

Euroformix update

Bug notification:

There is a bug in previous versions of EuroForMix (v1.10.0 and earlier) which causes a small miscalculation of the LR for some situations. There is a new version shortly to be issued (v1.11.3) which has the fix.

Description:

The bug fails to count the "number of previously typed alleles" in the Balding-Nichols formula for additional references than the person of interest (POI) (who is considered as the known non-contributor under Hd). There is no effect if $F_{st}=0$. If $F_{st}>0$ then the effect is present, and it increases as F_{st} increases.

Effect:

- 1) This has no effect on situations like Hp: "POI + unknown" vs Hd: "2 unknowns", or Hp: "POI + 2 unknowns" vs Hd: "3 unknowns", etc.
- 2) There is no effect for any situation where $F_{st}=0$.
- 3) If $F_{st}>0$, e.g. $F_{st}=0.01$, then there is a small effect for situations where known individuals are conditioned under Hd. For example: Hp: "POI + victim" vs Hd: "unknown + victim".
- 4) To give an idea of the magnitude of change after the bug-fix, see figure 1. Here the hypotheses Hp: "POI + Victim1" vs Hd: "1 unknown + Victim1" is considered for #Contr=2 results, and the hypotheses Hp: "POI + Victim1 + Victim2" vs Hd: "1 unknown + Victim1 + Victim2" is considered for #Contr=3 results.

Figure 1 gives the absolute differences in terms of $\log_{10}LR$ (v1.10.0 – v1.11.3). This shows that there is a slight anticonservative bias that increases as the LR and F_{st} increases. So for $LRs < 10e6$ and $F_{st}=0.01$ there is a maximum bias of $\log_{10}LR=+0.1$, increasing to about +0.3 for LRs up to $10e25$. It is always much less than an order of magnitude. The bias increases for higher F_{sts} .

I hope this doesn't cause any inconvenience but let me know if there is anything I can do to help.

I found the bug in the R-function prepareC at line 109 and 116 where the "tmp" variable does not succeed in counting alleles of additional references other than the POI. The bug was found after an external user contacted me.

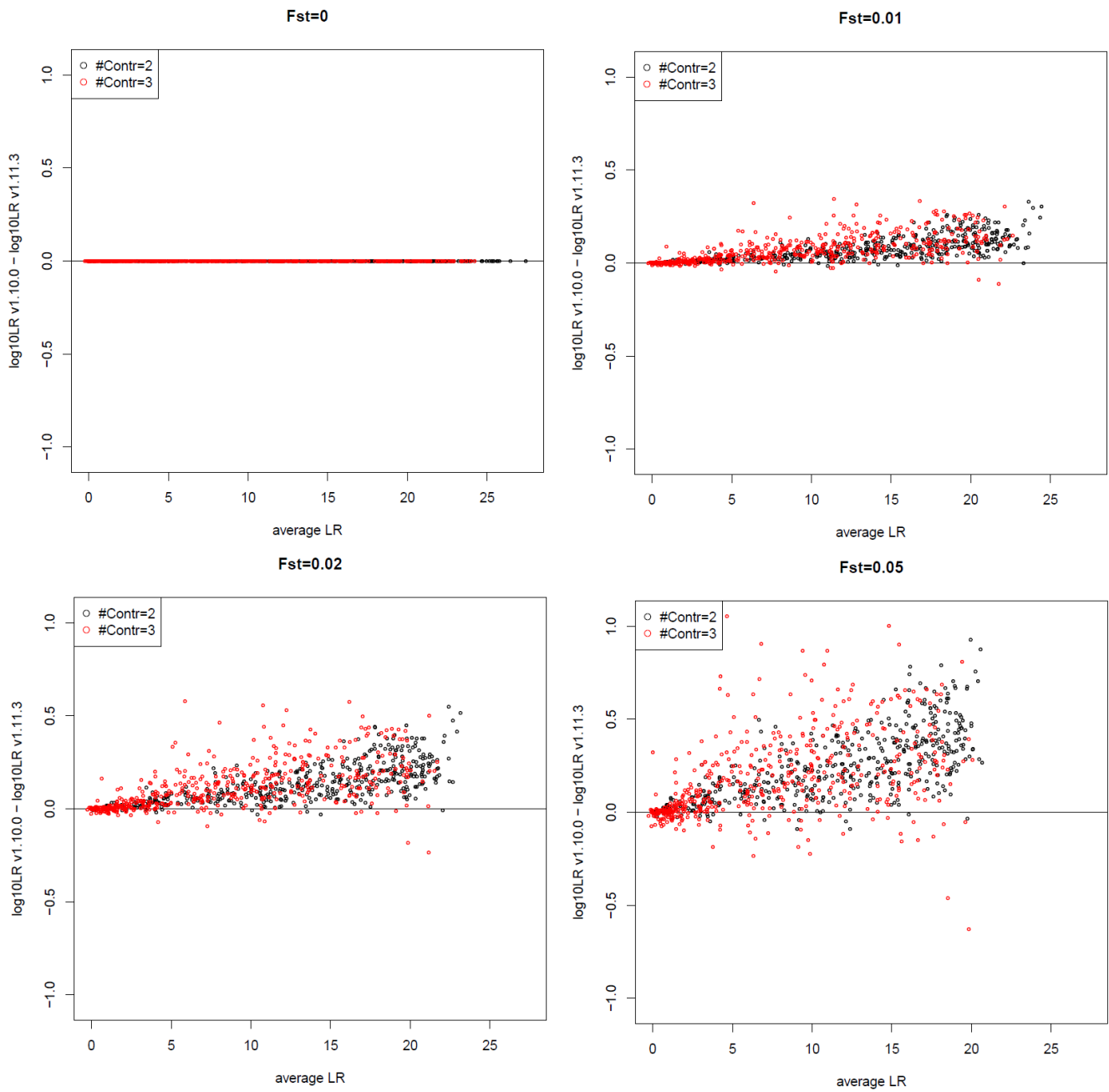


Figure 1: The figures show the LR differences between the original version (v1.10.0) and the fixed version (v1.11.3) for $F_{st}=0,0.01,0.02,0.05$. The hypotheses considered are H_p : "POI + Victim1" vs H_d : "1 unknown + Victim1" for $\#Contr=2$ results, and the hypotheses H_p : "POI + Victim1 + Victim2" vs H_d : "1 unknown + Victim1 + Victim2" for $\#Contr=3$ results. 500 random samples were used for simulation.

Attachment: R-code for simulation (for a given version).

```
library(euroformix)

M = 500 #number of random samples
popFreq = freqImport(paste(path.package("euroformix", quiet =
FALSE),"tutorialdata","FreqDatabases","ESX17_Norway.csv",sep=.Platform$file.sep))[[1]] #popfreq to use:

#Model setup:
threshT = 200 #detection threshold
kit0="ESX17" #selected kit
pC = 0.05 #Pr(dropin)
lam0 = 0.01 #dropin P.H. rate
nDone = 8 #number of successive optimizations

#Sim model setup:
mu0 = 1000 #P.H. expectation
sig0 = 0.4 #rel. P.H. variability
bet0 = 0.75 #degradation slope
stutt0 = 0 #no stutter considered

nCv = 2:3 #number of contributors
FSTv = c(0,0.01,0.02,0.05) #fst values

tab = numeric()
for(nC in nCv) {
  condhp = 1:nC #condition on all under Hp
  condhd = c(0,1:(nC-1)) #not condition on C1 under Hd

  for(m in 1:M) { #for each iter
    print(m)
    set.seed(m)
    simdat <- genDataset(nC,popFreq,mu=mu0,sigma=sig0,beta=bet0,stutt=0, threshT = threshT,kit=kit0,prC=0,nrep=1)
    #plotEPG(simdat$samples,kitname=kit0) #plot observations
    Qobj <- Qassignate(simdat$samples, popFreq, simdat$refData, incS=FALSE, incR=FALSE)
    mxPOI <- simdat$theta$mx[1]

    for(fst in FSTv) {
      hpfit <-
      contLikMLE(nC,samples=simdat$samples,refData=Qobj$refData,popFreq=Qobj$popFreq,threshT=threshT,nDone=nDone,verbose=FAL
SE,xi=0,kit=kit0,condOrder=condhp,prC=pC,lambda=lam0,fst=fst,maxIter = 100)
      hdfit <-
      contLikMLE(nC,samples=simdat$samples,refData=Qobj$refData,popFreq=Qobj$popFreq,threshT=threshT,nDone=nDone,verbose=FAL
SE,xi=0,kit=kit0,condOrder=condhd,prC=pC,lambda=lam0,fst=fst,maxIter = 100,knownRef=1)
      LRmle <- exp(hpfit$fit$loglik - hdfit$fit$loglik)
      tab = rbind(tab, c(nC,m,mxPOI,fst,LRmle) )
    }
  }
}
colnames(tab) <- c("K","id","mx","fst","LR")

ver <- packageVersion("euroformix")
write.table(tab,paste0("EFMres_",ver,".txt"),row.names=FALSE,sep=";")
```