



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES

# 番茄*SICDPK2*的进化保守性与 抗逆功能关联性的系统解析

指导老师：罗静初

汇报人：李腾腾

小组成员：林婷、李腾腾、张娜、马百川



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES



目录

CONTENTS

系统发生树构建  
与进化分析

RNA-seq 分析

序列特征分析、多序列比对

蛋白质结构与功能分析

研究背景

明德格物 博学笃行



Part 1

# 研究背景

**科学挑战**：全球气候变化加剧，干旱、盐碱等非生物胁迫严重威胁番茄生产。

解析植物抗逆机制是培育抗逆品种的关键。

**核心靶点**：钙依赖性蛋白激酶（CDPK）是植物解码钙信号、启动抗逆响应的核心枢纽。

SICDPK2基因被报道参与胁迫响应，但其进化根源、结构功能与调控网络仍不清晰。

**系统解析SICDPK2：**

进化维度：多序列比对与系统发生树，揭示其进化保守性。

结构维度：蛋白质结构分析，预测功能关键位点。

调控维度：RNA-seq发现其在番茄中存在表达。



Part 2

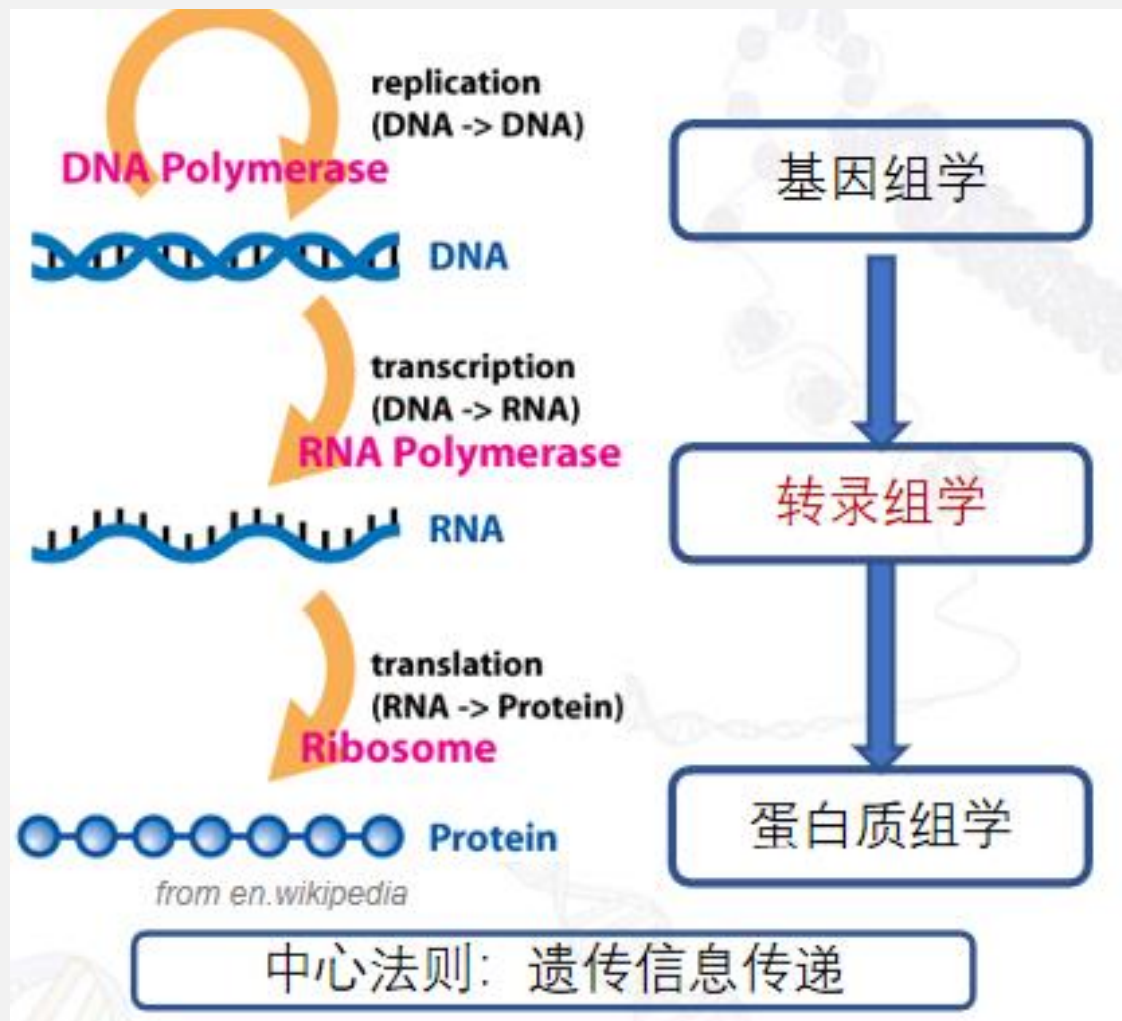
# RNA-seq 分析

# 转录组学概述



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES

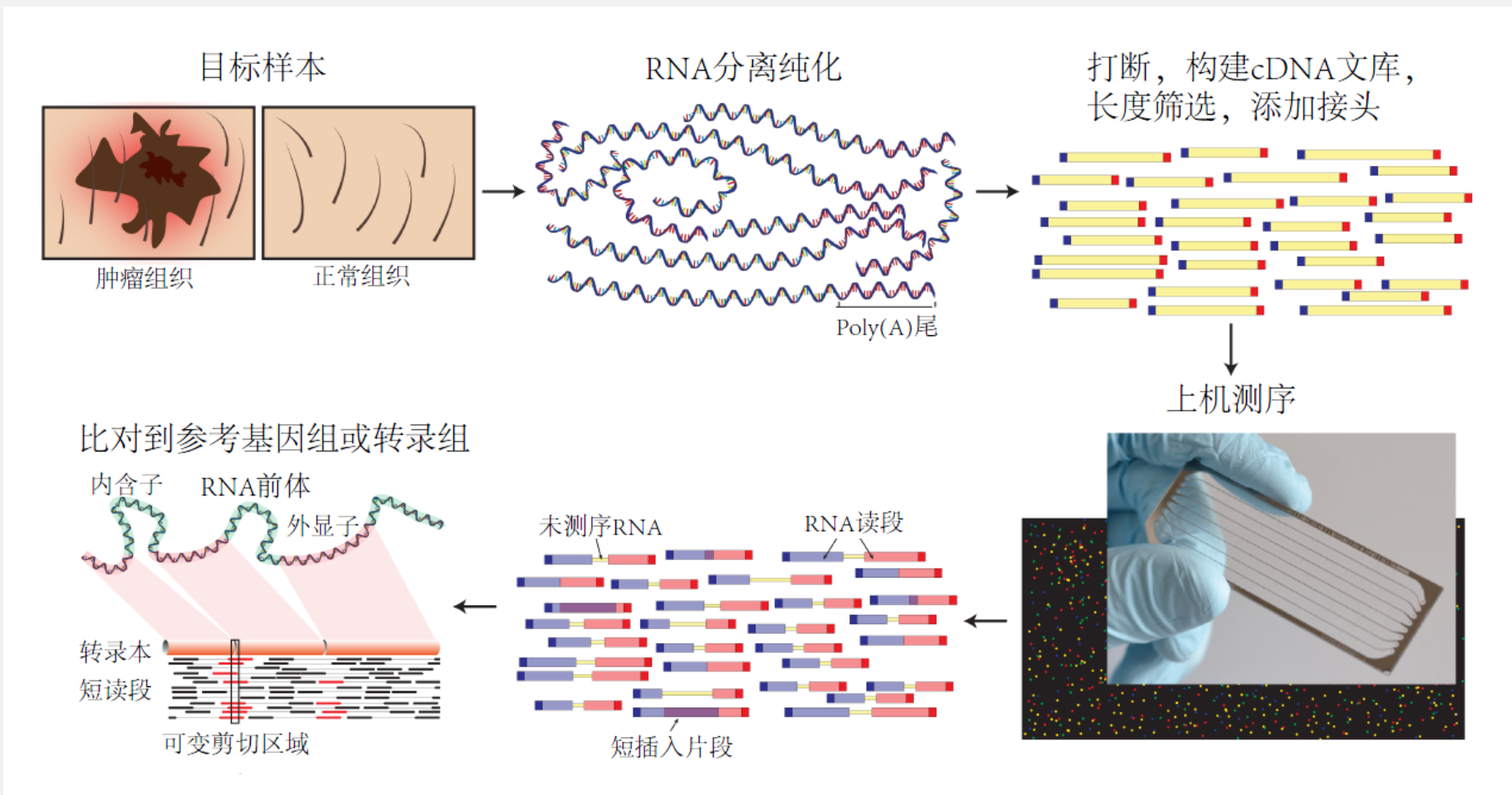


# 测序流程



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES

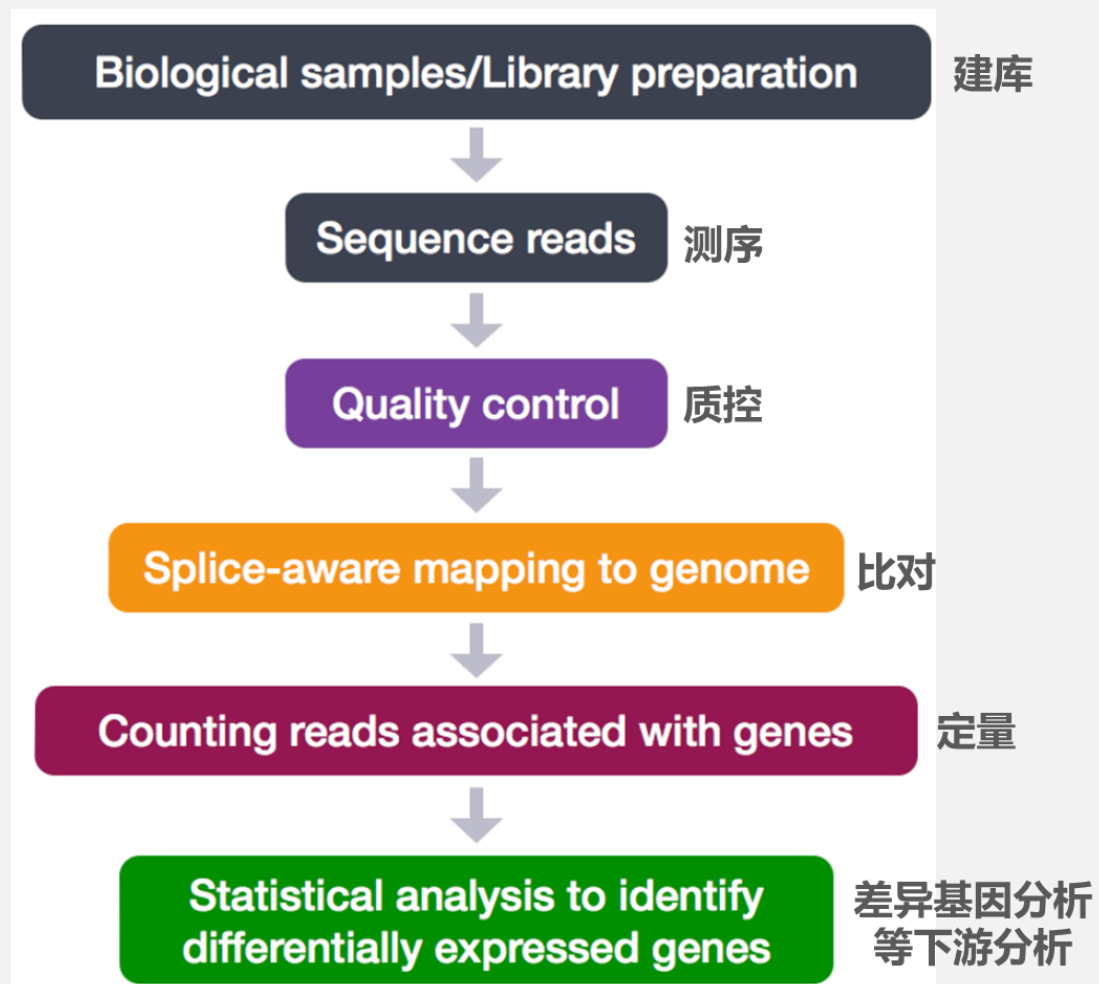


# 数据分析流程



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES





## 参考基因组文件 基因注释文件

```
GCF_036512215.1_SLM_r2.1_genomic.fna  
GCF_036512215.1_SLM_r2.1_genomic.gff  
GCF_036512215.1_SLM_r2.1_genomic.gtf
```

## 转录组数据文件

```
Control-CJ-1_1.fq.gz  
Control-CJ-1_2.fq.gz
```

## FastQC—测序质量评估:

### (1) 去除含有接头序列的Reads

(Reads中接头序列大于5bp)

### (2) 去除低质量的Reads

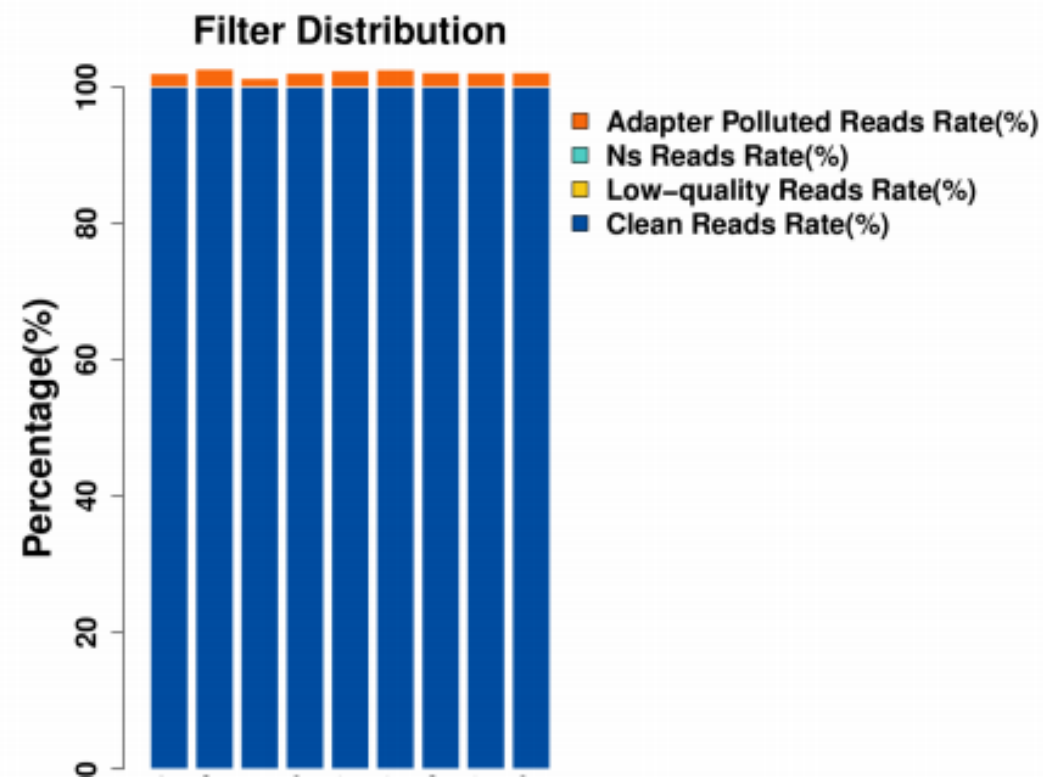
(质量值 $Q \leq 20$ 的碱基占总碱基的15%以上)

### (3) 去除含N比例大于5%的Reads

数据质量: 合格:  $\geq 80\%$

良好:  $\geq 85-90\%$

优秀:  $\geq 95\%$



- 橙色: Adapter Polluted Reads Rate (%) (接头污染Reads比例) —— 测序读段中因“接头 (Adapter) 污染”导致的无效读段比例。
- 青绿色: Ns Reads Rate (%) (含未知碱基N的Reads比例) —— 测序读段中包含未知碱基 (通常表示测序错误或低质量区域) 的比例。
- 黄色: Low - quality Reads Rate (%) (低质量Reads比例) —— 测序读段中质量值低于阈值的无效读段比例。
- 蓝色: Clean Reads Rate (%) (干净Reads比例) —— 经过滤后, 保留的高质量、无污染的有效读段比例。

# 序列比对

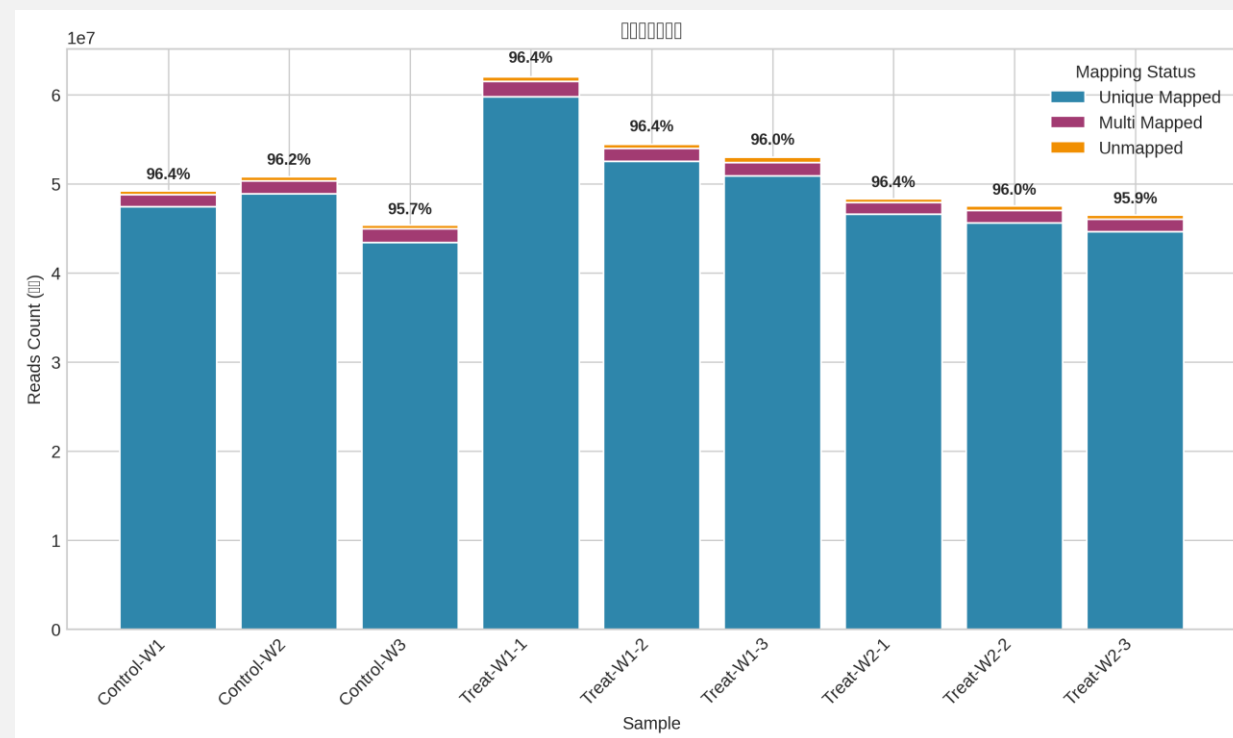
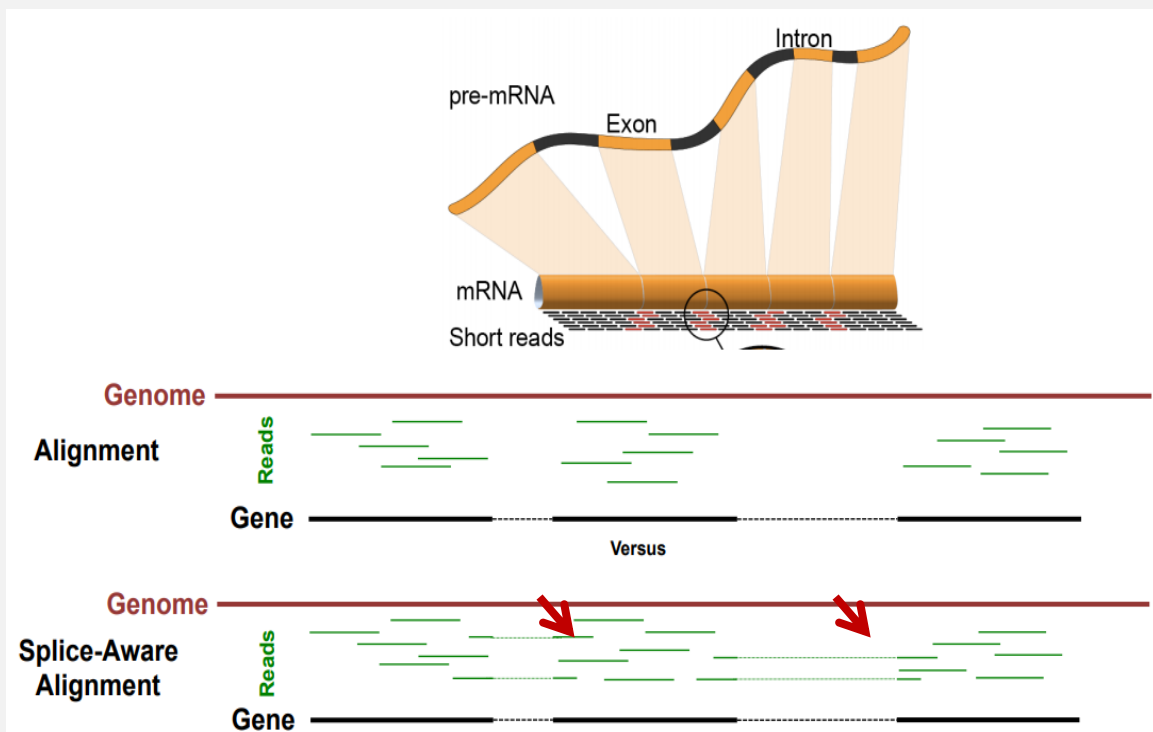


中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES

- 目的：找到测序reads在基因组上的来源
- 软件：Hisat2, Bowtie2, STAR

合格：≥ 70%  
良好：≥ 80-85%  
优秀：≥ 90%



# 基因表达定量



中国农业科学院

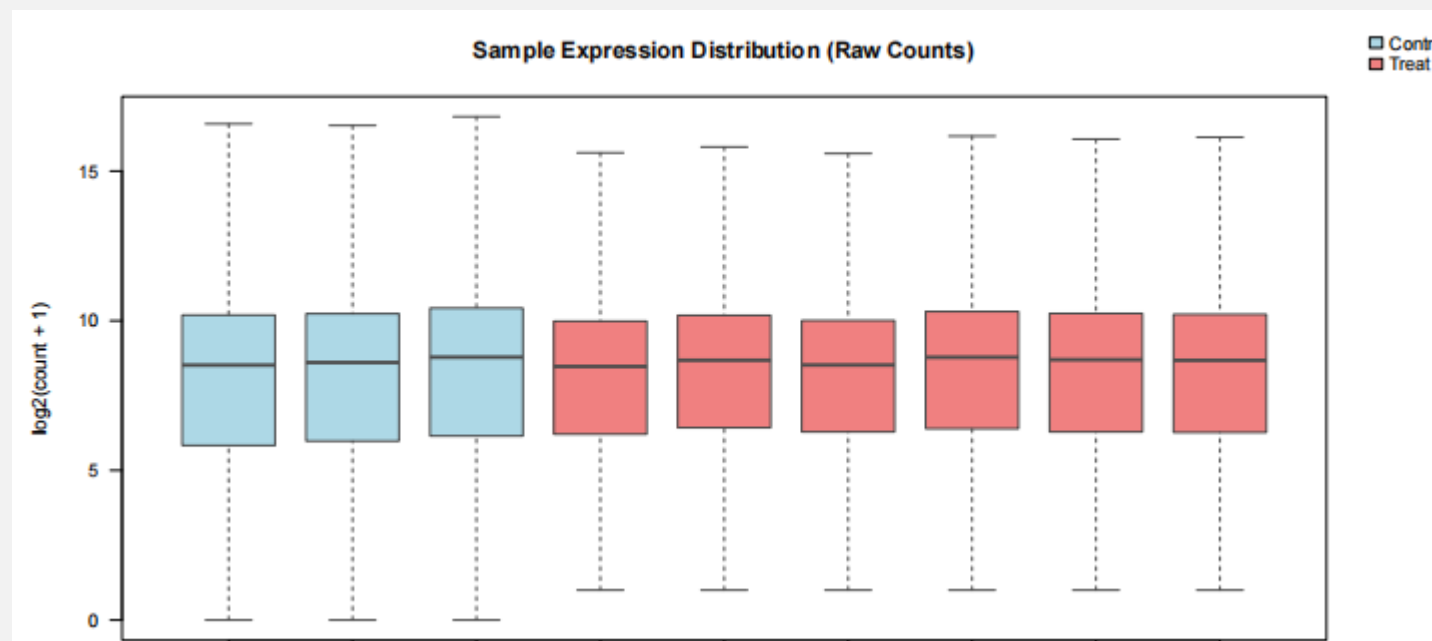
CHINESE ACADEMY OF AGRICULTURAL SCIENCES

**原始计数是基础：** Read Counts是下游所有分析的起点。

**标准化是关键：** 用FPKM比较样本内/间基因丰度，

用DESeq2/edgeR的标准化方法做差异分析。

(用FeatureCounts、stringtie等进行基因表达定量)



# 基因表达定量标准化



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES

挑战	FPKM如何解决	公式中的对应部分
1. 基因长度不同	长基因在测序时会被随机打断成更多片段，即使转录本数量相同，它产生的Reads也更多。这会造成“长基因表达量虚高”。	除以基因长度（每千碱基） 把总读数标准化到相同的长度单位上。
2. 样本测序深度不同	样本A测了5千万条Reads，样本B测了1亿条。B中所有基因的Read Counts天然就可能是A的两倍。	除以总比对Reads数（每百万条） 把总读数标准化到相同的测序量级上。
3. 要比较“表达丰度”	我们想知道一个基因在总转录本池子里占多大比例，而不是绝对个数。	“每百万条”这个概念，本质就是计算一个基因的Reads对丰度。

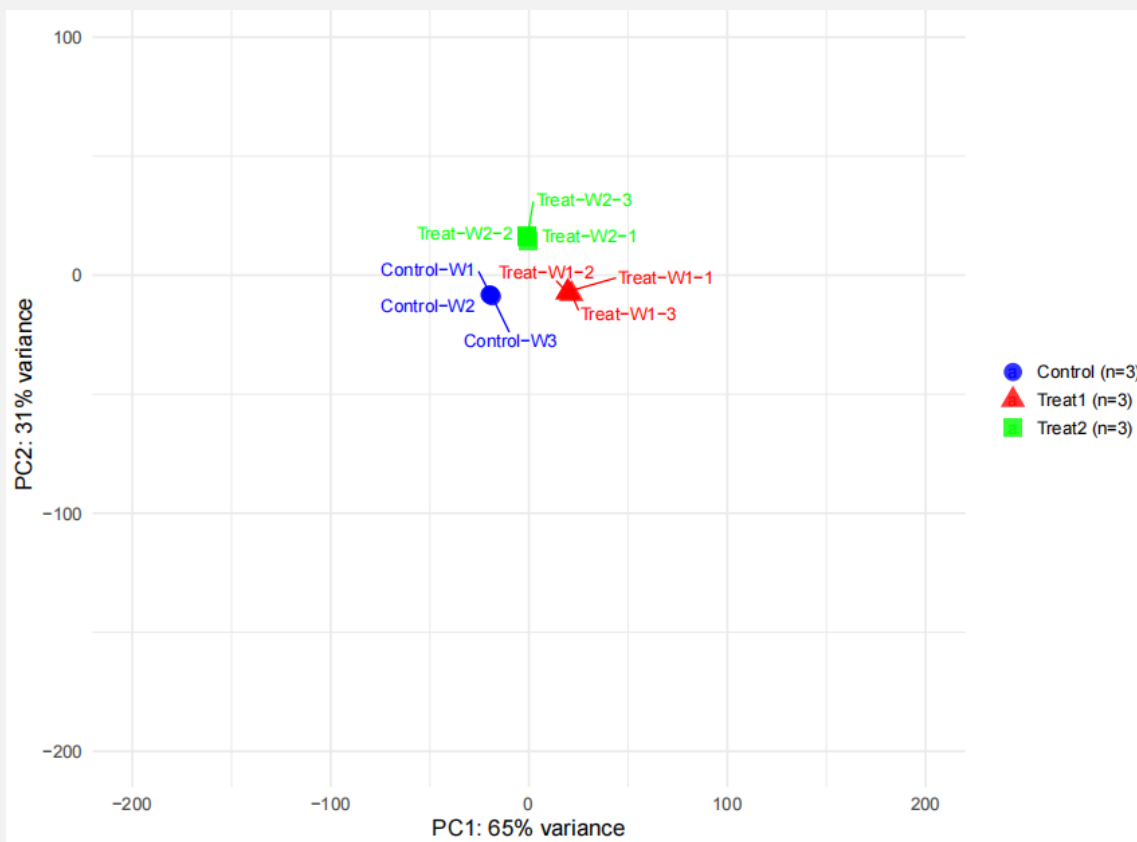
LOC101253982	0.601569	0.687597	0.701076	0.984753	1.03443	0.990941	0.83254	0.892721	0.873355
LOC101252661	2.68892	3.12646	3.08044	3.41439	3.72397	4.02982	3.72795	3.89651	3.4615
LOC101248129	0	0	0	0	0	0	0	0	0
CDPK2	23.0513	22.2915	22.8665	22.9831	22.8783	21.9034	21.3394	21.6173	22.1006
LOC101252662	9.87855	9.80484	9.78563	7.66134	8.90611	10.0994	10.5118	8.98095	9.93848
LOC101251341	8.33498	9.48986	8.45228	8.16345	8.33637	8.54229	7.02945	6.47631	6.53224
LOC101253983	21.0142	21.838	20.1207	19.1483	20.6016	18.8203	24.0812	24.9448	22.2606

# 评估数据质量

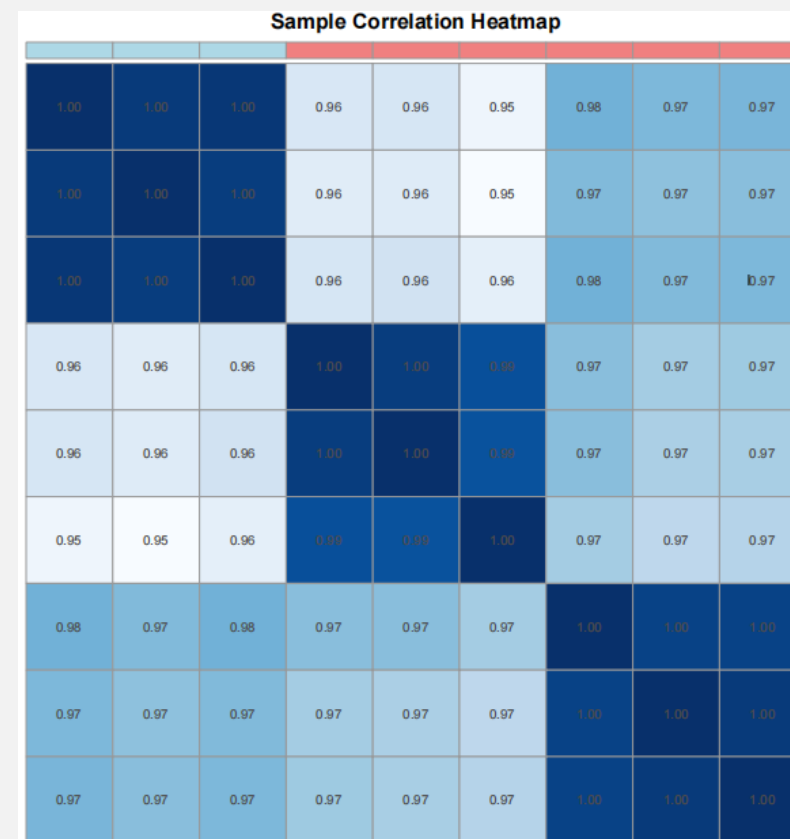


中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES



PCA图



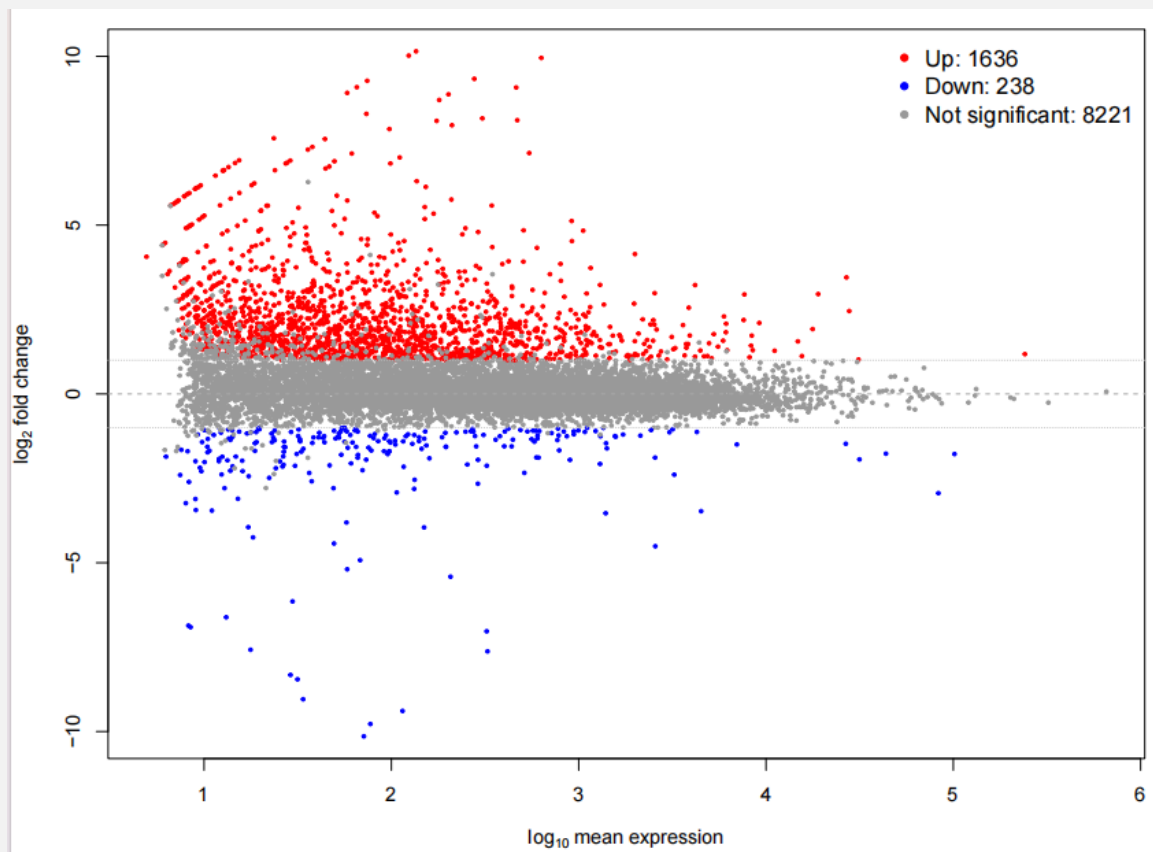
相关性热图

# 差异基因分析

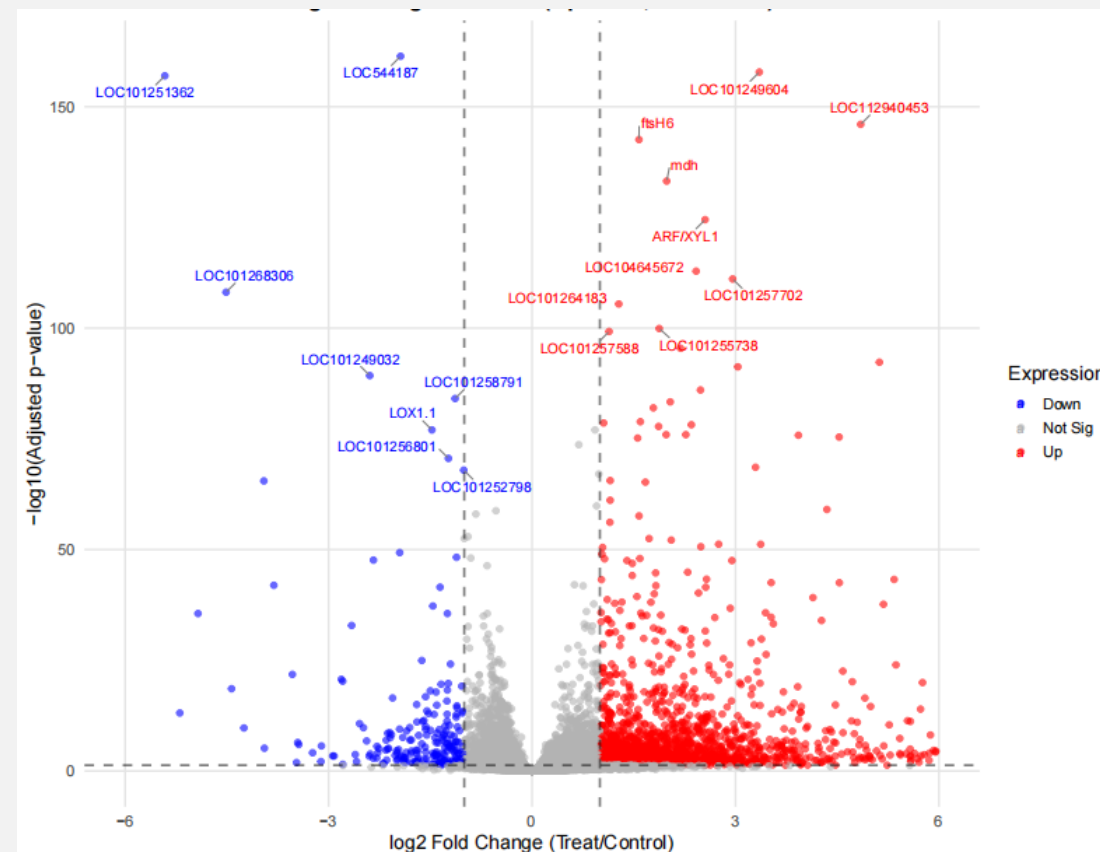


中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES



MA图



火山图



Part **3**

# 序列特征分析、多序列比对

# SICDPK2序列特征分析 (TBtools实现)



中国农业科学院

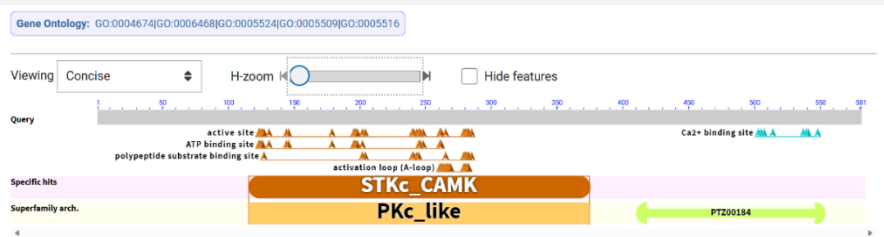
CHINESE ACADEMY OF AGRICULTURAL SCIENCES

## Output Stat Info

Total\_Len: 5957  
 Total\_Seq\_Num : 1  
 Total\_N\_Counts: 0  
 Total\_LowCase\_Counts: 0  
 Total\_GC\_content: 0.37  
 Minimum Len: 5957  
 Maximum Len: 5957  
 Mean Len: 5,957  
 Median Len: 5,957  
 N50: 5957

A	B	C	D	E	F	G
Sequence ID	Number of Amino Acid	Molecular Weight	Theoretical pI	Instability Index	Aliphatic Index	Grand Average of Hydropathicity
NP_001234784.1	581	64602.41	5.54	35.82	78.33	-0.407

序列特征	分析结果	生物学意义
核酸序列长度	5957 bp	符合植物CDPK基因典型长度范围
蛋白质序列长度	581 aa	与已知功能CDPK蛋白长度相近
分子量	64.6 kD	推测为可溶性蛋白，便于胞内信号传导
等电点 (pI)	5.54	提示其可能的胞内定位环境（酸性/碱性区域）
亲疏水性 (GRAVY)	-0.407	负值提示为亲水性蛋白，可能定位于细胞质或细胞膜表面
功能结构域预测	激酶催化模块 (PKc_like)、钙感应与调节模块	符合CDPK家族典型结构特征，具备Ca <sup>2+</sup> 结合与激酶催化功能

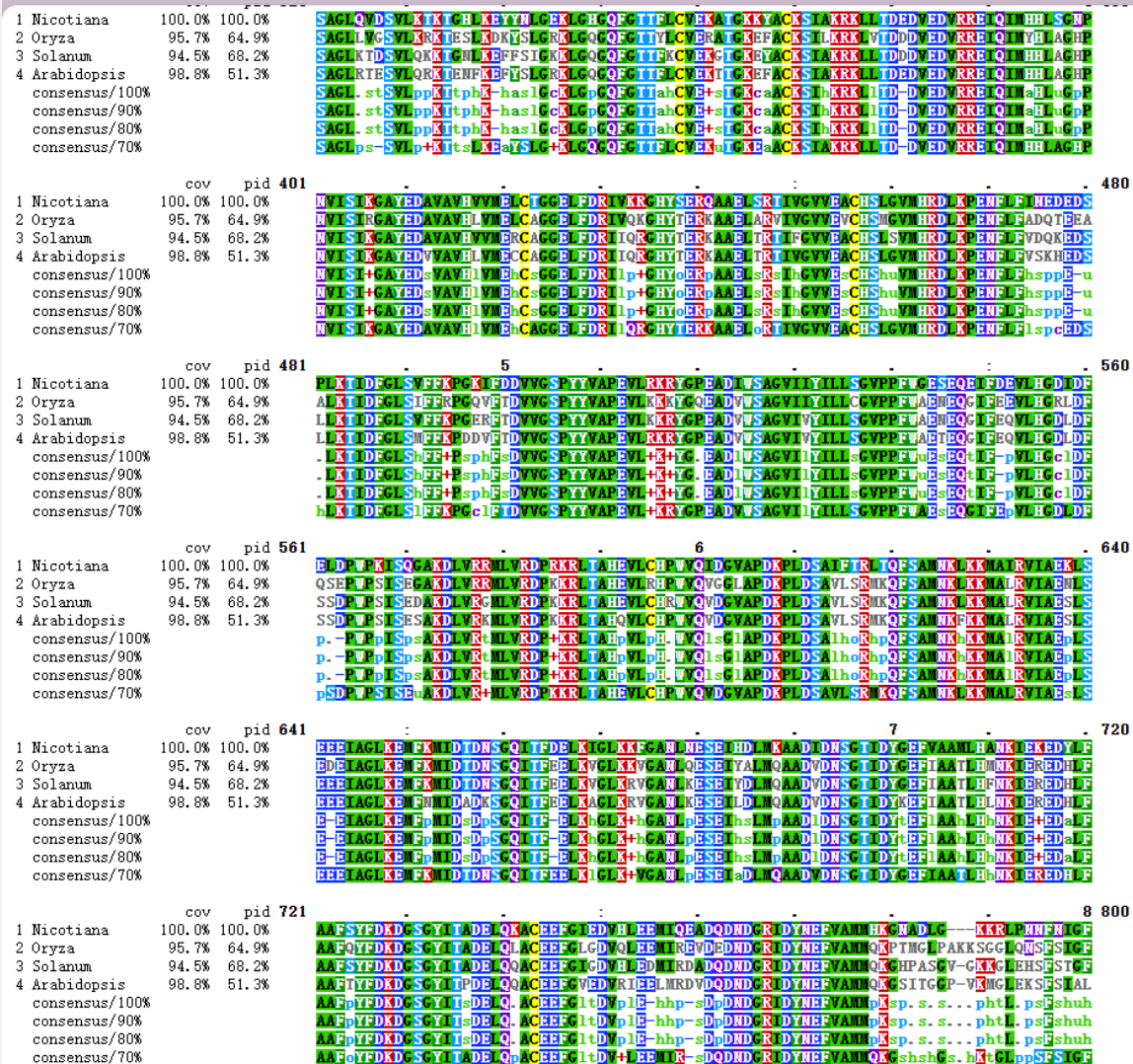


# 同源序列搜索与多序列比对



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES



利用 Clustal Omega 对番茄及近缘物种同源蛋白进行多序列比对，发现番茄CDPK2与烟草序列一致性达 100%、与拟南芥达 98.8%，且 401-480、561-640 区域高度保守；这表明该蛋白为番茄功能核心蛋白，可借鉴近缘物种功能信息与实验工具开展后续研究



Part **4**

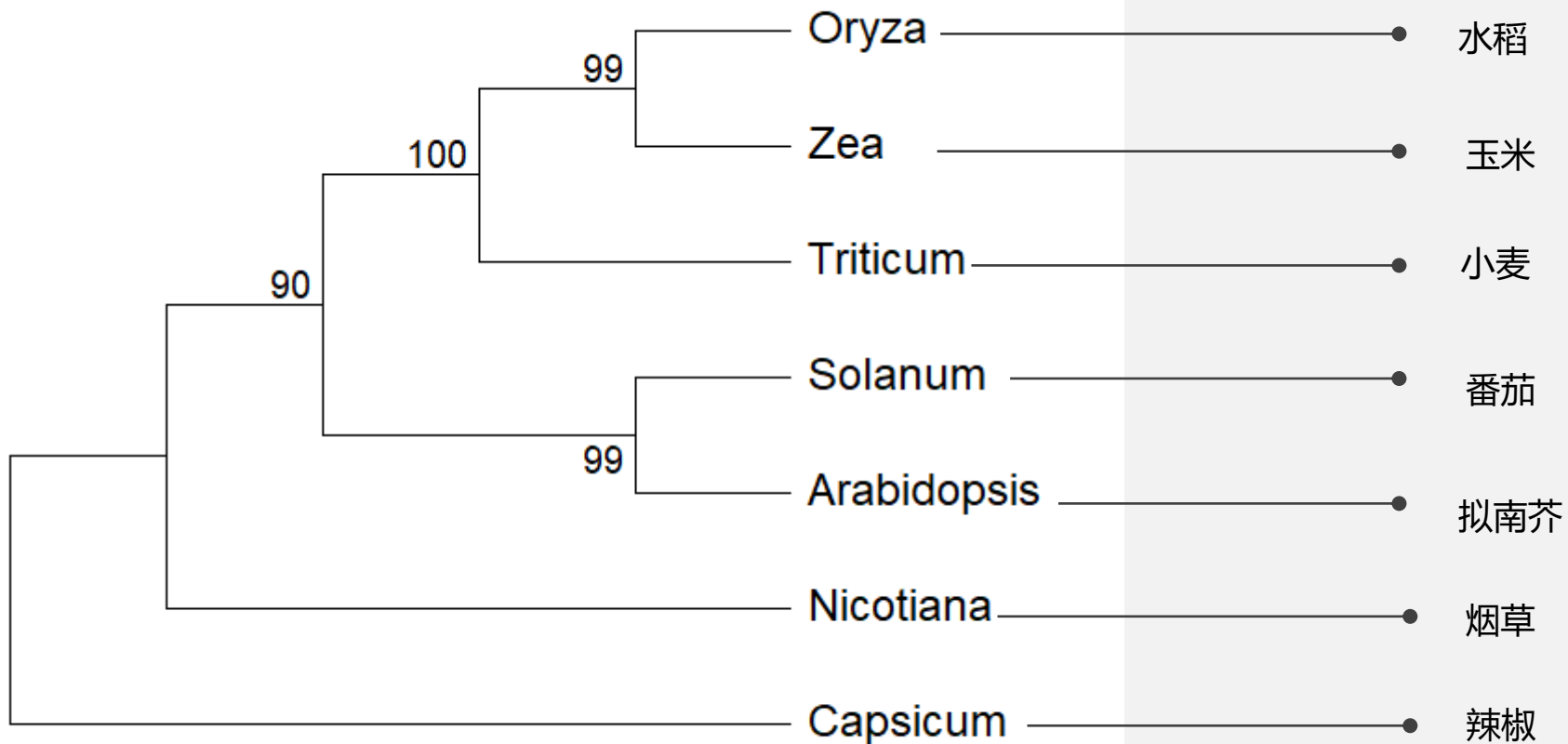
# 系统发生树构建与进化分析

# 系统发育树



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES



# 建树分析

## 1. 从进化树看 SICDPK2 的保守性

CDPK基因的进化保守性：番茄SICDPK2与拟南芥中CDPK家族基因可能具有共同祖先，序列同源性较高。  
参考价值：拟南芥CDPK的抗逆功能（如抗干旱、盐胁迫）研究较成熟，可通过进化关系推测SICDPK2可能继承了类似的抗逆功能结构域（如激酶域、钙调素类似域）

## 2. 不同物种 CDPK 的抗逆功能关联

番茄SICDPK2与禾本科CDPK虽亲缘远，但可能存在抗逆功能的趋同进化（比如都响应钙信号调控逆境胁迫）。  
烟草NtCDPK、辣椒CaCDPK参与抗病、抗盐反应；番茄SICDPK2作为茄属成员，更可能与这些同科物种的CDPK具有功能保守性

## 3. 进化保守性对 SICDPK2 抗逆功能的研究意义

从进化树的亲缘关系可指导实验：  
对比Solanum（番茄）与Oryza（水稻）CDPK的序列差异，定位抗逆特异性结构域（比如茄科特有的功能位点，决定SICDPK2在番茄中特有的抗逆响应）



Part **5**

# 蛋白质结构与功能分析

Number of amino acids: 581

Theoretical pI: 5.54

Molecular weight: 64602.41

Amino acid composition:

Ala (A)	40	6.9%
Arg (R)	28	4.8%
Asn (N)	15	2.6%
Asp (D)	44	7.6%
Cys (C)	7	1.2%
Gln (Q)	20	3.4%
Glu (E)	48	8.3%
Gly (G)	49	8.4%
His (H)	15	2.6%
Ile (I)	27	4.6%
Leu (L)	43	7.4%
Lys (K)	48	8.3%
Met (M)	18	3.1%
Phe (F)	28	4.8%
Pro (P)	23	4.0%
Ser (S)	39	6.7%
Thr (T)	23	4.0%
Trp (W)	5	0.9%
Tyr (Y)	12	2.1%
Val (V)	49	8.4%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%

(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 92

Total number of positively charged residues (Arg + Lys): 76

### Atomic composition:

Carbon	C	2855
Hydrogen	H	4507
Nitrogen	N	783
Oxygen	O	875
Sulfur	S	25

Formula: C<sub>2855</sub>H<sub>4507</sub>N<sub>783</sub>O<sub>875</sub>S<sub>25</sub>

Total number of atoms: 9045

### Extinction coefficients:

Extinction coefficients are in units of M<sup>-1</sup> cm<sup>-1</sup>, at 280 nm measured in water.

Ext. coefficient 45755

Abs 0.1% (=1 g/l) 0.708, assuming all pairs of Cys residues form cystines

Ext. coefficient 45380

Abs 0.1% (=1 g/l) 0.702, assuming all Cys residues are reduced

### Estimated half-life:

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

>20 hours (yeast, in vivo).

>10 hours (Escherichia coli, in vivo).

### Instability index:

The instability index (II) is computed to be 35.82

This classifies the protein as stable.

Aliphatic index: 78.33

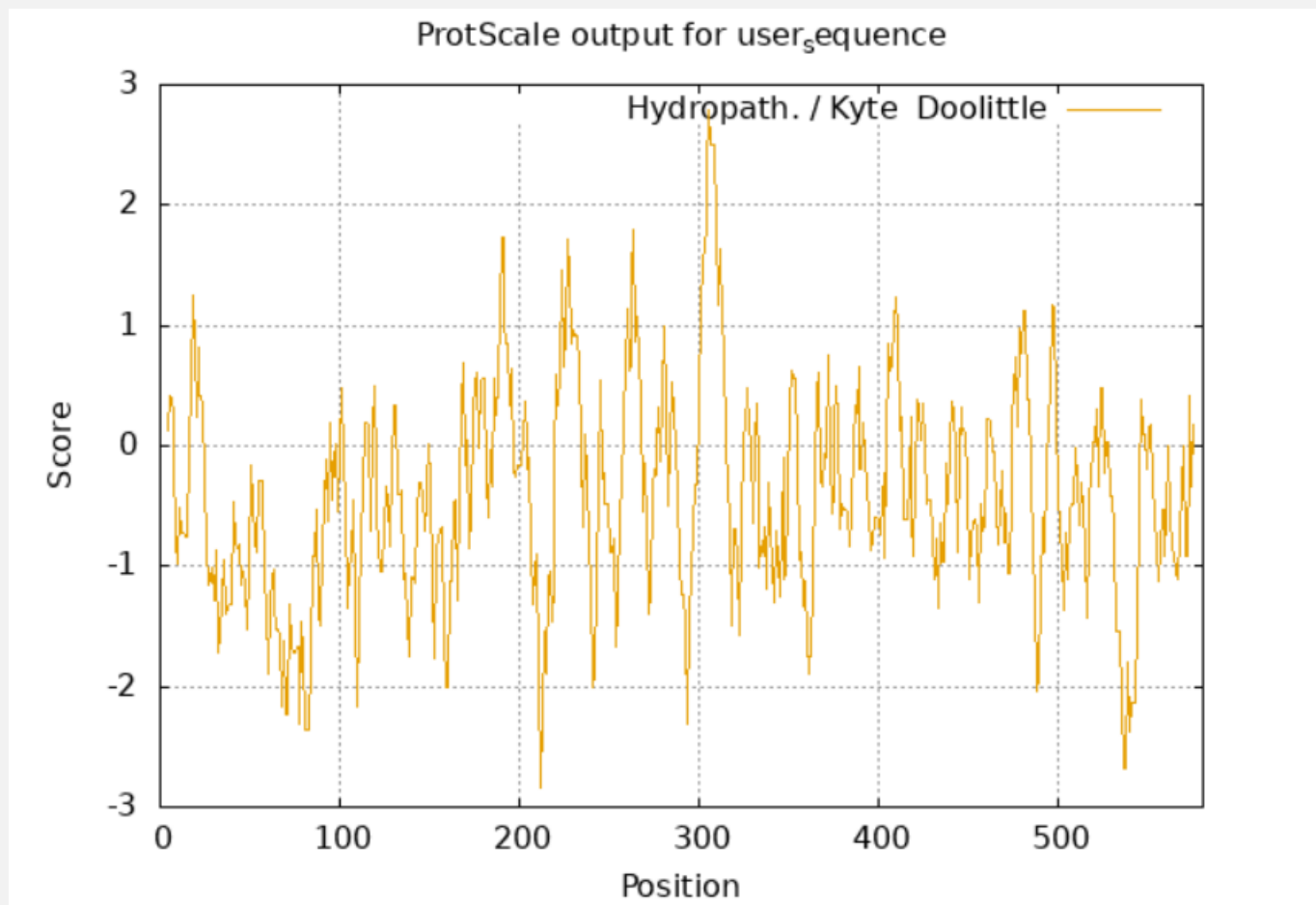
Grand average of hydropathicity (GRAVY): -0.407

# 亲疏水性分析



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES



**CDPK2 是亲水性蛋白**

# 亚细胞定位预测



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES

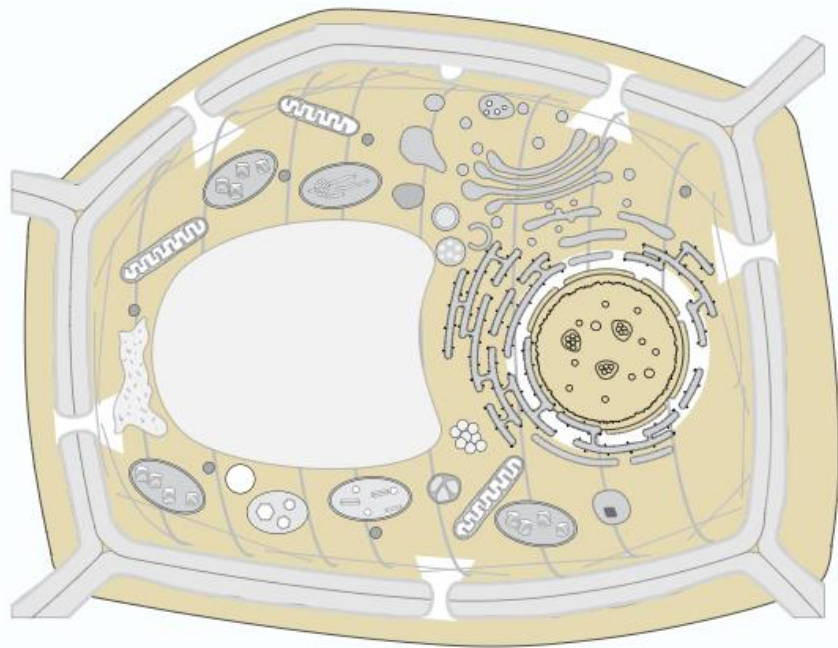
## Subcellular Location<sup>i</sup>

UniProt Annotation

GO Annotation

- cytoplasm [↗](#) Source:GO\_Central
- membrane [↗](#) Source:UniProtKB-KW
- nucleus [↗](#) Source:GO\_Central

[Complete GO annotation on QuickGO ↗](#)

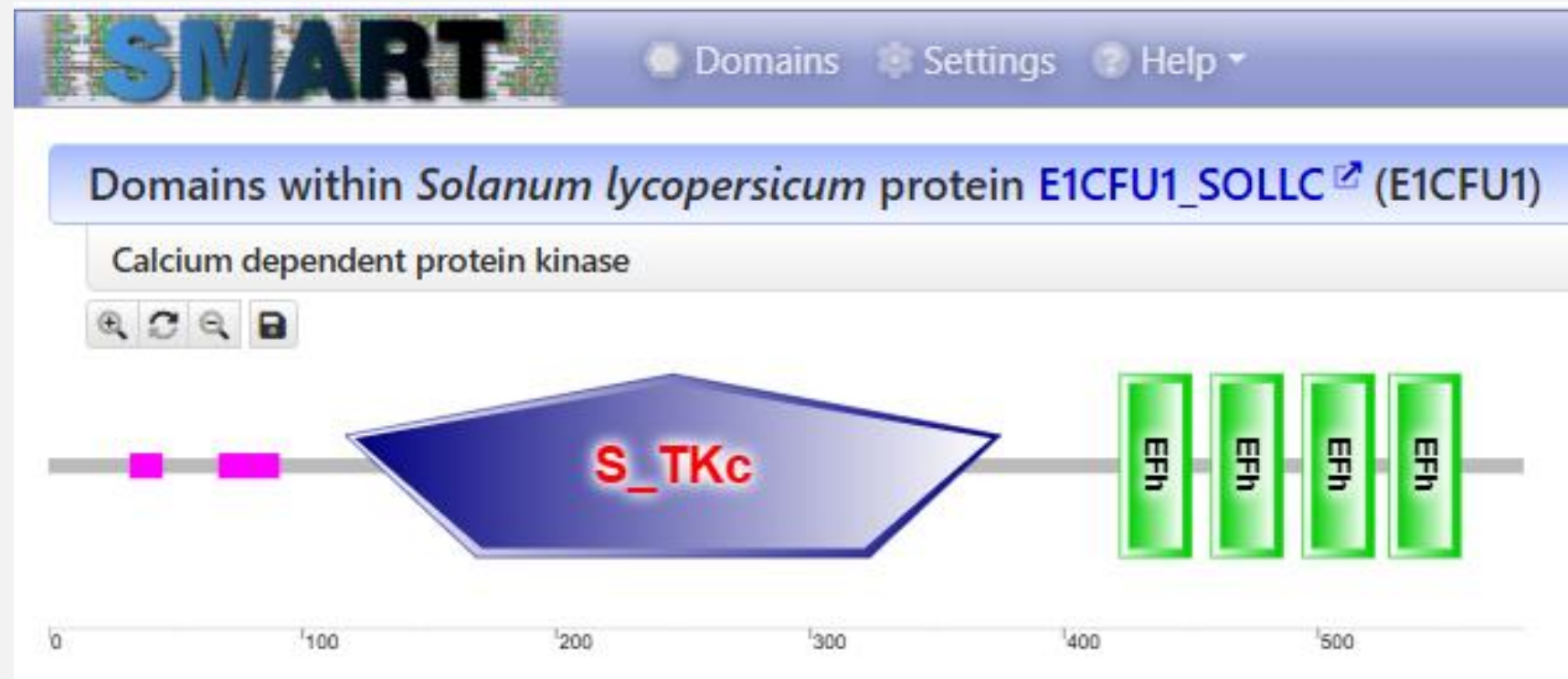


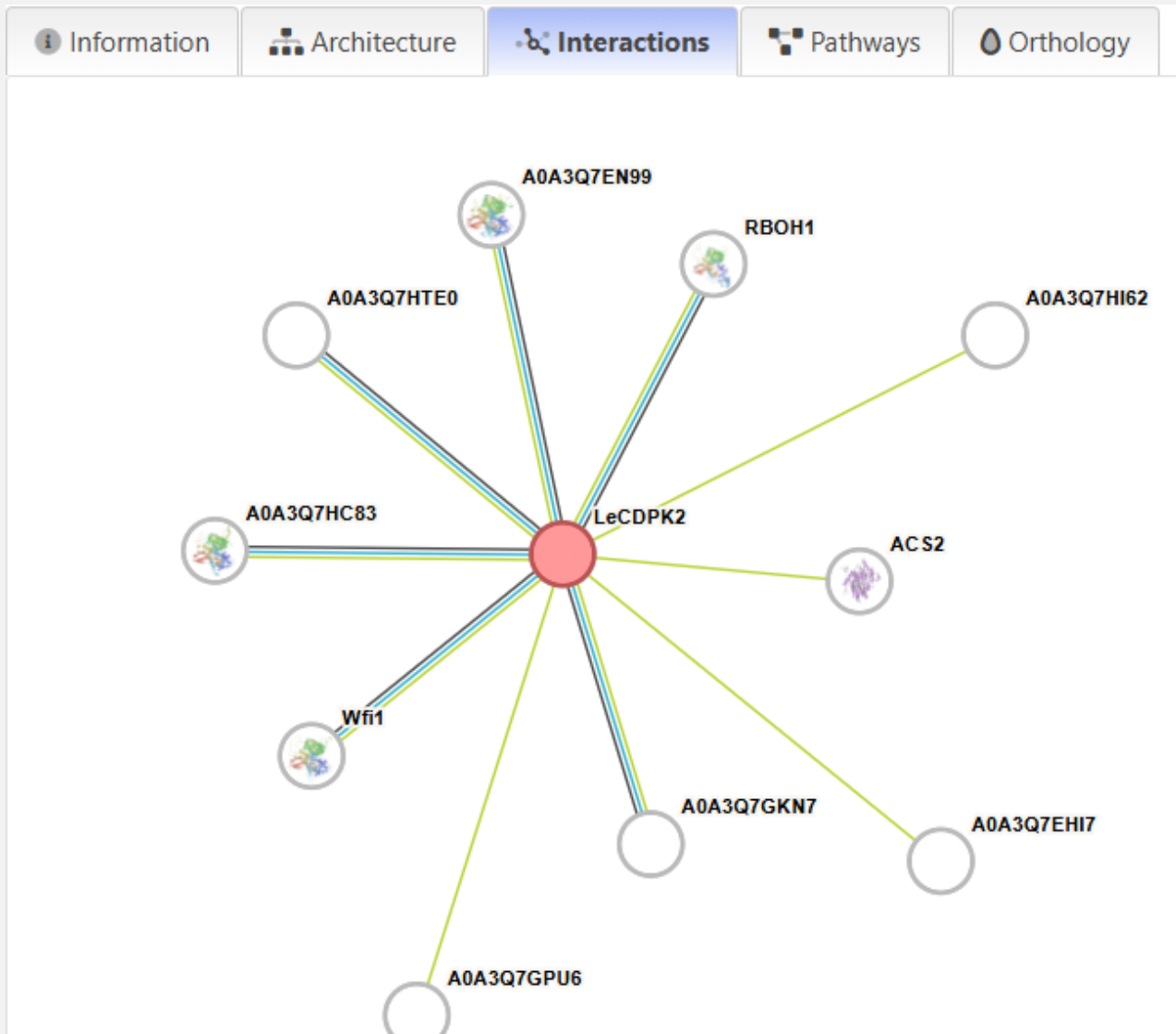
亚细胞定位预测:  
细胞质、细胞核

与亲疏水性预测  
中CDPK2不含跨  
膜结构域相对应

**S\_TKc 结构域：CDPK  
的催化核心，负责 ATP  
结合与底物蛋白的磷酸  
化**

**EFh 结构域：钙离子结  
合的关键基序**





**RBOH1: CDPK2 可能参与胁迫  
响应中的 ROS 信号通路**

**AC S2: CDPK2 可能参与激素  
信号 (乙烯) 的调控**



## KEGG orthologous groups

KO	Name	Description
<a href="#">K13412</a>	CPK	calcium-dependent protein kinase [EC:2.7.11.1]

## KEGG pathways

Pathway	Description
<a href="#">map05145</a>	Toxoplasmosis
<a href="#">map04626</a>	Plant-pathogen interaction

**CDPK2 被归类到K13412 (钙依赖蛋白激酶) 家族, CDPK2 的“钙依赖激酶”功能在进化中高度保守。**

# 可爱的G02家族成员



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES



**G02A 林婷** 序列特征分析、多序列比对

**G02B 李腾腾** rnaseq分析

**G02C 张娜** 蛋白质结构与功能分析

**G02D 马百川** 系统发生树构建与进化分析



中国农业科学院

CHINESE ACADEMY OF AGRICULTURAL SCIENCES

谢谢大家！

