

# 卓越工程师工作室

## 技术分享第四期

### DenovoAS\_Finder和DeepBSA的开发与应用

李昭

导师：李林教授

2023.11.10

## 求学经历：华中农业大学本硕博

2014.9-2018.6：14级农科张之洞班

2018.9-2020.6：作物遗传育种，导师李林教授

2020.9-至今：作物信息学，导师李林教授

## 研究方向

玉米野生近缘种大刍草的基因组组装及分析；

QTL定位方法的开发；

玉蜀黍属多年生性状变异的遗传机制初探。



## 座右铭

道路是曲折的，前途是光明的。  
--毛泽东

## DenovoAS\_Finder

**功能：**实现无参下鉴定基因选择性剪接。

**背景：**2020年开发，无大刍草参考基因组，需要进行其选择性剪接的鉴定。

**结果：**Li *et al.*, 2021, *Molecular Ecology Resources*

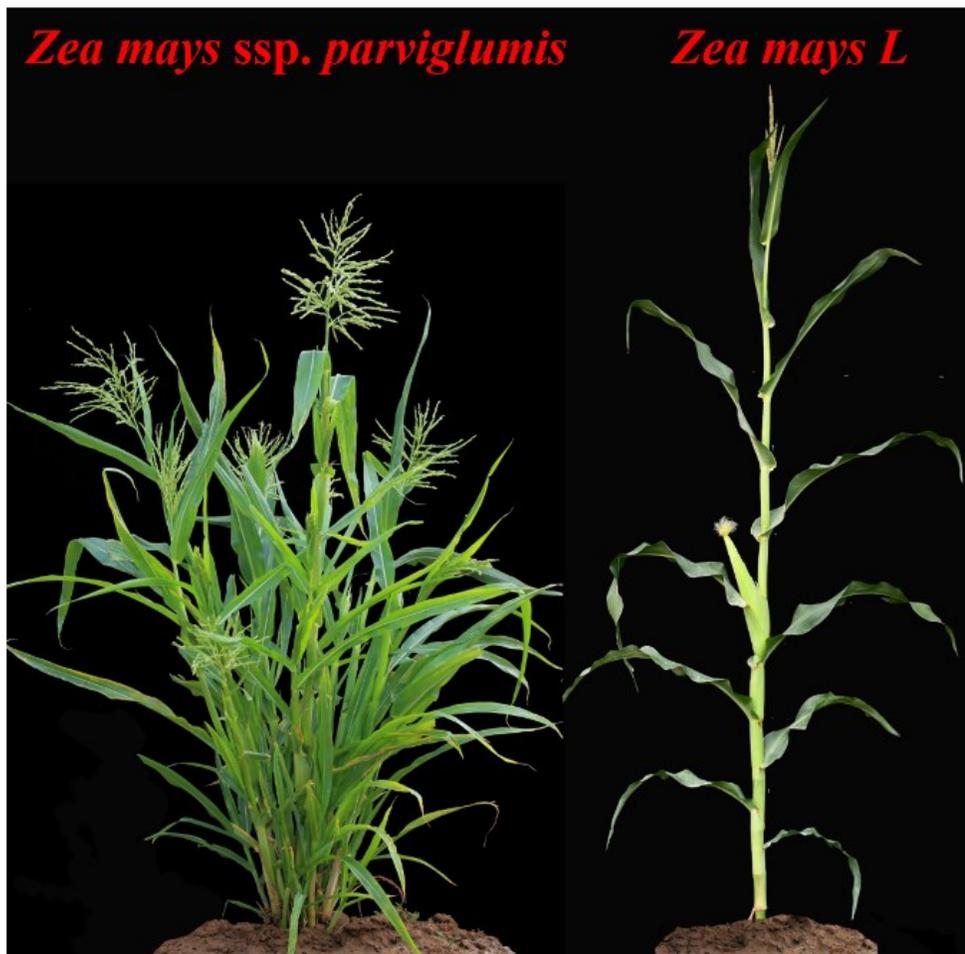
## DeepBSA

**功能：**实现混池测序分析。

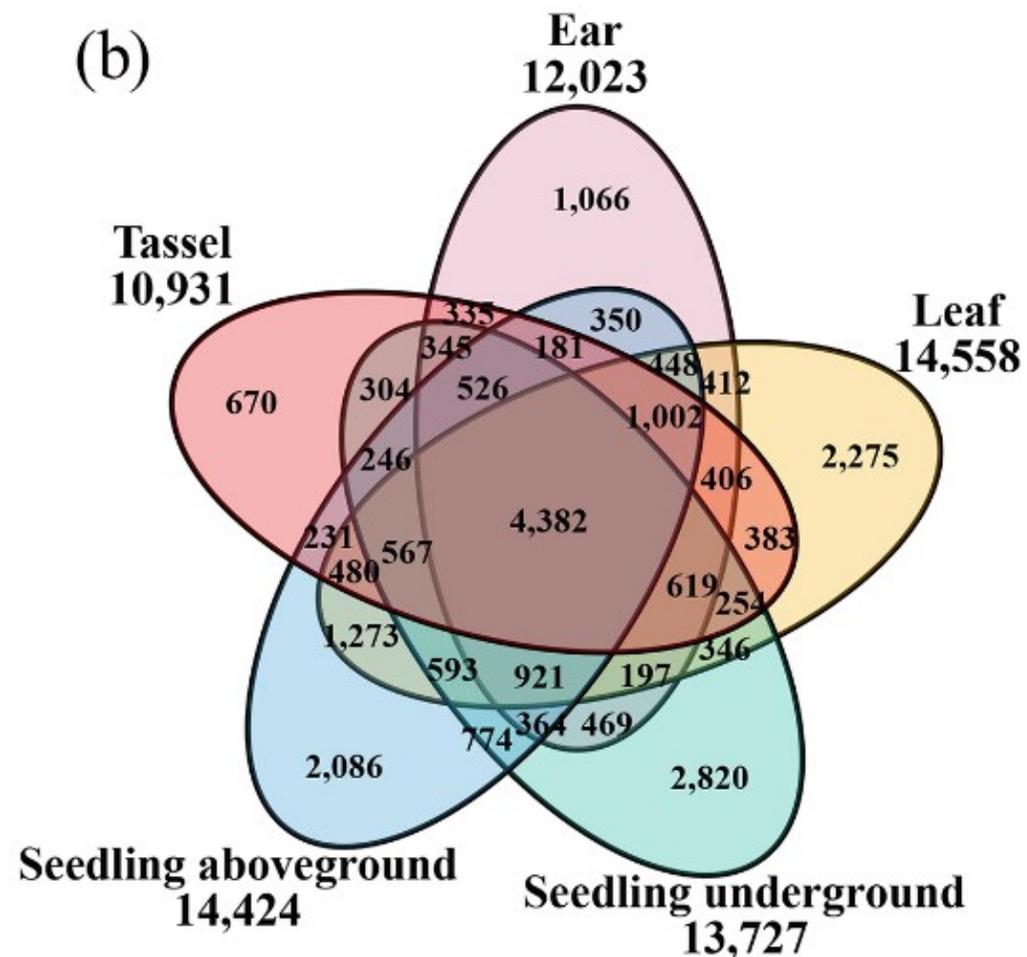
**背景：**2021年开发，需要进行多年生性状的快速定位。

**结果：**Li *et al.*, 2022, *Molecular Plant*

玉米与大刍草的表型差异

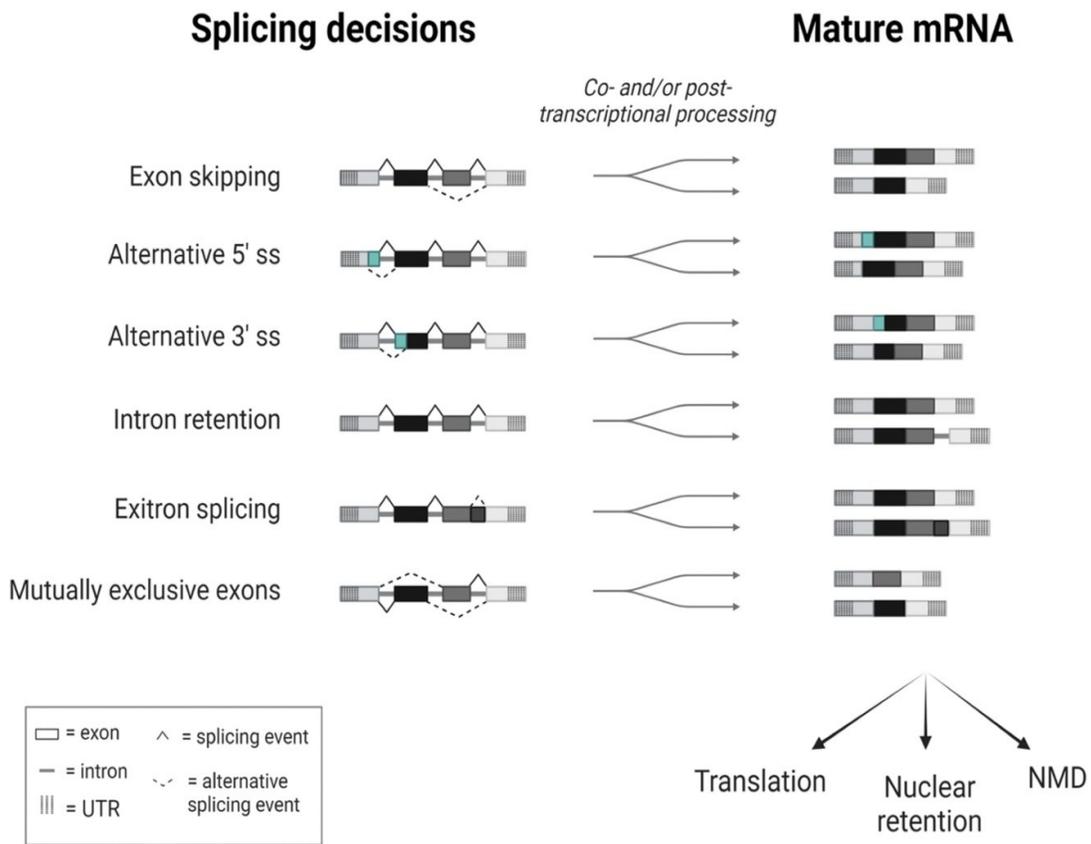


测取大刍草五个组织的转录组

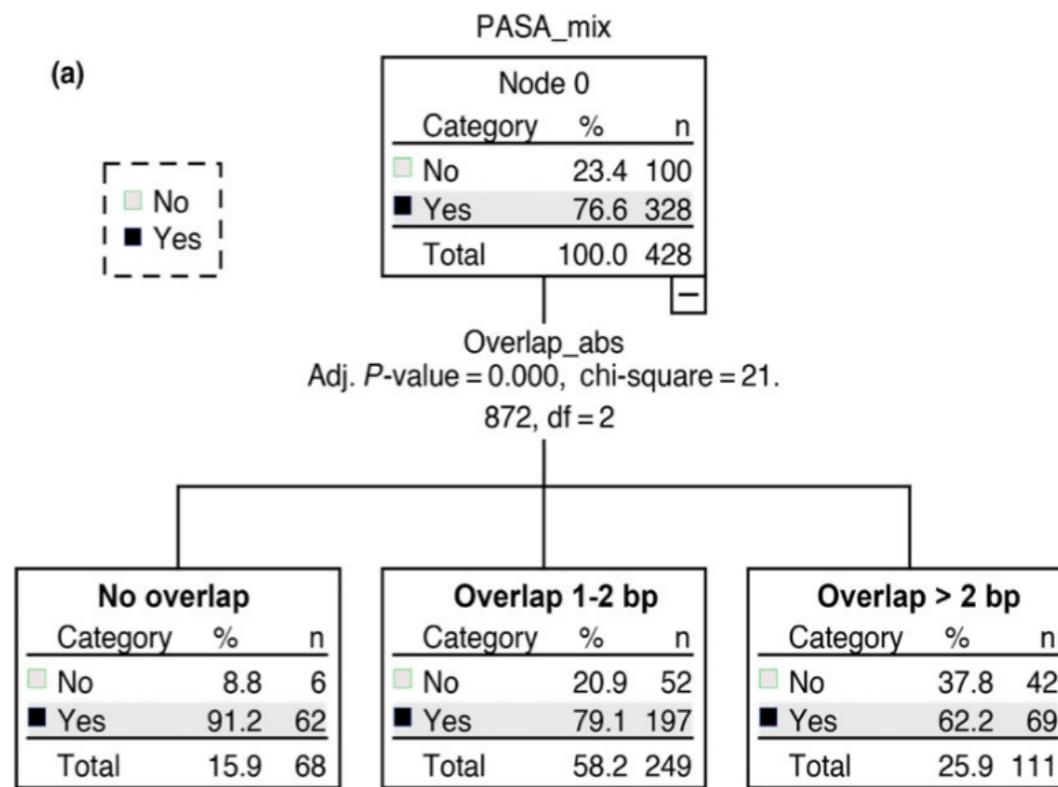


通过比对到基因组鉴定选择性剪接

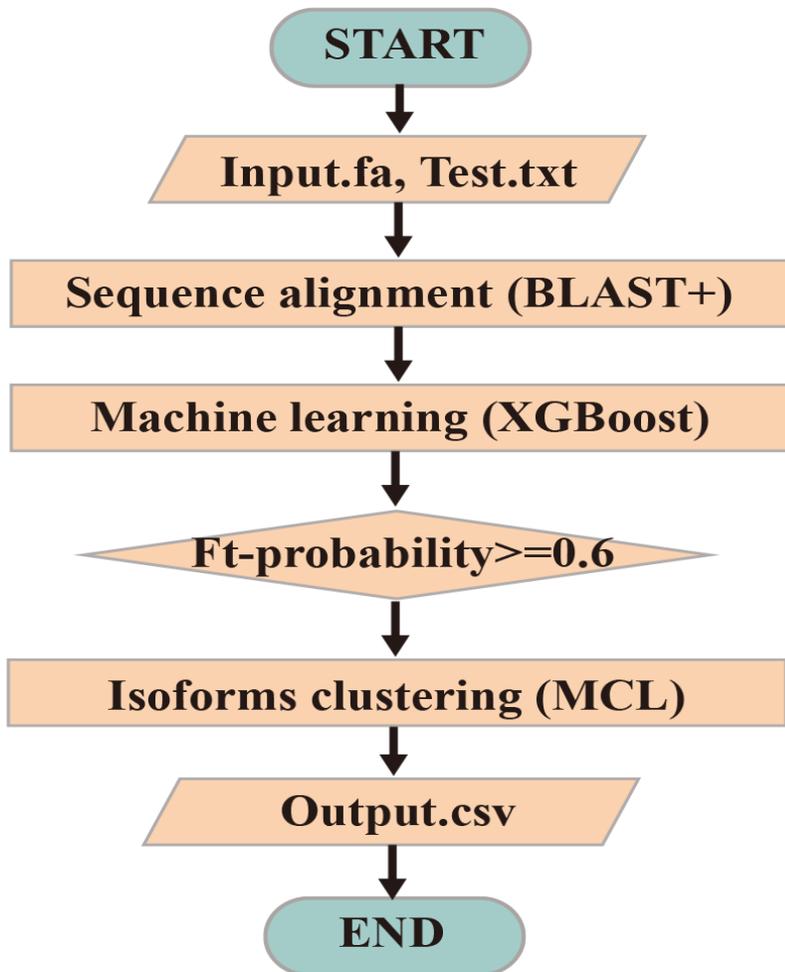
无参工具正确率只有66%–76%



Tognacca et al 2023



Liu et al 2017

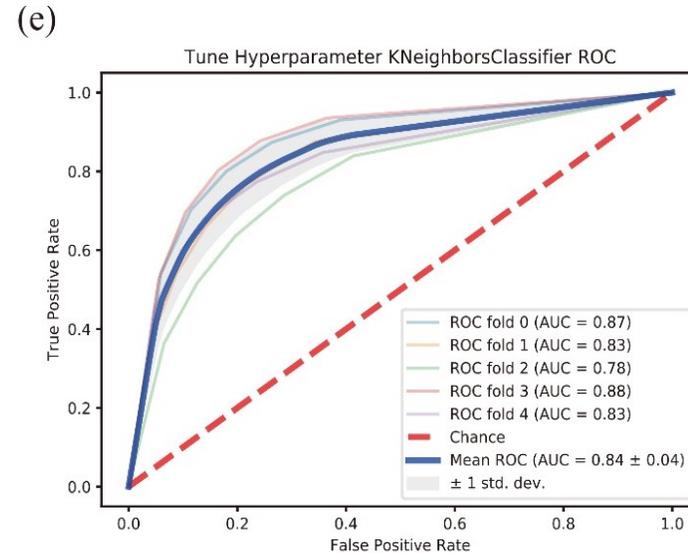
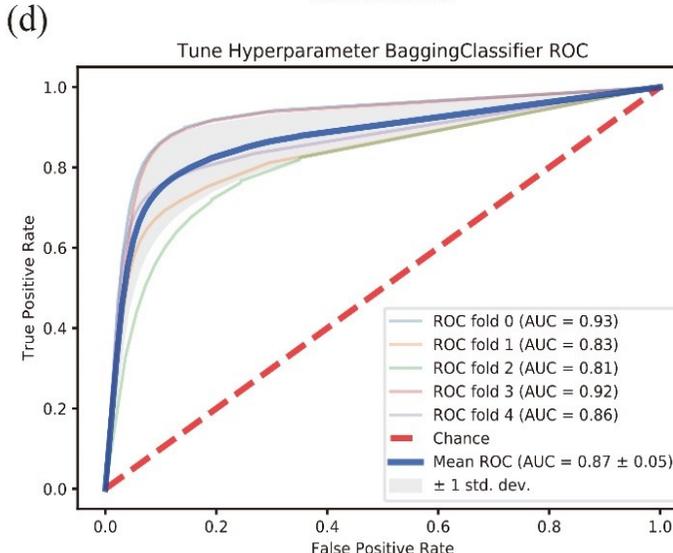
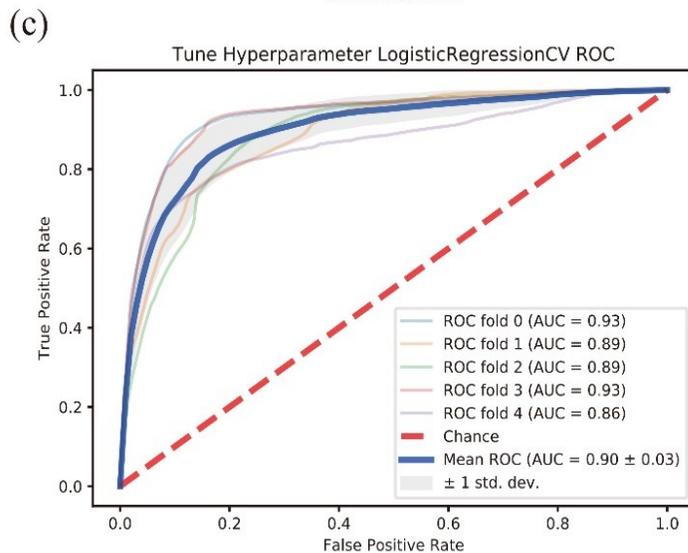
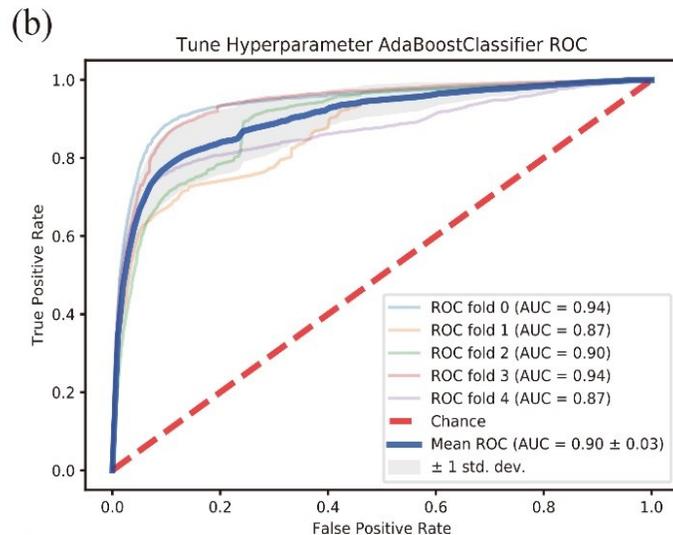
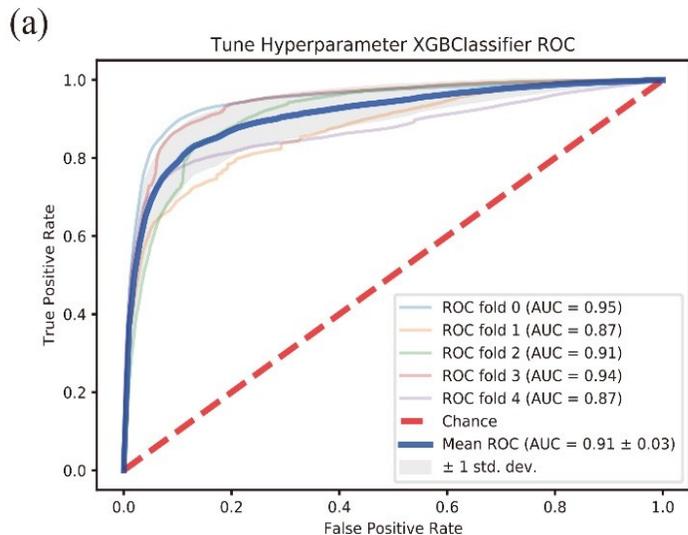


[https://github.com/lizhao007/DenovoAS\\_Finder](https://github.com/lizhao007/DenovoAS_Finder)

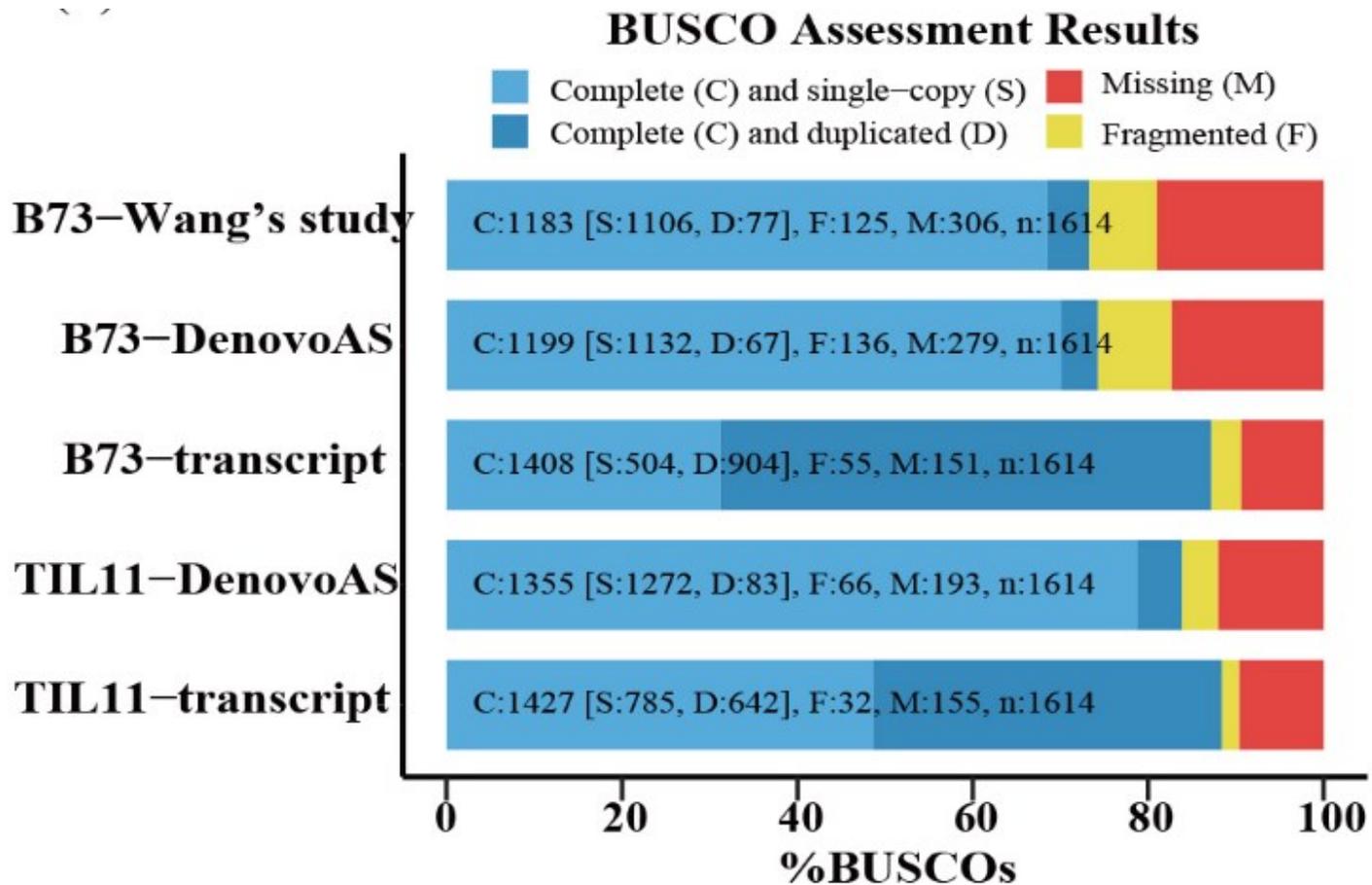
The screenshot shows the GitHub repository page for 'DenovoAS\_Finder' by user 'lizhao007'. The repository is public and contains several files:

File Name	Commit Message	Author	Date	Commits
DenovoAS_Finder.py	Add files via upload	da11255	on Nov 15, 2022	8
README.md	Update README.md	da11255	on Nov 15, 2022	8
data.fa	Add files via upload	da11255	on Nov 15, 2022	8
test.txt	Add files via upload	da11255	on Nov 15, 2022	8

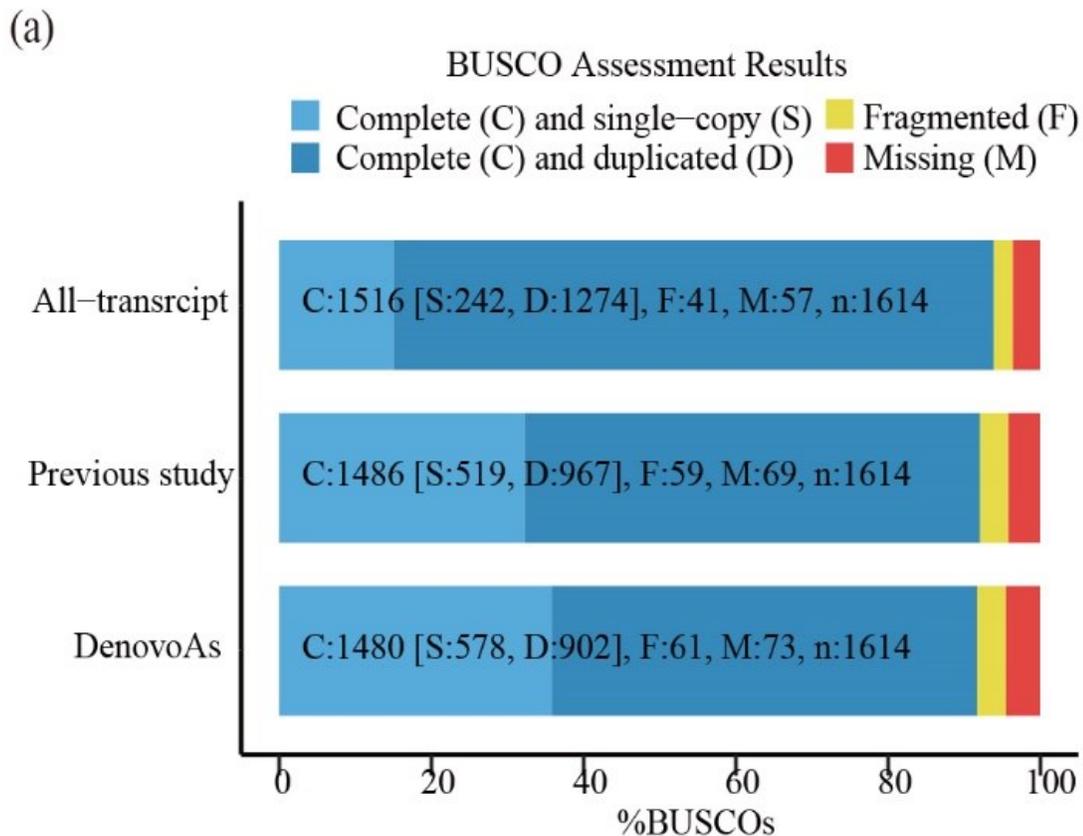
The repository also shows a 'main' branch with 1 branch and 0 tags. The README.md file is currently selected for viewing.



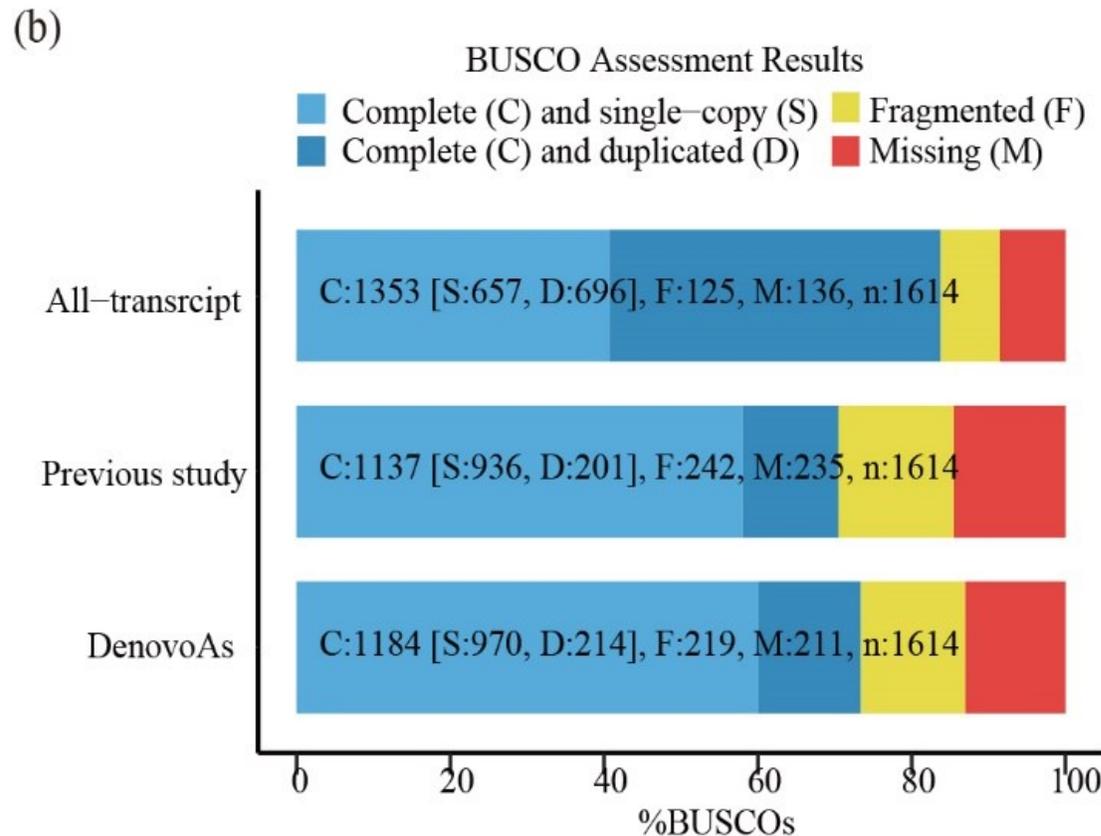
- (a) XGBoost
- (b) AdaBoost
- (c) LogisticRegression
- (d) Bagging
- (e) KNeighbors



禾本科二倍体物种  
*Panicum virgatum* L



锦葵科四倍体物种  
*Gossypium barbadense* L. cultivar 3-79



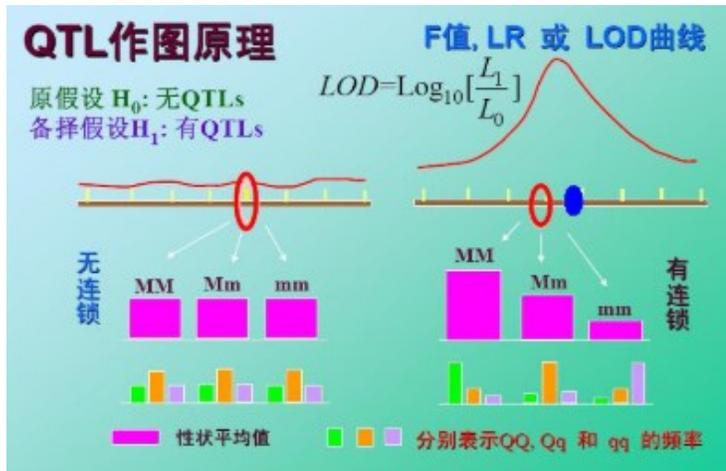
功能基因定位方法

连锁分析

(标记与位点的共分离)

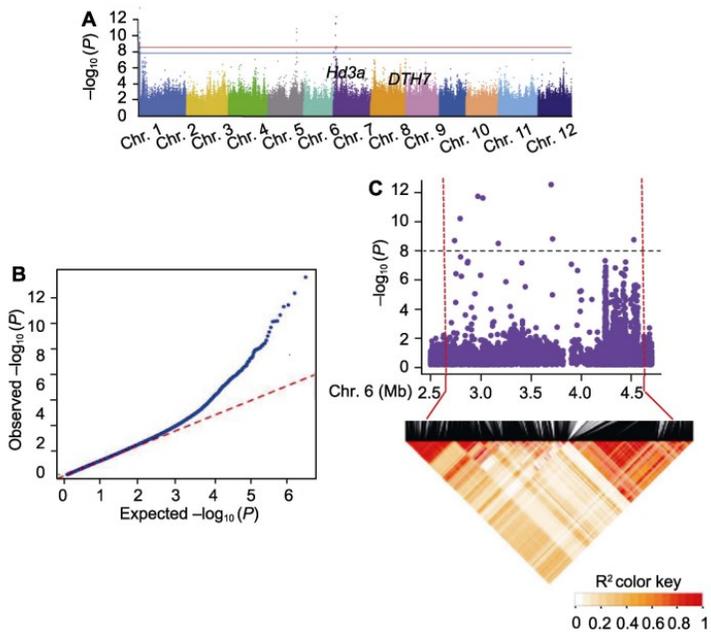
关联分析

(连锁不平衡)

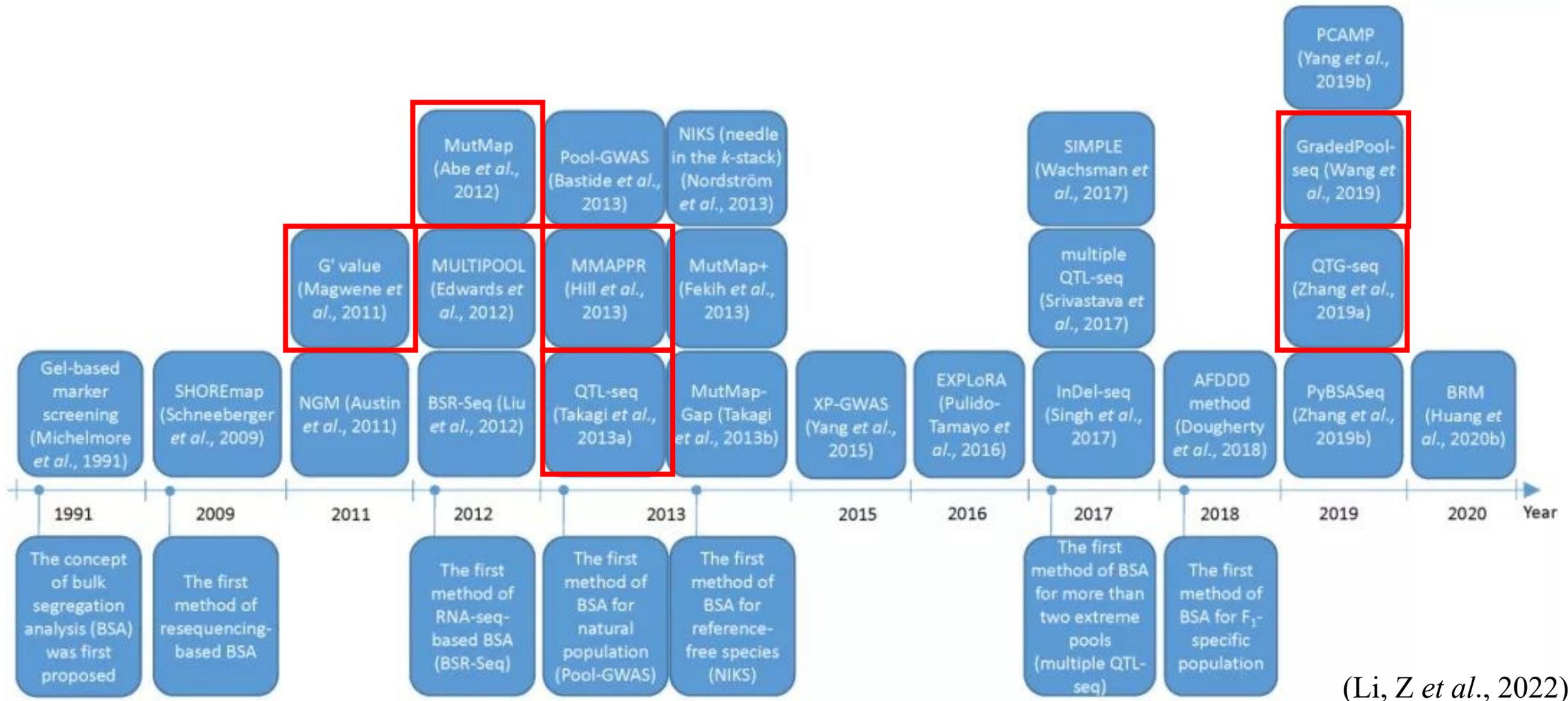


高密度遗传图谱

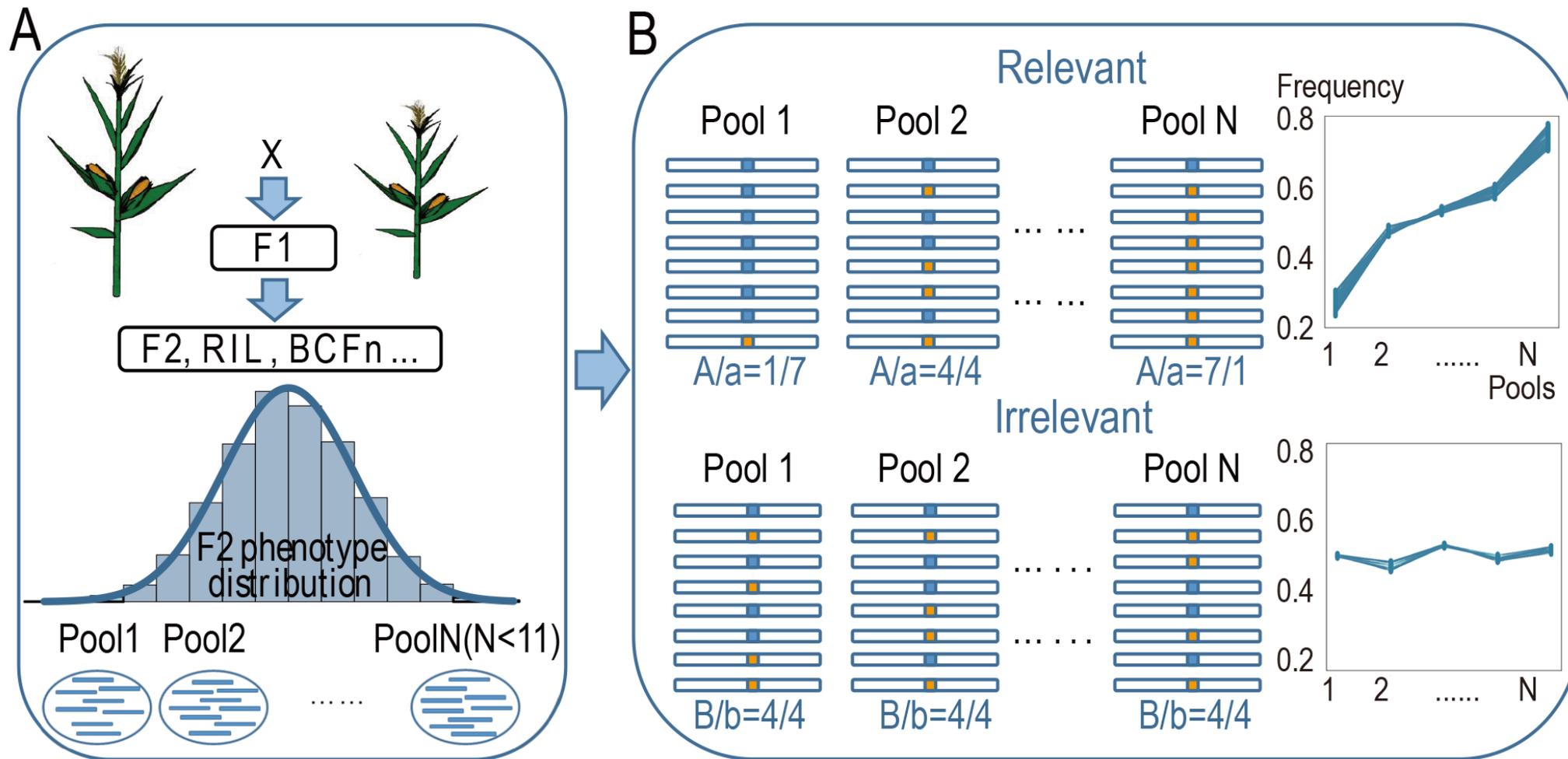
混池测序



赵宇慧等, 2020, 植物学报

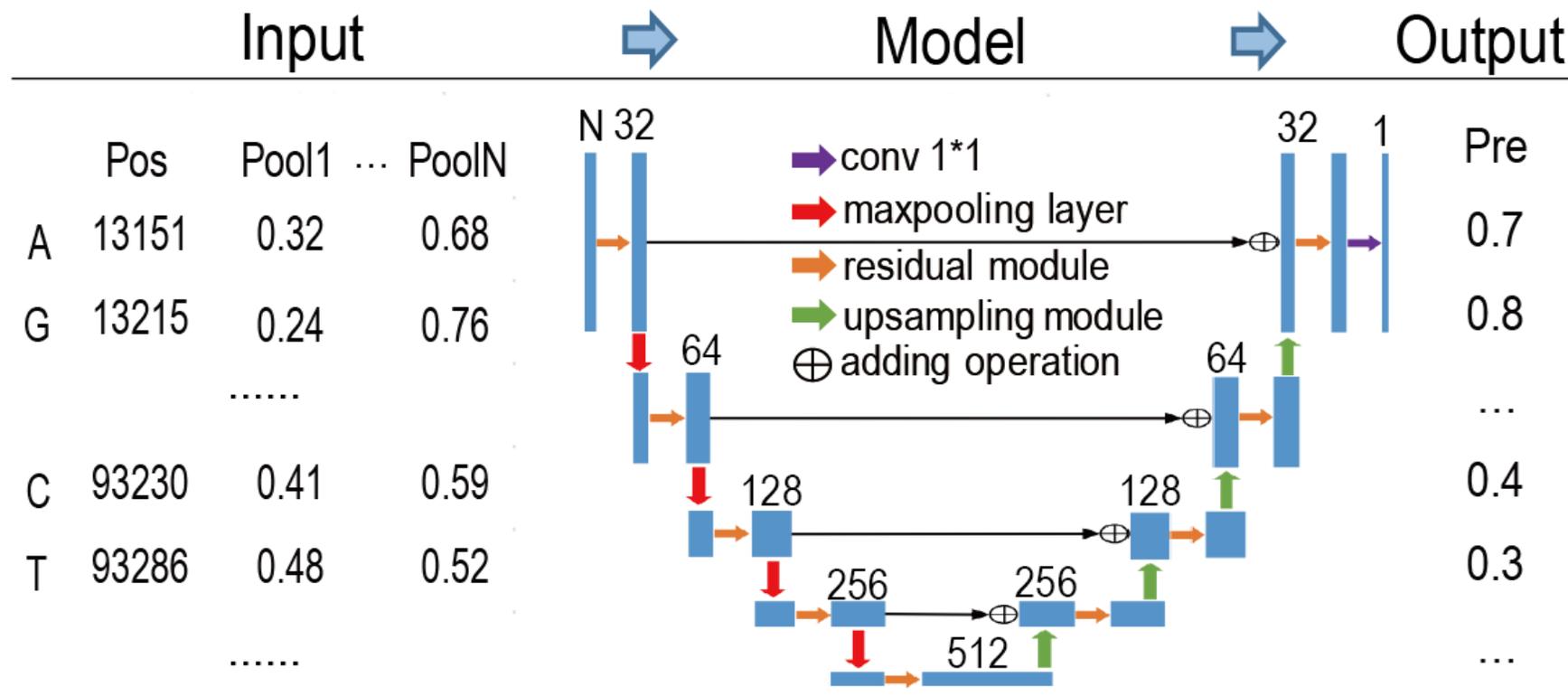


(Li, Z et al., 2022)



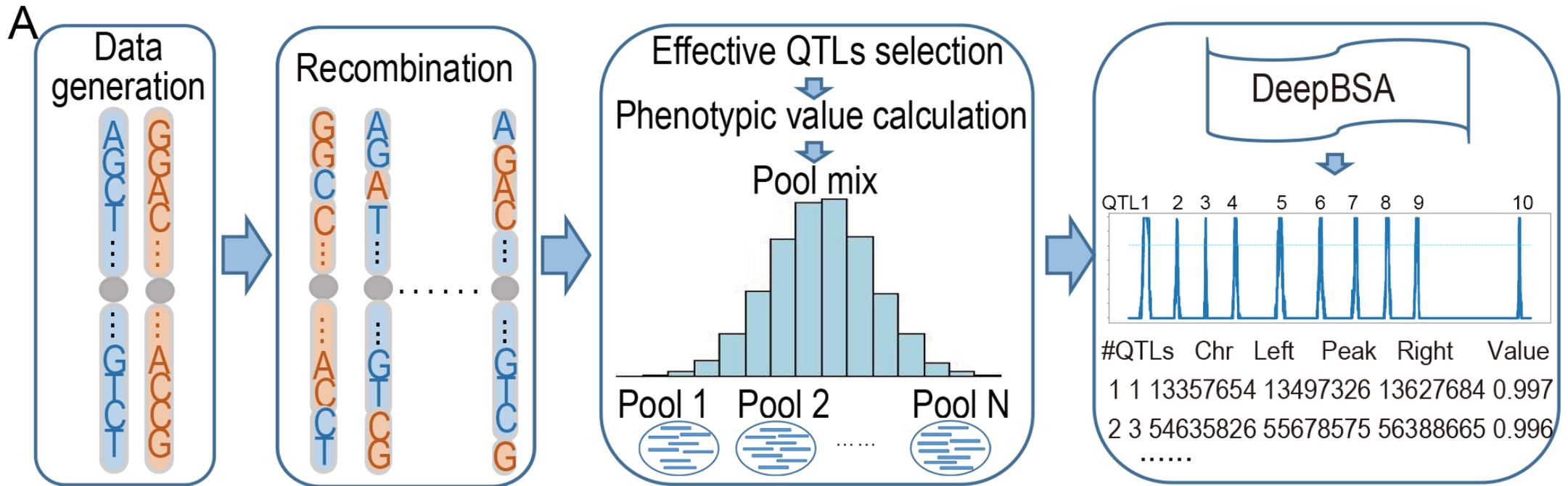
F<sub>2</sub>群体构建, 表型统计及混池测序

SNP检测及正负样本划分



深度学习模型的架构及输入输出

模拟真实的混池数据生成



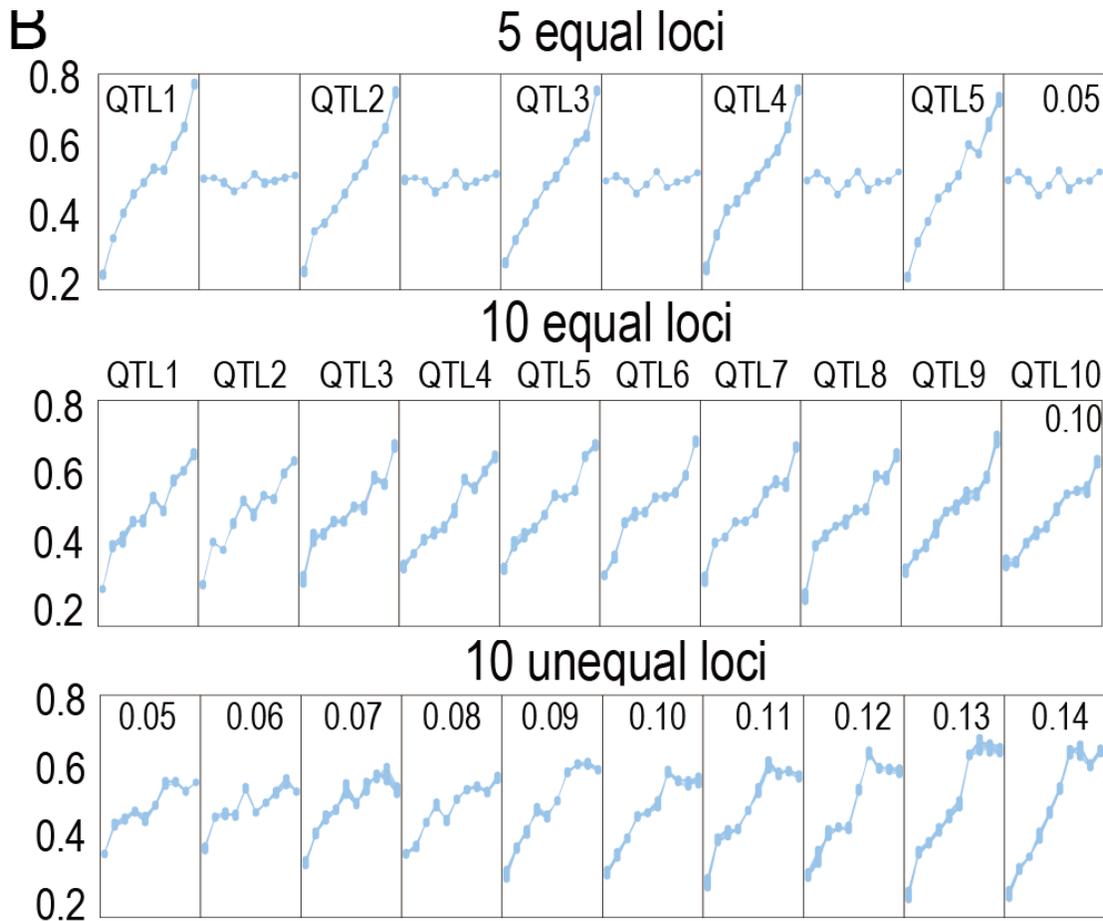
生成一个二倍体基因组

模拟配子的重组交换，重组次数25

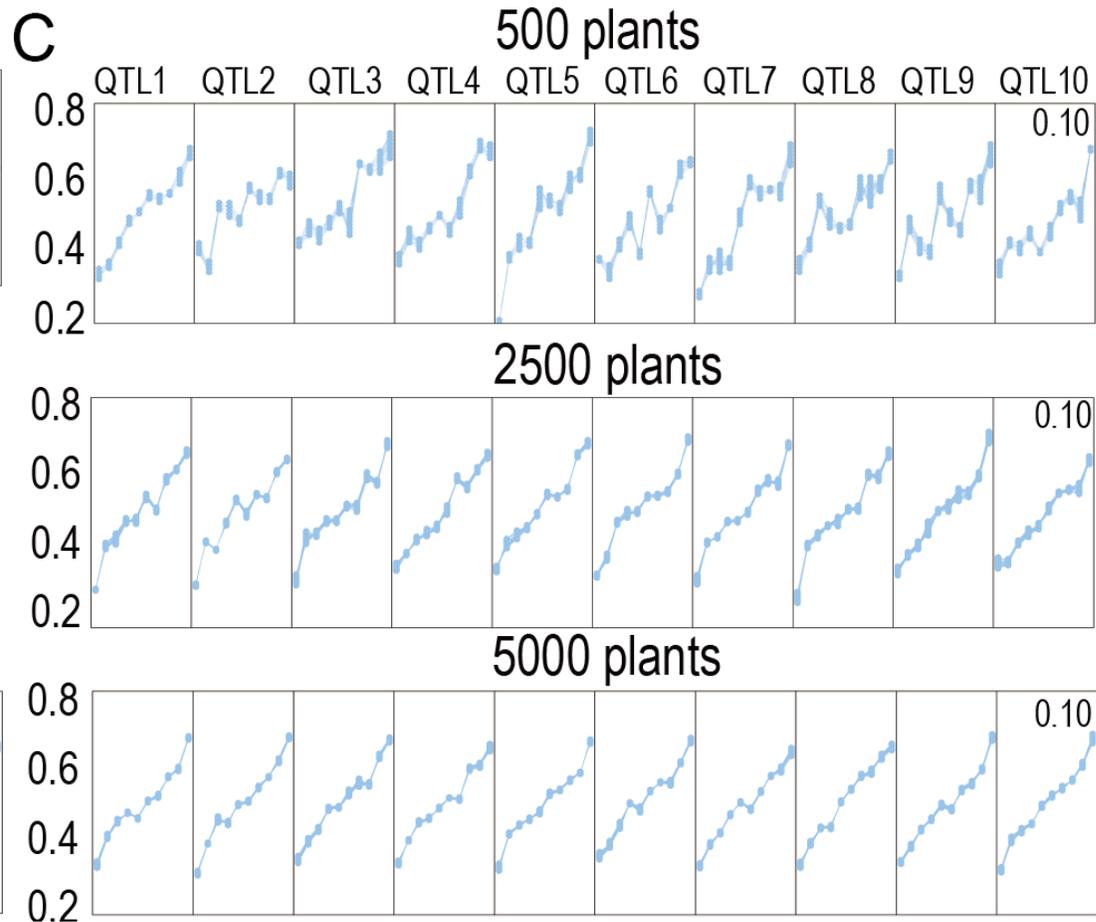
设置效应位点，统计单株表型值，模拟混池，产生数据

运行DeepBSA得到结果

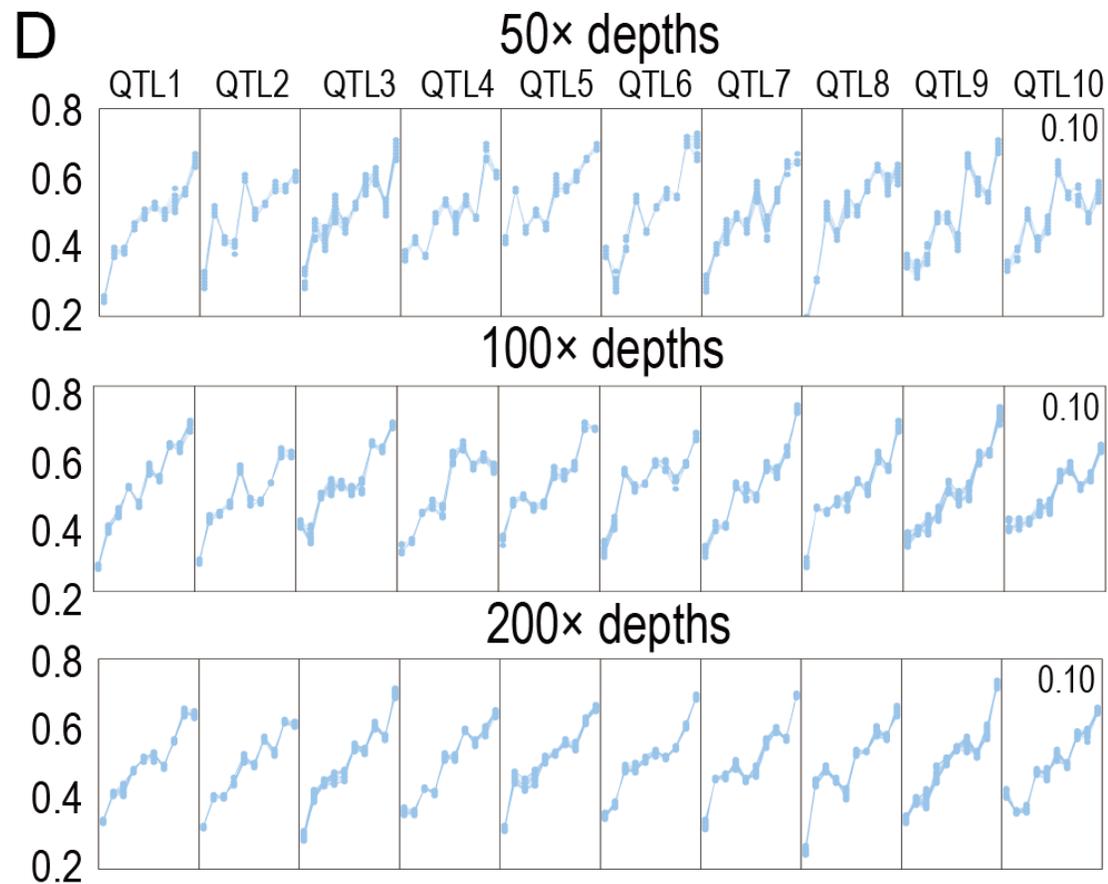
# 深度学习算法 (DL) 适用于复杂性状的定位



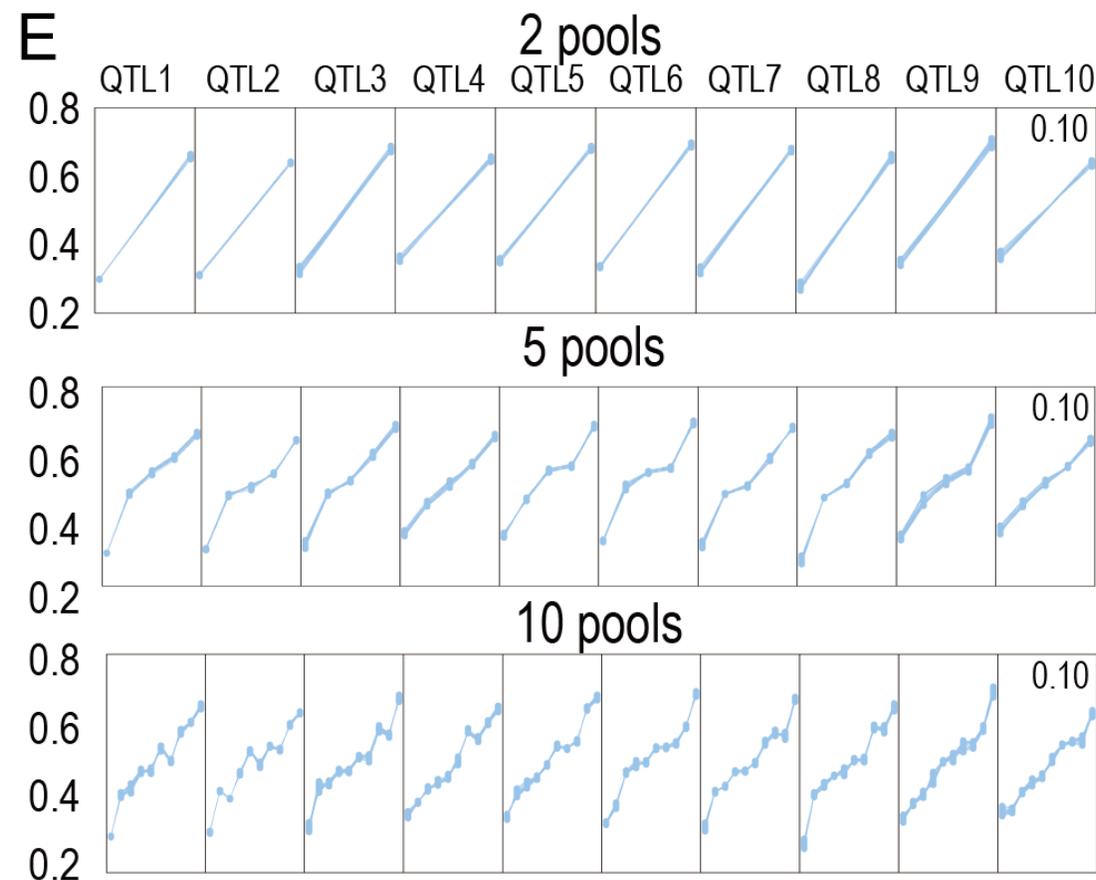
PVE低至5%的QTL能被有效检测到



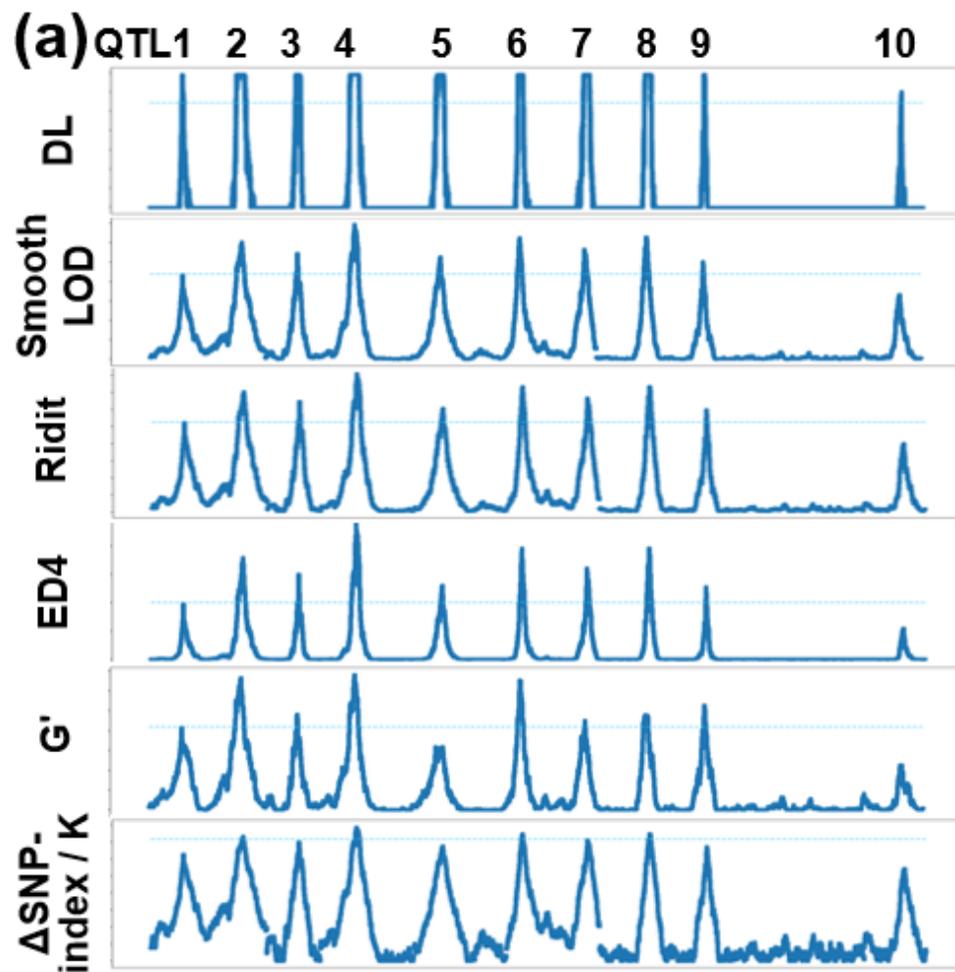
能检测到低至500单株的QTL



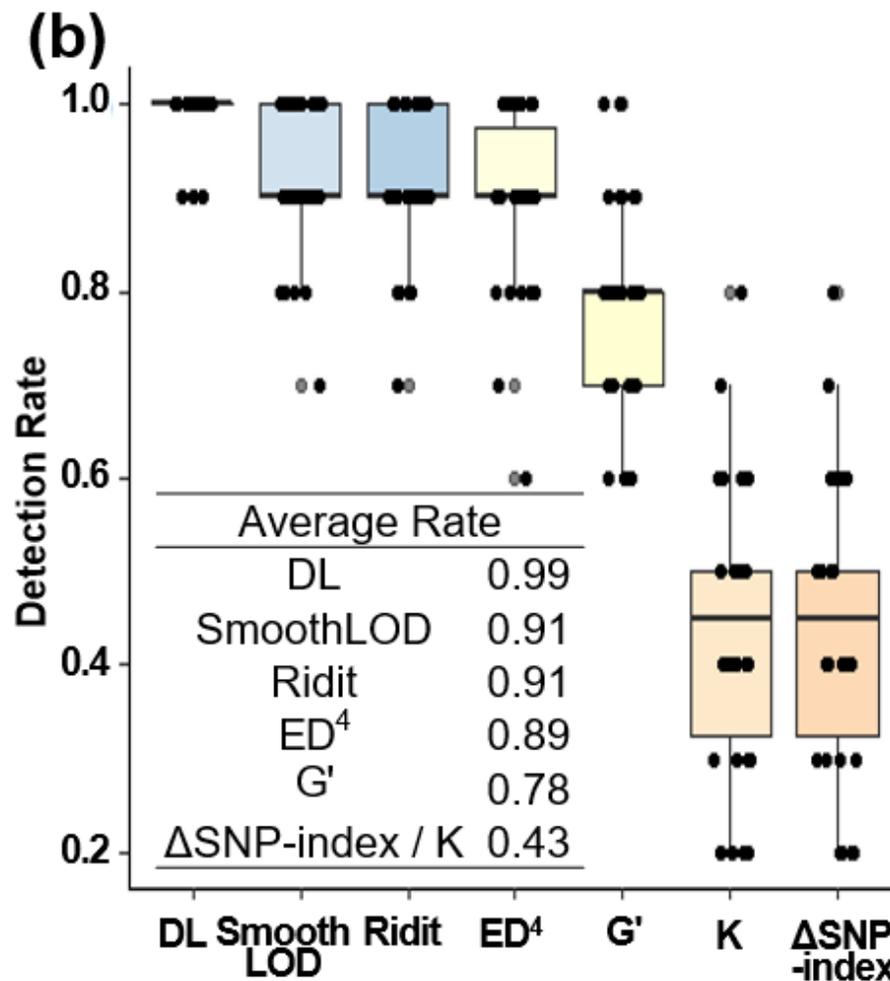
测序深度50×以上检测效果较好



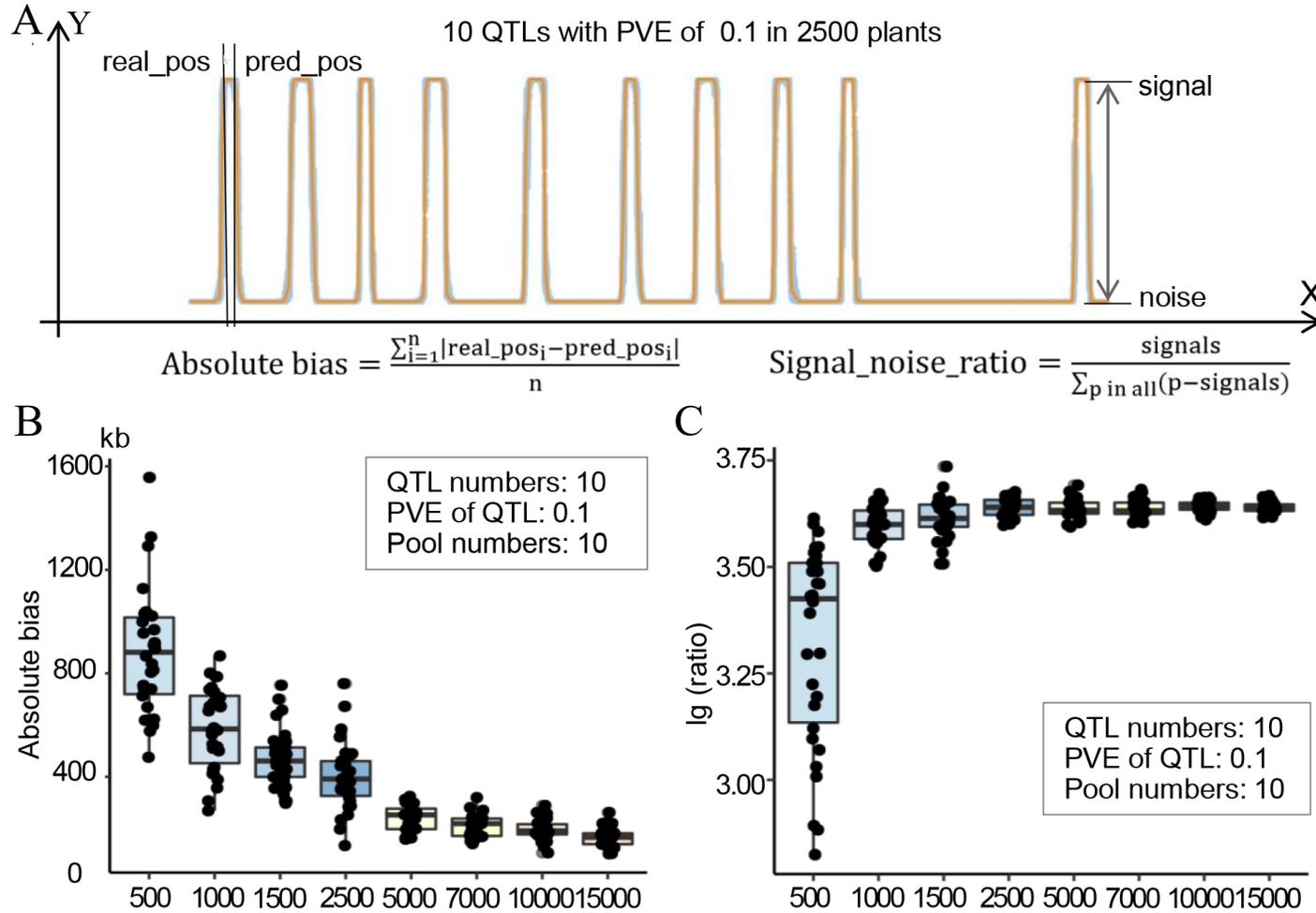
不同混池数下都能检测到目标位点



各个方法的检测结果图

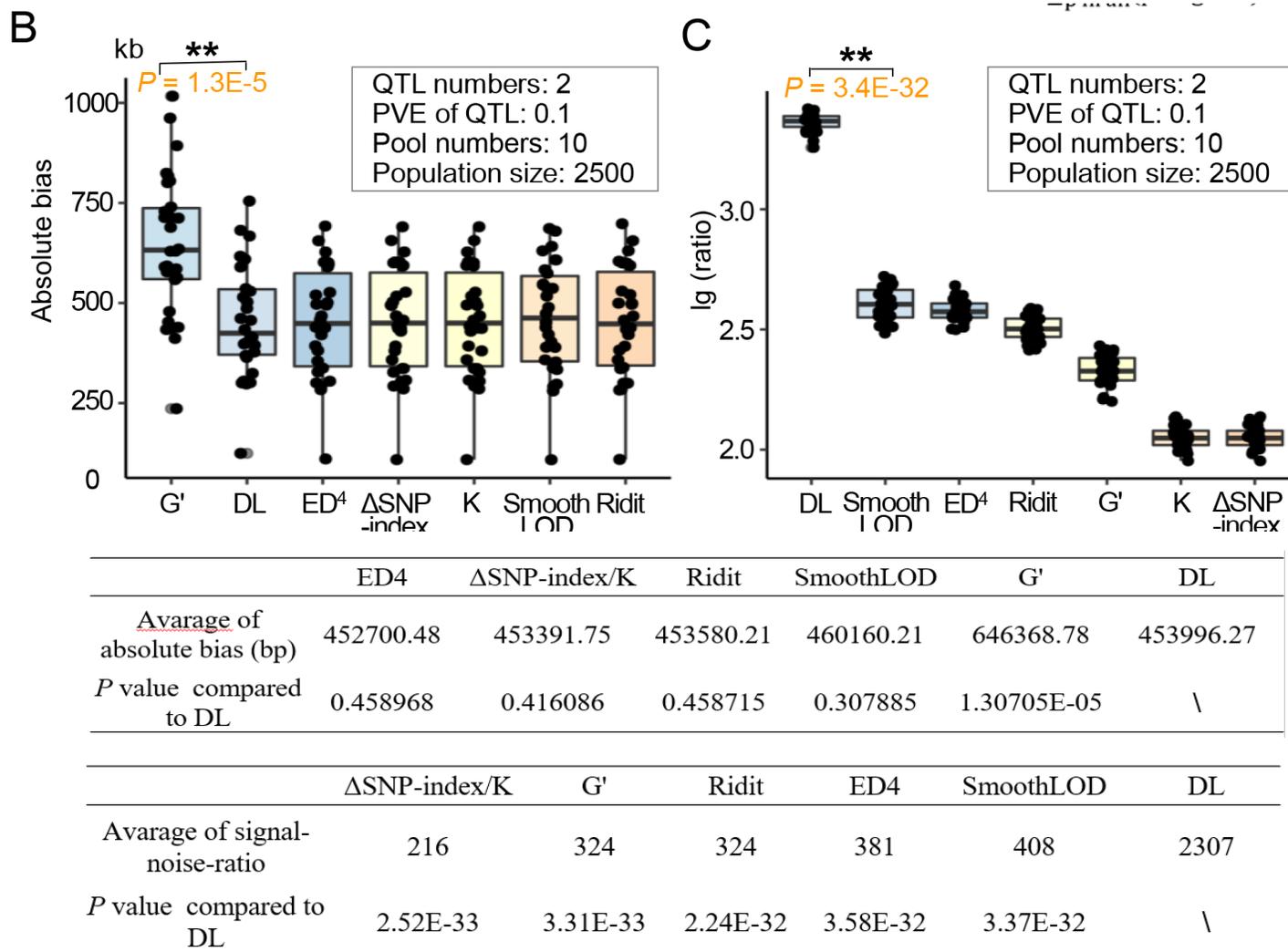


30次模拟下各个方法的QTL检出率



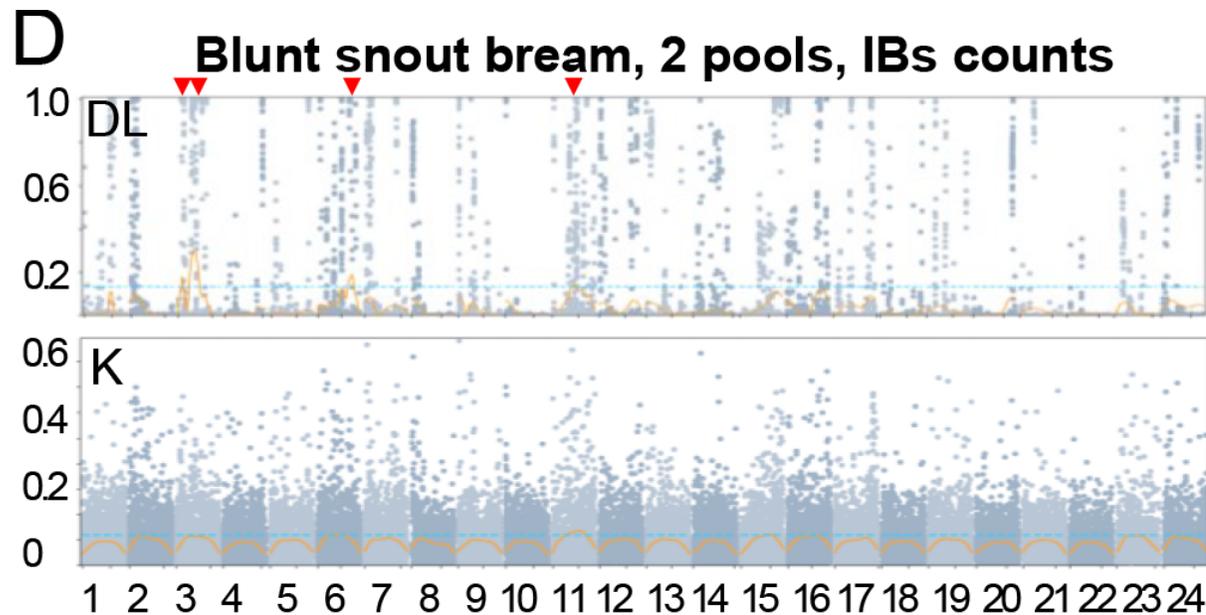
A. 偏离度和信噪比的指标制定； B. DL的偏离度随群体增大而减小； C. DL的信噪比随群体增大而增大。

# DL的偏离度和信噪比优于另五种BSA算法



B.DL的偏离度比其它方法更小； C.DL的信噪比比其它方法更大。

武昌鱼有无肌间刺性状定位，两个混池

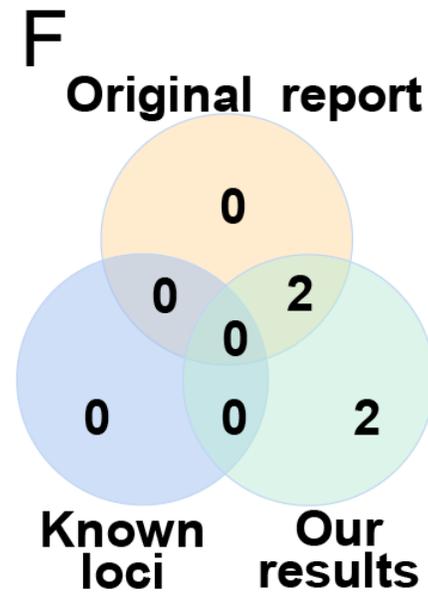


定位的DL和K峰图

Figure E shows the DL localization regions for four chromosomes. The table below summarizes the data:

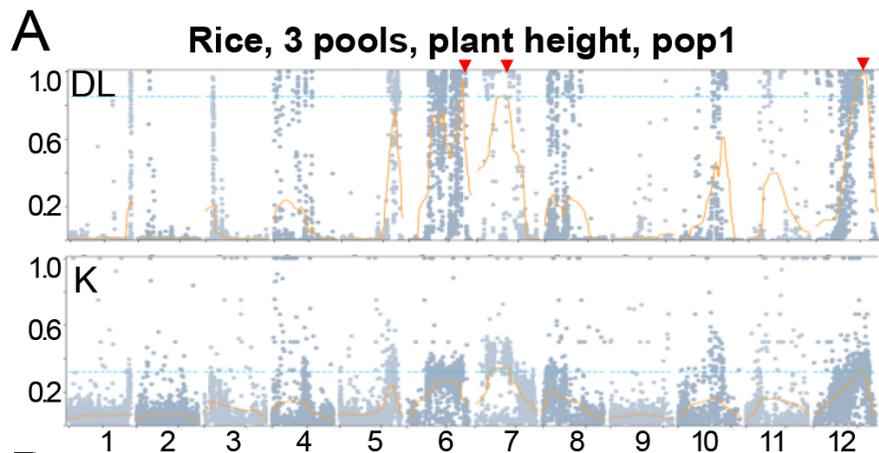
Chr	Region(Mb) ASM v1	Known loci
3	5.04-9.36	
3	16.90-32.41	
6	32.32-39.13	
11	15.25-23.07	

DL的定位区间



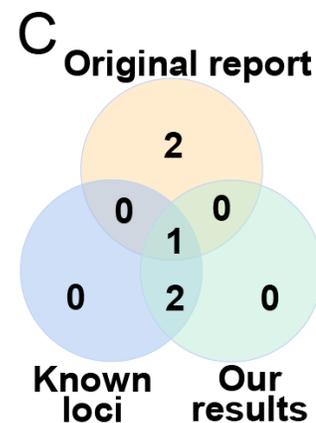
DL方法、原报道和已知基因的韦恩图

### 水稻株高性状定位, 三个混池 (Wang, C. et al. 2019)

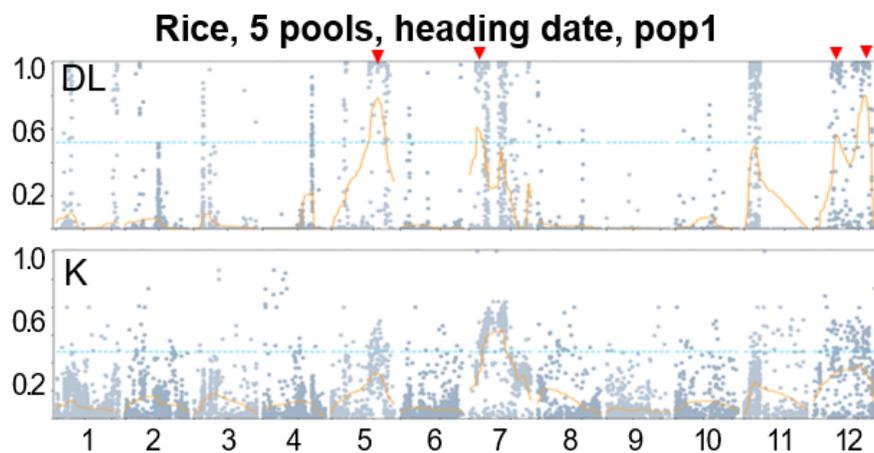


**B**

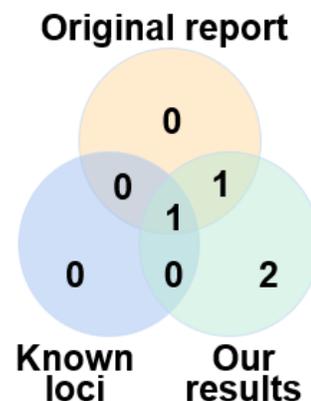
Chr	Region(Mb) IRGSP 4.0	Known loci
6	26.10-27.73	<i>sdt</i>
7	8.34-13.58	<i>Ghd7</i>
12	17.50-23.47	<i>NRL1</i>



### 水稻开花期性状定位, 五个混池 (Wang, C. et al. 2019)

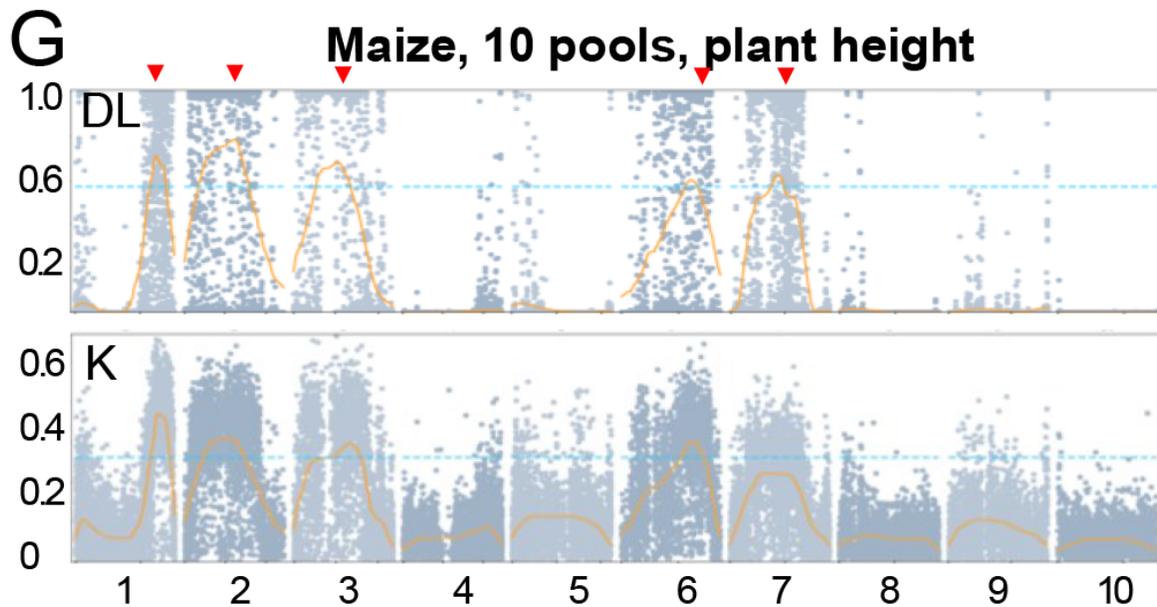


Chr	Region(Mb) IRGSP 4.0	Known loci
5	18.31-26.43	
7	6.20-9.95	<i>Ghd7</i>
12	9.29-11.33	
12	19.39-24.95	



DeepBSA能定位到已知基因和之前发表的区间, 且多出两个候选区间。

### 玉米株高性状定位，十个混池

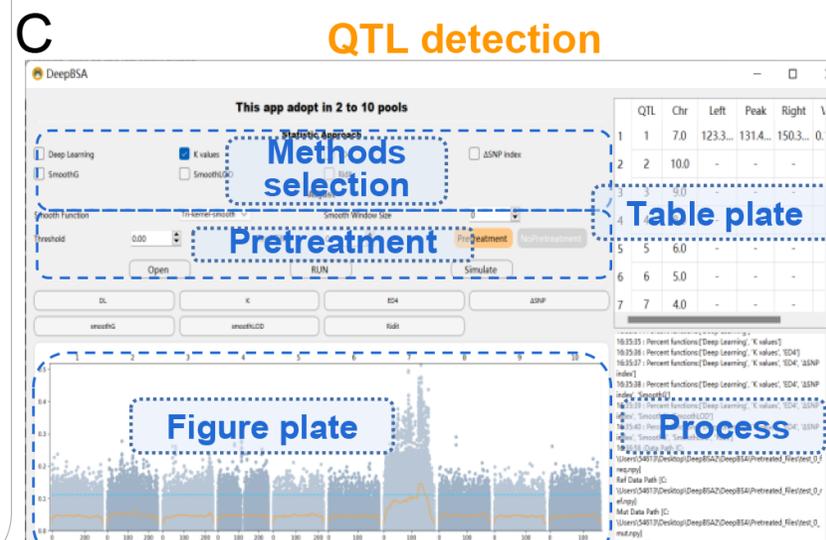
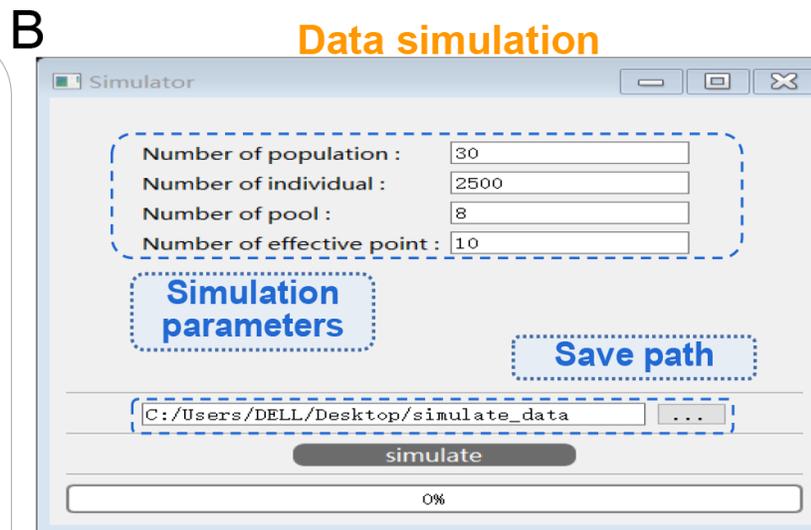
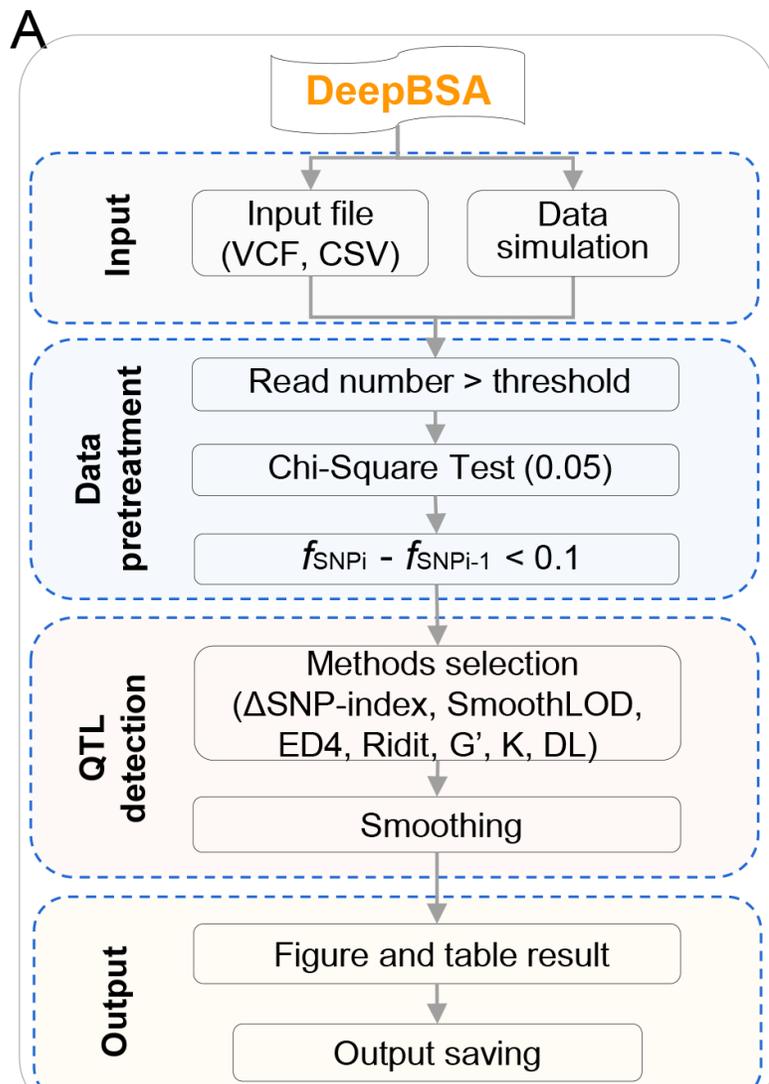


**H**

Chr	Region(Mb) AGP v4	Known loci
1	238.6-249.7	<i>an1</i> , <i>cyp8</i>
2	110.3-140.9	
3	88.4-118.9	
6	119.8-124.5	<i>FEA4</i>
7	84.28-90.15	

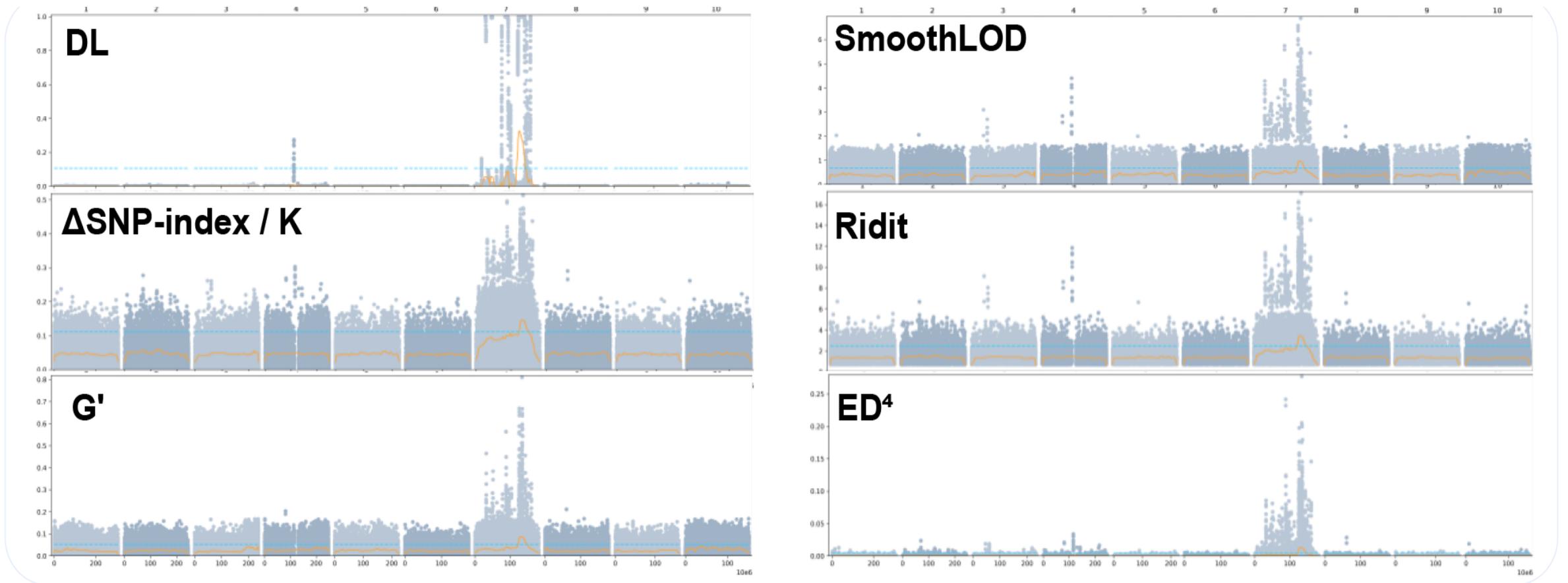


DeepBSA结果和三个已知基因区间重合，另发现两个新区间。



A.软件流程示意图;  
B.仿真数据生成模块;  
C.QTL定位模块。

## DeepBSA包含的七种算法结果示例 (Zhang, H. *et al.* 2019)



七种算法都能定位到目标位点(Chr7:135Mb)

下载地址: <http://zeasystemsbio.hzau.edu.cn/Tools/DeepBSA.zip>

This app adopt in 2 to 10 pools

**Methods selection**

Deep Learning     K values     ED4     SNP index  
 SmoothG     SmoothLOD

**Adjust Pretreatment**

Smooth Function: kernel-smo    2000  
Threshold: 0.10    Read Filter: 0    Pretreatment    NoPretreatment

Open    RUN    Simulate

DL    K    ED4    SNP  
smoothG    smoothLOD    None

**Figure plate**

QTL	Chr	Left	Peak	
1	1	7	12,091,3538	12,944,9034
2	2	10	-	-
3	3	9	-	-
4	4	8	-	-
5	5	6	-	-
6	6	-	-	-
7	7	4	-	-
8	8	3	-	-
9	9	2	-	-
10	10	1	-	-

**Process**

```
20:42:27 :  
Data Path [F:\DeepBAS\Pretreated_Files\pool-18_0_freq.npy]  
Ref Data Path [F:\DeepBAS\Pretreated_Files\pool-18_0_ref.npy]  
Mut Data Path [F:\DeepBAS\Pretreated_Files\pool-18_0_mut.npy]  
Pos Path [F:\DeepBAS\Pretreated_Files\pool-18_0_pos.npy]  
20:42:28 : Deep learning function is executing...  
20:42:50 : Deep learning function has done.  
20:42:50 : K function is executing...  
20:44:01 : K function has done.  
20:44:20 : Smooth Window Size:2000 - Threshold:0.1 - Read  
Number:0  
20:44:21 : Smooth Window Size:2000 - Threshold:0.1 - Read  
Number:0  
20:44:25 :  
Data Path [F:\DeepBAS\Pretreated_Files\pool-18_0_freq.npy]  
Ref Data Path [F:\DeepBAS\Pretreated_Files\pool-18_0_ref.npy]  
Mut Data Path [F:\DeepBAS\Pretreated_Files\pool-18_0_mut.npy]  
Pos Path [F:\DeepBAS\Pretreated_Files\pool-18_0_pos.npy]  
20:44:26 : Deep learning function is executing...  
20:44:47 : Deep learning function has done.  
20:44:47 : K function is executing...
```

欢迎大家下载使用，并提供宝贵建议！

李林教授

李伟夫副教授

刘浩博士

韩林倩博士

陈晓轩、施韶强

李孟助理、严师傅

刘雪美、周涛、宋泽萌、袁广

Lilab全体成员

