Overview of binQTL

Wen Yao and Shizhong Xu

2017-06-03

Contents

1 Introduction

Composite interval mapping (CIM) is the dominating algorithm used in QTL analysis of various phenotypes in experimental populations. With the development of next-generation sequening, genotyping of SNPs became much easier, which leads to the development of binmap. In binmap, the distance between adjacent markers were zero. As a result, the concept of "interval mapping" doesn't make sense in binmap. The **binQTL** package is designed to perform QTL analysis with binmap. In this vignette, we will rely on two simple, illustrative example datasets to explain the usage of **binQTL**.

To use the **binQTL** package, we need to load it into R first.

```
library(binQTL)
```

2 Use of binQTL with example dataset of 210 RILs

The phenotype in the first example dataset is the 1000-grain weight of 210 recombinant inbred lines (RIL) in the year of 1999. The genetic map of the 210 RILs is a binmap composed of 1619 bins.

2.1 Read in the phenotype file of 210 RILs

The phenotype file for the **binQTL** package should be a text file containing two columns. The first column is the identifier of all accessions while the second column contains the phenotypic value of specific trait for each accession. The column names of this file can be any characters specified by the User.

```
ril.phe <- read.csv(system.file("examples/ril.phe.csv", package="binQTL"), as.is=T)
dim(ril.phe)
```

```
## [1] 210   2
```

```
head(ril.phe)
```

```
##    ril  kgw
## 1 R001 24.4
## 2 R002 24.6
## 3 R003 24.5
## 4 R004 22.8
## 5 R005 25.2
## 6 R006 21.4
```

```
par(mar = c(5, 4, 1, 1))
hist(ril.phe[, 2], breaks=40, xlab="KGW (1999)", main="")
```

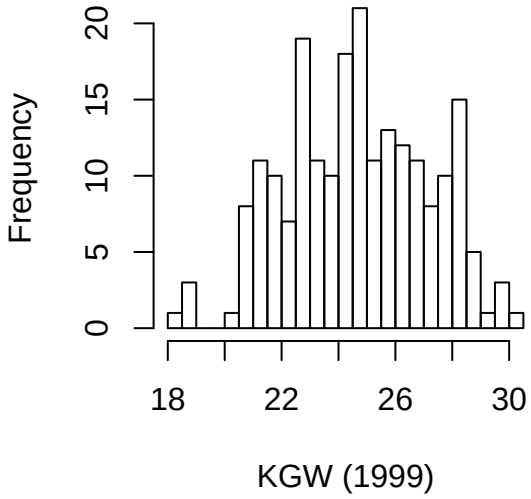

Figure 1: KGW of 210 RILs

2.2 Read in the genotype file of 210 RILs

The genotype file for the **binQTL** package should be a text file containing multiple columns and rows. Each row of this file is a bin. The first four columns contain the information of each bin including the identifier, the chromosome, the start coordinate and the end coordinate. The names of the first four columns should be "Bin", "Chr", "Start" and "Stop". Each of the rest columns gives the genotype of each accession at different bins. The names of the rest columns can be any characters specified by the User.

```
ril.geno <- read.csv(system.file("examples/ril.geno.csv", package="binQTL"), as.is=T)
dim(ril.geno)
```

```
## [1] 1619  214
```

```
ril.geno[1:2, 1:9]
```

```
##    Bin   Chr Start  Stop R001 R002 R003 R004 R005
```

```
## 1 Bin1 chr01 0.000 0.565   P1   P2   P2   P2   P1
## 2 Bin2 chr01 0.565 0.599   P1   P2   P2   P2   P1
```

```
## Not run
# par(mar = c(3, 4, 1, 1))
# plotBinmap(ril.geno, xlab="Genomic position", ylab="RILs", cex.axis=0.6)
## Not run
```

2.3 QTL mapping of RILs using ANOVA

```
qtl.res.1 <- aovQTL(phenotype = ril.phe, genotype = ril.geno)
head(qtl.res.1)
```

```
##    Bin   Chr Start  Stop         p     -logp
## 1 Bin1 chr01 0.000 0.565 0.2563842 0.5911088
## 2 Bin2 chr01 0.565 0.599 0.2143561 0.6688641
## 3 Bin3 chr01 0.599 0.922 0.2102838 0.6771941
## 4 Bin4 chr01 0.922 1.075 0.1410645 0.8505823
## 5 Bin5 chr01 1.075 1.147 0.1202287 0.9199918
## 6 Bin6 chr01 1.147 1.221 0.1240959 0.9062425
```

2.4 QTL mapping of RILs using binQTL

```
qtl.res.2 <- binQTLScan(phenotype = ril.phe, genotype = ril.geno)
head(qtl.res.2)
```

```
##    Bin   Chr Start  Stop p -logp
## 1 Bin1 chr01 0.000 0.565 1     0
## 2 Bin2 chr01 0.565 0.599 1     0
## 3 Bin3 chr01 0.599 0.922 1     0
## 4 Bin4 chr01 0.922 1.075 1     0
## 5 Bin5 chr01 1.075 1.147 1     0
## 6 Bin6 chr01 1.147 1.221 1     0
```

2.5 Comparison of different algotrithms

Let's get a glimpse of the QTL mapping results of different algorithms.

```
par(mar = c(3, 4, 2, 1))
par(mfrow=c(2, 1))
plotQTL(qtl.res.1[, c(1:4, 6)], main="ANOVA", cex.axis=0.6, ylab=expression(-log[10](p)))
plotQTL(qtl.res.2[, c(1:4, 6)], main="binQTL", cex.axis=0.6, ylab=expression(-log[10](p)))
```

3 Use of binQTL with example dataset of 278 IMF2s

The phenotype in the second example dataset is the 1000-grain weight of 278 immortalized F2 (IMF2) in the year of 1999. The genetic map of the 278 IMF2s is a binmap composed of 1619 bins.

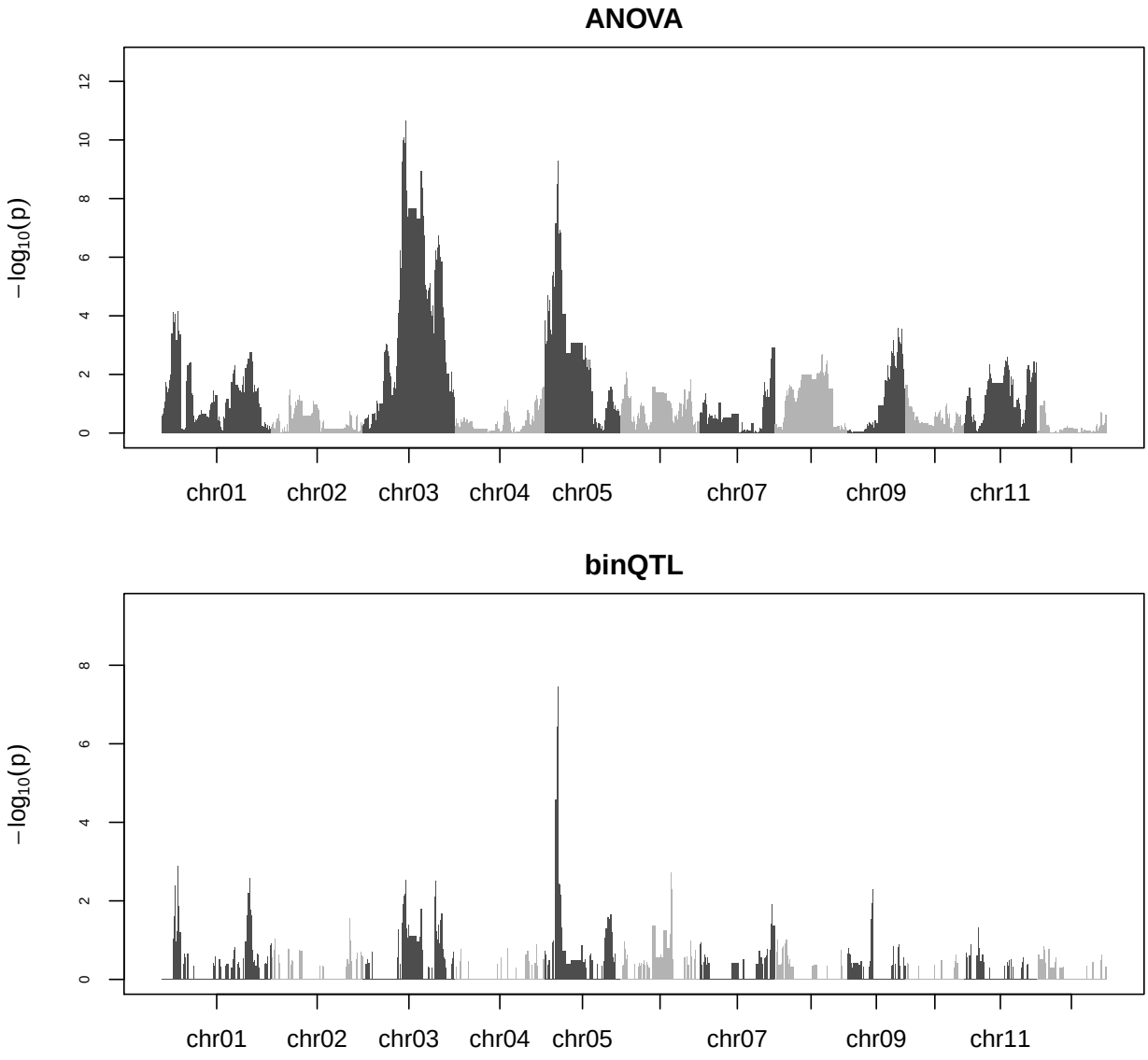

Figure 2: QTL mapping of 210 RILs using different algorithms.

3.1 Read in the phenotype file of 278 IMF2s

```
imf2.phe <- read.csv(system.file("examples/imf2.phe.csv", package="binQTL"), as.is=T)
dim(imf2.phe)
```

```
## [1] 278   2
```

```
head(imf2.phe)
```

```
##   cross  kgw99
## 1  F001 25.334
## 2  F002 22.432
## 3  F003 26.411
## 4  F005 25.029
## 5  F006 23.573
## 6  F008 26.463
```

```
par(mar = c(5, 4, 1, 1))
hist(imf2.phe[, 2], breaks=40, xlab="KGW (1999)", main="")
```

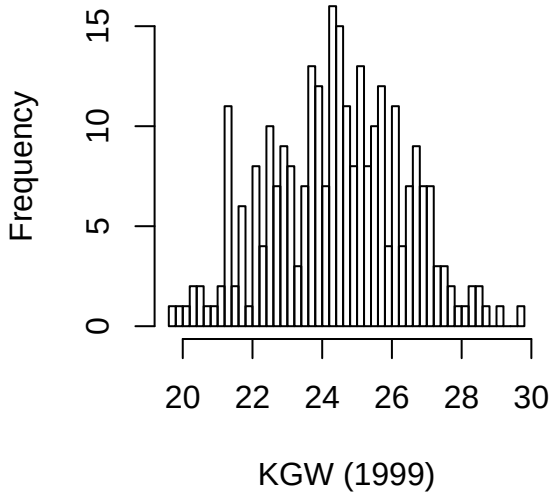

Figure 3: KGW of 278 IMF2s

3.2 Read in the genotype file of 278 IMF2s

```
imf2.geno <- read.csv(system.file("examples/imf2.geno.csv", package="binQTL"), as.is=T)
dim(imf2.geno)
```

```
## [1] 1619  282
```

```
imf2.geno[1:2, 1:9]
```

```
##    Bin   Chr Start  Stop F001 F002 F003 F005 F006
## 1 Bin1 chr01 0.000 0.565   P1   P2   P2    H   P2
## 2 Bin2 chr01 0.565 0.599   P1   P2   P2    H   P2
```

```
## Not run
# par(mar = c(3, 4, 1, 1))
# plotBinmap(imf2.geno, xlab="Genomic position", ylab="IMF2s", cex.axis=0.6)
## Not run
```

3.3 QTL mapping of IMF2s using ANOVA

```
qtl.res.3 <- aovQTL(phenotype = imf2.phe, genotype = imf2.geno)
head(qtl.res.3)
```

```
##    Bin   Chr Start  Stop           p    -logp
## 1 Bin1 chr01 0.000 0.565 0.011152487 1.952628
## 2 Bin2 chr01 0.565 0.599 0.007299795 2.136689
## 3 Bin3 chr01 0.599 0.922 0.008330228 2.079343
## 4 Bin4 chr01 0.922 1.075 0.010369449 1.984244
## 5 Bin5 chr01 1.075 1.147 0.004231273 2.373529
## 6 Bin6 chr01 1.147 1.221 0.004798506 2.318894
```

3.4 QTL mapping of IMF2s using binQTL

```
qtl.res.4 <- binQTLScan(phenotype = imf2.phe, genotype = imf2.geno, population="f2")
head(qtl.res.4)
```

```
##    Bin   Chr Start  Stop p -logp
## 1 Bin1 chr01 0.000 0.565 1     0
## 2 Bin2 chr01 0.565 0.599 1     0
## 3 Bin3 chr01 0.599 0.922 1     0
## 4 Bin4 chr01 0.922 1.075 1     0
## 5 Bin5 chr01 1.075 1.147 1     0
## 6 Bin6 chr01 1.147 1.221 1     0
```

3.5 Comparison of different algotrithms

Let's get a glimpse of the QTL mapping results of different algorithms.

```
par(mar = c(3, 4, 2, 1))
par(mfrow=c(2, 1))
plotQTL(qtl.res.3[, c(1:4, 6)], main="ANOVA", cex.axis=0.6, ylab=expression(-log[10](p)))
plotQTL(qtl.res.4[, c(1:4, 6)], main="binQTL", cex.axis=0.6, ylab=expression(-log[10](p)))
```

4 Session Information

The version number of R and packages loaded for generating the vignette were:

```
sessionInfo()
```

```
## R version 3.3.3 (2017-03-06)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 14393)
##
## locale:
```

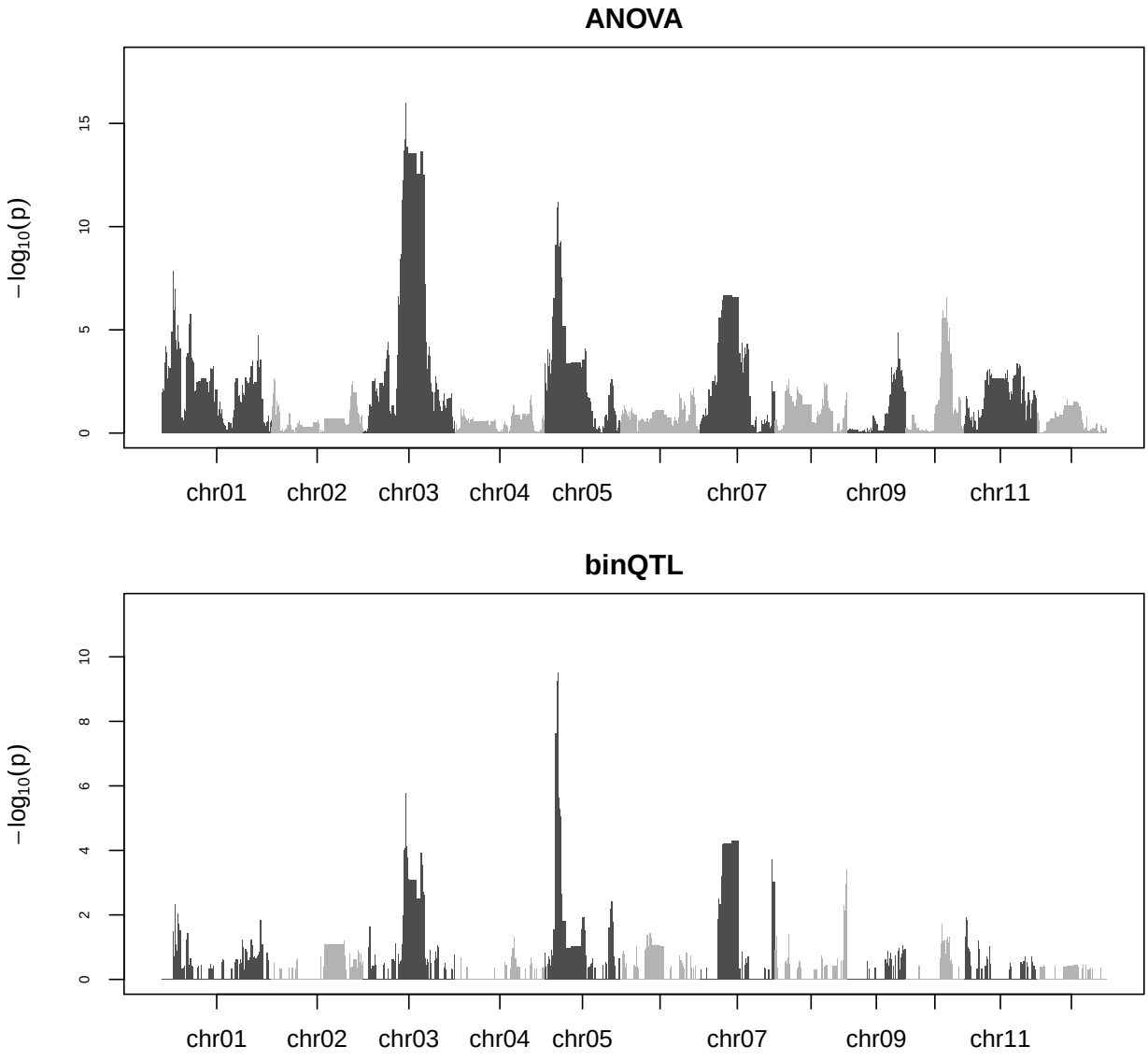

Figure 4: QTL mapping of 278 IMF2s using different algorithms.

```
## [1] LC_COLLATE=C
## [2] LC_CTYPE=Chinese (Simplified)_China.936
## [3] LC_MONETARY=Chinese (Simplified)_China.936
## [4] LC_NUMERIC=C
## [5] LC_TIME=Chinese (Simplified)_China.936
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] binQTL_0.1.0 MASS_7.3-47
##
## loaded via a namespace (and not attached):
##  [1] backports_1.0.5 magrittr_1.5    rprojroot_1.2   htmltools_0.3.5
##  [5] tools_3.3.3     yaml_2.1.14     Rcpp_0.12.10    rmarkdown_1.5
##  [9] stringi_1.1.5   highr_0.6       knitr_1.15.1    digest_0.6.12
## [13] stringr_1.2.0   evaluate_0.10
```