Supplemental Materials

Appendix A. Plant growth and bacterial isolation procedures

Plants. We planted seeds from 190 unique *M. lupulina* individuals in a greenhouse to obtain leaf tissue for DNA extraction. To prepare seeds, we first scarified seeds by removing the black seed pod and cutting a small nick at the edge of the seed opposite the hilum (around 0.5 mm). We then sterilized seeds in 95% ethanol for 30 seconds and in bleach for an additional four minutes. We hydrated seeds by immersing scarified seeds in distilled water for 25-30 minutes. We placed soaked seeds on 0.08% agar plates using sterile technique and stored plates inverted and covered in foil at 4°C for 7 – 8 days (stratification). We stimulated radicle growth by placing seeds at room temperature for 15 hour and after radicle length reached 1 cm, we exposed seeds to light for 1.5 hours to induce chlorophyll production. Germinated seeds were grown in a greenhouse with a day temperature of 22°C for 16 hours and a night temperature of 18°C for 8 hours (Moreau *et al.* 2006). For the first two weeks of growth, we top watered plants every other day with a spray bottle. For the rest of the growing period we bottom watered plants every other day. We harvested two to three young leaves (totaling 0.02 to 0.07 g) from four to five week old plants. We performed DNA extractions on the harvested leaf tissue according to the instructions of the Qiagen DNeasy Plant Tissue Mini Protocol, eluting extractions twice in a total of 100 µl of elution buffer. DNA was sent to Cornell University for GBS; samples were digested with the restriction enzyme EcoT22I.

Bacteria. We harvested one nodule from each individual plant. To sterilize the nodules, we hydrated nodules in distilled water for one hour and soaked nodules in 95% ethanol for 20 seconds and then in bleach for an additional 20 seconds. We crushed nodules in sterile water

droplets using flame sterilized forceps and plated the crushed nodules onto nutrient agar plates (TY media). We incubated the plates for two to four days at 30°C until sufficient bacterial growth was achieved and colonies were visible. A single colony from each plate was re-streaked onto new nutrient agar plates four times to remove contamination and isolate a single strain. We grew the microbial cultures in sterile liquid TY media at 30°C and 200 rpm for approximately 48 hours or until cultures reached a cell density of 8×10^8 cells/ml (optical density (OD) = 1). We performed DNA extractions on liquid microbial cultures using the MoBio UltraClean Microbial DNA Isolation Kit, eluting DNA once in 50 µl of distilled water.

References

Moreau D, Groves E, Ruffel S, Lepetit M (2006) Growing *M. truncatula*: choice of substrates and growth conditions. *Medicago truncatula* handbook, **version November 2006**, 1-26.

Tables

Table 1. Locations of *M. lupulina* and *Ensifer* populations used in population genetic analysis.

Population ID	VCF ID	Province/State	Country	Latitude	Longitude
15	AIL	Ontario	Canada	43.143207	-81.491089
6	AV	Pennsylvania	USA	40.200904	-76.763663
17	BLU	Ontario	Canada	43.373892	-81.715794
24	BR	Ontario	Canada	43.74994	-79.639721
11	BUF	New York	USA	42.855817	-78.754292
31	CAL 1	Ontario	Canada	43.939625	-80.006304
27	CAL 2	Ontario	Canada	43.840718	-79.755077
33	COB	Ontario	Canada	44.103365	-78.156738
34	COO	Ontario	Canada	44.202389	-79.647639
1	DE	Delaware	USA	38.686381	-75.07443
16	DUN	Ontario	Canada	43.269472	-79.934721
14	FOR	Ontario	Canada	43.090579	-82.023067
29	GA	Ontario	Canada	43.872622	-80.322633
35	GIL	Ontario	Canada	44.219369	-79.56007
25	GOD	Ontario	Canada	43.759751	-81.702318
30	HA 2	Ontario	Canada	43.91227	-80.877399
2	HL	Ontario	Canada	38.920588	-75.448694
36	HO	Ontario	Canada	44.219923	-81.049833
38	KA	Ontario	Canada	44.555739	-78.838577
18	KIT	Ontario	Canada	43.446111	-80.499766
32	KSR	Ontario	Canada	43.958223	-79.506683
26	MAP	Ontario	Canada	43.836307	-80.636623
19	MTC	Ontario	Canada	43.460894	-81.204951
10	NEY	Ontario	Canada	42.371386	-77.353299
28	NH	Ontario	Canada	43.861701	-81.360455
23	NY	Ontario	Canada	43.713177	-79.462159
4	PA1	Pennsylvania	USA	39.930768	-75.584929
8	PAN	Pennsylvania	USA	40.830615	-76.84046
7	PAR	New York	USA	40.659773	-76.918652
5	PT	Pennsylvania	USA	40.07175	-75.435691
20	PUS	Ontario	Canada	43.466937	-80.183973
12	SAR	Ontario	Canada	42.960129	-82.368429
37	SC	Ontario	Canada	44.332505	-80.993443
39	SEG	Ontario	Canada	45.225519	-79.682214
9	SIN	Pennsylvania	USA	41.457507	-77.135632
22	TO	Ontario	Canada	43.661197	-79.401163
3	UBR	Delaware	USA	39.626937	-75.675749

21	WA	Ontario	Canada	43.596353	-80.625336
13	WAR	Ontario	Canada	43.0054	-81.856813

Table 2. NCBI gene reference numbers for nodulation and nitrogen fixation genes in *E. meliloti* and *E. medicae* used in nucleotide diversity analysis.

Gene name	***E. meiloti* gene number**	***E. medicae* gene number**
nodA	1235511	5320486
nodB	1235510	5320487
nodC	1235509	5320488
nifA	1235479	5320534
nifB	1235478	5320535
nifD	1235486	5320526
nifE	1235488	5320524
nifH	1235485	5320527
nifK	1235487	5320525
nifN	1235514	5320483
nifX	1235489	5320523
Type III effector 1279	-	5322126
Type III effector 4319	-	5319168
exoU glucosyltransferase	1237402	-

Table 3. Heterozygosity statistics and selfing rate estimates of *M. lupulina* populations.

Population	H observed	H expected	F_{IS}	Selfing rate (s)
AIL	0.022	0.031	0.290322581	0.45
AV	0.012	0.089	0.865168539	0.927710843
BLU	0.005	0.062	0.919354839	0.957983193
BR	0.037	0.11	0.663636364	0.797814208
BUF	0.004	0.002	-	-
CAL1	0.025	0.099	0.747474747	0.855491329
CAL2	0.043	0.096	0.552083333	0.711409396
COB	0.032	0.091	0.648351648	0.786666667
COO	0.038	0.112	0.660714286	0.795698925
DE	0.014	0.07	0.8	0.888888889
DUN	0.051	0.115	0.556521739	0.715083799
FOR	0.014	0.129	0.891472868	0.942622951
GA	0.006	0.083	0.927710843	0.9625
GIL	0.007	0.082	0.914634146	0.955414013
GOD	0.031	0.15	0.793333333	0.884758364
HA1	0.006	0.129	0.953488372	0.976190476
HA2	0.068	0.123	0.447154472	0.617977528
HL	0.018	0.077	0.766233766	0.867647059
HO	0.033	0.129	0.744186047	0.853333333
KA	0.026	0.091	0.714285714	0.833333333
KIT	0.025	0.147	0.829931973	0.907063197
KSR	0.048	0.11	0.563636364	0.720930233
MAP	0.043	0.102	0.578431373	0.732919255
MTC	0.095	0.112	0.151785714	0.263565891
NEY	0.01	0.075	0.866666667	0.928571429
NH	0.004	0.062	0.935483871	0.966666667
NY	0.006	0.12	0.95	0.974358974
PA	0.029	0.124	0.766129032	0.867579909
PAR	0.026	0.027	0.037037037	0.071428571
PT	0.021	0.12	0.825	0.904109589
PUS	0.054	0.131	0.58778626	0.740384615
SAR	0.009	0.09	0.9	0.947368421
SC	0.029	0.141	0.794326241	0.885375494
SEG	0.022	0.111	0.801801802	0.89
SIN	0.043	0.091	0.527472527	0.690647482
TO	0.005	0.057	0.912280702	0.95412844

UBR	0.012	0.006	-	-
WA	0.011	0.096	0.885416667	0.939226519
WAR	0.012	0.082	0.853658537	0.921052632

Table 4. Spatial genetic autocorrelation analyses in *M. lupulina*, in all *Ensifer* samples, and in each *Ensifer* species separately. *n* = sample size; *r* = spatial genetic autocorrelation. Bolded values of *r* are significantly different from the null expectation of a random distribution of genotypes in space at the $\alpha = 0.05$ level.

M. lupulina								
Distance Class	**100**	**200**	**300**	**400**	**500**	**600**	**700**	**800**
n	4066	4928	2179	1316	1072	1693	1949	732
r	**0.044**	**0.006**	**-0.004**	**-0.011**	**-0.028**	**-0.028**	**-0.030**	**-0.038**
P ($r_{random} \geq r_{observed}$)	0.001	0.001	0.985	1.000	1.000	1.000	1.000	1.000
P ($r_{random} \leq r_{observed}$)	1.000	1.000	0.016	0.001	0.001	0.001	0.001	0.001
All *Ensifer* samples								
Distance Class	**100**	**200**	**300**	**400**	**500**	**600**	**700**	**800**
n	779	961	500	330	302	392	398	163
r	**0.024**	**0.030**	**0.030**	0.012	-0.015	**-0.052**	**-0.046**	**-0.095**
P ($r_{random} \geq r_{observed}$)	0.015	0.001	0.006	0.148	0.861	0.999	0.997	0.999
P ($r_{random} \leq r_{observed}$)	0.986	1.000	0.995	0.853	0.140	0.002	0.004	0.002
E. medicae								
Distance Class	**200**	**400**	**600**	**800**				
n	469	241	46	24				
r	0.002	0.010	-0.024	0.071				
P ($r_{random} \geq r_{observed}$)	0.378	0.334	0.852	0.116				
P ($r_{random} \leq r_{observed}$)	0.623	0.667	0.149	0.885				
E. meliloti								
Distance Class	**200**	**400**	**600**	**800**				
n	502	219	310	194				
r	0.001	-0.006	0.004	-0.002				
P ($r_{random} \geq r_{observed}$)	0.343	0.793	0.312	0.648				
P ($r_{random} \leq r_{observed}$)	0.658	0.208	0.689	0.353				

Table 5. Summary statistics on alignment of rhizobia samples to the *E. meliloti* reference genome.

Individual	**Percentage of reads aligned to *E. meliloti***	**Mean depth per individual**
AIL_1_3	87.94	109.819
AV_1_1	81.48	26.1715
AV_1_4	87.84	95.6372
BLU1_2	78.91	65.6438
BLU1_6	82.18	133.294
BUF_1_3	86.65	122.791
CAL_1_2	85.63	125.700
CAL_1_6	85.78	89.1354
CAL_2_1	87.36	75.6606
CAL_2_2	84.9	130.457
COO_1_4	83.91	95.0817
COO_1_5	86.18	115.610
DE_1_2	88.86	72.648
DE_1_3	94.02	104.408
DE_1_4	87.51	49.6715
DUN_1_2	88.63	110.245
DUN_1_6	71.04	103.814
FOR_1_4	86.74	77.9982
GA_1_5	86.38	108.029
GIL_1_1	84.66	81.7762
GIL_1_2	81.81	24.1137
HA_1_2	84.24	103.43

HA_2_4	81.66	69.9097
HL_1_4	87.93	91.4477
KIT_1_6	84.56	84.778
KSR_1_1	87.19	95.8881
MTC_1_3	87.27	95.769
NEY_1_5	83.11	75.7076
NEY_1_7	81.91	121.482
NEY_1_8	88.28	40.6065
NH_1_5	82.4	22.3953
PAN_1_2	86.23	114.585
PAN_1_3	86.85	82.0668
PAN_1_6	87.44	66.0487
PAR_1_1	87.11	122.511
PA_1_6	83.5	105.081
PA_1_7	85.63	99.8014
PA_1_8	87.08	79.1282
PT_1_10	86.51	76.7455
PT_1_4	84.44	47.1245
PT_1_5	85.84	43.2383
PUS_1_3	75.49	20.6311
SC_1_3	83.82	68.8989
SC_1_4	77.81	60.9116
SIN_1_1	82.49	54.8574
SIN_1_2	82.68	69.37
TO_3_1	84.97	68.7274
TO_3_2	84.56	52.4982
UBR_1_2	83.36	56.0072

UBR_1_3	87.54	28.9675

Table 6. Summary statistics on alignment of rhizobia samples to the *E. medicae* reference genome.

Individual	**Percentage of reads aligned to *E. medicae***	**Mean depth per individual**
AIL_1_1	85.54	131.722
BR_1_2	74.38	57.8205
BR_1_3	89.78	76.8122
BUF_1_4	87.80	71.099
BUF_1_6	69.95	52.9103
COB_1_4	82.77	60.1591
COB_1_5	91.73	80.0176
DUN_1_8	81.42	76.556
FOR_1_1	82.72	66.4653
GA_1_1	82.41	53.4903
GIL_1_3	87.68	63.2636
GOD_1_7	69.32	62.2766
HA_2_2	87.19	79.296
HO_1_1	87.28	64.6651
HO_1_5	83.79	64.0444
KA_1_4	82.74	74.5689
KA_1_5	81.14	64.9704
KA_1_6	77.17	54.2248
KA_1_7	79.77	57.8899
KIT_1_1	83.33	63.1471
KSR_1_3	65.82	50.7817
MAP_1_13	80.45	76.9371

MAP_1_6	84.96	61.4672
MTC_1_4	85.39	57.6762
NEY_1_4	81.73	56.2387
NEY_1_6	82.14	72.408
NH_1_1a	86.10	65.6901
NY_1_1	92.48	53.5809
NY_1_7	83.35	73.4052
SAR_1_1	79.40	53.1739
SAR_1_4	88.22	70.6346
SAR_1_8	83.16	46.5088
SEG_1_5	88.54	26.5504
UBR_1_8	82.39	27.9963
WAR_1_4	86.49	32.5708
WAR_1_8	83.18	55.0148
WA_1_4	83.58	26.0675
WA_1_5	85.28	35.0268

Figures

Figure 1. Structure plot of all *Ensifer* samples. Sequences from all *Ensifer* samples were aligned to the *E. meliloti* reference genome.

Figure 2. Neighbour joining tree of all *Ensifer* samples. Sequences from all *Ensifer* samples were aligned to the *E. meliloti* reference genome. Red branches represent *E. meliloti* samples and blue branches represent *E. medicae*.

Figure 3. Structure plot of all *Ensifer* samples. Sequences from all *Ensifer* samples were aligned to the *E. medicae* reference genome.

Figure 4. Spatial genetic autocorrelation analysis. A) *M. lupulina*. B) All *Ensifer* samples. C) *E. medicae*. D) *E. meliloti*.

Figure 5. Minor allele frequency plots of a) *E. meliloti*, and b) *E. medicae*. Proportion (y-axis) represents the portion of minor alleles that are assigned to each frequency bin.

Figure 6. Phylogenetic analysis of *E. meliloti* genome structures. A) Phylogeny estimated using chromosome SNPs. B) Phylogeny estimated using pSymA SNPs. C) Phylogeny estimated using pSymB SNPs. Neighbour joining tree with posterior support given at each node. Scale bar represents number of nucleotide differences.

Figure 7. Phylogenetic analysis of *E. medicae* genome structures. A) Phylogeny estimated using chromosome SNPs. B) Phylogeny estimated using pSMED01 SNPs. C) Phylogeny estimated using pSMED02 SNPs. Neighbour joining tree with posterior support given at each node. Scale bar represents number of nucleotide differences.

Figure 1.

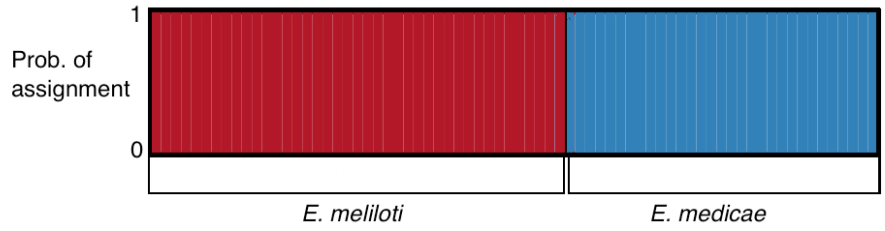


Figure 2.

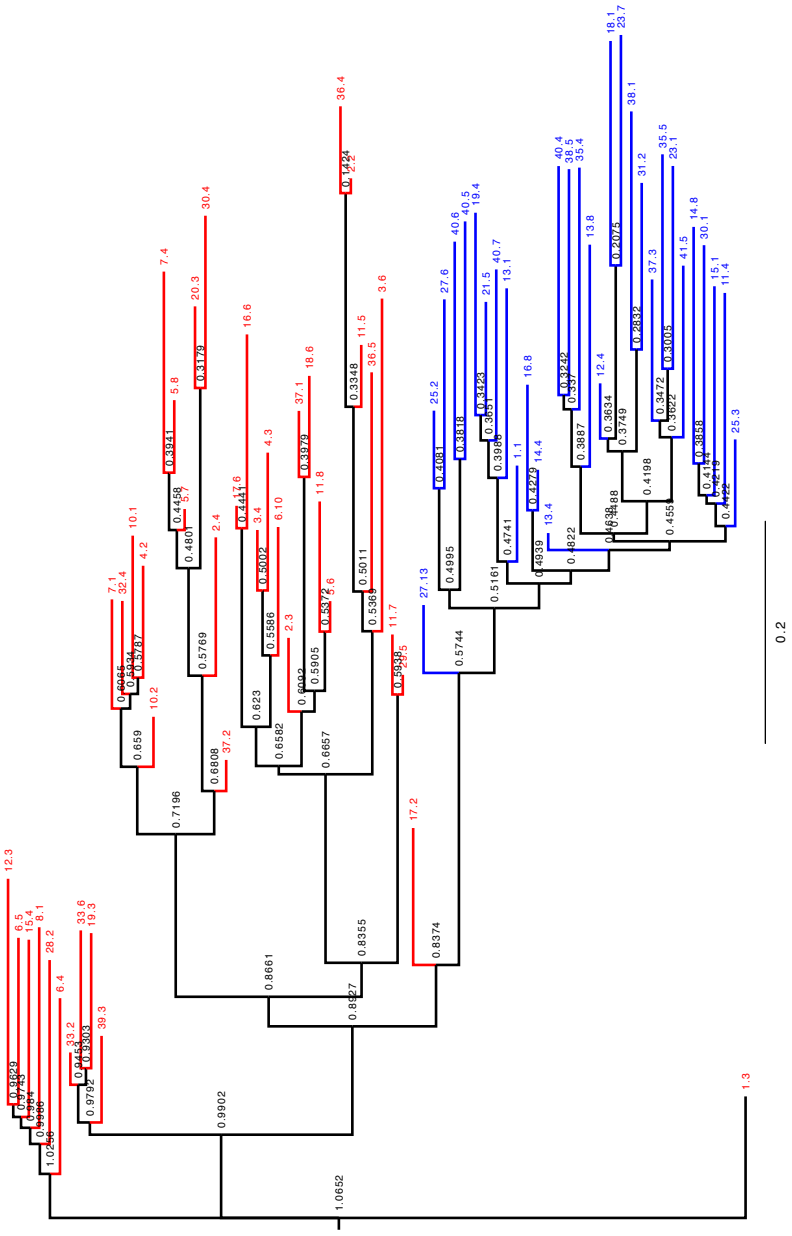


Figure 3.

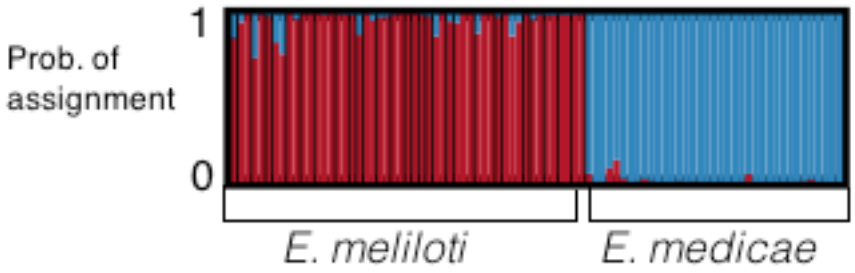
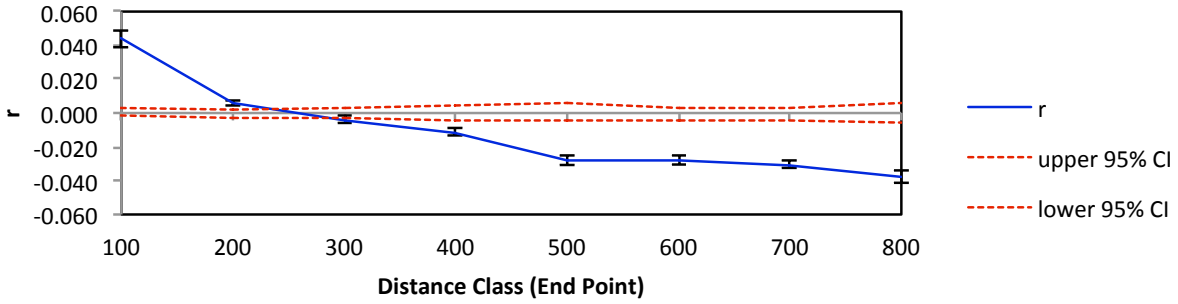
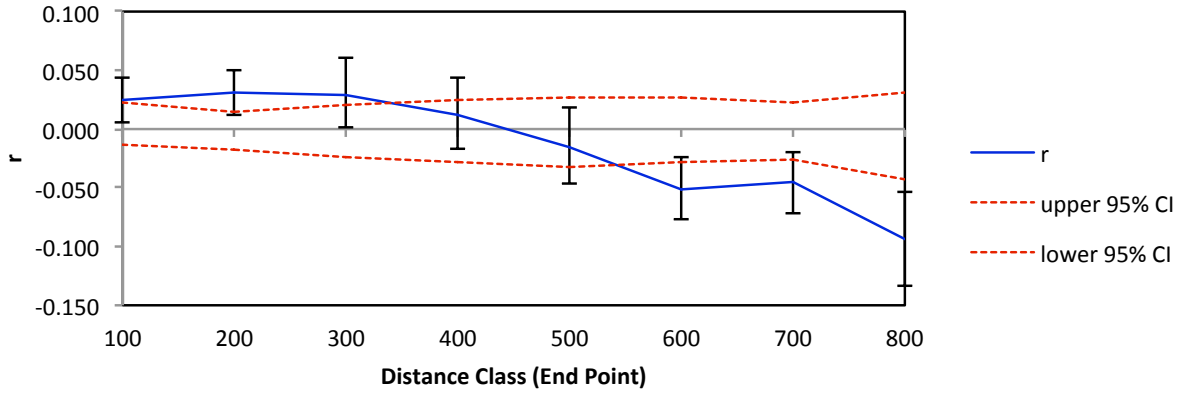


Figure 4.

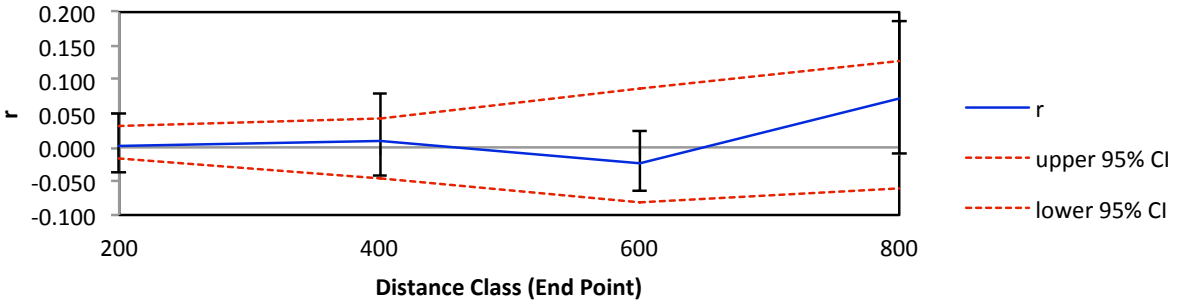
(A) *M. lupulina*



(B) All *Ensifer* samples



(C) *E. medicae*



(D) *E. meliloti*

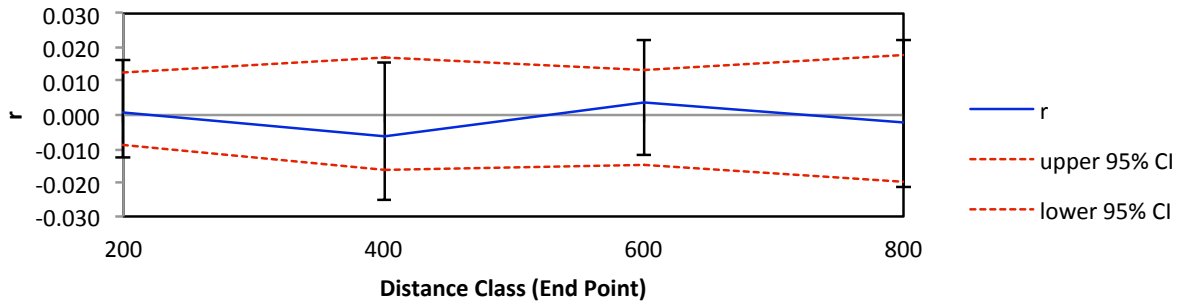


Figure 5.

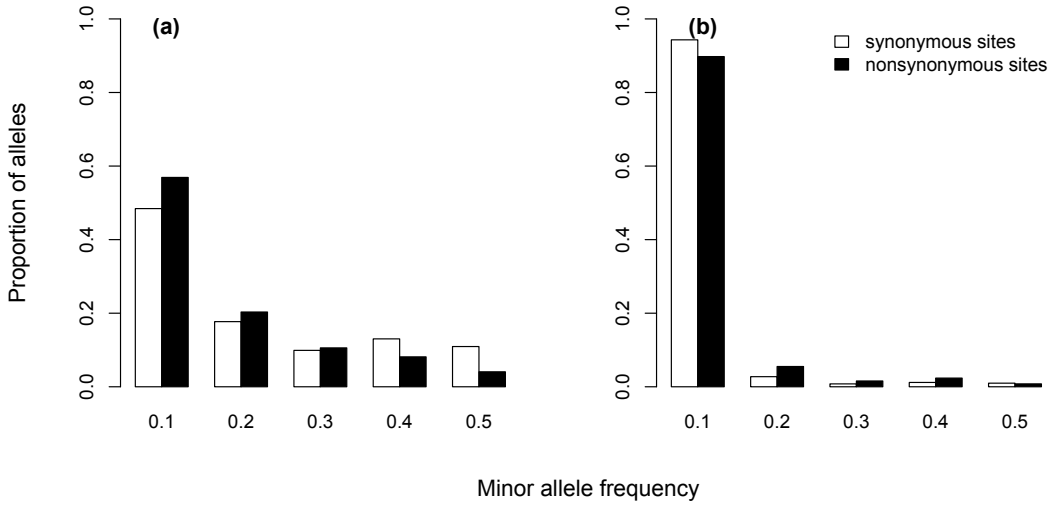

(a)
(b)
synonymous sites
nonsynonymous sites
Proportion of alleles
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5
Minor allele frequency

Figure 6.

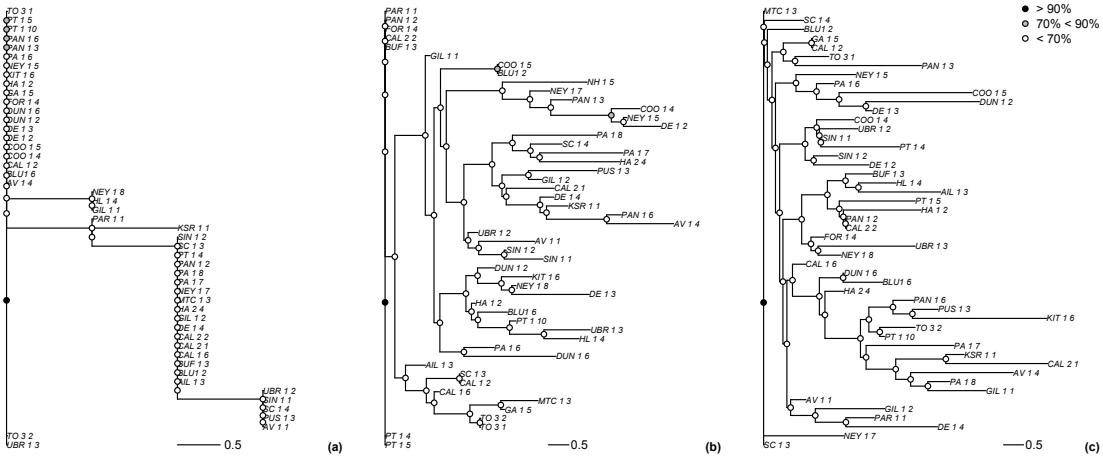
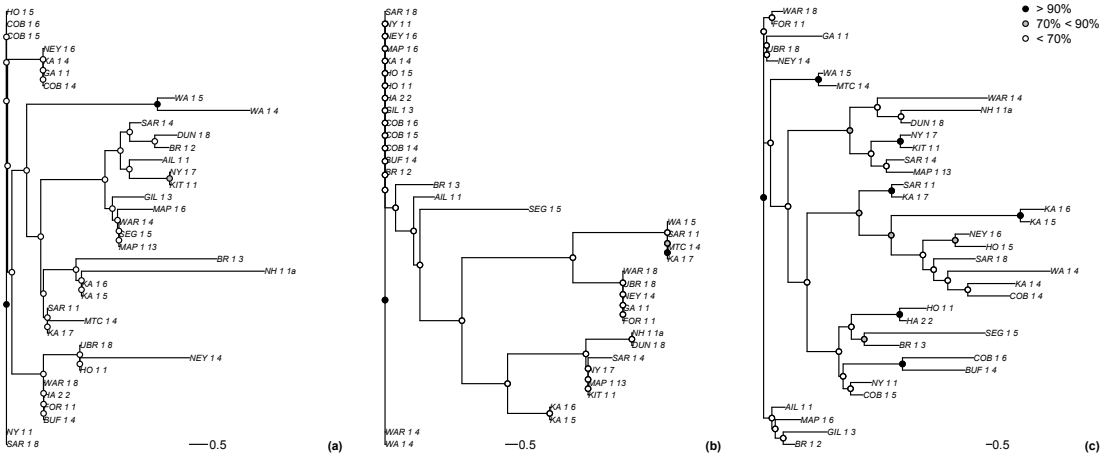

TO 3 1
PT 1 5
PT 1 10
PAN 1 6
PAN 1 3
PA 1 6
NEY 1 5
KIT 1 6
HA 1 2
SA 1 5
FOR 1 4
DUN 1 6
DUN 1 2
DE 1 3
DE 1 2
COO 1 5
COO 1 4
CAL 1 2
BLU1 6
AV 1 4
NEY 1 8
HL 1 4
GIL 1 1
PAR 1 1
KSR 1 1
SIN 1 2
SC 1 3
PT 1 4
PAN 1 2
PA 1 8
PA 1 7
NEY 1 7
MTC 1 3
HA 2 4
GIL 1 2
DE 1 4
CAL 2 2
CAL 2 1
CAL 1 6
BUF 1 3
BLU1 2
AIL 1 3
UBR 1 2
SIN 1 1
SC 1 4
PUS 1 3
AV 1 1
TO 3 2
UBR 1 3
0.5
(a)
PAR 1 1
PAN 1 2
FOR 1 4
CAL 2 2
BUF 1 3
GIL 1 1
COO 1 5
BLU1 2
NH 1 5
NEY 1 7
PAN 1 3
COO 1 4
NEY 1 5
DE 1 2
PA 1 8
SC 1 4
PA 1 7
HA 2 4
PUS 1 3
GIL 1 2
CAL 2 1
DE 1 4
KSR 1 1
PAN 1 6
AV 1 4
UBR 1 2
AV 1 1
SIN 1 2
SIN 1 2
SIN 1 1
DUN 1 2
KIT 1 6
NEY 1 8
DE 1 3
HA 1 2
BLU1 6
PT 1 10
UBR 1 3
HL 1 4
PA 1 6
DUN 1 6
AIL 1 3
SC 1 3
CAL 1 2
CAL 1 6
MTC 1 3
GA 1 5
TO 3 2
TO 3 1
PT 1 4
PT 1 5
0.5
(b)
MTC 1 3
SC 1 4
BLU1 2
GA 1 5
CAL 1 2
TO 3 1
PAN 1 3
NEY 1 5
PA 1 6
COO 1 5
DUN 1 2
DE 1 3
COO 1 4
UBR 1 2
SIN 1 1
PT 1 4
SIN 1 2
DE 1 2
BUF 1 3
HL 1 4
AIL 1 3
PT 1 5
HA 1 2
PAN 1 2
CAL 2 2
FOR 1 4
UBR 1 3
NEY 1 8
CAL 1 6
DUN 1 6
BLU1 6
HA 2 4
PAN 1 6
PUS 1 3
KIT 1 6
TO 3 2
PT 1 10
PA 1 7
KSR 1 1
CAL 2 1
AV 1 4
PA 1 8
GIL 1 1
AV 1 1
GIL 1 2
PAR 1 1
DE 1 4
NEY 1 7
SC 1 3
0.5
(c)
> 90%
70% < 90%
< 70%

Figure 7.


HO 1 5
COB 1 6
COB 1 5
NEY 1 6
KA 1 4
GA 1 1
COB 1 4
WA 1 5
WA 1 4
SAR 1 4
DUN 1 8
BR 1 2
AIL 1 1
NY 1 7
KIT 1 1
GIL 1 3
MAP 1 6
WAR 1 4
SEG 1 5
MAP 1 13
BR 1 3
NH 1 1a
KA 1 6
KA 1 5
SAR 1 1
MTC 1 4
KA 1 7
UBR 1 8
NEY 1 4
HO 1 1
WAR 1 8
HA 2 2
FOR 1 1
BUF 1 4
NY 1 1
SAR 1 8
0.5
(a)
SAR 1 8
NY 1 1
NEY 1 6
MAP 1 6
KA 1 4
HO 1 5
HO 1 1
HA 2 2
GIL 1 3
COB 1 6
COB 1 5
COB 1 4
BUF 1 4
BR 1 2
BR 1 3
AIL 1 1
SEG 1 5
WA 1 5
SAR 1 1
MTC 1 4
KA 1 7
WAR 1 8
UBR 1 8
NEY 1 4
GA 1 1
FOR 1 1
NH 1 1a
DUN 1 8
SAR 1 4
NY 1 7
MAP 1 13
KIT 1 1
KA 1 6
KA 1 5
WAR 1 4
WA 1 4
0.5
(b)
WAR 1 8
FOR 1 1
GA 1 1
UBR 1 8
NEY 1 4
WA 1 5
MTC 1 4
WAR 1 4
NH 1 1a
DUN 1 8
NY 1 7
KIT 1 1
SAR 1 4
MAP 1 13
SAR 1 1
KA 1 7
KA 1 6
KA 1 5
NEY 1 6
HO 1 5
SAR 1 8
WA 1 4
KA 1 4
COB 1 4
HO 1 1
HA 2 2
SEG 1 5
BR 1 3
COB 1 6
BUF 1 4
NY 1 1
COB 1 5
AIL 1 1
MAP 1 6
GIL 1 3
BR 1 2
0.5
(c)
> 90%
70% < 90%
< 70%