Rumbling orchids: How to assess divergent evolution between the nuclear host and chloroplast endosymbionts

Oscar Alejandro Pérez-Escobar; Juan Antonio Balbuena; Marc Gottschling

Supporting information

including: User guide – managing the pipeline in R. A tutorial to execute the pipeline described in the main text is provided, using as a working example conflicting chloroplast and nuclear phylogenies of *Satyrium* (Orchidaceae).

including: additional Figures S1–S6.

including: additional Tables S1–S5.

USER GUIDE – MANAGING THE PIPELINE IN R

An R (R Development core team 2015) script is presented to carry out tests of phylogenetic congruence, and detection of outlier associations, between trees derived from organellar and nuclear loci. The script stands as a pipeline to execute PACo (Procustes Approach to Phylogeny: Balbuena et al. 2013) and ParaFit (Legendre et al. 2002) that are traditionally employed in coevolutionary studies. In addition, it also includes a set of functions useful to spot outliers in trees.

To be executed, the pipeline requires two sets of posterior probability trees derived from Bayesian inference or Maximum Likelihood (ML) phylogenies, corresponding to the organellar and nuclear trees, respectively. The user can decide to run PACo and ParaFit either with phylograms or unit branch length trees as input, in order to take into account and compare the effects of considering evolutionary rates. PACo yields a residual sum of squares (m_{XY}^2) that measures congruence between phylogenies and uses a permutation approach to test significance. Squared residual values (e_i^2) provide a direct measure of each 'o'-/'n'-association's contribution to the global value m_{XY}^2. This estimate can be normalized as a proportion of m_{XY}^2 (i.e., $\varepsilon_i^2 = e_i^2 / m_{XY}^2$). In case of complete congruence between both phylogenies, the ε_i^2's are expected to follow a uniform distribution with expected mean 1/N, where N = number of 'o'-/'n'-associations. Therefore, 1/N provides a threshold value and any ε_i^2 linked to a conflicting association is expected to be > 1/N.

As for ParaFit, the pipeline computes the ParaFitLink2 statistic ($pfl2_i$), which also evaluates the contribution of each link association and is more appropriate than ParaFitLink1 in one-to-one association scenarios (Legendre et al. 2002). The $pfl2_i$ value of a given association is inversely proportional to the phylogenetic pattern observed. Therefore, outlier sequences are expected to have $pfl2_i \approx 0$. The pipeline produces plots of the median and 95% empirical confidence intervals of ε_i^2 and $pfl2_i$ values, and outlier associations can be identified by comparison with a given cut-off value. Because in all simulations and real data set analyses PACo performed better than the *pfl2* statistic, the respective ε_i^2 value of each association only is plotted independently onto the nuclear and organelle phylogenies, thus providing a visual detection of outliers for the end-user.

In order to assist users with little or no experience about R, we provide herein a tutorial to the pipeline. All analyses can be executed by cutting and pasting the syntax in an R

console. The text in red represents parameters that should be set by the user in order to adapt the analysis to specific purposes. The tutorial demonstrates the efficiency of PACo and the pipeline to detect outlier associations and to test for congruence using the plastid (*mat*K, *trn*L–*trn*F, *trn*S–*trn*G) and nuclear-ribosomal (ITS) phylogenies of *Satyrium* Sw. (Orchidaceae), for which topological conflicts between trees derived from nuclear and plastid data sets have been reported (van der Niet and Linder 2008). We have made available separate chloroplast and nuclear derived posterior probability trees (Dryad repository, doi:10.5061/dryad.q6s1f) used throughout this tutorial, and a chloroplast-/nuclear concatenated alignment is available at TreeBASE (Study ID S1221).

RUNNING PROCEDURE

In addition to the basic R installation, five dedicated packages need to be installed to implement the pipeline, namely "ape", "cluster", "gplots", "phytools", and "vegan" (see http://cran.r-project.org/doc/manuals/R-admin.html#Installingpackages for details). For every running analysis, libraries required to execute the pipeline must be loaded.

```
library (ape)
library (cluster)
library (gplots)
library (phytools)
library (vegan)
```

PACo application

A complete description of PACo is provided by Balbuena et al (2013), and we refer to this study for details describing syntaxes of functions. To execute PACo and ParaFit, a set of functions have to be defined first. In both cases, the method proposed by de Vienne et al. (2011) is used to transform of patristic distances into Euclidean space.

```
PACo.dV <- function (H.dist, P.dist, HP.bin) {
  HP.bin <- which(HP.bin > 0, arr.in=TRUE)
  H.PCo <- pcoa(sqrt(H.dist), correction="none")$vectors
  P.PCo <- pcoa(sqrt(P.dist), correction="none")$vectors
  H.PCo <- H.PCo[HP.bin[,1],]
  P.PCo <- P.PCo[HP.bin[,2],]
  list (H.PCo = H.PCo, P.PCo = P.PCo)
}
```

The function *D.wrapper* will execute PACo and ParaFit for each of the trees included in the tree data sets (see below). It also allows the end-user to compare the influence of evolutionary distances in Procrustes and ParaFit analyses by executing PACo using either phylograms or unit branch length trees as input data. Unit branch length trees are obtained by computing branch lengths values of 1 to each branch of the tree data sets.

```
D.wrapper <- function(n) {
  DH.add <- cophenetic(treeH[[n]])
  DP.add <- cophenetic(treeP[[n]])
  DH.top <- cophenetic(compute.brlen(treeH[[n]], 1))
  DP.top <- cophenetic(compute.brlen(treeP[[n]], 1))
  DH.add <- DH.add[rownames(NCP),rownames(NCP)]
  DP.add <- DP.add[colnames(NCP), colnames(NCP)]
  DH.top <- DH.top[rownames(NCP),rownames(NCP)]
  DP.top <- DP.top[colnames(NCP), colnames(NCP)]

  PACo.add <- PACo.dV(DH.add, DP.add, HP)
  Proc.add <- procrustes(PACo.add$H.PCo, PACo.add$P.PCo)
  add.res <- residuals(Proc.add)
  HostX <- Proc.add$X
  ParY <- Proc.add$Yrot
  colnamesPACo <- paste(rownames(HostX),rownames(ParY), sep="_")

  PACo.top <- PACo.dV(DH.top, DP.top, HP)
  Proc.top <- procrustes(PACo.top$H.PCo, PACo.top$P.PCo)
  top.res <- residuals(Proc.top)

  PF.add <- parafit(sqrt(DH.add), sqrt(DP.add), HP, nperm=1,
test.links=TRUE, silent=TRUE)
  PFL2.add <- c(PF.add$link.table[,5])

  PF.top <- parafit(sqrt(DH.top), sqrt(DP.top), HP, nperm=1,
test.links=TRUE, silent=TRUE)
  PFL2.top <- c(PF.top$link.table[,5])

  write (add.res, file="PACo_res_add.txt", ncolumns = NLinks ,
append=TRUE, sep="\t")
  write (top.res, file="PACo_res_top.txt", ncolumns = NLinks ,
append=TRUE, sep="\t")
```

```
  write (PFL2.add, file="PFL2_add.txt", ncolumns = NLinks ,
append=TRUE, sep="\t")
  write (PFL2.top, file="PFL2_top.txt", ncolumns = NLinks ,
append=TRUE, sep="\t")
  write (colnamesPACo, "colnamesPACo.txt", ncolumns=NLinks,
sep="\t")
}
```

Data input

In order to execute the global test of congruence, two files must be loaded, namely consensus trees derived from the organellar and nuclear data sets. For example, the consensus trees produced by the MrBayes application are to be used in this step. In addition, a set of posterior probability trees obtained from Bayesian analysis or ML trees derived independently from the organellar and nuclear data sets are required for detection of outlier associations. Using a tree set and not consensus tree for outlier detection is preferred, because the former option allows for inclusion of phylogenetic uncertainty into the analysis. Trees may be uploaded in either Nexus or Newick format. A third file required to execute PACo and ParaFit is a binary matrix, in which corresponding pairs of organellar and nuclear Operational Taxonomic units (OTUs) are associated. However, this matrix is readily generated by the pipeline (see below) when both data sets share exactly the same number and names of OTUs. The user should ensure that sequence names in the binary association matrix match exactly with those of the trees. (Note also that the order of the taxa in the phylogenies should match with that of the binary matrix, but the pipeline includes a sorting algorithm to ensure this and no user intervention is required in this regard.) If data sets contain unequal numbers of sequences, then end-users must generate and upload the association matrix manually. Note that input files should include OTU labels that match exactly in all files, and we recommend the use of short name labels for the sake of the interpretation of graphical outputs. Use the following syntax to load trees in R:

```
NTree <- read.tree("myfilename.t")
CPTree <- read.tree("myfilename.t")
```

If input phylogenies are instead in Nexus format:

```
NTree <- read.nexus("myfilename.t")
CPTree <- read.nexus("myfilename.t")
```

For large data sets (e.g., trees with more than 200 OTUs), manual generation of the binary association matrix comprising organellar and nuclear OTUs can be time-consuming. The binary matrix can be generated by the following code:

```
NTaxa <- sort(NTree$tip.label)
CPTaxa <- sort(CPTree$tip.label)
NCP <- as.matrix(table(NTaxa, CPTaxa))
```

However, if small trees (e.g., trees with less than 50 OTUs) are being analyzed, or if the user already has a text file with the association matrix, it can be loaded into R:

```
NCP <- as.matrix(read.table("myfilename.txt", header=TRUE))
```

In order to accommodate for phylogenetic uncertainty into the analysis, a sets of trees in either Nexus or Newick format is required for detection of outlier sequences (see above):

```
ByH <- "myfilename.t"
ByP <- "myfilename.t"
```

Trees in Newick format

```
treeH <- read.tree(file= ByH)
treeP <- read.tree(file= ByP)
```

Trees in Nexus format

```
treeH <- read.nexus(file= ByH)
treeP <- read.nexus(file= ByP)
```

Using the following script, the end-user may set a given number of trees to be discarded (burn-in) from the tree data set, in this example the first 18,000 trees are discarded:

```
treeH <- treeH[18001: length(treeH)]
treeP <- treeP[18001: length(treeP)]

NLinks = sum(NCP)
HP <- diag(NLinks)
```

Testing cophylogeny between nuclear and chloroplast phylogenies

To execute the global test of congruence between organellar and nuclear data sets, PACo requires patristic distances to obtain a global m^2_{XY} value. Therefore, consensus organellar and nuclear trees (see data input) must be transformed into matrices of patristic distances:

```
N.D <- cophenetic (NTree)
CP.D <- cophenetic (CPTree)
```

The organellar and nuclear matrices of patristic distances are then sorted to match the rows and the columns of the binary association matrix:

```
N.D <- N.D[rownames(NCP),rownames(NCP)]
CP.D <- CP.D [colnames(NCP), colnames(NCP)]
```

Finally, to apply PACo:

```
PACo.fit <- PACo.dV(N.D, CP.D, NCP)
NCP.proc <- procrustes(PACo.fit$H.PCo, PACo.fit$P.PCo)
```

The following syntax computes the residual sum of squares m^2_{XY} and randomizes the 'o'-/'n'-association matrix to determine, whether the probability p under H_o ('similarity between trees not higher than expected by chance', see main text) is rejected. The user must set a number of random permutations of the organelle-/host nucleus-matrix. Although we employed 100,000 in all analyses, a number ≤ 10,000 should be sufficient to obtain comparable results.

```
m2.obs <- NCP.proc$ss
N.perm = 10000
P.value = 0
set.seed(2)
for (n in c(1:N.perm))
{
  if (NLinks <= nrow(NCP) | NLinks <= ncol(NCP))
  {  flag2 <- TRUE
     while (flag2 == TRUE)  {
       NCP.perm <- t(apply(NCP,1,sample))
       if(any(colSums(NCP.perm) == NLinks)) flag2 <- TRUE else
flag2 <- FALSE
     }
  } else { NCP.perm <- t(apply(NCP,1,sample))}
```

```
  PACo.perm <- PACo.dV(N.D, CP.D, NCP.perm)
  m2.perm <- procrustes(PACo.perm$H.PCo, PACo.perm$P.PCo)$ss
  if (m2.perm <= m2.obs)
  {P.value = P.value + 1}
}
P.value <- P.value/N.perm

cat(" The observed m2 is ", m2.obs, "\n", "P-value = ", P.value,
" based on ", N.perm," permutations.")
```

Note that `set.seed(2)` sets a reproducible set of test permutations. Changing the integer value will produce a different set, but should not change the *p* value substantially. R will print out the *p* value and m^2_{XY}:

```
The observed m2 is 0.4655883
P-value = 0.0001 based on 1000 permutations.
```

Thus, the significance value at which H_0 is rejected is 0.0001. This shows that, despite the presence of outliers in the phylogenies, organellar and nuclear data sets in *Satyrium* reflect cophylogeny to some degree.

Detecting outlier associations

The contribution (e^2_i) to the global squared residual value (m^2_{XY}) and the $pfl2_i$ (see methods) of each association, using phylograms and unit branch length trees is computed using:

```
lapply(1:length(treeH), D.wrapper)
```

At execution, tables containing e^2_i and $pfl2_i$ values for each association (for both PACo and ParaFit analyses using phylograms and unit branch length trees) will be generated and saved in your working directory (files `PACo_res_add.txt`, `PACo_res_top.txt`, `PFL2_add.txt` and `PFL2_top.txt`). These tables are required by the pipeline (see below) to spot outlier sequences onto the phylogenies and can be loaded onto the workspace:

```
colnamesPACo <- read.table(file="colnamesPACo.txt", header=TRUE)
colnamesPACo <- colnames(colnamesPACo)
```

```
pac.add <- read.table(file="PACo_res_add.txt", header=FALSE,
col.names=colnamesPACo)
pac.top <- read.table(file="PACo_res_top.txt", header=FALSE,
col.names=colnamesPACo)
pf2.add <- read.table(file="PFL2_add.txt", header=FALSE,
col.names=colnamesPACo)
pf2.top <- read.table(file="PFL2_top.txt", header=FALSE,
col.names=colnamesPACo)
```

Next, outlier associations will be spotted by the pipeline using a threshold value (1/N). The following syntax will transform the e_i^2's into ε_i^2's obtained from either phylograms or unit branch length trees and will compute their respective median. Given the asymmetric distribution of the ε_i^2's, the median value was preferred over the mean as central tendency estimate:

```
m2A <- apply(pac.add, 1, sum)
pac.norm.add <- pac.add/m2A

m2T <- apply(pac.top, 1, sum)
pac.norm.top <- pac.top/m2T
```

To plot the median ε_i^2 and its 95% empirical confidence intervals obtained from sequences in phylograms and unit branch lengths, and to spot outlier taxa according to the threshold value (1/N), use the following script:

```
op <- par(oma=c(3,2,1,1))
par (mfrow=c(1,1),mar = c(4,4,1,1))

mA <- apply(pac.norm.add, 2, median)
uCI.A <- apply(pac.norm.add, 2, quantile, probs = 0.975)
lCI.A <- apply(pac.norm.add, 2, quantile, probs = 0.025)
cols <- c("lightgreen", "mistyrose")[(mA > 1/NLinks) + 1]
barplot2(mA, main = "PAco squared residuals - additive trees",
xlab="Association", ylab="Normalized PACo sqr. residuals",
        cex.axis=0.5, col=cols, border="lightgrey",
names.arg=colnamesPACo, las=2, cex.names=0.5, plot.ci=T,
ci.l=lCI.A, ci.u=uCI.A, ci.color="blue")
abline(h=1/NLinks, col="red")

mA <- apply(pac.norm.top, 2, median)
uCI.A <- apply(pac.norm.top, 2, quantile, probs = 0.975)
```

```
lCI.A <- apply(pac.norm.top, 2, quantile, probs = 0.025)
cols <- c("lightgreen", "mistyrose")[(mA > 1/NLinks) + 1]
barplot2(mA, main = "PAco squared residuals - unit branch length
trees", xlab="Association", ylab="Normalized PACo
sqr.residuals", cex.axis=0.5, col=cols, border="lightgrey",
names.arg=colnamesPACo, las=2, cex.names=0.5, plot.ci=T,
ci.l=lCI.A, ci.u=uCI.A, ci.color="blue")
abline(h=1/NLinks, col="red")
```

Two plots (Fig. S3, data with ε_i^2's obtained from unit branch length trees not shown) of all squared residual values determined from each 'o'-/'n'-association, and obtained from phylograms and unit branch lengths as well, will be plotted, respectively. Associations with ε_i^2's scores above the red line (i.e., 1/N threshold value) represent putative outlier sequences especially, if the lower bound of the associated 95% confidence interval is above the threshold. In the working example of *Satyrium*, 15 'o'-/'n'-associations were retrieved as outlier (Fig. S3). Eleven of such links presented indeed contrasting phylogenetic positions on chloroplast and nuclear trees (red bars in Fig. S3). All outlier associations detected by PACo as potentially outliers are shown in Figure S4. Names in red correspond to associations retrieved by PACo that are true outliers, whereas names in black are associations identified by PACo as potential outliers, even though they did not recover conflicting phylogenetic positions. In our simulations and real data set analyses *pfl2*$_i$ yielded suboptimal results, but the user may also wish to plot the *pfl2*$_i$'s for comparative purposes:

```
mA <- apply(pf2.add, 2, median)
uCI.A <- apply(pf2.add, 2, quantile, probs = 0.975)
lCI.A <- apply(pf2.add, 2, quantile, probs = 0.025)
cols <- c("lightgreen", "mistyrose")[(mA > 0) + 1]
barplot2(mA, main = "pfl2 statistic - additive trees",
xlab="Association", ylab="Normalized PACo sqr. residuals",
          cex.axis=0.5,       col=cols,       border="lightgrey",
names.arg=colnamesPACo,    las=2,    cex.names=0.5,    plot.ci=T,
ci.l=lCI.A,
          ci.u=uCI.A, ci.color="blue")
abline(h=0, col="red")

mA <- apply(pf2.top, 2,  median)
uCI.A <- apply(pf2.top, 2, quantile, probs = 0.975)
lCI.A <- apply(pf2.top, 2, quantile, probs = 0.025)
cols <- c("lightgreen", "mistyrose")[(mA > 0) + 1]
```

```
barplot2(mA, main = "pfl2 statistic - unit branch length trees",
xlab="Association", ylab="Normalized PACo sqr. residuals",
cex.axis=0.5, col=cols, border="lightgrey",
names.arg=colnamesPACo, las=2, cex.names=0.5, plot.ci=T,
ci.l=lCI.A, ci.u=uCI.A, ci.color="blue")
abline(h=0, col="red")
```

Validating classifications of outlier and congruent terminals with PAM

Cluster analysis using the Partition Around Medoids (PAM) algorithm (Kaufman and Rousseeuw 1990) allows the end-user to determine the extent of properly classified associations into outlier or congruent OTUs in relation to the total number of OTUs and the proportion of outlier/congruent OTUs. Our pipeline offers two alternatives to carry out clustering analyses, namely 1) using median ε_i^2 and $pfl2_i$ values combined and 2) using median ε_i^2's alone. Our simulations and real data set analyses show that the latter strategy yields stronger cluster structures, but comparison between the two approaches can still be useful to reveal doubtful associations. Clustering starts by standardizing both statistics (ε_i^2 and $pfl2_i$):

```
sum.pac.add <- apply(pac.add, 1, sum)
pac.add <- pac.add/sum.pac.add - 1/NLinks
sum.pac.top <- apply(pac.top, 1, sum)
pac.top <- pac.top/sum.pac.top - 1/NLinks

im.paco.add <- apply(pac.add, 2, median)
im.paco.top <- apply(pac.top, 2, median)
im.pf2.add <- apply(pf2.add, 2, median)
im.pf2.top <- apply(pf2.top, 2, median)

x.paco.add <- mean(im.paco.add) ; x.pf2.add <- mean(im.pf2.add)
sd.paco.add<- sd(im.paco.add)  ; sd.pf2.add <- sd(im.pf2.add)
im.paco.stadd <- (x.paco.add - im.paco.add)/sd.paco.add
im.pf2.stadd <- (x.pf2.add - im.pf2.add)/sd.pf2.add
metrics.stadd <- data.frame(im.paco.stadd, im.pf2.stadd)

x.paco.top <- mean(im.paco.top) ; x.pf2.top <- mean(im.pf2.top)
sd.paco.top <- sd(im.paco.top)  ; sd.pf2.top <- sd(im.pf2.top)
im.paco.sttop <- (x.paco.top - im.paco.top)/sd.paco.top
im.pf2.sttop <- (x.pf2.top - im.pf2.top)/sd.pf2.top
metrics.sttop <- data.frame(im.paco.sttop, im.pf2.sttop)
```

The user must specify the number of clusters (*k*). Initially, one should set *k*=2, as PAM is expected to separate the 'o'-/'n'-associations into non-conflicting and outlier. However, in some situations *pfl2* tends to split non-conflicting associations into two unnatural clusters, and *k* has to be set to 3 in order to retrieve the group of outlier associations.

```
nclust = my k
```

To apply clustering analysis using PACo in combination with *pfl2* with both phylograms and unit branch length trees use the following commands:

```
par (mfrow=c(2,1))
K.PAM <- pam(metrics.stadd, nclust, diss=FALSE)
plot(im.paco.add,im.pf2.add,
col=c("red","blue")[K.PAM$clustering])
title(main=list("PACo-Parafit - additive trees", cex=0.8))
SPaPf.add <- silhouette(K.PAM)
cat(summary(SPaPf.add)$avg.width)
SPaPf.add  <- summary(SPaPf.add)$avg.width
cat("\n")

K.PAM <- pam(metrics.sttop, nclust, diss=FALSE)
plot(im.paco.top,im.pf2.top,
col=c("red","blue")[K.PAM$clustering])
title(main=list("PACo-pf2 - unit branch length trees", cex=0.8))
SPaPf.top <- silhouette(K.PAM)
cat(summary(SPaPf.top)$avg.width)
SPaPf.top  <- summary(SPaPf.top)$avg.width
cat("\n")
```

In contrast, the end-user might want to apply clustering analysis using solely PACo with phylograms and unit branch length trees:

```
K.PAM <- pam(metrics.stadd[1], nclust, diss=FALSE)
plot(im.paco.add,im.pf2.add,
col=c("red","blue")[K.PAM$clustering])
title(main=list("PACo + additive trees", cex=0.8))
SPa.add <- silhouette(K.PAM)
cat(summary(SPa.add)$avg.width)
SPa.add <- summary(SPa.add)$avg.width
cat("\n")
```

```
K.PAM <- pam(metrics.sttop[1], nclust, diss=FALSE)
plot(im.paco.top,im.pf2.top,
col=c("red","blue")[K.PAM$clustering])
title(main=list("PACo - unit branch length trees", cex=0.8))
SPa.top <- silhouette(K.PAM)
cat(summary(SPa.top)$avg.width)
SPa.top <- summary(SPa.top)$avg.width
cat("\n")
```

All silhouette values from all clustering analysis on a single table can also be save on disk:

```
Sall <- rbind(SPaPf.add, SPa.add, SPaPf.top, SPa.top)
rownames(Sall) <- c("Silhouette PACo-Parafit additive",
"Silhouette PACo additive", "Silhouette PAco-Parafit unit branch
length", "Silhouette PACo unit branch length ")
write.table(Sall, "Silhouette_values_all.txt")
```

Spotting outlier associations on trees

In order to allow the end-user a better representation of potential outlier associations on trees, our pipeline finally produces a cophylogenetic plot of organellar and nuclear trees with outlier OTUs directly labeled on trees by means of a color scale:

```
op <- par(oma=c(1,1,1,1))
par (mfrow=c(1,2),mar = c(1,1,1,1))

mA <- apply(pac.norm.add, 2, median)
mA[mA > 1/NLinks] <- 1
mA[mA < 1/NLinks] <- 0
mA <- as.data.frame(mA)
out <- mA$mA
names(out) <- NTree$tip.label
out

plotTree(NTree, setEnv = T, offset=0.5, fsize=0.5, lwd=1)
title(main="Nuclear tree of Gene 1 - PACo potential conflicting
associations", font.main=1, cex.main=0.8)
tiplabels(pie = to.matrix(out, sort(unique(out))), piecol =
c("lightgreen", "lightcoral"), cex = 0.5)
legend("bottomleft", c("Congruent", "Conflicting"),
```

```
        cex=0.9, pch=16, col=c("lightgreen", "lightcoral"))

plotTree(CPTree, setEnv = T, offset=0.5, fsize=0.5, lwd=1)
title(main="Chloroplast tree of Gene 2 - PACo potential
conflicting associations", font.main=1, cex.main=0.8)
tiplabels(pie = to.matrix(out, sort(unique(out))), piecol =
c("lightgreen", "lightcoral"), cex = 0.5)
```

This script will plot the consensus trees of each data set analyzed, with the corresponding OTUs names. Their individual ε_i^2 scores are color-coded according to their values (conflicting or congruent). The color scale can be bespoke, by replacing the argument `"piecol"` with any alternative allowed by the function. In the working example (results with unit branch length trees not shown), the cophylogenetic plot of the consensus chloroplast and nuclear trees, together with their color-coded ε_i^2 scores (Fig. S5), largely reflects the results observed in the confidence interval plot (Fig. S3). The script also allows to easily spot outlier OTUs in large phylogenies (see Figs S6, S7 for a barplot with PACo squared residual values and plotted simulated trees of 200 OTUs showing outlier associations highlighted by PACo as potential outliers, respectively).

LITERATURE CITED

Baldwin B.G. 1992. Phylogenetic utility of the Internal Transciber Spacers of nuclear ribosomal DNA in plants: An example from the Compositae. Mol. Phylogenet. Evol. 1: 3–16.

De Vienne D.M., Ollier S., Aguileta G. 2012. Phylo-MCOA: A fast and effective method to detect outlier genes and species in phylogenomics using multiple co-inertia analysis. Mol. Biol. Evol. 29: 1587–1598.

Baldwin B.G., Markos S. 1998. Phylogenetic utility of the External Transcriber Spacer (ETS) of 18S-26S rDNA: Congurence of ETS and ITS trees of *Calycadenia* (Compositae). Mol. Phylogenet. Evol. 10: 449–463.

Górniak M., Paun O., Chase M.W. 2010. Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, *Xdh*: Congruence with organellar and nuclear ribosomal DNA results. Mol. Phylogenet. Evol. 56: 784–795.

Hamilton M.B. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. Mol. Ecol. 8: 521–523.

Kaufman L., Rousseeuw P.J. 1990. Finding groups in data: An introduction to cluster analysis. Wiley, New York.

Legendre P., Desdevises Y., Bazin E. 2002. A statistical test for host-parasite coevolution. Syst. Biol. 51: 217–234.

Monteiro S.H., Selbach-Schnadelbach A., de Oliveira R.P., van den Berg C. 2010. Molecular phylogenetics of *Galeandra* (Orchidaceae: Catasetinae) based on plastid and nuclear DNA sequences. Syst. Bot. 35: 476–486.

Neubig K.M., Whitten W.M., Carlsward B.S., Blanco M.A., Endara L., Williams N.H., Moore M. 2009. Phylogenetic utility of *ycf*1 in orchids: A plastid gene more variable than *mat*K. Plant Syst. Evol. 277: 75–84.

R Development Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

van der Niet T., Linder H.P. 2008. Dealing with incongruence in the quest of the species tree: a case of study from the orchid genus *Satyrium*. Mol. Phylogenet. Evol. 47:154–174.

FIGURES

Figure S1. Vector diagrams of squared residual values ε_i^2 and ParaFitLink2 statistic (*pfl2*) obtained by PACo and ParaFit, respectively, using simulated additive trees. Vector magnitude and orientation are related to the topological degree congruence of each 'o'-/'n'-association. Outlier associations are shown in red and non-conflicting in black. Trees with 50 terminals including a) 5 outliers (10%); b) 10 outliers (20%); c) 15 outliers (30%); d) 20 outliers (40%); with 100 terminals including e) 10% outliers; f) 20% outliers; g) 30% outliers; h) 40% outliers; with 200 terminals including i) 10% outliers; j) 20% outliers; k) 30% outliers; l) 40% outliers.

Figure S2. Vector diagrams of squared residual values ε_i^2 and ParaFitLink2 statistic (*pfl2*) using simulated unit branch length trees. Vector magnitude and orientation are related to the topological degree congruence of each 'o'-/'n'-association. Outlier associations are shown in red, non-conflicting in black. Trees with 50 terminals including a) 5 outliers (10%); b) 10 outliers (20%); c) 15 outliers (30%); d) 20 outliers (40%); with 100 terminals including e) 10% outliers; f) 20% outliers; g) 30% outliers; h) 40% outliers; with 200 terminals including i) 10% outliers; j) 20% outliers; k) 30% outliers; l) 40% outliers.

Figure S3. Normalized squared residual values ε_i^2) of individual 'o'-/'n'-associations obtained by PACo using additive trees. Pink bars indicate potential outlier associations identified by the pipeline. Taxa names in black, bold, and underlined represent OTUs retrieved by PACo that do not actually demonstrate phylogenetic distortion as in truly outlier associations.

Figure S4. Cophylogenetic plot showing the nuclear (ITS, left) and chloroplast (*matK*, *trnL–trnF*, right) phylogenies of *Satyrium*. Bayesian posterior probabilities > 0.95 are shown above corresponding branches. Terminals in red, bold, and underlined represent associations identified by PACo as outliers that are indeed conflicting sequences. Terminals in black, bold, and underlined represent associations retrieved by PACo that do not actually demonstrate phylogenetic distortion as in truly conflicting associations.

Figure S5. Cophylogenetic plot of nuclear (right) and chloroplast (left) trees of *Satyrium* showing outlier associations detected by PACo. Scale-color (bottom left) correspond to squared residual values ε_i^2 of individual 'o'-/'n'-associations. Potential outlier associations are indicated in purple, blue and light blue (see cutoff value 0.024 in Fig. S4).

Figure S6. Normalized squared residual values ε_i^2 of individual associations obtained by PACo using simulated additive trees of 200 terminals, which 20% of those are conflicting. Pink bars indicate potential outlier associations identified by the pipeline, whereas light-green bars represent non-conflicting associations.

Figure S7. Cophylogenetic plot of two simulated gene trees showing outlier associations detected by PACo. Red circles on tips correspond to potential outliers, whose squared residual values ε_i^2 of individual associations are higher than the cutoff value (1/N). Non-conflicting associations are indicated in light-green circles.

TABLES

Table S1. Primers and PCR settings used for amplifying chloroplast and nuclear DNA loci.

Table S2. Species names and voucher information for material used in this study. Taxa sequenced in this study are indicated in bold letters.

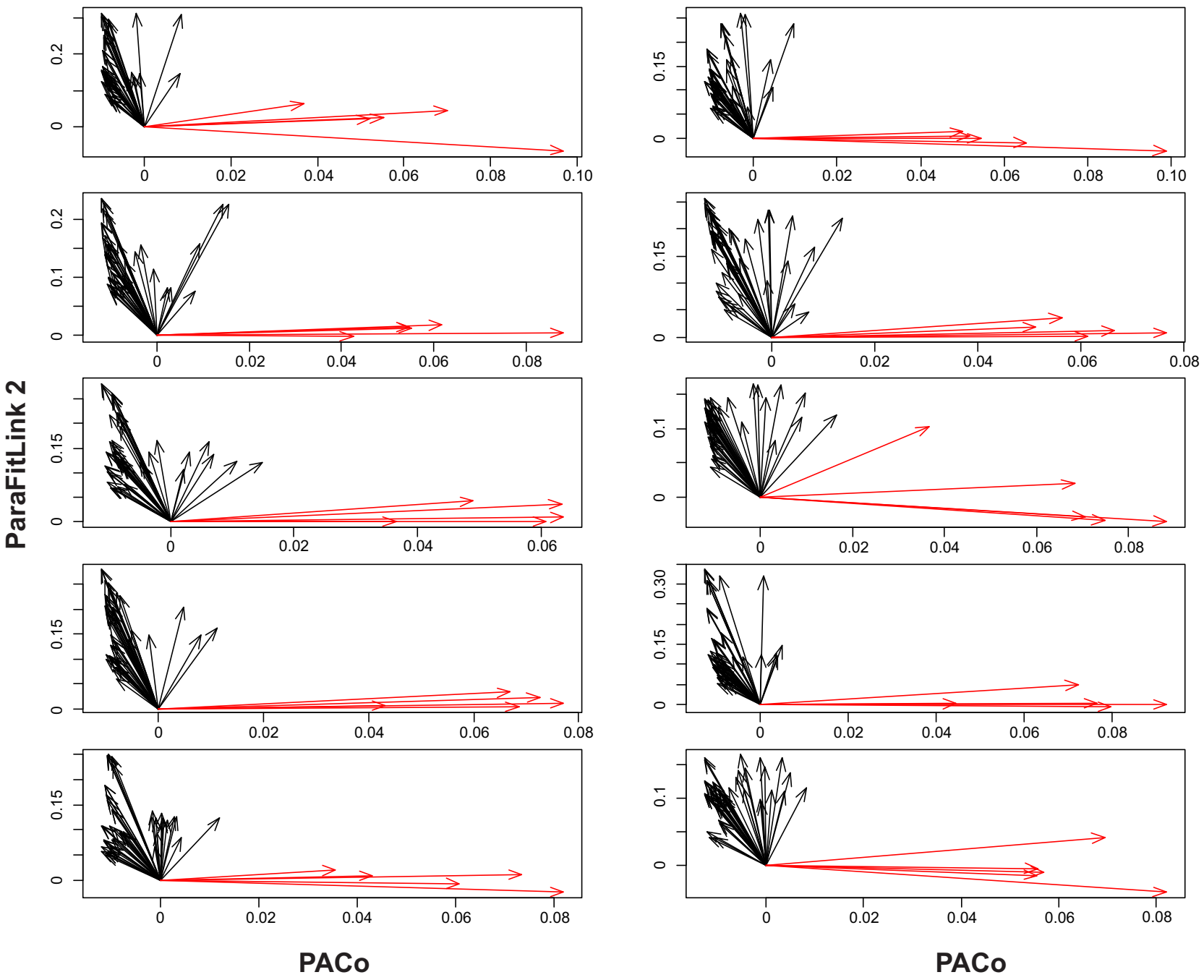
Table S3. Results of jModel test.

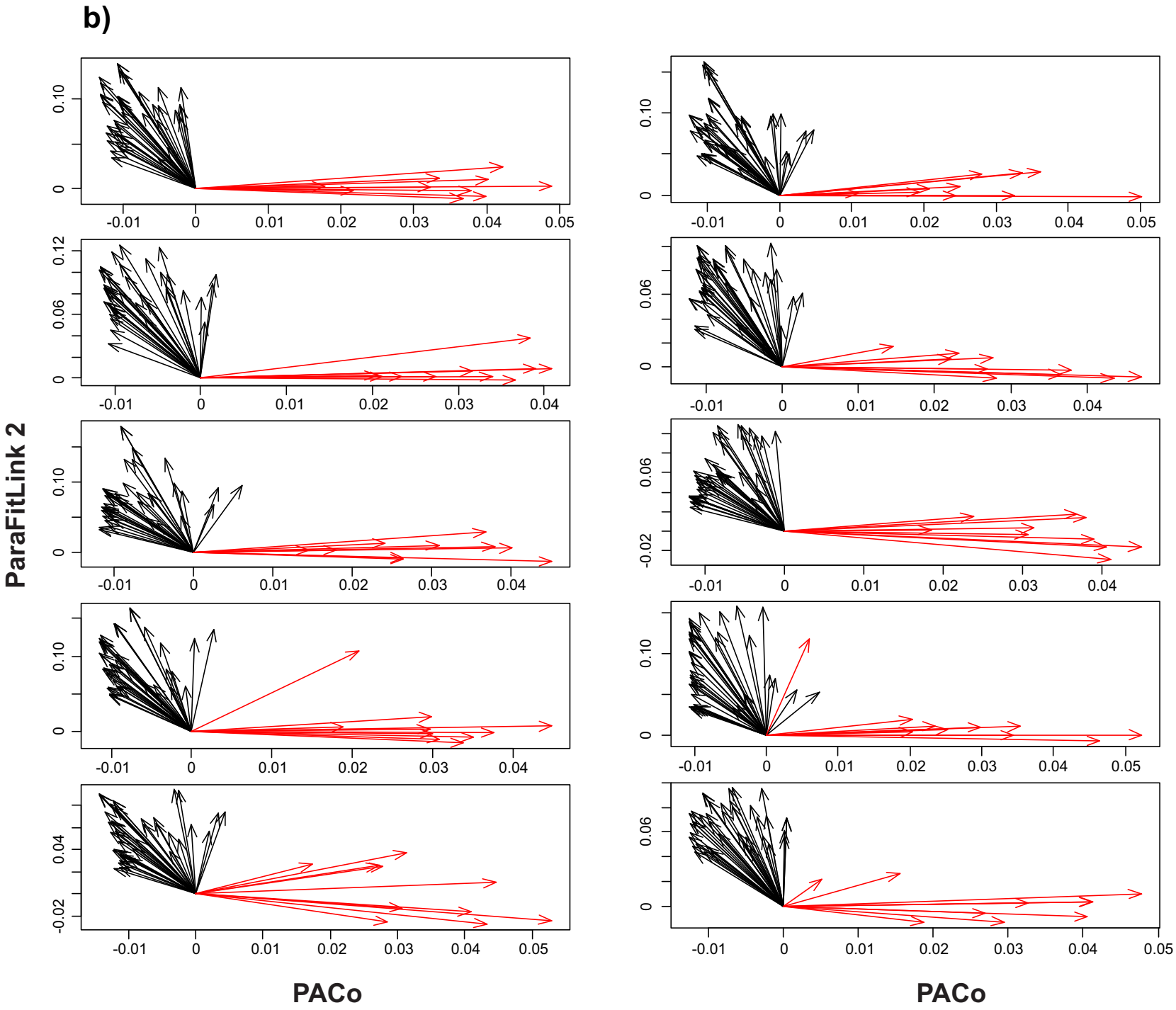
Table S4. Number of misclassified congruent ('c') and outlier ('x') associations in 10 pairs of simulated additive and unit branch length gene trees based on the median values of PACo and ParaFitLink2 (PFL2) statistics using the Partionioning Around Medioids algorithm (PAM). Trees were simulated with a) 50, b) 100 and c) 200 and a corresponding number of 10%, 20%, 30% and 40% of outlier OTUs, respectively. For each pair of trees, PACo and ParaFit were applied to 1000 sets of post burn-in trees obtained from Bayesian inferences by computing median statistics. PAM was applied for separation between 'c' and 'o' links using PACo in combination with ParaFit, or only the PACo statistic. Values of the average silhouette width (S) for each tree are also reported, as well as the total number of misidentified associations (Mis.T) and Average Silhouette width value (Av.S). Boldfaced values correspond to cases where the PAM algorithm required k=3 to separate 'x' associations, given that PFL2 tended to separate 'c' associations into two artificial clusters.

Table S5. Alignment characterization.

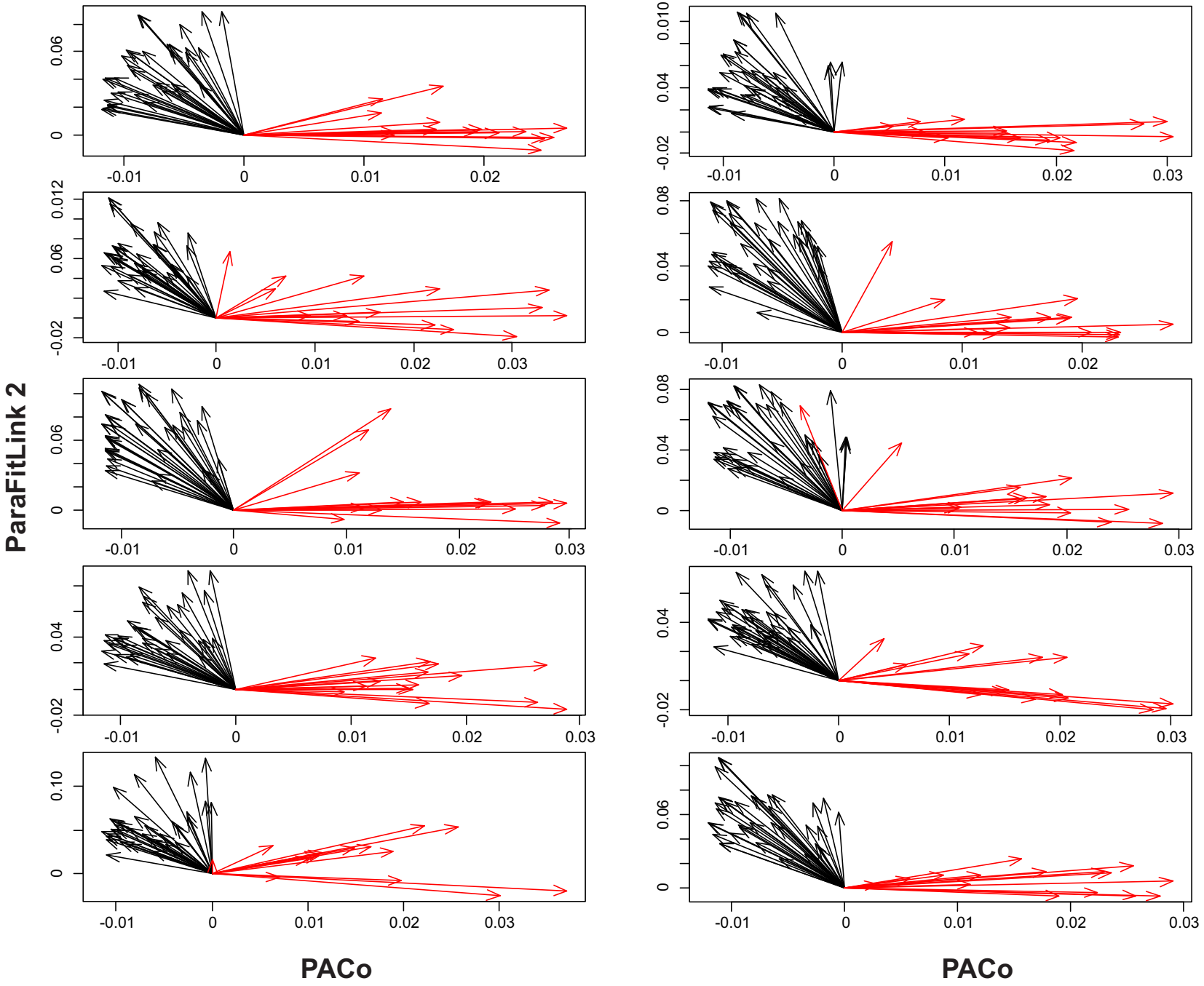
Fig. S1

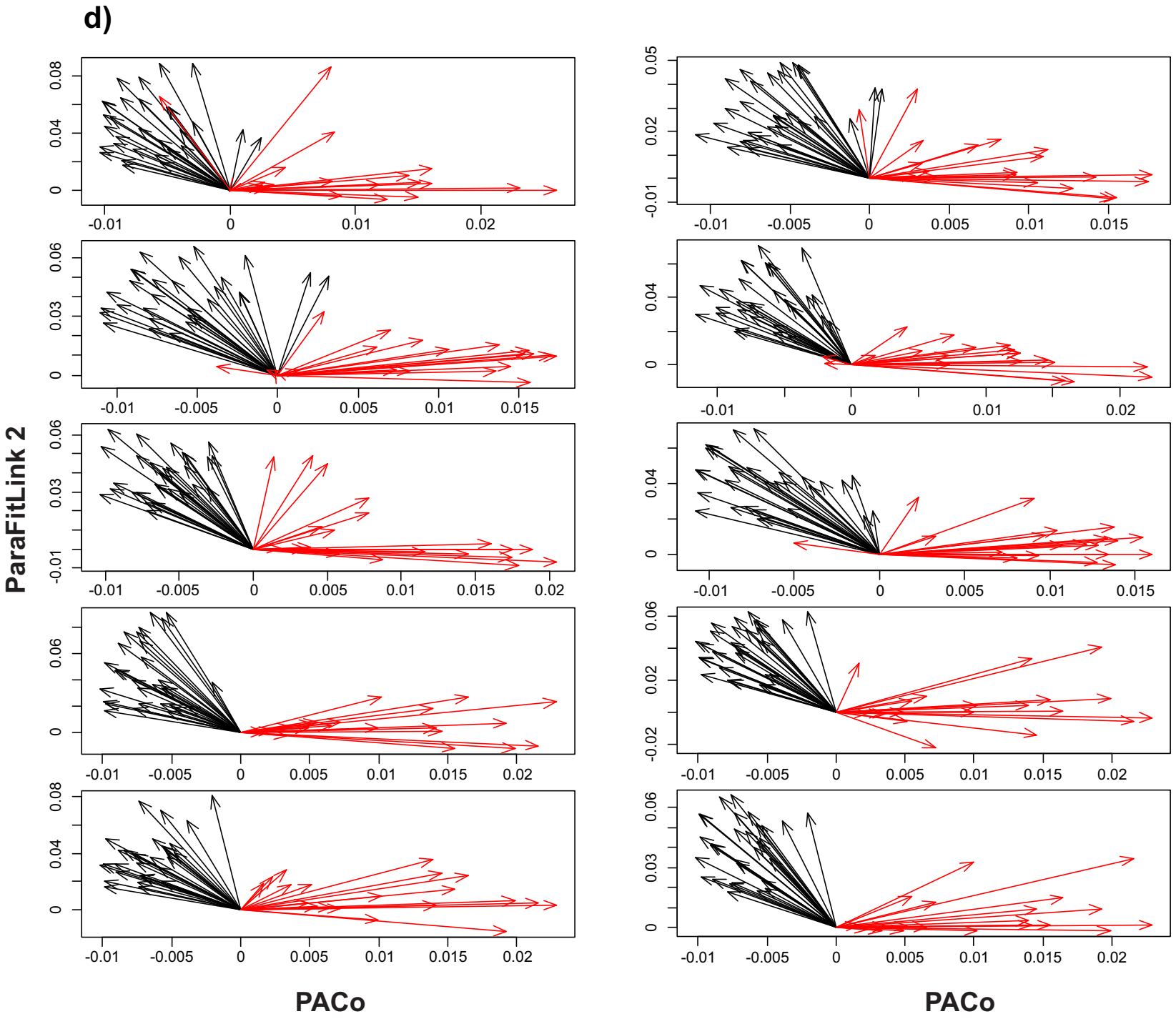
a)

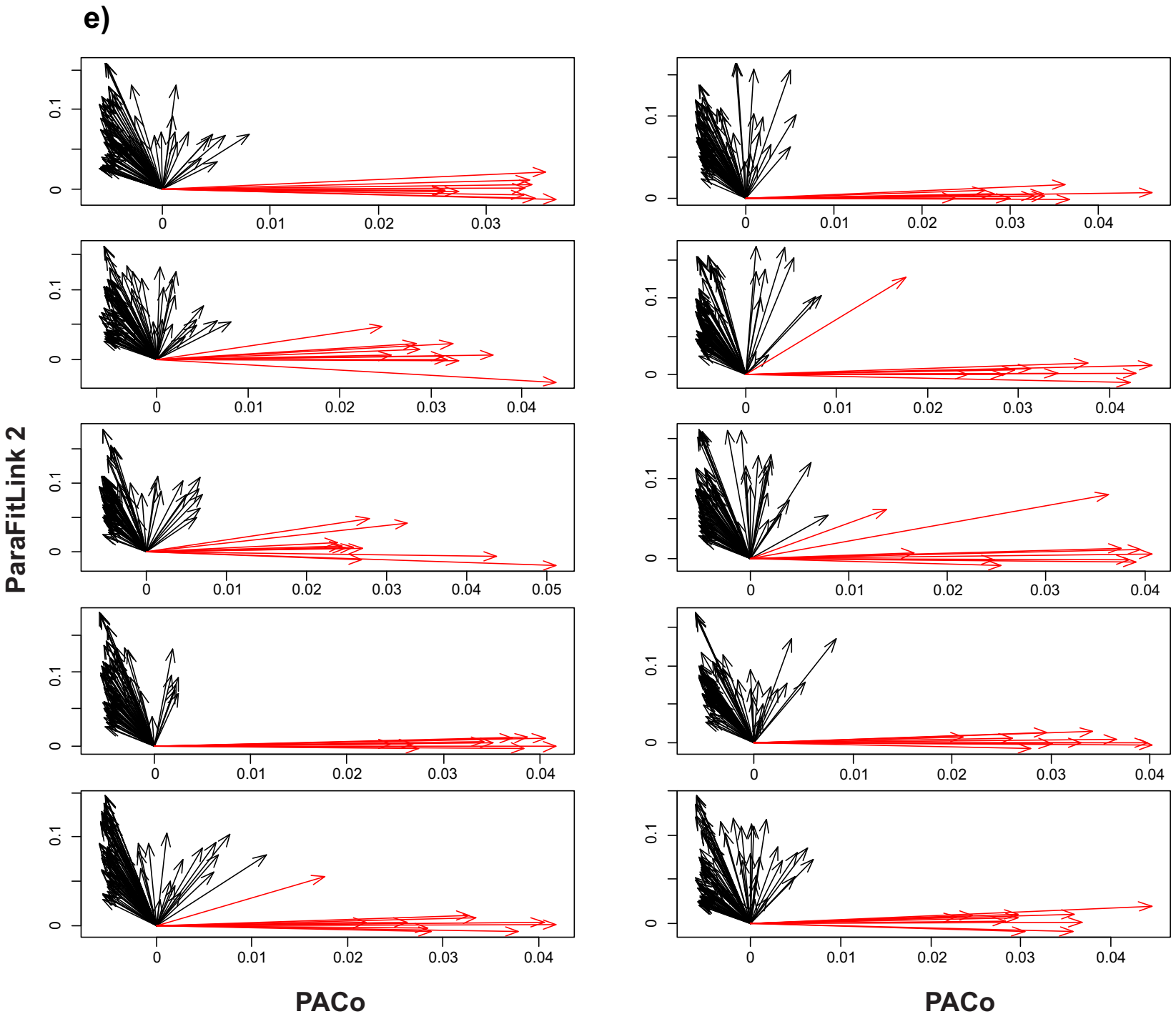



b)
ParaFitLink 2
PACo
PACo

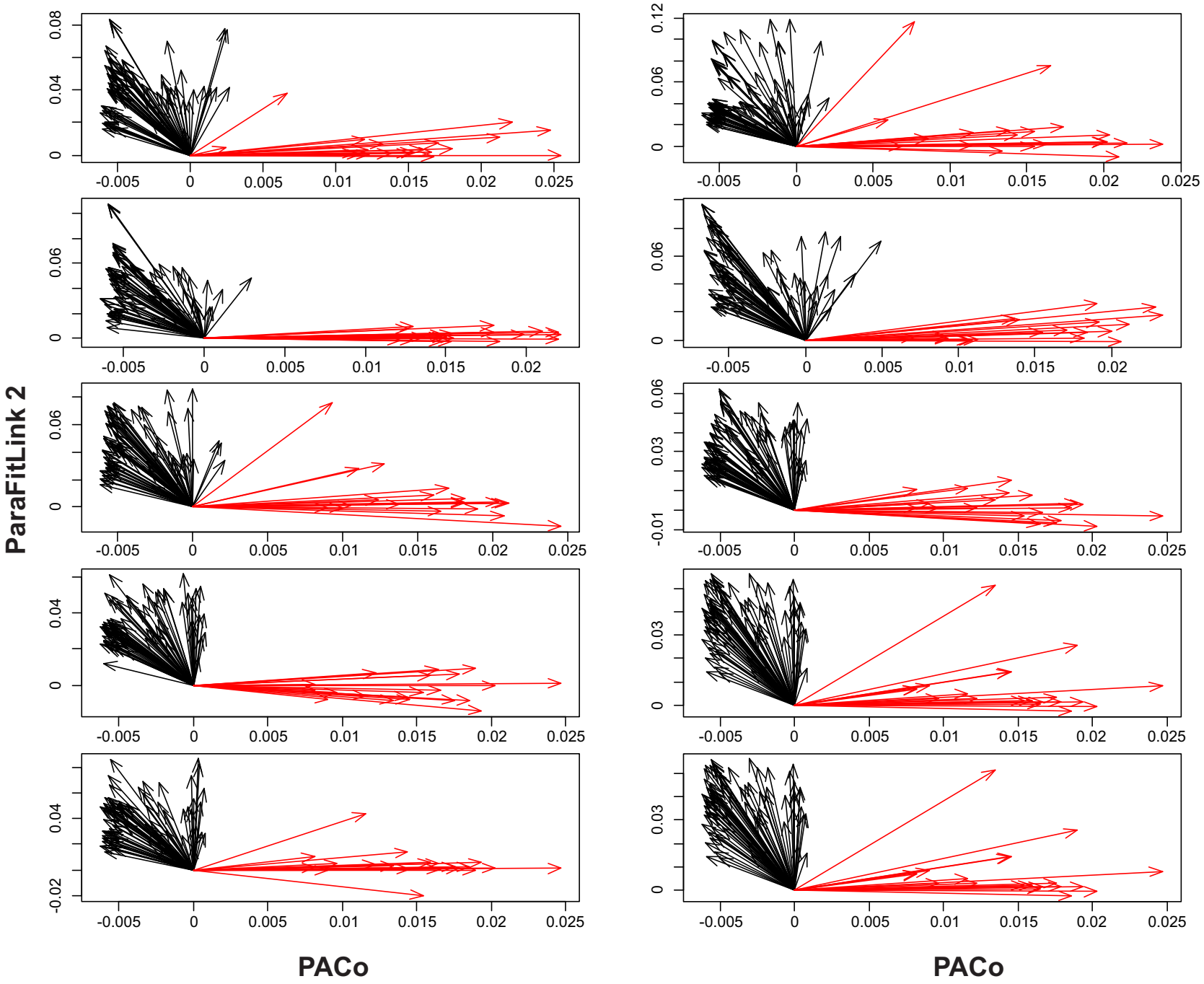
c)



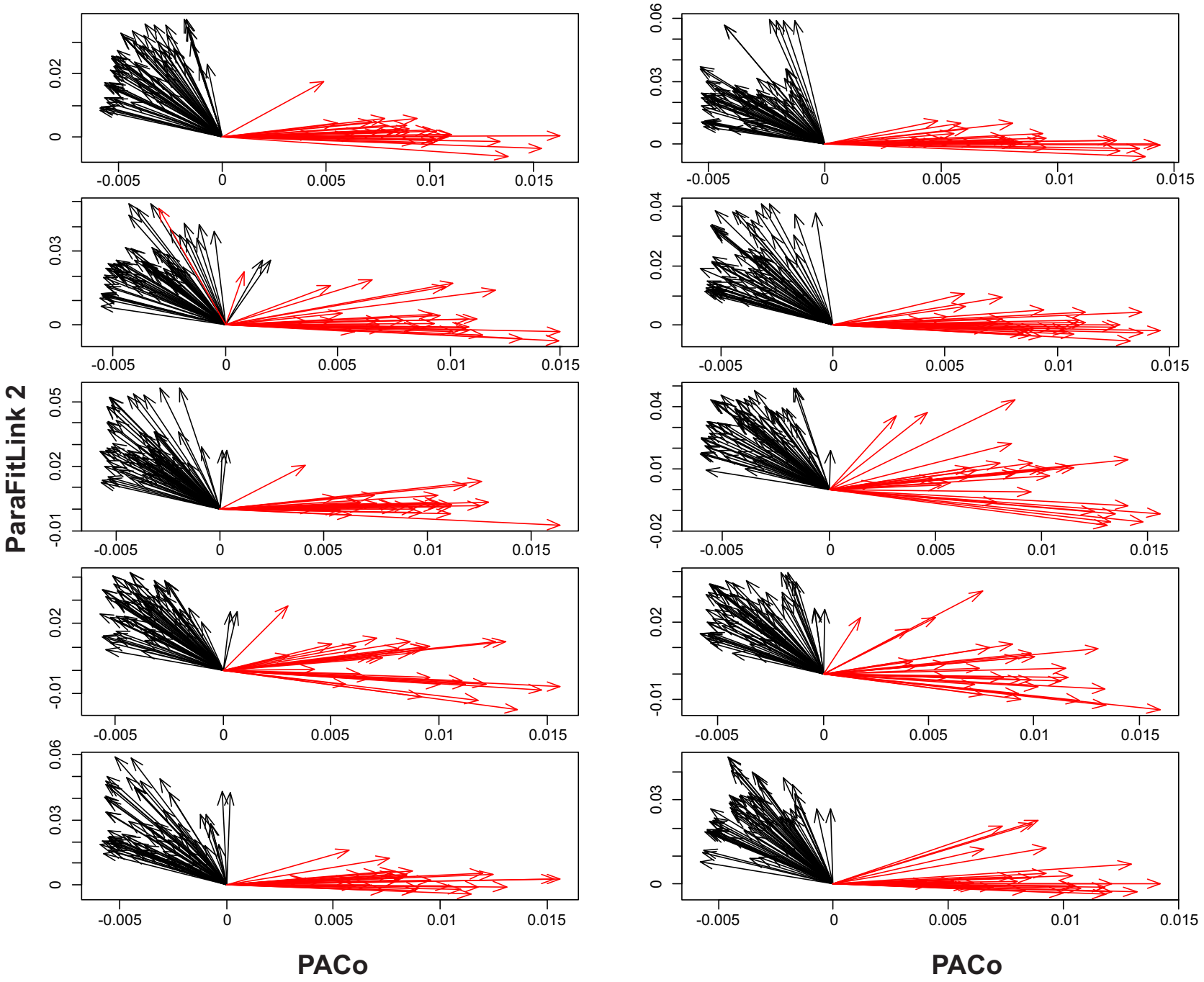

d)
ParaFitLink 2
PACo
PACo


e)
ParaFitLink 2
PACo
PACo

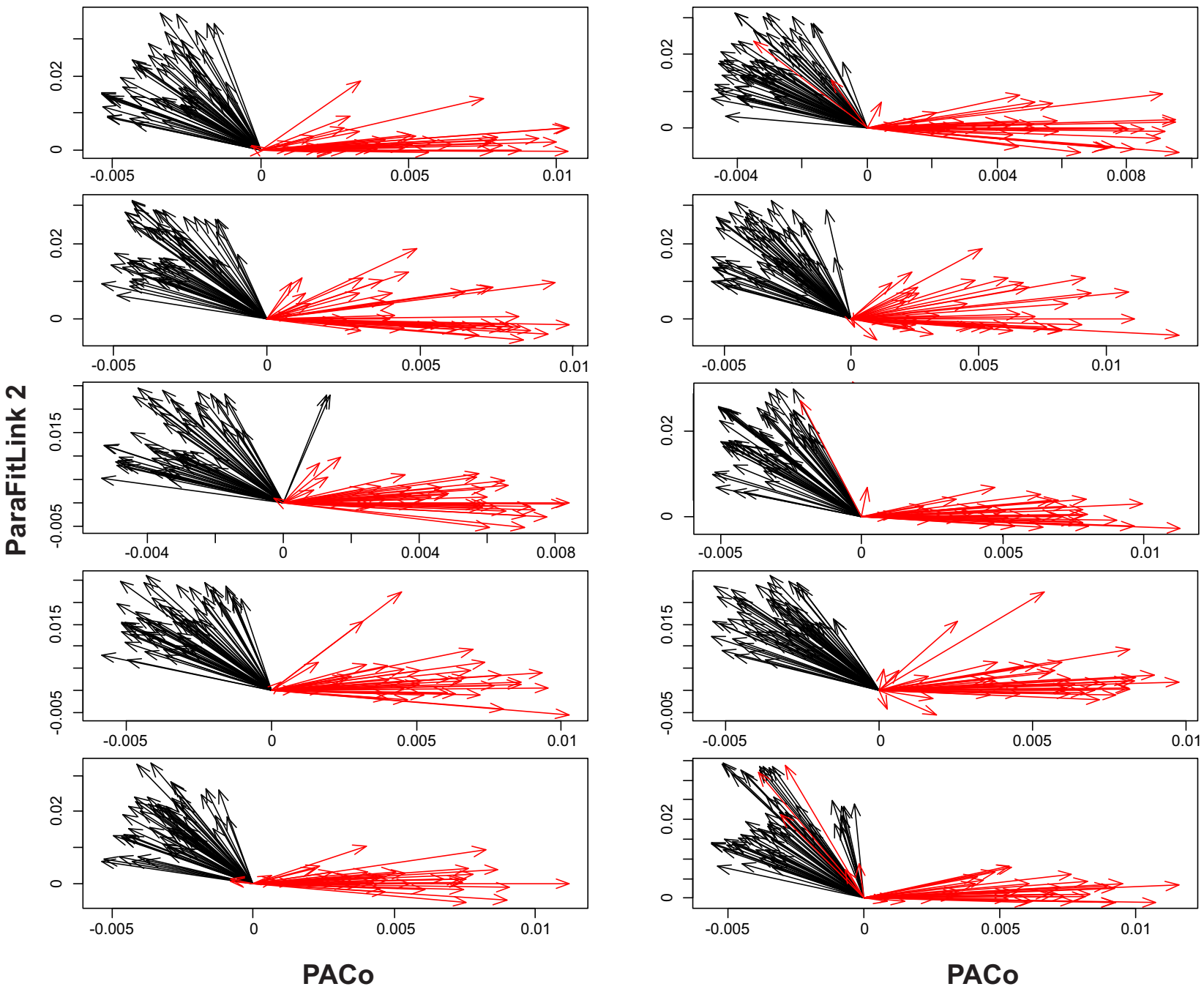
f)

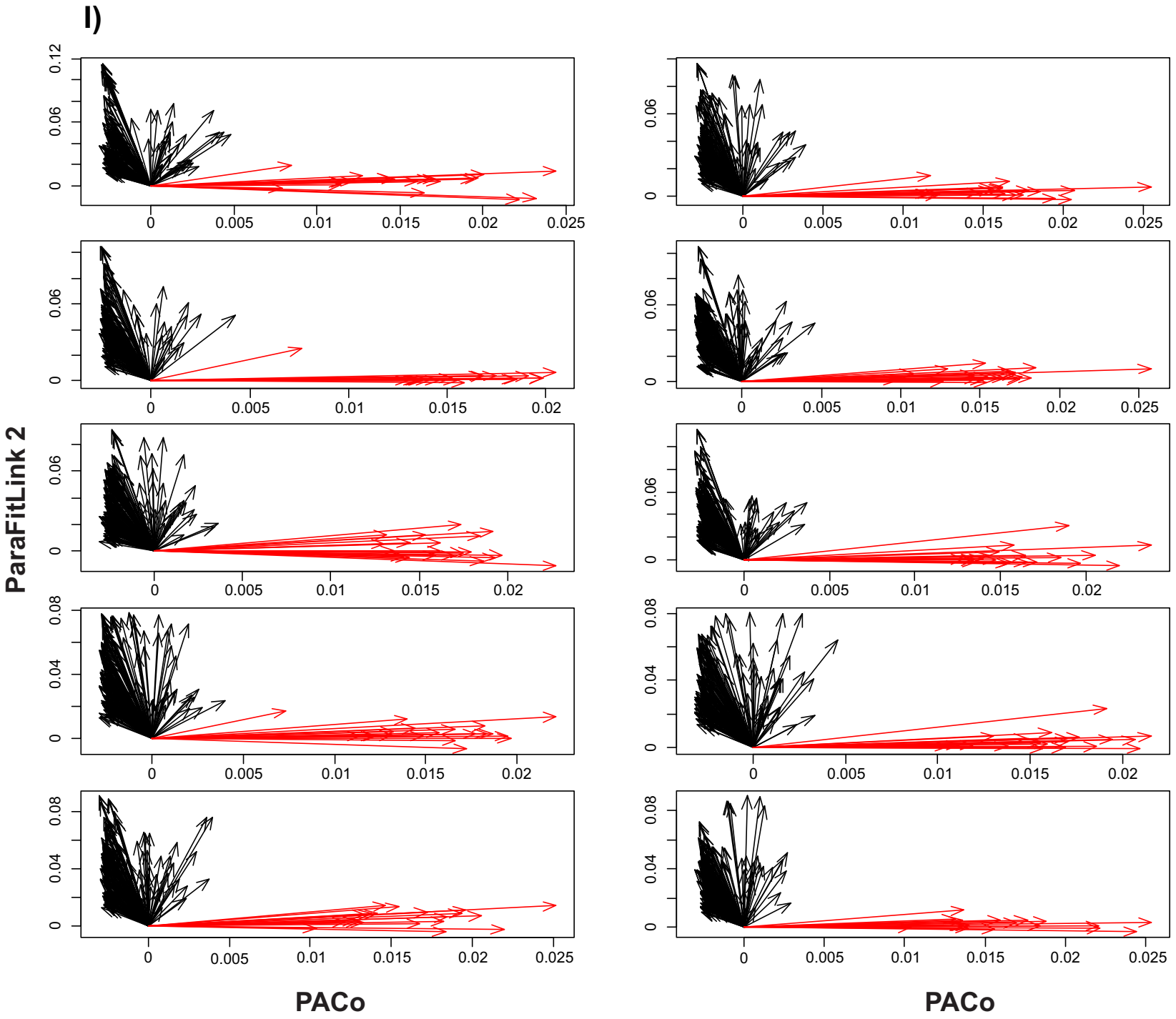


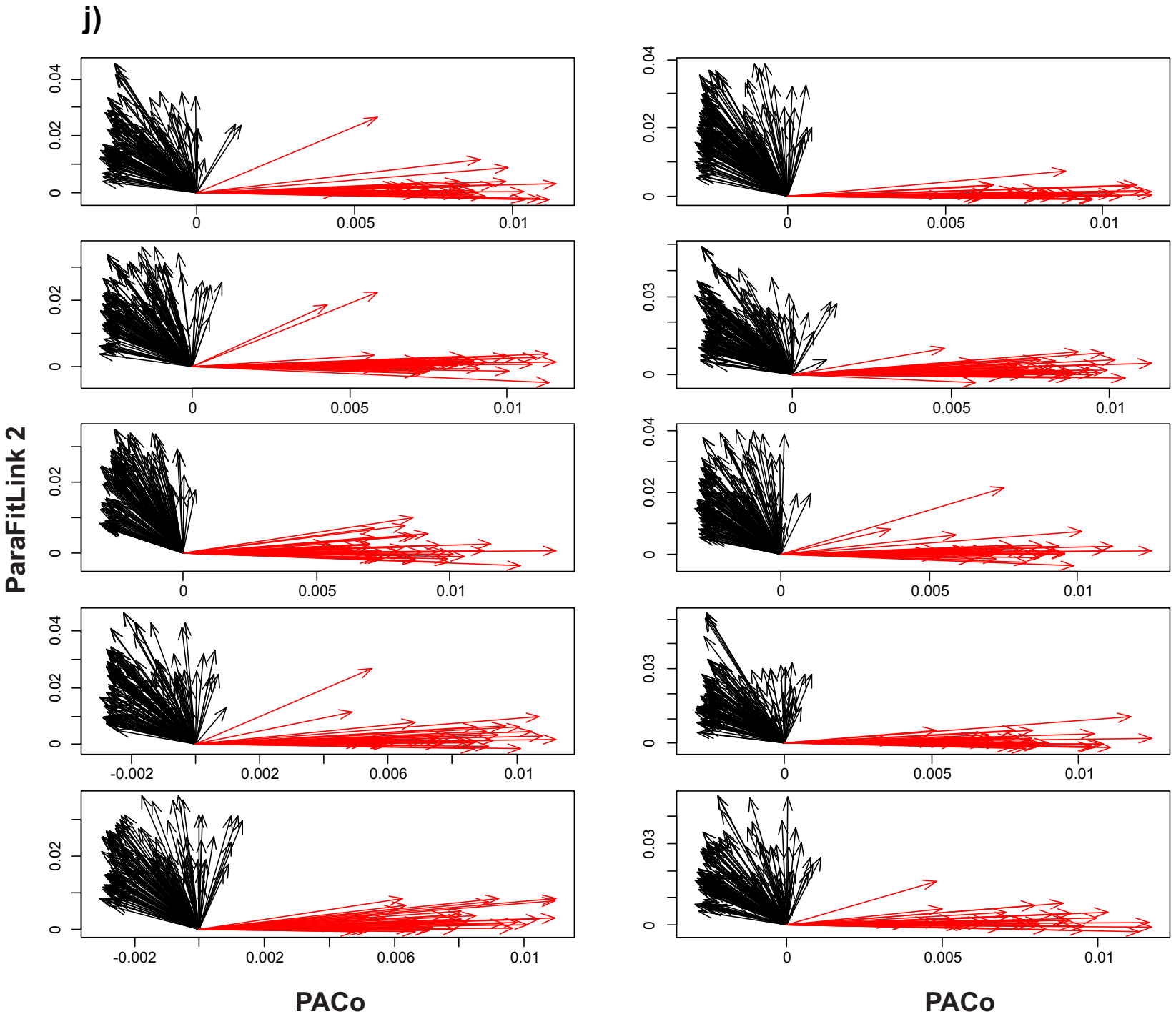
g)

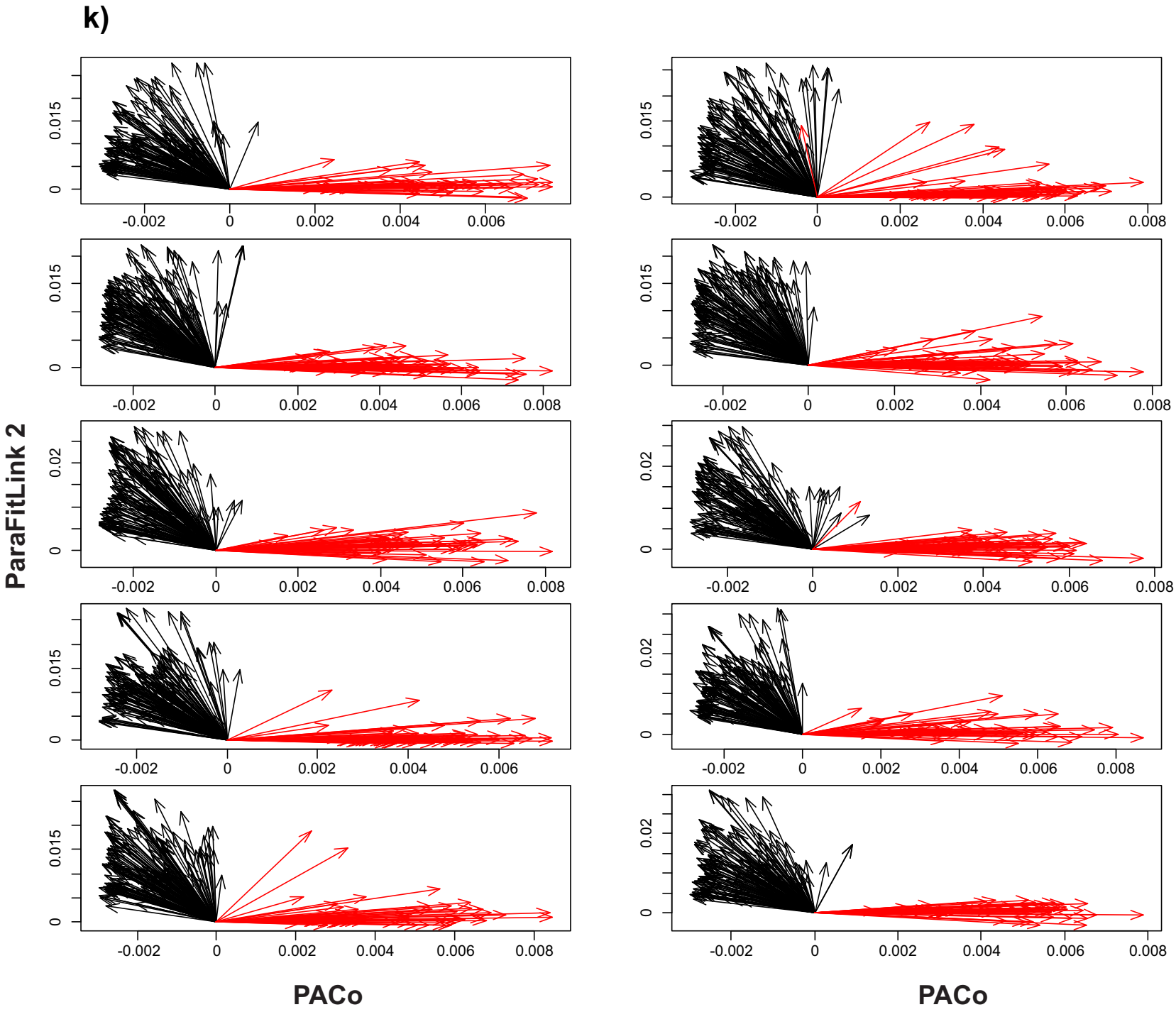


h)




I)
ParaFitLink 2
PACo
PACo


j)
ParaFitLink 2
PACo
PACo


k)
ParaFitLink 2
PACo
PACo

I)

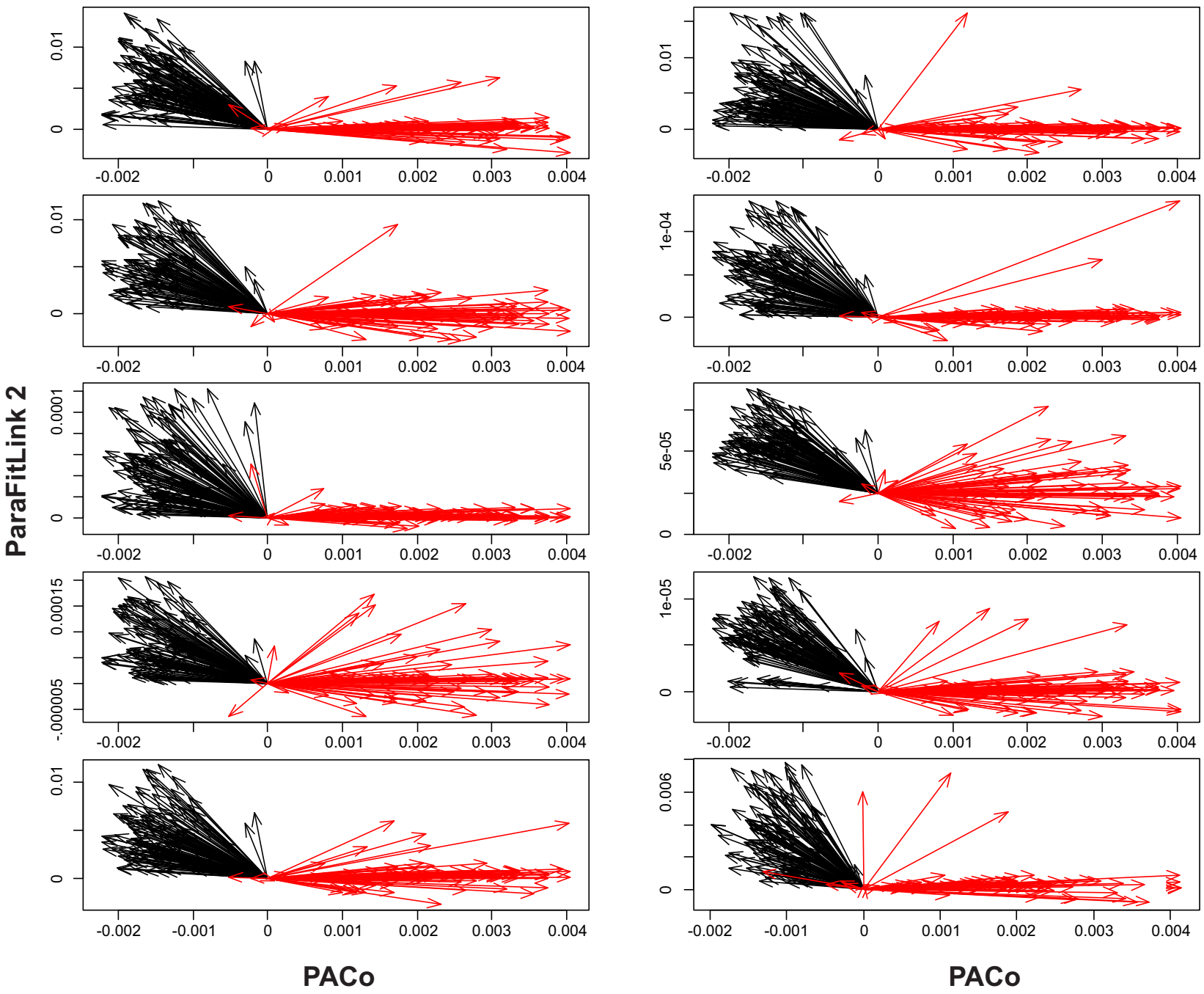
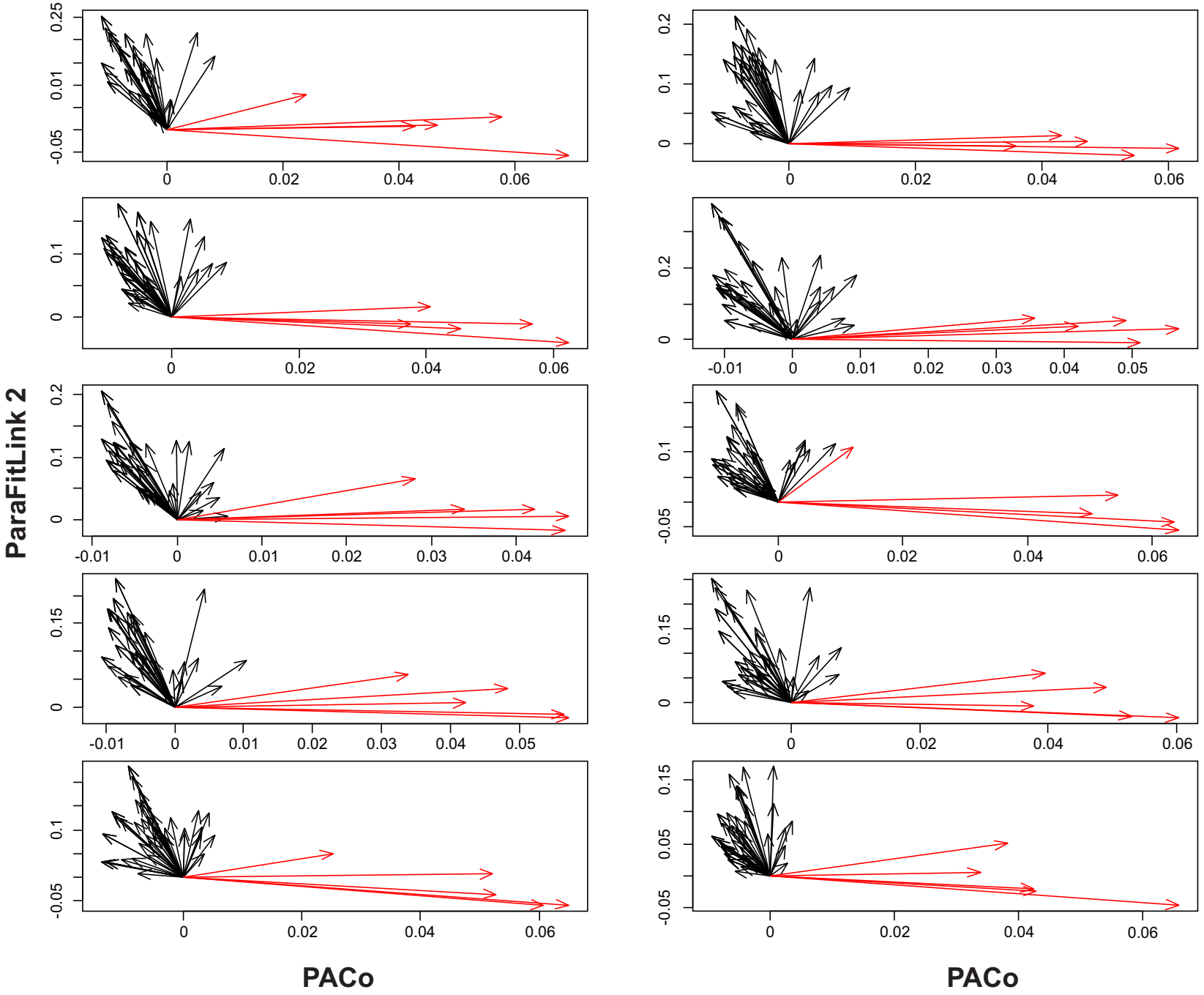
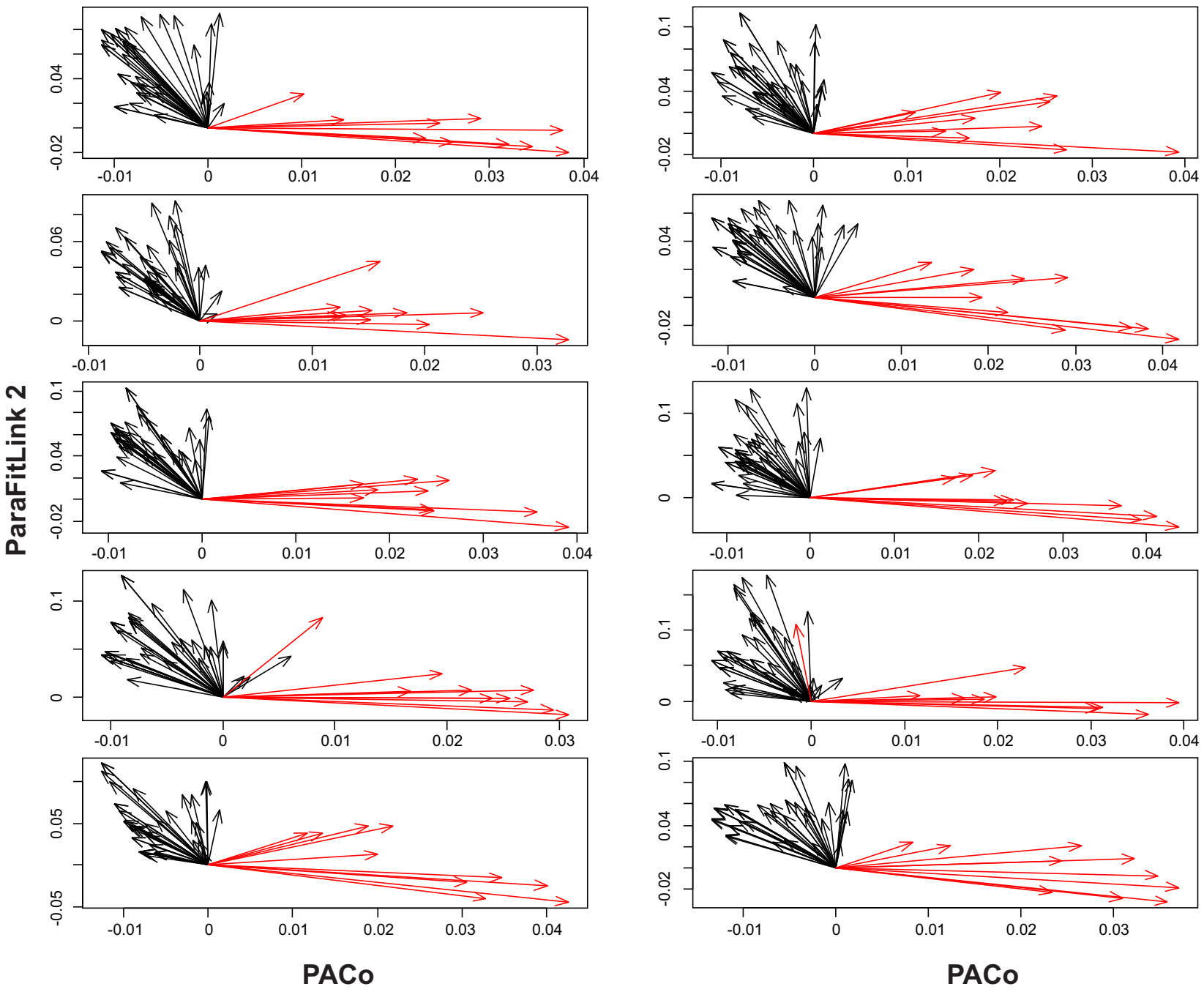

ParaFitLink 2
PACo
PACo
-0.002
-0.001
0
0.001
0.002
0.003
0.004
0.01
0
0.0001
-.000005
0.00015
1e-04
5e-05
1e-05
0.006

Fig. S2

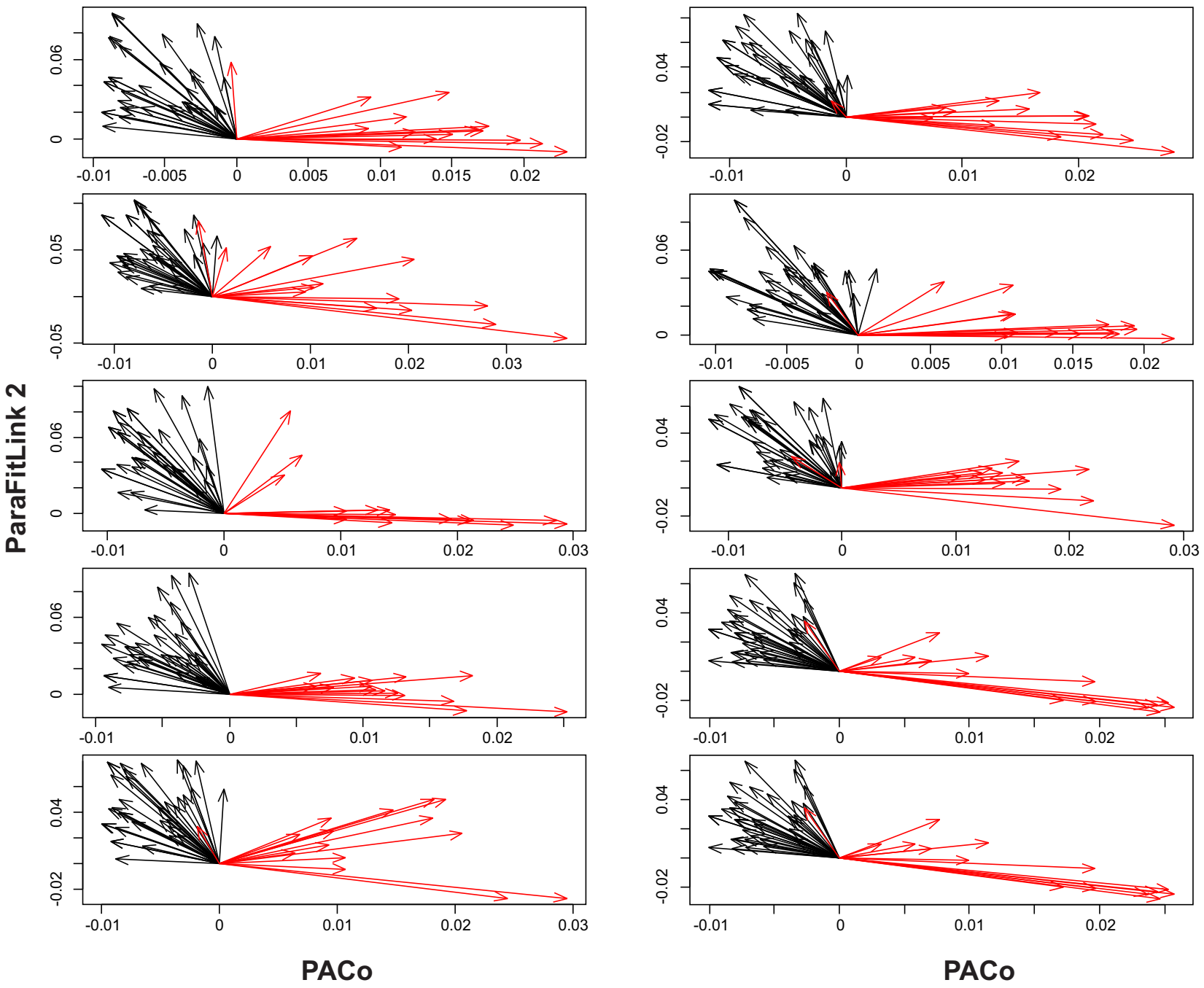
a)



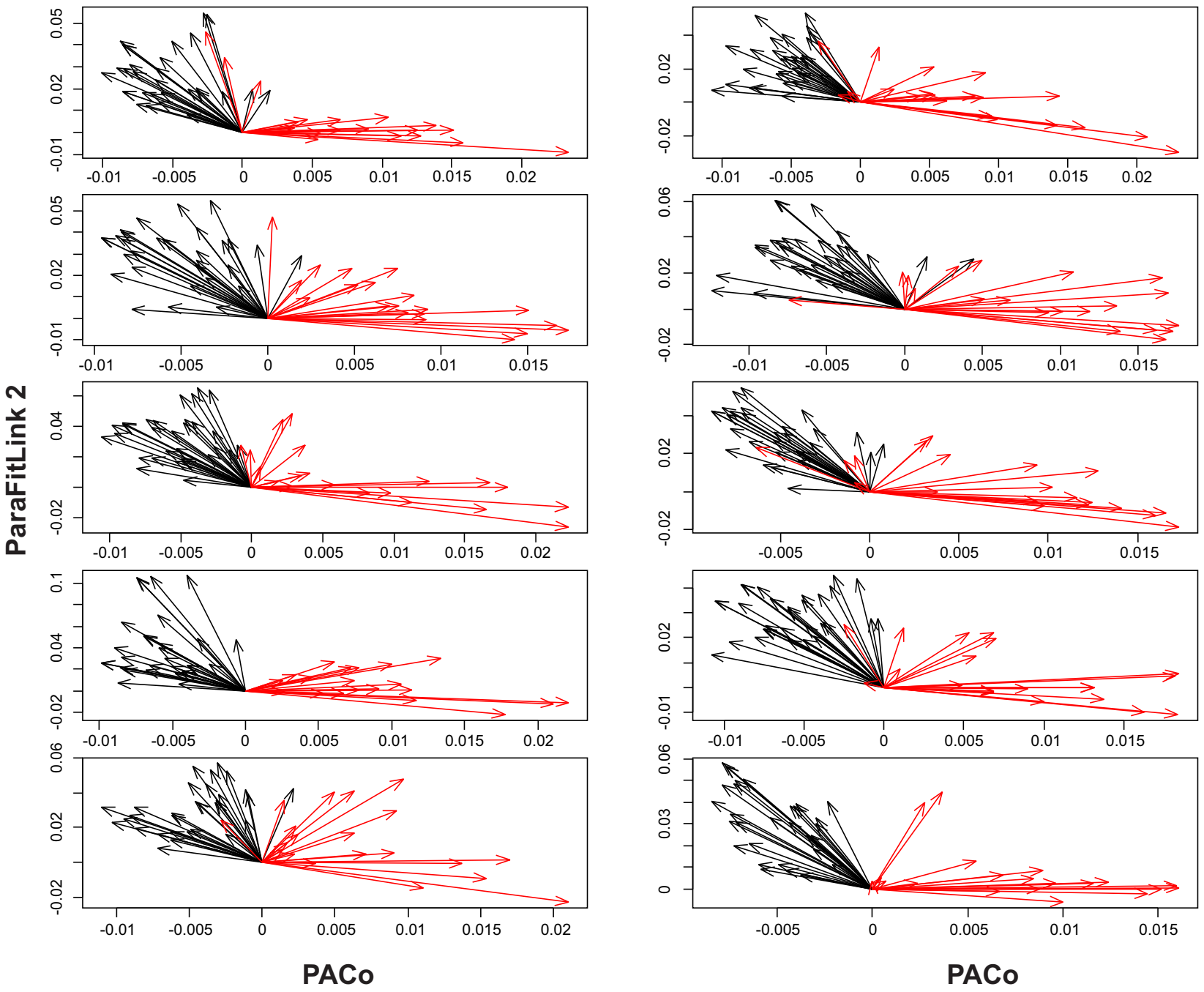
b)

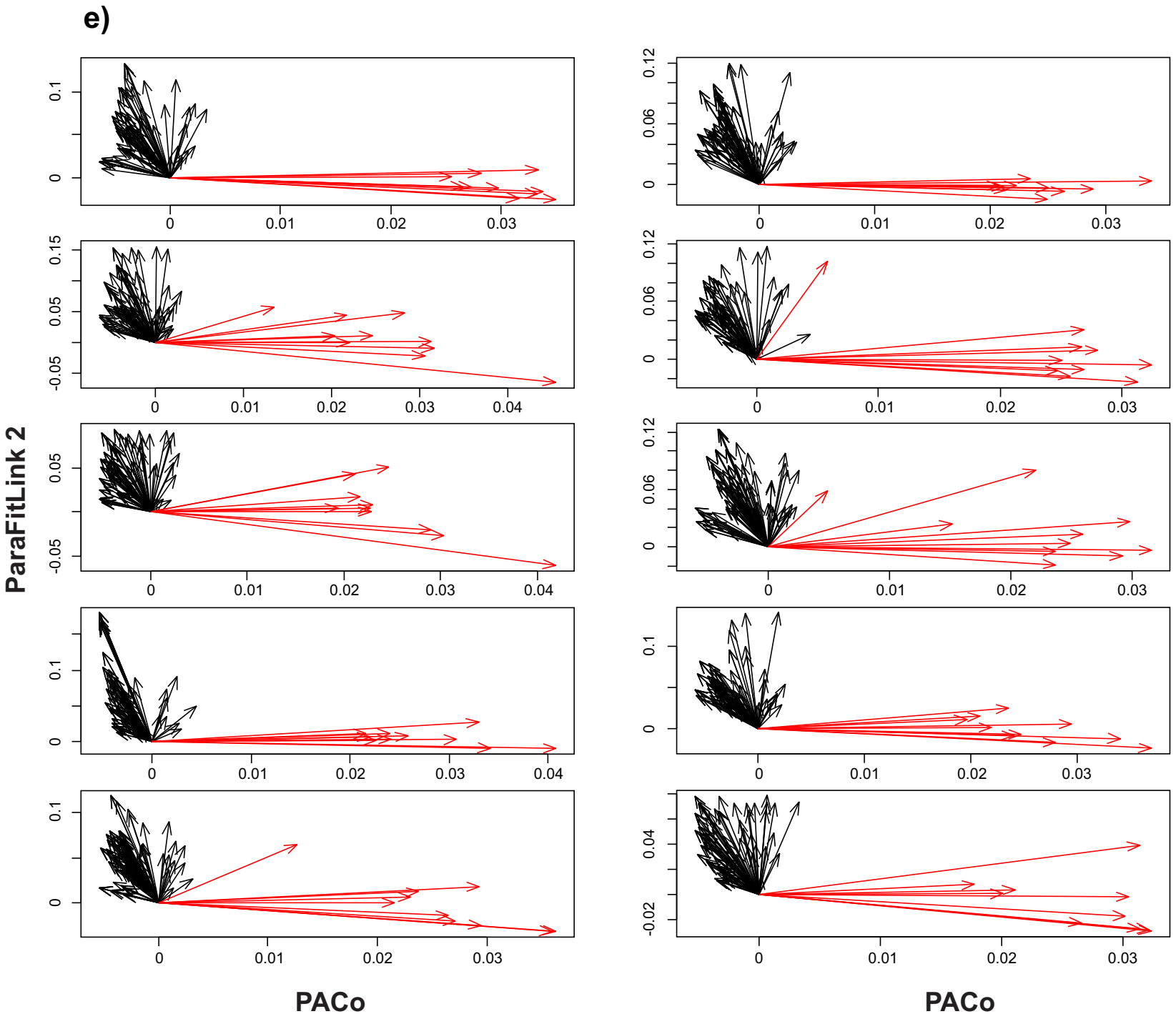


c)

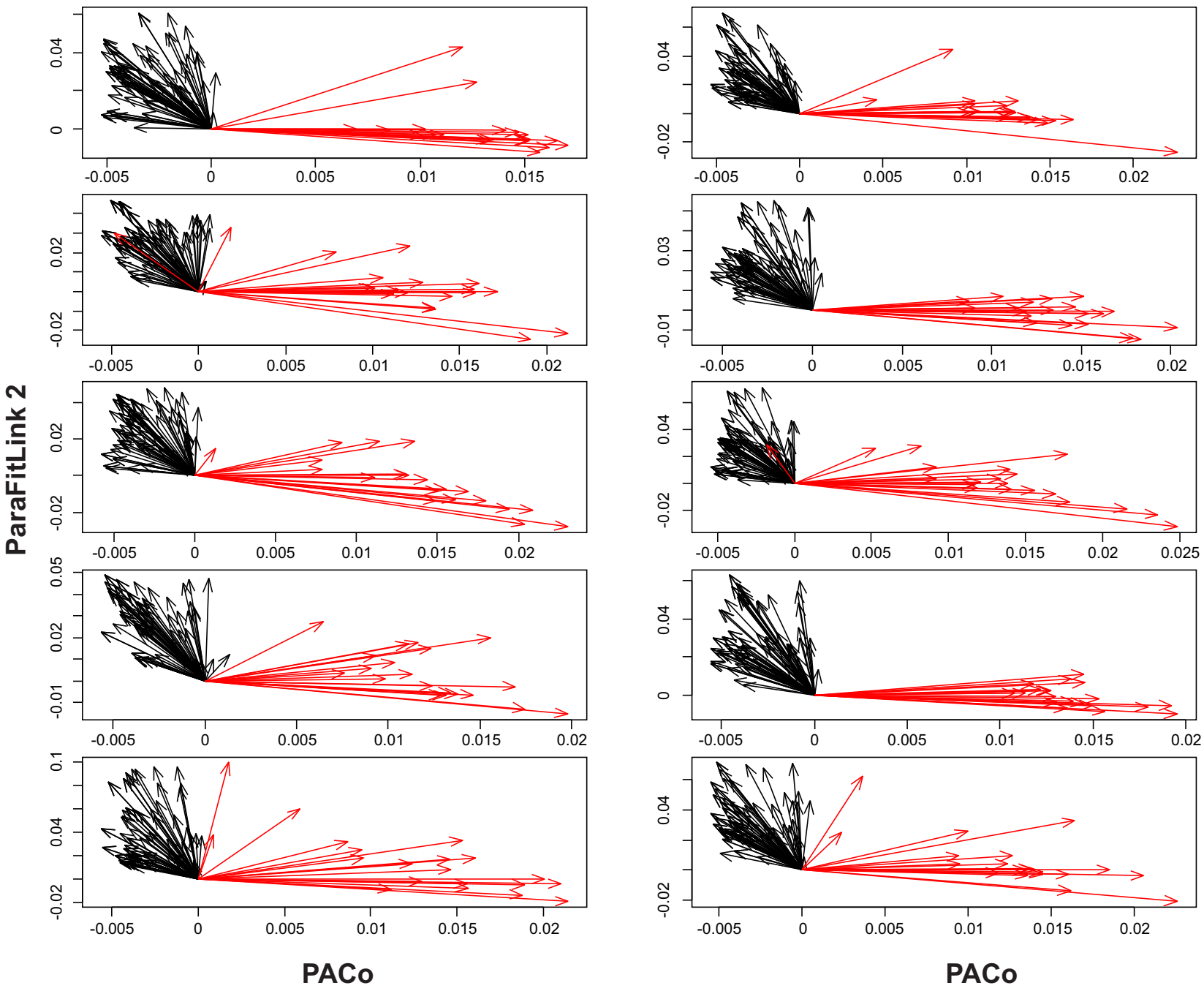


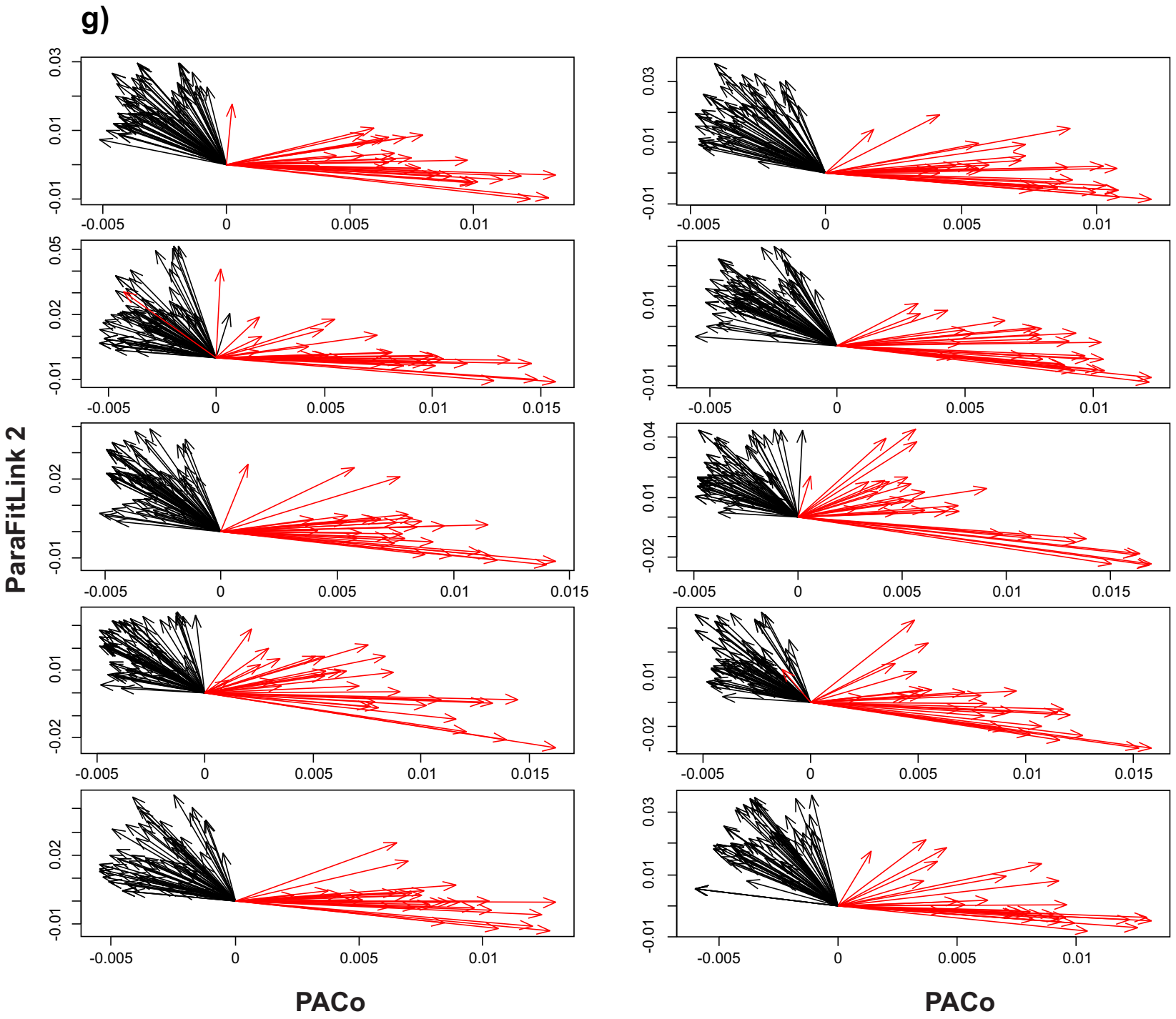
d)



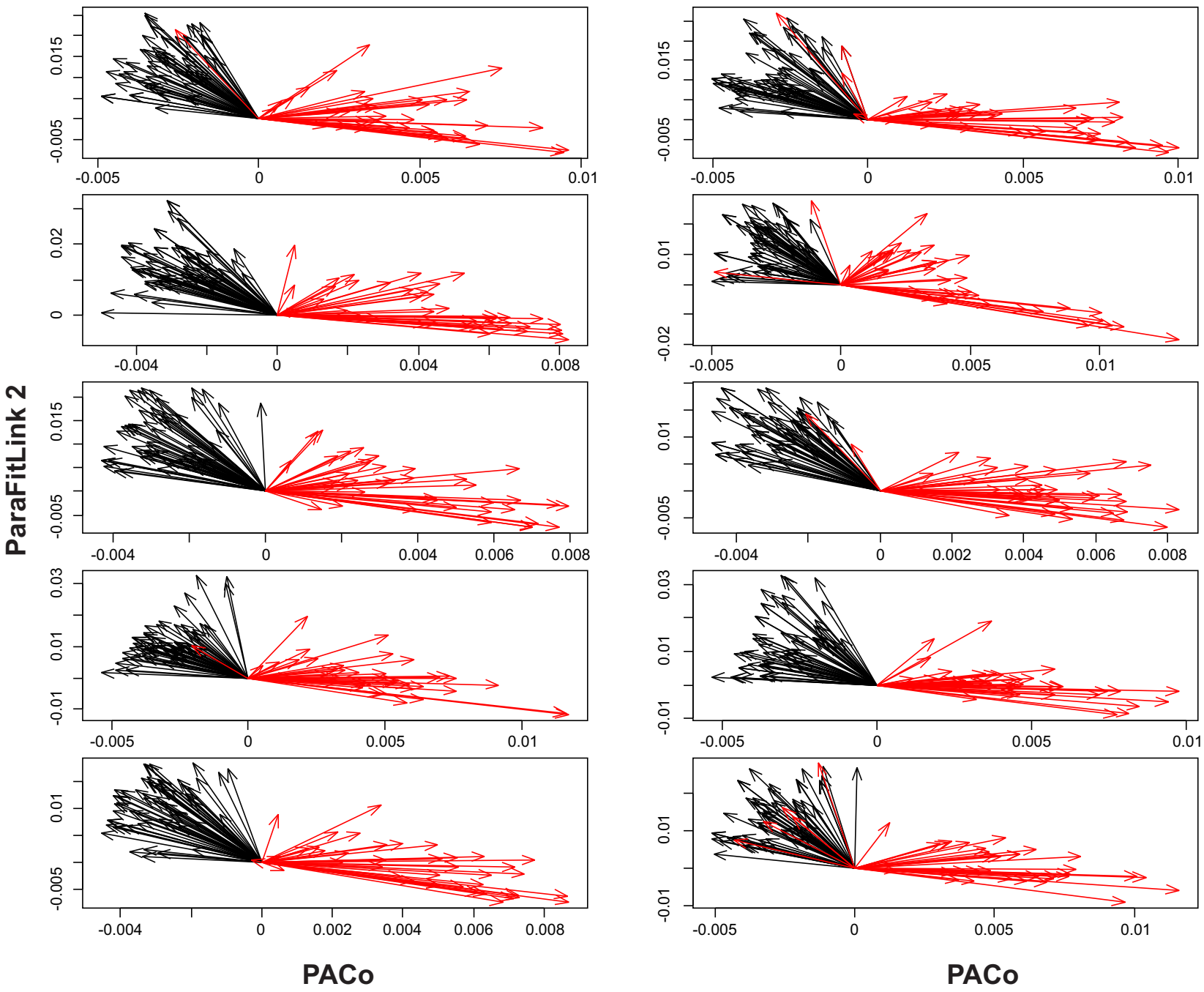

e)
ParaFitLink 2
PACo
PACo

f)

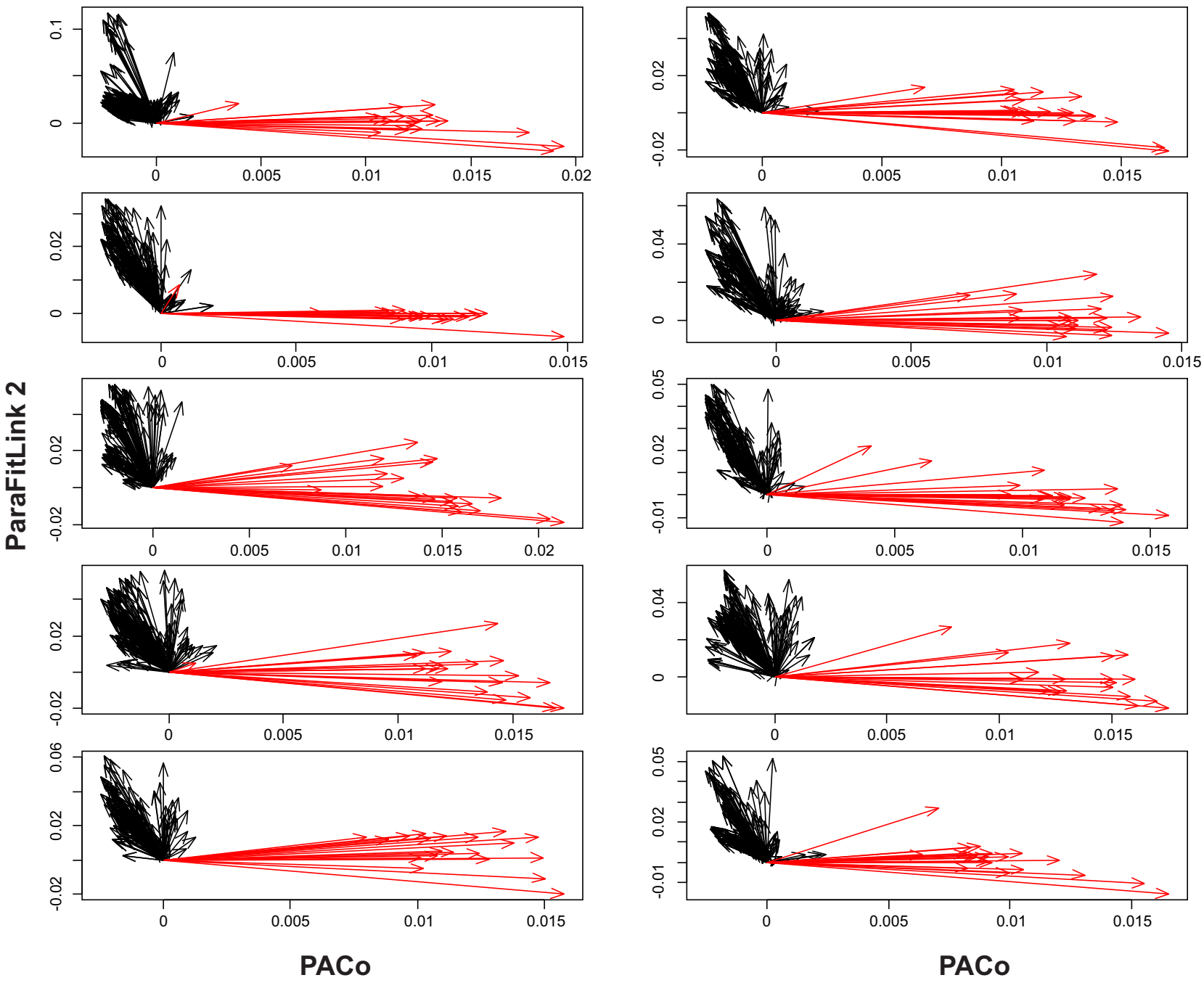


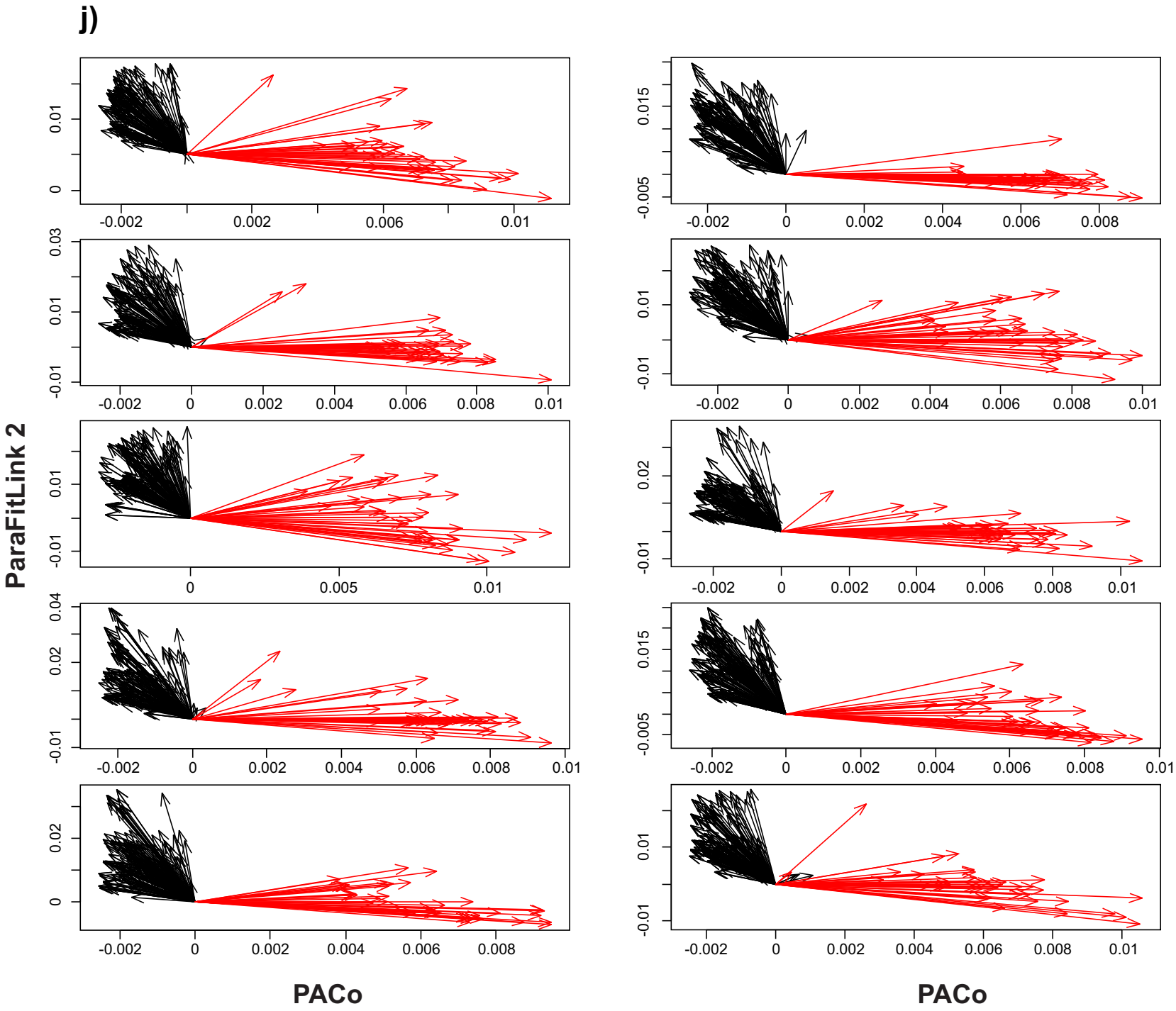

g)
ParaFitLink 2
PACo
PACo

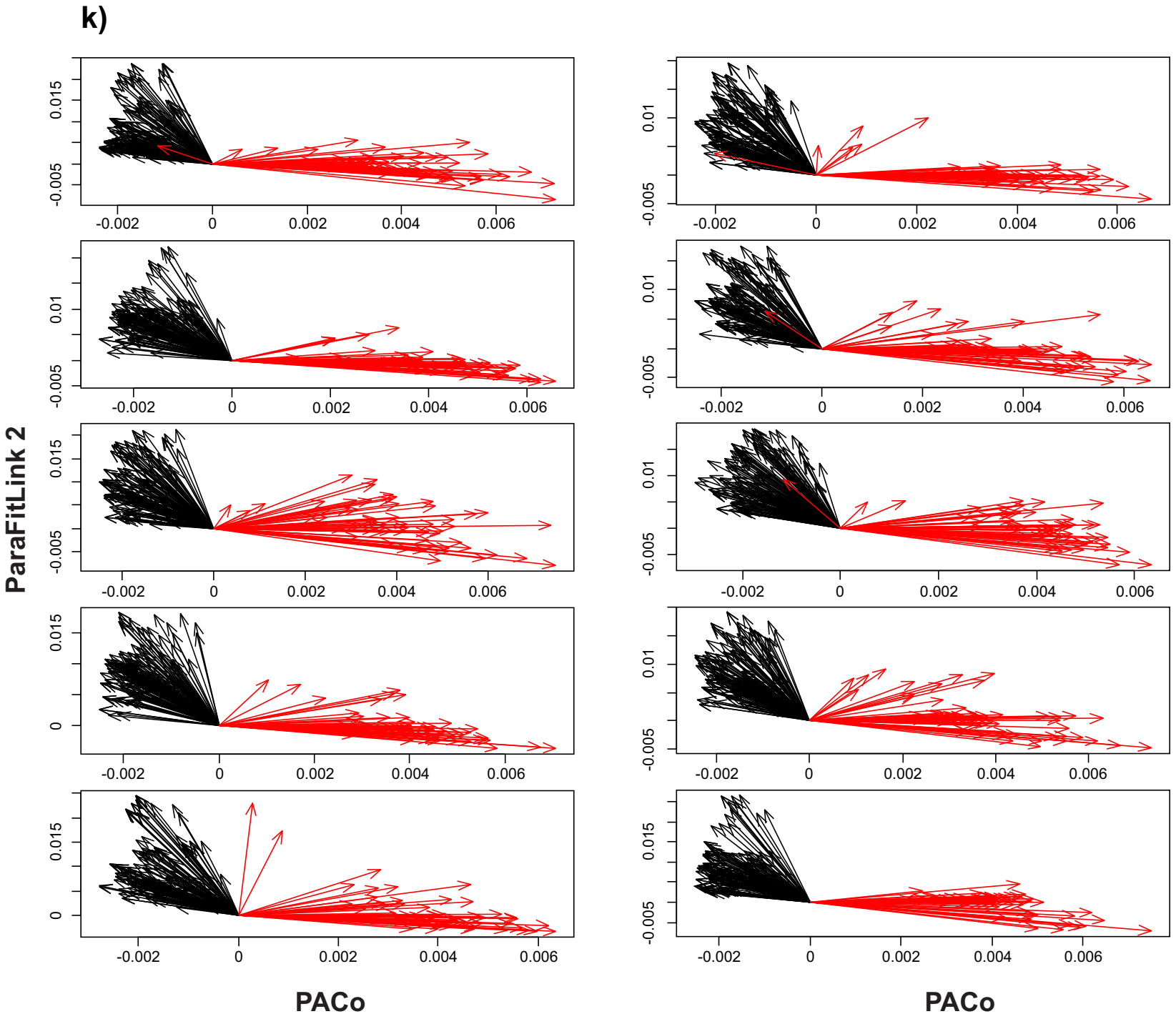
h)



I)




j)
ParaFitLink 2
PACo
PACo


k)
ParaFitLink 2
PACo
PACo

I)

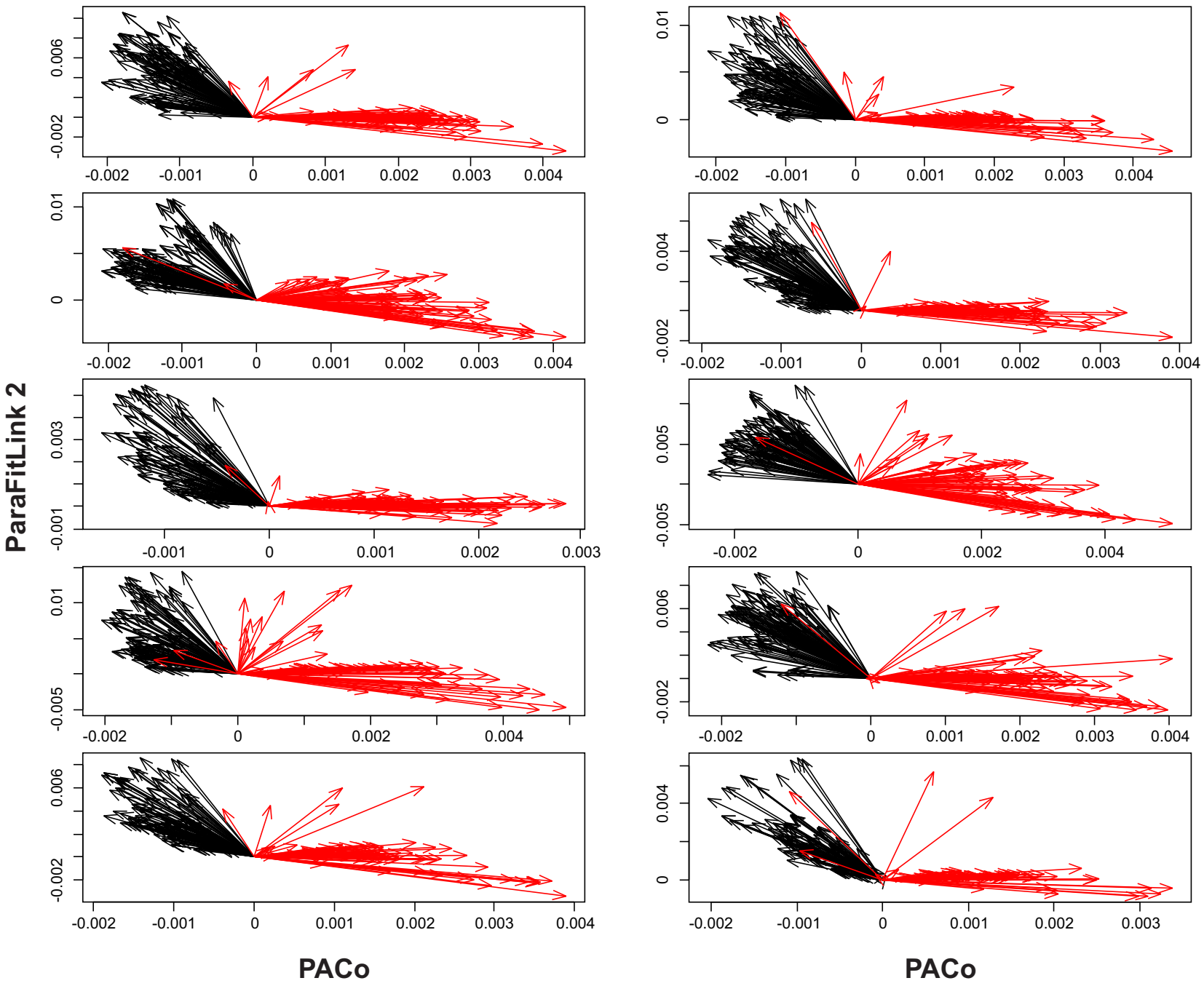


Fig. S3

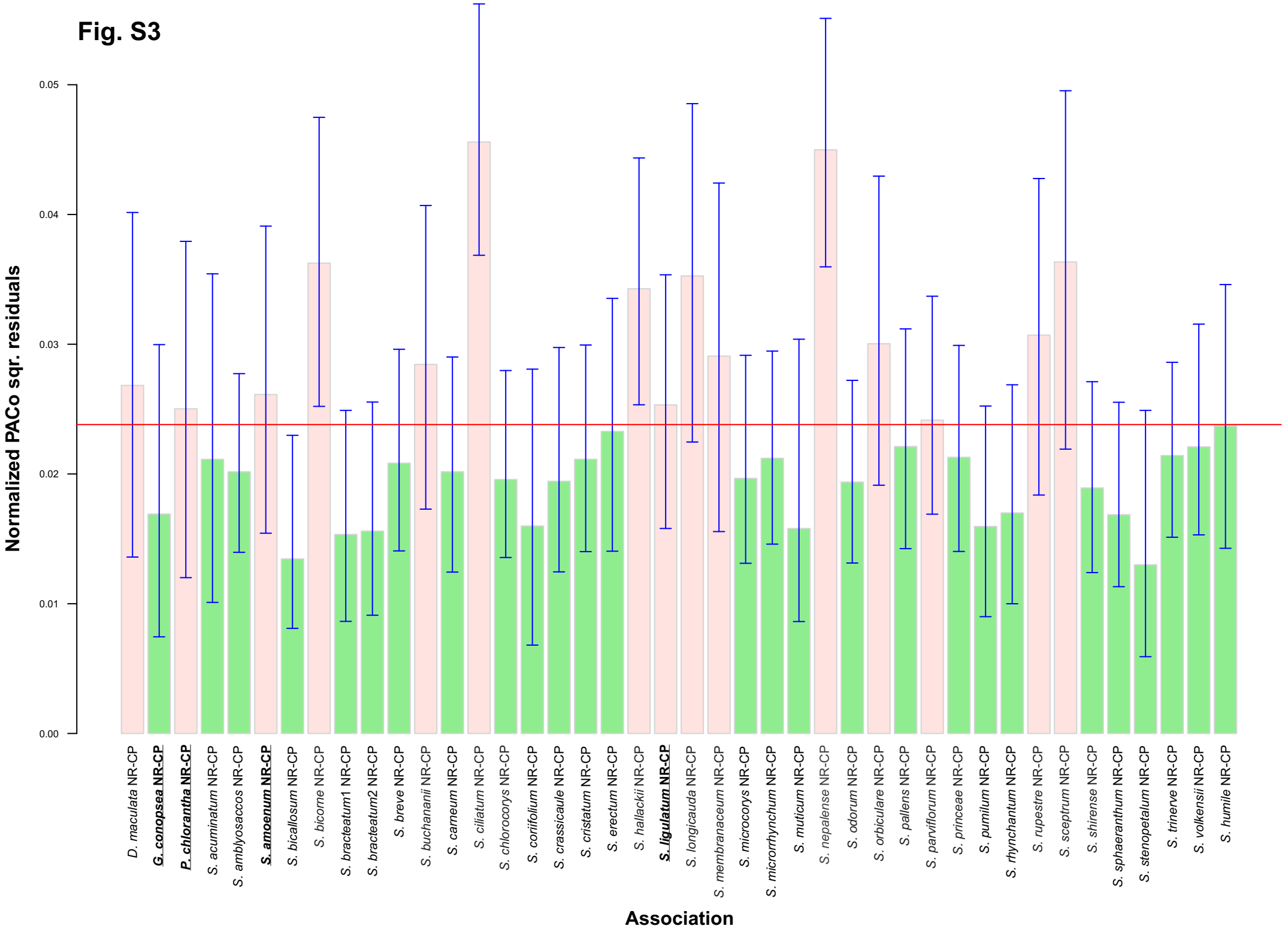

Normalized PACo sqr. residuals
0.00
0.01
0.02
0.03
0.04
0.05
Association
D. maculata NR-CP
G. conopsea NR-CP
P. chlorantha NR-CP
S. acuminatum NR-CP
S. amblyosaccos NR-CP
S. amoenum NR-CP
S. bicallosum NR-CP
S. bicorne NR-CP
S. bracteatum1 NR-CP
S. bracteatum2 NR-CP
S. breve NR-CP
S. buchananii NR-CP
S. carneum NR-CP
S. ciliatum NR-CP
S. chlorocorys NR-CP
S. corifolium NR-CP
S. crassicaule NR-CP
S. cristatum NR-CP
S. erectum NR-CP
S. hallackii NR-CP
S. ligulatum NR-CP
S. longicauda NR-CP
S. membranaceum NR-CP
S. microcorys NR-CP
S. microrrhynchum NR-CP
S. muticum NR-CP
S. nepalense NR-CP
S. odorum NR-CP
S. orbiculare NR-CP
S. pallens NR-CP
S. parviflorum NR-CP
S. princeae NR-CP
S. pumilum NR-CP
S. rhynchantum NR-CP
S. rupestre NR-CP
S. sceptrum NR-CP
S. shirense NR-CP
S. sphaeranthum NR-CP
S. stenopetalum NR-CP
S. trinerve NR-CP
S. volkensii NR-CP
S. humile NR-CP

Fig. S4

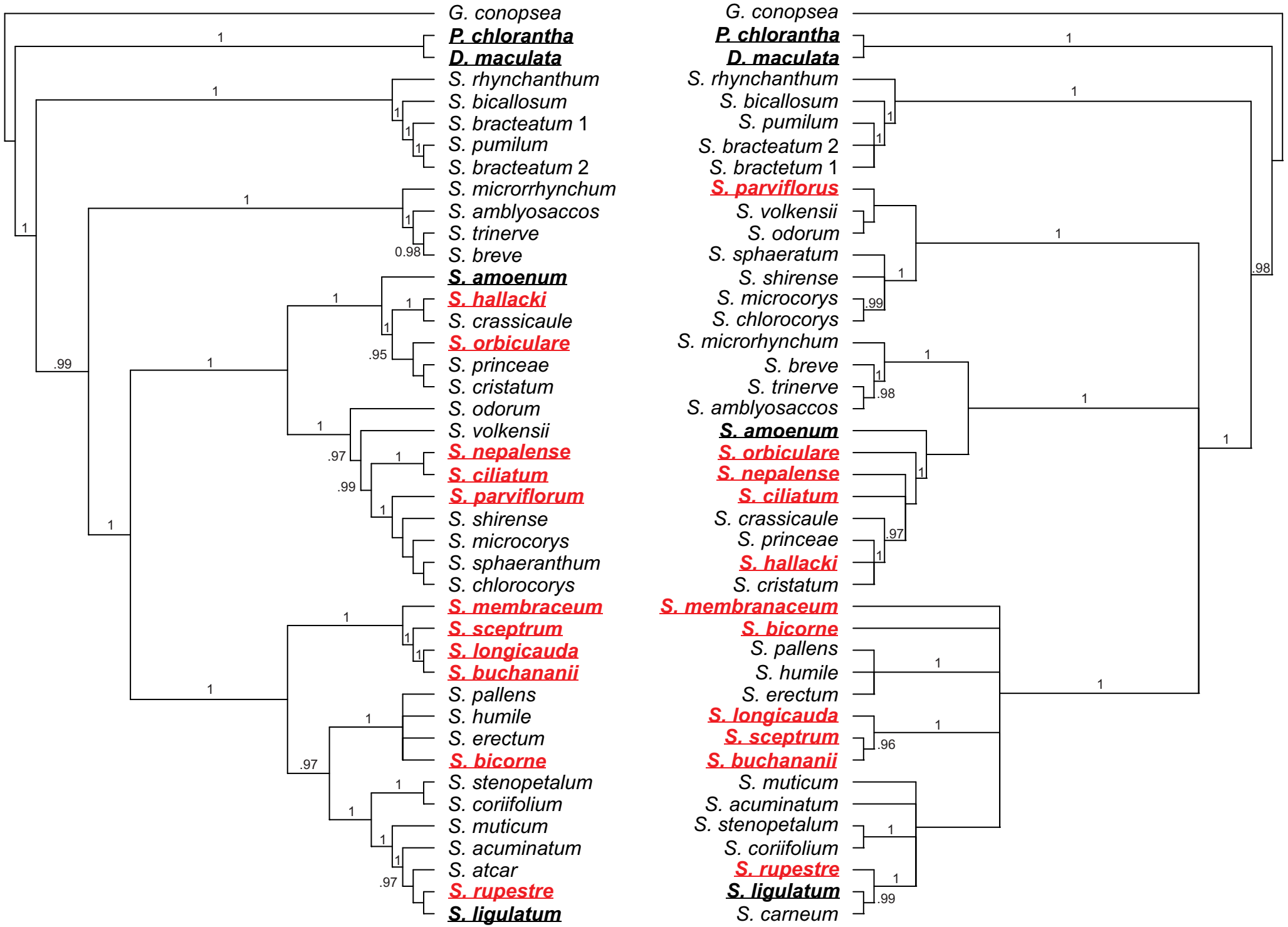

G. conopsea
P. chlorantha
D. maculata
S. rhynchanthum
S. bicallosum
S. bracteatum 1
S. pumilum
S. bracteatum 2
S. microrrhynchum
S. amblyosaccos
S. trinerve
S. breve
S. amoenum
S. hallacki
S. crassicaule
S. orbiculare
S. princeae
S. cristatum
S. odorum
S. volkensii
S. nepalense
S. ciliatum
S. parviflorum
S. shirense
S. microcorys
S. sphaeranthum
S. chlorocorys
S. membraceum
S. sceptrum
S. longicauda
S. buchananii
S. pallens
S. humile
S. erectum
S. bicorne
S. stenopetalum
S. coriifolium
S. muticum
S. acuminatum
S. atcar
S. rupestre
S. ligulatum
G. conopsea
P. chlorantha
D. maculata
S. rhynchanthum
S. bicallosum
S. pumilum
S. bracteatum 2
S. bractetum 1
S. parviflorus
S. volkensii
S. odorum
S. sphaeratum
S. shirense
S. microcorys
S. chlorocorys
S. microrhynchum
S. breve
S. trinerve
S. amblyosaccos
S. amoenum
S. orbiculare
S. nepalense
S. ciliatum
S. crassicaule
S. princeae
S. hallacki
S. cristatum
S. membranaceum
S. bicorne
S. pallens
S. humile
S. erectum
S. longicauda
S. sceptrum
S. buchananii
S. muticum
S. acuminatum
S. stenopetalum
S. coriifolium
S. rupestre
S. ligulatum
S. carneum

Fig. S5

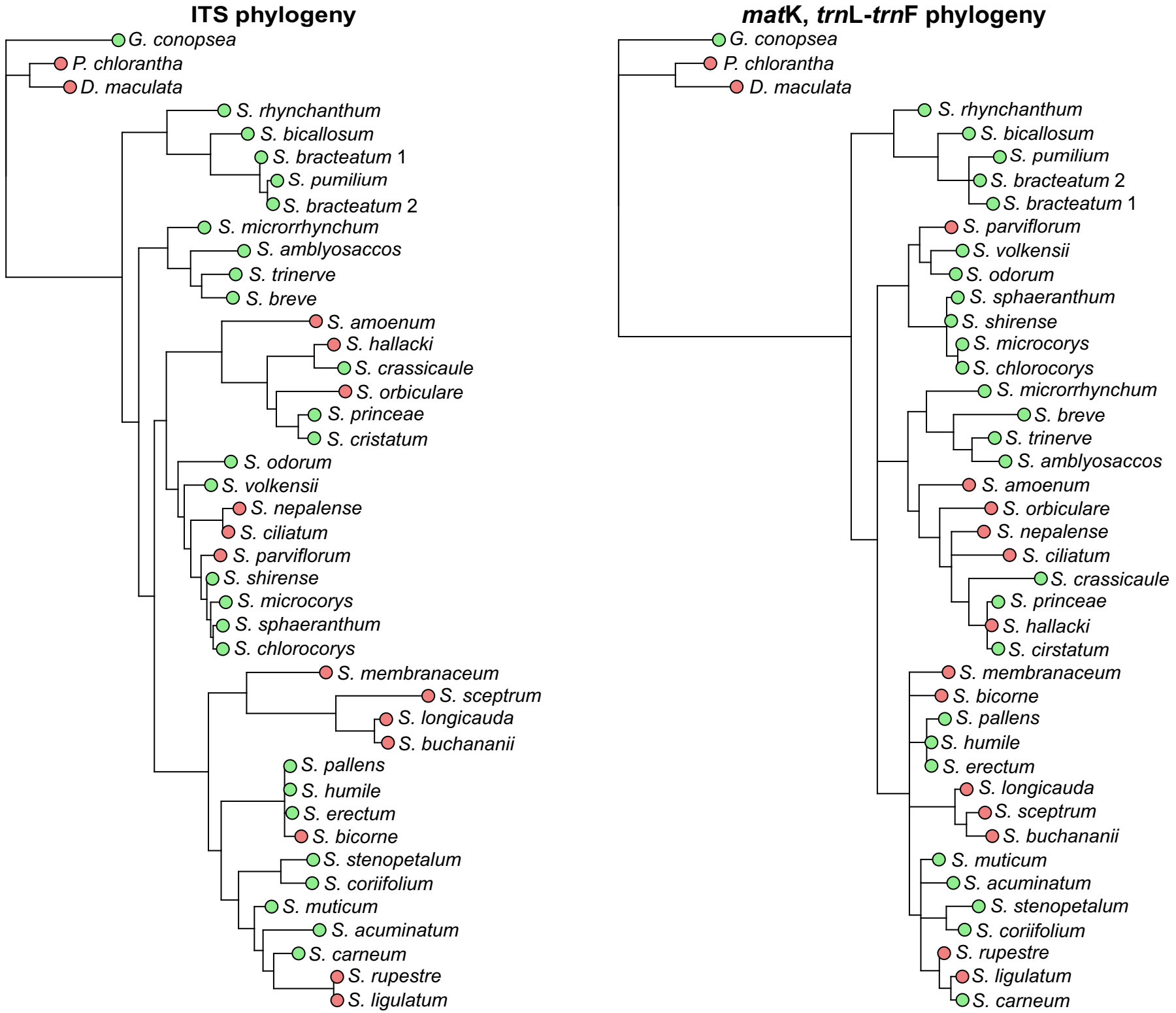

ITS phylogeny
G. conopsea
P. chlorantha
D. maculata
S. rhynchanthum
S. bicallosum
S. bracteatum 1
S. pumilium
S. bracteatum 2
S. microrrhynchum
S. amblyosaccos
S. trinerve
S. breve
S. amoenum
S. hallacki
S. crassicaule
S. orbiculare
S. princeae
S. cristatum
S. odorum
S. volkensii
S. nepalense
S. ciliatum
S. parviflorum
S. shirense
S. microcorys
S. sphaeranthum
S. chlorocorys
S. membranaceum
S. sceptrum
S. longicauda
S. buchananii
S. pallens
S. humile
S. erectum
S. bicorne
S. stenopetalum
S. coriifolium
S. muticum
S. acuminatum
S. carneum
S. rupestre
S. ligulatum
matK, trnL-trnF phylogeny
G. conopsea
P. chlorantha
D. maculata
S. rhynchanthum
S. bicallosum
S. pumilium
S. bracteatum 2
S. bracteatum 1
S. parviflorum
S. volkensii
S. odorum
S. sphaeranthum
S. shirense
S. microcorys
S. chlorocorys
S. microrrhynchum
S. breve
S. trinerve
S. amblyosaccos
S. amoenum
S. orbiculare
S. nepalense
S. ciliatum
S. crassicaule
S. princeae
S. hallacki
S. cirstatum
S. membranaceum
S. bicorne
S. pallens
S. humile
S. erectum
S. longicauda
S. sceptrum
S. buchananii
S. muticum
S. acuminatum
S. stenopetalum
S. coriifolium
S. rupestre
S. ligulatum
S. carneum

Fig. S6

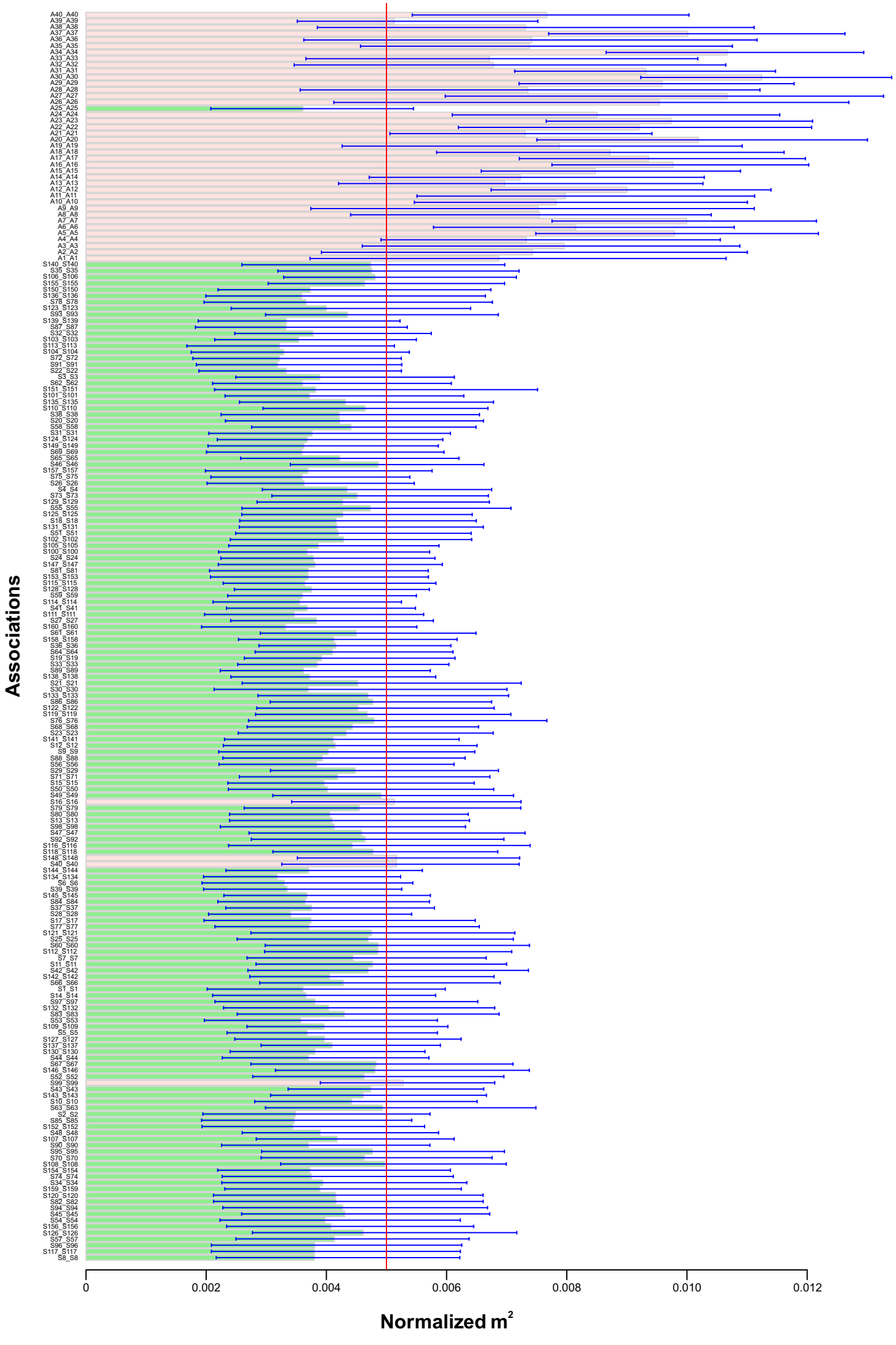

Associations
A40_A40
A39_A39
A38_A38
A37_A37
A36_A36
A35_A35
A34_A34
A33_A33
A32_A32
A31_A31
A30_A30
A29_A29
A28_A28
A27_A27
A26_A26
A25_A25
A24_A24
A23_A23
A22_A22
A21_A21
A20_A20
A19_A19
A18_A18
A17_A17
A16_A16
A15_A15
A14_A14
A13_A13
A12_A12
A11_A11
A10_A10
A9_A9
A8_A8
A7_A7
A6_A6
A5_A5
A4_A4
A3_A3
A2_A2
A1_A1
S140_S140
S35_S35
S106_S106
S155_S155
S150_S150
S136_S136
S78_S78
S123_S123
S93_S93
S139_S139
S87_S87
S32_S32
S103_S103
S113_S113
S104_S104
S72_S72
S91_S91
S22_S22
S3_S3
S62_S62
S151_S151
S101_S101
S135_S135
S110_S110
S38_S38
S20_S20
S58_S58
S31_S31
S124_S124
S149_S149
S69_S69
S65_S65
S46_S46
S157_S157
S75_S75
S26_S26
S4_S4
S73_S73
S129_S129
S55_S55
S125_S125
S18_S18
S131_S131
S51_S51
S102_S102
S105_S105
S100_S100
S24_S24
S147_S147
S81_S81
S153_S153
S115_S115
S128_S128
S59_S59
S114_S114
S41_S41
S111_S111
S27_S27
S160_S160
S61_S61
S158_S158
S36_S36
S64_S64
S19_S19
S33_S33
S89_S89
S138_S138
S21_S21
S30_S30
S133_S133
S86_S86
S122_S122
S119_S119
S76_S76
S68_S68
S23_S23
S141_S141
S12_S12
S9_S9
S88_S88
S56_S56
S29_S29
S71_S71
S15_S15
S50_S50
S49_S49
S16_S16
S79_S79
S80_S80
S13_S13
S98_S98
S47_S47
S92_S92
S116_S116
S118_S118
S148_S148
S40_S40
S144_S144
S134_S134
S6_S6
S39_S39
S145_S145
S94_S94
S37_S37
S28_S28
S17_S17
S77_S77
S121_S121
S25_S25
S60_S60
S112_S112
S7_S7
S11_S11
S42_S42
S142_S142
S66_S66
S1_S1
S14_S14
S97_S97
S132_S132
S83_S83
S53_S53
S109_S109
S5_S5
S127_S127
S137_S137
S130_S130
S44_S44
S67_S67
S146_S146
S52_S52
S99_S99
S43_S43
S143_S143
S10_S10
S63_S63
S2_S2
S85_S85
S152_S152
S48_S48
S107_S107
S90_S90
S95_S95
S70_S70
S108_S108
S154_S154
S74_S74
S34_S34
S159_S159
S120_S120
S82_S82
S94_S94
S45_S45
S54_S54
S156_S156
S126_S126
S57_S57
S96_S96
S117_S117
S8_S8
0
0.002
0.004
0.006
0.008
0.010
0.012
Normalized m²

Fig. S7

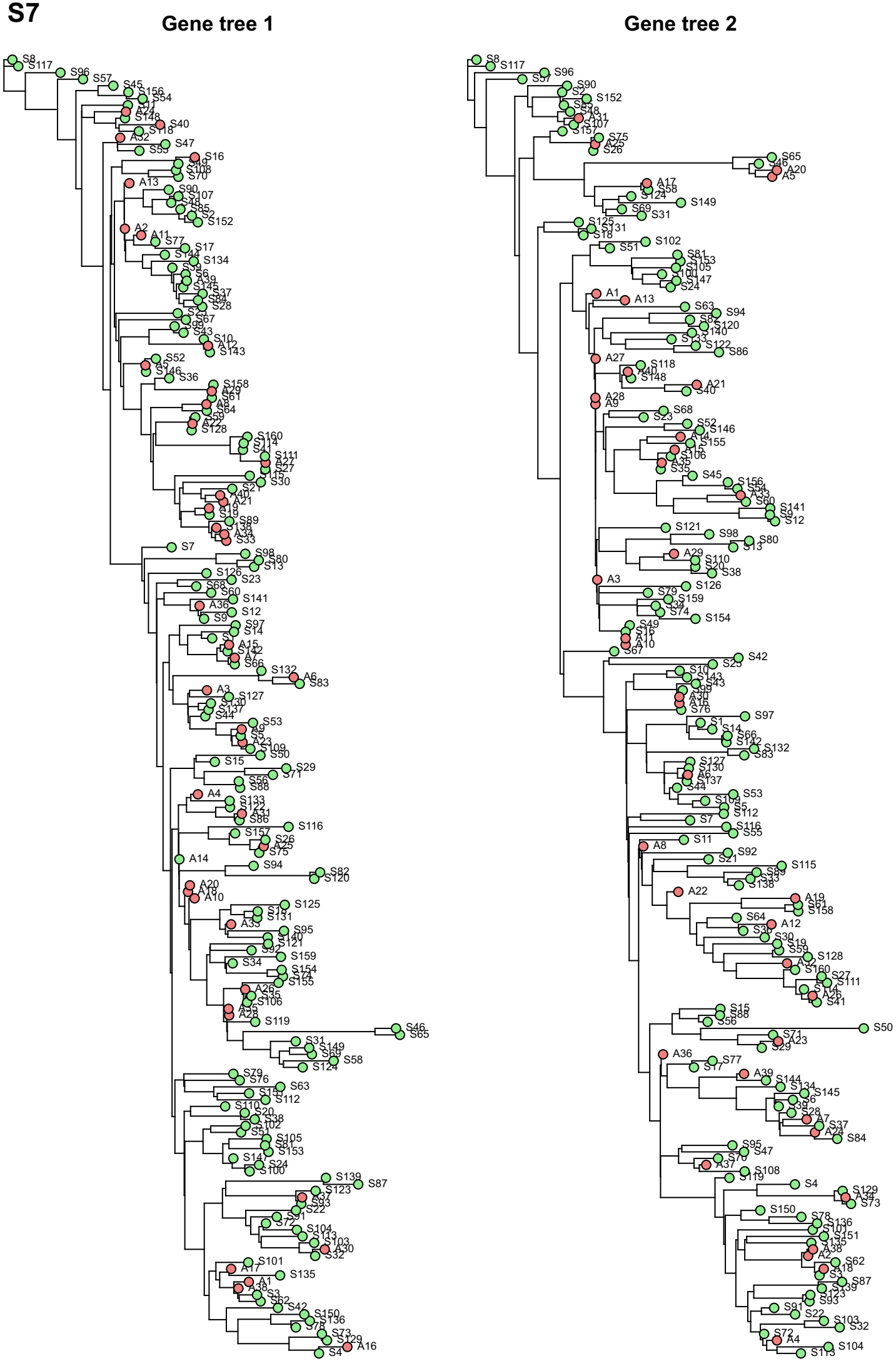


Table S1

Loci	Primer	Sequence	Reference	Pre-melt	Amplification	Final extention	Number of amplification cycles
ITS	ITS 4	TCC-TCC-GCT-TAT-TGA-TAT-GC	Baldwin (1992)	95°C (3 min)	95°C (30 secs) + 52°C (1 min) + 68°C (1 min)	68°C (10 min)	39
	ITS 5	GGA-AGT-AAA-AGT-CGT-AAC-AAG-G		95°C (3 min)	95°C (30 secs) + 52°C (1 min) + 68°C (1 min)	68°C (10 min)	39
ETS	EST-Orchid	CAT-ATG-AGT-TGT-TGC-GGA-CC (AT)-T	Monteiro et al (2010)	95°C (3 min)	95°C (30 secs) + 52°C (1 min) + 68°C (1 min)	68°C (10 min)	39
	18-IGS	AGA-CAA-GCA-TAT-GAC-TAC-TGG-CAG-G	Markos and Balwin (1998)	95°C (3 min)	95°C (30 secs) + 52°C (1 min) + 68°C (1 min)	68°C (10 min)	39
Xdh	X502F	TGT-GAT-GTC-GAT-GTA-TGC	Górniak et al (2010)	95°C (3 min)	95°C (30 secs) + 53°C (1 min) + 68°C (1.5 min)	68°C (10 min)	39
	X1599R	G(AT)G-AGA-GAA-A(CT)TG-GAG-CAA-C		95°C (3 min)	95°C (30 secs) + 53°C (1 min) + 68°C (1.5 min)	68°C (10 min)	39
Ycf1	3720F	TAC-GTA-TGT-AAT-GAA-CGA-ATG-G	Neubig et al (2009)	95°C (3 min)	95°C (30 secs) + 54°C (1 min) + 68°C (1.5 min)	68°C (10 min)	39
	5500R	GCT-GTT-ATT-GGC-ATC-AAA-CCA-ATA-GCG		95°C (3 min)	95°C (30 secs) + 54°C (1 min) + 68°C (1.5 min)	68°C (10 min)	39
trn S-G	trn-S(GCU)	GCC-GCT-TTA-GTC-CAC-TCA-GC	Hamilton (1999)	95°C (3 min)	95°C (30 secs) + 51.5°C (1 min) + 68°C (1.5 min)	68°C (10 min)	39
	trn-G(UCC)	GAA-CGA-ATC-ACA-CTT-TTA-CCA-C		95°C (3 min)	95°C (30 secs) + 51.5°C (1 min) + 68°C (1.5 min)	68°C (10 min)	39

Table S2

			Nuclear - ribosomal dataset			Chloroplast dataset		
Taxon	**DNA Source - voucher**	**Distribution**	**ITS spacer**	**ETS spacer**	***Xdh* gene**	***matK* gene**	***Trn* S-G spacer**	***ycf1* gene**
***Catasetum collare* Cogn.**	cult. BGM[1] 5/1000 (M)	Brasil, Colombia, Ecuador, Venezuela	KT768384	KT768350	KT768454	-	KT768421	KT768491
***Catasetum juruenense* Hoehne**	cult. BGM 5/1223 (M)	Brazil	KT768385	KT768351	KT768455	-	KT768422	KT768492
***Catasetum macrocarpum* Rich. ex Kunth**	cult. BGM 96/3071 (M)	Brazil-Venezuela	KT768386	KT768352	KT768456	-	KT768423	KT768493
***Catasetum meeae* Pabst**	cult. BGM 97/3836 (M)	Brazil	KT768387	KT768353	KT768457	-	-	-
***Catasetum* x *roseoalbum* (Hook.) Lindl.**	cult. BGM 6/2496 (M)	Venezuela	KT768388	KT768354	KT768458	-	KT768424	KT768494
Catasetum sp. 1	ML086	-	JF692010	-	-	-	-	JF692138
Catasetum sp. 2	ML301	-	JF692017	-	-	-	-	JF692140
Catasetum sp. 3	SR1153	-	JF691914	-	-	-	-	JF692061
Catasetum sp. 4	SR1203	-	JF691923	-	-	-	-	JF692066
Catasetum sp. 5	SR1213	-	JF691925	-	-	-	-	JF692067
Catasetum sp. 6	SR1463	-	JF691960	-	-	-	-	JF692150
***Clowesia russelliana* (Hook.) Dodson**	cult. BGM 98/2889 (M)	Central America, Colombia, Venezuela	KT768389	-	-	-	KT768425	KT768495
Clowesia sp. 1	SR0703	-	JF69204	-	-	-	-	JF692131
Clowesia sp. 2	SR0716	-	JF692041	-	-	-	-	JF692154
Clowesia sp. 3	SR0726	-	JF692042	-	-	-	-	JF692155
Cyanaeorchis arundinae (Rchb. f.) Barb. Rodr.	Klein 126	Brazil	KF771817	-	-	KF771821	-	-
Cyanaeorchis minor Schltr.	Klein 124	Brazil	KF771818	-	-	KF771822	-	-
Cyanaeorchis praetermisa J.A.N.Bat. & Bianch.	Batista et al. 3041 (BHCB)	Brazil	KF771819	-	-	KF771823	-	-
***Cycnoches aureum* Lindl. & Paxton**	Pérez & Gerlach 1473 (M)	Panama	KT768390	KT768355	KT768459	-	KT768426	KT768496
***Cycnoches barthiorum* G.F.Carr & Christenson**	cult. BGM 12/1476 (M)	Colombia	KT768391	KT768356	KT768460	-	KT768427	KT768497
***Cycnoches chlorochilon* Klotzch**	cult. BGM 94/981 (M)	Panama, Colombia, Venezuela	KT768392	KT768357	KT768461	-	KT768428	KT768498

***Cycnoches cooperi* Rolfe**	Whitten W3591 (FLAS)	Brazil, Peru	KT768393	KT768358	KT768462	-	KT768429	KT768499
***Cycnoches densiflorum* Rolfe**	cult. BGH[2] Kusibab 5/2004	Colombia, Panama	KT768394	KT768359	KT768463	-	KT768430	KT768500
***Cycnoches dianae* Rchb. f.**	Pérez & Gerlach 1468 (M)	Panama	KT768395	KT768360	KT768464	-	KT768431	KT768501
***Cycnoches egertonianum* Bateman**	(1) Franke s.n. (MEXU)	Southern Mexico, Guatemala, Belize, Honduras	KT768397	KT768362	KT768466	-	KT768433	KT768503
	(2) cult. BGM 12/1471 (M)	Southern Mexico, Guatemala, Belize, Honduras	KT768396	KT768361	KT768465	-	KT768432	KT768502
***Cycnoches guttulatum* Schltr.**	Pérez & Gerlach 1476 (M)	Panama	KT768398	KT768363	KT768467	-	KT768434	KT768504
***Cycnoches haagii* Barb. Rodr.**	cult. BGH Brock 10/72	Surinam, Venezuela, Colombia, Ecuador, Brazil, Peru, Bolivia	KT768399	KT768364	KT768468	-	KT768435	KT768505
***Cycnoches herrenhusanum* Jenny & G.A. Romero**	cult. BGH Hubein 1/78	Colombia	KT768400	KT768365	KT768469	-	KT768436	KT768506
***Cycnoches lehmannii* Rchb. f.**	cult. BGH Portilla T1/97	Ecuador, Peru	KT768401	KT768366	KT768470	-	KT768437	KT768507
***Cycnoches loddigesii* Lindl.**	cult. BGH H9/70	Colombia, Surinam, Venezuela	KT768402	KT768367	KT768471	-	KT768438	KT768508
***Cycnoches manoelae* V.P. Castro & Campacci**	cult. BGM 12/2255 (M)	Brazil	KT768403	KT768368	KT768472	-	KT768439	KT768509
***Cycnoches pachydactylon* Schltr.**	Pérez & Gerlach 1469 (M)	Panama	KT768404	KT768369	KT768473	-	KT768440	KT768510
***Cycnoches pentadactylon* Lindl.**	cult. BGH Kusibab 1/11	Brazil, Peru	-	KT768370	KT768474	-	KT768441	KT768511
***Cycnoches peruvianum* Rolfe**	(1) cult. BGM 12/0839 (M)	Ecuador, Peru, Colombia	KT768406	KT768372	KT768475	-	KT768443	KT768513

	(2) cult. BGH Kusibab 5/04	Ecuador, Peru, Colombia	KT768405	KT768371	-	-	KT768442	KT768512
***Cycnoches suarezii* Dodson**	cult. BGM 12/0836 (M)	Ecuador	KT768408	KT768374	KT768476	-	KT768444	KT768515
***Cycnoches ventricosum* Bateman**	cult. BGM 3/3992 (M)	Southern Mexico, Guatemala, Belize, Honduras, northern Nicaragua	KT768409	KT768375	KT768477	-	KT768445	KT768516
***Cycnoches warszewiczii* Rchb. f.**	cult. BGH H1/73	Southern Nicaragua, Costa Rica, Panama	KT768410	KT768376	KT768478	-	KT768446	KT768517
***Cycnoches* sp. 1**	Rodríguez s.n. (M)	-	KT768407	KT768373	-	-	-	KT768514
Cycnoches sp. 2	SR1106	-	JF691909	-	-	-	-	JF692056
Cycnoches sp. 3	SR1139a	-	JF691912	-	-	-	-	JF692059
Dressleria dilecta (Rchb.f.) Dodson	Whitten 1019 (FLAS)	Colombia, Panama, Costa Rica, Nicaragua	AF239411	-	-	AF239507	-	EU490731.1
***Dressleria* sp.**	cult. BGM 11/1194 (M)	-	KT768413	KT768377	-	-	-	KT768521
Galeandra devoniana R.H. Schomb. ex Lindl.	(1) Silva 1373 (HUEFS); (2) Pupulin 1133 (JBL)	Brazil, Colombia, Guyana, Venezuela	(1) EU877142	(2) EU877125	-	(2) KF660268	-	(2) KF660330
Galeandra sp.	ML092	-	JF692011	-	-	-	-	JF692079
Grobya galeata Lindl.	MWC295	Brazil	AF470487	-	-	AF47045	-	-
***Mormodes badia* Rolfe ex Watson**	cult. BGM 2/2480 (M)	Mexico	KT768415	KT768380	KT768484	-	KT768450	KT768525
***Mormodes ephippilabia* Fowlie**	cult. BGM 3/0775 (M)	Honduras	KT768416	KT768381	KT768485	-	-	KT768526
***Mormodes luxata* Lindl.**	cult. BGM 92/3103 (M)	Mexico	KT768417	KT768382	KT768486	-	-	KT768527
***Mormodes punctata* Rolfe**	Pérez & Gerlach 1483 (M)	Panama	KT768418	KT768383	KT768487	-	-	KT768528
Outgroup								

***Ansellia africana* Lindl.**	cult. BGM X/0021 (M)	Sub-saharan Africa	-	-	KT768453	-	KT768420	KT768490
***Cymbidium eburneum* Lindl.**	cult. BGM (M)	Burma, China, India, Nepal, Vietnam	KT768411	-	KT768479	-	KT768447	KT768518
***Cymbidium tracyanym* Rolfe**	cult. BGM (M)	Burma, China, Thailand, Vietnam	KT768412	-	KT768480	-	-	KT768519
Cyrtopodium andersonii (Lamb. ex Andrews) R. Br.	(1) Chase O-341; (2) Chase "no voucher" (K)	Brazil, Colombia, Guyana, Surinam, Venezuela	(1) AF470490	-	-	(1) AF470460	-	(2) KF660329
Cyrtopodium punctatum (L.) Lindl.	Chase O-126 (K)	Middle-north South America to Mexico	AF239412	-	-	AF239508	-	-
***Eulophia petersii* Rchb. f.**	cult. BGM 11/3892 (M)	South Africa	-	-	KT768481	-	KT768448	KT768522
***Grammatophyllum measuresianum* Sander**	cult. BGM Stoch 6/95 (M)	Philippines	-	KT768379	KT768483	-	KT768449	KT768524
***Oeceoclades maculata* (Lindl.) Lindl.**	cult. BGM 96/4473 (M)	Tropical America, Africa	-	-	KT768488	-	KT768451	KT768529
***Oeceoclades pulchra* (Thouars) M.A.Clem. & P.J. Cribb**	cult. BGM X/434 (M)	Tropical Asia, Asutralia	KT768414	-	KT768482	-	-	KT768523
***Oncidium luteum* Rolfe**	cult. BGM 13/0100 (M)	Costa Rica - Panama	KT768419	-	KT768489	-	KT768452	KT768530

[1]Material cultivated at the Botanic Garden Munich (Baviera, Germany), [2]Material cultivated at the Botanic Garden Hanover (Lower Saxony, Germany).

Table S3

Data partition	AIC	LRT
ITS	GTR+Γ	GTR+Γ
ETS	TPM2uf+Γ	GTR+Γ
Xdh	HKY+Γ	GTR+Γ
matK	TVM+Γ	GTR+Γ
trnS-trnG	TVM1+Γ	GTR+Γ
ycf1	TVM+Γ	GTR+Γ

Table S4

(A)

	Proportion of incongruent associations (%)																											
	10								20								30								40			
	Additive tree				Branch lengths = 1				Additive tree				Branch lengths = 1				Additive tree				Branch lengths = 1				Additive tree			
	PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo	
Tree	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**
1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	2	0	2	2	3	0	2	2
S	**0.68**		0.91		0.68		0.87		0.73		0.87		**0.65**		0.83		0.66		0.84		0.6		0.81		0.56		0.67		0.6		0.66	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	2	0	1	0	4	0	4	0
S	0.75		0.91		0.73		0.89		0.66		0.8		0.62		0.8		0.65		0.77		0.62		0.77		0.6		0.69		0.55		0.72	
3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	0	2	0	5	0	2	0	1	0	5	0	1	4	1	0
S	0.73		0.89		0.73		0.89		0.76		0.86		0.62		0.83		0.64		0.8		0.59		0.76		0.61		0.71		0.51		0.67	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	2	0	1	0	0	0	3	0	5	0	4	0
S	0.73		0.89		0.66		0.86		0.73		0.83		0.7		0.81		0.65		0.8		0.63		0.78		0.6		0.75		0.59		0.71	
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	3	0	0	0	3	0	0	0	6	0	5	0
S	0.71		0.89		0.68		0.86		0.68		0.82		0.73		0.85		0.65		0.8		0.65		0.76		0.65		0.7		0.57		0.63	
6	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	2	0	1	0	2	0	2	0	2	0	2	0	6	0	4	0
S	0.8		0.89		0.75		0.9		0.72		0.89		0.67		0.85		0.67		0.77		0.6		0.8		0.64		0.77		0.63		0.72	
7	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0.75		0.93		**0.6**		0.88		0.72		0.87		0.67		0.79		0.67		0.82		0.61		0.8		0.56		0.72		0.53		0.73	
8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	6	0	2	0	3	0	3	0	3	0	4	0
S	**0.74**		0.92		**0.65**		0.88		0.62		0.82		0.59		0.81		0.69		0.8		0.66		0.78		0.64		0.79		0.6		0.71	
9	0	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0	0	0	1	0	1	0	1	0	0	0	2	1	6	0	1	1
S	0.71		0.87		0.74		0.86		0.67		0.84		0.62		0.8		0.56		0.74		0.53		0.77		0.54		0.7		0.48		0.61	
10	0	0	0	0	0	0	0	0	0	0	1	0	2	0	1	0	0	0	2	0	1	0	3	0	0	0	3	0	2	0	3	0
S	0.79		0.91		0.7		0.91		0.73		0.82		0.7		0.79		0.66		0.77		0.6		0.78		0.61		0.73		0.56		0.71	
Mis. T	1	0	0	0	2	0	1	0	1	0	1	0	9	0	3	0	8	0	8	0	22	0	13	0	13	0	21	3	36	4	28	3
Av. S	0.739		0.901		0.692		0.88		0.702		0.842		0.657		0.816		0.65		0.791		0.609		0.781		0.601		0.723		0.562		0.687	

(B)

	Proportion of incongruent associations (%)																													
	10								20								30								40					
	Additive tree				Branch lengths = 1				Additive tree				Branch lengths = 1				Additive tree				Branch lengths = 1				Additive tree				Branch lengths = 1			
	PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo	
Tree	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**
1	0	0	0	0	0	0	0	0	1	0	1	0	2	0	2	0	1	0	0	0	1	0	1	0	1	0	2	0	3	0	3	0
S	0.73		0.92		0.74		0.93		0.69		0.83		0.68		0.86		0.72		0.84		0.72		0.84		0.62		0.72		0.61		0.78	
2	0	0	0	0	0	0	0	0	1	0	0	0	3	0	2	0	0	0	0	0	2	0	1	0	3	0	3	0	3	0	5	0
S	0.75		0.92		0.74		0.91		0.63		0.82		0.63		0.86		0.66		0.8		0.68		0.83		0.66		0.76		0.61		0.75	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	5	0	4	0	3	0	2	0	6	0	3	0
S	0.7		0.9		**0.58**		0.8		0.7		0.87		0.71		0.88		0.68		0.89		0.6		0.79		0.63		0.76		0.61		0.77	
4	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	3	0	0	0	1	0	1	0	11	0	3	0
S	**0.63**		0.91		0.75		0.92		0.66		0.8		0.67		0.85		0.7		0.85		0.7		0.83		0.61		0.72		0.55		0.74	
5	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	2	0	1	0	2	0	1	2	2	0	0	0
S	0.72		0.9		0.7		0.9		0.68		0.85		0.66		0.86		0.69		0.83		0.67		0.84		0.63		0.75		0.59		0.76	
6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	3	0	1	0	4	0	1	0	1	0	3	0	8	2	6	5
S	0.7		0.89		0.7		0.89		0.71		0.83		0.69		0.86		0.66		0.81		0.57		0.77		0.71		0.78		0.51		0.6	
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	2	0	1	0	2	0	1	0	5	0	2	0
S	0.74		0.94		0.7		0.92		0.76		0.88		0.72		0.88		0.68		0.81		0.58		0.77		0.67		0.76		0.61		0.77	
8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	4	0	1	0	3	0	1	0	2	0	3	0	2	0	0	0
S	0.76		0.91		0.75		0.91		0.71		0.82		0.7		0.85		0.68		0.81		0.65		0.88		0.69		0.78		0.6		0.79	
9	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	4	0	2	0	4	0
S	0.71		0.88		0.74		0.91		0.69		0.87		0.67		0.88		0.65		0.8		0.62		0.82		0.65		0.76		0.64		0.77	
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	1	0	5	0	7	0	6	0	5	0
S	0.7		0.9		0.74		0.91		0.69		0.86		0.69		0.87		0.7		0.84		0.68		0.81		0.66		0.76		0.64		0.77	
Mis. T	0	0	0	0	2	0	1	0	3	0	1	0	9	0	7	0	12	0	4	0	27	0	11	0	20	0	27	2	48	2	31	5
Av. S	0.714		0.907		0.714		0.9		0.692		0.843		0.682		0.865		0.682		0.828		0.647		0.818		0.653		0.755		0.597		0.75	

(C)

	Proportion of outliers (%)																											
	10								20								30								40			
	Additive tree				Branch lengths = 1				Additive tree				Branch lengths = 1				Additive tree				Branch lengths = 1				Additive tree			
	PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo		PACo+PFL2		PACo	
Tree	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**	**x**	**c**
1	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	2	0	2	0	2	0	4	0	4	0	4	0
S	0.69		0.89		0.73		0.91		0.7		0.88		0.69		0.88		0.66		0.82		0.64		0.86		0.65		0.8		0.68		0.82	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	0	7	0	6	0	3	0	4	0	4	0	2	0
S	0.72		0.92		0.72		0.92		0.73		0.88		0.73		0.91		0.66		0.81		0.7		0.87		0.62		0.8		0.64		0.81	
3	0	0	0	0	1	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	1	0	1	0
S	0.74		0.93		0.74		0.91		0.71		0.88		0.68		0.88		0.7		0.81		0.73		0.84		0.66		0.81		0.63		0.8	
4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	0	1	0	2	19	8	1	24	0	2	0
S	0.72		0.91		0.69		0.9		0.69		0.87		0.7		0.88		0.68		0.82		0.67		0.85		0.59		0.62		0.58		0.78	
5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	3	0	2	0	33	1	22	5	2	0	3	0
S	0.72		0.92		0.54		0.92		**0.72**		0.89		0.68		0.88		0.65		0.82		0.61		0.86		0.5		0.63		0.57		0.76	
6	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	1	0	1	1	3	0	2	0	5	0	3	4	7	0	2	0
S	0.73		0.93		0.71		0.92		0.68		0.87		0.7		0.89		0.66		0.82		0.67		0.87		0.63		0.7		0.67		0.82	
7	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	2	0	1	0	15	5	9	5	13	0	8	0
S	0.71		0.91		0.73		0.91		0.68		0.87		0.68		0.88		0.76		0.84		0.73		0.87		0.5		0.65		0.57		0.77	
8	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	1	0	3	4	3	4	4	0	4	0
S	0.71		0.9		0.75		0.91		0.7		0.86		0.69		0.89		0.66		0.81		0.66		0.84		0.63		0.7		0.69		0.82	
9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	2	0	0	0	2	0	2	0	2	0	5	0	4	0	1	0
S	0.71		0.91		0.72		0.92		0.7		0.84		0.75		0.9		0.67		0.81		0.67		0.86		0.58		0.76		0.6		0.76	
10	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	4	0	8	0	4	11	3	0
S	0.73		0.92		0.5		0.89		**0.67**		0.84		0.69		0.89		0.67		0.81		0.69		0.88		0.59		0.76		0.54		0.72	
Mis. T	0	0	0	0	5	0	5	0	4	0	0	0	6	0	4	0	7	0	2	1	26	0	17	0	70	29	68	19	67	11	30	0
Av. S	0.718		0.914		0.683		0.911		0.698		0.868		0.699		0.888		0.677		0.817		0.677		0.86		0.595		0.723		0.617		0.786	

Table S5

Loci	Lengh (bp)	Parsimony Informative Sites	Number of cells
ETS	475	149 / 32%	35/61
ITS	705	320 / 46%	57/61
Xdh	991	115 / 12%	37/61
matK	1721	76 / 4%	8/61
trn S-G	936	107 / 11%	34/61
ycf1	1643	209 / 8%	55/61