

doi:10.3969/j.issn.1672-5565.2014.04.02

# 苏太猪 *ELOVL7* 基因的生物信息学分析

章峰, 李景, 黄路生, 张志燕\*

(江西农业大学动物生物技术重点开放实验室, 南昌 330045)

**摘要:**不同于人、鼠等物种 *ELOVL7* 基因的高相似度, 不同品种的猪 *ELOVL7* 基因相似度较低。为了探究该基因的特性, 本研究运用生物信息学的方法对苏太猪 *ELOVL7* 基因及其氨基酸序列的同源性、理化性质、保守结构域、亚细胞定位、信号肽、跨膜结构域、亲水性/疏水性、二级结构、功能预测以及磷酸化位点等进行预测分析。结果表明: 在苏太猪中, *ELOVL7* 全长 2 324 bp, 编码区为 846 bp, 共编码 281 个氨基酸。其结构稳定, 分子量为 33 387.4 Da, 带正电荷, 偏碱性。该基因所编码的蛋白质最可能位于细胞膜上, 主要的功能是运输和结合, 为跨膜、非分泌型疏水蛋白质, 含有 1 个 GNS1/SUR4 家族的保守结构域, 并有 15 个丝氨酸激酶、15 个苏氨酸激酶和 20 个酪氨酸激酶潜在磷酸化位点。 $\alpha$  螺旋是 *ELOVL7* 二级结构和三级结构中最主要的结构元件。另外 *ELOVL7* 与大部分物种的氨基酸序列相似性达 90% 以上, 且亲缘关系较近。分析 *ELOVL7* 基因及其氨基酸序列的特征, 能够为进一步挖掘该基因内的突变对长链脂肪酸表型的影响以及合成、代谢机理提供分子依据。

**关键词:** *ELOVL7*; 脂肪酸; 生物信息学

**中图分类号:** Q754    **文献标志码:** A    **文章编号:** 1672-5565(2014)-04-242-07

## Bioinformatics analysis of *ELOVL7* gene in Sutai pig

ZHANG Feng, LI Jing, HUANG Lusheng, ZHANG Zhiyan\*

(Jiangxi Agricultural University, Key Laboratory for Animal Biotechnology of Jiangxi Province and the Ministry of Agriculture of China, Nanchang 330045, China)

**Abstract:** Differ to the high similarity between human and mice, the similarity of gene *ELOVL7* among pig breeds is low. To uncover this specific character, we analyzed some character of the *ELOVL7* gene and its encoding amino acid in Sutai pigs by bioinformatics method, including homology, physical and chemical properties, conserved domains, subcellular localization, signal peptide, transmembrane domain, hydrophilic/hydrophobic, secondary structure, function and kinase-specific phosphorylation site. The results showed that the length of *ELOVL7* is 2 324 bp and its 846 bp coding region encodes a protein with 281 amino acids. The structure of *ELOVL7* is stable, with 33 387.4 Da molecular weight, positive charge and basicity. The encoded protein of *ELOVL7* gene is a transmembrane, nonsecretory and hydrophobic protein, which most probably is located in cell membrane, contains a conserved domain belonging to GNS1/SUR4 family and has 15 serine, 15 threonine and 20 tyrosine underlying phosphorylated sites. Moreover the similarity of amino acid sequence is over 90% with high homology and its function is mainly involved with transport and binding. The corresponding secondary structure and tertiary structure are primarily composed of  $\alpha$ -helix. Analyzing some characters of the *ELOVL7* gene and its amino acid would benefit the study of impaction of potential mutation of *ELOVL7* to the syntetic of arachidic acid, and provide insights into the molecular basis of fatty acid metabolism.

**Keywords:** *ELOVL7*; Fatty acids; Bioinformatics

收稿日期: 2014-09-26; 修回日期: 2014-10-20.

基金项目: 国家自然科学基金(31200926)。

作者简介: 章峰, 男, 在读博士, 研究方向: 生物信息学; E-mail: fengzhang0709@hotmail.com.

\* 通信作者: 张志燕, 男, 博士, 研究方向: 数量遗传学; E-mail: bioducklily@hotmail.com.

脂肪酸(FA)是细胞的基本组成部分,在机体内对存储能量,信息传递及代谢调控等方面起着重要的作用。FA是由一条线性长的碳氢链(疏水尾)和一个末端羧基(亲水头)组成的羧酸。根据其碳氢链内是否含有双键又可分饱和脂肪酸和不饱和脂肪酸两类。高含量的单不饱和脂肪酸和多不饱和脂肪酸,如人体内必需脂肪酸 $\omega-3$ 系列,有利于人类的心血管健康<sup>[1]</sup>。然而高含量的饱和脂肪酸则会增加得心脏疾病的风险<sup>[2]</sup>。因此,FA是维持机体内的内稳态的重要成分。

由于家猪与人类的生理生化特性高度地相似,因此常常被当作一种研究人类遗传疾病的模式动物。杨斌等利用282头苏太猪进行全基因组关联分析,鉴别到影响花生酸的候选基因是位于16号染色体上的*ELOVL7*基因<sup>[3]</sup>。*ELOVL*家族是超长链脂肪酸延长关键步骤的催化剂<sup>[4]</sup>。此外,过量的摄入脂肪会促进*ELOVL7*过表达的前列腺癌细胞的生长,敲除*ELOVL7*基因后前列腺癌细胞则会衰减<sup>[5]</sup>。因此*ELOVL7*可能是阐明脂肪摄入量与前列腺癌关联的关键因素。

为了深入挖掘苏太猪中影响花生酸的因果突变位点,我们对苏太猪的互补脱氧核糖核酸(cDNA)进行测序。结果发现苏太猪cDNA与欧洲野猪cDNA(Ensemble)的相似性只有86%,差异性较大,具有一定的种群特异性,但在人、鼠之间的相似度高达90%以上。为了研究苏太猪*ELOVL7*基因的相关性特性,本研究运用生物信息学的方法对苏太猪*ELOVL7*基因及其氨基酸序列的同源性、理化性质、保守结构域、亚细胞定位、信号肽、跨膜结构域、亲水性/疏水性、二级结构、功能预测以及磷酸化位点等进行预测分析,旨在为花生酸因果突变位点的筛选以及机体内脂肪酸的内稳态机制提供分子依据。

## 1 材料和方法

### 1.1 *ELOVL7* 基因序列的获取

#### 1.1.1 实验样品

本研究所涉及的实验动物为苏太猪,是杜洛克公猪与梅山猪杂交25代的子代。有关苏太猪详细的特征及管理可见章峰等<sup>[6]</sup>。简单来说,所有苏太猪均进行标准化饲养和管理,并在240日龄时进行屠宰。采集适量腹脂,装入RNA free的微型管,并放入-80℃冰箱中保存。

#### 1.1.2 主要实验步骤

(1)使用TRNzol-A+总RNA提取试剂盒(TIANGEN)进行RNA的提取,得到总RNA;

(2)使用PrimeScript™ RT reagent Kit with gDNA Eraser试剂盒(TaKaRa)进行反转录反应,得到全基因组cDNA;

(3)以Ensembl网站Sus scrofa *ELOVL7*的cDNA序列为模板设计引物,进行聚合酶链式反应(PCR)。

(4)将得到的扩增产物委托生工生物工程(上海)股份有限公司进行双向测序,最终得到苏太猪*ELOVL7*的cDNA序列。

### 1.2 生物信息学分析

运用NCBI和ProtParam进行核酸和氨基酸序列组成和理化性质分析;运用NCBI、PSORT、SignalP、TMHMM、ProtScale和KinasePhos、ProFun等在线工具进行保守结构域、亚细胞定位、信号肽、跨膜结构、亲水性/疏水性和激酶磷酸化修饰位点、蛋白质功能预测;利用PSIPRED和Phyre2进行二级结构预测和三级结构预测,并用Pymol绘制三级分子结构图;运用NCBI和MEGA进行氨基酸序列的序列比对和同源性分析。所用软件或在线网站有关信息见表1。

表1 预测基因结构和功能应用的网站及软件

Table 1 The websites and software for predicting the gene structure and function

| 编号 | 名称                         | 网址   | 主要功能                  |
|----|----------------------------|--|-----------------------|
| 1  | ProtParam                  | <a href="http://web.expasy.org/protparam/">http://web.expasy.org/protparam/</a>  | 对蛋白质理化性质的预测           |
| 2  | NCBI                       | <a href="http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi">http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi</a>  | 分析并找到序列的ORF区          |
| 3  | NCBI                       | <a href="http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi">http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi</a><br><a href="http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi">http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi</a> | 保守结构域预测<br>氨基酸序列相似性分析 |
| 4  | PSORT <sup>[7]</sup>       | <a href="http://psort.hgc.jp/form.html">http://psort.hgc.jp/form.html</a>  | 亚细胞定位预测               |
| 5  | SignalP <sup>[8]</sup>     | <a href="http://www.cbs.dtu.dk/services/SignalP/">http://www.cbs.dtu.dk/services/SignalP/</a>  | 信号肽预测                 |
| 6  | TMHMM <sup>[9]</sup>       | <a href="http://www.cbs.dtu.dk/services/TMHMM/">http://www.cbs.dtu.dk/services/TMHMM/</a>  | 跨膜结构域预测               |
| 7  | ProtScale                  | <a href="http://web.expasy.org/protscale/">http://web.expasy.org/protscale/</a>  | 亲水性/疏水性分析             |
| 8  | KinasePhos <sup>[10]</sup> | <a href="http://kinasephos2.mbc.nctu.edu.tw/index.html">http://kinasephos2.mbc.nctu.edu.tw/index.html</a>  | 激酶磷酸化修饰位点分析           |
| 9  | ProtFun <sup>[11]</sup>    | <a href="http://www.cbs.dtu.dk/services/ProtFun/">http://www.cbs.dtu.dk/services/ProtFun/</a>  | 蛋白质功能预测               |
| 10 | PSIPRED <sup>[12]</sup>    | <a href="http://bioinf.cs.ucl.ac.uk/psipred/">http://bioinf.cs.ucl.ac.uk/psipred/</a>  | 蛋白质二级结构预测             |
| 11 | Phyre2 <sup>[13]</sup>     | <a href="http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index">http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index</a>  | 蛋白质三级结构预测             |
| 12 | Pymol                      | <a href="http://www.pymol.org/">http://www.pymol.org/</a>  | 蛋白质三维分子结构图            |
| 13 | MEGA <sup>[14]</sup>       | <a href="http://www.megasoftware.net/">http://www.megasoftware.net/</a>  | 氨基酸序列多重比对             |

## 2 结果与分析

### 2.1 ELOVL7 基因序列

ELOVL7 基因序列全长 2 324 bp, 有一个完整的开放阅读框, 为 846 bp, 编码 281 个氨基酸(图 1)。

```

atggccttcagtgatccttacagctaggactgtgctctttatgat
M A F S D L T A R T V R L Y D
aatggatcaaaagatgctgatccaagagttgaagactggctgctc
N W I K D A D P R V E D W L L
atgtcctcgccctctgccacaaacatcatcctgggactctatgtc
M S S P L P Q T I I L G L Y V
tattttgtcacttctctgggaccaaagttcatggagaatcgaag
Y F V T S L G P K F M E N R K
ccctttgaactcaagaaacgatgataacatacaattttttcata
P F E L K K A M I T Y N F F I
gtactcttttctatataatgttatgagttgtgatgtctggc
V L F S M Y I C Y E F V M S G
tggggtcacagttattcgtttcgatgtgaaattgtgactattca
W G T G Y S F R C E I V D Y S
cagtcacactactgcactgaggatggtagcacctgctggctttat
Q S P T A L R M V R T C W L Y
tacttctccaaattttattgagctattagatactatctttttgtt
Y F S K F I E L L D T I F F V
ctgctgaagaaaaatagtcgaagtacttctgcatgtcttccat
L R K K N S Q V T F L H V F H
cataccatcatgccatggacctgggtgttggagtcgaaatttct
H T I M P W T W F G V K F A
gcaggtgtttggggacgttccacgctttataaaatacacagtgta
A G G L G T F H A F I N T A V
catgtagtcatgtattgctactatggactgtgtgcatgggacca
H V V M Y C Y Y G L C A L G P
gcctaccagaaatacttgggtggaaaaaatctgacgacatta
A Y Q K Y L W W K K Y L T T L
cagcttgtccagtttatttattgtccacatccacataggtcagttc
Q L V Q F V I V T I H I G Q F
tttttcatggaggattgcaagtaccagatccagtcctttctgtac
F F M E D C K Y Q Y P V F L Y
atcattatgagttatgggtgcatctttcttctgctcttctctcat
I I M S Y G C I F L L L F L H
ttttgtactgtgcttacaccaaaggtcagaggctacccaaagct
F W Y C A Y T K G Q R L P K A
gtgaaaaatggaatcgcaaaacaaagaccactga 846
V K N G I G K N K D H *
    
```

图 1 ELOVL7 编码区及推导的氨基酸序列

Fig.1 The coding region sequence and deduced amino acid sequence

### 2.2 ELOVL7 氨基酸一级结构及理化性质预测

通过 ExPASy ProtParam 在线预测 ELOVL7 基因编码的氨基酸序列的组成成分和理化性质。结果显示这 281 个氨基酸原子组成为 C1595H2336N3680382S19, 共 4 700 个原子, 相对分子质量为 33 387.4 Da。其

带正电荷氨基酸数目(27)大于其带负电荷氨基酸数目(16), 等电点 (PI) 为 9.17, 偏碱性。ELOVL7 在 280 nm 的波长下的消光系数为  $88\ 280\ \text{M}^{-1}\cdot\text{cm}^{-1}$ , 其半衰期为 10 小时, 不稳定系数分值为 32.28, 小于 40, 推测此蛋白质结构比较稳定。

### 2.3 ELOVL7 信号肽预测

根据 Signal P4.1 Server 软件预测 ELOVL7 信号肽的结果如表 2 所示。其中 C 代表剪切位点的记分, 越高表示出现剪切位点的可能性就越大。S 表示信号肽记分, 高分位置表示相应的氨基酸位于信号肽部分概率较大, 低分位置表示位于成熟蛋白部分。Y 是基于 C 和 S 的综合记分, 通常被视为最理想的剪切位点, 具有最高的 C 值, 同时又是 S 值由高转低的位点。D 记分值是 S 和 Y 的加权平均值。ELOVL7 的 C、S 和 Y 的最高得分分别为 0.113、0.110 和 0.132, 均远低于 0.5, 因此可以推测出 ELOVL7 基因所编码的蛋白不存在信号肽。根据信号肽假说, 可以推测出此蛋白不是分泌蛋白。

表 2 ELOVL7 信号肽预测

Table 2 Signal peptide prediction of ELOVL7

| 指标     | 位点   | 分值    |
|--------|------|-------|
| max.C  | 51   | 0.113 |
| max.Y  | 24   | 0.11  |
| max.S  | 36   | 0.132 |
| mean S | 1~23 | 0.099 |
| D      | 1~23 | 0.105 |

### 2.4 ELOVL7 保守结构域预测

通过 NCBI 保守结构域数据库 (Conserved DomainDatabase, CDD), 分析 ELOVL7 中的保守结构域, 结果显示 ELOVL7 只含有 1 个属于 GNS1/SUR4 家族的保守结构域(见图 2)。

### 2.5 ELOVL7 亚细胞定位预测

运用 PSORT 工具预测 ELOVL7 的亚细胞定位, ELOVL7 可能位于细胞膜、高尔基体、线粒体内膜、内质网(膜)上(见表 2)。其在细胞膜的概率最高为 0.600, 而定位在高尔基体(0.400)和线粒体内膜的概率(0.396)相差不大, 在内质网(膜)上的概率最小(0.300)。

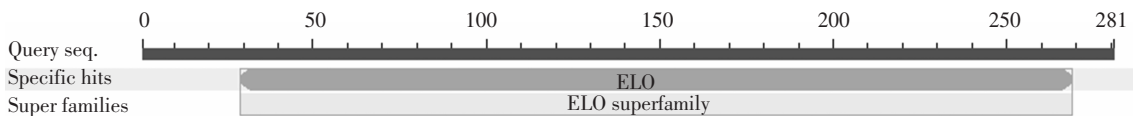


图 2 ELOVL7 保守结构域预测

Fig.2 Prediction of ELOVL7 conserved domains

表 3 *ELOVL7* 的亚细胞定位  
Table 3 Subcellular location of *ELOVL7*

| 亚细胞定位                                   | 概率    |
|---|-------|
| 细胞膜 (plasma membrane)                   | 0.600 |
| 高尔基体 (Golgi body)                       | 0.400 |
| 线粒体内膜 (mitochondrial inner membrane)    | 0.396 |
| 内质网(膜) endoplasmic reticulum (membrane) | 0.300 |

2.6 *ELOVL7* 跨膜结构域预测

跨膜结构域是横跨膜内和膜外的一段蛋白序列,是膜内在蛋白与膜脂相结合的主要部位,它固着于细胞膜上起锚定作用,一般由 20 个左右的疏水氨基酸组成。通过 ExPASy 软件中的 TMHMM Server v 2.0 工具预测 *ELOVL7* 基因的跨膜螺旋区。结果显示其 7 个跨膜螺旋区预测位置分别在 29~51,72~94,114~136,143~162,172~194,206~228,238~260,因此推测 *ELOVL7* 基因所编码的蛋白为跨膜蛋白(见图 3)。

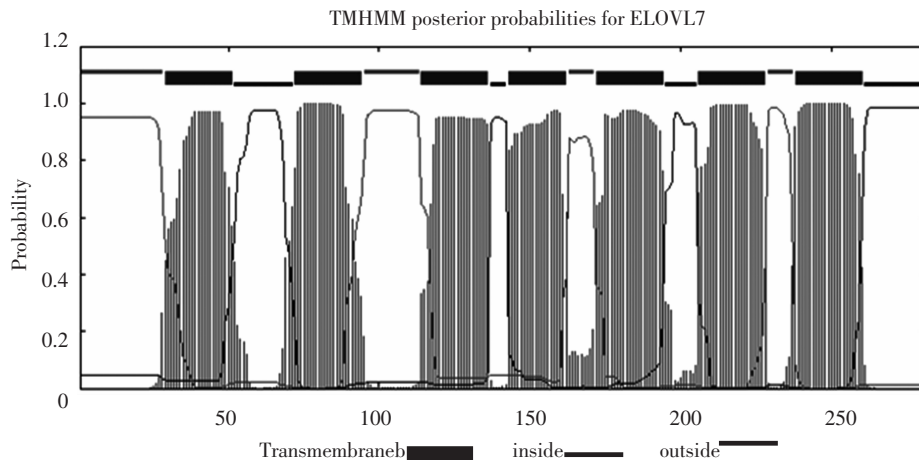


图 3 *ELOVL7* 跨膜结构域预测

Fig.3 The prediction of *ELOVL7*

2.7 *ELOVL7* 亲水性/疏水性分析

通过 ExPASy ProtScale 在线软件,对 *ELOVL7* 氨基酸序列的疏水性/亲水性预测结果如图 4,通过分析发现,在整条链中,最高的分值为 2.114,为排在 247 位的半胱氨酸 (Cys),代表其疏水性最强;最低的分值为 -1.510,为排在 271 位的缬氨酸 (Val),代表其亲水性最强。GRAVY (Grand average of

hydropathicity) 值为 0.301。

GRAVY 是测蛋白质的亲疏水性标准,其定义为序列中所有氨基酸亲水值的总和与氨基酸数量的比值,通常 GRAVY 值分布在 2 至 -2 之间,负值越大表示亲水性越好,正值越大表示疏水性越强。另外从整体看整条肽链的疏水性氨基酸残基多于亲水性氨基酸残基,因此可以推测出 *ELOVL7* 是一种非溶性蛋白。

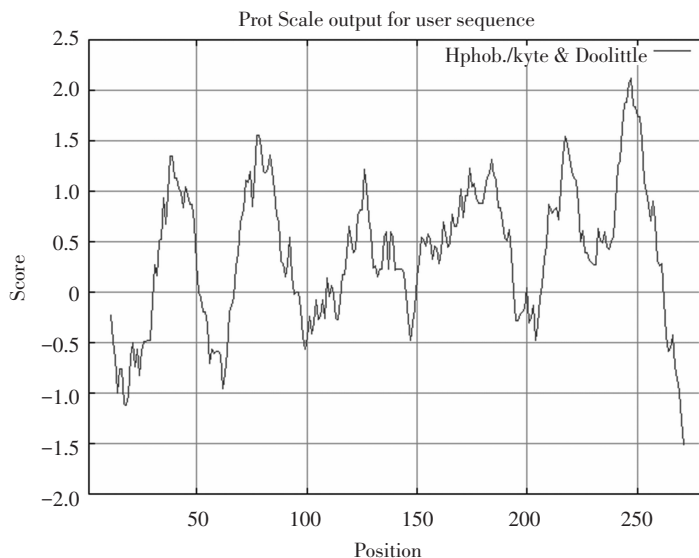


图 4 *ELOVL7* 亲水性/疏水性分析

Fig.4 The analysis of hydrophilic/hydrophobic

2.8 ELOVL7 激酶磷酸化修饰位点分析

蛋白质的磷酸化则是指在蛋白激酶催化作用下把 ATP 或 GTP 的磷酸基转移到底物蛋白质氨基酸残基的过程。蛋白质链的磷酸化一般只发生在丝氨酸 (serine, S), 苏氨酸 (threonine, T) 或酪氨酸 (tyrosine, Y) 这三个残基上。通过 KinasePhos 网站预测, 结果显示 ELOVL7 有 15 个丝氨酸激酶、15 个苏氨酸激酶和 20 个酪氨酸激酶潜在磷酸化位点。

2.9 ELOVL7 蛋白的功能预测

ELOVL7 蛋白功能预测结果如表 4 所示, 该蛋白的最可能的功能是运输和结合, 但也有一定可能参与翻译、嘌呤与嘧啶的合成、能量新陈代谢等过程。

表 4 ELOVL7 蛋白功能预测

Table 4 The function prediction of ELOVL7 protein

| Functional              | Prob  | Odds  |
|-------------------------|-------|-------|
| Energy metabolism       | 0.135 | 1.496 |
| Purines and pyrimidines | 0.339 | 1.397 |
| Translation             | 0.105 | 2.38  |
| Transport and binding   | 0.804 | 1.96  |

2.10 ELOVL7 二级结构预测

蛋白质的二级结构是指多肽链主链骨架盘绕折叠而形成的构象, 借氢键维系。主要有  $\alpha$ -螺旋、 $\beta$ -折叠、 $\beta$ -转角、无规卷曲。通过 PredictProtein 在线软件对 ELOVL7 二级结构进行分析。在 ELOVL7 的二级结构中,  $\alpha$  螺旋结构所占比例最大, 为 59.43%, 无规则卷曲结构占 37.01%, 延伸链只占 3.56% (见图 5)。由此可以推断出,  $\alpha$  螺旋结构是 ELOVL7 二级结构中最大的结构元件, 且分散于整个蛋白质中。

2.11 ELOVL7 同源性分析

通过 NCBI 数据库中的 Blastp 程序, 将苏太猪与犬、猫、牛、人类、马、小鼠、羊、鸡、鸭的 ELOVL7 氨基酸序列进行同源性分析, 数据显示其与氨基酸序列相似性分别为 93%、93%、92%、91%、91%、90%、90%、75%、73%。利用 Mega 5.05 软件对 ELOVL7 氨基酸序列和其它物种 ELOVL7 的氨基酸序列进行多重比对, 并构建系统进化树 (见图 6), 结果表明苏太猪 ELOVL7 基因与大部分物种亲缘关系非常近, 进一步验证了其同源性。

2.12 ELOVL7 三级结构预测

ELOVL7 蛋白质三维结构预测结果如图 7 所示: 空间结构以  $\alpha$  螺旋为主, 无规则卷曲和延伸链分散于整个蛋白质中。将该蛋白与人类、犬和小鼠中 ELOVL7 蛋白的三级空间结构图进行比较, 结果显示苏太猪与犬的三维分子结构非常相似, 但是与人和小鼠的差别较大。说明尽管氨基酸序列极其相似, 但是翻译加工成蛋白质的过程具有复杂的变化。

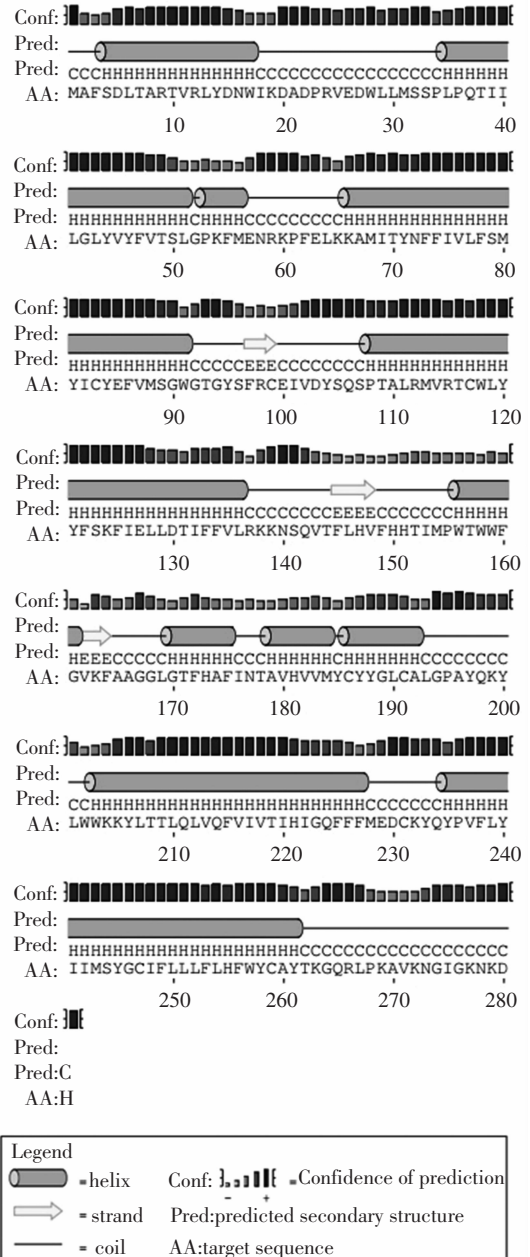


图 5 ELOVL7 二级结构预测

Fig.5 The secondary structure prediction of ELOVL7

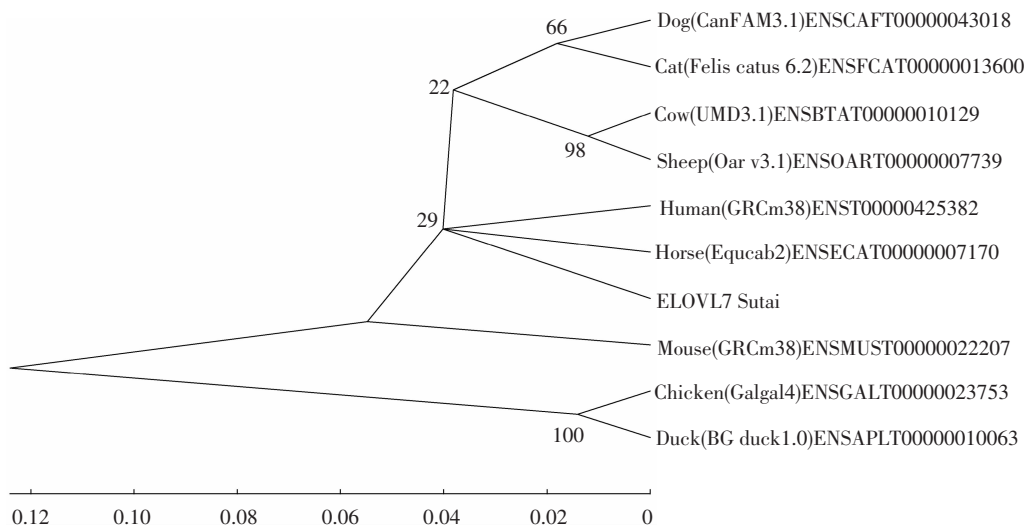


图6 不同物种的系统进化树

Fig.6 The phylogenetic tree of different species

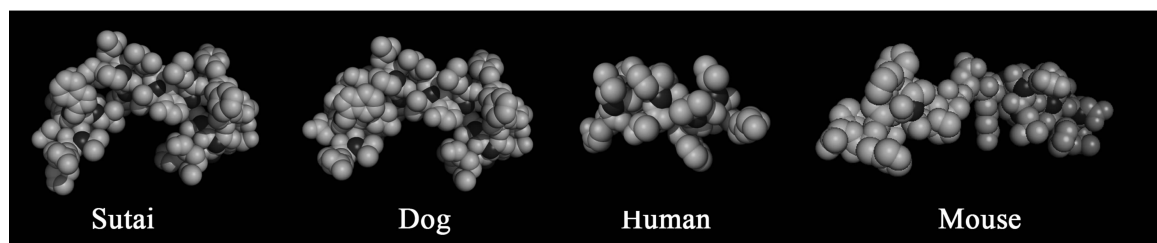


图7 三级分子结构图

Fig.7 3D molecular structures

### 3 结论

花生酸参与机体脂肪酸代谢循环,是人体内重要的饱和脂肪酸。就目前所知,ELVOL7 是调控花生酸的强候选基因。因此本研究通过生物信息学手段,系统地研究了苏太猪 *ELOVL7* 基因及其编码蛋白质的特征,为苏太猪脂肪酸调控网络提供理论依据。

*ELOVL7* 全长 2 324 bp,编码区为 846 bp,共编码 281 个氨基酸。其结构稳定,分子量为 33 387.4 Da,带正电荷,偏碱性。该基因所编码的蛋白质最可能位于细胞膜上,含有 1 个 GNS1/SUR4 家族的保守结构域,为跨膜、非分泌型疏水蛋白质,并有 15 个丝氨酸激酶、15 个苏氨酸激酶和 20 个酪氨酸激酶潜在磷酸化位点。 $\alpha$  螺旋是其二级结构中最主要的结构元件,无规则卷曲和延伸链分散于整个蛋白质中。

苏太猪 *ELOVL7* 同源性分析显示与犬、猫、牛、人类、马、小鼠和羊极为相似,相似性高达 90% 以上,从进化树也可以看出他们十分相近。但其三级空间结构图与人和小鼠的差别较大,说明 *ELOVL7*

基因从氨基酸序列翻译加工成蛋白质的过程具有复杂的变化。

### 参考文献(References)

- [1] SIMOPOULOS A P. The importance of the ratio of omega-6/omega-3 essential fatty acids [J]. *Biomedicine & Pharmacotherapy*, 2002, 56(8): 365-379.
- [2] WEBB E C, O'NEILL H A. The animal fat paradox and meat quality [J]. *Meat Science*, 2008, 80(1): 28-36.
- [3] YANG B, ZHANG W, ZHANG Z, et al. Genome-wide association analyses for fatty acid composition in porcine muscle and abdominal fat tissues [J]. *PLoS One*, 2013, 8(6): e65554.
- [4] NAGANUMA T, SATO Y, SASSA T, et al. Biochemical characterization of the very long-chain fatty acid elongase *ELOVL7* [J]. *FEBS Letters*, 2011, 585(20): 3337-3341.
- [5] TAMURA K, MAKINO A, HULLIN-MATSUDA F, et al. Novel lipogenic enzyme *ELOVL7* is involved in prostate cancer growth through saturated long-chain fatty acid metabolism [J]. *Cancer Research*, 2009, 69(20): 8133-8140.

- [6] ZHANG F, ZHANG Z, YAN X, et al. Genome-wide association studies for hematological traits in Chinese Sutan pigs [J]. *BMC Genetics*, 2014, 15(1): 41.
- [7] NAIR R, ROST B. Better prediction of sub-cellular localization by combining evolutionary and structural information [J]. *Proteins*, 2003, 53(4): 917-930.
- [8] PETERSEN T N, BRUNAK S, VON HEIJNE G, et al. SignalP 4.0: discriminating signal peptides from transmembrane regions [J]. *Nature Methods*, 2011, 8(10): 785-786.
- [9] KROGH A, LARSSON B, VON HEIJNE G, et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes [J]. *Journal of Molecular Biology*, 2001, 305(3): 567-580.
- [10] WONG Y H, LEE T Y, LIANG H K, et al. KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns [J]. *Nucleic Acids Research*, 2007, 35(suppl 2): W588-W594.
- [11] JENSEN L J, GUPTA R, STAERFELDT H H, et al. Prediction of human protein function according to Gene Ontology categories [J]. *Bioinformatics*, 2003, 19(5): 635-642.
- [12] BUCHAN D W, MINNECI F, NUGENT T C, et al. Scalable web services for the PSIPRED protein analysis workbench [J]. *Nucleic Acids Research*, 2013, 41(W1): W349-W357.
- [13] KELLEY L A, STERNBERG M J. Protein structure prediction on the Web: a case study using the Phyre server [J]. *Nature Protocols*, 2009, 4(3): 363-371.
- [14] TAMURA K, PETERSON D, PETERSON N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods [J]. *Molecular Biology and Evolution*, 2011, 28(10): 2731-279.