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., $e_i^2 = e_i^2/m_{XY}^2$). In case of complete congruence between both phylogenies, the $e_i^2$’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 $e_i^2$ 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 $e_i^2$ 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 $pfl2_i$ statistic, the respective $e_i^2$ 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 \((matK, trnL-trnF, trnS-trnG)\) 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.

```R
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.

```R
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[rownames(NCP), colnames(NCP)]
  DH.top <- DH.top[rownames(NCP),rownames(NCP)]
  DP.top <- DP.top[rownames(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 organelar 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 set 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 $m_{XY}^2$ value. Therefore, consensus organellar and nuclear trees (see data input) must be transformed into matrices of patristic distances:

\[
N.D \leftarrow \text{cophenetic} (\text{NTree})
\]
\[
CP.D \leftarrow \text{cophenetic} (\text{CPTree})
\]

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

\[
N.D \leftarrow N.D[\text{rownames(NCP)}, \text{rownames(NCP)}]
\]
\[
CP.D \leftarrow CP.D [\text{colnames(NCP)}, \text{colnames(NCP)}]
\]

Finally, to apply PACo:

\[
PACo.fit \leftarrow \text{PACo.dV}(N.D, CP.D, NCP)
\]
\[
NCP.proc \leftarrow \text{procrustes}(\text{PACo.fit$H.PCo}, \text{PACo.fit$P.PCo})
\]

The following syntax computes the residual sum of squares $m_{XY}^2$ 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.

\[
m2.obs \leftarrow NCP.proc$ss
\]
\[
N.perm = 10000
\]
\[
P.value = 0
\]
\[
\text{set.seed}(2)
\]
\[
\text{for (n in c(1:N.perm))}
\]
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 \( (e_i^2) \) 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 \( e_i^2 \) 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)
Next, outlier associations will be spotted by the pipeline using a threshold value (1/N). The following syntax will transform the $\epsilon_i^2$'s into $\epsilon_i^2$'s obtained from either phylograms or unit branch length trees and will compute their respective median. Given the asymmetric distribution of the $\epsilon_i^2$'s, the median value was preferred over the mean as central tendency estimate:

\[
m_2A <- \text{apply}(\text{pac.add}, 1, \text{sum})
\]

\[
pac.norm.add <- \text{pac.add}/m_2A
\]

\[
m_2T <- \text{apply}(\text{pac.top}, 1, \text{sum})
\]

\[
pac.norm.top <- \text{pac.top}/m_2T
\]

To plot the median $\epsilon_i^2$ 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:

\[
op <- \text{par}(\text{oma}=c(3,2,1,1))
\]

\[
\text{par (mfrow=}c(1,1),\text{mar} = c(4,4,1,1))
\]

\[
mA <- \text{apply}(\text{pac.norm.add}, 2, \text{median})
\]

\[
uCI.A <- \text{apply}(\text{pac.norm.add}, 2, \text{quantile}, \text{probs} = 0.975)
\]

\[
lCI.A <- \text{apply}(\text{pac.norm.add}, 2, \text{quantile}, \text{probs} = 0.025)
\]

\[
cols <- c("lightgreen", "mistyrose")[(mA > 1/NLinks) + 1]
\]

\[
\text{barplot2}(mA, \text{main} = "PAco squared residuals - additive trees", \\
\text{xlab}="Association", \text{ylab}="Normalized PAco sqr. residuals", \\
\text{cex.axis}=0.5, \text{col}=cols, \text{border}="lightgrey", \\
\text{names.arg}=\text{colnamesPACo}, \text{las}=2, \text{cex.names}=0.5, \text{plot.ci}=T, \\
\text{ci.l}=lCI.A, \text{ci.u}=uCI.A, \text{ci.color}="blue")
\]

\[
\text{abline}(h=1/NLinks, \text{col}="red")
\]

\[
mA <- \text{apply}(\text{pac.norm.top}, 2, \text{median})
\]

\[
uCI.A <- \text{apply}(\text{pac.norm.top}, 2, \text{quantile}, \text{probs} = 0.975)
\]
Two plots (Fig. S3, data with $\varepsilon_i^2$’s obtained from unit branch length trees not shown) of all squared residual values determined from each ‘o’-‘n’-association, and obtained from phylogenograms and unit branch lengths as well, will be plotted, respectively. Associations with $\varepsilon_i^2$’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:

```r
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")
```

```r
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=1/NLinks, 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 $\varepsilon_i^2$ and $pfl2_i$ values combined and 2) using median $\varepsilon_i^2$’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 ($\varepsilon_i^2$ and $pfl2_i$):

```r
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.

\[
nclust = \text{my } k
\]

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

```r
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:

```r
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"),
This script will plot the consensus trees of each data set analyzed, with the corresponding OTUs names. Their individual $\varepsilon_i^2$ 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 $\varepsilon_i^2$ 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


**Figure S1.** Vector diagrams of squared residual values $\varepsilon_i^2$ 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 $\varepsilon_i^2$ 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 $\varepsilon_i^2$ 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 $\varepsilon_i^2$ 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 $\varepsilon_i^2$ 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 $\epsilon_i^2$ 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 Partitioning Around Medoids 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)
ParaFitLink 2

PACo

PACo
Fig. S2

a)
ParaFitLink 2

PACo

PACo
<table>
<thead>
<tr>
<th>Loci</th>
<th>Primer</th>
<th>Sequence</th>
<th>Reference</th>
<th>Pre-melt</th>
<th>Amplification</th>
<th>Final extension</th>
<th>Number of amplification cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITS</td>
<td>ITS 4</td>
<td>TCC-TCC-GCT-TAT-TGA-TAT-GC</td>
<td>Baldwin (1992)</td>
<td>95°C (3 min)</td>
<td>95°C (30 secs) + 52°C (1 min) + 68°C (1 min)</td>
<td>68°C (10 min)</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>ITS 5</td>
<td>GGA-AGT-AAA-AGT-CGT-AAC-AAG-G</td>
<td></td>
<td>95°C (3 min)</td>
<td>95°C (30 secs) + 52°C (1 min) + 68°C (1 min)</td>
<td>68°C (10 min)</td>
<td>39</td>
</tr>
<tr>
<td>ETS</td>
<td>EST-Orchid</td>
<td>CAT-ATG-AGT-TGT-TGC-GGA-CC (AT)-T</td>
<td>Monteiro et al (2010)</td>
<td>95°C (3 min)</td>
<td>95°C (30 secs) + 52°C (1 min) + 68°C (1 min)</td>
<td>68°C (10 min)</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>18-IGS</td>
<td>AGA-CAA-GCA-TAT-GAC-TAC-TGG-CAG-G</td>
<td>Markos and Balwin (1998)</td>
<td>95°C (3 min)</td>
<td>95°C (30 secs) + 52°C (1 min) + 68°C (1 min)</td>
<td>68°C (10 min)</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>X502F</td>
<td>TGT-GAT-GTC-GAT-GTA-TGC</td>
<td>Górnia et al (2010)</td>
<td>95°C (3 min)</td>
<td>95°C (30 secs) + 53°C (1 min) + 68°C (1.5 min)</td>
<td>68°C (10 min)</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>X1599R</td>
<td>G(AT)G-AGA-GAA-A(CT)TG-GAG-CAA-C</td>
<td></td>
<td>95°C (3 min)</td>
<td>95°C (30 secs) + 53°C (1 min) + 68°C (1.5 min)</td>
<td>68°C (10 min)</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>3720F</td>
<td>TAC-GTA-TGT-AAT-GAA-CGA-ATG-G</td>
<td>Neubig et al (2009)</td>
<td>95°C (3 min)</td>
<td>95°C (30 secs) + 54°C (1 min) + 68°C (1.5 min)</td>
<td>68°C (10 min)</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>5500R</td>
<td>GCT-GTT-ATT-GGC-ATC-AAA-CCA-ATA-GCG</td>
<td></td>
<td>95°C (3 min)</td>
<td>95°C (30 secs) + 54°C (1 min) + 68°C (1.5 min)</td>
<td>68°C (10 min)</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>trn-S(GCU)</td>
<td>GCC-GCT-TTA-GTC-CAC-TCA-GC</td>
<td>Hamilton (1999)</td>
<td>95°C (3 min)</td>
<td>95°C (30 secs) + 51.5°C (1 min) + 68°C (1.5 min)</td>
<td>68°C (10 min)</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>trn-G(UCC)</td>
<td>GAA-CGA-ATC-ACA-CTT-TTA-CCA-C</td>
<td></td>
<td>95°C (3 min)</td>
<td>95°C (30 secs) + 51.5°C (1 min) + 68°C (1.5 min)</td>
<td>68°C (10 min)</td>
<td>39</td>
</tr>
<tr>
<td>Taxon</td>
<td>DNA Source - voucher</td>
<td>Distribution</td>
<td>Nuclear - ribosomal dataset</td>
<td>Chloroplast dataset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>----------------------</td>
<td>--------------------------------</td>
<td>-----------------------------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Catasetum collare Cogn.</strong></td>
<td>cult. BGM(^1) 5/1000 (M)</td>
<td>Brasil, Colombia, Ecuador, Venezuela</td>
<td>KT768384</td>
<td>KT768350</td>
<td>KT768454</td>
<td>-</td>
<td>KT768421</td>
</tr>
<tr>
<td><strong>Catasetum juruenense Hoehne</strong></td>
<td>cult. BGM 5/1223 (M)</td>
<td>Brazil</td>
<td>KT768385</td>
<td>KT768351</td>
<td>KT768455</td>
<td>-</td>
<td>KT768422</td>
</tr>
<tr>
<td><strong>Catasetum macrocarpum Rich. ex Kunth</strong></td>
<td>cult. BGM 96/3071 (M)</td>
<td>Brazil-Venezuela</td>
<td>KT768386</td>
<td>KT768352</td>
<td>KT768456</td>
<td>-</td>
<td>KT768423</td>
</tr>
<tr>
<td><strong>Catasetum meae Pabst</strong></td>
<td>cult. BGM 97/3836 (M)</td>
<td>Brazil</td>
<td>KT768387</td>
<td>KT768353</td>
<td>KT768457</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Catasetum x roseoalbum (Hook.) Lindl.</strong></td>
<td>cult. BGM 6/2496 (M)</td>
<td>Venezuela</td>
<td>KT768388</td>
<td>KT768354</td>
<td>KT768458</td>
<td>-</td>
<td>KT768424</td>
</tr>
<tr>
<td><strong>Clowesia russelliana (Hook.) Dodson</strong></td>
<td>cult. BGM 98/2889 (M)</td>
<td>Central America, Colombia, Venezuela</td>
<td>KT768389</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>KT768425</td>
</tr>
<tr>
<td><strong>Clowesia sp. 1</strong></td>
<td>ML086</td>
<td>-</td>
<td>JF692010</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Clowesia sp. 2</strong></td>
<td>ML301</td>
<td>-</td>
<td>JF692017</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Clowesia sp. 3</strong></td>
<td>SR1153</td>
<td>-</td>
<td>JF691914</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Clowesia sp. 4</strong></td>
<td>SR1203</td>
<td>-</td>
<td>JF691923</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Clowesia sp. 5</strong></td>
<td>SR1213</td>
<td>-</td>
<td>JF691925</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Clowesia sp. 6</strong></td>
<td>SR1463</td>
<td>-</td>
<td>JF691960</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Cymnoches aureum Lindl. &amp; Paxton</strong></td>
<td>Pt. BGM 1473 (M)</td>
<td>Panama</td>
<td>KT768390</td>
<td>KT768355</td>
<td>KT768459</td>
<td>-</td>
<td>KT768426</td>
</tr>
<tr>
<td><strong>Cymnoches barthiorum G.F.Carr &amp; Christenson</strong></td>
<td>Pt. BGM 12/1476 (M)</td>
<td>Colombia</td>
<td>KT768391</td>
<td>KT768356</td>
<td>KT768460</td>
<td>-</td>
<td>KT768427</td>
</tr>
<tr>
<td><strong>Cymnoches chlorochilon Klotzch</strong></td>
<td>Pt. BGM 94/981 (M)</td>
<td>Panama, Colombia, Venezuela</td>
<td>KT768392</td>
<td>KT768357</td>
<td>KT768461</td>
<td>-</td>
<td>KT768428</td>
</tr>
<tr>
<td><strong>Cycnoches cooperi</strong> Rolfe</td>
<td>Whitten W3591 (FLAS)</td>
<td>Brazil, Peru</td>
<td>KT768393</td>
<td>KT768358</td>
<td>KT768462</td>
<td>-</td>
<td>KT768429</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------</td>
<td>--------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----</td>
<td>----------</td>
</tr>
<tr>
<td><strong>Cycnoches densiflorum</strong> Rolfe</td>
<td>cult. BGH² Kusibab 5/2004</td>
<td>Colombia, Panama</td>
<td>KT768394</td>
<td>KT768359</td>
<td>KT768463</td>
<td>-</td>
<td>KT768430</td>
</tr>
<tr>
<td><strong>Cycnoches dianae</strong> Rchb. f.</td>
<td>Pérez &amp; Gerlach 1468 (M)</td>
<td>Panama</td>
<td>KT768395</td>
<td>KT768360</td>
<td>KT768464</td>
<td>-</td>
<td>KT768431</td>
</tr>
<tr>
<td><strong>Cycnoches egertonianum</strong> Bateman</td>
<td>(1) Franke s.n. (MEXU)</td>
<td>Southern Mexico, Guatemala, Belize, Honduras</td>
<td>KT768397</td>
<td>KT768362</td>
<td>KT768466</td>
<td>-</td>
<td>KT768433</td>
</tr>
<tr>
<td></td>
<td>(2) cult. BGM 12/1471 (M)</td>
<td>Southern Mexico, Guatemala, Belize, Honduras</td>
<td>KT768396</td>
<td>KT768361</td>
<td>KT768465</td>
<td>-</td>
<td>KT768432</td>
</tr>
<tr>
<td><strong>Cycnoches guttulatum</strong> Schltr.</td>
<td>Pérez &amp; Gerlach 1476 (M)</td>
<td>Panama</td>
<td>KT768398</td>
<td>KT768363</td>
<td>KT768467</td>
<td>-</td>
<td>KT768434</td>
</tr>
<tr>
<td><strong>Cycnoches haagii</strong> Barb. Rodr.</td>
<td>cult. BGH Brock 10/72</td>
<td>Surinam, Venezuela, Colombia, Ecuador, Brazil, Peru, Bolivia</td>
<td>KT768399</td>
<td>KT768364</td>
<td>KT768468</td>
<td>-</td>
<td>KT768435</td>
</tr>
<tr>
<td><strong>Cycnoches herrenhusanum</strong> Jenny &amp; G.A. Romero</td>
<td>cult. BGH Hubein 1/78</td>
<td>Colombia</td>
<td>KT768400</td>
<td>KT768365</td>
<td>KT768469</td>
<td>-</td>
<td>KT768436</td>
</tr>
<tr>
<td><strong>Cycnoches lehmannii</strong> Rchb. f.</td>
<td>cult. BGH Portilla T1/97</td>
<td>Ecuador, Peru</td>
<td>KT768401</td>
<td>KT768366</td>
<td>KT768470</td>
<td>-</td>
<td>KT768437</td>
</tr>
<tr>
<td><strong>Cycnoches loddigesii</strong> Lindl.</td>
<td>cult. BGH H9/70</td>
<td>Colombia, Surinam, Venezuela</td>
<td>KT768402</td>
<td>KT768367</td>
<td>KT768471</td>
<td>-</td>
<td>KT768438</td>
</tr>
<tr>
<td><strong>Cycnoches manoelae</strong> V.P. Castro &amp; Campacci</td>
<td>cult. BGM 12/2255 (M)</td>
<td>Brazil</td>
<td>KT768403</td>
<td>KT768368</td>
<td>KT768472</td>
<td>-</td>
<td>KT768439</td>
</tr>
<tr>
<td><strong>Cycnoches pachydactylon</strong> Schltr.</td>
<td>Pérez &amp; Gerlach 1469 (M)</td>
<td>Panama</td>
<td>KT768404</td>
<td>KT768369</td>
<td>KT768473</td>
<td>-</td>
<td>KT768440</td>
</tr>
<tr>
<td><strong>Cycnoches pentadactylon</strong> Lindl.</td>
<td>cult. BGH Kusibab 1/11</td>
<td>Brazil, Peru</td>
<td>-</td>
<td>KT768370</td>
<td>KT768474</td>
<td>-</td>
<td>KT768441</td>
</tr>
<tr>
<td><strong>Cycnoches peruvianum</strong> Rolfe</td>
<td>(1) cult. BGM 12/0839 (M)</td>
<td>Ecuador, Peru, Colombia</td>
<td>KT768406</td>
<td>KT768372</td>
<td>KT768475</td>
<td>-</td>
<td>KT768443</td>
</tr>
<tr>
<td>Species</td>
<td>Cultivar</td>
<td>Location of Origin</td>
<td>GenBank Accession Numbers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cycnoches suarezii</em> Dodson</td>
<td>cult. BGM 12/0836 (M)</td>
<td>Ecuador</td>
<td>KT768408, KT768374, KT768476, KT768444, KT768515</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cycnoches ventricosum</em> Bateman</td>
<td>cult. BGM 3/3992 (M)</td>
<td>Southern Mexico, Guatemala, Belize, Honduras, northern Nicaragua</td>
<td>KT768409, KT768375, KT768477, KT768445, KT768516</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cycnoches warszewiczii</em> Rchb. f.</td>
<td>cult. BGH H1/73</td>
<td>Southern Nicaragua, Costa Rica, Panama</td>
<td>KT768410, KT768376, KT768478, KT768446, KT768517</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cycnoches sp. 1</em> Rodríguez s.n. (M)</td>
<td></td>
<td></td>
<td>KT768407, KT768373</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cycnoches sp. 2</em> SR1106</td>
<td></td>
<td></td>
<td>JF691909</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cycnoches sp. 3</em> SR1139a</td>
<td></td>
<td></td>
<td>JF691912</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Dressleria ditecta</em> (Rchb.f.) Dodson</td>
<td>Whitten 1019 (FLAS)</td>
<td>Colombia, Panama, Costa Rica, Nicaragua</td>
<td>AF239411</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Dressleria sp.</em></td>
<td>cult. BGM 11/1194 (M)</td>
<td></td>
<td>KT768413, KT768377</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Galeandra devoniana</em> R.H. Schomb. ex Lindl.</td>
<td>(1) Silva 1373 (HUEFS); (2) Pupulin 1133 (JBL)</td>
<td>Brazil, Colombia, Guyana, Venezuela</td>
<td>(1) EU877142, (2) EU877125, (2) KF660268</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Galeandra sp.</em> ML092</td>
<td></td>
<td></td>
<td>JF692011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Grobya galeata</em> Lindl.</td>
<td>MWC295</td>
<td>Brazil</td>
<td>AF470487</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mormodes badia</em> Rolfe ex Watson</td>
<td>cult. BGM 2/2480 (M)</td>
<td>Mexico</td>
<td>KT768415, KT768380, KT768484, KT768450, KT768525</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mormodes ephippilabia</em> Fowlie</td>
<td>cult. BGM 3/0775 (M)</td>
<td>Honduras</td>
<td>KT768416, KT768381, KT768485, KT768526</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mormodes luxata</em> Lindl.</td>
<td>cult. BGM 92/3103 (M)</td>
<td>Mexico</td>
<td>KT768417, KT768382, KT768486, KT768527</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mormodes punctata</em> Rolfe</td>
<td>Pérez &amp; Gerlach 1483 (M)</td>
<td>Panama</td>
<td>KT768418, KT768383, KT768487, KT768528</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Outgroup
<table>
<thead>
<tr>
<th>Species</th>
<th>Cultivar</th>
<th>Origin</th>
<th>Accession Numbers</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ansellia africana</strong> Lindl.</td>
<td>BGM X/0021 (M)</td>
<td>Sub-saharan Africa</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Cymbidium eburneum</strong> Lindl.</td>
<td>BGM (M)</td>
<td>Burma, China, India, Nepal, Vietnam</td>
<td>KT768411</td>
<td>-</td>
</tr>
<tr>
<td><strong>Cymbidium tracyanym</strong> Rolfe</td>
<td>BGM (M)</td>
<td>Burma, China, Thailand, Vietnam</td>
<td>KT768412</td>
<td>KT768480</td>
</tr>
<tr>
<td><strong>Cyrtopodium andersonii</strong> (Lamb. ex Andrews) R. Br.</td>
<td>(1) Chase O-341; (2) Chase &quot;no voucher&quot; (K)</td>
<td>Brazil, Colombia, Guyana, Surinam, Venezuela</td>
<td>(1) AF470490</td>
<td>(1) AF470460</td>
</tr>
<tr>
<td><strong>Cyrtopodium punctatum</strong> (L.) Lindl.</td>
<td>Chase O-126 (K)</td>
<td>Middle-north South America to Mexico</td>
<td>AF239412</td>
<td>AF239508</td>
</tr>
<tr>
<td><strong>Eulophia petersii</strong> Rchb. f.</td>
<td>BGM 11/3892 (M)</td>
<td>South Africa</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Grammatophyllum measuresianum</strong> Sander</td>
<td>BGM Stoch 6/95 (M)</td>
<td>Philippines</td>
<td>-</td>
<td>KT768379</td>
</tr>
<tr>
<td><strong>Oeceoclades maculata</strong> (Lindl.) Lindl.</td>
<td>BGM 96/4473 (M)</td>
<td>Tropical America, Africa</td>
<td>-</td>
<td>KT768848</td>
</tr>
<tr>
<td><strong>Oeceoclades pulchra</strong> (Thouars) M.A.Clem. &amp; P.J. Cribb</td>
<td>BGM X/434 (M)</td>
<td>Tropical Asia, Australia</td>
<td>KT768414</td>
<td>KT768482</td>
</tr>
<tr>
<td><strong>Oncidium luteum</strong> Rolfe</td>
<td>BGM 13/0100 (M)</td>
<td>Costa Rica - Panama</td>
<td>KT768419</td>
<td>KT768489</td>
</tr>
</tbody>
</table>

1Material cultivated at the Botanic Garden Munich (Baviera, Germany), 2Material cultivated at the Botanic Garden Hanover (Lower Saxony, Germany).
<table>
<thead>
<tr>
<th>Data partition</th>
<th>AIC</th>
<th>LRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITS</td>
<td>GTR+Γ</td>
<td>GTR+Γ</td>
</tr>
<tr>
<td>ETS</td>
<td>TPM2uf+Γ</td>
<td>GTR+Γ</td>
</tr>
<tr>
<td>Xdh</td>
<td>HKY+Γ</td>
<td>GTR+Γ</td>
</tr>
<tr>
<td>matK</td>
<td>TVM+Γ</td>
<td>GTR+Γ</td>
</tr>
<tr>
<td>trnS-trnG</td>
<td>TVM1+Γ</td>
<td>GTR+Γ</td>
</tr>
<tr>
<td>ycf1</td>
<td>TVM+Γ</td>
<td>GTR+Γ</td>
</tr>
</tbody>
</table>
Table S4

(A) Table highlighting the proportion of incongruent associations (%).

<table>
<thead>
<tr>
<th>Tree</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Additive tree</td>
<td>Branch lengths = 1</td>
<td>Additive tree</td>
<td>Branch lengths = 1</td>
</tr>
<tr>
<td></td>
<td>PACo+PFL2</td>
<td>PACo</td>
<td>PACo+PFL2</td>
<td>PACo</td>
</tr>
<tr>
<td>S 1</td>
<td>0.68</td>
<td>0.91</td>
<td>0.68</td>
<td>0.87</td>
</tr>
<tr>
<td>S 2</td>
<td>0.75</td>
<td>0.91</td>
<td>0.73</td>
<td>0.89</td>
</tr>
<tr>
<td>S 3</td>
<td>0.73</td>
<td>0.89</td>
<td>0.73</td>
<td>0.89</td>
</tr>
<tr>
<td>S 4</td>
<td>0.73</td>
<td>0.89</td>
<td>0.66</td>
<td>0.86</td>
</tr>
<tr>
<td>S 5</td>
<td>0.71</td>
<td>0.89</td>
<td>0.68</td>
<td>0.86</td>
</tr>
<tr>
<td>S 6</td>
<td>0.8</td>
<td>0.89</td>
<td>0.75</td>
<td>0.9</td>
</tr>
<tr>
<td>S 7</td>
<td>0.75</td>
<td>0.93</td>
<td>0.6</td>
<td>0.88</td>
</tr>
<tr>
<td>S 8</td>
<td>0.74</td>
<td>0.92</td>
<td>0.65</td>
<td>0.88</td>
</tr>
<tr>
<td>S 9</td>
<td>0.71</td>
<td>0.87</td>
<td>0.74</td>
<td>0.86</td>
</tr>
<tr>
<td>S 10</td>
<td>0.79</td>
<td>0.91</td>
<td>0.7</td>
<td>0.91</td>
</tr>
<tr>
<td>Mis. T</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Av. S</td>
<td>0.739</td>
<td>0.901</td>
<td>0.692</td>
<td>0.88</td>
</tr>
<tr>
<td>Tree</td>
<td>x</td>
<td>c</td>
<td>x</td>
<td>c</td>
</tr>
<tr>
<td>S</td>
<td>0.73</td>
<td>0.92</td>
<td>0.74</td>
<td>0.93</td>
</tr>
<tr>
<td>2</td>
<td>0.75</td>
<td>0.92</td>
<td>0.74</td>
<td>0.91</td>
</tr>
<tr>
<td>S</td>
<td>0.7</td>
<td>0.9</td>
<td>0.58</td>
<td>0.8</td>
</tr>
<tr>
<td>4</td>
<td>0.63</td>
<td>0.91</td>
<td>0.75</td>
<td>0.92</td>
</tr>
<tr>
<td>S</td>
<td>0.72</td>
<td>0.9</td>
<td>0.7</td>
<td>0.9</td>
</tr>
<tr>
<td>6</td>
<td>0.7</td>
<td>0.89</td>
<td>0.7</td>
<td>0.89</td>
</tr>
<tr>
<td>S</td>
<td>0.74</td>
<td>0.94</td>
<td>0.7</td>
<td>0.92</td>
</tr>
<tr>
<td>8</td>
<td>0.76</td>
<td>0.91</td>
<td>0.75</td>
<td>0.91</td>
</tr>
<tr>
<td>S</td>
<td>0.71</td>
<td>0.88</td>
<td>0.74</td>
<td>0.91</td>
</tr>
<tr>
<td>10</td>
<td>0.7</td>
<td>0.9</td>
<td>0.74</td>
<td>0.91</td>
</tr>
<tr>
<td>S</td>
<td>0.714</td>
<td>0.907</td>
<td>0.714</td>
<td>0.9</td>
</tr>
<tr>
<td>Tree</td>
<td>Additive tree</td>
<td>Branch lengths = 1</td>
<td>Additive tree</td>
<td>Branch lengths = 1</td>
</tr>
<tr>
<td>------</td>
<td>---------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>1</td>
<td>0.69</td>
<td>0.89</td>
<td>0.73</td>
<td>0.91</td>
</tr>
<tr>
<td>2</td>
<td>0.72</td>
<td>0.92</td>
<td>0.72</td>
<td>0.92</td>
</tr>
<tr>
<td>3</td>
<td>0.74</td>
<td>0.93</td>
<td>0.74</td>
<td>0.91</td>
</tr>
<tr>
<td>4</td>
<td>0.72</td>
<td>0.91</td>
<td>0.69</td>
<td>0.89</td>
</tr>
<tr>
<td>5</td>
<td>0.72</td>
<td>0.92</td>
<td>0.54</td>
<td>0.92</td>
</tr>
<tr>
<td>6</td>
<td>0.73</td>
<td>0.93</td>
<td>0.71</td>
<td>0.92</td>
</tr>
<tr>
<td>7</td>
<td>0.71</td>
<td>0.91</td>
<td>0.73</td>
<td>0.91</td>
</tr>
<tr>
<td>8</td>
<td>0.71</td>
<td>0.9</td>
<td>0.75</td>
<td>0.91</td>
</tr>
<tr>
<td>9</td>
<td>0.71</td>
<td>0.9</td>
<td>0.72</td>
<td>0.92</td>
</tr>
<tr>
<td>10</td>
<td>0.73</td>
<td>0.92</td>
<td>0.5</td>
<td>0.89</td>
</tr>
<tr>
<td>Mis. T</td>
<td>0.71</td>
<td>0.91</td>
<td>0.68</td>
<td>0.91</td>
</tr>
<tr>
<td>Loci</td>
<td>Length (bp)</td>
<td>Parsimony Informative Sites</td>
<td>Number of cells</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-----------------------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>ETS</td>
<td>475</td>
<td>149 / 32%</td>
<td>35/61</td>
<td></td>
</tr>
<tr>
<td>ITS</td>
<td>705</td>
<td>320 / 46%</td>
<td>57/61</td>
<td></td>
</tr>
<tr>
<td>Xdh</td>
<td>991</td>
<td>115 / 12%</td>
<td>37/61</td>
<td></td>
</tr>
<tr>
<td>matK</td>
<td>1721</td>
<td>76 / 4%</td>
<td>8/61</td>
<td></td>
</tr>
<tr>
<td>trn S-G</td>
<td>936</td>
<td>107 / 11%</td>
<td>34/61</td>
<td></td>
</tr>
<tr>
<td>ycf1</td>
<td>1643</td>
<td>209 / 8%</td>
<td>55/61</td>
<td></td>
</tr>
</tbody>
</table>