Supplementary appendix to:

Graphs in phylogenetic comparative analysis: Anscombe's quartet revisited

Liam J. Revell[1,2,*], Klaus Schliep[2], Eugenio Valderrama[3], and James E. Richardson[1,4]

[1]Programa de Biología, Universidad del Rosario, Bogotá, Colombia

[2]Department of Biology, University of Massachusetts Boston, Boston MA

[3]Universidad el Bosque, Bogotá, Colombia

[4]Tropical Diversity Section, Royal Botanic Garden Edinburgh, Edinburgh, Scotland

[*]Correspondence: +1 314-580-3045, liam.revell@phytools.org

Supplementary appendix I: R code to produce Figure 1 of the article.

```
library(datasets)
attach(anscombe)
anscombe
fits<-list()
for(i in 1:4){
      model<-paste("y",i,"~x",i,sep="")
      cat(paste("Model",i,"\n--------",model,"--------\n"))
      fits[[i]]<-lm(model)
      cat("\n")
      print(coef(fits[[i]]))
      print(anova(fits[[i]]))
      cat("\n----------------------\n\n")
}
pdf(file="Figure1.pdf")
par(mfrow=c(2,2),mar=c(4.6,4.1,2.6,1.1))
for(i in 1:4){
      model<-paste("y",i,"~x",i,sep="")
      plot(as.formula(model),pch=21,
            bg=phytools::make.transparent("orange",0.5),cex=2,xlim=c(3,19),
            ylim=c(3,13))
      abline(fits[[i]],lwd=2,col=phytools::make.transparent("blue",0.25))
      mtext(paste(letters[i],")",sep=""),adj=-0.1,line=1)
}
dev.off()
```

Supplementary appendix II: R code to generate phylogenetic Anscombe datasets, Tables 1 & 2, and Figures 2 through 6 of this article.

```
SEED<-56 ## different seeds will result in different datasets that also
         ## satisfy the property of those in this study.
set.seed(SEED)
library(phytools)
library(geiger)

## generate tree
tree<-untangle(ladderize(pbtree(n=20,scale=1,tip.label=LETTERS[1:20])),
      "read.tree") ## ladderized for plotting purposes only
## end generate tree

## Figure 2
pdf(file="Figure2.pdf")
plotTree(tree,fsize=1.4,ftype="i")
dev.off()
## end Figure 2

## generate data
y1<-setNames(sort(fastBM(tree)),tree$tip.label)
y2<-fastBM(tree,a=3,sig2=0.5)
y3<-setNames(runif(n=Ntip(tree),min=0,max=6),tree$tip.label)
y4<-fastBM(tree,a=3,sig2=0.1)

f1<-function(theta,sig2=1,a=0,y,C){
      y<-y/theta[1]-theta[2]
      N<-nrow(C)
      obj<-phyl.vcv(as.matrix(y),C,1)
      (obj$R[1,1]*(N-1)/N-sig2)^2+(obj$alpha[1,1]-a)^2
}
f2<-function(theta,sig2=1,a=0,y,C){
      y<-round(y/theta[1]-theta[2],2)
      N<-nrow(C)
      obj<-phyl.vcv(as.matrix(y),C,1)
      (obj$R[1,1]*(N-1)/N-sig2)^2+(obj$alpha[1,1]-a)^2
}
fit<-optim(c(1,0),f1,sig2=0.5,a=3.0,y=y1,C=vcv(tree),method="L-BFGS-B",
      lower=c(1e-12,-Inf),upper=rep(Inf,2))
y1<-y1/fit$par[1]-fit$par[2]

fit<-optim(c(1,0),f1,sig2=0.5,a=3.0,y=y2,C=vcv(tree),method="L-BFGS-B",
      lower=c(1e-12,-Inf),upper=rep(Inf,2))
y2<-y2/fit$par[1]-fit$par[2]

fit<-optim(c(1,0),f1,sig2=0.5,a=3.0,y=y3,C=vcv(tree),method="L-BFGS-B",
      lower=c(1e-12,-Inf),upper=rep(Inf,2),control=list(maxit=2000))
y3<-y3/fit$par[1]-fit$par[2]
```

```
f1<-function(theta,sig2=1,a=0,y,C){
      y<-setNames(c(y,theta[1]),rownames(C))-theta[2]
      N<-nrow(C)
      obj<-phyl.vcv(as.matrix(y),C,1)
      (obj$R[1,1]*(N-1)/N-sig2)^2+(obj$alpha[1,1]-a)^2
}
f2<-function(theta,sig2=1,a=0,y,C){
      y<-round(setNames(c(y,theta[1]),rownames(C))-theta[2],2)
      N<-nrow(C)
      obj<-phyl.vcv(as.matrix(y),C,1)
      (obj$R[1,1]*(N-1)/N-sig2)^2+(obj$alpha[1,1]-a)^2
}
fit<-optim(c(y4[20],0),f1,sig2=0.5,a=3,y=y4[1:19],C=vcv(tree),
      method="L-BFGS-B",lower=c(-Inf,-Inf),upper=c(Inf,Inf))
y4[20]<-fit$par[1]
y4<-y4-fit$par[2]

## Table 1
ind<-LETTERS[1:20]
round(data.frame(y1,y2,y3,y4)[ind,],2)
write.csv(file="Table1.csv",round(data.frame(y1,y2,y3,y4)[ind,],2))
## end Table 1

## Table 2
fit1<-fitContinuous(tree,y1)
fit2<-fitContinuous(tree,y2)
fit3<-fitContinuous(tree,y3)
fit4<-fitContinuous(tree,y4)
obj<-data.frame(
      sig2=c(fit1$opt$sigsq,fit2$opt$sigsq,fit3$opt$sigsq,fit3$opt$sigsq),
      a=c(fit1$opt$z0,fit2$opt$z0,fit3$opt$z0,fit3$opt$z0),
      logL=c(fit1$opt$lnL,fit2$opt$lnL,fit3$opt$lnL,fit3$opt$lnL),
      AIC=c(fit1$opt$aic,fit2$opt$aic,fit3$opt$aic,fit3$opt$aic),
      k=c(fit1$opt$k,fit2$opt$k,fit3$opt$k,fit3$opt$k))
signif(obj,4)
write.csv(file="Table2.csv",signif(obj,4))
## end Table 2

## Figure 3
pdf("Figure3.pdf")
par(mfrow=c(2,2),mar=c(4.6,4.1,3.1,1.1))
labels<-setNames(seq(min(c(y1,y2,y3,y4)),max(c(y1,y2,y3,y4)),
    by=diff(range(c(y1,y2,y3,y4)))/(Ntip(tree)-1)),
    names(sort(y1)))
phenogram(tree,y1,ftype="off",ylim=range(c(y1,y2,y3,y4)),
      label.pos=labels,colors="lightblue",ylab=expression(y[1]))
mtext("a)",adj=-0.1,line=1)
labels<-setNames(seq(min(c(y1,y2,y3,y4)),max(c(y1,y2,y3,y4)),
    by=diff(range(c(y1,y2,y3,y4)))/(Ntip(tree)-1)),
    names(sort(y2)))
phenogram(tree,y2,ftype="off",ylim=range(c(y1,y2,y3,y4)),
      label.pos=labels,colors="lightgreen",ylab=expression(y[2]))
mtext("b)",adj=-0.1,line=1)
labels<-setNames(seq(min(c(y1,y2,y3,y4)),max(c(y1,y2,y3,y4)),
    by=diff(range(c(y1,y2,y3,y4)))/(Ntip(tree)-1)),
    names(sort(y3)))
```

```
phenogram(tree,y3,ftype="off",ylim=range(c(y1,y2,y3,y4)),
      label.pos=labels,colors="orange",ylab=expression(y[3]))
mtext("c)",adj=-0.1,line=1)
labels<-setNames(seq(min(c(y1,y2,y3,y4)),max(c(y1,y2,y3,y4)),
    by=diff(range(c(y1,y2,y3,y4)))/(Ntip(tree)-1)),
    names(sort(y4)))
phenogram(tree,y4,ftype="off",ylim=range(c(y1,y2,y3,y4)),
      label.pos=labels,colors="red",ylab=expression(y[4]))
mtext("d)",adj=-0.1,line=1)
dev.off()
## end Figure 3

## Figure 4
pdf(file="Figure4.pdf")
par(mfrow=c(2,2))
c1<-setMap(contMap(tree,y1,plot=FALSE),invert=T)
plot(c1,leg.txt="",ftype="off",xlim=c(-0.1,1.1),mar=c(1.1,1.1,3.1,1.1),
      lwd=3)
text(x=0.25,y=0.7,expression(y[1]))
mtext("a)",adj=0.05,line=1)
c2<-setMap(contMap(tree,y2,plot=FALSE),invert=T)
plot(c2,leg.txt="",ftype="off",xlim=c(-0.1,1.1),mar=c(1.1,1.1,3.1,1.1),
      lwd=3)
text(x=0.25,y=0.7,expression(y[2]))
mtext("b)",adj=0.05,line=1)
c3<-setMap(contMap(tree,y3,plot=FALSE),invert=T)
plot(c3,leg.txt="",ftype="off",xlim=c(-0.1,1.1),mar=c(1.1,1.1,3.1,1.1),
      lwd=3)
text(x=0.25,y=0.7,expression(y[3]))
mtext("c)",adj=0.05,line=1)
c4<-setMap(contMap(tree,y4,plot=FALSE),invert=T)
plot(c4,leg.txt="",ftype="off",xlim=c(-0.1,1.1),mar=c(1.1,1.1,3.1,1.1),
      lwd=3)
text(x=0.25,y=0.7,expression(y[4]))
mtext("d)",adj=0.05,line=1)
dev.off()
## end Figure 4

## Figure 5
pdf(file="Figure5.pdf")
par(mfrow=c(2,4))
foo<-function(tip,tree,x) fitContinuous(drop.tip(tree,tip),
      x[-which(tree$tip.label==tip)])$opt$sigsq
sig2<-setNames(sapply(tree$tip.label,foo,tree=tree,x=y1),tree$tip.label)
plotTree.barplot(tree,sig2,add=TRUE,args.plotTree=list(ftype="off",lwd=2,
      mar=c(4.1,3.1,3.1,0)),args.barplot=list(xlab=expression(sigma^2),
      xlim=c(0,0.75),mar=c(4.1,0,3.1,1.1)))
abline(v=0.5,col="red")
text(x=0.525,y=33.5,expression(paste(sigma^2,"=0.5",sep="")),pos=4)
mtext("a)",adj=-1,line=1)
sig2<-setNames(sapply(tree$tip.label,foo,tree=tree,x=y2),tree$tip.label)
plotTree.barplot(tree,sig2,add=TRUE,args.plotTree=list(ftype="off",lwd=2,
      mar=c(4.1,3.1,3.1,0)),args.barplot=list(xlab=expression(sigma^2),
      xlim=c(0,0.75),mar=c(4.1,0,3.1,1.1)))
abline(v=0.5,col="red")
text(x=0.525,y=33.5,expression(paste(sigma^2,"=0.5",sep="")),pos=4)
mtext("b)",adj=-1,line=1)
```

```
sig2<-setNames(sapply(tree$tip.label,foo,tree=tree,x=y3),tree$tip.label)
plotTree.barplot(tree,sig2,add=TRUE,args.plotTree=list(ftype="off",lwd=2,
      mar=c(4.1,3.1,3.1,0)),args.barplot=list(xlab=expression(sigma^2),
      xlim=c(0,0.75),mar=c(4.1,0,3.1,1.1)))
abline(v=0.5,col="red")
text(x=0.525,y=33.5,expression(paste(sigma^2,"=0.5",sep="")),pos=4)
mtext("c)",adj=-1,line=1)
sig2<-setNames(sapply(tree$tip.label,foo,tree=tree,x=y4),tree$tip.label)
plotTree.barplot(tree,sig2,add=TRUE,args.plotTree=list(ftype="off",lwd=2,
      mar=c(4.1,3.1,3.1,0)),args.barplot=list(xlab=expression(sigma^2),
      xlim=c(0,0.75),mar=c(4.1,0,3.1,1.1)))
abline(v=0.5,col="red")
text(x=0.525,y=33.5,expression(paste(sigma^2,"=0.5",sep="")),pos=4)
mtext("d)",adj=-1,line=1)
dev.off()
## end Figure 5


## Figure 6
pdf(file="Figure6.pdf")
X<-fastBM(tree,a=3,sig2=0.5,nsim=9)
par(mfrow=c(3,3))
nulo<-apply(X,2,phenogram,tree=tree,colors="lightblue",ftype="off")
dev.off()
## end Figure 6
```