

数量形状位点定位与全基因组关联分析

日期: 2022-12-7

实验者: 生信 2001 张子栋

[MarkdownNotes/软件第6次作业.md at main · Bluuur/MarkdownNotes \(github.com\)](#)

[生物信息学原理/软件第6次作业.md · blur/MarkdownNotes - 码云 - 开源中国 \(gitee.com\)](#)

实验目的

1. 了解 R 语言工作环境
2. 熟悉 QTL 和 GWAS 分析的一般流程
3. 掌握至少一种常用 QTL 和 GWAS 分析软件

实验内容

1. 使用 R 包 qtl2 进行 QTL 分析
2. 使用软件 TASSEL 进行 GWAS 分析
3. 熟悉基因型、表型等数据格式, 掌握 QTL 和 GWAS 分析结果的解读和可视化

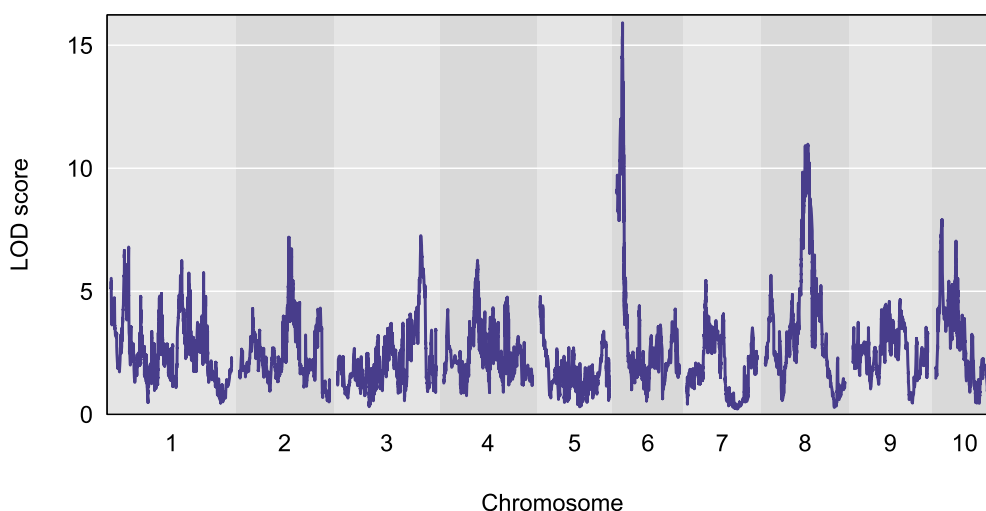
实验操作

1. 利用玉米 MAGIC 群体数据 (https://github.com/rqtl/qtl2data/blob/main/MaizeMAGIC/maize_magic.zip), 根据前面 R/qtl2 使用方法和流程, 使用 R/qtl2 完成以下 QTL 分析:

1. 本群体包含哪些数据?

- 基因型标记文件 `maize_magic_geno.csv`
- 表型文件 `maize_magic_pheno.csv`
- 遗传图谱 `maize_magic_gmap.csv`
- 物理图谱 `maize_magic_pmap.csv`
- 表型协变量文件 `maize_magic_phenocovar.csv`

2. 绘制株高 (PH) 的 LOD 曲线



3. 给出株高对应的 QTL

1	lodindex	lodcolumn	chr	pos	lod	ci_lo	ci_hi
2	4	2	PH	6 12.28468	15.910467	11.009798	12.30230
3	5	2	PH	8 90.14981	10.979198	79.445938	93.54998
4	6	2	PH	10 13.94991	7.916526	9.585541	43.14987

代码实现:

```
1 > # 安装 R 包并导入
2 > if (!require('BiocManager')){
3 +   install.packages("BiocManager")
4 +   library("BiocManager")
5 + }
6 > if (!require('qt12')){
7 +   install.packages("qt12")
8 +   library("qt12")
9 + }
10 > # 加载数据
11 > maizeMagic <- read_cross2(file.choose())
12 > # 查看数据
13 > summary(maizeMagic)
14 Object of class cross2 (crosstype "genri19")
15
16 Total individuals           529
17 No. genotyped individuals   529
18 No. phenotyped individuals  529
19 No. with both geno & pheno  529
20
21 No. phenotypes              4
22 No. covariates              0
23 No. phenotype covariates    1
24
25 No. chromosomes            10
26 Total markers               41324
27
28 No. markers by chr:
29   1    2    3    4    5    6    7    8    9   10
30 6594 4788 4675 4663 4458 3295 3380 3567 3025 2879
31 > # 标记插入遗传图, 获取假定 QTL; 以 1 cM 为间隔插入伪标记
32 > maizeMagicMap <- qt12::insert_pseudomarkers(map = maizeMagic$gmap,
33 + step = 1)
34 > # 计算基因型概率, 假定基因分型误差概率 0.002
35 > maizeMagicPr <- qt12::calc_genoprob(cross = maizeMagic, map =
36 + maizeMagicMap, error_prob = 0.002)
37 > # 可视化查看基因型效率
38 > # 参数以此为 基因型效率, marker 图, 要查看的个体编号, 要查看的染色体号
39 > # 染色体的坐标在横轴上, 基因型在纵轴上, 较高的基因型概率表示为暗色
40 > qt12::plot_genoprob(maizeMagicPr, maizeMagicMap, ind = 1, chr = 1)
41 > # 运用 Haley-Knott regression 进行基因组扫描
42 > # 可加协变量, 此处不加, 输出 LOD 分数矩阵 (positions \times phenotypes)
43 > maizeMagicScan1Out <- qt12::scan1(maizeMagicPr, maizeMagic$pheno,
44 + cores = 0)
```

```

42 > # 绘制 LOD 曲线, 指定一列 表型 PH
43 > qt12::plot_scan1(maizeMagicScan1Out, map = maizeMagicMap, lodcolumn =
  "PH")
44 > # 查看对 PH 而言 LOD 分数最高的伪标记, LOD 得分最高的基因型标记
45 > which.max(maizeMagicScan1Out[, "PH"])
46 PZE.106020122
47 26298
48 > # permutation test 说明scan结果的统计学意义
49 > # 识别随机下可能出现的最大 LOD 分数, 使用 1000 种排列
50 > maizeMagicOperm <- qt12::scan1perm(genoprobs = maizeMagicPr, pheno =
  maizeMagic$pheno, n_perm = 1000)
51 > # 显著性阈值, 默认 5% 水平
52 > # PH 为 7.45 期望 LOD 得分低于 7.45 是偶然事件
53 > summary(maizeMagicOperm, alpha = 0.05)
54 LOD thresholds (1000 permutations)
55 PS PH EH GYrad
56 0.05 7.33 7.53 7.82 7.75
57 > # 寻峰, 95% 置信区间
58 > thr <- summary(maizeMagicOperm)
59 > maizeMagicPeaks <- qt12::find_peaks(scan1_output = maizeMagicScan1Out,
  map = maizeMagicMap, threshold = thr, prob = 0.95, expand2markers =
  FALSE)
60 > # 查看 PH 表型对应多少满足阈值的峰(QTL), 分别在哪儿
61 > maizeMagicPeaks[maizeMagicPeaks$lodcolumn == "PH", ]
62 lodindex lodcolumn chr pos lod ci_lo ci_hi
63 4 2 PH 6 12.28468 15.910467 11.009798 12.30230
64 5 2 PH 8 90.14981 10.979198 79.445938 93.54998
65 6 2 PH 10 13.94991 7.916526 9.585541 43.14987

```

2. 使用 TASSEL 对自带数据集 (安装目录下 TutorialData 子目录中), 对其它两个性状 (EarHT 和 EarDia) 中任意一个, 进行关联分析。

至少需要给出以下结果:

1. 与 EarHT 或 EarDia 最显著关联的 Top 5 位点信息 (包括染色体号、位置和 P 值)

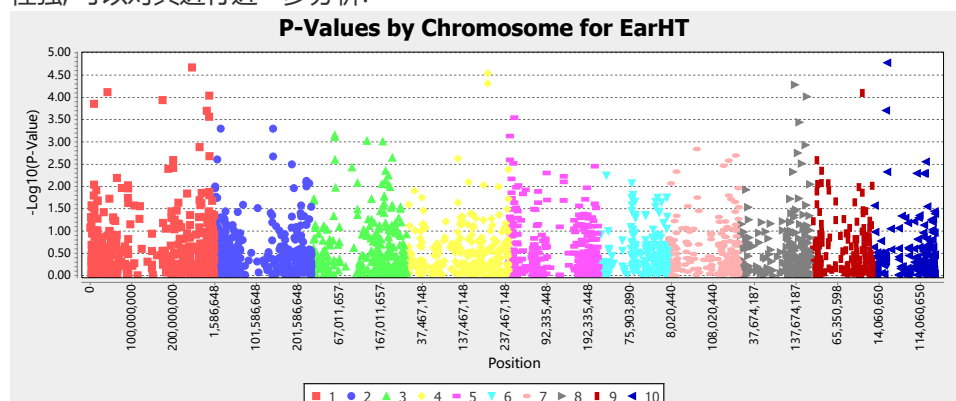
■ 选择性状为 EarHT

Trait	Marker	Chr	Pos	df	F	p	add_effect	add_F	add_p	dom_effect	dom_F	dom_p	errortf	MarkerR2	Genetic Var	Residual Var	-2LnLikelihood
EarHT	None			0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	283	NaN	131.93544	119.64002	2244.66218
EarHT	PZA01597.1	10	61603191	2	1.1403E-4	0.99989	6.5767E-4	2.3484E-7	0.99981	-1.5862E-1	2.2805E-4	0.98796	251.0	9.167E-7	131.93544	119.64002	2244.66218
EarHT	PZA00106	1	10069145	2	2.1144E-4	0.99979	0.01187	4.4577E-5	0.99468	-2.3476E-1	3.9181E-4	0.98422	260.0	1.6072E-6	131.93544	119.64002	2244.66218
EarHT	PZB01389.1	8	134723842	1	1.0094E-7	0.99975	NaN	NaN	NaN	NaN	NaN	NaN	262.0	3.7753E-10	131.93544	119.64002	2244.66218
EarHT	PZA03613.1	1	2914171	2	3.4808E-4	0.99965	0.03152	3.9754E-4	0.98411	0.17382	2.8005E-4	0.98666	264.0	2.5792E-6	131.93544	119.64002	2244.66218
EarHT	PZA01688.3	3	223670423	2	4.1419E-4	0.99959	0.01811	1.5051E-4	0.99022	-2.6208E-1	6.3325E-4	0.97994	256.0	3.1298E-6	131.93544	119.64002	2244.66218

2. 两张图 (曼哈顿图、QQ 图) 及相应的解释 (绘图结果说明什么)

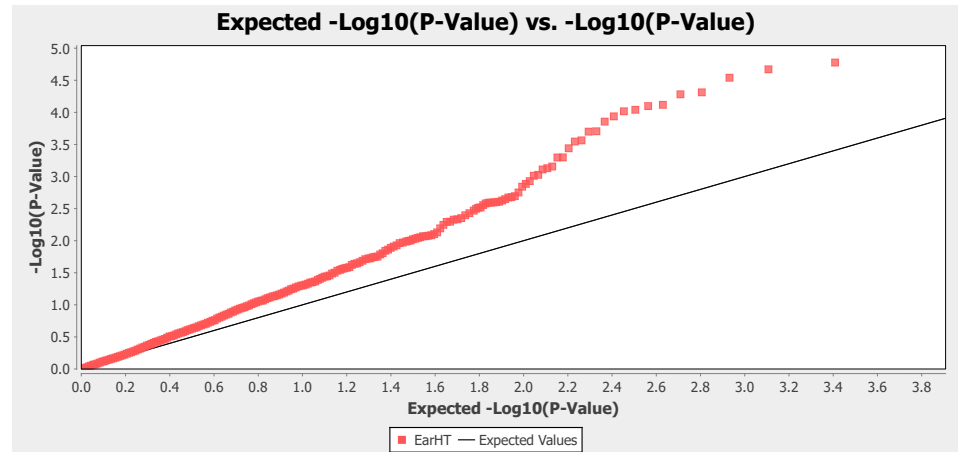
■ 曼哈顿图

■ 一号染色体 五号染色体出现明显堆积峰, 说明这些区域的基因型与 EarHT 表型关联性很强, 可以对其进行进一步分析。



- QQ 图

- QQ 图后半部分脱离均匀分布, 说明基因型与表型之间存在显著相关的自然选择作用.



讨论

在这次上机操作中, 熟悉了 R/qtl 包, TASSEL 的使用.