

Comparison of two trees

General information

Both trees are **rooted**. Both trees contain **91 nodes**. Both trees contain **46 leaves**. Both trees have branch lengths. 40 leaves are shared between the two trees; **6 leaves** are present only in the first tree, while **6** are present only in the second tree. These are summarised in the following table:

In both trees

Only in first tree

Only in second tree

Acaryochloris.marina.MBIC1 1017	Anabaena.cylindrica.PCC712 2	Amborella.trichopoda.vascC T
Angiopteris.evecta.vascCT	Anabaena.variabilis.ATCC29 413	Arthrospira.sp.PCC8005
Anthoceros.formosae.strCT	Chlorella.vulgaris.cCT	Cryptomeria.japonica.vascC T
Chaetosphaeridium.globosum .strCT	Fistulifera.sp.JPCC.DA0580 .bacCT	Nostoc.punctiforme.PCC7310 2
Coccomya.sp.C169	Pseudanabaena.biceps.PCC74 29	P.marinus.subsp.marinus.st r.CCMP1375
Coffea.arabica.vascCT	Trichodesmium.erythraeum.I MS101	Pleurocapsa.sp.PCC7327
Cyanobium.gracile.PCC6307		
Cyanophora.paradoxa.glaucT		
Cyanothece.sp.CCY0110		
Cyanothece.sp.PCC7424		
Cyanothece.sp.PCC7425		
Fischerella.sp.JSC11		
Floydiella.terrestris.cCT		
Gnetum.parvifolium.vascCT		
Illicium.oligandrum.vascCT		
Lyngbya.majuscula.3L		
Mesostigma.viride.strCT		
Microcystis.aeruginosa.NIE S843		
Nephroselmis.olivacea.cCT		
Nodosilinea.nodulosa.PCC71 04		
Nostoc.azollae.0708		
Oscillatoria.sp.PCC6506		
P.marinus.str.AS9601		
P.marinus.str.MIT9303		
P.marinus.str.MIT9515		
Pedinomonas.minor.cCT		
Pinus.contorta.vascCT		
Porphyra.purpurea.rCT		
Pseudanabaena.sp.PCC7367		
Stigeoclonium.helveticum.c CT		
Synechococcus.sp.JA33Ab		
Synechococcus.sp.PCC7002		
Synechococcus.sp.RCC307		
Synechococcus.sp.RS9916		
Synechococcus.sp.WH7803		
Synechococcus.sp.WH8102		
Synechococcus.spongiarum.S H4		
Synedra.acus.bacCT		
Thalassiosira.pseudonana.b acCT		
Zygnema.circumcarinatum.st rCT		

The following analyses are based on the subset of leaves shared between the two trees.

The trees have different topologies. The Robinson-Foulds distance between them is **38**. The weighted Robinson-Foulds distance between them is **0.7975**. There are **60 splits in common** between the two trees. There are **19 splits** that only appear in the first tree, and **19 splits** that only appear in the second tree.

Considering the 60 splits in common between the two trees, the average difference in length between a split from the first tree and the corresponding split from the second tree is **0.01187** (89% highest-density interval: -0.02723 — 0.09020). [Figure 1](#) shows the distribution of the difference in length between splits from the first tree and the

corresponding splits from the second tree.

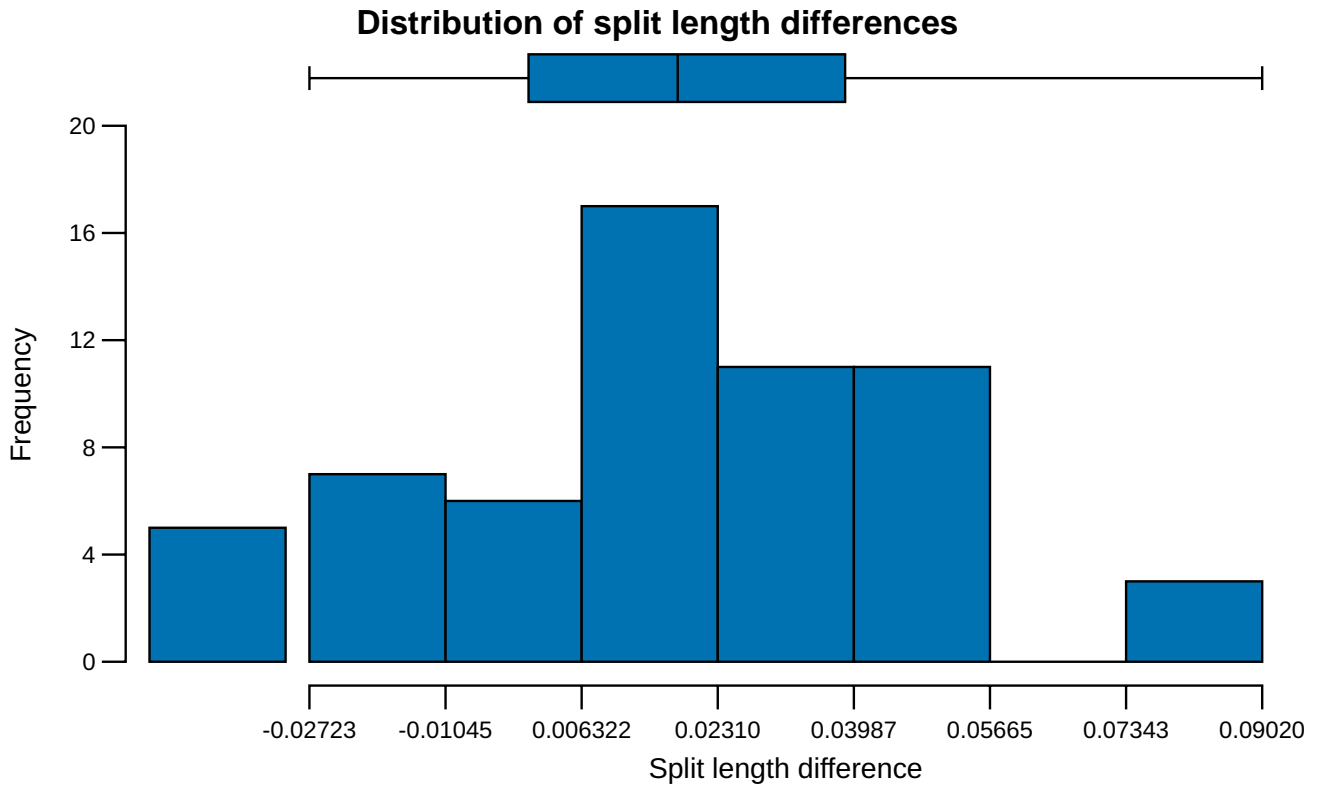


Figure 1. Distribution of split length differences. The histogram shows the distribution of the difference in length between splits from the first tree and the corresponding splits from the second tree. 5 values smaller than -0.02723 are shown in the underflow bin. The box and whisker plot at the top represents the median branch length, the interquartile range, and the 89% HDI.

Tree shape statistics

The **Sackin indices** of the trees are **321** (YHK model normalisation: 0.1449, PDA model normalisation: 1.03) and **259** (YHK model normalisation: -1.2029, PDA model normalisation: 0.83). The difference between the Sackin indices of the trees is not significantly different than expected under the YHK model. The difference between the Sackin indices of the trees is not significantly different than expected under the PDA model. [Figure 2](#) shows the expected distribution of Sackin index differences under the YHK and PDA models.

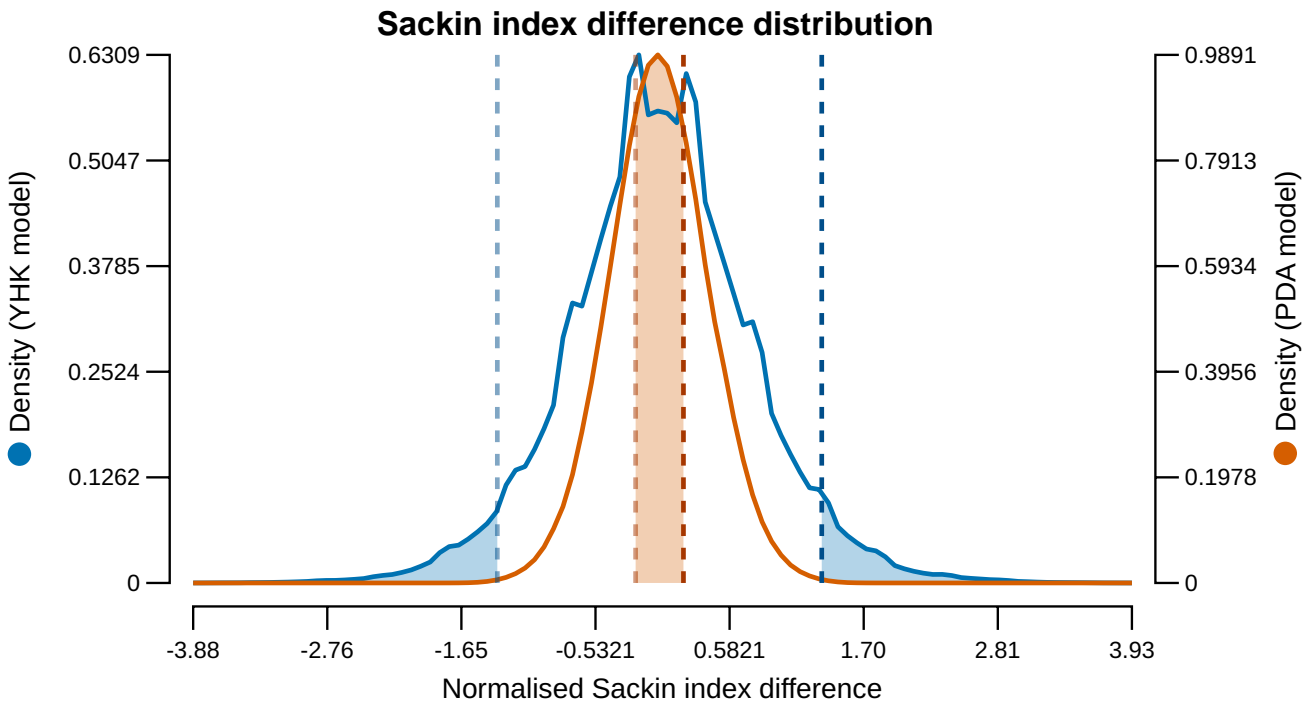


Figure 2. Distribution of Sackin index differences. The curve shows the expected distribution of (normalised) Sackin index differences under the null hypothesis for two trees containing 40 tips, computed using a Monte Carlo approach (2000 random trees were sampled according to the YHK or PDA model, and all possible pairings between them were considered). The dashed line represents the observed value of the Sackin index difference, normalised according to the YHK or PDA model. The blue plot refers to the YHK model, while the orange plot to the PDA model.

The **Colless indices** of the trees are **173** (YHK model normalisation: 0.9454, PDA model normalisation: 0.5545) and **91** (YHK model normalisation: -0.8372, PDA model normalisation: 0.2917). The difference between the Colless indices of the trees is not significantly different than expected under the YHK model. The difference between the Colless indices of the trees is not significantly different than expected under the PDA model. [Figure 3](#) shows the expected distribution of Colless index differences under the YHK and PDA models.

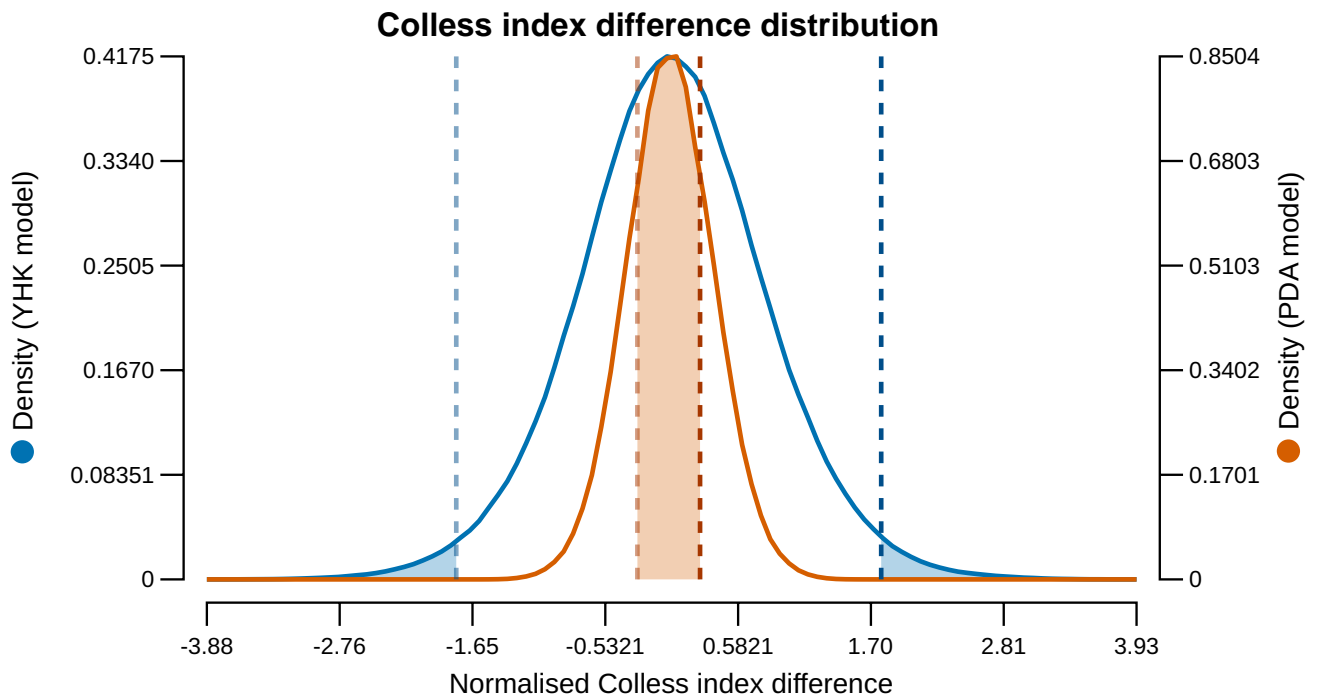


Figure 3. Distribution of Colless index differences. The curve shows the expected distribution of (normalised) Colless index differences under the null hypothesis for two trees containing 40 tips, computed using a Monte Carlo approach (2000 random trees were sampled according to the YHK or PDA model, and all possible pairings between them were considered). The dashed line represents the observed value of the Colless index difference, normalised according to the YHK or PDA model. The blue plot refers to the YHK model, while the orange plot refers to the PDA model.

The first tree has **10** cherries (YHK model normalisation: -3.73, PDA model normalisation: -1.14), while the second tree has **12** cherries (YHK model normalisation: -2.33, PDA model normalisation: 0.06215). The difference between the numbers of cherries does not differ significantly from the value expected under the YHK model. The difference between the numbers of cherries does not differ significantly from the number expected under the PDA model. [Figure 4](#) shows the expected distribution of the difference between the numbers of cherries under the YHK and PDA models.

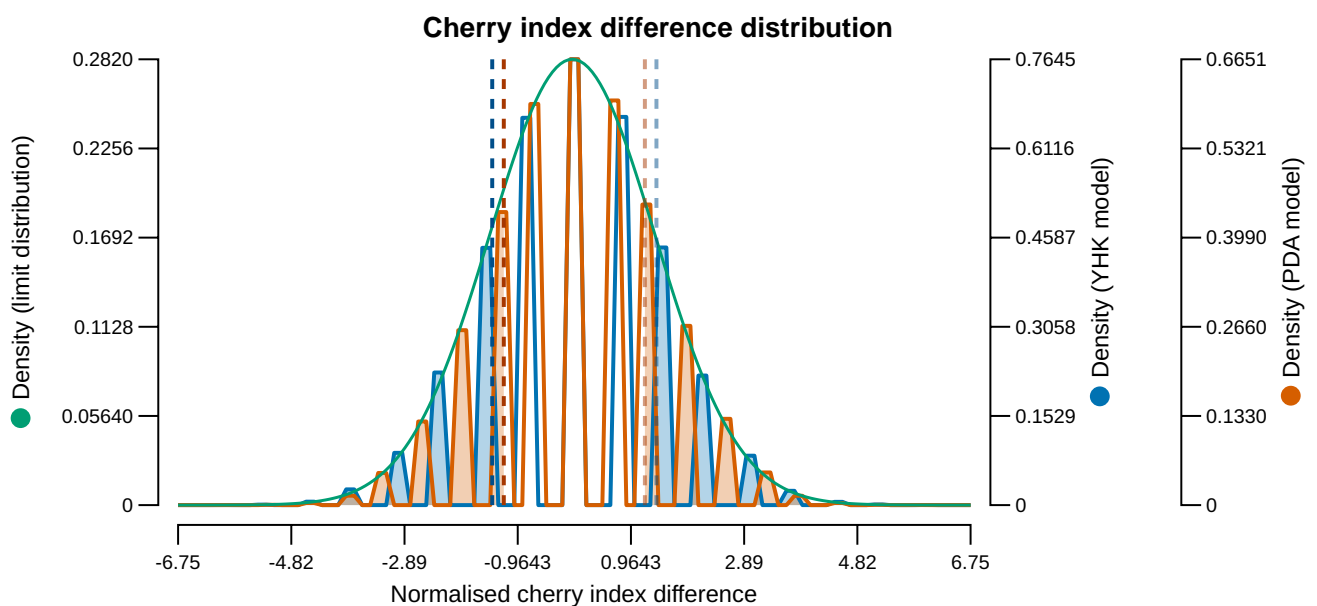


Figure 4. Distribution of the difference between the numbers of cherries. The curve shows the expected distribution of the (normalised) difference between the numbers of cherries under the null hypothesis for two trees containing 40 tips, computed using a Monte Carlo approach (2000 random trees were sampled according to the YHK or PDA model, and all possible pairings between them were considered). The dashed line represents

the observed value of the difference between the numbers of cherries, normalised according to the YHK or PDA model. The blue plot refers to the YHK model, while the orange plot to the PDA model. The green curve represents the limit distribution for the (normalised) difference between the numbers of cherries as the number of tips grows to infinity.