

APPENDIX S1 Implementation in R of the greedy algorithm used to determine the optimal order of sample accumulation for maximum rate of discovery of new species. See Table 1 for times taken to conduct different sampling methods.

```
rm(list=ls(all=TRUE))
Species<-read.table(file="C:\\Research\\Current projects\\Kalsum - ant surveys in the high
canopy\\Greedy\\Species input.txt")
Times<-c(rep(66,20),rep(42.833,20),rep(30.833,20)) #Times taken for each kind of sampling
prior to any other kind being conducted on that tree
nsp<-length(Species[,1]) #Total number of species in dataset
output.names<-matrix(nrow=60,ncol=1,data=0) #Vector to receive order of samples
output.numbers<-matrix(nrow=60,ncol=6,data=0) #Matrix to receive data relating to species
accumulation
colnames(output.numbers)<-
c("New.species","Sampling.hours","Cum.species","Rate","Index","Cum.hours")
for (i in 1:60){ #Loop to accumulate all sixty samples (20 of each kind)
  #deciding which one to add next:
  rbind(Times,colSums(Species)/Times)->time.rate
  subset(time.rate[2,],time.rate[2,]==max(time.rate[2,]))->mx #Find the sample that gives
the most new species per hour
  which.max(time.rate[2,])->index #Location in Species of maximum for this loop
  output.names[i]<-names(mx[1])
  output.numbers[i,1]<-colSums(Species)[index]
  output.numbers[i,2]<-Times[index]
  output.numbers[i,3]<-if(i==1) output.numbers[i,1] else output.numbers[i-
1,3]+output.numbers[i,1]
  output.numbers[i,4]<-mx[1]
  output.numbers[i,5]<-index
  output.numbers[i,6]<-if(i==1) output.numbers[i,2] else output.numbers[i-
1,6]+output.numbers[i,2]
  output.names
  output.numbers
  for (j in 1:nsp){
    if (Species[j,index]==1) Species[j,]<-0 #deletes all species occurrences from
species already discovered so species already discovered are discounted in future calculations of
rates of species discovery
  }
  Species[,index]<--1 #stops that row from being sampled again by giving it a
negative rate of spp accumulation (compared to unsampled, but unproductive columns)
  if(index<21) if(Times[index+20]==42.833) Times[index+20]<-26.833 #These six lines
alter the time taken for taking other samples on the same tree
  if(index<21) if(Times[index+40]==30.833) Times[index+40]<-14.833
  if(20<index) if(index<41) Times[index-20]<-50
  if(20<index) if(index<41) Times[index+20]<-7.5
  if(40<index) Times[index-40]<-50
  if(40<index) Times[index-20]<-19.5
```

```

}

par(family="serif",lend="square",lwd=2,xaxs="i",yaxs="i")
plot(output.numbers[,3]~output.numbers[,6],xlab="Researcher hours",ylab="Species
collected",ylim=c(0,180),xlim=c(0,1400),type="n",las=1,bty="l") #plots graph of accumulation
with different colours for different sample types
lines(c(output.numbers[1,6],0),c(output.numbers[1,3],0),col="black",lwd=6,lend="square")
for (i in 2:60){
  if(output.numbers[i,5]<21) lines(c(output.numbers[i,6],output.numbers[i-
1,6]),c(output.numbers[i,3],output.numbers[i-1,3]),col="black",lwd=6,lend="square")
  if(20<output.numbers[i,5]) if(output.numbers[i,5]<41)
lines(c(output.numbers[i,6],output.numbers[i-1,6]),c(output.numbers[i,3],output.numbers[i-
1,3]),col="grey55",lwd=6,lend="square")
  if(40<output.numbers[i,5]) lines(c(output.numbers[i,6],output.numbers[i-
1,6]),c(output.numbers[i,3],output.numbers[i-1,3]),col="light grey",lwd=6,lend="square")
}
legend(700,50,c("Fogging","Purse-string trap","Baited pitfall trap"), lwd=6,
col=c("black","grey55","light grey"),bty="n",cex=1)

```

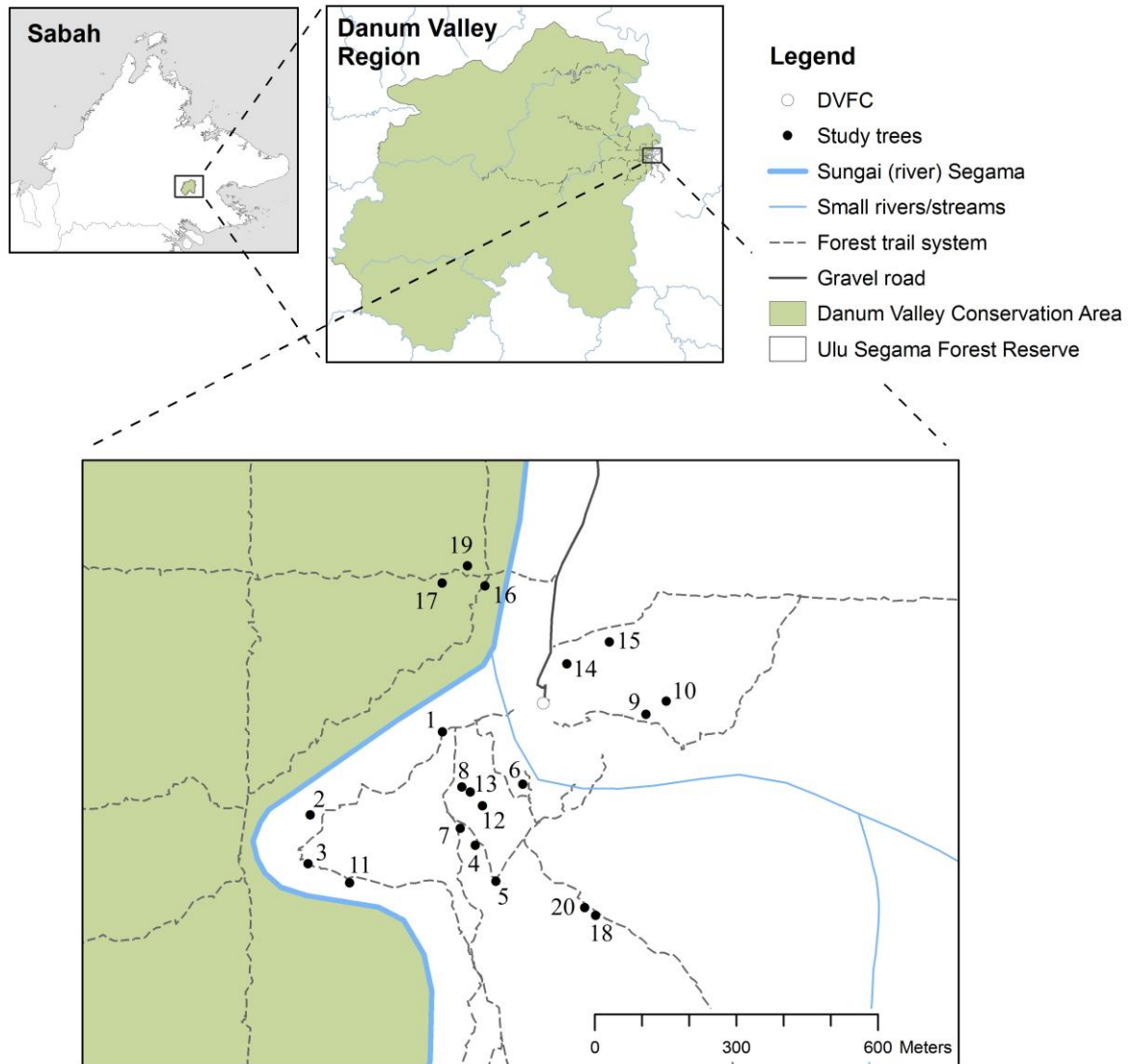


FIGURE S1. Locations of the trees sampled in the primary forest in the Danum Valley Conservation Area and the Ulu Segama Forest Reserve.



FIGURE S2. Mixed photography of sampling in the high canopy for ants.

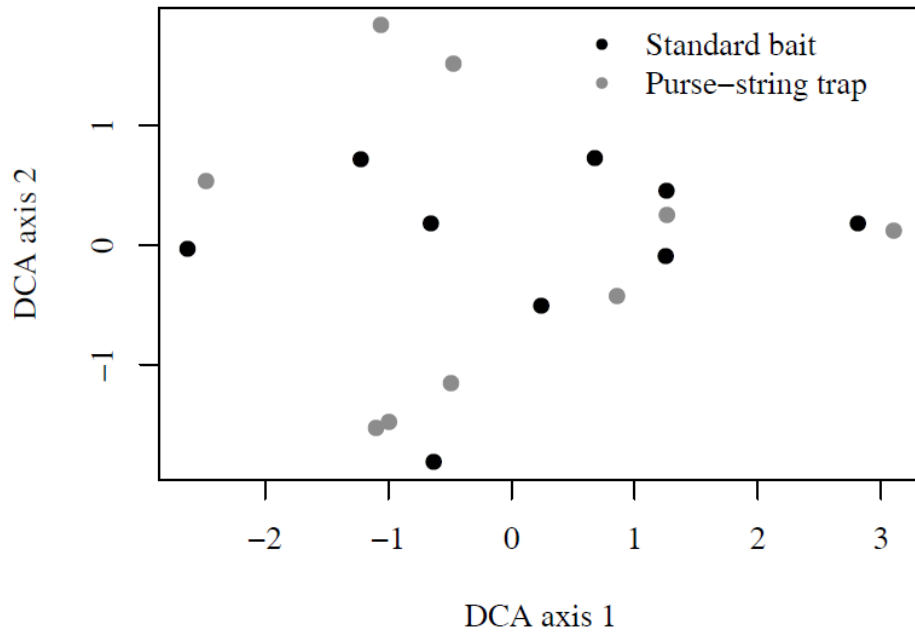


FIGURE S3. Unconstrained Detrended Correspondence Analysis (DCA) ordination of species composition for samples resulting from purse-string trapping and standard baiting.

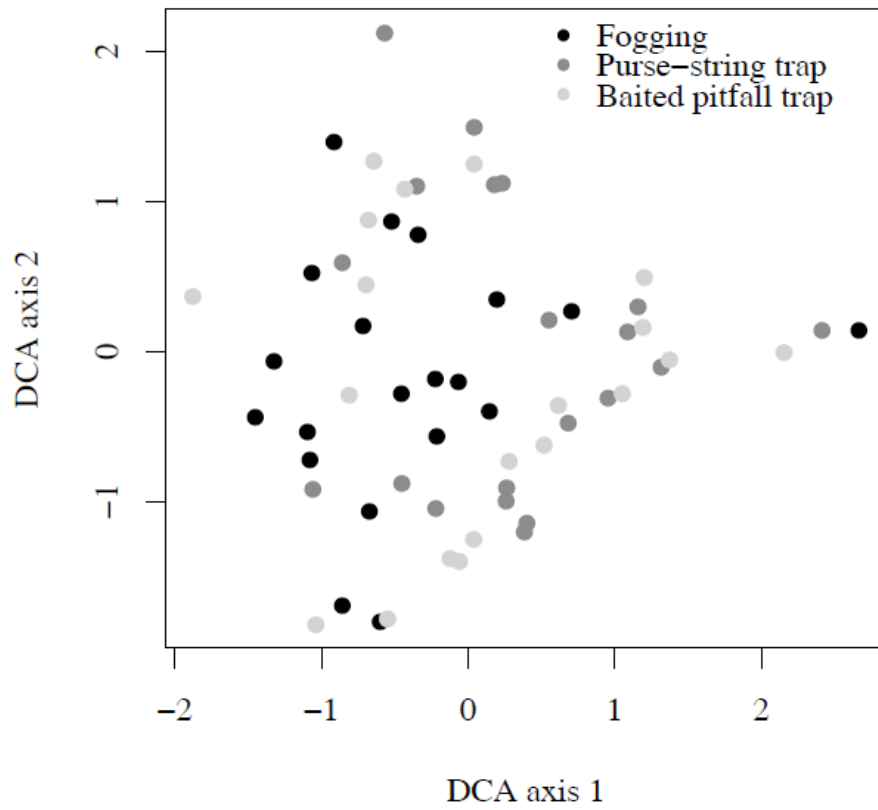


FIGURE S4. Unconstrained Detrended Correspondence Analysis (DCA) ordination of species composition for samples resulting from purse-string trapping, baited pitfall trapping and fogging.