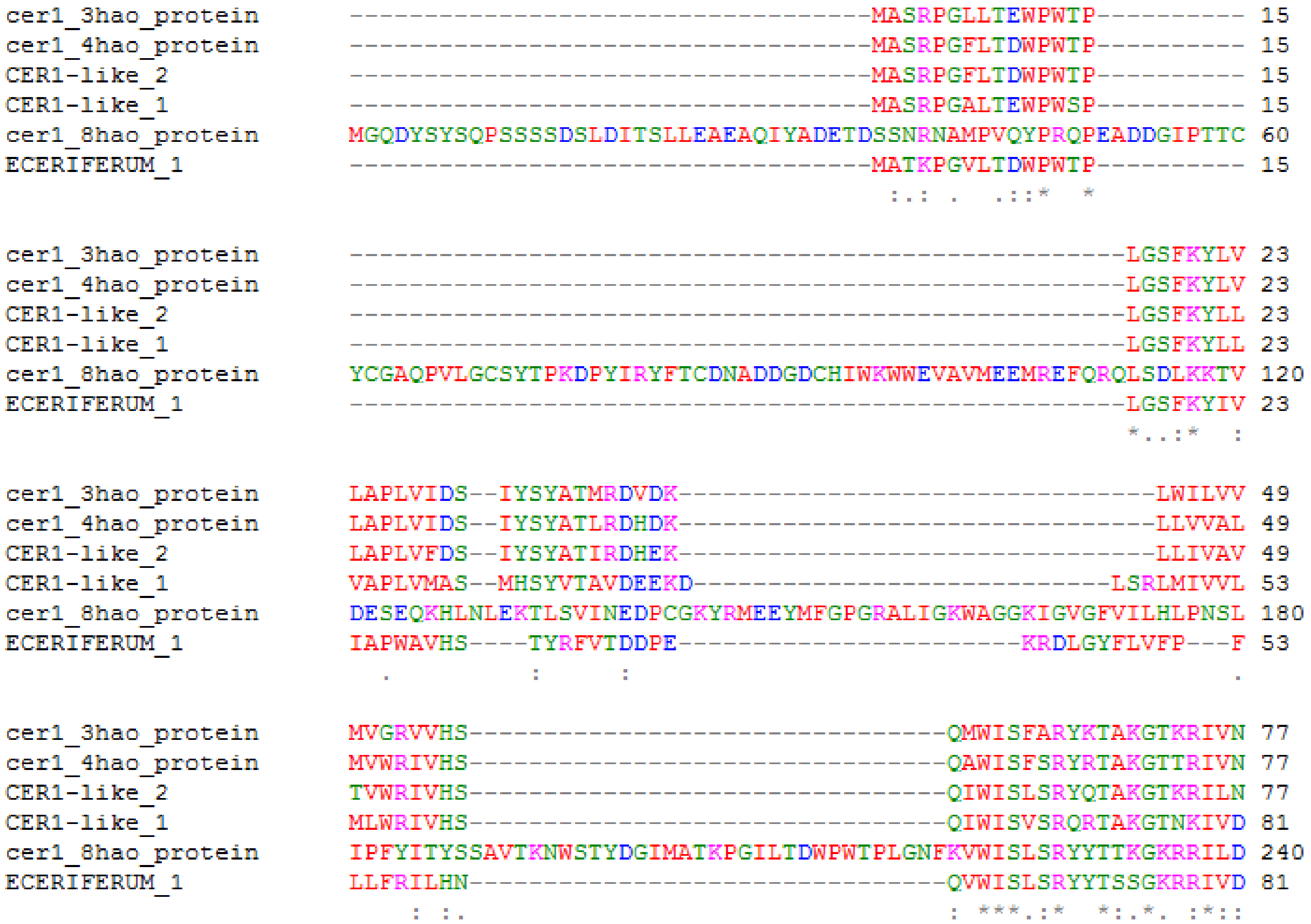
3号染色体上的含614个aa，预测基因含4个保守结构域，同拟南芥cer1基因一样

Cer1_4号染色体



此序列包含619个aa，4个保守序列，同拟南芥cer1一样。

多序列比对

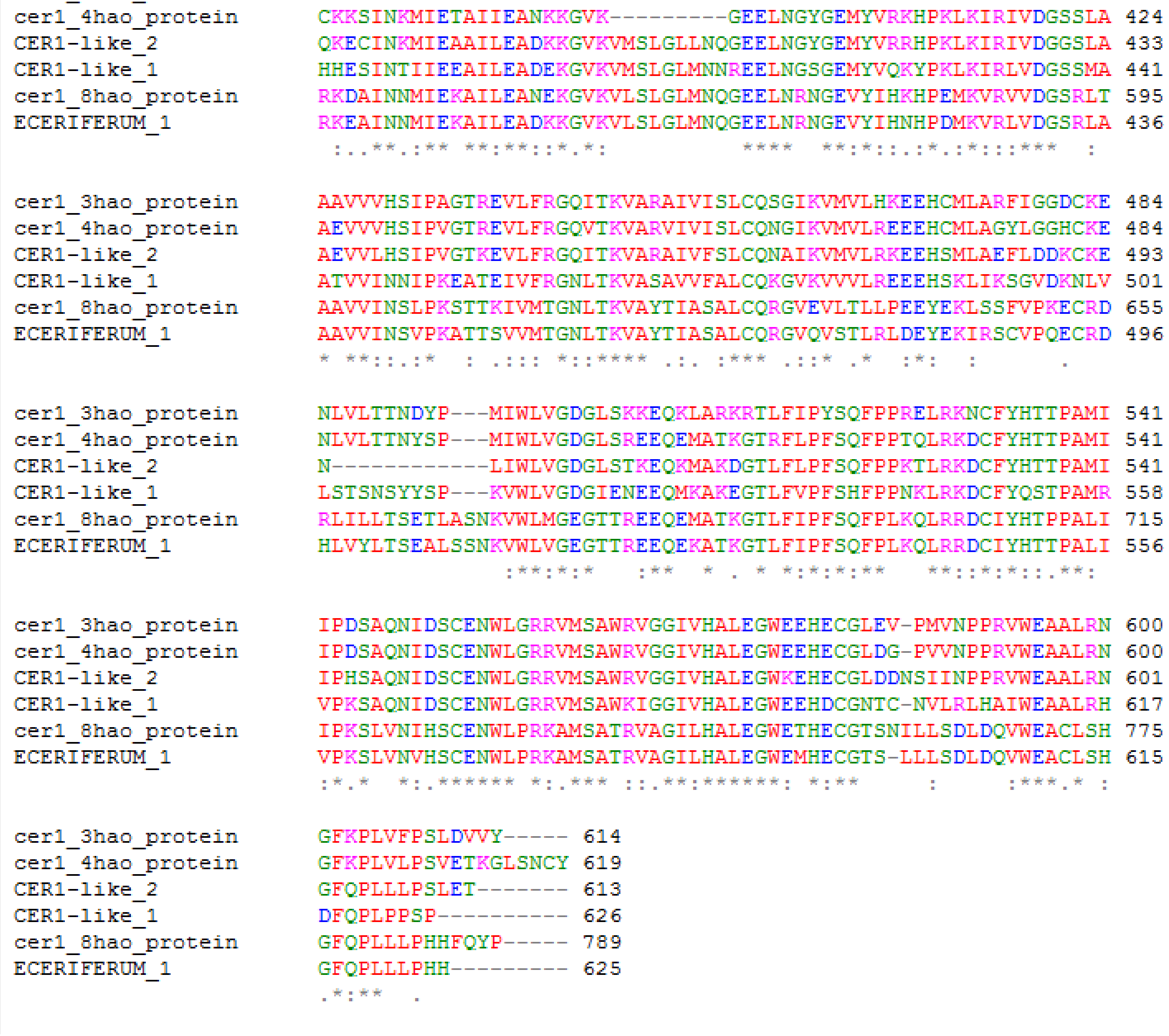


```
cer1_3hao_protein      -----------------------------------MASRPGLLTEWPWTP---------- 15
cer1_4hao_protein      -----------------------------------MASRPGFLTDWPWTP---------- 15
CER1-like_2            -----------------------------------MASRPGFLTDWPWTP---------- 15
CER1-like_1            -----------------------------------MASRPGALTEWPWSP---------- 15
cer1_8hao_protein      MGQDYSYSQPSSSSDSLDITSLLEAEAQIYADETDSSNRNAMPVQYPRQPEADDGIPTTC 60
ECERIFERUM_1           -----------------------------------MATKPGVLTDWPWTP---------- 15
                                                           :.: .  .::*  *

cer1_3hao_protein      ----------------------------------------------------LGSFKYLV 23
cer1_4hao_protein      ----------------------------------------------------LGSFKYLV 23
CER1-like_2            ----------------------------------------------------LGSFKYLL 23
CER1-like_1            ----------------------------------------------------LGSFKYLL 23
cer1_8hao_protein      YCGAQPVLGCSYTPKDPYIRYFTCDNADDGDCHIWKWWEVAVMEEMREFQRQLSDLKKTV 120
ECERIFERUM_1           ----------------------------------------------------LGSFKYIV 23
                                                                           *..:*  :

cer1_3hao_protein      LAPLVIDS--IYSYATMRDVDK--------------------------------LWILVV 49
cer1_4hao_protein      LAPLVIDS--IYSYATLRDHDK--------------------------------LLVVAL 49
CER1-like_2            LAPLVFDS--IYSYATIRDHEK--------------------------------LLIVAV 49
CER1-like_1            VAPLVMAS--MHSYVTAVDEEKD----------------------------LSRLMIVVL 53
cer1_8hao_protein      DESEQKHLNLEKTLSVINEDPCGKYRMEEYMFGPGRALIGKWAGGKIGVGFVILHLPNSL 180
ECERIFERUM_1           IAPWAVHS----TYRFVTDDPE-----------------------KRDLGYFLVFP---F 53
                         .         :     :                                        .

cer1_3hao_protein      MVGRVVHS--------------------------------QMWISFARYKTAKGTKRIVN 77
cer1_4hao_protein      MVWRIVHS--------------------------------QAWISFSRYRTAKGTTRIVN 77
CER1-like_2            TVWRIVHS--------------------------------QIWISLSRYQTAKGTKRILN 77
CER1-like_1            MLWRIVHS--------------------------------QIWISVSRQRTAKGTNKIVD 81
cer1_8hao_protein      IPFYITYSSAVTKNWSTYDGIMATKPGILTDWPWTPLGNFKVWISLSRYYTTKGKRRILD 240
ECERIFERUM_1           LLFRILHN--------------------------------QVWISLSRYYTSSGKRRIVD 81
                           : :.                                : ***.:*  *:.*. :*::
```

多序列比对

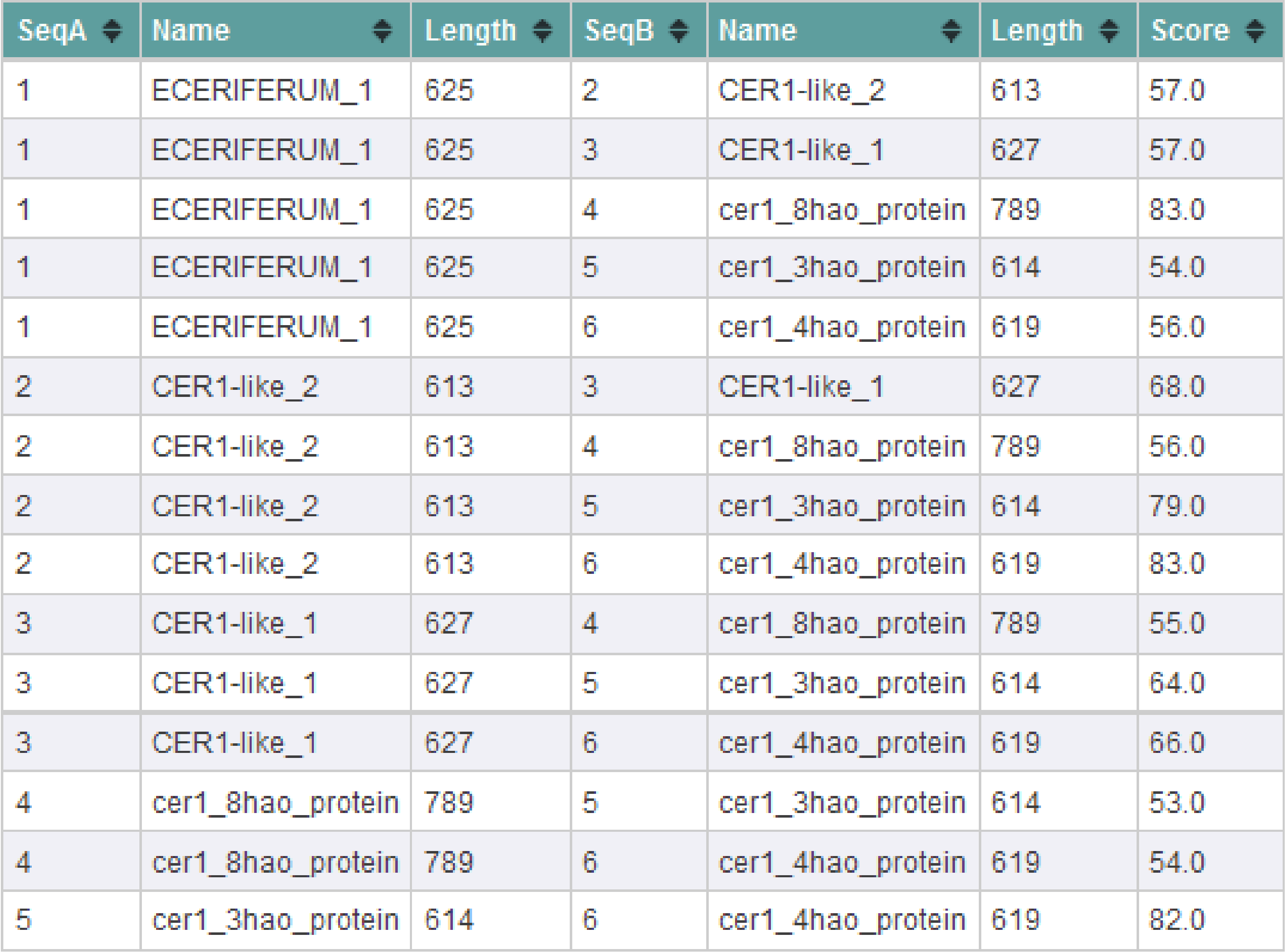
```
cer1_3hao_protein      KSIEFNQVDRERTWDDQVIFNTLVIYLAKVYVLGTNTLPFWRLDGVVQVALLHAGPVEFI 137
cer1_4hao_protein      KSIEFEQVDRERTWDDQIIFNTIIVYLVKAYVIRNNPVPFWRLDGVVLTVLLHAGPVEFI 137
CER1-like_2            KSIEFDQVDRERTWDDQIIFNTLIVYLTKVYVSGTSTIPFWRTDGVILVALLHAGPVEFI 137
CER1-like_1            KPIEFEQVDRERTWDDQVIFNTLLMYLANIKLPGASHLPPWRLDGAILMALLHAGPVEFL 141
cer1_8hao_protein      KGIDFNQVDRETNWDDQILFNGLLFYIGIMLLPQAKQLPWWRTDGVLMAAMLHAGPVEFL 300
ECERIFERUM_1           KGIDFNQVDRETNWDDQILFNGVLFYIGINLLPEAKQLPWWRTDGVLMAALIHTGPVEFL 141
                       * *:*:***** .****::** ::.*:    :   . :* ** **.:  .::*:*****:

cer1_3hao_protein      YYWFHRALHHHFLYSRYHSHHHSSIVTEPITSVVHPFAEHIGYTLILGIPLVMTLLCGTV 197
cer1_4hao_protein      YYWFHRALHHHYLYSRYHSHHHSSIVTEPITSVVHPFAEHIGYTLILGIPLITCLLCGTV 197
CER1-like_2            YYWFHRALHHHFLYSRYHSHHHSSIVTEPITSVVHPFAEHIGYTLILGLPLITTFMCGTV 197
CER1-like_1            YYWFHRALHHHFLYSRYHSHHHSSIVTEPITSVVHPFAEHIAYTLLFAIPMVTASLCGIL 201
cer1_8hao_protein      YYWLHKALHHHFLYSRYHSHHHSSIVTEPITSVIHPFAEHIAYFILFAIPLLTTLLTKTA 360
ECERIFERUM_1           YYWLHKALHHHFLYSRYHSHHHSSIVTEPITSVIHPFAEHIAYFILFAIPLLTTLLTKTA 201
                       ***:*:*****:*********************:*******.* :::.:*::   :

cer1_3hao_protein      SAASVTLYITYIDFMNNLGHCNFELIPRSFFSLFPHIKYLCYTPSFHSLHHTQFRTNYSL 257
cer1_4hao_protein      SVASLFLYLTYIDFMNNLGHCNFELIPKPFFSLFPPLKFICYTPSFHSLHHTQFRTNYSL 257
CER1-like_2            SVVSIALYLTYIDFMNNMGHCNFELIPKFLFSLLPPLKFLCYTPSFHSLHHTQFRTNYSL 257
CER1-like_1            SIVSIMGYITYIDFMNNMGHCNFELFPKRLFHLFPPLKFLCYTPSFHSLHHTQFRTNYSL 261
cer1_8hao_protein      SIASFSGYVIYIDFMNNMGHCNFELVPKRLFHLFPPLKYLCYTPSFHSLHHTQFRTNYSL 420
ECERIFERUM_1           SIISFAGYIIYIDFMNNMGHCNFELIPKRLFHLFPPLKFLCYTPSYHSLHHTQFRTNYSL 261
                       *  *.  *: *******:*******.*: :* *:* :*::*****:**************

cer1_3hao_protein      FMPMYDYIYGTNDKCSDSLYESLLEQEEEKPEAIHLTHLTSLDSIYHLRLGFASFSSHPL 317
cer1_4hao_protein      FMPMYDYIYGTNDKCSDSLYETSLEKEEEKPDAIHLTHLTSLDSIYQLRLGFASLSSHPL 317
CER1-like_2            FMPMYDYIYGTTDECSDSLYETSLEKEEEKPDAIHLTHLTSLDSIYHLRLGFASLSSHPL 317
CER1-like_1            FMPIYDFIYGTTDNLTDSLYERSLEIEEESPDVIHLTHLTTHNSIYQMRLGFPSLSSCPL 321
cer1_8hao_protein      FMPLYDYIYGTMDETSDTLYEKSLERGEDRVDVVHLTHLTTPESIYHLRIGLASFASYPF 480
ECERIFERUM_1           FMPLYDYIYGTMDESTDTLYEKTLERGDDIVDVVHLTHLTTPESIYHLRIGLASFASYPF 321
                       ***:**:**** *: :*:***  **  ::  :.:******: :***::*:*:.*::* *:

cer1_3hao_protein      SSR---FYLVLMKPITLIISFVLTS-FFSRTFVFERNRFGDLTLHSHLLPKFSSHYKSQQ 373
cer1_4hao_protein      SSR---CYLLLMRPFTLILSFILTS-FSFRTFVFERNRFRDLTIHSHLLPKFSSHYISQQ 373
CER1-like_2            SSR---CYLFLMKPFALILSFILRS-FSFQTFVVERNRFRDLTLHSHLLPKFSSHYMSHQ 373
CER1-like_1            WSRPPWYLTCFMWPFTLLCSFALTSAIPLRTFVFERNRLRDLTVHSHLLPKFSFHYKSQR 381
cer1_8hao_protein      SYR---WFMRLLWPFTSLS--MLFTLFYASLFVSERNSFEKLNLQSWIIPRYNLQYLLKW 535
ECERIFERUM_1           AYR---WFMRLLWPFTSLS--MIFTLFYARLFVAERNSFNKLNLQSWVIPRYNLQYLLKW 376
                         *       :: *:: :    : : :    ** *** : .*.::* ::*::. :*  :
```

多序列比对



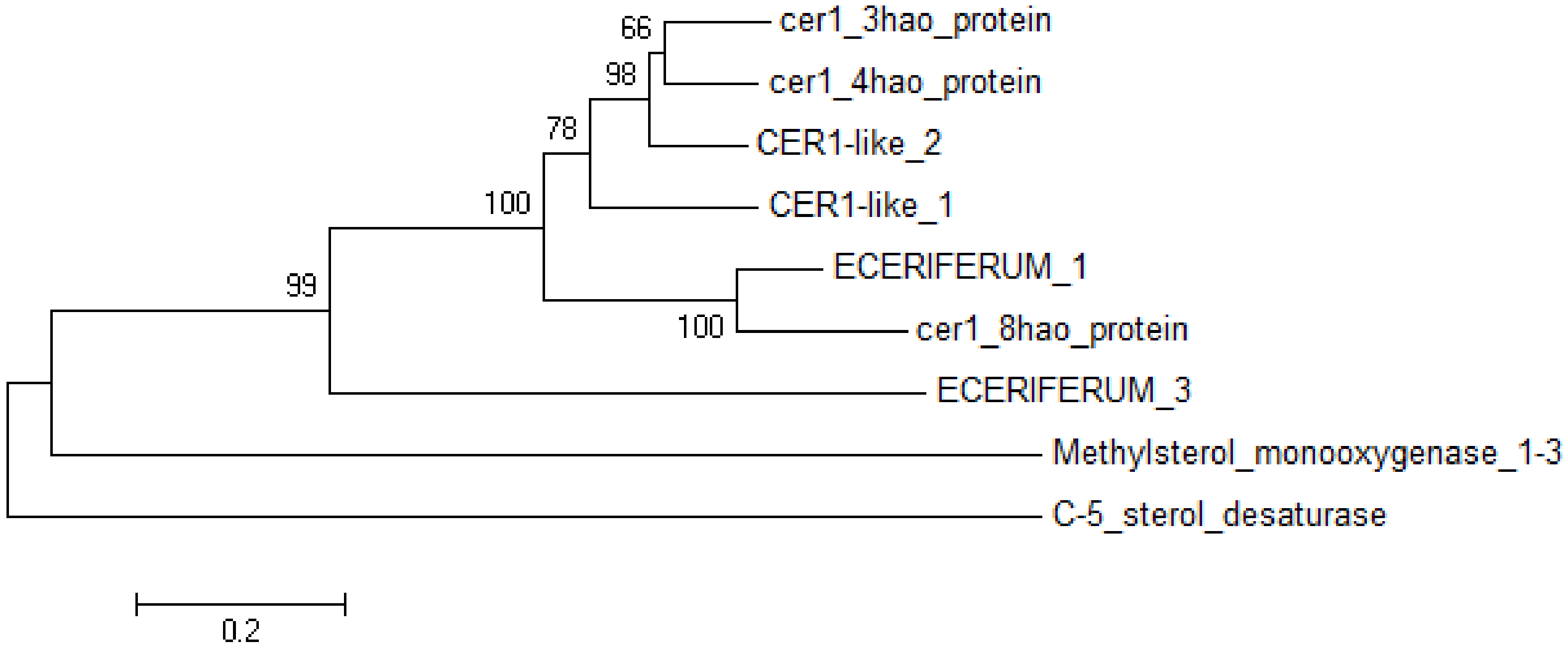
```
cer1_3hao_protein      QKESINKLIETAILEAEKKDVR---------GEELNGYGEMYVRKYPKLKIKIVDGTSLD 424
cer1_4hao_protein      CKKSINKMIETAIIEANKKGVK---------GEELNGYGEMYVRKHPKLKIRIVDGSSLA 424
CER1-like_2            QKECINKMIEAAILEADKKGVKVMSLGLLNQGEELNGYGEMYVRRHPKLKIRIVDGGSLA 433
CER1-like_1            HHESINTIIEEAILEADEKGVKVMSLGLMNNREELNGSGEMYVQKYPKLKIRLVDGSSMA 441
cer1_8hao_protein      RKDAINNMIEKAILEANEKGVKVLSLGLMNQGEELNRNGEVYIHKHPEMKVRVVDGSRLT 595
ECERIFERUM_1           RKEAINNMIEKAILEADKKGVKVLSLGLMNQGEELNRNGEVYIHNHPDMKVRLVDGSRLA 436
                        :..**.:** **:**::*.*:          ****  **:*::.:*.:*:::***  :

cer1_3hao_protein      AAVVVHSIPAGTREVLFRGQITKVARAIVISLCQSGIKVMVLHKEEHCMLARFIGGDCKE 484
cer1_4hao_protein      AEVVVHSIPVGTREVLFRGQVTKVARVIVISLCQNGIKVMVLREEEHCMLAGYLGGHCKE 484
CER1-like_2            AEVVLHSIPVGTKEVLFRGQITKVARAIVFSLCQNAIKVMVLRKEEHSMLAEFLDDKCKE 493
CER1-like_1            ATVVINNIPKEATEIVFRGNLTKVASAVVFALCQKGVKVVVLREEEHSKLIKSGVDKNLV 501
cer1_8hao_protein      AAVVINSLPKSTTKIVMTGNLTKVAYTIASALCQRGVEVLTLLPEEYEKLSSFVPKECRD 655
ECERIFERUM_1           AAVVINSVPKATTSVVMTGNLTKVAYTIASALCQRGVQVSTLRLDEYEKIRSCVPQECRD 496
                       * **::.:*  : .::: *::**** .:. :*** .::* .*  :*:  :      .

cer1_3hao_protein      NLVLTTNDYP---MIWLVGDGLSKKEQKLARKRTLFIPYSQFPPRELRKNCFYHTTPAMI 541
cer1_4hao_protein      NLVLTTNYSP---MIWLVGDGLSREEQEMATKGTRFLPFSQFPPTQLRKDCFYHTTPAMI 541
CER1-like_2            N------------LIWLVGDGLSTKEQKMAKDGTLFLPFSQFPPKTLRKDCFYHTTPAMI 541
CER1-like_1            LSTSNSYYSP---KVWLVGDGIENEEQMKAKEGTLFVPFSHFPPNKLRKDCFYQSTPAMR 558
cer1_8hao_protein      RLILLTSETLASNKVWLMGEGTTREEQEMATKGTLFIPFSQFPLKQLRRDCIYHTPPALI 715
ECERIFERUM_1           HLVYLTSEALSSNKVWLVGEGTTREEQEKATKGTLFIPFSQFPLKQLRRDCIYHTTPALI 556
                                     :**:*:*   :**  * . * *:*:*:**   **::*:*::.**:

cer1_3hao_protein      IPDSAQNIDSCENWLGRRVMSAWRVGGIVHALEGWEEHECGLEV-PMVNPPRVWEAALRN 600
cer1_4hao_protein      IPDSAQNIDSCENWLGRRVMSAWRVGGIVHALEGWEEHECGLDG-PVVNPPRVWEAALRN 600
CER1-like_2            IPHSAQNIDSCENWLGRRVMSAWRVGGIVHALEGWKEHECGLDDNSIINPPRVWEAALRN 601
CER1-like_1            VPKSAQNIDSCENWLGRRVMSAWKIGGIVHALEGWEEHDCGNTC-NVLRLHAIWEAALRH 617
cer1_8hao_protein      IPKSLVNIHSCENWLPRKAMSATRVAGILHALEGWETHECGTSNILLSDLDQVWEACLSH 775
ECERIFERUM_1           VPKSLVNVHSCENWLPRKAMSATRVAGILHALEGWEMHECGTS-LLLSDLDQVWEACLSH 615
                       :*.*  *:.****** *:.*** ::.**:******: *:**     :     :***.* :

cer1_3hao_protein      GFKPLVFPSLDVVY----- 614
cer1_4hao_protein      GFKPLVLPSVETKGLSNCY 619
CER1-like_2            GFQPLLLPSLET------- 613
CER1-like_1            DFQPLPPSP---------- 626
cer1_8hao_protein      GFQPLLLPHHFQYP----- 789
ECERIFERUM_1           GFQPLLLPHH--------- 625
                       .*:**  .
```

多序列比对



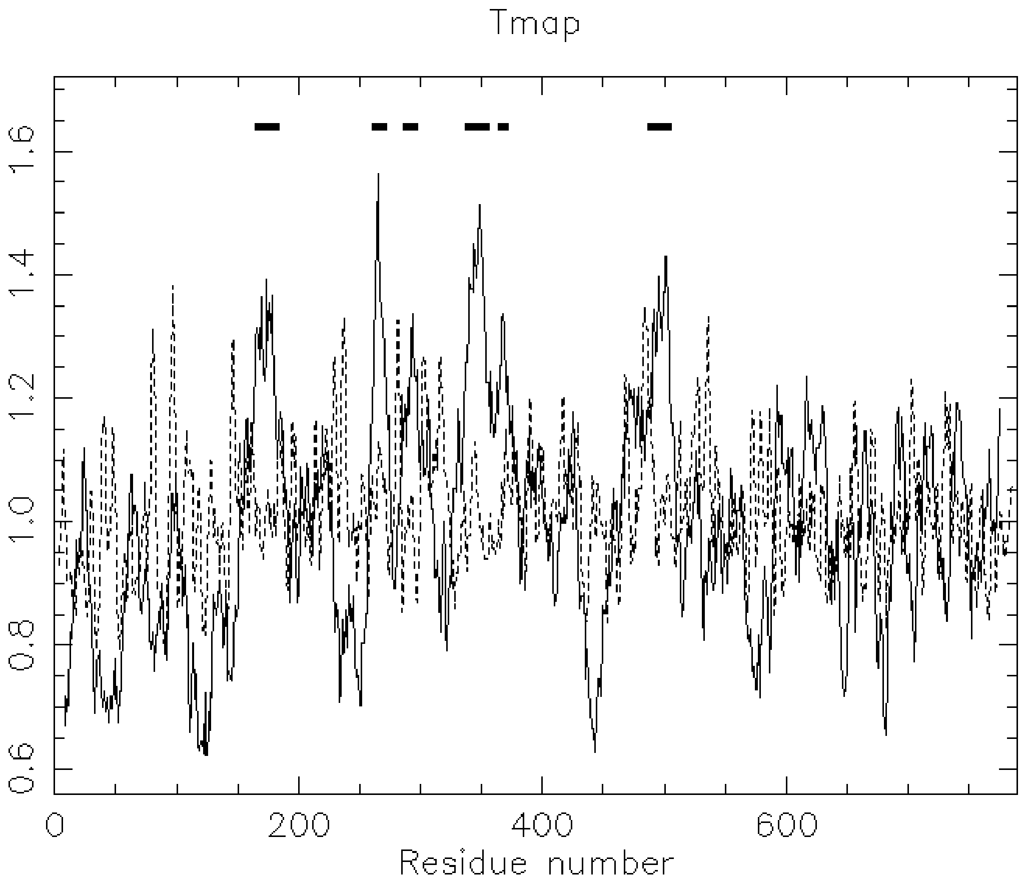
SeqA	Name	Length	SeqB	Name	Length	Score
1	ECERIFERUM_1	625	2	CER1-like_2	613	57.0
1	ECERIFERUM_1	625	3	CER1-like_1	627	57.0
1	ECERIFERUM_1	625	4	cer1_8hao_protein	789	83.0
1	ECERIFERUM_1	625	5	cer1_3hao_protein	614	54.0
1	ECERIFERUM_1	625	6	cer1_4hao_protein	619	56.0
2	CER1-like_2	613	3	CER1-like_1	627	68.0
2	CER1-like_2	613	4	cer1_8hao_protein	789	56.0
2	CER1-like_2	613	5	cer1_3hao_protein	614	79.0
2	CER1-like_2	613	6	cer1_4hao_protein	619	83.0
3	CER1-like_1	627	4	cer1_8hao_protein	789	55.0
3	CER1-like_1	627	5	cer1_3hao_protein	614	64.0
3	CER1-like_1	627	6	cer1_4hao_protein	619	66.0
4	cer1_8hao_protein	789	5	cer1_3hao_protein	614	53.0
4	cer1_8hao_protein	789	6	cer1_4hao_protein	619	54.0
5	cer1_3hao_protein	614	6	cer1_4hao_protein	619	82.0

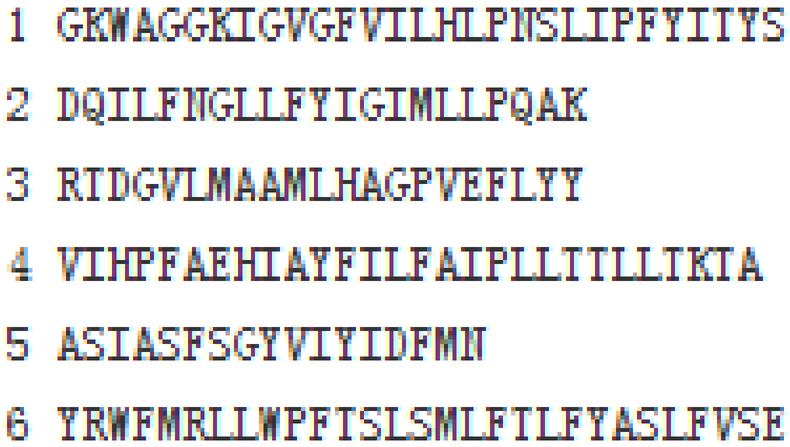
进化分析



Cer1-like_2和cer1-like_1是拟南芥中cer1的另外2个拷贝
固醇脱氢酶，甲基甾醇单加氧酶，cer3是另一个与相关蛋白。

甘蓝cer1_8hao结构预测





1 GKWAGGKIGVGFVILHLPNSLIPFYITYS

2 DQILFNGLLFYIGIMLLPQAK

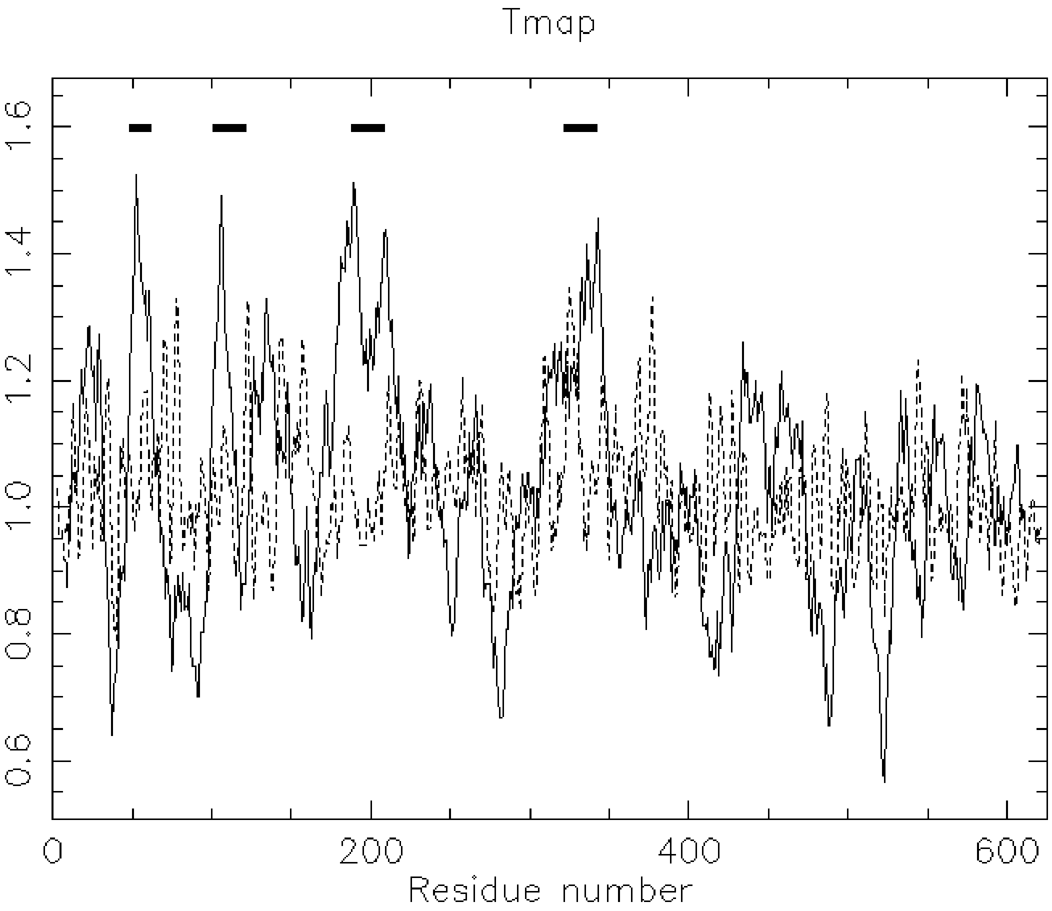
3 RIDGVLMAAMLHAGPVEFLYY

4 VIHPFAEHIAYFILFAIPLLITILLIKTA

5 ASIASFSGYVIYIDFMN

6 YRWFMRLLWPFTSLSMLFTILFYASLFVSE

Cer1疏水结构预测



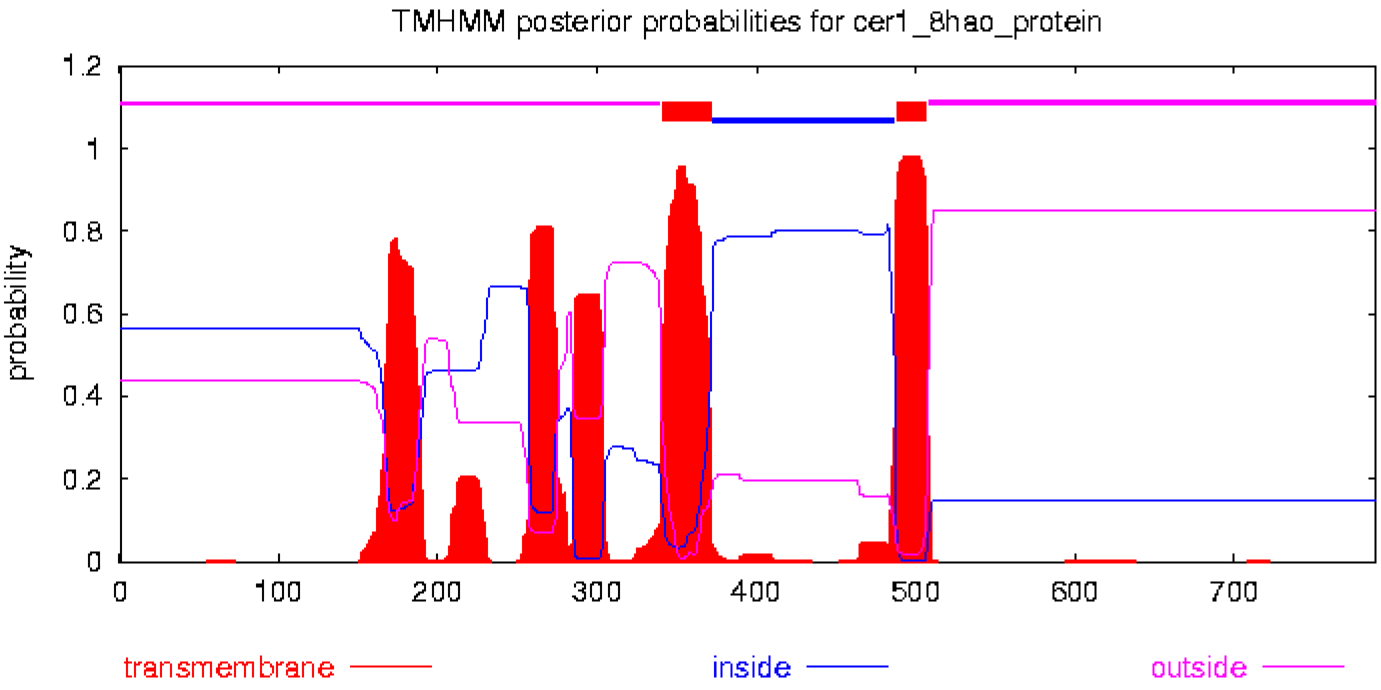
1 DLGYFLVFPFLLFRILHNQVWI

2 DQILFNGVLFYIGINLLPEAKQLPWWRID

3 YFILFAIPLLITILLIKTASIISFAGYIIY

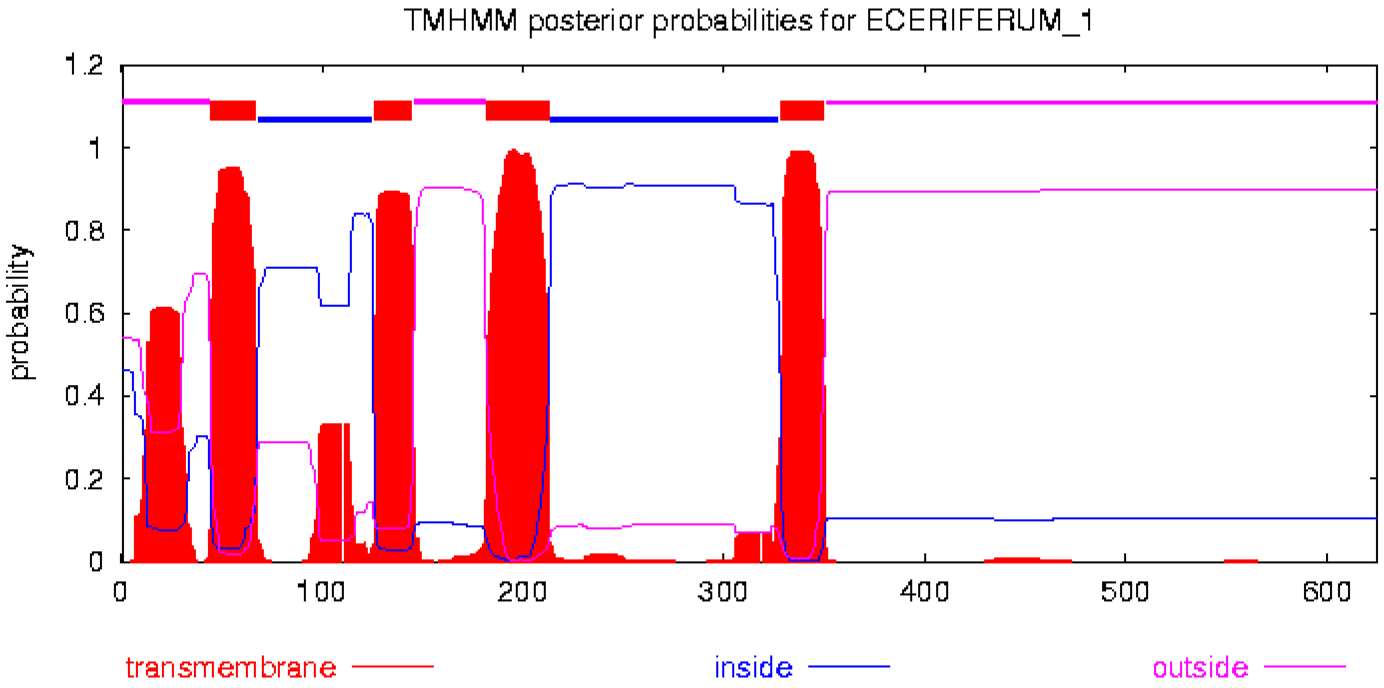
4 SYPFAYRWFMRLLWPFTSLSMIFILFYAR

甘蓝cer1_8hao跨膜结构分析



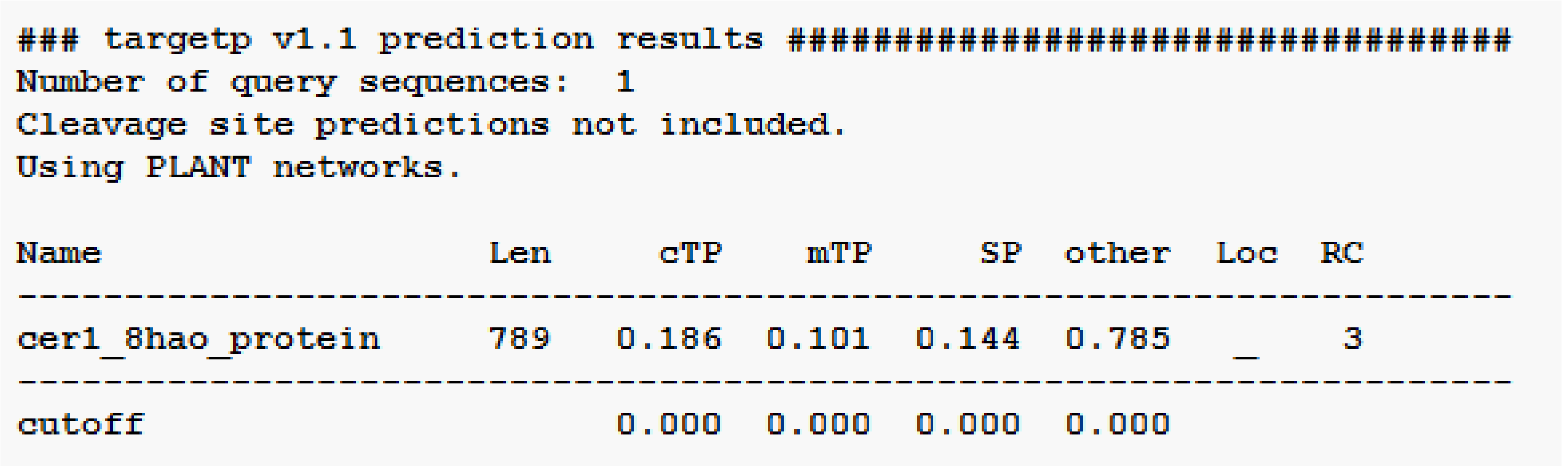
由图可知，此蛋白具有2个跨膜结构，1-340outside，341-372TMhelix，373-487inside，488-507TMhelix，508-789outside。

Cer1跨膜结构



如图，包含4个跨膜结构，cer1亚细胞定位在ER上。

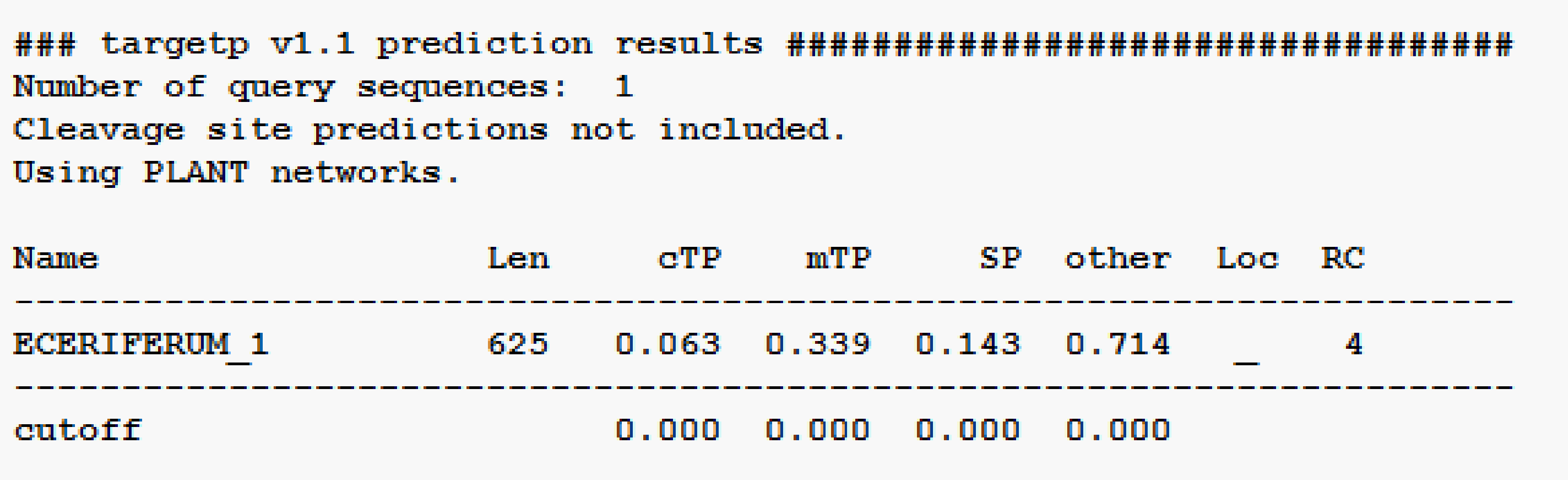
Cer1_8hao信号肽及亚细胞定位预测



```
### targetp v1.1 prediction results ###################################
Number of query sequences:  1
Cleavage site predictions not included.
Using PLANT networks.

Name                  Len     cTP    mTP     SP  other  Loc  RC
-----------------------------------------------------------------------
cer1_8hao_protein     789   0.186  0.101  0.144  0.785   _    3
-----------------------------------------------------------------------
cutoff                      0.000  0.000  0.000  0.000
```

不含信号肽，

Cer1信号肽及亚细胞定位预测



```
### targetp v1.1 prediction results ##################################
Number of query sequences:  1
Cleavage site predictions not included.
Using PLANT networks.

Name                  Len     cTP    mTP     SP  other  Loc  RC
----------------------------------------------------------------------
ECERIFERUM_1          625   0.063  0.339  0.143  0.714   _     4
----------------------------------------------------------------------
cutoff                      0.000  0.000  0.000  0.000
```

无信号肽

谢谢