

1 **THE IMPROBABLE JOURNEYS OF EPIPHYTIC PLANTS ACROSS THE ANDES:**
2 **HISTORICAL BIOGEOGRAPHY OF *CYCNOCHES* (CATASETINAE,**
3 **ORCHIDACEAE)**

4

5

6 Oscar Alejandro Pérez-Escobar*, Marc Gottschling, Guillaume Chomicki, Fabien L.

7 Condamine, Bente Klitgård, Emerson Pansarin and Günter Gerlach

8

9 **Appendix S1**

10 **Including:** Extended material and methods, Tables S1-S5, Figures S1-S6.

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27 **Extended materials and methods**

28 *Phylogenetic incongruence analysis*

29 PACo builds upon co-phylogenetic approaches to assess the similarities between two
30 trees by comparison of Euclidean distance matrices derived from phylogenies¹. Additionally,
31 it determines the contribution of every association (i.e., pair of terminals of the analysed
32 phylogenies) to the co-phylogenetic pattern observed thus, it has the potential to statistically
33 identify specific potential conflicting terminals. PACo requires as input data a set of Posterior
34 Probability or Likelihood Bootstrap phylograms derived from every genomic compartment
35 (i.e., ‘nuc’ and ‘cp’) and a binary association matrix, in which same taxa present in both
36 datasets are linked. We executed PACo using 1000 Bayesian Posterior Probabilities obtained
37 from Mr.Bayes analysis computed from concatenated ‘nuc’ and ‘cp’ datasets, respectively.

38 Plastids are maternally inherited in orchids^{2,3} and hence, they reflect maternal
39 evolutionary relationships between species only. In contrast, nuclear loci show bi-parental
40 inheritance⁴ thus revealing maternal and paternal evolutionary histories. Our nuclear derived
41 phylogeny strongly reflects morphological relationships, whereas the plastid tree revealed an
42 apparent geographical pattern (see *Results* and *Discussion*). Following these principles,
43 conflicting ‘cp’ sequences were excluded from the dataset, and non-conflicting sequences of
44 each locus were re-aligned and concatenated.

45

46 *Orchidaceae phylogeny molecular clock dating*

47 We compiled 264 sequences from nuclear ITS locus, 297 sequences from *matK*, 276
48 from *rbcL*, and 217 from *trnL-trnF* (Table S4). Molecular clock dating relied on the
49 uncorrelated lognormal relaxed clock model implemented in BEAST v.2.0⁵. We used the
50 GTR + G nucleotide substitution model with 6 rate categories, a Yule tree prior, and a
51 Markov chain Monte Carlo (MCMC) chain length of 100 million generations, sampling every
52 10000 generations. Four fossil calibrations were used. The oldest known fossil of the

53 Asparagales (105 Mya⁶) was used as a maximal calibration point for the Orchidaceae
54 following⁷ and was given a normal prior (offset = 105.3; SD = 8). A gamma distribution was
55 used on each of the three ingroup fossil constraints (the offset value was set to the minimum
56 age of each fossil and the standard deviation was set such that the maximum age was included
57 in the 97.5% quantile). The fossil orchid †*Meliochis caribea* (15–20 Mya) has been assigned
58 to the Goodyerinae⁸, and provided a minimum age for the subtribe (offset = 15; SD= 5). Two
59 recently discovered orchid macrofossils (†*Dendrobium winkaphyllum* and †*Earina*
60 *fouldenensis*, both 20–23 Mya⁹, were used as minimum constraints within Epidendroideae,
61 with gamma priors (offset = 20; SD = 4.5). The trees were summarized using
62 TREEANNOTATOR v.1.8.0 (included in the BEAST package), with a 10% burn-in and a
63 0.98 posterior probability limit.

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82 **References**

- 83 1. Balbuena, J. A., Míguez-Lozano, R. & Blasco-Costa, I. PACo: a novel procrustes
84 application to cophylogenetic analysis. *PLoS One* **8**, e61048 (2013).
- 85 2. Chang, S.-B., Chen, W.-H., Chen, H.-H., Fu, Y.-M. & Lin, Y.-S. RFLP and inheritance
86 patterns of chloroplast DNA in intergenic hybrids of *Phalaenopsis* and *Doritis*. *Bot.*
87 *Stud.* **41**, 219–223 (2000).
- 88 3. Cafasso, D., Widmer, A. & Cozzolino, S. Chloroplast DNA inheritance in the orchid
89 *Anacamptis palustris* using single-seed polymerase chain reaction. *J. Hered.* **96**, 66–70
90 (2005).
- 91 4. Petit, R. J. *et al.* Comparative organization of chloroplast , mitochondrial and nuclear
92 diversity in plant populations. *Mol. Ecol.* **14**, 689–701 (2005).
- 93 5. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis.
94 *PLoS Comput. Biol.* **10**, 1–6 (2014).
- 95 6. Walker, J. W. & Walker, A. G. Ultrastructure of lower Cretaceous angiosperm pollen
96 and early evolution of flowering plants. *Ann. Missouri Bot. Gard.* **71**, 464–521 (1984).
- 97 7. Gustafsson, A. L. S., Verola, C. F. & Antonelli, A. Reassessing the temporal evolution
98 of orchids with new fossils and a Bayesian relaxed clock , with implications for the
99 diversification of the rare South American genus *Hoffmannseggella* (Orchidaceae :
100 Epidendroideae). *BMC Evol. Biol.* **10**, 177–190 (2010).
- 101 8. Ramírez, S. R., Gravendeel, B., Singer, R. B., Marshall, C. R. & Pierce, N. E. Dating
102 the origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature* **448**,
103 1042–1045 (2007).
- 104 9. Conran, J. G., Bannister, J. M. & Lee, D. E. Earliest orchid macrofossils: Early
105 Miocene *Dendrobium* and *Earina* (Orchidaceae: Epidendroideae) from New Zealand.
106 *Am. J. Bot.* **96**, 466–474 (2009).
- 107 10. Baldwin, B. G. Phylogenetic utility of the internal transcribed spacers of nuclear

108 ribosomal DNA in plants: An example from the compositae. *Mol. Phylogenet. Evol.* **1**,
109 3–16 (1992).

110

111 **Supplementary tables**

112 **Table S1.** Species names and voucher information, including herbarium specimens of
113 voucher deposition, for material used in this study to build *Cycnoches* molecular dataset. Taxa
114 sequenced for this study are shown in boldface.

115 **Table S2.** Primers and PCR settings used for amplifying chloroplast and nuclear DNA loci.

116 **Table S3.** Evolutionary models used for phylogenetic reconstructions, as selected by the
117 software jModelTest.

118 **Table S4.** Species names and voucher information, including herbarium specimens of
119 voucher deposition, for material used in this study to build Orchidaceae molecular dataset.

120 **Table S5.** Alignment characterization.

121

122 **Supplementary figures**

123 **Figure S1.** Conflicting nuclear-plastid associations (i.e. pair of nuclear / plastid taxa) obtained
124 by PACo analysis, using posterior probability trees. Taxa with normalized squared residual
125 values above the cut-off threshold (red line; see species with red bars) indicate potential
126 conflicting associations.

127 **Figure S2.** Phylogenetic relationships of *Cycnoches* independently derived 59 from nuclear
128 (ITS, ETS, *Xdh*) and plastid loci (*trnS-G*, *ycf1*) datasets. Potential outlier, congruent and
129 excluded terminals (i.e. taxa with sequences included on the plastid loci but missing on the
130 nuclear dataset prior analysis) are shown in red, green and blue, respectively.

131 **Figure S3.** Chronogram for Orchidaceae obtained under a relaxed clock model, applied to a
132 non-conflicting, concatenated nuclear ITS *Xdh* and plastid *matK*, *rbcL* loci. Mean absolute

133 ages and 95% Confidence Intervals are shown. The MRCA of Catasetinae is labelled with a
134 red star.

135 **Figure S4.** Distribution map of *Cycnoches* and number of species per biogeographical area.
136 (Inset) Number of species reported for every biogeographic region. Geopolitical boundaries
137 map generated by ArcMAP software (<http://www.esri.com>). Political divisions and elevation
138 data from DIVA-GIS (<http://www.diva-gis.org/gdata>).

139

140 **Figure S5.** Chronogram for *Cycnoches* obtained under a relaxed clock model, applied to a
141 non-conflicting, concatenated nuclear (ITS, ETS, *Xdh*) and plastid (*trnS-trnG*, *ycf1*) loci.
142 Mean absolute ages and 95% Confidence Intervals are shown.

143 **Figure S6.** Probabilities of ancestral areas estimated from a chronogram of *Cycnoches* and the
144 BayArea-like model including the founding speciation process. (Inset) Coded areas used for
145 biogeographical analysis. Geopolitical boundaries map generated by ArcMAP software
146 (<http://www.esri.com>). Political divisions and elevation data from DIVA-GIS
147 (<http://www.diva-gis.org/gdata>).

148

149

150

151

152

153

154

155

156

157

158