iScience, Volume *25*

Supplemental information

A complex interplay of evolutionary forces continues to shape ancient co-occurring symbiont genomes

Yumary M. Vasquez and Gordon M. Bennett

This supplemental file includes

Supplementary Figures 1 to 5

Supplementary Table 2

Supplementary Table 3

References

Other supplemental material for this manuscript includes the following:

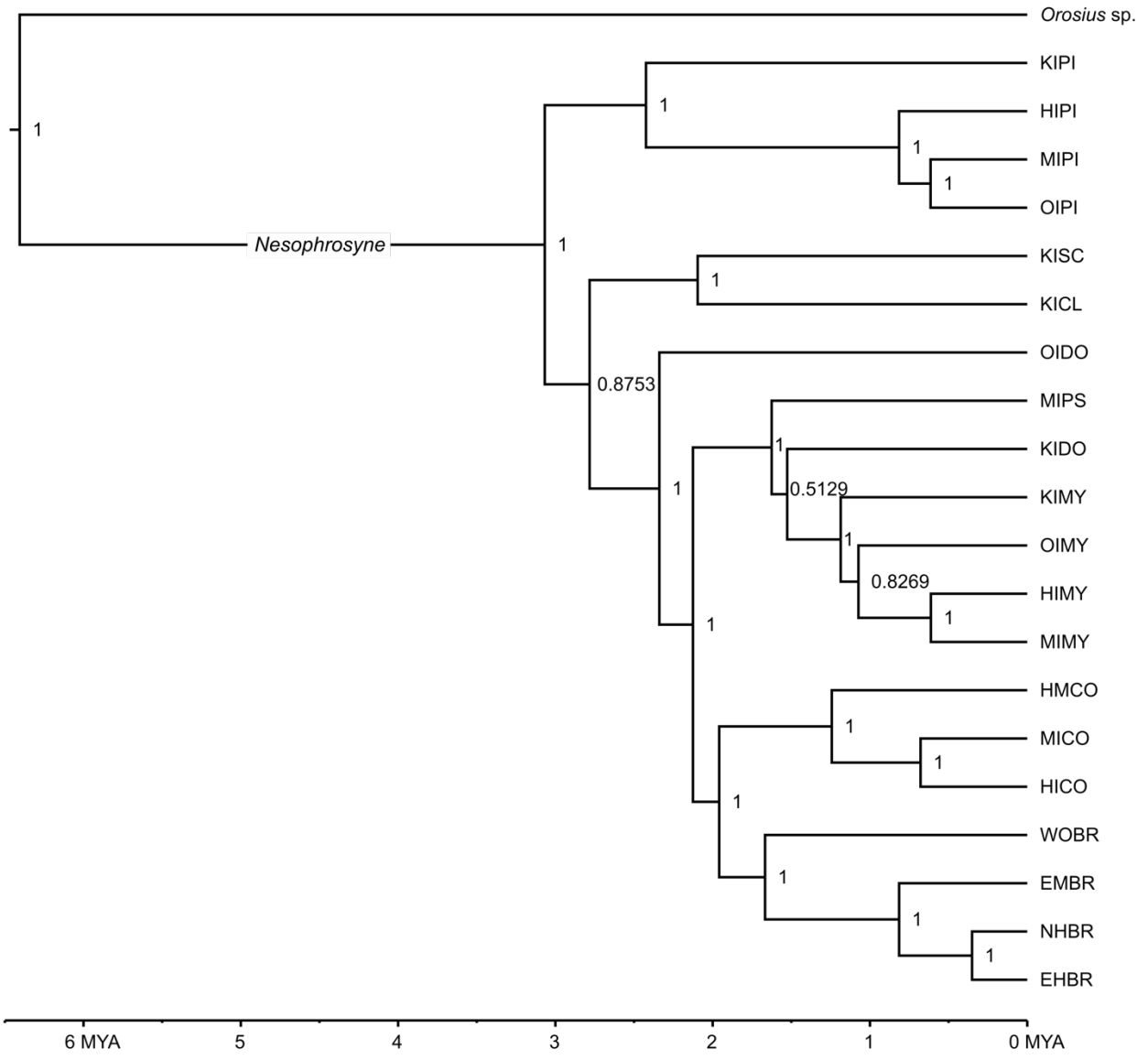
Supplementary Table 1 (provided in separate excel file)

Supplementary table

Supplementary Table 1. Statistics for fully assembled dual-obligate symbiont genomes, *Sulcia* and *Nasuia*, from endemic Hawaiian leafhopper species (Cicadellidae: *Nesophrosyne*). Provided as separate excel file. Related to Table 1.
Supplementary Table 2. Summary of global substitutions rates across all plant groups and within plant groups.
Supplementary Table 3. Nonsynonymous to synonymous (dN/dS) rates for genes not universally retained within *Nesophrosyne*.

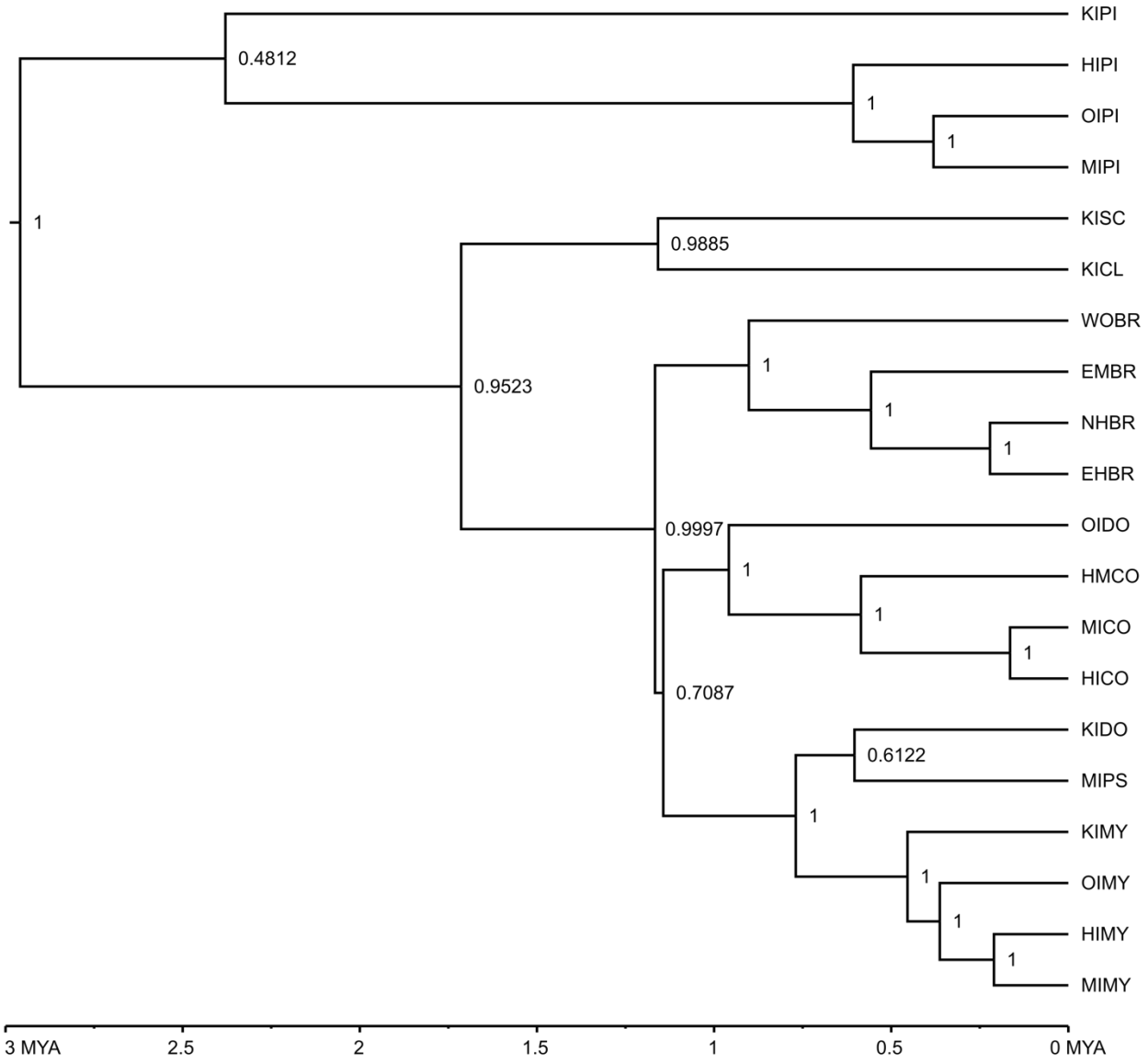
Supplementary figures

Supplementary Fig 1. Time-calibrated phylogeny of *Nesophrosyne* leafhopper host mitochondrial sequences.
Supplementary Fig 2. Time-calibrated phylogeny of *Sulcia* sequences.
Supplementary Fig 3. Time-calibrated phylogeny of *Nasuia* sequences.
Supplementary Fig 4. Ancestral state reconstruction of *Sulcia* genes convergently loss among *Nesophrosyne* species.
Supplementary Fig 5. Ancestral state reconstruction of *Nasuia* genes convergently loss among *Nesophrosyne* species.



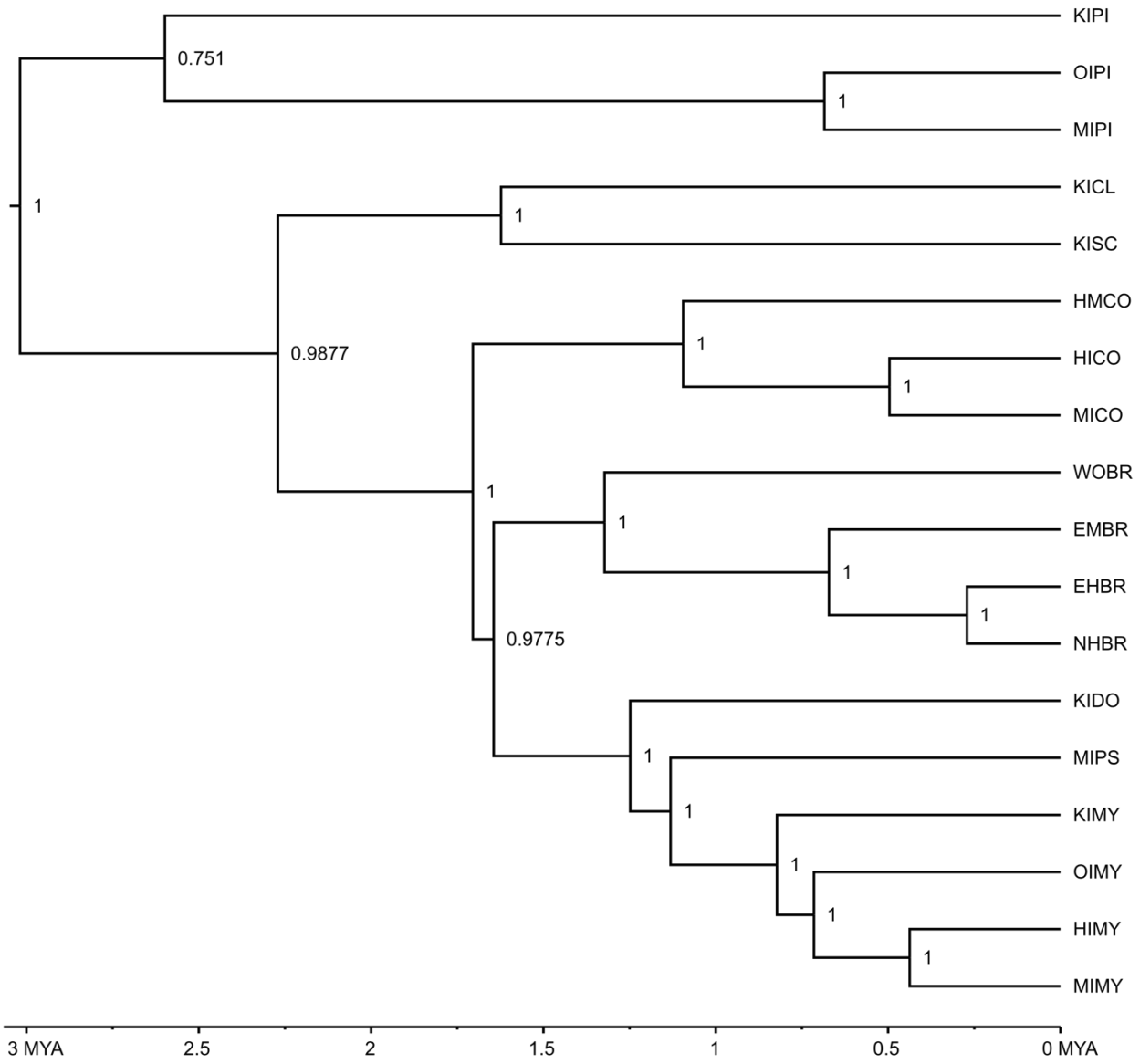
Supplementary Fig 1. Time-calibrated phylogeny of *Nesophrosyne* leafhopper host mitochondrial sequences, Related to Figure 1. Complete mitochondrial tree generated in BEAST v.1.10.4. The resulting phylogenetic data sets included concatenated protein coding and ribosomal genes for a total of 15 genes (14,304 sites) from leafhopper mitochondria. The tree prior included the yule process speciation with a random starting tree. Five internal node calibrations were selected following our previous phylogenetic study of the *Nesophrosyne* (Bennett and O'Grady, 2013). Node dates were determined from *Nesophrosyne* species divergences that match the sequential geological formation of the Hawaiian Islands (i.e., progression rule) and applied with a normal prior distribution since absolute species divergence could have occurred earlier or after island formation. Multiple chains were sampled every 1000 generations following Bayesian recommendations (two chains with four million generations; Huelsenbeck et al., 2002). Runs were performed with an uncorrelated relaxed clock with a

lognormal distribution. Posterior probabilities are provided. Scale bar indicates time calibration in millions of years. See Table 1 for species shorthand nomenclature. (MYA = million years ago)



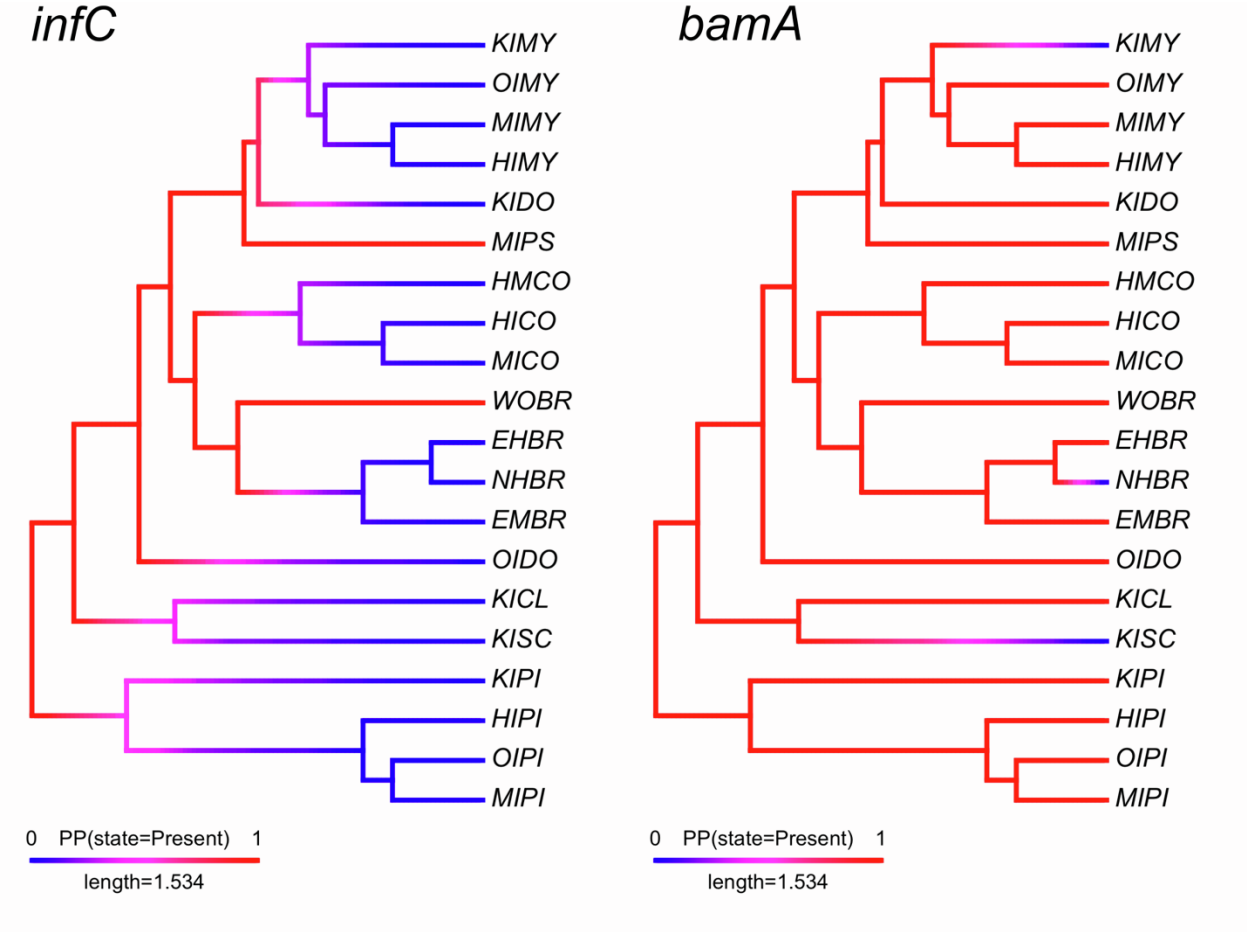
Supplementary Fig 2. Time-calibrated phylogeny of *Sulcia* sequences, Related to Figure 1. *Sulcia* tree generated in BEAST v.1.10.4. The resulting phylogenetic data sets included concatenated protein coding and ribosomal genes for a total of 184 genes (181,781 sites). The tree prior included the yule process speciation with a random starting tree. Five internal node calibrations were selected following our previous phylogenetic study of the *Nesophrosyne* (Bennett and O'Grady, 2013). Node dates were determined from *Nesophrosyne* species divergences that match the sequential geological formation of the Hawaiian Islands (i.e., progression rule) and applied with a normal prior distribution since absolute species divergence could have occurred earlier or after island formation. Multiple chains were sampled every 1000 generations following Bayesian recommendations (two chains with four million generations; Huelsenbeck et al., 2002). Runs were performed with an uncorrelated relaxed clock with a

lognormal distribution. Posterior probabilities are provided. Scale bar indicates time calibration in millions of years. See Table 1 for species shorthand nomenclature. (MYA = million years ago)

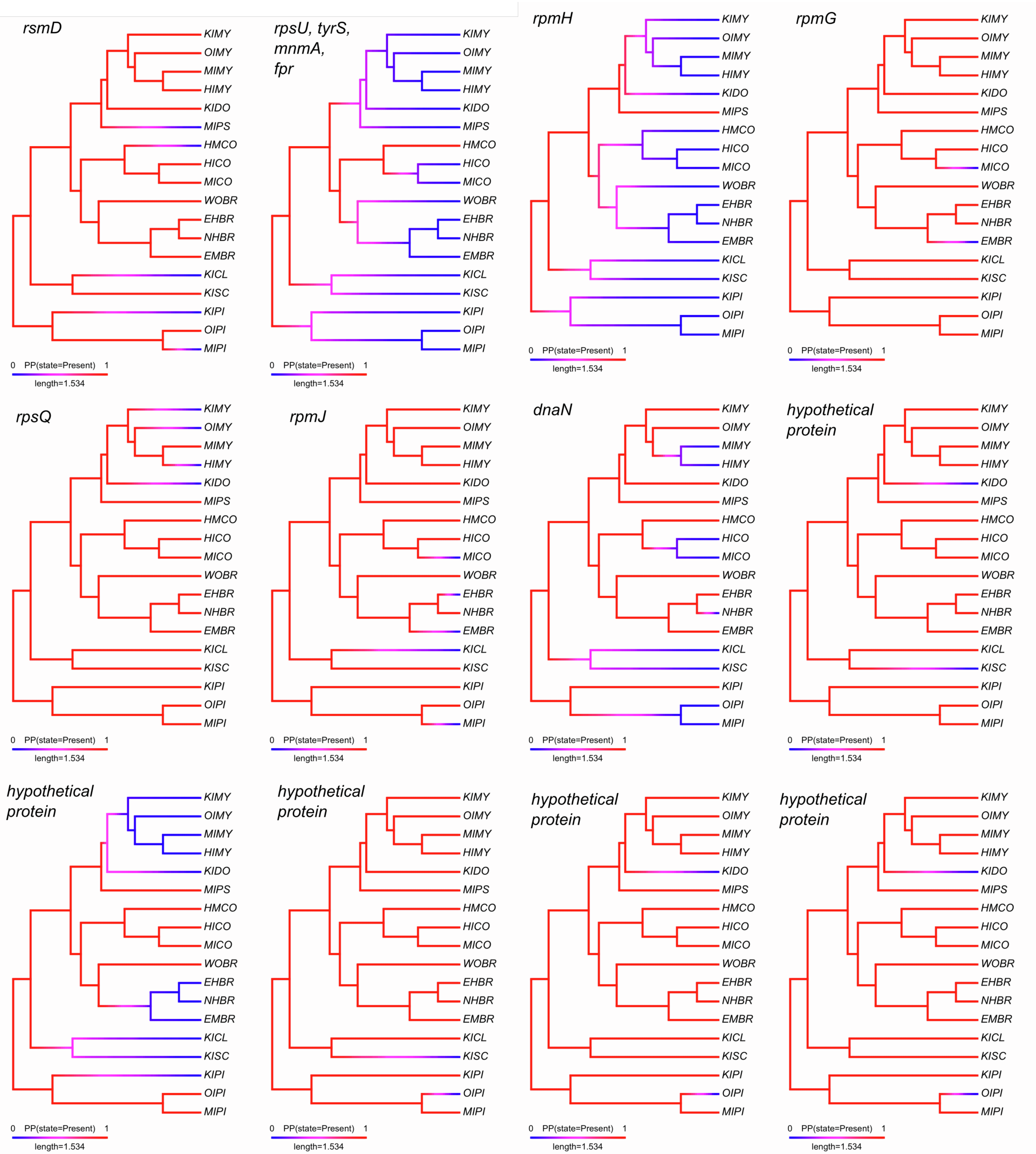


Supplementary Fig 3. Time-calibrated phylogeny of *Nasuia* sequences, Related to Figure 1. *Nasuia* tree generated in BEAST v.1.10.4. The resulting phylogenetic data sets included concatenated protein coding and ribosomal genes for a total of 99 genes (86,095 sites). The tree prior included the yule process speciation with a random starting tree. Five internal node calibrations were selected following our previous phylogenetic study of the *Nesophrosyne* (Bennett and O'Grady, 2013). Node dates were determined from *Nesophrosyne* species divergences that match the sequential geological formation of the Hawaiian Islands (i.e., progression rule) and applied with a normal prior distribution since absolute species divergence could have occurred earlier or after island formation. Multiple chains were and sampled every 1000 generations following Bayesian recommendations (two chains with four million

generations; (Huelsenbeck et al., 2002). Runs were performed with an uncorrelated relaxed clock with a lognormal distribution. Posterior probabilities are provided. Scale bar indicates time calibration in millions of years. See Table 1 for species shorthand nomenclature.



Supplementary Fig 4. Ancestral state reconstruction of *Sulcia* genes convergently loss among *Nesophrosyne* species, Related to Figure 1. Ancestral state reconstruction from phytools v.1.0-1 package on genes convergently lost in *Sulcia* (Revell, 2012). Convergent loss of genes was mapped to the mitochondrial host phylogeny using a custom model that only allows for the loss of genes to occur (no gene gain) to account for the lack of environmental phase (and hence lack of gene uptake) for symbiotic bacteria. A stochastic character mapping was generated for each gene with the custom model and simulating 100 trees. Colors represent posterior density of stochastic maps, blue = absence and red = presence.



rsmD
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534
rpsU, tyrS, mnmA, fpr
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534
rpmH
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534
rpmG
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534
rpsQ
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534
rpmJ
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534
dnaN
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534
hypothetical protein
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534
hypothetical protein
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534
hypothetical protein
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534
hypothetical protein
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534
hypothetical protein
KIMY
OIMY
MIMY
HIMY
KIDO
MIPS
HMCO
HICO
MICO
WOBR
EHBR
NHBR
EMBR
KICL
KISC
KIPI
OIPI
MIPI
0 PP(state=Present) 1
length=1.534

Supplementary Fig 5. Ancestral state reconstruction of *Nasuia* genes convergently loss among *Nesophrosyne* species, Related to Figure 2 and Figure 3. Ancestral state reconstruction from phytools v.1.0-1 package on genes convergently lost in *Nasuia* (Revell, 2012). Convergent loss of genes was mapped to the mitochondrial host phylogeny using a custom model that only allows for the loss of genes to occur (no gene gain) to account for the lack of environmental phase (and hence lack of gene uptake) for symbiotic bacteria. A stochastic character mapping was generated for each gene with the custom model and simulating 100 trees. Colors represent posterior density of stochastic maps, blue = absence and red = presence. Note: the last two plots are two different hypothetical genes lost in both KIDO and OIPI species.

Averages of substitutions/site/year across all plant groups		
	Genome	Average substitutions/site/year
	Mitochondria	2.15E-07
	Nasuia	1.13E-07
	Sulcia	3.24E-09
Averages of substitutions/site/year across each plant group		
Host Plant	Genome	Average substitutions/site/year
Broussaisia	Mitochondria	2.38E-07
Coprosma	Mitochondria	1.80E-07
Myoporum	Mitochondria	2.35E-07
Pipturus	Mitochondria	2.06E-07
Broussaisia	*Nasuia*	9.72E-08
Coprosma	*Nasuia*	9.91E-08
Myoporum	*Nasuia*	1.23E-07
Pipturus	*Nasuia*	1.31E-07
Broussaisia	*Sulcia*	5.12E-09
Coprosma	*Sulcia*	1.67E-09
Myoporum	*Sulcia*	2.79E-09
Pipturus	*Sulcia*	3.16E-09

Supplementary Table 2. Summary of global substitutions rates across all plant groups and within plant groups, Related to Figure 4. Summary of global substitutions rates across all plant groups and within plant groups. Substitutions rates per gene were calculated by their

pairwise distances in MEGAX v.10.2.4 and divided by their time of divergence (Kumar et al., 2018). Rates are averaged across each genome, as well as by each associated host-plant.

Genome: *Nasuia*		
Gene	dN/dS	# of sequences
dnaN	0.18276	7
rpmG	0.03875	16
rpmJ	0.03449	13
rpsQ	0.03311	14
rsmD	0.02502	13
Genome: *Sulcia*		
Gene	dN/dS	# of sequences
bamA	0.22209	17
infC	0.0001	2

Supplementary Table 3. Nonsynonymous to synonymous (dN/dS) rates for genes not universally retained within *Nesophrosyne*, Related to Figures 1,2 and 4. Nonsynonymous to synonymous (dN/dS) rates for genes not universally retained within *Nesophrosyne*. Rates per gene were calculated using the M0 model from codeml (Yang, 2007).