```
neptune<-read.table(file.choose(),sep="\t",dec=".",header=TRUE,colClasses=c(rep("numeric",3), rep("integer",2), rep("character",2),
rep("numeric",2), "character", rep("factor",2)))
#The Neptune file is a data frame including all the radiolarian occurences present on the Neptune database below 40°S, younger than
24Ma, with the synonymies resolved. It consists of 12 columns: 'Sample.Age' (the age of the sample in Ma), 'Sample.depth' (the depth of the
sample on the core in meters below sea floor), 'Water.Depth' (the depth of the site in meters below sea level), "Leg", "Site", "Hole",
"Hole.ID", "Latitude", "Longitude", "Ocean", "Taxon.Abundance" (not used here: it can be semi-quantitative information or truly
quantitative information depending on the source) and finally "Taxon.Name".

bin<-seq(0,24,by=.5)                #set bins (here one bin for half a million year)
length(bin)-1->nb
binmid<-numeric(nb)                 #set the midpoints of each bin (for plotting purposes)
for(i in 1:nb){binmid[i]<-(bin[i]+bin[i+1])/2}

#OCCURENCE & COLLECTION BINNING

occ<-numeric(nb)                    #Nb of Occurences per bin
rawsp<-numeric(nb)                  #Raw species count per bin
occbinned<-list()                   #List of Occurences per bin
ncoll<-numeric(nb)                  #Nb of Collections per bin
collbinned<-list()                  #List of Collections per bin
mostfrequent<-numeric(nb)           #Most abundant species in the bin
ab<-list()                          #Binned abundances
freq<-list()                        #Binned frequencies
u<-numeric(nb)                      #Good's u
single<-numeric(nb)                 #Number of singletons
for (i in 1:nb){
    occbinned[[i]]<-subset(neptune,bin[i]<Sample.Age & bin[i+1]>Sample.Age)
    occ[i]<-nrow(occbinned[[i]])
    rawsp[i]<-length(unique(occbinned[[i]]$Taxon.Name))
    unique(neptune[neptune$Sample.Age<bin[i+1] & neptune$Sample.Age>bin[i],][,c(5,2)])->collbinned[[i]]
    collbinned[[i]][order(collbinned[[i]][,1]),]->collbinned[[i]]
    nrow(collbinned[[i]])->ncoll[i]
    summary(occbinned[[i]]$Taxon.Name,maxsum=1000)->ab[[i]]
    ab[[i]][ab[[i]]!=0]->ab[[i]]
    single[i]<-sum(ab[[i]]==1)
    u[i]<- 1- single[i]/occ[i]
    if(length(ab[[i]])!=0){mostfrequent[i]<-names(ab[[i]])[ab[[i]]==max(ab[[i]],na.rm=T)]}
    freq[[i]]<-ab[[i]]/occ[i]
    }

#BOUNDARY CROSSER on SQS SUBSAMPLED DATA
q<-0.6                              #quota for SQS subsampling
dominant<-"include"                 #can be set to 'exclude' in order to exclude the dominant taxon from the subsampling
trials<-100                         #number of trials
SQSspecies<-array(0,dim=c(trials,nb))   #In-bin number of subsampled species
bcSQS<-array(0,dim=c(trials,nb))    #Boundary crossers after subsampling
rateSQS<-array(0,dim=c(trials,nb))  #Extinction rate
origSQS<-array(0,dim=c(trials,nb))  #Origination rate
pba<-0                              #Progress bar counter
pb<-txtProgressBar(0,nb*trials,style=3)     #Set up progress bar
for(t in 1:trials){
    seen<-array(0,dim=c(length(unique(neptune$Taxon.Name)),nb))      #Occurence matrix after subsampling
    for(i in 1:nb){
        if(nrow(occbinned[[i]])!=0){
            pool<-occbinned[[i]]                #'Picking jar'
            left<-nrow(pool)                    #Nb of occurences left to be picked
            sumfreq<-0                          #Cumulative frequencies
            if(q<=u[i]){
            while(sumfreq<q){
                x<-floor(runif(1,min=1,max=left+1))
                if(seen[as.integer(pool$Taxon.Name[x]),i]==0){
                    if(pool$Taxon.Name[x]!=mostfrequent[i] | dominant=="include"){sumfreq<-sumfreq+as.numeric(freq[[i]]
[names(freq[[i]])==pool$Taxon.Name[x]])}
                    seen[as.integer(pool$Taxon.Name[x]),i]<-1
                    SQSspecies[t,i]<-SQSspecies[t,i]+1
                    }
                pool[x,]<-pool[left,]
                left<-left-1
                }}
            pba<-pba+1
            Sys.sleep(0.01)
            setTxtProgressBar(pb,pba)
            }
        }
    bL<-bt<-Ft<-numeric(nb)         #Respectively taxa disapearing in the bin, taxa crossing the bin, and taxa appearing in the bin
    for (k in 1:nrow(seen)){
        if(seen[k,1]==1 & sum(seen[k,2:nb])>0){bL[1]<-bL[1]+1}
        if(seen[k,nb]==1 & sum(seen[k,1:(nb-1)])>0){Ft[nb]<-Ft[nb]+1}
        for(j in 2:(nb-1)){
            if(sum(seen[k,1:(j-1)])>0 & sum(seen[k,(j+1):nb])>0){bt[j]<-bt[j]+1}
            if(seen[k,j]==1){
            if(sum(seen[k,1:(j-1)])==0 & sum(seen[k,(j+1):nb])>0){bL[j]<-bL[j]+1}
            if(sum(seen[k,1:(j-1)])>0 & sum(seen[k,(j+1):nb])==0){Ft[j]<-Ft[j]+1}
```

```
                }
            }
        }
        bcSQS[t,]<-((bL+bt)+(bt+Ft))/2    # Standing mean diversity, i. e. the mean between the bottom boundary and the top boundary
crossers
        rateSQS[t,]<- -log(bt/(bL+bt))        # Foote's extinction rate
        origSQS[t,]<- -log(bt/(Ft+bt))        # Foote's origination rate
    }
bcSQSM<-SQSspeciesM<-rateSQSM<-origSQSM<-numeric(nb)            # Mean value for each trial of the smd, the in-bin diversity, the
                                                               # extinction and the origination rates

for(i in 1:nb){
    mean(bcSQS[,i],na.rm=T)->bcSQSM[i]
    mean(rateSQS[,i],na.rm=T)->rateSQSM[i]
    mean(origSQS[,i],na.rm=T)->origSQSM[i]
    mean(SQSspecies[,i],na.rm=T)->SQSspeciesM[i]
    }
rbind("SQS subsampled diversity"=SQSspeciesM,"Boundary Crossers on SQS"=bcSQSM)->sqsmat            #Output matrix
for(i in 1:nb){if(sqsmat[2,i]==0){sqsmat[2,i]<-NA}}
SQS<-list("SQS Subsampling summary"=sqsmat,"SQS subsampling value for 100 trials"=SQSspecies, "SQS Boundary crossers in 100
trials"=bcSQS, "Extinction rate"=rateSQSM, "Origination rate"=origSQSM)

#BOUNDARY CROSSER on UW BY-COLLECTION SUBSAMPLED DATA

quota<-10                                #Quota on the number of collections per bin
trials<-100
bcUW<-uwSpecies<-rateUW<-origUW<-array(0,dim=c(trials,nb))
pba<-0
pb<-txtProgressBar(0,nb*trials,style=3)
for(t in 1:trials){
    seen<-array(0,dim=c(length(levels(neptune$Taxon.Name)),nb))
    for(i in 1:nb){
        temp<-collbinned[[i]]
        collsamp<-0                            #Nb of Collections already picked
        uwsampled<-c()                         #Picked occurences cumulative storage
        ncollLeft<-ncoll[i]
        if(ncoll[i]<quota){uwSpecies[t,i]<-NA}
        else{
            while(collsamp<quota){
                cellNo<-floor(runif(1,min=1,max=ncollLeft+1))        #Pick randomly a sample out of the remaining samples
                neptune[(neptune$Site==temp[cellNo,1] & neptune$Sample.Depth==temp[cellNo,2]),]->picked
                rbind(uwsampled,picked)->uwsampled
                seen[picked$Taxon.Name,i]<-1            #Fill in the subsampled occurence matrix
                temp[cellNo,]<-temp[ncollLeft,]        #To avoid picking the same cell again and again
                collsamp<-collsamp+1
                ncollLeft<-ncollLeft-1
            }
            uwSpecies[t,i]<-length(unique(uwsampled$Taxon.Name))        #Number of taxa picked
        }
    }
    pba<-pba+1                            #Progress Bar counter
    Sys.sleep(0.01)
    setTxtProgressBar(pb,pba)
    }
    bL<-Ft<-bt<-numeric(nb)                    #Respectively:taxa disappearing, taxa appearing and taxa crossing the bin
    for (k in 1:nrow(seen)){
        if(seen[k,1]==1 & sum(seen[k,2:nb])>0){bL[1]<-bL[1]+1}            #For the first and last bins
        if(seen[k,nb]==1 & sum(seen[k,1:(nb-1)])>0){Ft[nb]<-Ft[nb]+1}
        for(j in 2:(nb-1)){
            if(sum(seen[k,1:(j-1)])>0 & sum(seen[k,(j+1):nb])>0){bt[j]<-bt[j]+1}        #For all the other bins
            if(seen[k,j]==1){
            if(sum(seen[k,1:(j-1)])==0 & sum(seen[k,(j+1):nb])>0){bL[j]<-bL[j]+1}
            if(sum(seen[k,1:(j-1)])>0 & sum(seen[k,(j+1):nb])==0){Ft[j]<-Ft[j]+1}
            }
        }
    }
    bcUW[t,]<-((bL+bt)+(bt+Ft))/2                        #Boundary Crossers (standing mean diversity)
    rateUW[t,]<- -log(bt/(bL+bt))
    origUW[t,]<- -log(bt/(Ft+bt))
}
uwSpeciesM<-bcUWM<-rateUWM<-origUWM<-numeric(nb)                        #mean richness & mean bc on t trials
for(i in 1:nb){
    mean(uwSpecies[,i],na.rm=T)->uwSpeciesM[i]
    mean(bcUW[,i],na.rm=T)->bcUWM[i]
    mean(rateUW[,i],na.rm=T)->rateUWM[i]
    mean(origUW[,i],na.rm=T)->origUWM[i]
    }
rbind("UW subsampled diversity"=uwSpeciesM,"Boundary Crossers on UW"=bcUWM)->uwmat                #Output matrix
for(i in 1:nb){if(uwmat[2,i]==0){uwmat[2,i]<-NA}}
UW<-list("UW Subsampling summary"=uwmat,"UW subsampling value for 100 trials"=uwSpecies, "UW Boundary crossers in 100
trials"=bcUW, "Extinction rate"=rateUWM, "Origination rate"=origUWM)

#BOUNDARY CROSSER on OW BY-COLLECTION SUBSAMPLED DATA

quota<-100                # Quota on the number of occurences per bin
```

```
trials<-100
bcOW<-owSpecies<-array(0,dim=c(trials,nb))
pba<-0
pb<-txtProgressBar(0,nb*trials,style=3)
for(t in 1:trials){
    seen<-array(0,dim=c(length(levels(neptune$Taxon.Name)),nb))
    for(i in 1:nb){
        temp<-collbinned[[i]]
        occsamp<-0
        owsampled<-c()
        ncollLeft<-ncoll[i]
        if(occ[i]<quota){owSpecies[t,i]<-NA}
        else{
            while(occsamp<quota){
                cellNo<-floor(runif(1,min=1,max=ncollLeft+1))
                neptune[(neptune$Site==temp[cellNo,1] & neptune$Sample.Depth==temp[cellNo,2]),]->picked
                rbind(owsampled,picked)->owsampled
                seen[picked$Taxon.Name,i]<-1
                temp[cellNo,]<-temp[ncollLeft,]
                occsamp<-occsamp+nrow(picked)
                ncollLeft<-ncollLeft-1
            }
            owSpecies[t,i]<-length(unique(owsampled$Taxon.Name))
        }
    }
    pba<-pba+1
    Sys.sleep(0.01)
    setTxtProgressBar(pb,pba)
    }
    bL<-bt<-Ft<-numeric(nb)
        for (k in 1:nrow(seen)){
        if(seen[k,1]==1 & sum(seen[k,2:nb])>0){bL[1]<-bL[1]+1}
        if(seen[k,nb]==1 & sum(seen[k,1:(nb-1)])>0){Ft[nb]<-Ft[nb]+1}
        for(j in 2:(nb-1)){
            if(sum(seen[k,1:(j-1)])>0 & sum(seen[k,(j+1):nb])>0){bt[j]<-bt[j]+1}
            if(seen[k,j]==1){
            if(sum(seen[k,1:(j-1)])==0 & sum(seen[k,(j+1):nb])>0){bL[j]<-bL[j]+1}
            if(sum(seen[k,1:(j-1)])>0 & sum(seen[k,(j+1):nb])==0){Ft[j]<-Ft[j]+1}
                }
            }
        }
    bcOW[t,]<-((bL+bt)+(bt+Ft))/2
}
owSpeciesM<-bcOWM<-numeric(nb)
for(i in 1:nb){
    mean(owSpecies[,i],na.rm=T)->owSpeciesM[i]
    mean(bcOW[,i],na.rm=T)->bcOWM[i]
    }
rbind("OW subsampled diversity"=owSpeciesM,"Boundary Crossers on OW"=bcOWM)->owmat
for(i in 1:nb){if(owmat[2,i]==0){owmat[2,i]<-NA}}
OW<-list("OW Subsampling summary"=owmat,"OW subsampling value for 100 trials"=owSpecies, "OW Boundary crossers in 100
trials"=bcOW)

#BOUNDARY CROSSER on OW-squared BY-COLLECTION SUBSAMPLED DATA

quota<-500                  # Quota on the number of occurences (squared) per bin
trials<-100
bcOW2<-ow2Species<-array(0,dim=c(trials,nb))
pba<-0
pb<-txtProgressBar(0,nb*trials,style=3)
occ2<-numeric(nb)          # Number of occurence-squared per bin
for(i in 1:nb){
    for(k in 1:nrow(collbinned[[i]])){
        occ2[i]+nrow(subset(neptune,Site==collbinned[[i]][k,1] & Sample.Depth==collbinned[[i]][k,2]))^2->occ2[i]
    }
}
for(t in 1:trials){
    seen<-array(0,dim=c(length(levels(neptune$Taxon.Name)),nb))
    for(i in 1:nb){
        temp<-collbinned[[i]]
        occsamp<-0
        ow2sampled<-c()
        ncollLeft<-ncoll[i]
        if(occ2[i]<quota){ow2Species[t,i]<-NA}
        else{
            while(occsamp<quota){
                cellNo<-floor(runif(1,min=1,max=ncollLeft+1))
                subset(neptune,Site==temp[cellNo,1] & Sample.Depth==temp[cellNo,2])->picked
                rbind(ow2sampled,picked)->ow2sampled
                seen[picked$Taxon.Name,i]<-1
                temp[cellNo,]<-temp[ncollLeft,]
                occsamp<-occsamp+nrow(picked)^2
                ncollLeft<-ncollLeft-1
            }
```

```r
                ow2Species[t,i]<-length(unique(ow2sampled$Taxon.Name))
            }
        pba<-pba+1
        Sys.sleep(0.01)
        setTxtProgressBar(pb,pba)
        }
        bL<-bt<-Ft<-numeric(nb)
        for (k in 1:nrow(seen)){
            if(seen[k,1]==1 & sum(seen[k,2:nb])>0){bL[1]<-bL[1]+1}
            if(seen[k,nb]==1 & sum(seen[k,1:(nb-1)])>0){Ft[nb]<-Ft[nb]+1}
            for(j in 2:(nb-1)){
                if(sum(seen[k,1:(j-1)])>0 & sum(seen[k,(j+1):nb])>0){bt[j]<-bt[j]+1}
                if(seen[k,j]==1){
                if(sum(seen[k,1:(j-1)])==0 & sum(seen[k,(j+1):nb])>0){bL[j]<-bL[j]+1}
                if(sum(seen[k,1:(j-1)])>0 & sum(seen[k,(j+1):nb])==0){Ft[j]<-Ft[j]+1}
                    }
                }
            }
        bcOW2[t,]<-((bL+bt)+(bt+Ft))/2
}
ow2SpeciesM<-bcOW2M<-numeric(nb)
for(i in 1:nb){
    mean(ow2Species[,i],na.rm=T)->ow2SpeciesM[i]
    mean(bcOW2[,i],na.rm=T)->bcOW2M[i]
    }
rbind("OW2 subsampled diversity"=ow2SpeciesM,"Boundary Crossers on OW2"=bcOW2M)->ow2mat
for(i in 1:nb){if(ow2mat[2,i]==0){ow2mat[2,i]<-NA}}
OW2<-list("OW2 Subsampling summary"=ow2mat, "OW2 subsampling value for 100 trials"=ow2Species, "OW2 Boundary crossers in 100
trials"=bcOW2)

#Classic Rarefaction

quota<-100
trials<-100
rarefied<-bcRar<-array(dim=c(trials,nb))
pba<-0
pb<-txtProgressBar(0,trials,style=3)
for(t in 1:trials){
    seen<-array(0,dim=c(length(levels(neptune$Taxon.Name)),nb))
    for(i in 1:nb){
        if(occ[i]<quota){rarefied[t,i]<-NA}
        else{
            occbinned[[i]][sample(nrow(occbinned[[i]]),quota),]->d
            length(unique(d$Taxon.Name))->rarefied[t,i]
            seen[unique(d$Taxon.Name),i]<-1
            }
        }
    bL<-bt<-Ft<-numeric(nb)
    for (k in 1:nrow(seen)){
        if(seen[k,1]==1 & sum(seen[k,2:nb])>0){bL[1]<-bL[1]+1}
        if(seen[k,nb]==1 & sum(seen[k,1:(nb-1)])>0){Ft[nb]<-Ft[nb]+1}
        for(j in 2:(nb-1)){
            if(sum(seen[k,1:(j-1)])>0 & sum(seen[k,(j+1):nb])>0){bt[j]<-bt[j]+1}
            if(seen[k,j]==1){
            if(sum(seen[k,1:(j-1)])==0 & sum(seen[k,(j+1):nb])>0){bL[j]<-bL[j]+1}
            if(sum(seen[k,1:(j-1)])>0 & sum(seen[k,(j+1):nb])==0){Ft[j]<-Ft[j]+1}
                }
            }
        }
    bcRar[t,]<-((bL+bt)+(bt+Ft))/2
    pba<-pba+1
    Sys.sleep(0.01)
    setTxtProgressBar(pb,pba)
}
rarefiedM<-bcRarM<-numeric(nb)
for(i in 1:nb){
    mean(rarefied[,i],na.rm=T)->rarefiedM[i]
    mean(bcRar[,i],na.rm=T)->bcRarM[i]
    }
rbind("Rarefied diversity (100 specimens)"=rarefiedM,"Boundary Crossers"=bcRarM)->Rarmat
for(i in 1:nb){if(Rarmat[2,i]==0){Rarmat[2,i]<-NA}}
Rar<-list("Rarefaction summary"=Rarmat, "Classic Rarefaction value for 100 trials"=rarefied, "Boundary crossers in 100 trials"=bcRar)

save(UW,OW,OW2,SQS,Rar,bin,binmid,file="NeptuneSubsampling.Rdata")
```