

黄瓜对WMV的抗性遗传分析



组别：G01

组员：朱萍萍 王晓伟 田桂丽 肖纯梅

报告人：田桂丽

2013.6.23

研究必要性



- 目前黄瓜是我国第一大保护地栽培作物
- 病毒病（WMMV、ZYMV和PRSV）发生严重
- 研究植物的相关抗病基因能够比使用药剂等进行更有效地防治，也是未来防治病害的重要途径。





可行性

- ▶ 利用表型和分子方法进行WMV抗性的遗传分析及基因定位，确定了其是有一对隐性单基因控制的。
- ▶ 比对黄瓜基因组测序结果，找出了其在染色体上的物理位置，并对目标区域内的隐性抗性基因进行分析，最终确定了抗性基因 *WMV-1* 的一个候选基因 *eIF4E*。
- ▶ *eIF4E* 属于真核生物翻译起始因子家族，在很多物种的研究中，以表明其与马铃薯Y病毒属的抗性有关。



一、植物自身抗病毒作用方式

- ▶ **主动抗性**：由显性抗性基因R基因介导的抗性。首先引起植物的超敏反应，再引起植物的系统获得抗性，它可以使植物远离病毒侵染位点的正常组织获得对相同或相近的免疫力。
- ▶ **被动抗性**：在植物中普遍存在，通过缺失或突变一些对病毒在植物体内生存起至关重要作用的寄主因子，从而达到使病毒在植物体内不能繁殖的作用，多由隐性抗性基因控制。



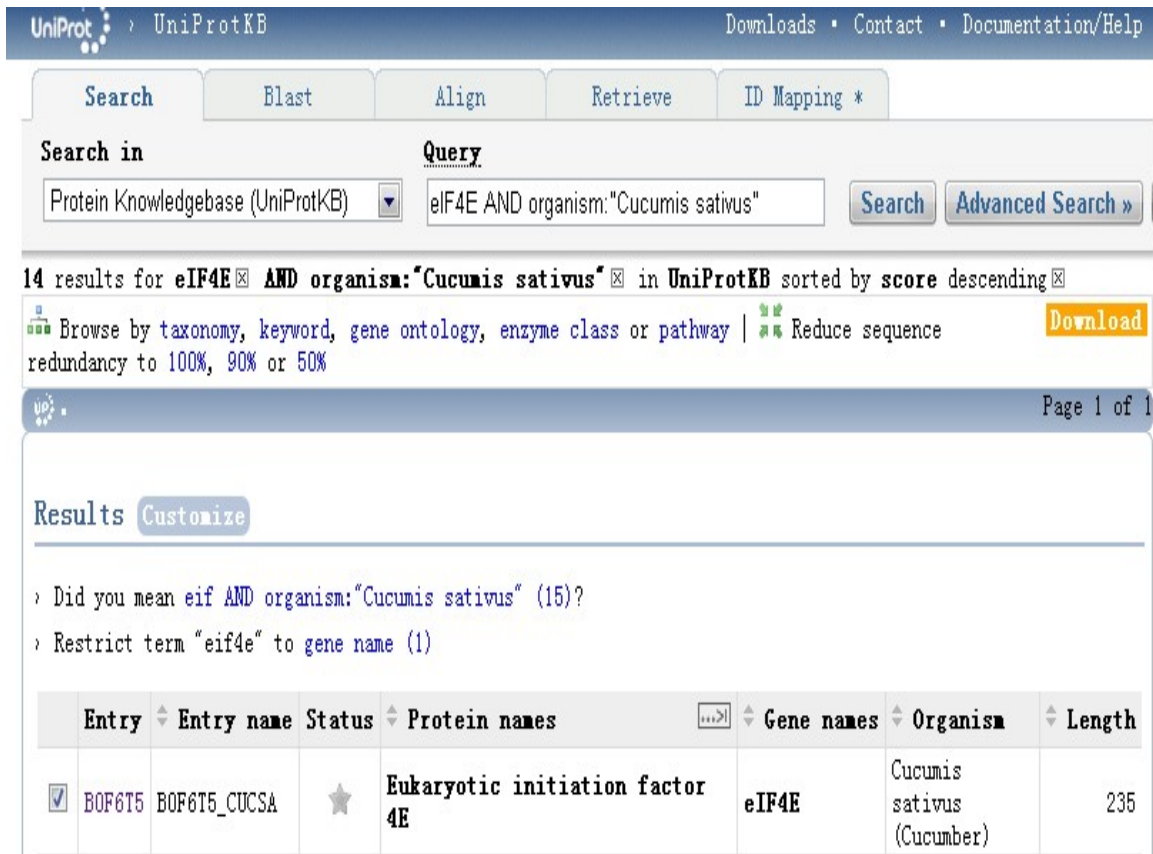
1.2 *eIF4E* 的抗病毒机理

- 所有真核细胞的mRNA 在5'端都有一个帽子结构。
- *eIF4E* 又称帽结合蛋白(cap binding protein, CBP), 能识别mRNA的帽结构, 起始mRNA翻译。
- 马铃薯Y 病毒属病毒的VPg 蛋白位于病毒mRNA的5'端, 其类似于真核生物的mRNA的5'端帽子结构, 与植物的*eIF4E* 结合, 起始病毒的翻译, 这是病毒感染的必须条件。

二、黄瓜中*eIF4E*基因的研究



在UniProt中查到该基因 *eIF4E* 的信息如下



The screenshot shows the UniProt search interface. The search query is "eIF4E AND organism: 'Cucumis sativus'". The results are sorted by score descending. The first result is for the entry B0F6T5, which is the Eukaryotic initiation factor 4E from Cucumis sativus (Cucumber). The protein length is 235 amino acids.

UniProt > UniProtKB Downloads • Contact • Documentation/Help

Search Blast Align Retrieve ID Mapping *

Search in Protein Knowledgebase (UniProtKB) Query eIF4E AND organism: "Cucumis sativus" Search Advanced Search »

14 results for eIF4E AND organism: "Cucumis sativus" in UniProtKB sorted by score descending

Browse by taxonomy, keyword, gene ontology, enzyme class or pathway | Reduce sequence redundancy to 100%, 90% or 50% Download

Page 1 of 1

Results Customize

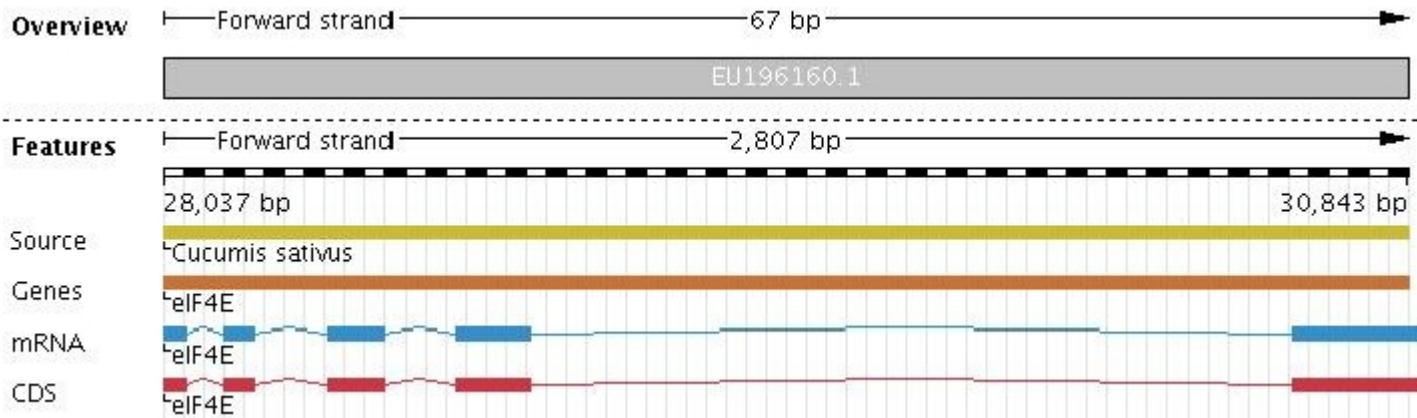
> Did you mean eif AND organism: "Cucumis sativus" (15)?

> Restrict term "eif4e" to gene name (1)

Entry	Entry name	Status	Protein names	Gene names	Organism	Length
<input checked="" type="checkbox"/> B0F6T5	B0F6T5_CUCSA	★	Eukaryotic initiation factor 4E	eIF4E	Cucumis sativus (Cucumber)	235

- 在葫芦科中是未经人工审核的一个翻译起始因子基因，翻译出的蛋白序列长度为235个氨基酸，在GenBank中可以找到其有关信息，其序列号为ABY56085.1。

这段序列是构建的黄瓜质粒文库中全基因序列的一部分，这个全基因序列中有5个编码区，而翻译起始因子 *eIF4E* 只是其中一个。

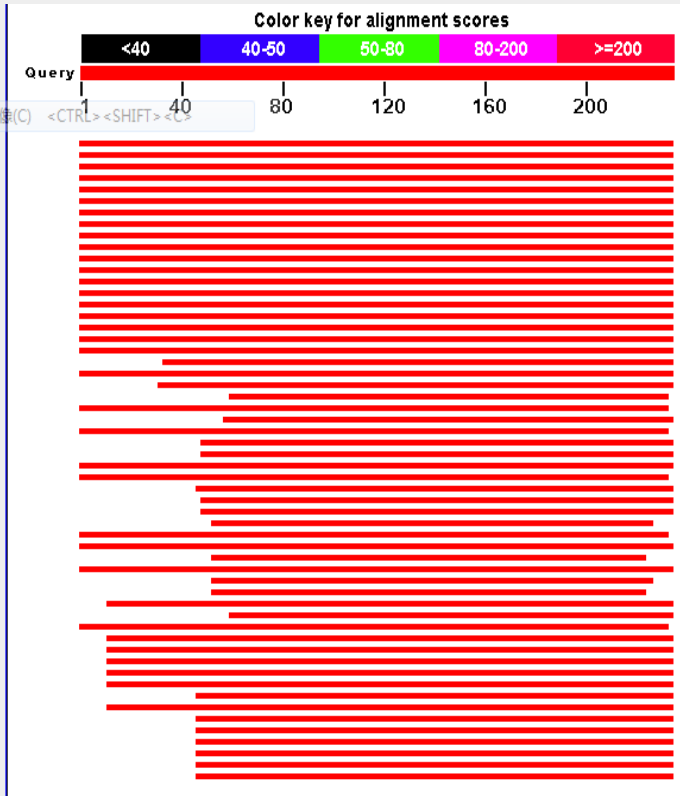
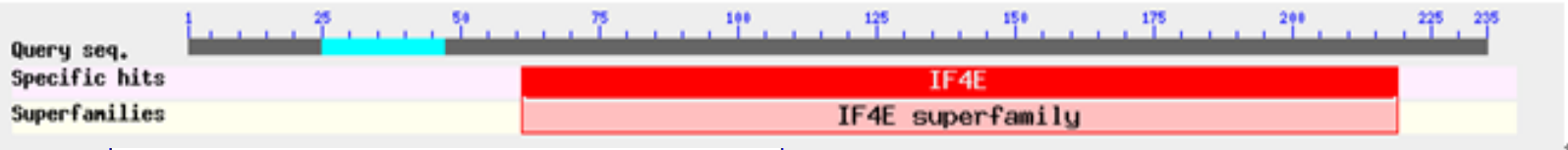


2.2 该基因的核酸序列以及对应氨基酸序列

```

SHOWORF of c from 1 to 708
-----|-----|-----|-----|
F1      1 ATGGTAGTTGAAGATACGAATCAAAGCTACATCGGGCGGAAGATCTTTCTAA 50
      1 M V V E E D T I K A T S A E D L S N 17
-----|-----|-----|-----|
F1     51 TTCCATTGCTAATCAAAAACCCCTAGAGGACGTGGCGGTGAGGAAAGATGAGG 100
      18 S I A N Q N P R G R G G E E D E E 34
-----|-----|-----|-----|
F1    101 AACTTGAGGAAGGTGAGATCGTCGGCGATGACGACCTTGACTCCTCCAAT 150
      35 L E E G E I V G D D D L D S S N 50
-----|-----|-----|-----|
F1    151 TTGACCGCGGCGCTTAGTTCATCAGCCTCACCCCTCTTGAGCACTCTTGGAC 200
      51 L T A A L V H Q P H F L E H S W T 67
-----|-----|-----|-----|
F1    201 CTTTTGGTTTGATAACCCATCTGCCAAATCCAAGCAAGCCACCTGGGGTG 250
      68 F W F D N P S A K S K Q A T W G A 84
-----|-----|-----|-----|
F1    251 CGTCTATGCGACCGATCTATACTTCTCCACTGTCGAGGAGTTCTGGAGT 300
      85 S M R P I Y T P S T V E E P W S 100
-----|-----|-----|-----|
F1    301 GTTACAACAATATTCAATCCCAAGCAAATTGGCGTTGAGGGCAGATT 350
     101 V Y N N I H H P S K L A L R A D L 117
-----|-----|-----|-----|
F1    351 G T A C T G C T T C A A A C A T A A A A T T G A G C C T A A A T G G G A A G A T C C T G T T T G T G 400
     118 Y C F K H K I E P K W E D P V C A 134
-----|-----|-----|-----|
F1    401 CGAGTGGAGGGGAAGTGGACTGTGAACTTTTCAAGGGGAAAATCTGATAAT 450
     135 S G G K W T V N F S R G K S D N 150
-----|-----|-----|-----|
F1    451 GCCTGGTTGTACAGCTGCTTGCCTATGATCGGAGAACAGTTTGACTGTGG 500
     151 G W L Y T L L A M I G E Q F D C G 167
-----|-----|-----|-----|
F1    501 TGATGAAATTTGTGGAGCAGTTGTTAATGTTAGATCTGGGCAGGATAAAA 550
     168 D E I C G A V V N V R S G Q D K I 184
-----|-----|-----|-----|
F1    551 TATCGATTTGGACAAAAAATGCTTCCAATGAAGCTGCGCAGGCGAGCATT 600
     185 S I W T K H A S N E A A Q A S I 200
-----|-----|-----|-----|
F1    601 GGAAAACAGTGGAAAGGAGTTTCTTGATTACAATGAGAGCATTGGCTTTAT 650
     201 G K Q W K E F L D Y N E S I G F I 217
-----|-----|-----|-----|
F1    651 ATTCCACGATGATGCAAAGAAATTTGATAGACATGCCGAGAATAAATATA 700
     218 F H D D A K K F D E H A K N K Y M 234
-----|-----|-----|-----|
F1    701 TGGTGGGA 708
     235 V * 235
  
```

利用黄瓜该基因的蛋白序列，在 NCBI 中进行BLAST，得到的结果如下

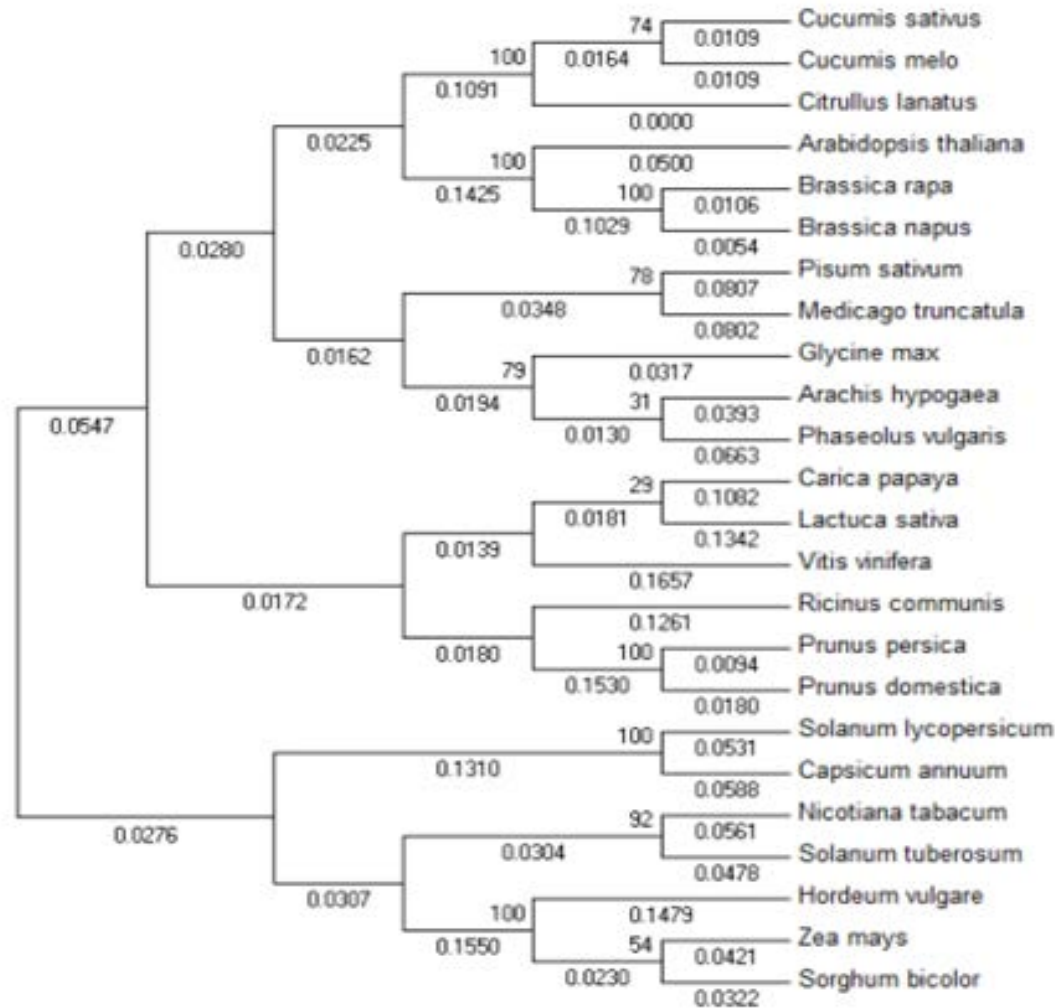


比对结果相似性较高的有很多个，所以可以看出该基因是在大多数生物中都普遍存在的。

根据E值、Query cover值选取了除黄瓜以外的23个物种，进行了系统发育分析。

Organism↵	Organism↵	Organism↵
<i>Cucumis sativus</i> 黄瓜↵	<i>Carica papaya</i> 番木瓜↵	<i>Arabidopsis thaliana</i> 拟南芥↵
<i>Cucumis melo</i> 甜瓜↵	<i>Nicotiana tabacum</i> 烟草↵	<i>Sorghum bicolor</i> 高粱↵
<i>Citrullus lanatus</i> 西瓜↵	<i>Brassica rapa</i> 白菜↵	<i>Solanum lycopersicum</i> 番茄↵
<i>Pisum sativum</i> 豌豆↵	<i>Brassica napus</i> 油菜↵	<i>Vitis vinifera</i> 葡萄↵
<i>Arachis hypogaea</i> 花生↵	<i>Prunus persica</i> 碧桃↵	<i>Ricinus communis</i> 蓖麻↵
<i>Glycine max</i> 大豆↵	<i>Prunus domestica</i> 欧洲李↵	<i>Solanum tuberosum</i> 马铃薯↵
<i>Medicago truncatula</i> 苜蓿↵	<i>Zea mays</i> 玉米↵	<i>Phaseolus vulgaris</i> 菜豆↵
<i>Lactuca sativa</i> 莴苣↵	<i>Capsicum annuum</i> 甜椒↵	<i>Hordeum vulgare</i> 大麦↵

为了推断该基因家族的起源和演化进行系统发育分析





系统发育树分析

- 与黄瓜该基因相似性较高的是甜瓜，两者的该基因可能来源于西瓜，都与其起源的祖先距离较远。
- 西瓜中该基因与其起源的祖先距离近很多。

*eIF4E*在黄瓜与西瓜的一致性高达94.89%

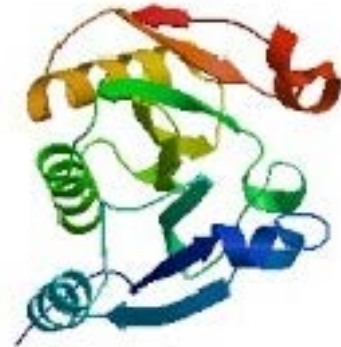
gi Citrullus_lanatus	E	T	T	D	40
gi Cucumis_sativus	D	A	S	E	40
Consensus	mvve tikats edlsn ianqnprgrgg edeeleegei				
gi Citrullus_lanatus		S	I	G	80
gi Cucumis_sativus		T	L	D	80
Consensus	vgddldssnl aa vhpplshswtfwf npsakskqa				
gi Citrullus_lanatus	I				120
gi Cucumis_sativus	M				120
Consensus	twgas rpiytfstveefwsvynnihhpsklalradlycf				
gi Citrullus_lanatus		N			160
gi Cucumis_sativus		S			160
Consensus	khkiepkwedpvca gqkwtvnfsrgksdngwlytllami				
gi Citrullus_lanatus					200
gi Cucumis_sativus					200
Consensus	geqfcdcgdeicgavvnvrsgqdkisiwtknasneaaqasi				
gi Citrullus_lanatus		D	E	S	234
gi Cucumis_sativus		E	D	M	234
Consensus	gkqwkefldyn sigfifh dakkfdrhaknky				

2.3 WMV-1三级结构的相似比对模型（西瓜）

在PDB

（DOI:[10.2210/pdb2wmc/pdb](https://doi.org/10.2210/pdb2wmc/pdb)）
数据库中，找到了西瓜的该蛋白的
三级结构

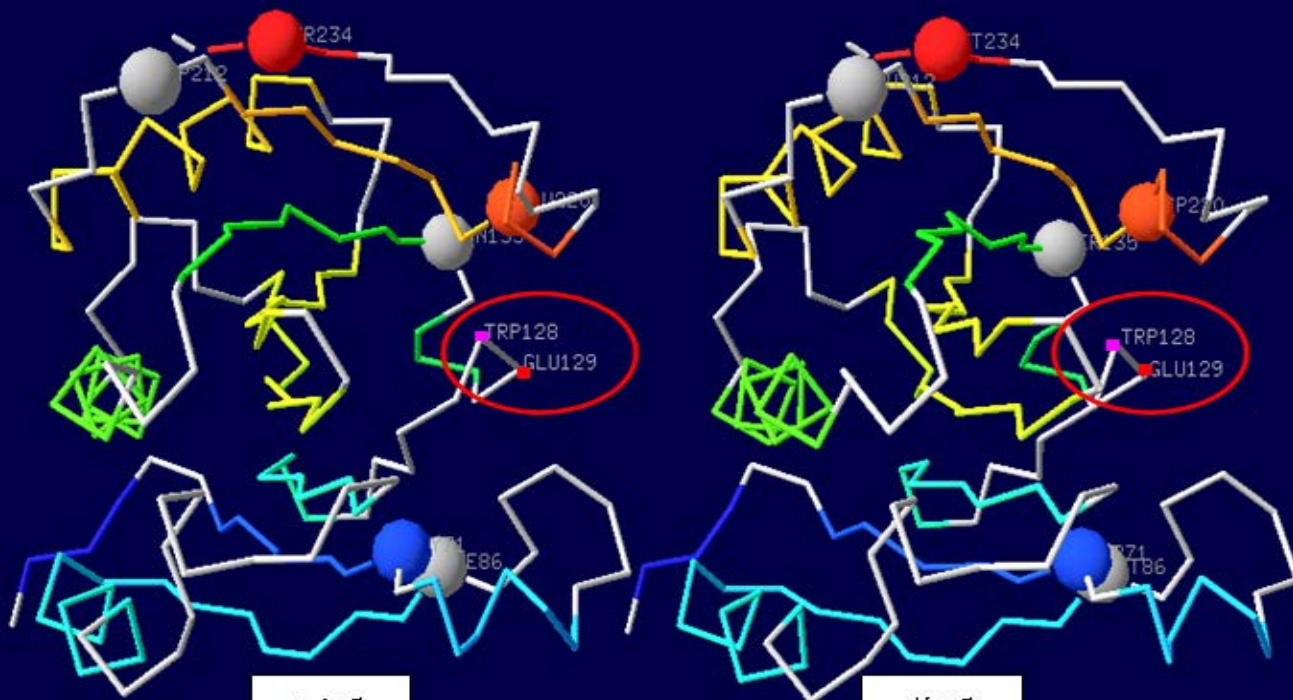
在Swiss Model中，预测
WMV-1蛋白的三级结构



各区域颜色不同表示所预测结构的可靠性，蓝色表示可靠性最高，红色表示可靠性最低。

eIF4E 黄瓜与西瓜的差异位点

Cucumis sativus (1155 x 554)



西瓜

黄瓜

活性位点及黄瓜与西瓜的差异位点

gi Pisum_sativum	MVVEDTPKSIITDDQITTNPM....RVIEDDMNLEEGER	35
gi Citrullus_lanatus	MVVEETIKATSTEDLSNTIAMONPRGRGGDEDEELEEGER	40
gi Cucumis_sativus	MVVEDTIKATSÆDLSNSIAMONPRGRGGEEDEELEEGER	40
Consensus	m vve t k d n r d leegei	
gi Pisum_sativum	LDDED..DSSATSKPWWHQPHLENSWTFWFDPAAKSKQA	73
gi Citrullus_lanatus	VGDDDLSSNLSAAIVHQPHLEHSWTFWFGMPSAKSKQA	80
gi Cucumis_sativus	VGDDDLSSNLTAAALVHQPHLEHSWTFWFDPMPAKSKQA	80
Consensus	d dss v hqph le swtfwf p akskqa	
gi Pisum_sativum	ANGSSMRPIYTFSTVEEFWSIYNNIHHPGKLVGADFYCF	113
gi Citrullus_lanatus	TWGASIRPIYTFSTVEEFWSVYNNIHHPKLAALRADLYCF	120
gi Cucumis_sativus	TWGASMRPIYTFSTVEEFWSVYNNIHHPKLAALRADLYCF	120
Consensus	w g s i r p i y t f s t v e e f w s y n n i h h p k l a a l r a d l y c f	
gi Pisum_sativum	KHKIEPKWEDPVCANGGKWTVMNFSRGKSDNGWLYTLLAMI	153
gi Citrullus_lanatus	KHKIEPKWEDPVCANGGKWTVMNFSRGKSDNGWLYTLLAMI	160
gi Cucumis_sativus	KHKIEPKWEDPVCANGGKWTVMNFSRGKSDNGWLYTLLAMI	160
Consensus	khkiepkwedpvcanggkwtvmnfsrgksdngwlytllami	
gi Pisum_sativum	GEQFDHGDEICGAVVNVRGRAEKISIWTKNASNEAAQVSI	193
gi Citrullus_lanatus	GEQFDCGDEICGAVVNVRSGQDKISIWTKNASNEAAQASI	200
gi Cucumis_sativus	GEQFDCGDEICGAVVNVRSGQDKISIWTKNASNEAAQASI	200
Consensus	geqfd gdeicgavvnvr kisiwtknasneaq si	
gi Pisum_sativum	GKQWKEFLDYNETMGFIFHIDAKKLDNRNAKNRYV	227
gi Citrullus_lanatus	GKQWKEFLDYNDISIGFIFHEDAKKFDNRHAKNRYV	234
gi Cucumis_sativus	GKQWKEFLDYNETMGFIFHIDAKKLDNRNAKNRYV	234
Consensus	gkqwkefldynetmgfifhidakkldrnaaknryv	

- 黄瓜与西瓜的差异位点的生化性质并无太大改变，因此 *eIF4E* 在黄瓜中的功能极有可能与西瓜的一致。



总结

- ▶ 通过生物信息技术分析初步推测 *WMV-1* 基因为翻译起始因子 *eIF4E*，与黄瓜 WMV 的病毒抗性有极其重要的关系。
- ▶ 下步利用基因工程的方法进行 *WMV-1* 基因功能的验证。

致谢

- 衷心感谢 **罗老师** 对我们的悉心指导和谆谆教诲，让我们发现和体会到了生物信息学的乐趣和价值！
- 非常感谢组员及同学们的支持和帮助，使我们的学习更轻松！

Thank you !

