Data Analysis

Azam et al.

Historical dataset analysis

```
library(asreml)
library(ggplot2)
library(ggpubr)
library(dplyr)
library(ggrepel)
```

Analysis of coefficient of variation

```
historical.pheno <- read.csv("FileS1.csv")
historical.pheno$id <- factor(historical.pheno$id)
historical.pheno$source <- factor(historical.pheno$source)

ans <- asreml(fixed=resistance.score~source,random=~id,data=historical.pheno)
stdev <- tapply(residuals(ans),historical.pheno$source,sd)

pred.ans <- predict(ans,classify="source")
ix <- match(names(stdev),pred.ans$pvals$source)
CV <- data.frame(CV=stdev/pred.ans$pvals$predicted.value[ix])

CV$Source <- pred.ans$pvals$source

TableS2 <- CV %>%
  arrange(desc(CV)) %>%
  relocate(Source, .before = 1)

kable(TableS2, row.names = FALSE)
```

Source	CV
NEIKER	0.2205701
Germicopa_netted_scab	0.1796829
Danespo	0.1758716
Potato_Research_Institute	0.1666340
Nordic_Genetic_Resource_Center	0.1608021
University_of_Pannonia	0.1543021
Semena	0.1534316
SASA	0.1509189

Source	CV
Vavilov_All-Russian	0.1485185
The_french_seed_potato	0.1482652
Germicopa_pitted_scab	0.1451225
INRA	0.1443325
IPK	0.1415453
Allotment_&_Gardens	0.1392023
AHDB	0.1368218
Henry_Doubleday_(HDRA)	0.1366739
Binst	0.1322782
GRIN_Czech	0.1310180
Agrico_Research	0.1284308
IHAR	0.1279433
Stet_2022-23	0.1253357
Canadian	0.1233105
Europlant	0.1128507
Parkland	0.1098207
Semagri	0.1092470
Dick_Bedlington_Farms	0.1066810
Kweekinstituut_Karna	0.1050922
Gopex_pitted_scab	0.1049304
Solana	0.1013507
Geniteurslijst_1954-85	0.1009125
Eurogrow	0.1002566
Bavaria_Saat	0.1001047
HZPC_Holland	0.0973861
Averis_2019-23	0.0972321
IPM	0.0935994
Meijer_Potato	0.0922875
NIVAP	0.0917549
Agroplant	0.0876380
Den_Hartigh_BV_website	0.0821312
Agrico_Research_website	0.0794331
De_Ninjs_potatoes	0.0794169
North_American	0.0725380
Meijer_Potato_website	0.0688263
HZPC_Holland_website	0.0683691
Gopex_netted_scab	0.0680887
Beschrijvende_Rassenlijst	0.0680750
Nederlandse_catalogus_1994	0.0679780
MIM	0.0624970
Cebeco_Seed_Innovations	0.0595409
Interseed_Potatoes_Website	0.0587424
The_potato_company	0.0522555
Fobek_BV	0.0401649

```
write.csv(TableS2, "TableS2.csv")
```

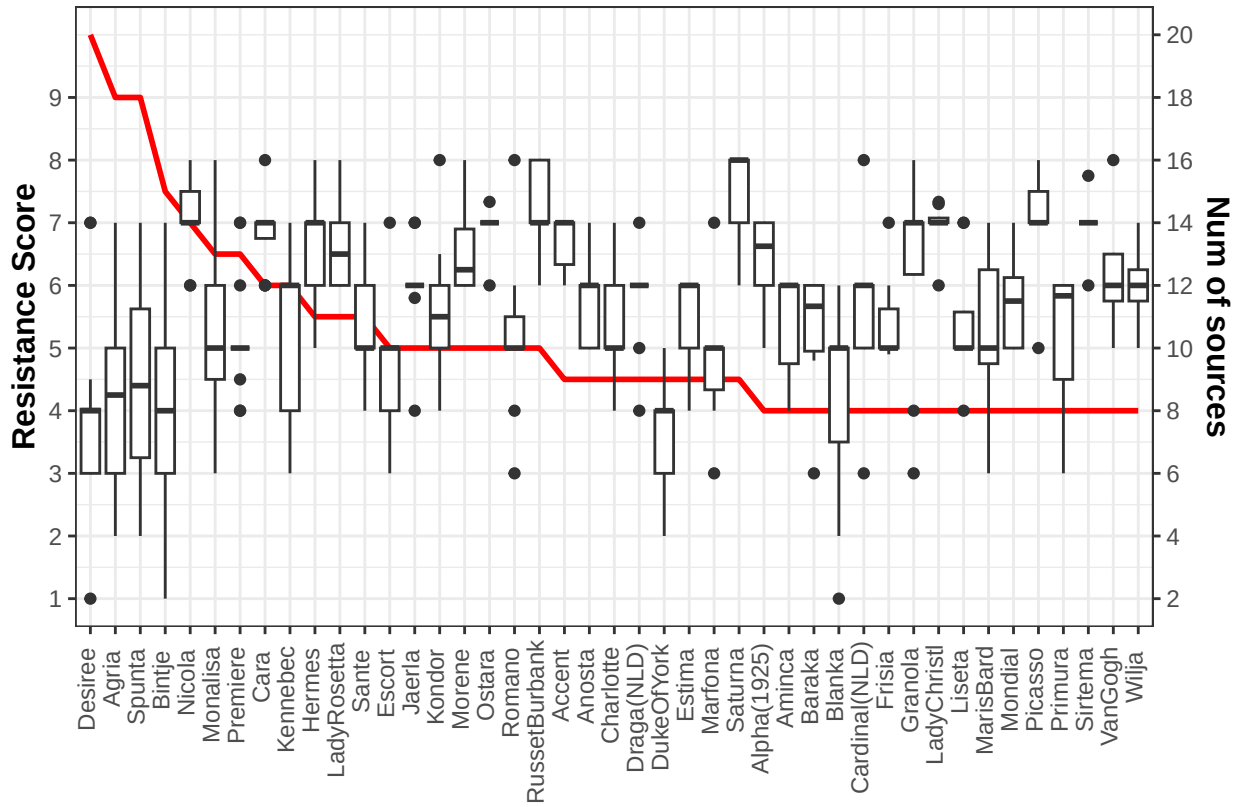
Consistency of resistance scores across different sources

```
freq_table <- historical.pheno %>%
  group_by(id) %>%
  filter(n() >= 8) %>%
  summarise(Frequency = n()) %>%
  arrange(desc(Frequency))

length(freq_table$id)
```

```
## [1] 43
```

We chose to take varieties that were replicated at least across 8 different sources.

```
merge_hist <- historical.pheno %>%
  filter(id %in% freq_table$id) %>%
  left_join(freq_table, by = "id")

coeff <- 2

ggplot(data = merge_hist,aes(x= reorder(id, -Frequency), y=resistance.score)) +
        geom_line(aes(x=reorder(id, -Frequency), y=Frequency/coeff, group = 1),
                  color = "red", linewidth = 1)+ geom_boxplot()+
        scale_y_continuous(name = "Resistance Score", sec.axis = sec_axis(~.*coeff,
                           name = "Num of sources",breaks = c(seq(0, 20, by = 2))),
                           limits = c(1, 10), breaks = c(seq(1, 9, by = 1))) + theme_bw() + theme(axis.
        axis.title=element_text(size=12,face="bold"))  + xlab("") +
        theme(legend.position="none")
```




```
ggsave("FigureS1.png",device="png",dpi=600,height=4,width=7)
```

Population structure

```
data <- read.csv("FileS5.csv",check.names=F)
geno <- as.matrix(data[,-(1:3)])
dim(geno)
```

```
## [1] 6133  707
```

```
rownames(geno) <- data$marker
id <- colnames(geno)

x <- scale(t(geno), scale=F, center=T)
x[is.na(x)] <- 0 #impute missing data with allele freq

pca <- prcomp(x, scale=F, center=T)
pca$sdev[1:2]^2/sum(pca$sdev^2) #PVE
```

```
## [1] 0.10268269 0.02825351
```

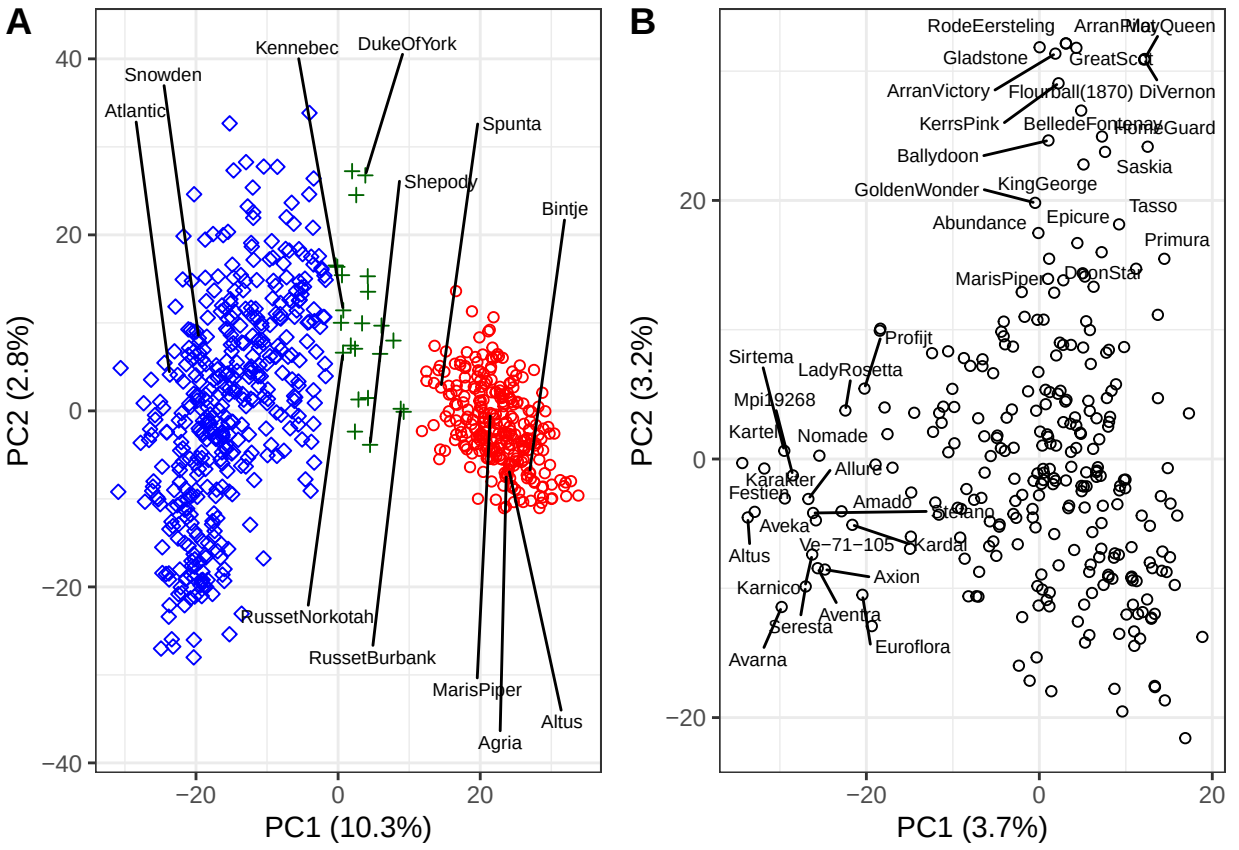
```
tmp <-as.integer(cut(pca$x[,1],breaks=c(-100,-1,10,100)))
pdat <- data.frame(id=id, x=pca$x[id,1], y=pca$x[id,2],
                   group=factor(c("US chip","Other","European")[tmp]))
pdat2 <- filter(pdat,id %in%
        c("Snowden","Atlantic","Kennebec","MarisPiper","DukeOfYork",
          "RussetNorkotah","RussetBurbank","Shepody","Altus","Agria",
          "Spunta","Bintje"))
pdat2 <- pdat2[order(pdat2$x),]

p1 <- ggplot(pdat,aes(x=x,y=y,col=group,shape=group)) + geom_point() +
  scale_shape_manual(values=c(1,3,5)) +
  xlab("PC1 (10.3%)") + ylab("PC2 (2.8%)") +
  scale_color_manual(values=c("red","darkgreen","blue")) + theme_bw() +
  geom_text_repel(data=pdat2,aes(x=x,y=y,label=id),colour="black",size=2.5,
                  nudge_x =c(-5,-5,-5,-5,5,10,0,10,-5,0,5,5),
                  nudge_y=c(30,30,-30,30,15,30,-30,30,-30,-30,-30,30))

ggsave("FigureS2.png",device="png",dpi=600,height=4,width=7)
```

The first PC separated the US chip and European germplasm groups, with North American russets and some historical varieties in between ("Other").

```
x2 <- x[pdat$id[pdat$group=="European"],]
pca2 <- prcomp(x2,scale=F,center=T)
pca2$sdev[1:2]^2/sum(pca2$sdev^2) #3.7%
```

```
## [1] 0.03679193 0.03238060
```

```
id2 <- rownames(pca2$x)

pdat3 <- data.frame(id=id2,x=pca2$x[id2,1],y=pca2$x[id2,2])
pdat4 <- filter(pdat3,x < -20 | y > 15)
pdat4 <- pdat4[order(pdat4$x),]
p2 <- ggplot(pdat3,aes(x=x,y=y)) + geom_point(shape=1) +
  xlab("PC1 (3.7%)") + ylab("PC2 (3.2%)") + theme_bw() +
  geom_text_repel(data=pdat4,aes(x=x,y=y,label=id),colour="black",size=2.5,max.overlaps=15)

p1 <- p1 + guides(color="none",shape="none")
ggarrange(p1,p2,ncol=2,labels=c("A","B"))
```




```
ggsave("Figure2.png",device="png",dpi=600,height=4,width=8)
```

For a PCA within the European group, the first PC separated out varieties for the starch processing market, and the second PC separated varieties from the UK.

The three groups in Fig. 2A were used to create a structured genomic relationship matrix in StageWise:

```
write.csv(data.frame(id=pdat$id, pop=pdat$group),"popfile.csv",row.names=F)

library(StageWise)
geno <- read_geno(filename="FileS5.csv", ploidy=4L, map=T, dominance=T,
                  pop.file="popfile.csv")
```

```
## Minor allele threshold = 5 genotypes
## Number of markers = 6116
## Number of genotypes = 707
```

StageWise analysis

Stage1 analysis by environment

Historical dataset

```
historical <- Stage1("FileS1.csv",solver="asreml",trait="resistance.score",
                     workspace = c("3gb","3gb"),
                     effects=data.frame(name="source",factor=T,fixed=F))
```

```
historical$fit
```

```
##          env H2.plot H2.entry    AIC
## 1 historical   0.566    0.658 4045.5
```

```
write.csv(historical$blues, "BLUEs_historical.csv")
```

Contemporary US dataset

```
US.pheno <- read.csv("FileS2.csv")
env <- unique(US.pheno$env)

blues <- NULL
vcov <- vector("list",6)
names(vcov) <- env
H2.entry <- numeric(6)
names(H2.entry) <- env

ans1 <- Stage1(filename="FileS2.csv",solver="asreml",env="Lakeview23",
               effects=data.frame(name="block",fixed=F,factor=T),
               traits="resistance.score")
```

```
## Warning in stat_boxplot(outlier.color = "red"): Ignoring unknown parameters:
## `outlier.colour`
```

```
blues <- rbind(blues,ans1$blues)
vcov[[1]] <- ans1$vcov[[1]]
H2.entry[1] <- ans1$fit$H2.entry

ans1 <- Stage1(filename="FileS2.csv",solver="asreml",env="Lakeview24",
               traits="resistance.score")
```

```
## Warning in stat_boxplot(outlier.color = "red"): Ignoring unknown parameters:
## `outlier.colour`
```

```
blues <- rbind(blues,ans1$blues)
vcov[[2]] <- ans1$vcov[[1]]
H2.entry[2] <- ans1$fit$H2.entry

ans1 <- Stage1(filename="FileS2.csv",solver="spats",
               env=c("Antigo24","Hancock21","Hancock22"),
               spline=c("row","range"),
               traits="resistance.score")
```

```
## Warning in stat_boxplot(outlier.color = "red"): Ignoring unknown parameters:
## `outlier.colour`
```

```
blues <- rbind(blues,ans1$blues)
vcov[3:5] <- ans1$vcov
H2.entry[3:5] <- ans1$fit$H2.entry

ans1 <- Stage1(filename="FileS2.csv",solver="spats",env="Hancock23",
               effects=data.frame(name="block",fixed=F,factor=T),
               spline=c("row","range"),
               traits="resistance.score")
```

```
## Warning in stat_boxplot(outlier.color = "red"): Ignoring unknown parameters:
## 'outlier.colour'
```

```
blues <- rbind(blues,ans1$blues)
vcov[[6]] <- ans1$vcov[[1]]
H2.entry[6] <- ans1$fit$H2.entry

H2.entry
```

```
## Lakeview23 Lakeview24   Antigo24  Hancock21  Hancock22  Hancock23
##      0.815      0.751      0.660      0.550      0.590      0.530
```

```
summary(H2.entry)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.5300  0.5600  0.6250  0.6493  0.7282  0.8150
```

```
write.csv(blues, "BLUEs_contemporary_US.csv")
```

Stage2 analysis

The US and historical datasets were treated as separate "locations" in StageWise to estimate genetic and residual variances separately, as well as genetic covariance.

```
stage1.blues <- rbind(data.frame(blues,loc="US"),
                      data.frame(historical$blues,loc="historical"))
stage1.vcov <- c(vcov, historical=historical$vcov[[1]])

write.csv(stage1.blues[,1:3], "BLUEs_combined.csv")
```

```
ans2 <- Stage2(data=stage1.blues, vcov=stage1.vcov, geno=geno,
               non.add="g.resid", workspace="3gb")
summary(ans2$vars,separate.loc=T)
```

```
## $var
##              historical    US
## env               0.000 0.279
## additive          0.132 0.201
## g.resid           0.000 0.132
## g x env           0.776 0.212
```

```
## Stage1.error      0.363 0.289
##
## $PVE
##              historical    US
## additive          0.104 0.240
## g.resid           0.000 0.158
## g x env           0.610 0.255
## Stage1.error      0.286 0.346
##
## $cor.mat
##            historical    US
## historical      1.000 0.999
## US              0.999 1.000
```

```
ans2$params$vc
```

```
##                          name     estimate         SE
## 1        additive:loc!loc!cor 9.990040e-01         NA
## 2 additive:loc!loc_historical 1.414577e-01 0.11351468
## 3         additive:loc!loc_US 2.150728e-01 0.06529161
## 4       id:loc!loc_historical 1.702386e-07         NA
## 5                id:loc!loc_US 1.325492e-01 0.05084460
## 6     loc_historical!residual 7.790112e-01 0.12831802
## 7              loc_US!residual 2.127897e-01 0.03490236
```

```
ans2$aic
```

```
## [1] 874.7593
```

```
ans2d <- Stage2(data=stage1.blues, vcov=stage1.vcov, geno=geno,
                non.add="dom", workspace="3gb")
```

```
## Warning in asreml(data = data, fixed = BLUE ~ dom:loc + env - 1, random =
## ~vm(id, : Some components changed by more than 1% on the last iteration
```

```
summary(ans2d$vars,separate.loc=T)
```

```
## $var
##              historical    US
## env               0.000 0.299
## additive          0.120 0.281
## dominance         0.071 0.057
## heterosis         0.000 0.012
## g x env           0.714 0.239
## Stage1.error      0.363 0.289
##
## $PVE
##              historical    US
## additive          0.094 0.320
## dominance         0.056 0.065
## heterosis         0.000 0.014
```

```
## g x env           0.563 0.272
## Stage1.error      0.287 0.329
##
## $cor.mat
##            historical    US
## historical      1.000 0.949
## US              0.949 1.000
```

```
ans2d$params$vc
```

```
##                           name   estimate         SE
## 1         additive:loc!loc!cor 0.94860855 0.53222292
## 2  additive:loc!loc_historical 0.12833630 0.11274672
## 3          additive:loc!loc_US 0.30154295 0.06911668
## 4 dominance:loc!loc_historical 0.07739228 0.11153917
## 5         dominance:loc!loc_US 0.06207967 0.04370238
## 6      loc_historical!residual 0.71621790 0.14887019
## 7              loc_US!residual 0.23967261 0.03505344
```

```
ans2d$aic
```

```
## [1] 875.2585
```

The additive genomic correlation was very close to 1. The model with dominance had essentially the same aic, so the simpler model with a line effect (g.resid) was chosen.

GEBV Reliability

Due to the perfect additive genomic correlation with the g.resid model, there is no difference in BLUPs computed with respect to either dataset ("location"). But there is a difference in GEBV reliability of the three population groups, due to the data quality and relationships.

```
pop <- read.csv("popfile.csv")

prep <- blup_prep(stage1.blues, stage1.vcov, geno, vars=ans2$vars)
GEBV <- blup(prep,geno,what="BV",index.coeff=c("US"=1,"historical"=1))
GEBV$pop <- pop$pop[match(GEBV$id,pop$id)]

ggplot(GEBV,aes(x=pop,y=r2)) + geom_violin(draw_quantiles = 0.5) +
  theme_bw() + xlab("") + ylab("GEBV Reliability")
```

```
## Warning: The ‘draw_quantiles‘ argument of ‘geom_violin()‘ is deprecated as of ggplot2
## 4.0.0.
## i Please use the ‘quantiles.linetype‘ argument instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```


```
ggsave("Figure3.png",device="png",dpi=600,height=5,width=4)
```

GWAS Analysis

Historical dataset

```
library(GWASpoly)

data_hist <- read.GWASpoly(ploidy=4, pheno.file="BLUEs_historical.csv",
                      geno.file="FileS3.csv",format="numeric", n.traits=1, delim=",")
```

```
## Number of polymorphic markers: 14150
## Missing marker data imputed with population mode
## N = 0 individuals with phenotypic and genotypic information
## Detected following fixed effects:
## id
## BLUE
## Detected following traits:
## env
```

```
K_hist <- set.K(data_hist, LOCO = TRUE)

N_hist <- 292

params_hist <- set.params(fixed = "env", fixed.type = "factor",
                          geno.freq=1 - 5/N_hist, P3D=TRUE)
```

```
data.scan_hist <- GWASpoly(K_hist, models = c("additive","1-dom","2-dom"),
                           traits = "BLUE", params = params_hist, n.core = 6)
```

```
M_hist <- set.threshold(data.scan_hist, method = "M.eff", level = 0.05)
```

```
## Thresholds
##      additive 1-dom-alt 1-dom-ref 2-dom-alt 2-dom-ref
## BLUE     5.17      5.13      4.57      5.01      4.83
```

```
QTL_hist <- get.QTL(M_hist, traits = "BLUE", bp.window = NULL)

QTL_hist
```

```
## data frame with 0 columns and 0 rows
```

No significant associations were found with historical dataset.

Contemporary US dataset

```
data_contemp.US <- read.GWASpoly(ploidy=4, pheno.file="BLUEs_contemporary_US.csv",
                      geno.file="FileS4.csv",format="numeric", n.traits=1, delim=",")
```

```
## Number of polymorphic markers: 14795
## N = 0 individuals with phenotypic and genotypic information
## Detected following fixed effects:
## id
## BLUE
## Detected following traits:
## env
```

```
K_contemp.US <- set.K(data_contemp.US, LOCO = TRUE)

N_contemp.US <- 416

params_contemp.US <- set.params(fixed = "env", fixed.type = "factor",
                          geno.freq=1 - 5/N_contemp.US, P3D=TRUE)
```

```
data.scan_contemp.US <- GWASpoly(K_contemp.US, models = c("additive","1-dom","2-dom"),
                           traits = "BLUE", params = params_contemp.US, n.core = 6)
```

```
M_contemp.US <- set.threshold(data.scan_contemp.US, method = "M.eff", level = 0.05)
```

```
## Thresholds
##      additive 1-dom-alt 1-dom-ref 2-dom-alt 2-dom-ref
## BLUE     5.13      5.06      4.78      5.02      4.92
```

```
QTL_contemp.US <- get.QTL(M_contemp.US, traits = "BLUE", bp.window = NULL)

QTL_contemp.US
```

```
## data frame with 0 columns and 0 rows
```

No significant associations were found with contemporary US dataset.

Combined dataset

```
data_comb <- read.GWASpoly(ploidy=4, pheno.file="BLUEs_combined.csv",
                      geno.file="FileS5.csv",format="numeric", n.traits=1, delim=",")
```

```
## Number of polymorphic markers: 6129
## Missing marker data imputed with population mode
## N = 0 individuals with phenotypic and genotypic information
## Detected following fixed effects:
## id
## BLUE
## Detected following traits:
## env
```

```
K_comb <- set.K(data_comb, LOCO = TRUE)

N_comb <- 707

params_comb <- set.params(fixed = "env", fixed.type = "factor",
                          geno.freq=1 - 5/N_comb, P3D=TRUE, n.PC = 1)

# The 1st PC was added as fixed effect factor as it explains 10% of variation (Fig.2A).
```

```
data.scan_comb <- GWASpoly(K_comb, models = c("additive","1-dom","2-dom"),
                           traits = "BLUE", params = params_comb, n.core = 6)
```

```
M_comb <- set.threshold(data.scan_comb, method = "M.eff", level = 0.05)
```

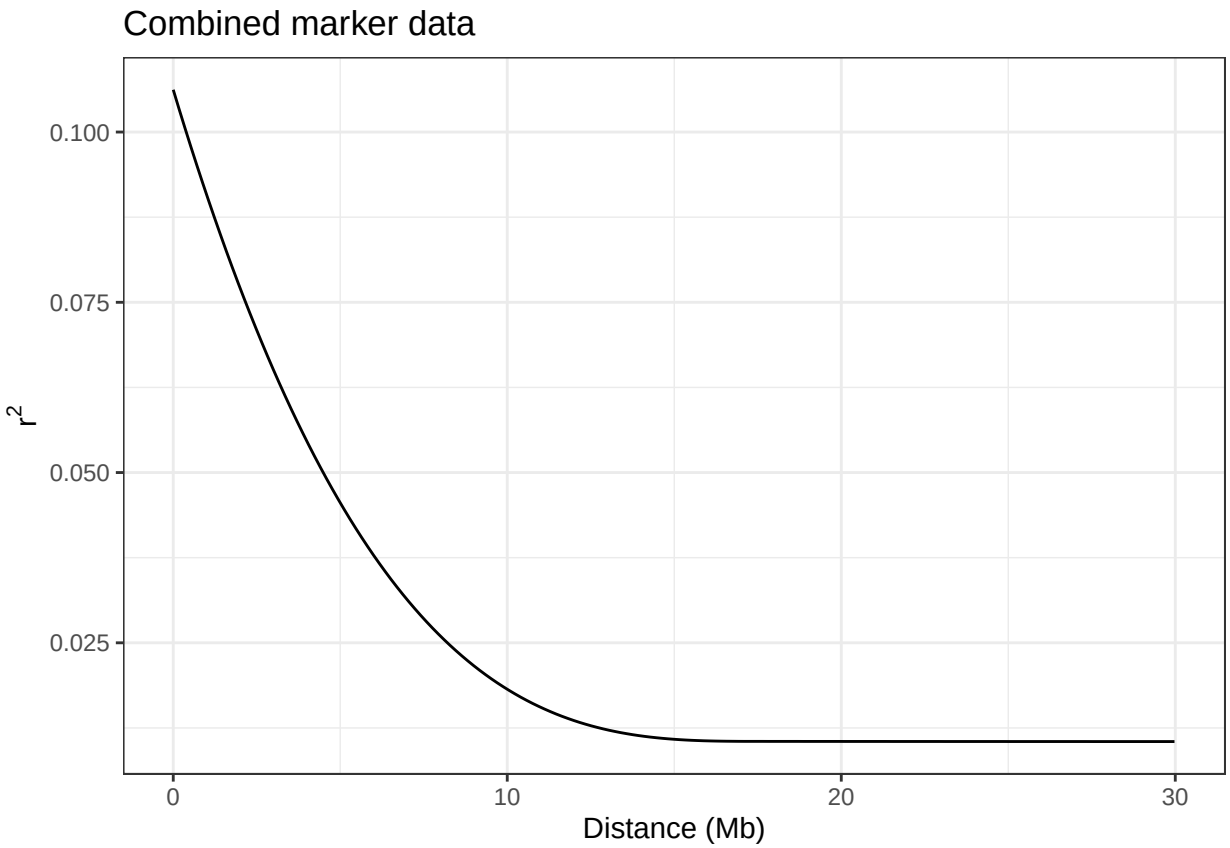
```
## Thresholds
##      additive 1-dom-alt 1-dom-ref 2-dom-alt 2-dom-ref
## BLUE     4.87      4.84       4.5      4.82       4.7
```

```
QTL_comb <- get.QTL(M_comb, traits = "BLUE", bp.window = NULL)

kable(QTL_comb, row.names = FALSE)
```

Trait	Model	Threshold	Marker	Chrom	Position	Ref	Alt	Score	Effect
BLUE	additive	4.87	PotVar0042350	chr01	76374420	0	1	5.34	0.1889764
BLUE	1-dom-alt	4.84	PotVar0041300	chr01	75095240	0	1	5.13	-0.3544253
BLUE	1-dom-ref	4.50	PotVar0099669	chr01	87350385	0	1	4.54	0.3461184
BLUE	2-dom-alt	4.82	PotVar0042350	chr01	76374420	0	1	6.56	0.3747933
BLUE	2-dom-alt	4.82	PotVar0117603	chr02	24247561	0	1	6.36	-0.4749061

```
LD_comb <- LD.plot(M_comb, max.loci = 1000)
LD_comb + xlim(0,30) + ggtitle("Combined marker data")
```

```
## Warning: Removed 329 rows containing missing values or values outside the scale range
## (`geom_line()`).
```




```
ggsave("FigureS3.png",device="png",dpi = 600,height=4,width=5)
```

```
## Warning: Removed 329 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

We chose a bp.window of 5e6 based on the LD plot.

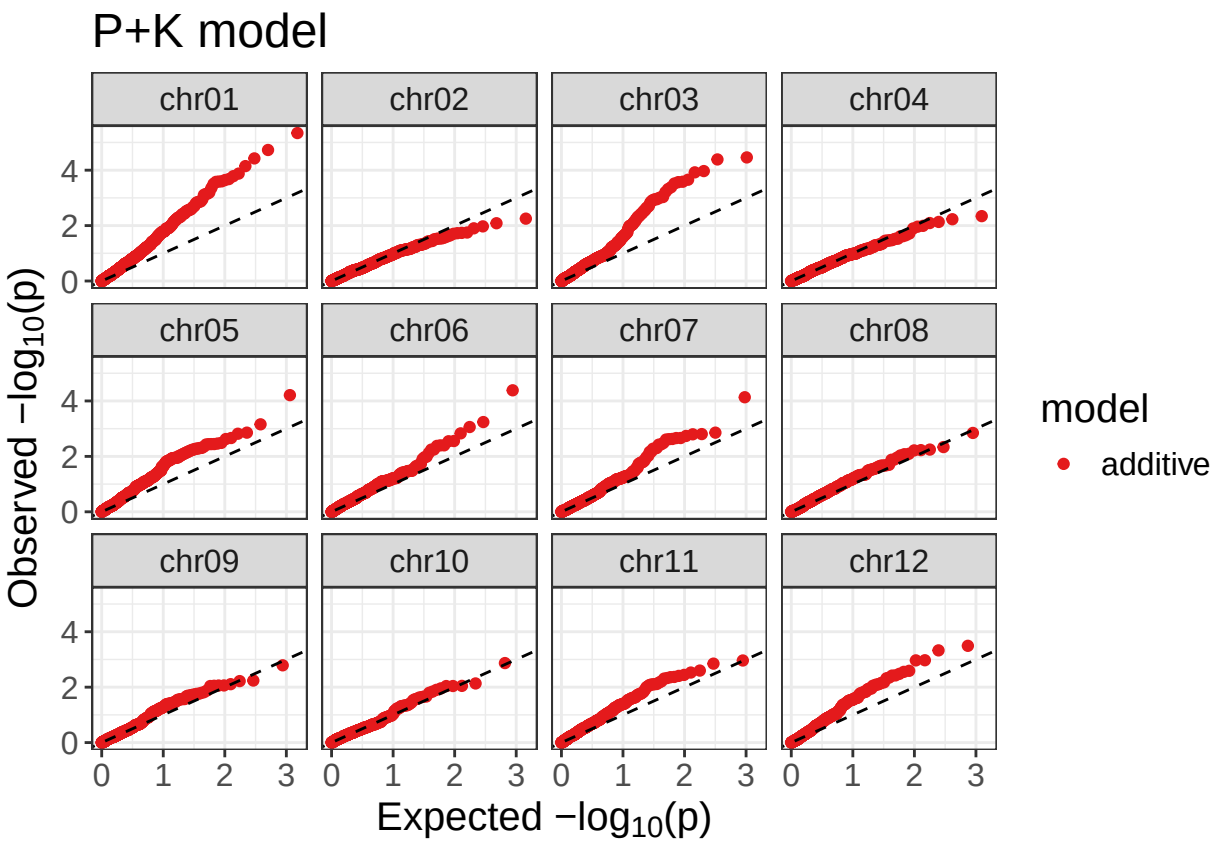
```
QTL_comb <- get.QTL(M_comb, traits = "BLUE", bp.window = 5e6)

kable(QTL_comb, row.names = FALSE)
```

Trait	Model	Threshold	Marker	Chrom	Position	Ref	Alt	Score	Effect
BLUE	additive	4.87	PotVar0042350	chr01	76374420	0	1	5.34	0.1889764
BLUE	1-dom-alt	4.84	PotVar0041300	chr01	75095240	0	1	5.13	-0.3544253
BLUE	1-dom-ref	4.50	PotVar0099669	chr01	87350385	0	1	4.54	0.3461184
BLUE	2-dom-alt	4.82	PotVar0042350	chr01	76374420	0	1	6.56	0.3747933
BLUE	2-dom-alt	4.82	PotVar0117603	chr02	24247561	0	1	6.36	-0.4749061

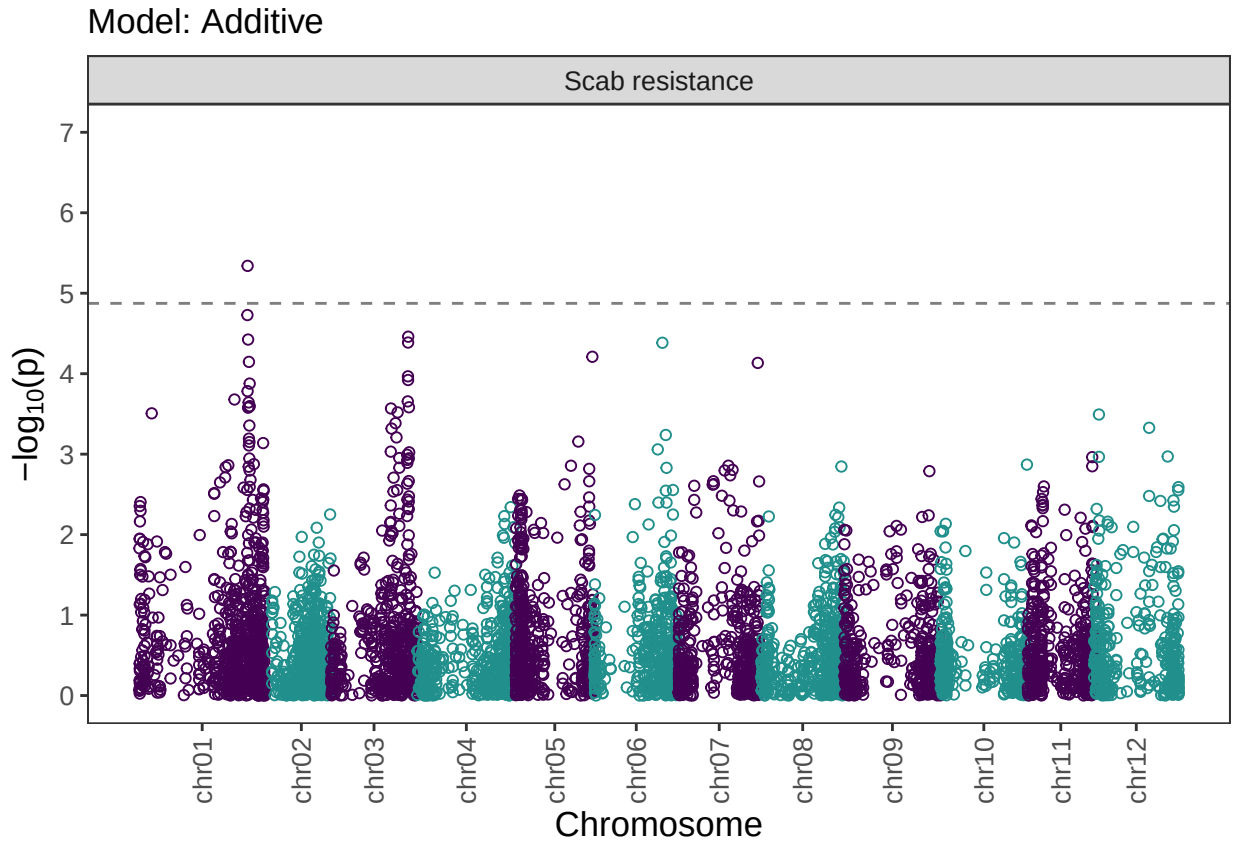
4 unique SNPs were identified with the combined dataset.

QQplots and Manhattan plots - combined dataset

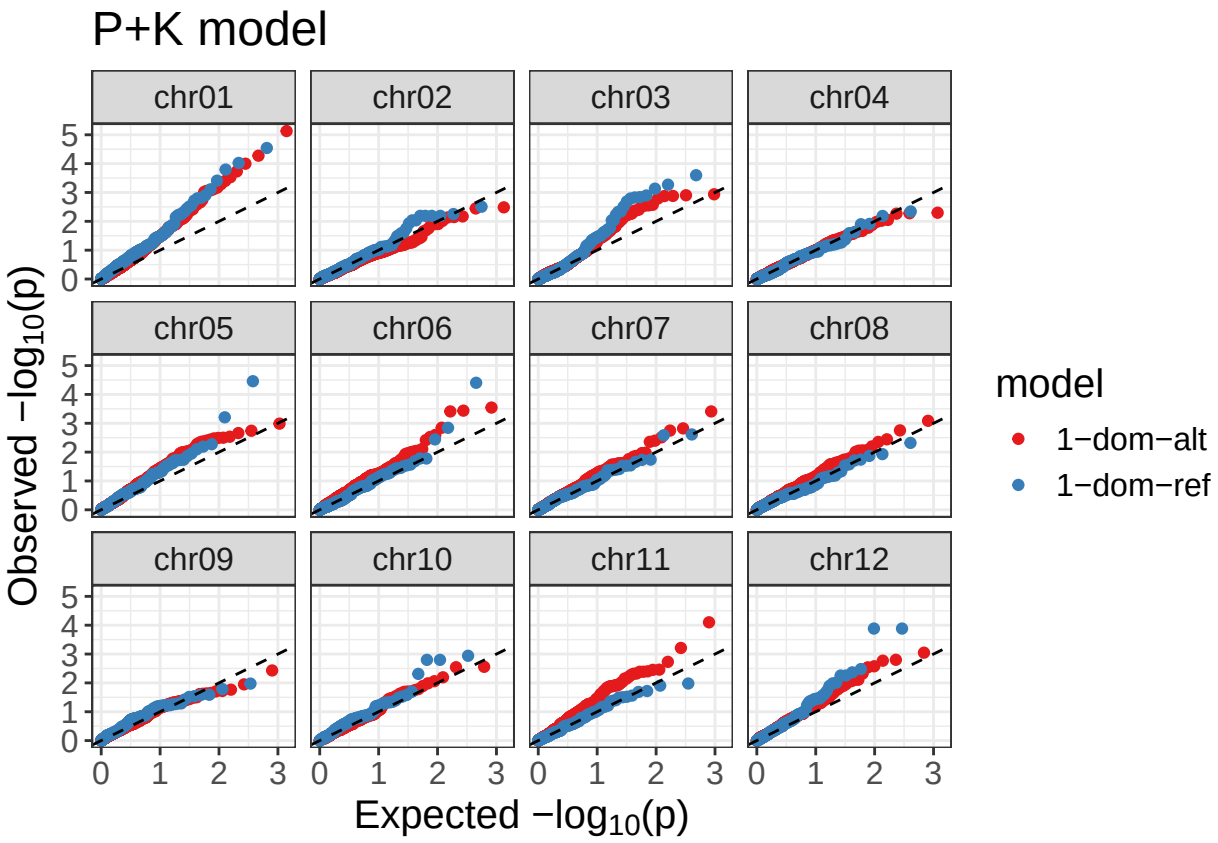
```
# QQplots

q1 <- qq.plot(data.scan_comb, trait="BLUE", models = "additive") +
  ggtitle(label="P+K model")
q2 <- qq.plot(data.scan_comb, trait="BLUE", models = "1-dom") +
  ggtitle(label="P+K model")
q3 <- qq.plot(data.scan_comb, trait="BLUE", models = "2-dom") +
  ggtitle(label="P+K model")

# Manhattan plots

p1 <- manhattan.plot(M_comb, traits = "BLUE", models = "additive") +
  theme(axis.text.x = element_text(angle=90,vjust=0.5),
  text = element_text(size = 13)) + scale_y_continuous(limits = c(0,7),
  breaks = 0:7) + theme(legend.position = "none") + labs(subtitle = "Model: Additive") +
  facet_wrap(~ trait,  labeller = labeller(trait = c("BLUE" = "Scab resistance")))
p2 <- manhattan.plot(M_comb, traits = "BLUE", models = "1-dom") +
  theme(axis.text.x = element_text(angle=90,vjust=0.5),
  text = element_text(size = 13)) + scale_y_continuous(limits = c(0,7),
  breaks = 0:7) + theme(legend.position = "none") + labs(subtitle = "Model: 1-dom") +
  facet_wrap(~ trait,  labeller = labeller(trait = c("BLUE" = "Scab resistance")))
p3 <- manhattan.plot(M_comb, traits = "BLUE", models = "2-dom") +
  theme(axis.text.x = element_text(angle=90,vjust=0.5),
  text = element_text(size = 13)) + scale_y_continuous(limits = c(0,7),
  breaks = 0:7) + theme(legend.position = "none") + labs(subtitle = "Model: 2-dom") +
  facet_wrap(~ trait,  labeller = labeller(trait = c("BLUE" = "Scab resistance")))

# Additive model
q1
```



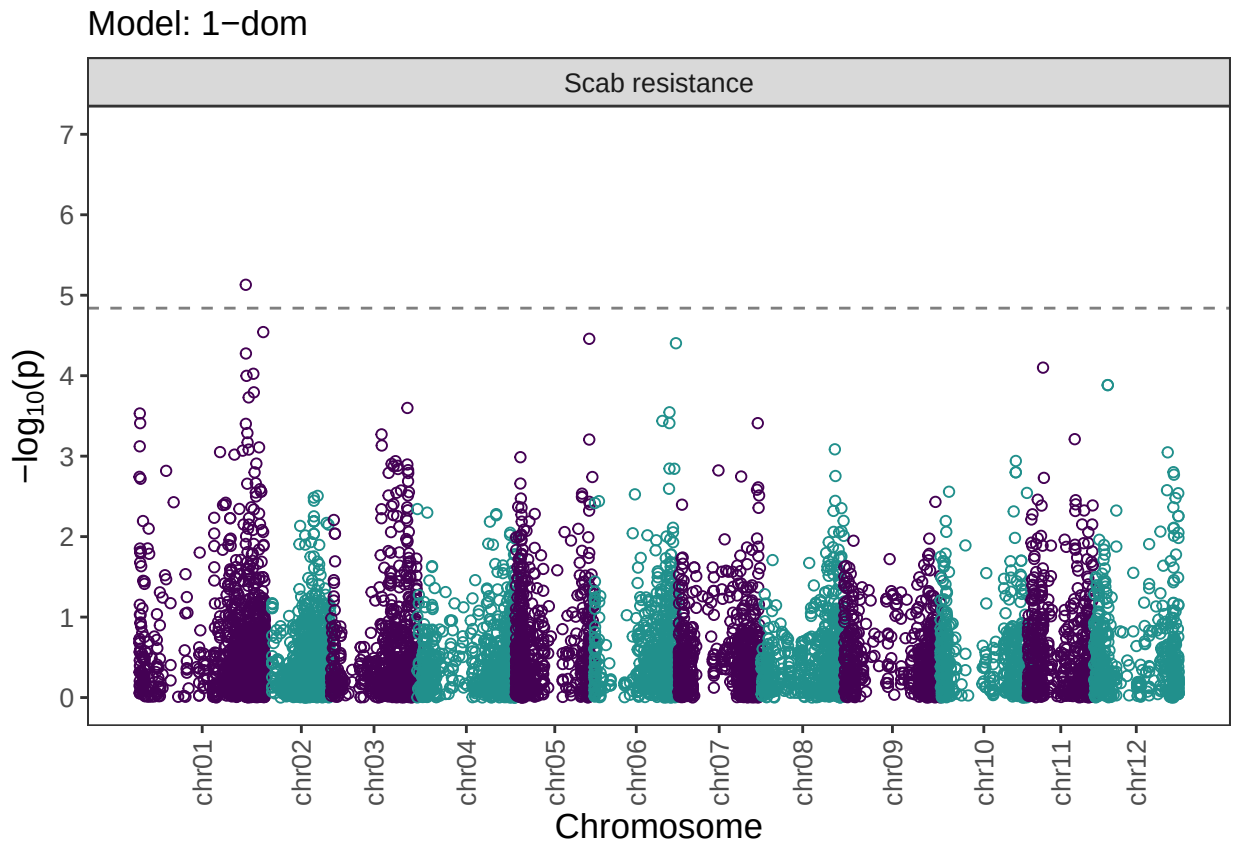

```
ggsave("FigureS4.1.png",device="png",dpi = 600,height=5,width=7)

p1
```



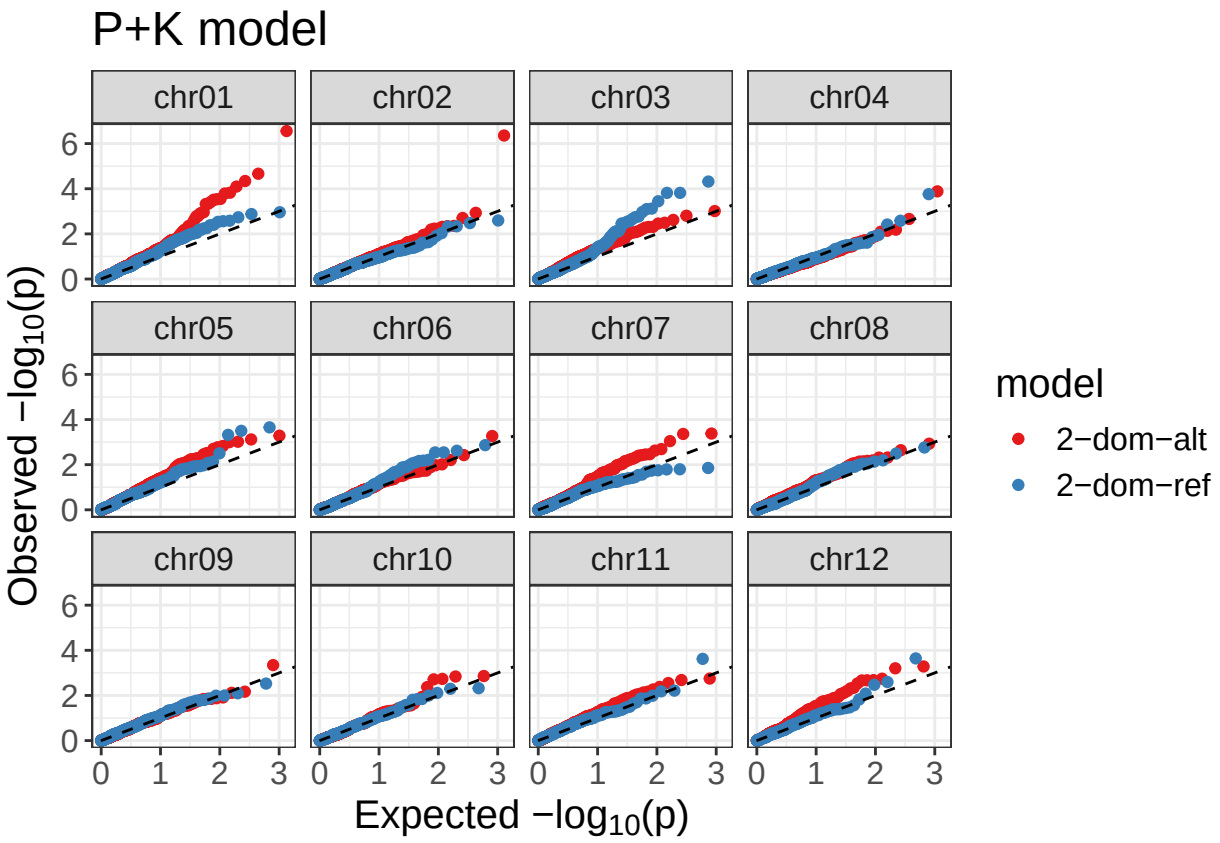

```
# 1-dom model
q2
```



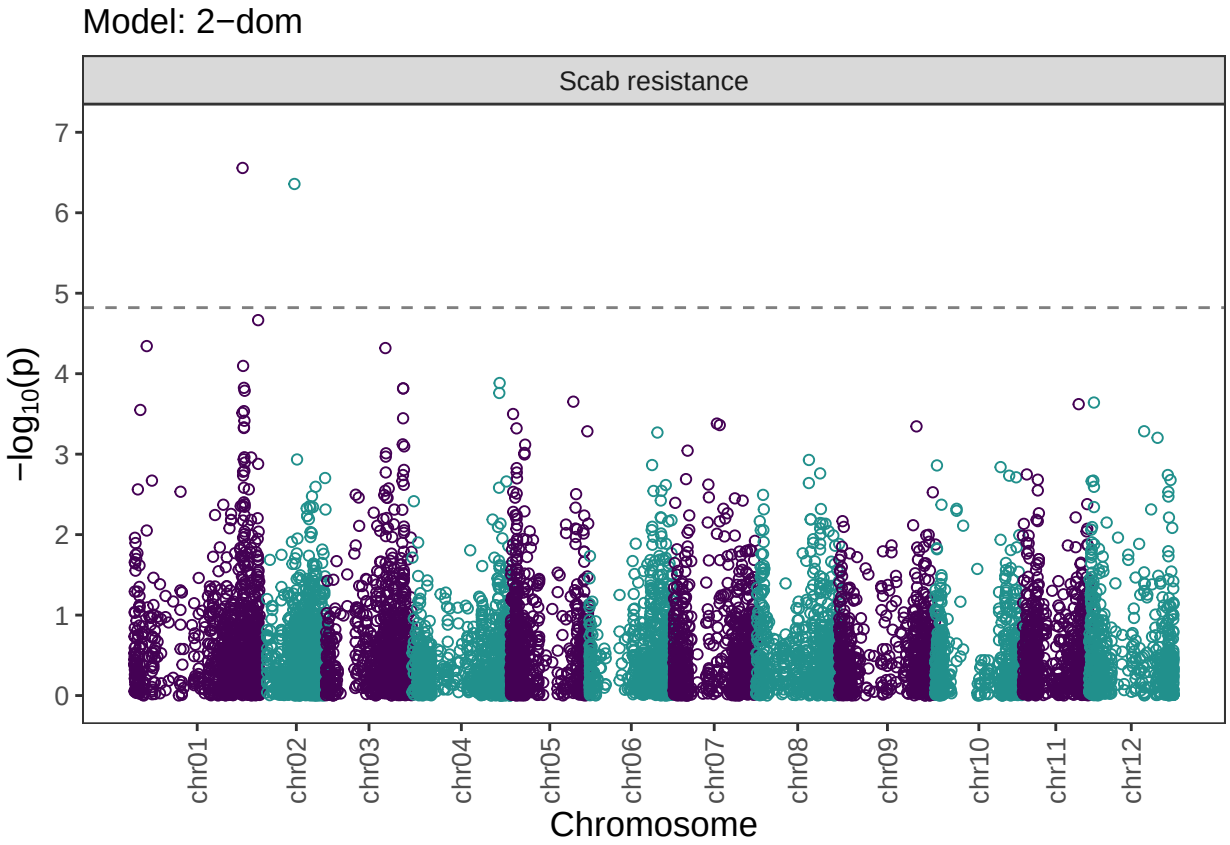

```
ggsave("FigureS4.2.png",device="png",dpi = 600,height=5,width=7)

p2
```




```
# 2-dom model
q3
```




```
ggsave("FigureS4.3.png",device="png",dpi = 600,height=5,width=7)

p3
```




```
Figure4 <- ggarrange(p1,p2,p3,ncol = (3), labels = c("A","B","C"))

ggsave("Figure4.png",device="png",dpi = 600,height=4,width=12)
```

MultiQTL model

Under backwards elimination, markers were removed sequentially from the full model starting with the highest p-value until all markers had a p-value less than or equal to 0.05.

```
fitQTL_comb_full <- fit.QTL(data = M_comb, trait = "BLUE",
                            qtl=QTL_comb[,c("Marker","Model")],
                            fixed=data.frame(Effect="env",Type="factor"))
```

```
kable(fitQTL_comb_full,row.names = FALSE)
```

Marker	Chrom	Position	Model	R2	pval
PotVar0042350	chr01	76374420	additive	0.0001181	0.7320809
PotVar0041300	chr01	75095240	1-dom-alt	0.0157329	0.0000730
PotVar0099669	chr01	87350385	1-dom-ref	0.0165614	0.0000470
PotVar0042350	chr01	76374420	2-dom-alt	0.0052595	0.0221855
PotVar0117603	chr02	24247561	2-dom-alt	0.0223728	0.0000022

```r
# We removed the 1st SNP and re-ran fitQTL.
fitQTL_comb_red <- fit.QTL(data = M_comb,
                           trait = "BLUE",qtl=QTL_comb[-1,c("Marker","Model")],
                           fixed=data.frame(Effect="env",Type="factor"))
```

```r
kable(fitQTL_comb_red, row.names = FALSE)
```

Marker	Chrom	Position	Model	R2	pval
PotVar0041300	chr01	75095240	1-dom-alt	0.0162320	5.60e-05
PotVar0099669	chr01	87350385	1-dom-ref	0.0164770	4.91e-05
PotVar0042350	chr01	76374420	2-dom-alt	0.0197420	8.70e-06
PotVar0117603	chr02	24247561	2-dom-alt	0.0223867	2.10e-06

```r
sum(fitQTL_comb_red$R2)
```

```
## [1] 0.07483785
```

After removing the SNP identified with the additive model, we ended up with 4 SNPs that altogether explain 7.5% of variation.