

1 Supplementary material

2

3 How to become a yucca moth: Minimal trait evolution needed to establish the
4 obligate pollination mutualism

5

6 *Accepted for the Biological Journal of the Linnean Society, 20 March 2010*

7

8 Jeremy B. Yoder,^{1,3} Christopher Irwin Smith,^{2,4} and Olle Pellmyr^{1,5}

9

10 ¹Dept. of Biological Sciences
11 University of Idaho
12 Moscow, ID 83844
13 USA

14

15 ²Dept. of Biology
16 Willamette University
17 Salem, OR 97301
18 USA

19

20 ³jbyoder@gmail.com

21 ⁴csmith@willamette.edu

22 ⁵pellmyr@uidaho.edu – *to whom correspondence should be addressed*

23

24

25 1073 words, 2 tables

26

26 **DNA sequence acquisition**

27 Prior to DNA extraction, we removed the head, wings, and abdomen from adult specimens and
28 retained them as vouchers. We extracted total genomic DNA from the remaining thorax using either the
29 DNEasy Blood and Tissue Kit (QIAGEN, Inc., Valencia, CA, USA) or the IsoQuick DNA Isolation Kit
30 (Orca Research, Inc., Bothell, WA, USA). For larvae, we extracted DNA from either the whole individual
31 or a ~25 mg portion of tissue, whichever was smaller.

32 Novel mitochondrial sequence data were collected for the deeply diverged species of *Lampronia*
33 given in Table S1. To assess variance in phylogenetic histories between independently assorting loci, we
34 compared data from the 2100-bp mitochondrial region with a 400-bp region of the nuclear locus coding
35 for arginine kinase (Arg-K) in a subset of the Prodoxidae (n = 11) including representatives from each
36 major clade of the mitochondrial phylogeny (Table S1). The Prodoxidae are deeply diverged from
37 members of the Lepidoptera for which more extensive genetic resources are developed, and the Arg-K
38 locus is, to date, the only nuclear locus to have been successfully amplified in every major Prodoxid clade.
39 Even this locus could only be completely amplified and sequenced in the limited sample of taxa we
40 present here.

41 **PCR and sequencing protocols**

42 For *COI-II*: We amplified the target region by PCR with four separate, partially overlapping
43 primer pairs. The amplified regions span positions 1461 to 3771 of the *Drosophila yakuba* mtDNA genome
44 (Clary & Wolstenholme, 1985). The primer pairs used include 1461F-2302R, 2331F-3020R, 2638F-3306R,
45 and 3252F-3771R (Clary & Wolstenholme, 1985; Simon, Frati, Beckenbach, Crespi, Liu & Flook, 1994). In
46 addition, we developed taxon-specific optimized primers for amplification of some regions in the
47 following species: *Lampronia fuscata*, *L. rupella*, and *L. oregonella* (Table S2). PCR was conducted in
48 30 μ L reaction volumes containing 50mM KCl, 10mM Tris buffer (ph = 9.0), 2.5mM MgCl₂, 0.2mM
49 dNTPs, 0.25 mM of each primer, one unit of Taq polymerase, and ~10ng genomic DNA. Reactions were

50 conducted with an optimized thermal cycler program of one cycle at 95°C for 5 min followed by 40 cycles
51 at 95°C for 1 min, 48°C for 1 min, 72°C for 90 s, and a final cycle at 72°C for 10 min.

52 *For Arg-K:* We amplified the target region by PCR with a single primer pair, which we developed
53 by identifying conserved regions of arginine kinase sequences from the basal lepidopteran genus
54 *Epicephala* and the *Bombyx mori* genome. Primer sequences are given in Table S3. PCR was conducted as
55 for COI-II, but using an optimized thermal cycler program of one cycle at 95°C for 5 min followed by 40
56 cycles at 95°C for 1 min, 55°C for 1 min, 72°C for 90 s, and a final cycle at 72°C for 10 min.

57 For both loci, PCR products were purified using QIAGEN PCR purification columns. Thermal
58 cycle sequencing was completed using the same primers as for the original PCR, with BigDye v3.1
59 (Applied Biosystems, Foster City, CA, USA). Sequencing products were cleaned using Centri-sep
60 Sephadex columns (Princeton Separations, Adelphia, NJ, USA), and both forward and reverse strands
61 were sequenced on an ABI 3730 automated DNA sequencer (Applied Biosystems).

62 **Phylogenetic reconstruction using MRBAYES**

63 Bayesian phylogenetic reconstruction was performed in MRBAYES version 3.1.2 (Huelsenbeck
64 and Ronquist 2001; Ronquist and Huelsenbeck 2003) implemented in parallel form (Altekar et al. 2004) on
65 a supercomputing cluster maintained by the Initiative for Bioinformatics and Evolutionary Studies
66 (IBEST) at the University of Idaho. The analysis was conducted under a GTR+I+ Γ model (i.e., allowing
67 free estimation of all parameters), using 32 Markov chains with heating parameter equal to 0.025 and 2
68 swaps per cycle. We set the Metropolis-coupled Markov chain Monte Carlo (MC³) analysis to terminate
69 when the standard deviation of split frequencies dropped below 0.01 (after 1.208×10^6 generations); we
70 selected a burn-in of 2.5×10^5 generations by examining stationarity of the log-likelihood scores. Finally,
71 we used PAUP* (Swofford 2002) to test for clocklike evolution on the MRBAYES consensus topology
72 (using all groupings compatible with the post-burn-in treespace sample) with branch lengths determined
73 by the GTR+I+ Γ model parameters selected by ModelTest (which were qualitatively similar to parameter

74 values estimated by MRBAYES); enforcement of a molecular clock under these conditions was strongly
75 rejected (difference in log-likelihood scores = 127.171; $p \ll 0.001$).

76 **Ancestral state reconstruction with BayesTraits**

77 For all trait reconstructions, we implemented the BayesTraits analysis with transition rate priors
78 drawn from an exponential distribution with means drawn in turn from a uniformly-distributed
79 hyperprior (Pagel et al., 2004). We performed each reconstruction in two independent replicate runs, and
80 assessed convergence by calculating the correlation between post-burn-in parameter value estimates
81 from the two runs; the independent runs were assumed to have converged when the correlation
82 coefficient exceeded 0.95. We calculated the posterior probability and the 95% confidence interval about
83 that value for each possible ancestral state in the MRCA of each group of interest by averaging across post-
84 burn-in estimates of the probability for that state.

85 To estimate host association for nodes of interest, we assigned extant taxa to one of eight known
86 host families (Table 1). We restricted the hyperprior distribution to between 0 and 10, and ran the
87 Markov chain for 10^8 generations, following 5×10^5 generations of burn-in, with samples taken every 10^5
88 generations. The correlation between post-burn-in parameter estimates from the two replicate runs was
89 0.9997. To reconstruct larval feeding habit, we assigned extant taxa to one of five known larval feeding
90 habits (Table 1). We restricted the hyperprior to between 0 and 15, and ran the Markov chain for 5×10^7
91 generations following 5×10^5 generations of burn-in, with samples taken every 10^5 generations. The
92 correlation between parameter estimates from the two replicate runs was 0.9995.

93

93 TABLE S1. Specimens used in phylogenetic reconstruction

Species	Host family	GenBank ID ¹	Collection locality
Basal prodoxids			
<i>Greya enchrysa</i> ²	Saxifragaceae	EU884123	USA: ID. Shoshone Co.
<i>G. mitellae</i> ²	Saxifragaceae	EU884120	USA: ID. Benewah Co.
<i>G. politella</i>	Saxifragaceae	U49021; GU393360	USA: CO. San Juan Co.
<i>G. solenobiella</i>	Apiaceae	AF150910	USA: CA. Monterey Co.
<i>G. variabilis</i>	unknown	AF150909	USA: WA. Clallam Co.
<i>Lampronia aenescens</i>	Rosaceae	AF150912	USA: WA. Garfield Co.
<i>L. capitella</i> ²	Grossulariaceae	EU884122; GU393369	Sweden. Nb. Sikfors
<i>L. corticella</i> ²	Rosaceae	EU884119	Sweden: Vb. Umeå
<i>L. fuscatella</i> ²	Betulaceae	EU884121; GU393368	UK: Surrey: Chobham Common
<i>L. oregonella</i> ²	Saxifragaceae	EU884118; GU393367	USA: ID: Idaho Co.
<i>L. rupella</i> ²	Geraniaceae	EU884124	Sweden: Hä. Ljungdalen
<i>L. sublustris</i> ²	Rosaceae	EU884125; GU393370	USA: ID. Latah Co.
<i>Mesepiola specca</i> ²	Nolinaceae	EU884116; GU393366	USA: AZ. Mohave Co.
<i>M. specca</i> ²	Nolinaceae	EU884117	USA: San Bernardino Co.
<i>Tetragma gei</i>	Rosaceae	AF150913	USA: ID. McCroskey Park
Bogus yucca moths			
<i>Prodoxus aenescens</i>	Agavaceae	AY737259	USA: CA. San Diego Co.
<i>P. atascosanellus</i>	Agavaceae	AY737270	USA: TX. Cameron Co.
<i>P. californicus</i>	Agavaceae	AY737271	USA: CA. San Diego Co.
<i>P. carnerosanellus</i>	Agavaceae	AY737273	USA: TX. Brewster Co.
<i>P. cinereus</i>	Agavaceae	AY737258	USA: CA: San Diego Co.
<i>P. coloradensis</i>	Agavaceae	AF150917; GU393362	USA: AZ. Mohave Co.

Species	Host family	GenBank ID¹	Collection locality
<i>P. decipiens</i>	Agavaceae	AY737268	USA: TN. Wilson Co.
<i>P. gypsicolor</i>	Agavaceae	AF150920	USA: CA. San Bernardino Co.
<i>P. intricatus</i>	Agavaceae	AY737274	MEX: Veracruz.
<i>P. mapimiensis</i>	Agavaceae	AY737266	USA: TX. Brewster Co.
<i>P. marginatus</i>	Agavaceae	AY737260	USA: CA. San Diego Co.
<i>P. ochrocarus</i>	Agavaceae	AY737263	USA: AZ. Cochise Co.
<i>P. pallida</i>	Agavaceae	AF150919	USA: CA. Riverside Co.
<i>P. phylloryctus</i>	Agavaceae	AY737272	USA: CO. Dolores Co.
<i>P. quinquepunctellus</i>	Agavaceae	AY737267	USA: AZ. Coconino Co.
<i>P. sordidus</i>	Agavaceae	AY737264	USA: CA. San Bernardino Co.
<i>P. sonorensis</i>	Agavaceae	AY737262	USA: AZ. Santa Cruz Co.
<i>P. tamaulipellus</i>	Agavaceae	AY737261	USA: AZ. Cameron Co.
<i>P. tehuacanensis</i>	Agavaceae	AY737269	MEX: Puebla
<i>P. weethumpi</i>	Agavaceae	AY737265; GU393361	USA: CA. San Bernardino Co.
<i>P. y-inversus</i>	Agavaceae	AF150918	USA: NV. Clark Co.
Pollinators			
<i>Parategeticula ecdysiastica</i>	Agavaceae	DQ924360	MEX: Baja California Sur
<i>P. elephantipella</i>	Agavaceae	AF150922	MEX: Chiapas
<i>P. martella</i>	Agavaceae	AF150924	MEX: Coahuila
<i>P. pollenifera</i>	Agavaceae	AF150921	USA: AZ. Santa Cruz Co.
<i>P. tzoyatllella</i>	Agavaceae	AF150923	MEX: Coahuila
<i>Tegeticula altiplanella</i>	Agavaceae	DQ924340	USA: AZ. Apache Co.
<i>T. antithetica</i>	Agavaceae	U49025	USA: NV. Grant Co.
<i>T. baccatella</i>	Agavaceae	DQ924342	USA: AZ. Pima Co.
<i>T. baja</i>	Agavaceae	DQ924334	MEX: Baja California Sur
<i>T. californica</i>	Agavaceae	DQ075470	USA: CA. San Diego Co.
<i>T. cassandra</i>	Agavaceae	DQ075512	USA: FL. Ocala Co
<i>T. carnerosanella</i>	Agavaceae	DQ924331	MEX: San Luis Potosi
<i>T. elatella</i>	Agavaceae	DQ924349	USA: TX. Brewster Co.
<i>T. maculata baja</i>	Agavaceae	AY004305	MEX: Baja California
<i>T. maculata extranea</i>	Agavaceae	U49023	USA: CA. Riverside Co.

Species	Host family	GenBank ID¹	Collection locality
<i>T. maculata maculata</i>	Agavaceae	U49024; GU393363	USA: CA. Tulare Co.
<i>T. maderae</i>	Agavaceae	DQ924337	MEX: Sonora
<i>T. mexicana</i>	Agavaceae	DQ924329	MEX: San Luis Potosi
<i>T. mojavella</i>	Agavaceae	DQ924338	USA: CA. San Bernardino Co.
<i>T. rostratella</i>	Agavaceae	DQ075478	USA: TX. Brewster Co.
<i>T. superficiella</i>	Agavaceae	DQ075500	USA: UT: Washington Co.
<i>T. synthetica</i>	Agavaceae	AY327144; GU393365	USA: CA. Los Angeles Co.
<i>T. tambasi</i>	Agavaceae	DQ924347	MEX: Michoacan
<i>T. tehuacana</i>	Agavaceae	DQ924324	MEX: Puebla
<i>T. yuccasella</i>	Agavaceae	DQ924356; GU393364	USA: TX. Sonora Co.
Cheaters			
<i>T. corruptrix C</i>	Agavaceae	DQ075516	USA: WY. Crook Co.
<i>T. corruptrix F</i>	Agavaceae	DQ924359	USA: TX. Brewster Co.
<i>T. intermedia</i>	Agavaceae	DQ924350	USA: NM: Valencia Co.
Outgroup			
<i>Adela septentrionella</i> ²	Rosaceae	EU884115	USA: ID. Benewah Co.

94 ¹ GenBank accession for COI-II given first, then Arg-K if collected

95 ² Sequence data newly acquired for this study

96

97

97 **TABLE S2. Optimized primers used to sequence portions of the 1461-2191 mtDNA region for**
 98 ***Lampronia fuscata*, *L. rupella*, and *L. oregonella*.**
 99

Primer name (mtDNA region)	Sequence
fus-COI-F (5' end of COI)	AATTCGTGCT GAATTAGGAA CTCCTAGTTC CTTAATTGGT
fus-COI-R	AAAGTATTGT AATAGCTCCA GCTAATACTG GTAATGAAAG
ore-midCOI-F (COI)	CTTATCTTCT AACATTTTCAC ATTCAGGAAG ATCTGTTGAT TTAA
ore-midCOI-R	ACTTCATAAT ATTGATGGAG TATAGTTAAT TTTAGTTCCA TGTAAG
rup-tRNAL-F (tRNA-leucine)	ATTTTTCCCA CAACATTTTT TAGGATTAAG AGGA
rup-tRNAL-R	ATTGATGTCC GATAGTTTTT AATGT
lamp-COII-F (COII)	AACATATACT CTAATATTTA TTGCTCTACC ATCATTACGA
lamp-COII-R	TACATCTGTT GCTGTGATTA AGATTCTGAAT TTGGTTGTTA

100

101 **TABLE S3. Primers used to sequence portions of the Arginine Kinase locus.**

Primer name	Sequence
ArgK-Int0-F	TTCCTSTTCA AGGARGGYGA CCGC
ArgK-R	ACGCCGCCCT CGGCCTCMGT GTGC

102

103

103 Table S4. Posterior probabilities of possible ancestral host associations for select nodes from the
 104 phylogeny of the Prodoxidae (Node numbers follow Figure 2).
 105

Posterior probability of possible ancestral host association \pm 95% CI ¹								
Node	Apiaceae	Saxifragaceae	Rosaceae	Betulaceae	Ruscaceae	Agavaceae	Grossulariaceae	Geraniaceae
1	0.07935 \pm 0.0023	0.2828 \pm 0.0059	0.1347 \pm 0.0039	0.0886 \pm 0.0025	0.1311 \pm 0.0035	0.0430 \pm 0.0019	0.1306 \pm 0.0036	0.1098 \pm 0.0030
2	0.0725 \pm 0.000	0.0661 \pm 0.0023	0.2730 \pm 0.0048	0.0738 \pm 0.0021	0.2051 \pm 0.0047	0.1521 \pm 0.0043	0.0776 \pm 0.0021	0.0799 \pm 0.0022
3	0.0635 \pm 0.0026	0.5328 \pm 0.0076	0.0102 \pm 0.0006	0.0825 \pm 0.0030	0.0344 \pm 0.0016	0.0020 \pm 0.0002	0.1622 \pm 0.0048	0.1123 \pm 0.0040
4	0.0331 \pm 0.0018	0.0157 \pm 0.0009	0.0104 \pm 0.0007	0.0319 \pm 0.0018	0.3542 \pm 0.0074	0.4872 \pm 0.0079	0.0320 \pm 0.0017	0.0354 \pm 0.0019
5	0.0124 \pm 0.0008	0.0063 \pm 0.0005	0.9299 \pm 0.0027	0.0127 \pm 0.0008	0.0125 \pm 0.0008	0.0009 \pm 0.0000	0.0123 \pm 0 .0008	0.0130 \pm 0.0008
6	0.0656 \pm 0.0023	0.3008 \pm 0.0070	0.0181 \pm 0.0008	0.1406 \pm 0.0037	0.0480 \pm 0.0017	0.0049 \pm 0.0027	0.2689 \pm 0.0055	0.1532 \pm 0.0041
7	0.0945 \pm 0.0030	0.6848 \pm 0.0052	0.0207 \pm 0.0007	0.0434 \pm 0.0014	0.0405 \pm 0.0013	0.0093 \pm 0.0003	0.0506 \pm 0.0016	0.0562 \pm 0.0017
8	0.0011 \pm 0.0001	0.0003 \pm 0.0000	0.0002 \pm 0.0000	0.0011 \pm 0.0002	0.0015 \pm 0.0001	0.9937 \pm 0.0003	0.0011 \pm 0.0001	0.0010 \pm 0.0001
9	0.0019 \pm 0.0002	0.0007 \pm 0.0000	0.0004 \pm 0.0000	0.0016 \pm 0.0001	0.9919 \pm 0.0005	0.0002 \pm 0.0000	0.0016 \pm 0.0002	0.0018 \pm 0.0002
10	0.0023 \pm 0.0002	0.0007 \pm 0.0000	0.0005 \pm 0.0000	0.0023 \pm 0.0002	0.0032 \pm 0.0024	0.9867 \pm 0.0006	0.0022 \pm 0.0002	0.0022 \pm 0.0002
11	0.0021 \pm 0.0002	0.0006 \pm 0.0000	0.0004 \pm 0.0000	0.0021 \pm 0.0002	0.0029 \pm 0.0002	0.9878 \pm 0.0006	0.0021 \pm 0.0002	0.0020 \pm 0.0002

106 ¹Most probable state indicated by **bold** posterior probability value

107 **Table S5. Posterior probabilities of possible ancestral oviposition habits for select nodes from the**
 108 **phylogeny of the Prodoxidae (Node numbers follow Figure 2).**
 109

Posterior probability of ancestral larval feeding habit \pm 95% CI ¹					
Node	Ovary	Fruit	Leaf	Twigs	Floral stem
1	0.6519 \pm 0.0078	0.0711 \pm 0.0023	0.1005 \pm 0.0029	0.1180 \pm 0.0037	0.0586 \pm 0.0025
2	0.6824 \pm 0.0069	0.0660 \pm 0.0023	0.0974 \pm 0.0027	0.0984 \pm 0.0029	0.0557 \pm 0.0022
3	0.4316 \pm 0.0075	0.1211 \pm 0.0030	0.1452 \pm 0.0032	0.1956 \pm 0.0043	0.1064 \pm 0.0026
4	0.6711 \pm 0.0077	0.0760 \pm 0.0033	0.0930 \pm 0.0035	0.0824 \pm 0.0034	0.0775 \pm 0.0033
5	0.8062 \pm 0.0045	0.0354 \pm 0.0015	0.0686 \pm 0.0024	0.0700 \pm 0.0025	0.0197 \pm 0.0008
6	0.3840 \pm 0.0067	0.1160 \pm 0.0026	0.1565 \pm 0.0031	0.2452 \pm 0.0047	0.0983 \pm 0.0021
7	0.3248 \pm 0.0056	0.1816 \pm 0.0029	0.1594 \pm 0.0029	0.1443 \pm 0.0028	0.1887 \pm 0.0030
8	0.3886 \pm 0.0075	0.1514 \pm 0.0047	0.1245 \pm 0.0046	0.1015 \pm 0.0042	0.2341 \pm 0.0058
9	0.9680 \pm 0.0012	0.0046 \pm 0.0003	0.0130 \pm 0.0007	0.0130 \pm 0.0008	0.0013 \pm 0.0001
10	0.0020 \pm 0.0058	0.2476 \pm 0.0064	0.0494 \pm 0.0020	0.0404 \pm 0.0018	0.6605 \pm 0.0076
11	0.9415 \pm 0.0022	0.0084 \pm 0.0006	0.0234 \pm 0.0013	0.0240 \pm 0.0014	0.0026 \pm 0.0002

110 ¹Most probable state indicated by **bold** posterior probability value

111 **LITERATURE CITED**

112 Altekar, G., S. Dwarkadas, J. P. Huelsenbeck, and F. Ronquist. 2004. Parallel Metropolis-coupled Markov
113 chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20:407-415.

114 Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees.
115 *Bioinformatics* 17:754-755.

116 Ronquist, F., and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed
117 models. *Bioinformatics* 19:1572-1574.

118 Swofford, D. 2002. PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer
119 Associates, Sunderland, Mass.

120

121

122