# F2 adaptations to dominant markers

*Cristiane Hayumi Taniguti - Statistical Genetics Lab - Department of Genetics - Luiz de Queiroz College of Agriculture - University of São Paulo*

OneMap until version 2.2.0 does not adequately deal with dominant markers in F2 populations. The modification was to apply phase estimation to F2 populations using the same background algorithms used for outcrossing populations. This modification was needed to estimate the correct phase between dominant and codominant markers. The modification also allows the estimation of the phase between the heterozygotes progeny in codominant markers. In this new version, we want to provide to users progeny haplotypes, and for that, it's essential to distinguish phase also in F2 intercross.

**Warning**: Because of all the modifications, users can find differences in estimations for this type of population between this and older versions of OneMap.

Here we perform some simulations to show these differences and the need for updates. Two scenarios will be simulated. One with only codominant markers and other with codominant and dominant markers. Three metacentric chromosomes compose the maps with 100 cM and population size of 200 individuals.

- Scenario 1: 150 codominant (A.H.B) markers, 50 per chromosome and no missing data
- Scenario 2: 50 codominant (A.H.B), 50 D.B markers, and 50 C.A markers, 50 markers per chromosome and no missing data

Both scenarios will be evaluated by the old and new approaches.

## Packages

```
library(onemap)
```

## Scenario 1

### Simulation

```
run_pedsim(chromosome = c("Chr1", "Chr02", "Chr03"), n.marker = c(50,50,50),
           tot.size.cm = c(100,100,100), centromere = c(50, 50, 50),
           n.ind = 200, mk.types = c("A.H.B"),
           n.types = c(150), pop = "F2",
           path.pedsim = "/home/cristiane/Programs/PedigreeSim/",
           name.mapfile = "mapfile.map", name.founderfile="founderfile.gen",
           name.chromfile="sim.chrom", name.parfile="sim.par",
           name.out="sim_cod_F2")
```

```
pedsim2raw(cross="f2 intercross", genofile = "sim_cod_F2_genotypes.dat",
           parent1 = "P1", parent2 = "P2", f1 = "F1",
           out.file = "sim_cod_F2.raw", miss.perc = 0)
```

## New approach

```
dataset <- read_onemap("sim_cod_F2.raw")

twopts <- rf_2pts(dataset)
seq1 <- make_seq(twopts, "all")
lod_sug <- suggest_lod(dataset)
lgs <- group(seq1, LOD= lod_sug) # Groups is correct

lg1 <- make_seq(lgs,1)
# Distance estimation
time <- system.time(map1 <- map(lg1))
save.obj <- list(map1, time)
save(save.obj, file = "save.obj4.RData")
```

```
##
## Printing map:
##
## Markers           Position          Parent 1       Parent 2
##
##  1 M001              0.00          a |  | b       a |  | b
##  2 M002              1.76          a |  | b       a |  | b
##  3 M003              4.29          a |  | b       a |  | b
##  4 M004              6.06          a |  | b       a |  | b
##  5 M005              8.60          a |  | b       a |  | b
##  6 M006             12.44          a |  | b       a |  | b
##  7 M007             13.19          a |  | b       a |  | b
##  8 M008             14.70          a |  | b       a |  | b
##  9 M009             16.73          a |  | b       a |  | b
## 10 M010             18.24          a |  | b       a |  | b
## 11 M011             20.26          a |  | b       a |  | b
## 12 M012             22.55          a |  | b       a |  | b
## 13 M013             24.32          a |  | b       a |  | b
## 14 M014             26.08          a |  | b       a |  | b
## 15 M015             27.60          a |  | b       a |  | b
## 16 M016             29.36          a |  | b       a |  | b
## 17 M017             31.39          a |  | b       a |  | b
## 18 M018             33.41          a |  | b       a |  | b
## 19 M019             34.92          a |  | b       a |  | b
## 20 M020             36.68          a |  | b       a |  | b
## 21 M021             38.19          a |  | b       a |  | b
## 22 M022             40.73          a |  | b       a |  | b
## 23 M023             43.00          a |  | b       a |  | b
## 24 M024             45.02          a |  | b       a |  | b
## 25 M025             48.33          a |  | b       a |  | b
## 26 M026             49.34          a |  | b       a |  | b
## 27 M027             51.10          a |  | b       a |  | b
## 28 M028             52.62          a |  | b       a |  | b
## 29 M029             54.13          a |  | b       a |  | b
## 30 M030             55.89          a |  | b       a |  | b
## 31 M031             57.40          a |  | b       a |  | b
## 32 M032             60.45          a |  | b       a |  | b
## 33 M033             61.45          a |  | b       a |  | b
## 34 M034             64.76          a |  | b       a |  | b
```

```
## 35 M035              67.80          a |  | b       a |  | b
## 36 M036              71.37          a |  | b       a |  | b
## 37 M037              72.63          a |  | b       a |  | b
## 38 M038              74.14          a |  | b       a |  | b
## 39 M039              76.93          a |  | b       a |  | b
## 40 M040              78.94          a |  | b       a |  | b
## 41 M041              81.48          a |  | b       a |  | b
## 42 M042              82.73          a |  | b       a |  | b
## 43 M043              85.01          a |  | b       a |  | b
## 44 M044              85.76          a |  | b       a |  | b
## 45 M045              87.02          a |  | b       a |  | b
## 46 M046              88.78          a |  | b       a |  | b
## 47 M047              90.55          a |  | b       a |  | b
## 48 M048              93.08          a |  | b       a |  | b
## 49 M049              93.83          a |  | b       a |  | b
## 50 M050              95.85          a |  | b       a |  | b
##
## 50 markers            log-likelihood: -1920.484
```

```
## Time spent
```

```
##    user  system elapsed
##   9.716   0.004   9.717
```

```
# Ordering
ug1 <- ug(lg1)
rcd1 <- rcd(lg1)
seriation1 <- seriation(lg1)
record1 <- record(lg1)
mds1 <- mds_onemap(lg1)
order1 <- order_seq(lg1)
order1 <- make_seq(order1, "force")

p_ug_cod1 <- rf_graph_table(ug1)
p_rcd_cod1 <- rf_graph_table(rcd1)
p_ser_cod1 <- rf_graph_table(seriation1)
p_rec_cod1 <- rf_graph_table(record1)
p_mds_cod1 <- rf_graph_table(mds1)
p_map_cod1 <- rf_graph_table(map1)
p_order_cod1 <- rf_graph_table(order1)
```

## Old approach

```
dataset <- read_onemap(inputfile = "sim_cod_F2.raw")

twopts <- rf_2pts(dataset)
seq1 <- make_seq(twopts, "all")
lod_sug <- suggest_lod(dataset)
lgs <- group(seq1, LOD= lod_sug) # Group is correct

lg1 <- make_seq(lgs,1)
# Distance estimation
time <- system.time(map1 <- map(lg1))
```

```r
save.obj <- list(map1, time)
save(save.obj, file = "save.obj3.RData")
```

Printing map:

| | Markers | Position |
|---|---|---|
| 1 | M001 | 0.00 |
| 2 | M002 | 1.76 |
| 3 | M003 | 4.29 |
| 4 | M004 | 6.06 |
| 5 | M005 | 8.60 |
| 6 | M006 | 12.44 |
| 7 | M007 | 13.19 |
| 8 | M008 | 14.70 |
| 9 | M009 | 16.73 |
| 10 | M010 | 18.24 |
| 11 | M011 | 20.26 |
| 12 | M012 | 22.55 |
| 13 | M013 | 24.32 |
| 14 | M014 | 26.08 |
| 15 | M015 | 27.60 |
| 16 | M016 | 29.36 |
| 17 | M017 | 31.39 |
| 18 | M018 | 33.41 |
| 19 | M019 | 34.92 |
| 20 | M020 | 36.68 |
| 21 | M021 | 38.19 |
| 22 | M022 | 40.73 |
| 23 | M023 | 43.00 |
| 24 | M024 | 45.02 |
| 25 | M025 | 48.33 |
| 26 | M026 | 49.34 |
| 27 | M027 | 51.10 |
| 28 | M028 | 52.62 |
| 29 | M029 | 54.13 |
| 30 | M030 | 55.89 |
| 31 | M031 | 57.40 |
| 32 | M032 | 60.45 |
| 33 | M033 | 61.45 |
| 34 | M034 | 64.76 |
| 35 | M035 | 67.80 |
| 36 | M036 | 71.37 |
| 37 | M037 | 72.63 |
| 38 | M038 | 74.14 |
| 39 | M039 | 76.93 |
| 40 | M040 | 78.94 |
| 41 | M041 | 81.48 |
| 42 | M042 | 82.73 |
| 43 | M043 | 85.01 |
| 44 | M044 | 85.76 |
| 45 | M045 | 87.02 |
| 46 | M046 | 88.78 |
| 47 | M047 | 90.55 |

```
48 M048              93.08
49 M049              93.83
50 M050              95.85

 50 markers            log-likelihood: -1920.484

Time spent

user  system elapsed
0.016   0.000   0.013
```

```
# Ordering
ug1 <- ug(lg1)
rcd1 <- rcd(lg1)
seriation1 <- seriation(lg1)
record1 <- record(lg1)
mds1 <- mds_onemap(lg1)
order1 <- order_seq(lg1)
order1 <- make_seq(order1, "force")

p_ug_cod_old1 <- rf_graph_table(ug1)
p_rcd_cod_old1 <- rf_graph_table(rcd1)
p_ser_cod_old1 <- rf_graph_table(seriation1)
p_rec_cod_old1 <- rf_graph_table(record1)
p_mds_cod_old1 <- rf_graph_table(mds1)
p_map_cod_old1 <- rf_graph_table(map1)
p_order_cod_old1 <- rf_graph_table(order1)
```

# Scenario 2

## Simulation

```
run_pedsim(chromosome = c("Chr1", "Chr02", "Chr03"), n.marker = c(50,50,50),
           tot.size.cm = c(100,100,100), centromere = c(50, 50, 50),
           n.ind = 200, mk.types = c("A.H.B", "C.A", "D.B"),
           n.types = c(50, 50, 50), pop = "F2",
           path.pedsim = "/home/cristiane/Programs/PedigreeSim/",
           name.mapfile = "mapfile.map", name.founderfile="founderfile.gen",
           name.chromfile="sim.chrom", name.parfile="sim.par",
           name.out="sim_F2")
```

```
pedsim2raw(cross="f2 intercross", genofile = "sim_F2_genotypes.dat",
           parent1 = "P1", parent2 = "P2", f1 = "F1",
           out.file = "sim_F2.raw", miss.perc = 0)
```

## New approach

```
dataset <- read_onemap(inputfile = "sim_F2.raw")

twopts <- rf_2pts(dataset)
seq1 <- make_seq(twopts, "all")
```

```r
lod_sug <- suggest_lod(dataset)
lgs <- group(seq1, LOD= lod_sug) # Do not group correctly

lg1 <- make_seq(twopts,1:50)
# Distance estimation
time <- system.time(map1 <- map(lg1))
save.obj <- list(map1, time)

save(save.obj, file = "save.obj2.RData")
```

```
##
## Printing map:
##
## Markers           Position           Parent 1       Parent 2
##
##  1 M001              0.00          a |  | o       o |  | o
##  2 M002              1.45          a |  | b       a |  | b
##  3 M003              4.62          o |  | o       a |  | o
##  4 M004              8.39          o |  | a       o |  | o
##  5 M005              8.94          o |  | o       a |  | o
##  6 M006             10.50          a |  | b       a |  | b
##  7 M007             11.56          o |  | o       a |  | o
##  8 M008             14.03          a |  | b       a |  | b
##  9 M009             16.86          o |  | o       a |  | o
## 10 M010             19.11          a |  | b       a |  | b
## 11 M011             21.25          a |  | o       o |  | o
## 12 M012             23.91          o |  | o       a |  | o
## 13 M013             24.85          o |  | o       a |  | o
## 14 M014             27.39          o |  | a       o |  | o
## 15 M015             29.00          o |  | a       o |  | o
## 16 M016             31.68          o |  | a       o |  | o
## 17 M017             35.97          o |  | o       a |  | o
## 18 M018             35.98          o |  | a       o |  | o
## 19 M019             40.72          o |  | o       a |  | o
## 20 M020             40.76          o |  | a       o |  | o
## 21 M021             44.54          o |  | o       a |  | o
## 22 M022             46.06          a |  | b       a |  | b
## 23 M023             49.11          o |  | o       a |  | o
## 24 M024             50.10          a |  | b       a |  | b
## 25 M025             52.43          o |  | o       a |  | o
## 26 M026             55.03          o |  | a       o |  | o
## 27 M027             56.34          o |  | a       o |  | o
## 28 M028             59.18          o |  | o       a |  | o
## 29 M029             60.35          a |  | b       a |  | b
## 30 M030             62.85          a |  | b       a |  | b
## 31 M031             63.64          a |  | o       o |  | o
## 32 M032             65.63          a |  | b       a |  | b
## 33 M033             67.38          a |  | b       a |  | b
## 34 M034             70.39          a |  | b       a |  | b
## 35 M035             72.64          a |  | b       a |  | b
## 36 M036             74.14          a |  | b       a |  | b
## 37 M037             74.69          o |  | o       a |  | o
## 38 M038             76.72          o |  | o       a |  | o
## 39 M039             79.27          o |  | o       a |  | o
```

```
## 40 M040              80.53          a |  | b       a |  | b
## 41 M041              83.79          a |  | o       o |  | o
## 42 M042              83.79          o |  | o       a |  | o
## 43 M043              85.23          o |  | o       a |  | o
## 44 M044              89.66          o |  | a       o |  | o
## 45 M045              92.78          o |  | a       o |  | o
## 46 M046              93.75          a |  | b       a |  | b
## 47 M047              94.23          o |  | o       a |  | o
## 48 M048              96.24          o |  | a       o |  | o
## 49 M049              97.73          o |  | o       a |  | o
## 50 M050              99.66          a |  | b       a |  | b
##
## 50 markers            log-likelihood: -1942.969

## Time spent

##    user  system elapsed
##  97.544   0.012  97.540
```

```
# Ordering
ug1 <- ug(lg1)
rcd1 <- rcd(lg1)
#seriation1 <- seriation(lg1) Error: There are
# too many ties in the ordering process - please,
# consider using another ordering algorithm.
record1 <- record(lg1)
mds1 <- mds_onemap(lg1)
order1 <- order_seq(lg1)
order1 <- make_seq(order1, "force")

p_ug1 <- rf_graph_table(ug1)
p_rcd1 <- rf_graph_table(rcd1)
#p_ser1 <- rf_graph_table(seriation1)
p_rec1 <- rf_graph_table(record1)
p_mds1 <- rf_graph_table(mds1)
p_map1 <- rf_graph_table(map1)
p_order1 <- rf_graph_table(order1)
save.image(file = "new_app.RData")
```

## Old approach

```
dataset <- read_onemap(inputfile = "sim_F2.raw")

twopts <- rf_2pts(dataset)
seq1 <- make_seq(twopts, "all")
lod_sug <- suggest_lod(dataset)
lgs <- group(seq1, LOD= lod_sug) # Do not group correctly

lg1 <- make_seq(twopts,1:50)
# Distance estimation
time <- system.time(map1 <- map(lg1)) # Time spent 2.150 sec

save.obj <- list(map1, time)
save(save.obj, file = "save.obj1.RData")
```

Printing map:

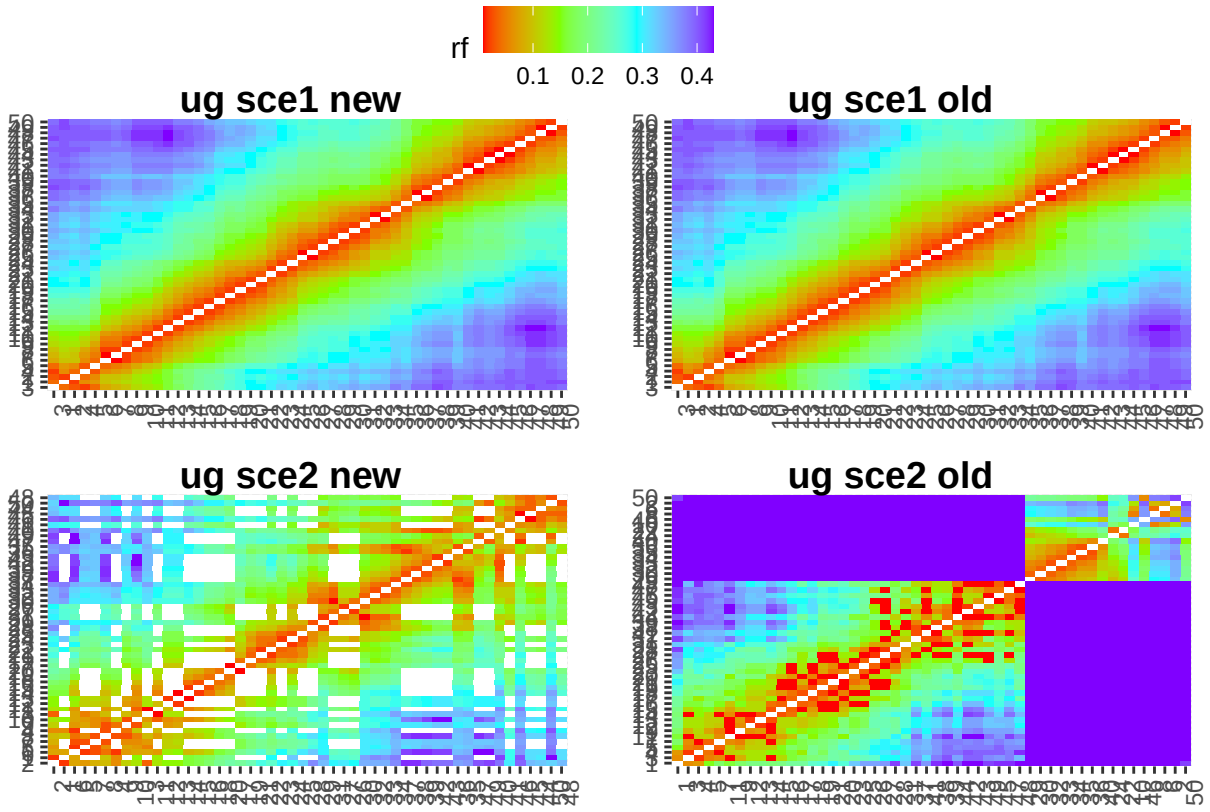
| | Markers | Position |
|---|---|---|
| 1 | M001 | 0.00 |
| 2 | M002 | 402.95 |
| 3 | M003 | 805.90 |
| 4 | M004 | 809.75 |
| 5 | M005 | 810.92 |
| 6 | M006 | 1213.88 |
| 7 | M007 | 1616.83 |
| 8 | M008 | 2019.78 |
| 9 | M009 | 2422.73 |
| 10 | M010 | 2825.69 |
| 11 | M011 | 3228.64 |
| 12 | M012 | 3230.64 |
| 13 | M013 | 3231.55 |
| 14 | M014 | 3235.24 |
| 15 | M015 | 3236.92 |
| 16 | M016 | 3239.69 |
| 17 | M017 | 3243.06 |
| 18 | M018 | 3244.23 |
| 19 | M019 | 3247.88 |
| 20 | M020 | 3249.37 |
| 21 | M021 | 3251.74 |
| 22 | M022 | 3654.69 |
| 23 | M023 | 4057.64 |
| 24 | M024 | 4460.59 |
| 25 | M025 | 4863.55 |
| 26 | M026 | 4867.19 |
| 27 | M027 | 4868.85 |
| 28 | M028 | 4868.90 |
| 29 | M029 | 5271.85 |
| 30 | M030 | 5274.38 |
| 31 | M031 | 5677.33 |
| 32 | M032 | 6080.28 |
| 33 | M033 | 6082.05 |
| 34 | M034 | 6085.09 |
| 35 | M035 | 6087.37 |
| 36 | M036 | 6088.87 |
| 37 | M037 | 6491.83 |
| 38 | M038 | 6493.81 |
| 39 | M039 | 6496.31 |
| 40 | M040 | 6899.26 |
| 41 | M041 | 7302.22 |
| 42 | M042 | 7302.22 |
| 43 | M043 | 7303.68 |
| 44 | M044 | 7308.27 |
| 45 | M045 | 7311.35 |
| 46 | M046 | 7714.31 |
| 47 | M047 | 8117.26 |
| 48 | M048 | 8119.71 |
| 49 | M049 | 8121.23 |
| 50 | M050 | 8524.18 |

```
 50 markers            log-likelihood: -4829.173

Time spent

user  system elapsed
1.780   0.000   1.785
```

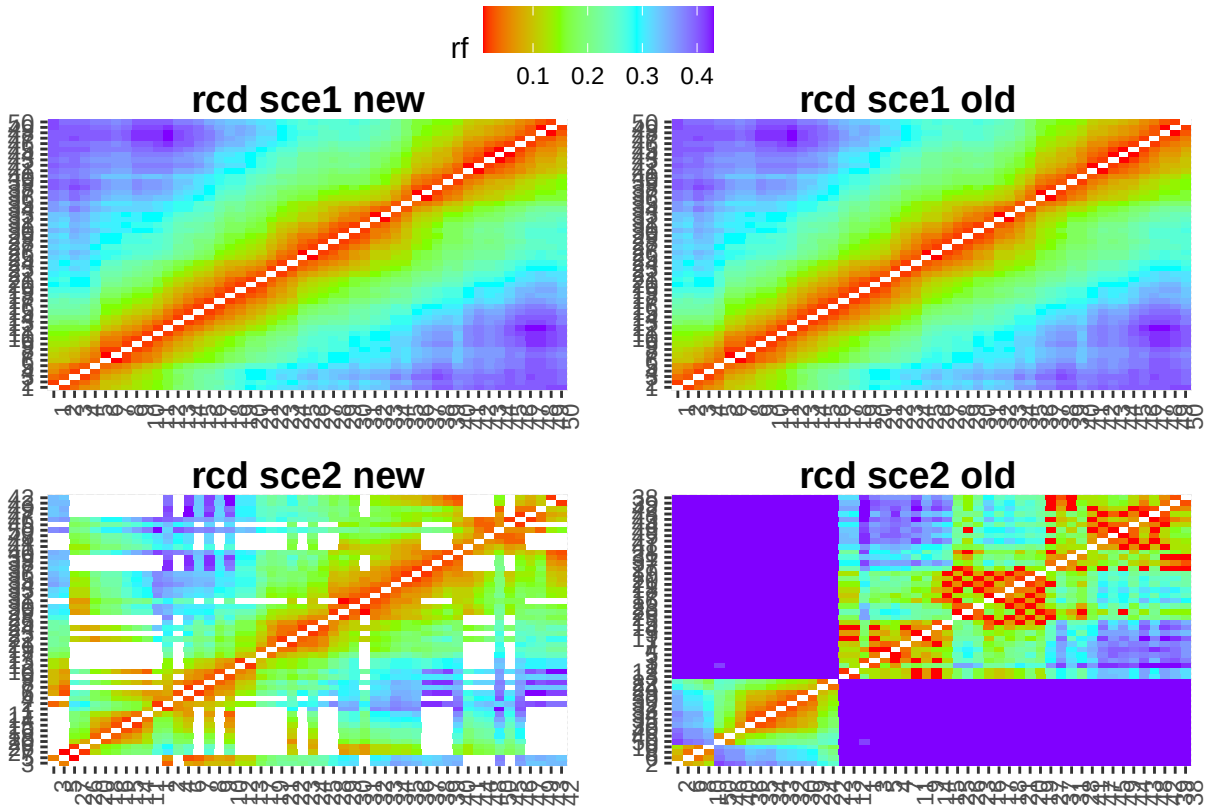
```
# Ordering
ug1 <- ug(lg1)
rcd1 <- rcd(lg1)
#seriation1 <- seriation(lg1) Error: There are
# too many ties in the ordering process - please,
# consider using another ordering algorithm.
record1 <- record(lg1)
mds1 <- mds_onemap(lg1)
order1 <- order_seq(lg1)
order1 <- make_seq(order1, "force")

p_ug_old1 <- rf_graph_table(ug1)
p_rcd_old1 <- rf_graph_table(rcd1)
#p_ser_old1 <- rf_graph_table(seriation1)
p_rec_old1 <- rf_graph_table(record1)
p_mds_old1 <- rf_graph_table(mds1)
p_map_old1 <- rf_graph_table(map1)
p_order_old1 <- rf_graph_table(order1)
save.image(file = "old_app.RData")
```

## Ordering comparision

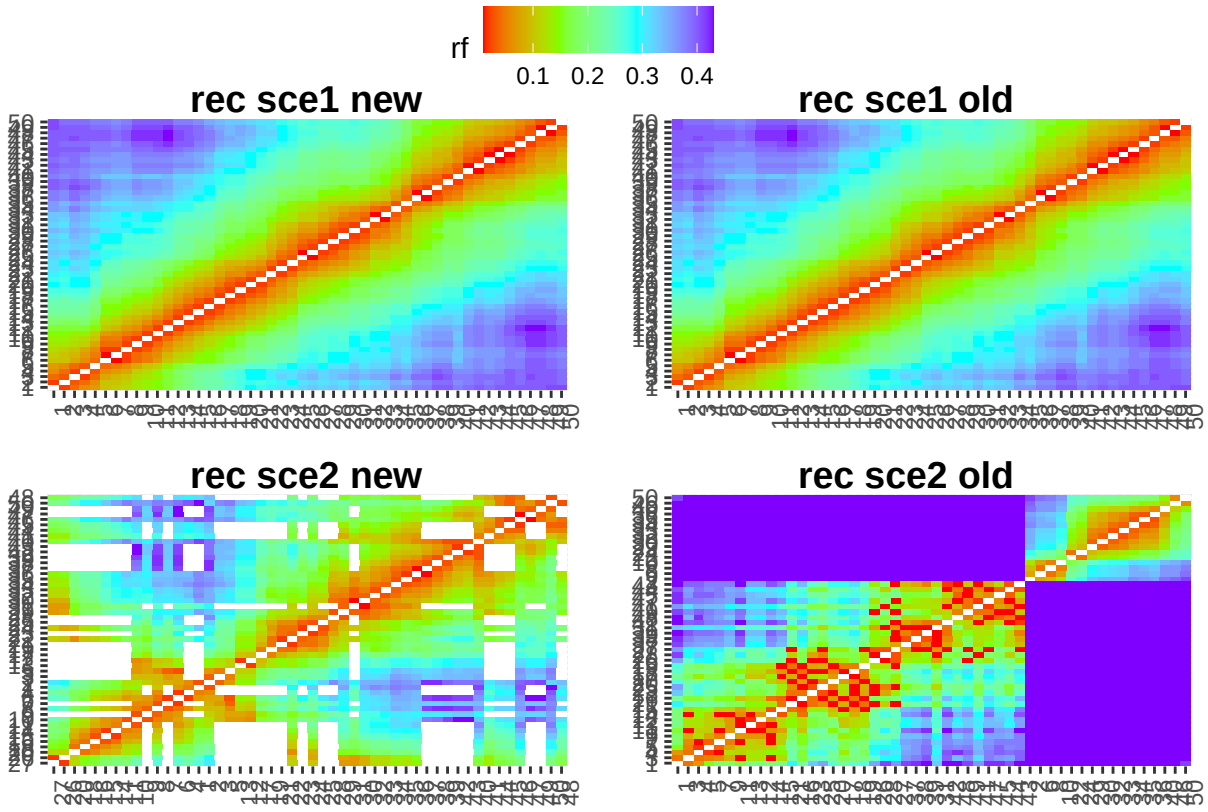
```
library(ggpubr)
load("old_app.RData")
load("new_app.RData")

p <- ggarrange(p_ug_cod1, p_ug_cod_old1, p_ug1,
               p_ug_old1 , common.legend = TRUE,
          labels = c("ug sce1 new", "ug sce1 old",
                     "ug sce2 new", "ug sce2 old"),
          vjust = 0.3,
          hjust= -1,
          ncol=2, nrow=2)
p
```



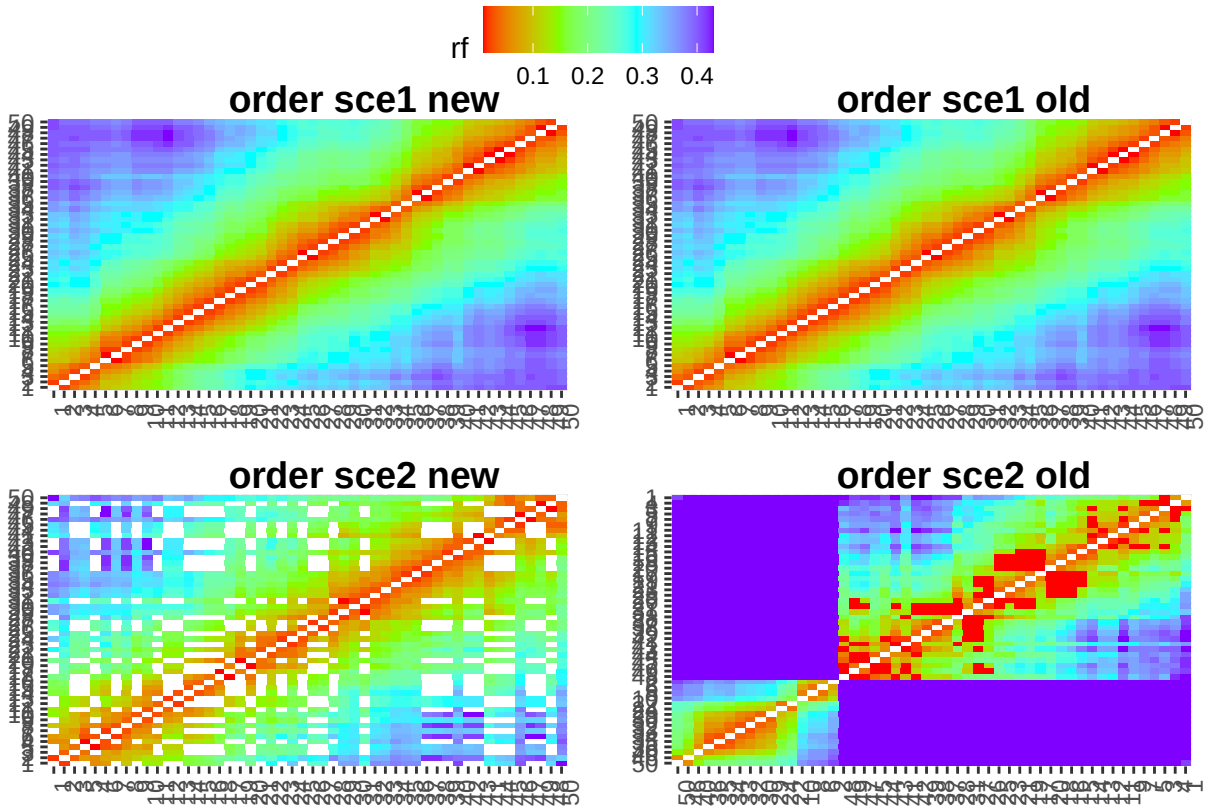

```
p <- ggarrange(p_rcd_cod1, p_rcd_cod_old1, p_rcd1,
               p_rcd_old1 , common.legend = TRUE,
          labels = c("rcd sce1 new", "rcd sce1 old",
                     "rcd sce2 new", "rcd sce2 old"),
          vjust = 0.3,
          hjust= -1,
          ncol=2, nrow=2)
p
```



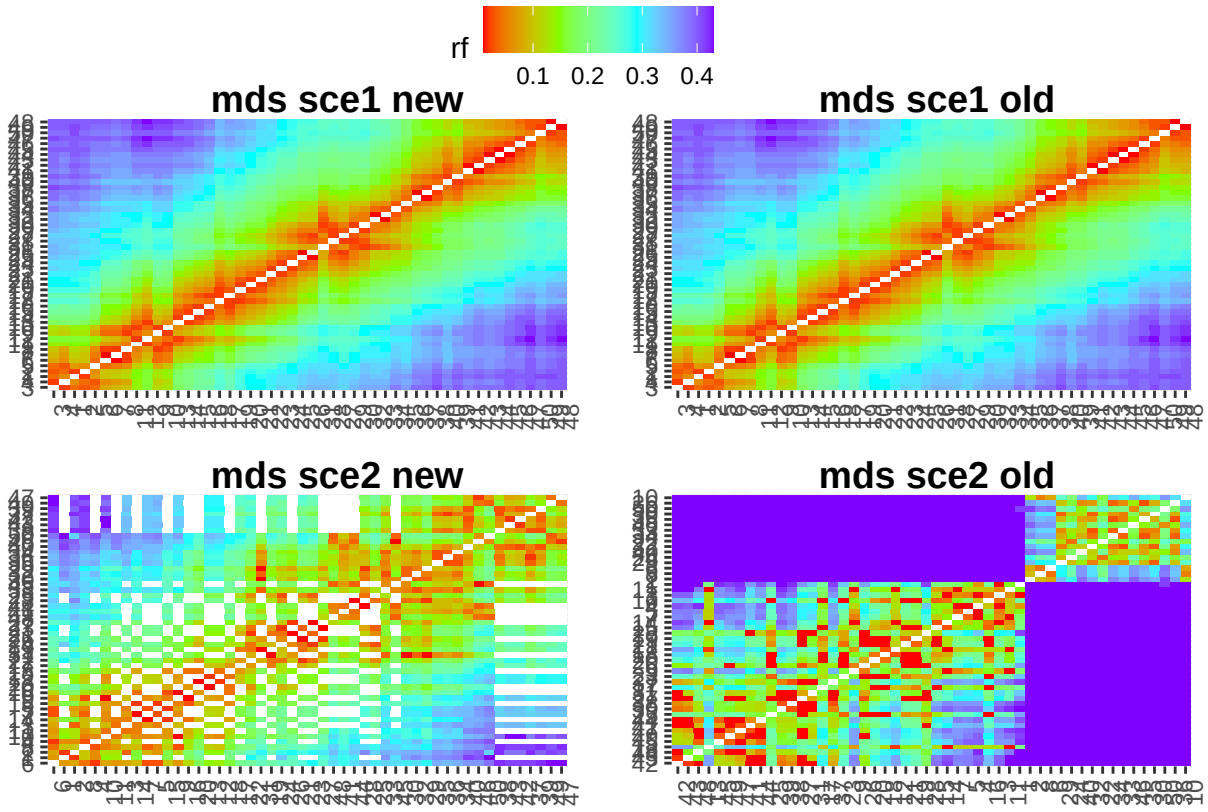

```
p <- ggarrange(p_rec_cod1, p_rec_cod_old1, p_rec1,
               p_rec_old1 , common.legend = TRUE,
          labels = c("rec sce1 new", "rec sce1 old",
                     "rec sce2 new", "rec sce2 old"),
          vjust = 0.3,
          hjust= -1,
          ncol=2, nrow=2)
p
```




```
p <- ggarrange(p_order_cod1, p_order_cod_old1, p_order1,
               p_order_old1 , common.legend = TRUE,
          labels = c("order sce1 new", "order sce1 old",
                     "order sce2 new", "order sce2 old"),
          vjust = 0.3,
          hjust= -1,
          ncol=2, nrow=2)
p
```




```
p <- ggarrange(p_mds_cod1, p_mds_cod_old1, p_mds1,
               p_mds_old1 , common.legend = TRUE,
          labels = c("mds sce1 new", "mds sce1 old",
                     "mds sce2 new", "mds sce2 old"),
          vjust = 0.3,
          hjust= -1,
          ncol=2, nrow=2)
p
```




## Conclusions

With only one repetition of the simulation, we can already see that the modification improves the distance estimation for dominant markers. Also, give more information for the group and ordering algorithms. The OneMap group function does not work properly for these cases; a better approach needs to be implemented. We can also see that ug and MDS algorithms built a better order compared to other ordering algorithms.

## References

Margarido, G. R. A., Souza, A. P., & Garcia, A. A. F. (2007). OneMap: software for genetic mapping in outcrossing species. Hereditas, 144(3), 78–79. https://doi.org/10.1111/j.2007.0018-0661.02000.x