Early Austronesians: Into and Out of Taiwan

Albert Min-Shan Ko, Chung-Yu Chen, Qiaomei Fu, Frederick Delfin, Mingkun Li, Hung-Lin Chiu, Mark Stoneking, and Ying-Chin Ko

Figure S1. Sampling locations
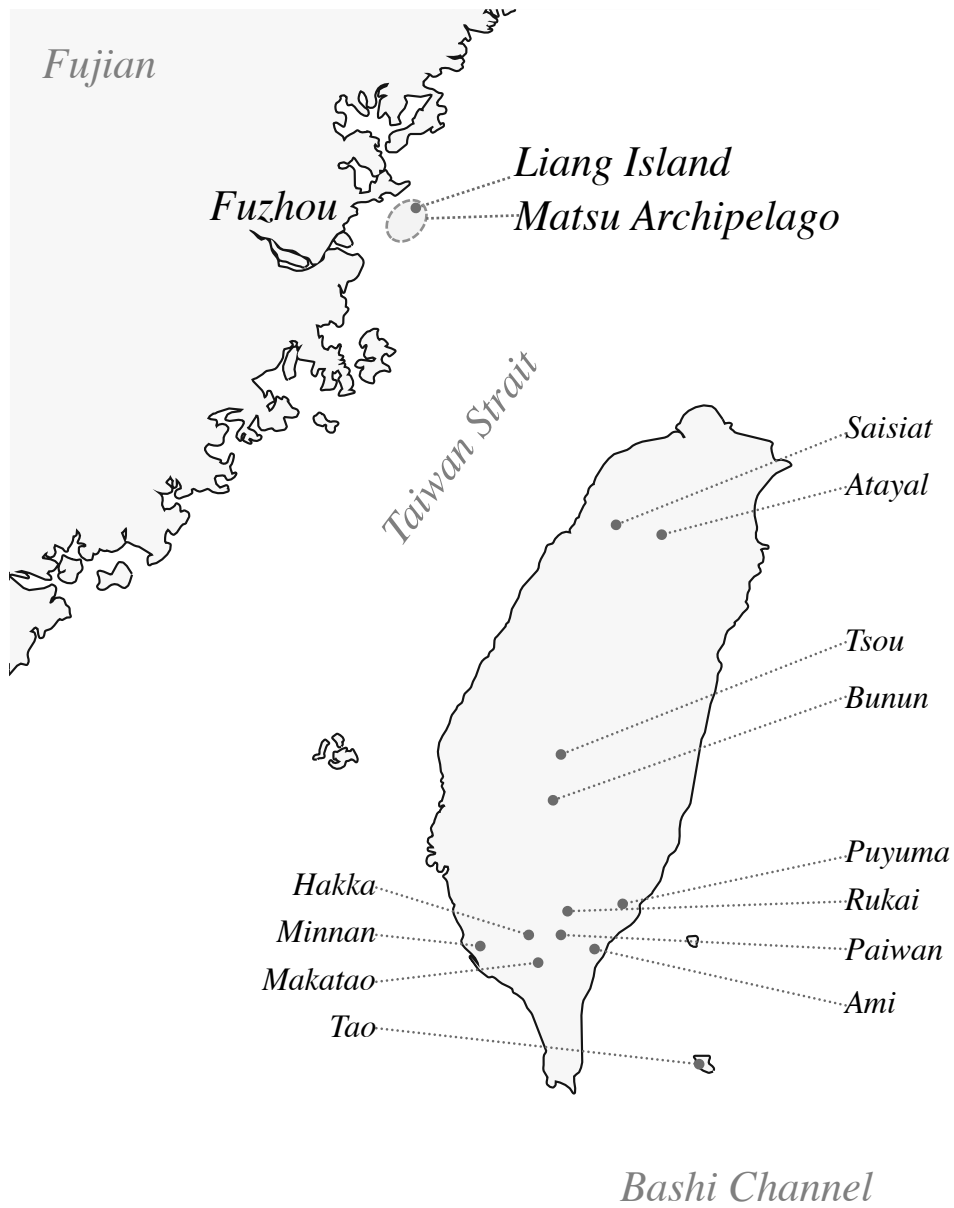
Location of the Liang Island (solid dot) within the Matsu Archipelago of Taiwan (shaded circle) and the 12 Taiwanese groups (solid dots; refer to Table S1 for coordinates)
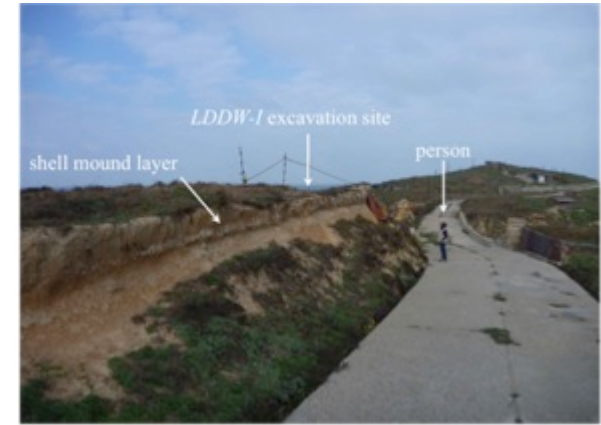
Figure S2. Archeology of the Liangdao Man

(a) From the LDDW-I site, three human parietal bone fragments exposed on the side of the shell mound, which later found to belong to the Liangdao Man

(b) LDDW-I site, the surface conditions prior to excavation, in the area within wooden poles; the road (right) is at a lower level

(c) View from the roadside

(d) LDDW-I site, the layer prior exposure of Liangdao Man, the side adjacent to the road has been exposed for access

(e) LDDW-I site, team wore protective clothing during excavating the burial, to minimize external contamination

(f) LDDW-I site, the burial *in situ*

Figure S3. The C14-AMS dating report of the Liangdao Man

A thoracic rib was sent in May 2012 to Beta Analytic INC., an ISO 17025-accredited radiocarbon-dating lab in Florida, USA. The C14-AMS date of the Liangdao Man is 8060-8320 Cal BP (at 95% probability).

Sample Data	Measured Radiocarbon Age	13C/12C Ratio	Conventional Radiocarbon Age(*)
Beta - 321640	7180 +/- 40 BP	-13.0 o/oo	7380 +/- 40 BP

SAMPLE : LDDW-I-P1 M01
ANALYSIS : AMS-PRIORITY delivery
MATERIAL/PRETREATMENT : (bone collagen): collagen extraction: with alkali
2 SIGMA CALIBRATION : Cal BC 6370 to 6210 (Cal BP 8320 to 8160) AND Cal BC 6140 to 6120 (Cal BP 8090 to 8060)

CALIBRATION OF RADIOCARBON AGE TO CALENDAR YEARS

(Variables: C13/C12=-13:lab. mult=1)

Laboratory number:	Beta-321640
Conventional radiocarbon age:	7380±40 BP
2 Sigma calibrated results: (95% probability)	Cal BC 6370 to 6210 (Cal BP 8320 to 8160) and Cal BC 6140 to 6120 (Cal BP 8090 to 8060)

Intercept data

Intercept of radiocarbon age with calibration curve:	Cal BC 6230 (Cal BP 8180)
1 Sigma calibrated results: (68% probability)	Cal BC 6330 to 6320 (Cal BP 8280 to 8270) and Cal BC 6250 to 6220 (Cal BP 8200 to 8170)
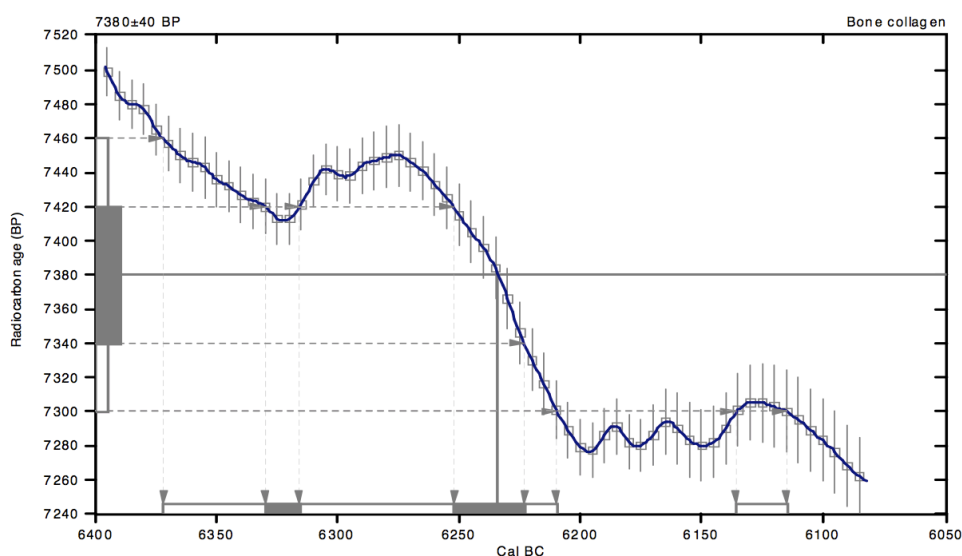
References:
Database used
    INTCAL09
References to INTCAL09 database
    Heaton,et.al.,2009, Radiocarbon 51(4):1151-1164, Reimer,et.al, 2009, Radiocarbon 51(4):1111-1150,
    Stuiver,et.al,1993, Radiocarbon 35(1):137-189, Oeschger,et.al.,1975,Tellus 27:168-192
Mathematics used for calibration scenario
    A Simplified Approach to Calibrating C14 Dates
    Talma, A. S., Vogel, J. C., 1993, Radiocarbon 35(2):317-322

Beta Analytic Radiocarbon Dating Laboratory
4985 S.W. 74th Court, Miami, Florida 33155 • Tel: (305)667-5167 • Fax: (305)663-0964 • E-Mail: beta@radiocarbon.com

Figure S4. Length distribution and base composition in different fragment lengths
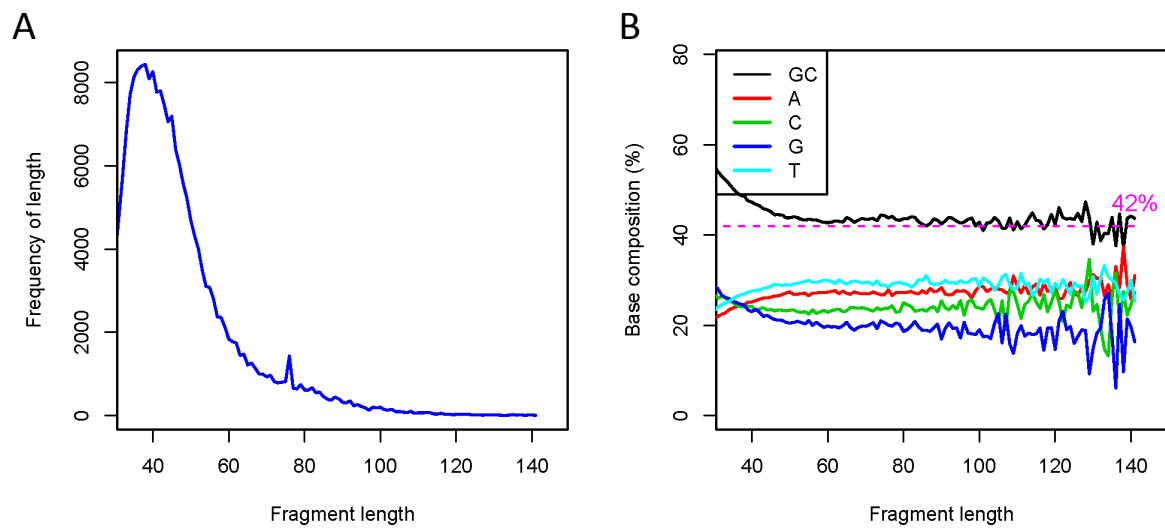
A

B

Figure S5. Proportional support with the consensus and coverage along the mtDNA
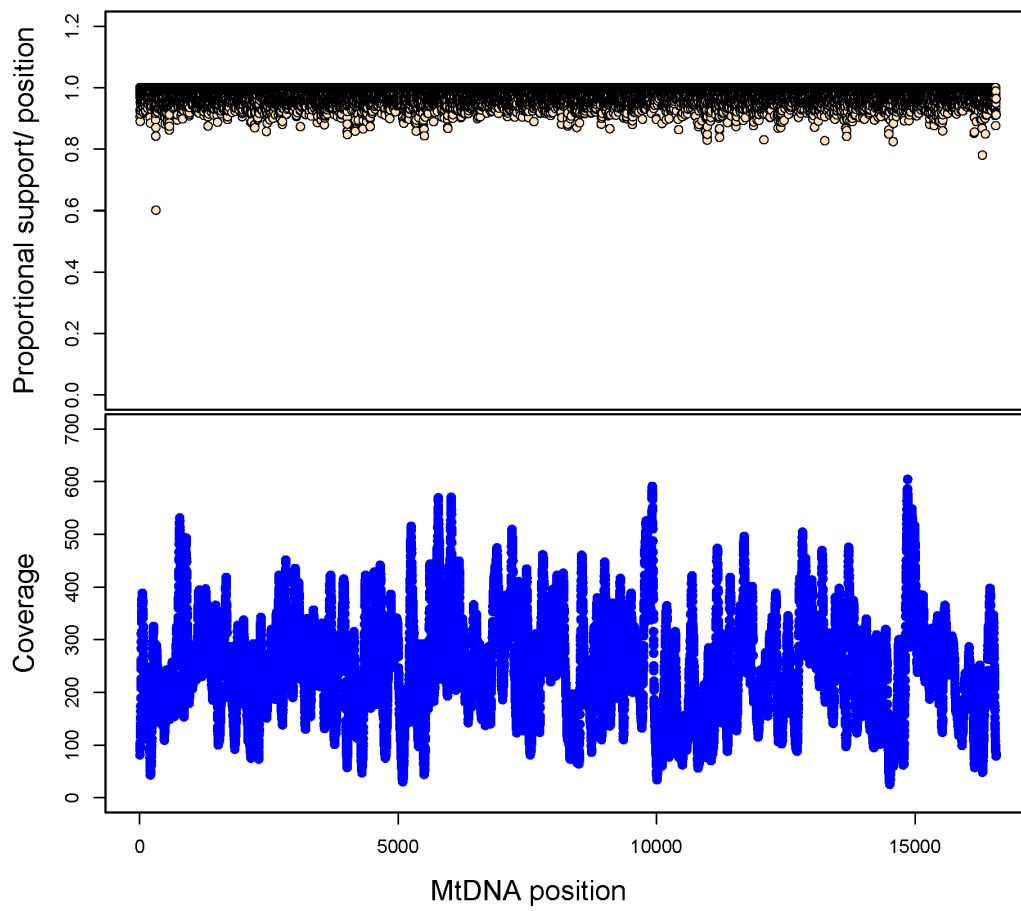
Figure S6. Damage pattern from the 5' end and 3'end of the fragment
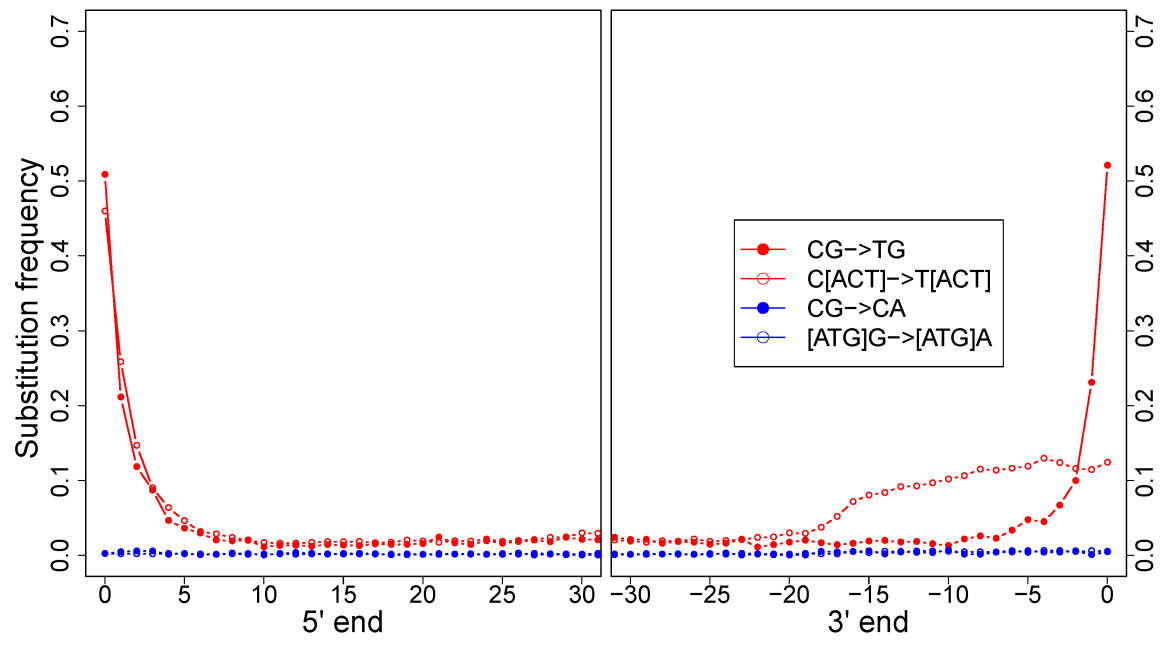
Figure S7. Plot of sequencing coverage per sample of the 12 Taiwanese groups
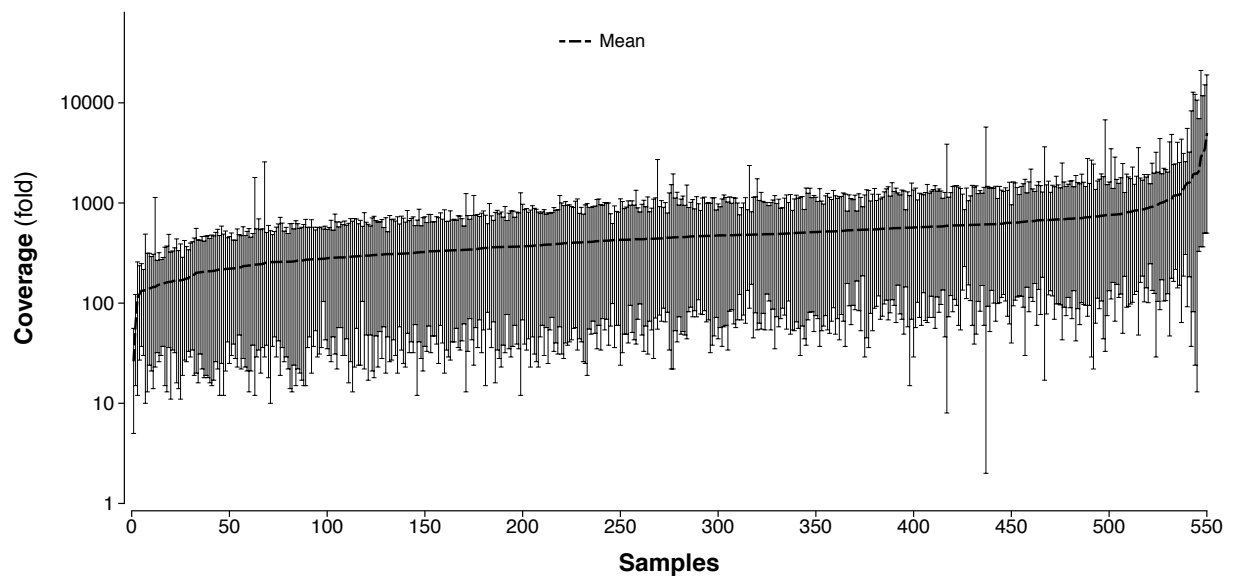
Figure S8. Resampling of Formosan and Han groups to observe sample size effects
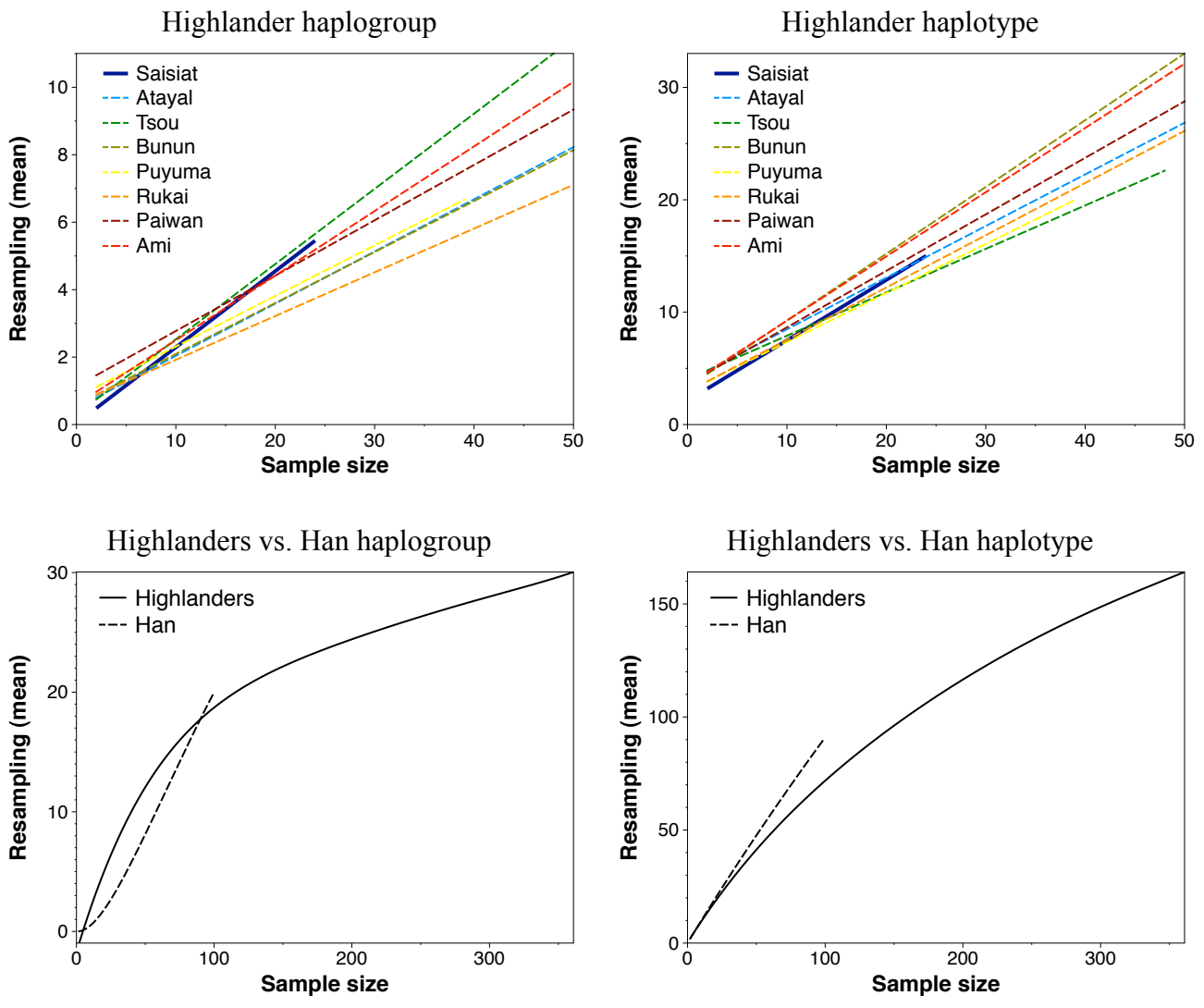
Figure S9. Non-metric multidimensional scaling (NMDS) plots based on Φst distances

(a)
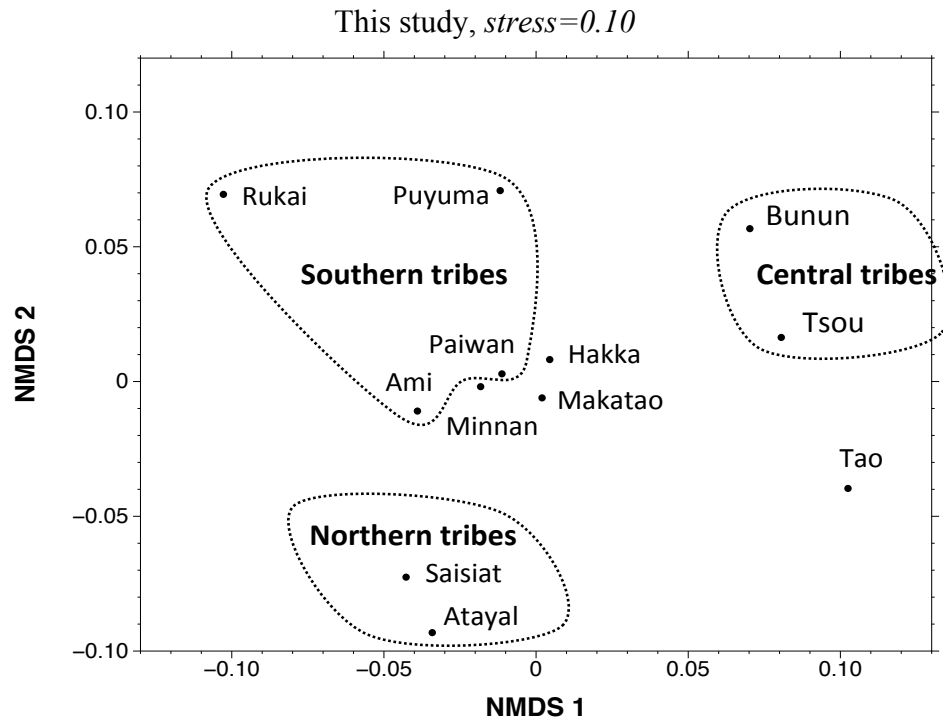
(b)
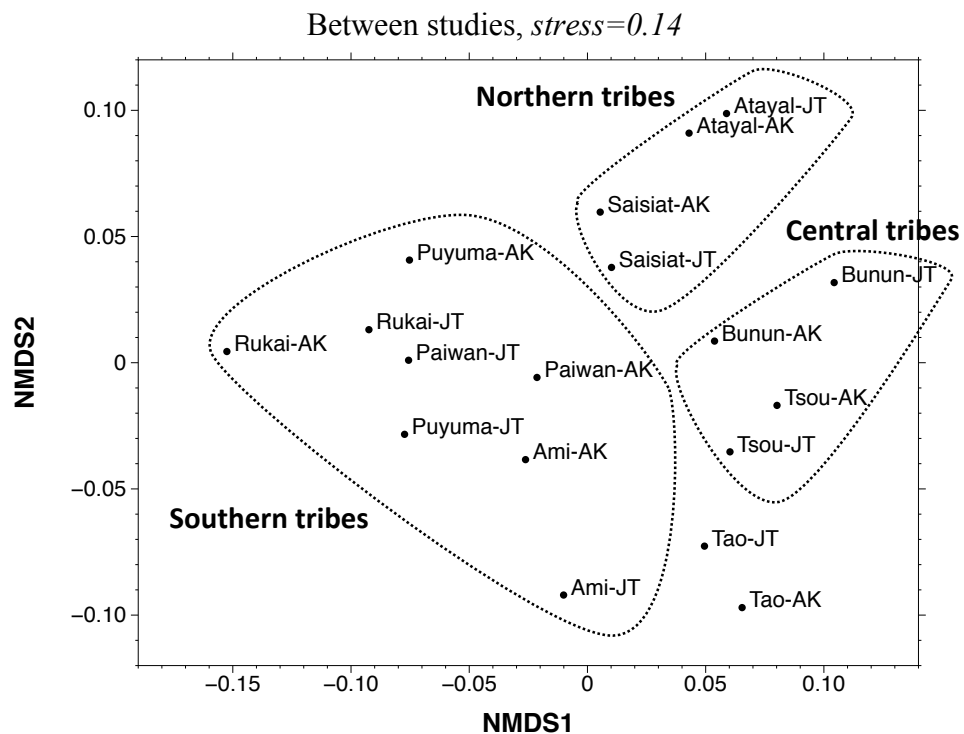
(a) The 12 Taiwanese groups; 550 complete mtDNA genomes (relevant groups circled), (b) Comparison of this study (suffix-AK) and Trejaut *et al.*[1] (suffix-JT) in a combined over 1000 Taiwanese aboriginal sequences, downscaled to 713 bp that include HVS-I and HVS-II.

Figure S10. Haplogroup E calibrated to the Liangdao Man

(a) Strict, modern mtDNA

(b) Relaxed, modern mtDNA

(c) Strict, calibrated by ancient DNA

(d) Relaxed, calibrated by ancient DNA
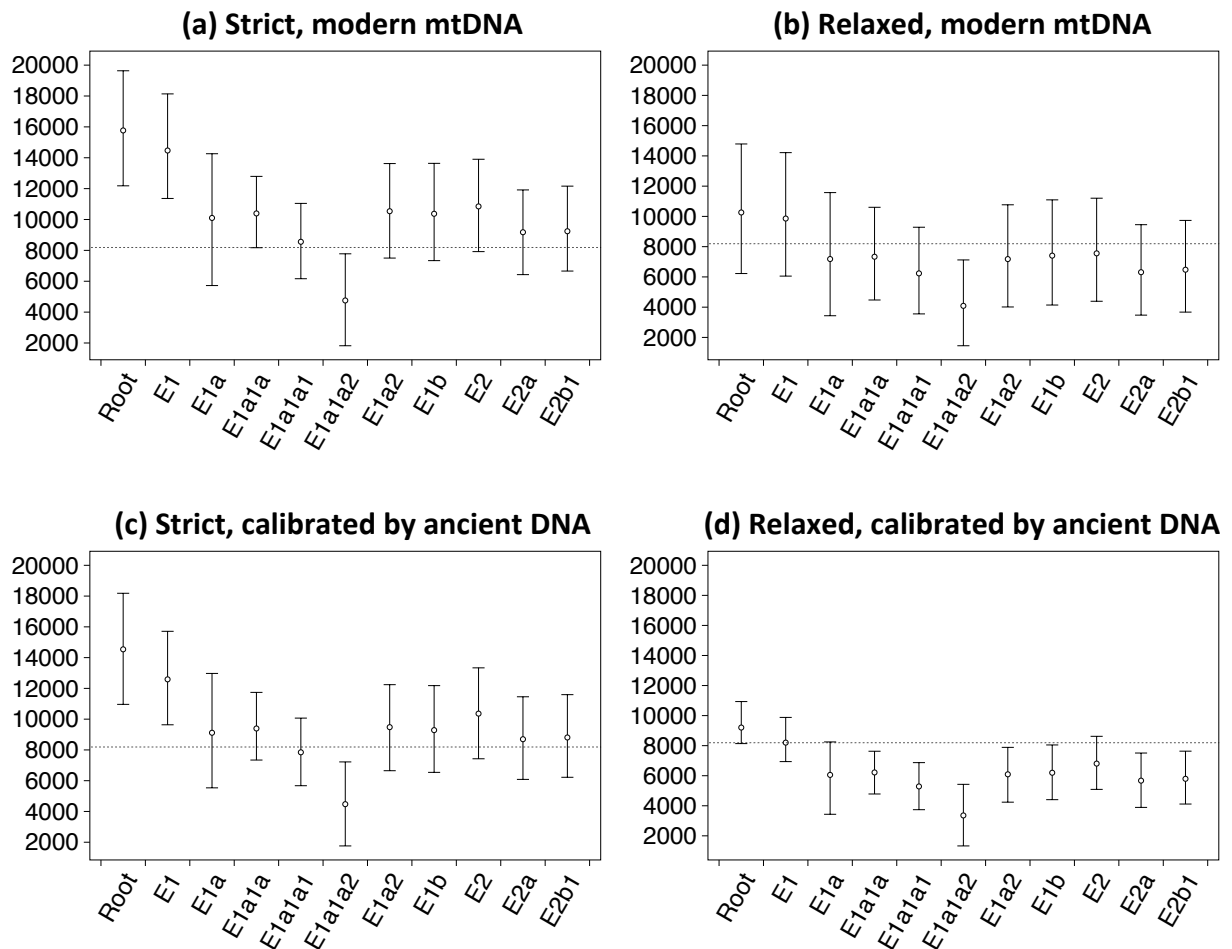
We compared 104 complete mtDNA haplogroup E sequences that are calibrated/uncalibrated to ancient DNA and under strict/relaxed clock methods. The calibrated relaxed clock (d) receives the best support, where many E lineages coalesce at a mean of about 6 ka. The exception is E1a1a2 that is so far observed in the south Formosans, thus reflecting expansion within Taiwan, and its coalescence time largely agrees with 'Formosan phylogeny' simulation of south to diversify 3376-1383 ya (Table S4). The mean (open circle); line is 95% HPD; horizontal dash is mean age of Liangdao Man (8190 ya); y-axis is years BP.

Figure S11. Bayesian skyline plot of the calibrated haplogroup E
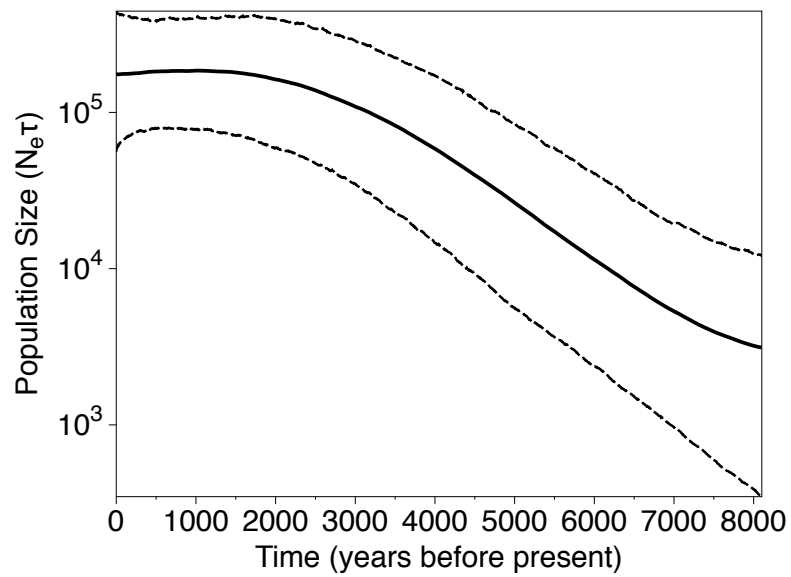
Population size is product of effective size (N_e) and generation time (τ, we assume 25 years/generation); solid line is mean; dash is 95% HPD.

Figure S12. Highland Formosan mtDNA genomes calibrated to the Liangdao Man

(a) Uncalibrated strict clock

(b) Calibrated relaxed clock

(c) Uncalibrated strict clock Haplogroup TMRCA

(d) Calibrated relaxed clock Haplogroup TMRCA
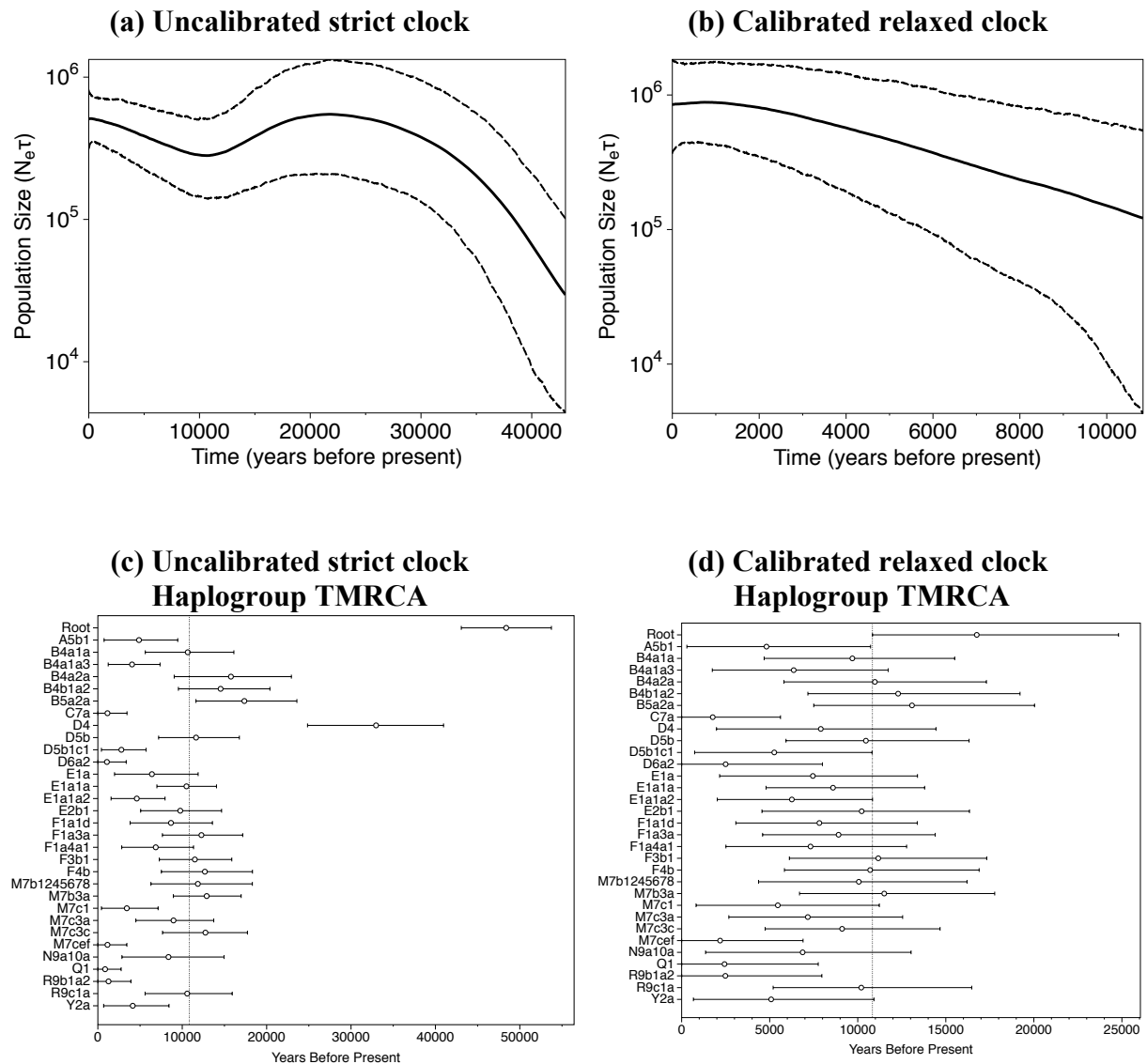
We compared 361 complete Formosan mtDNA sequences under uncalibrated/strict clock and calibrated/relaxed clock methods. The calibrated relax clock (b and d) receives the best support. *(a-b)* Bayesian skyline plot of population size changes over time; population size is product of effective size (N_e) and generation time (τ, we assume 25 years/generation); solid line is mean; dash is 95% HPD; *(c-d)* TMRCA of highland Formosan haplogroups; mean is open circle; line is 95% HPD; vertical dash is 10,835 ya is the start of population expansion from (b) and placed as a reference.

Figure S13. Proposed model for simulation 1: Into/Out-of-Taiwan

Into Taiwan   *Out-of-Taiwan*
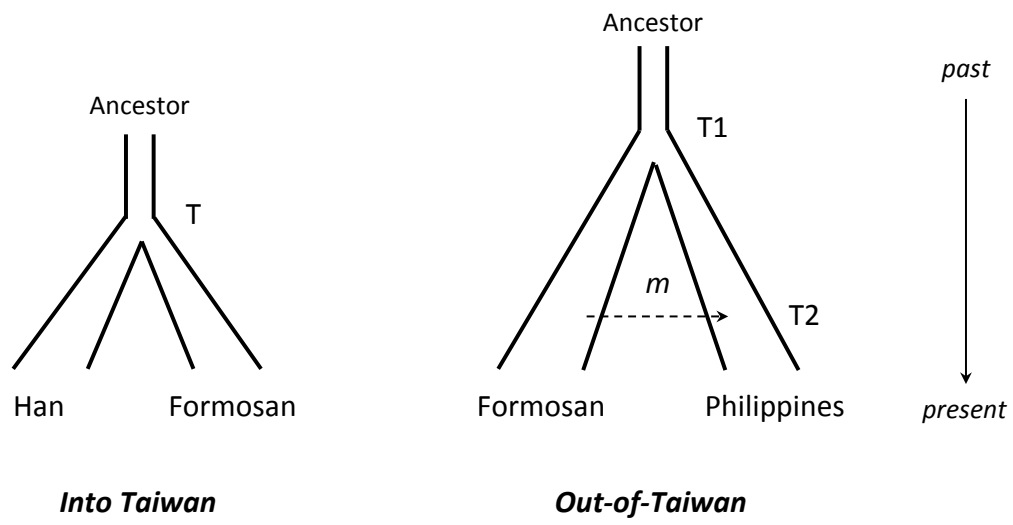
Figure S14. Determining the best shape for the Formosan phylogeny

(a)
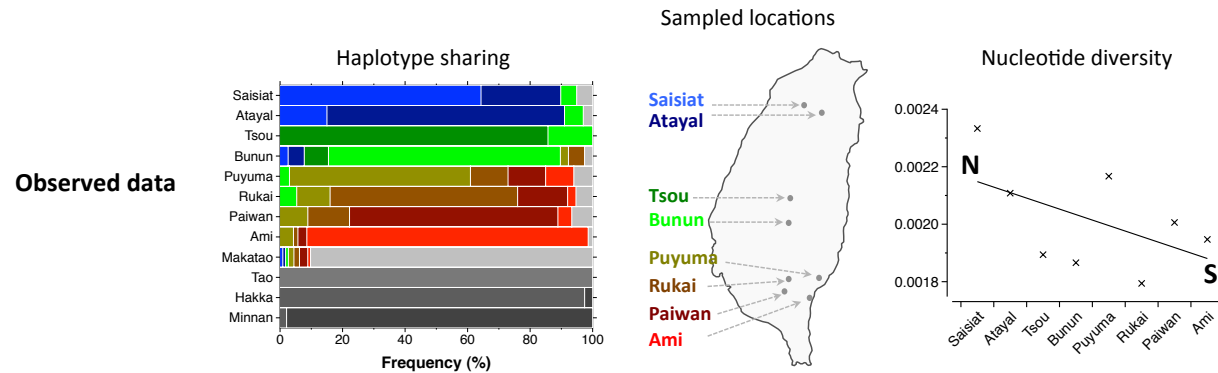
(b)
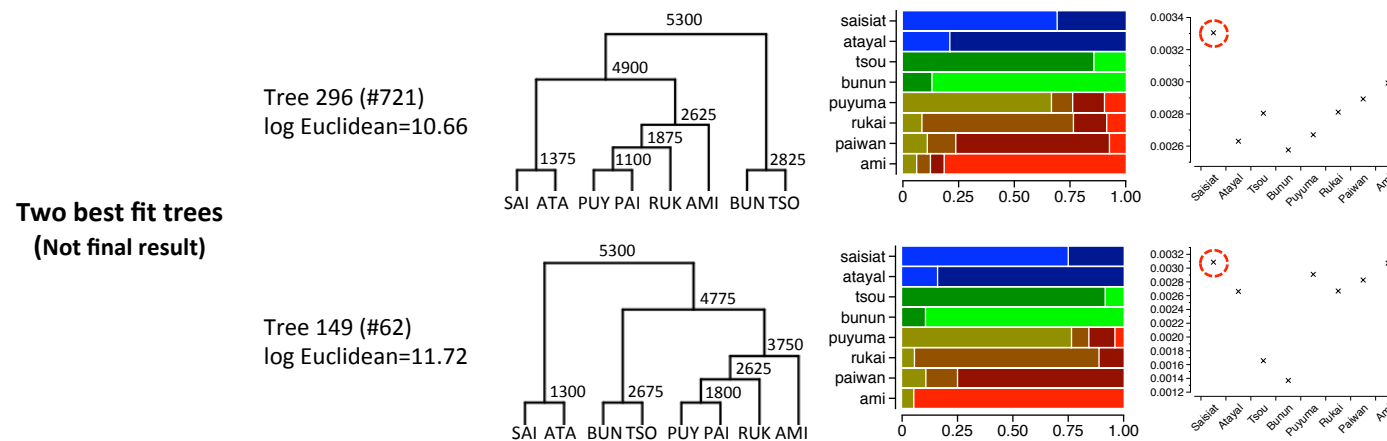
(a) Observed summary of 550 complete mtDNA sequences; each group is colored to sampled locations (center); color bars denote the frequency of haplotypes shared within the group and with other groups. (b) Two best fitting trees, the difference being whether central: tree 296 (#721) or north: tree 149 (#62), diverged first; *middle*, the branch order and divergence times (years BP) and shared haplotypes, *right*, nucleotide diversity, which we required the Saisiat tribe (red circle) to show the highest diversity, however fit is not perfect (log Euclidean distance, *left*).

Figure S15. Improve fit to observed nucleotide diversity using the best shape tree 149(#62)

Top-to-bottom, simulations increasingly distant from the observed

1: log Euclidean=0.503

100: log Euclidean=0.647

1000: log Euclidean=0.807

10,000: log Euclidean=1.085
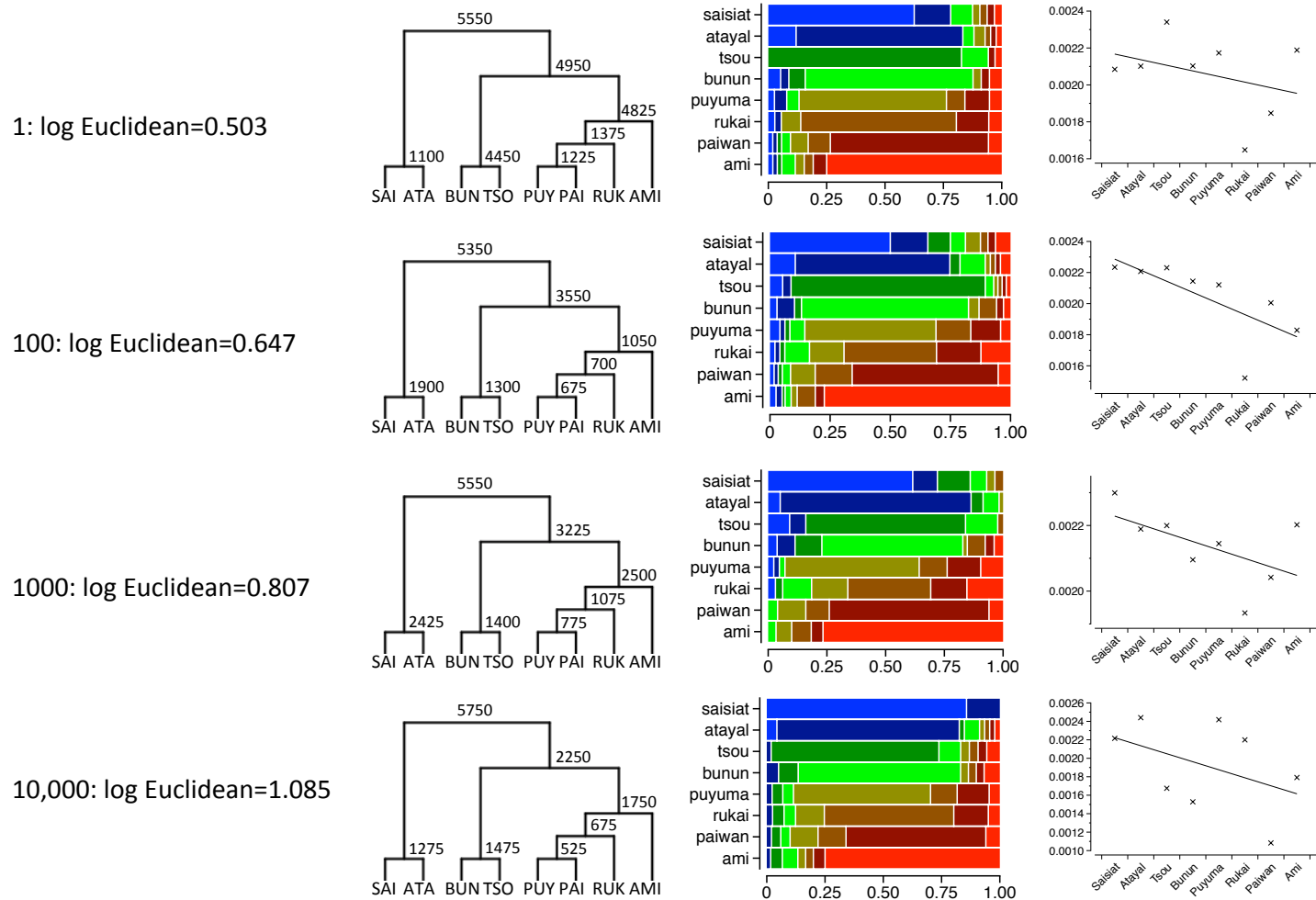
Figure S16. Tree 149(#62) to represent the Formosan phylogeny and Bayesian posterior tree root heights of the four geographic groups
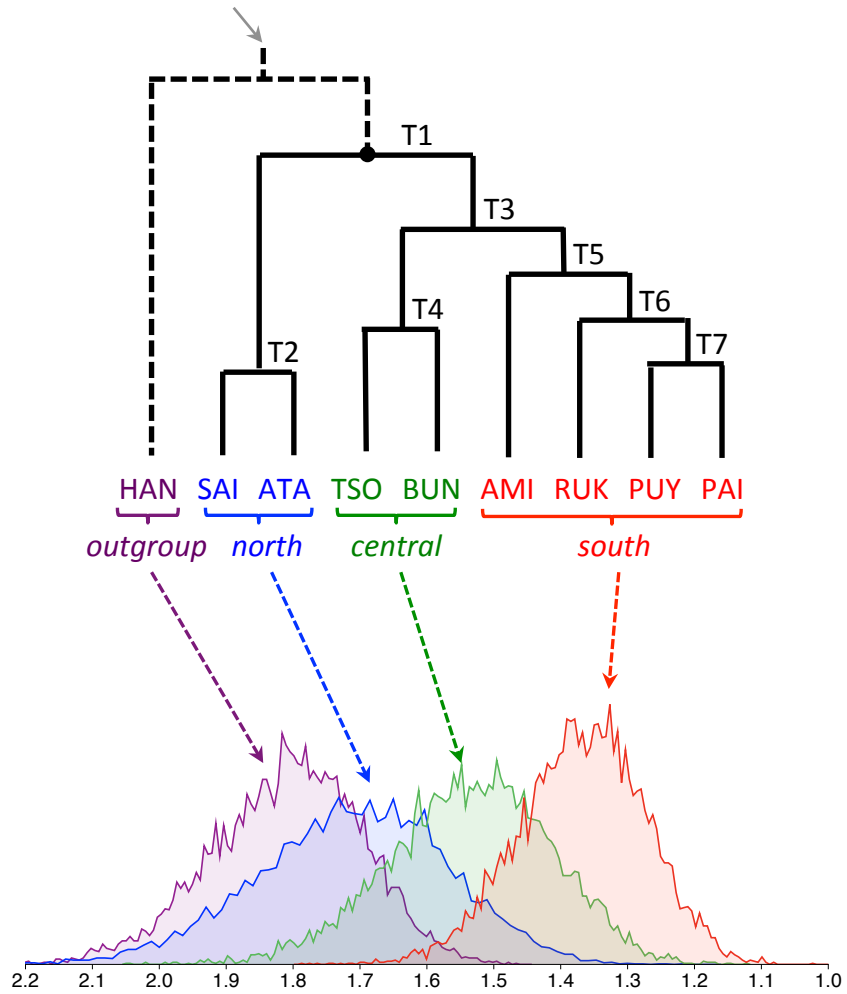
Southward groups show progressively younger coalescent ages (time is relative on the x-axis).

Figure S17. Posterior distribution (red) of the best tree 149 (#62) shows improved fit to the observed (black bar) than prior (blue)
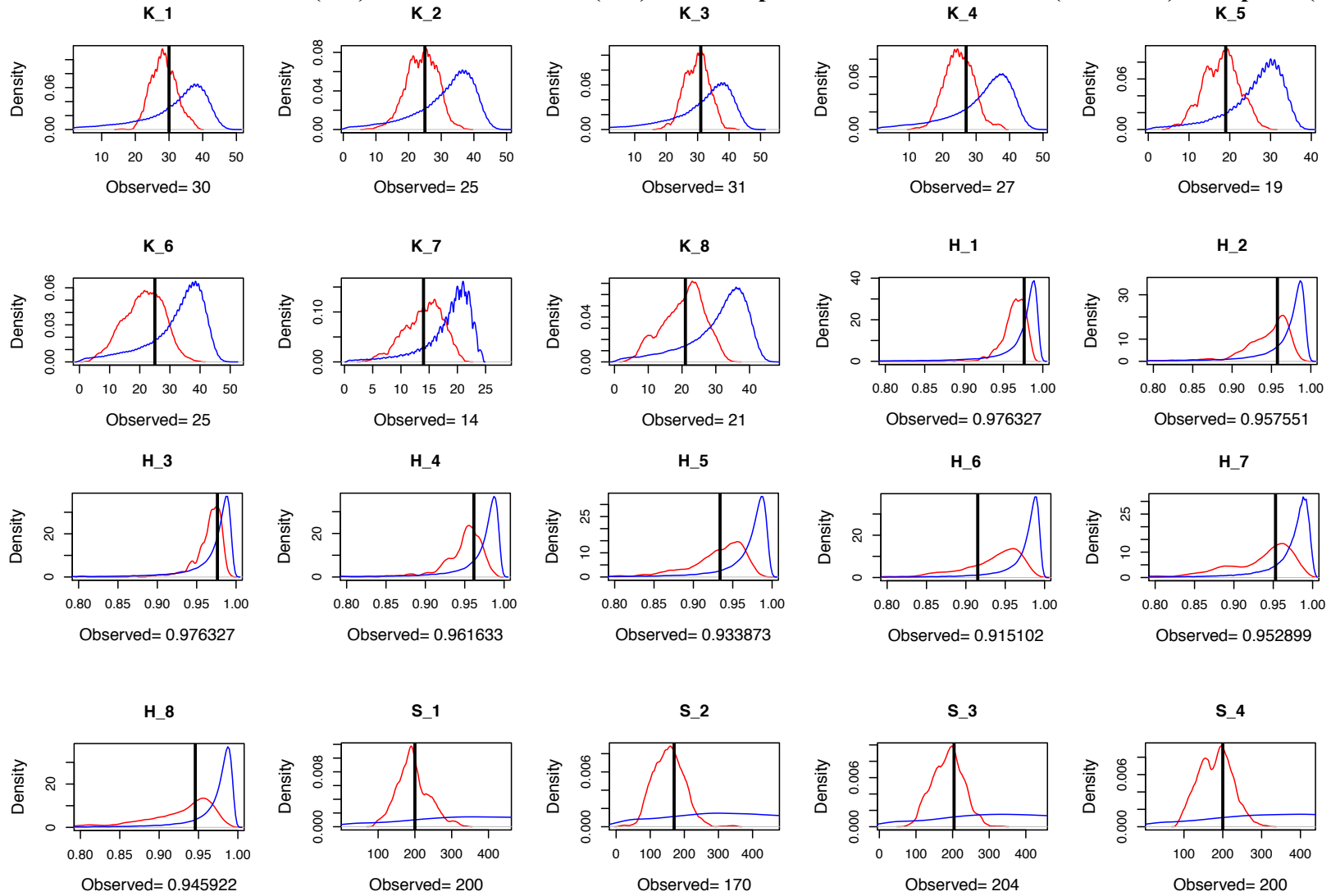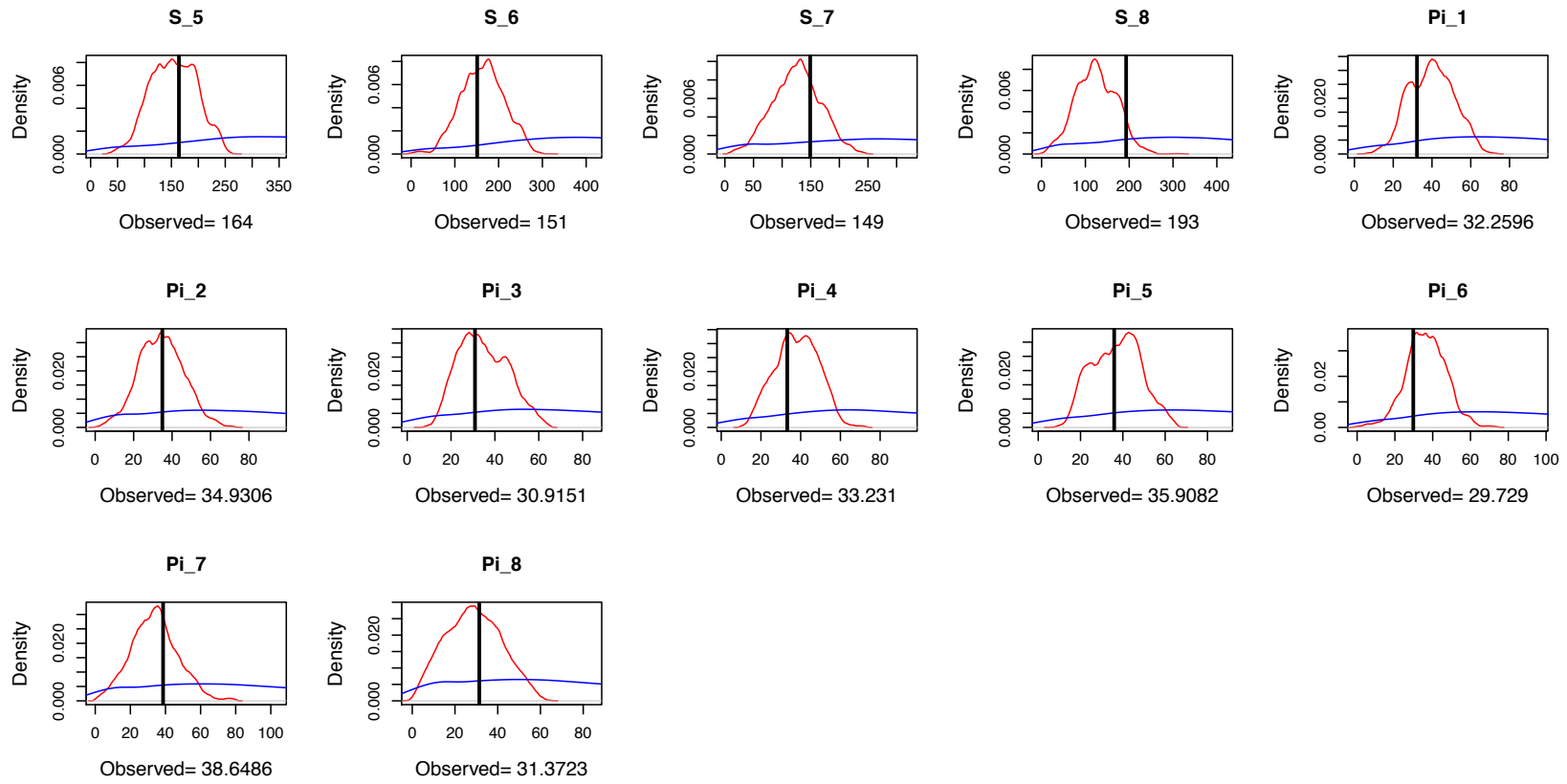
Figure S18. Linguistic relationships used in this study

Blust, 1999   Li, 1998   Ross, 1994   Sagart, 2004
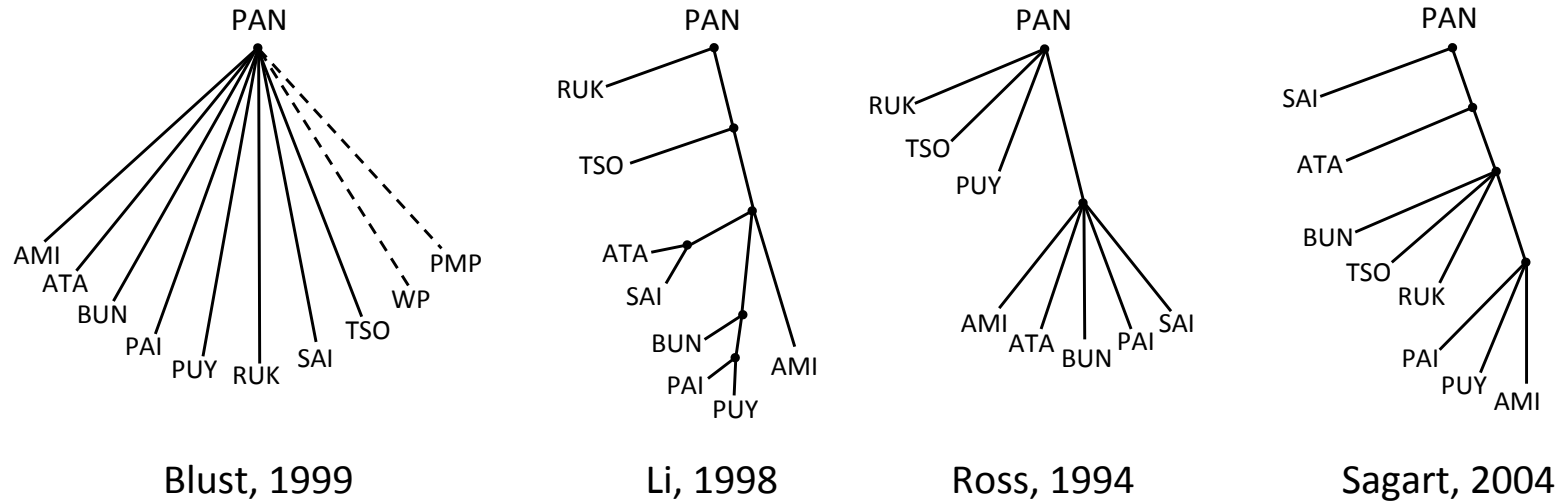
Blust's model[2] has ten branches descended from one common ancestor, here shown eight highlanders (solid lines) and groups not included (dashed lines) are the Western Plainsmen (WP) and Proto-Malayo-Polynesian (PMP); Li's model[3] has two early branches: Rukai and Tsou; Ross's model[4] has Rukai, Tsou and Puyuma as first branch and others under 'Nuclear Austronesian' (expanded for illustration); Sagart's model[5] is based on evolution of the Proto-Austronesian numeral system. Note: tribal placement deriving from a single point does not convey any order or geographic information; the root of each model is Proto-Austronesian (PAN).

Figure S19. Mismatch distributions and Bayesian skyline plots of the 12 Taiwanese groups. Highland Formosans are organized by descending latitude; *left*, mismatch distribution (frequency vs. pairwise differences); *right*, population size changes (effective size × generation time; solid line is mean; dash is 95% HPD).

Saisiat (Northern Highland Formosan)

Atayal (Northern Highland Formosan)

Tsou (Central Highland Formosan)

Bunun (Central Highland Formosan)
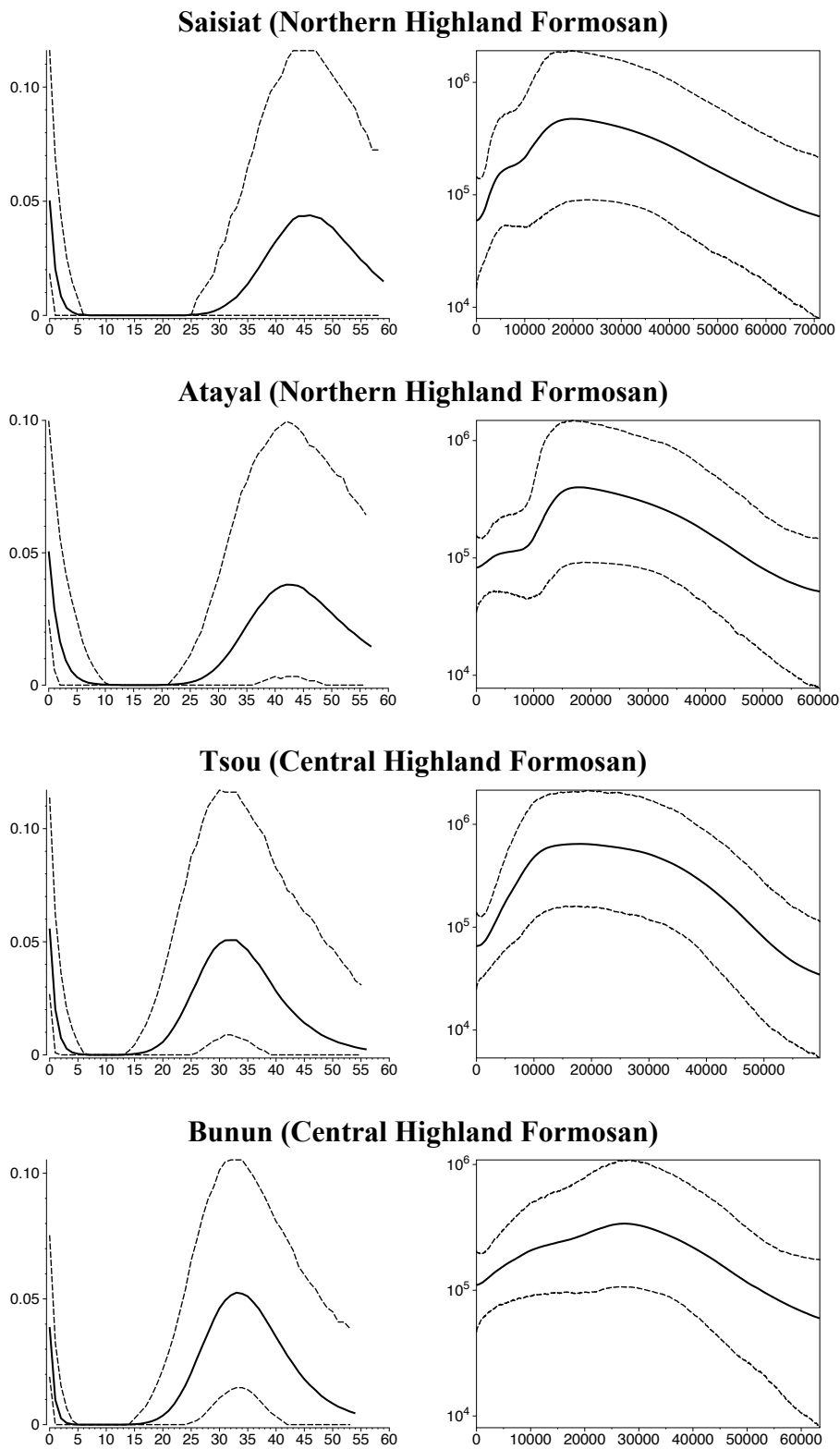
Puyuma (Southern Highland Formosan)
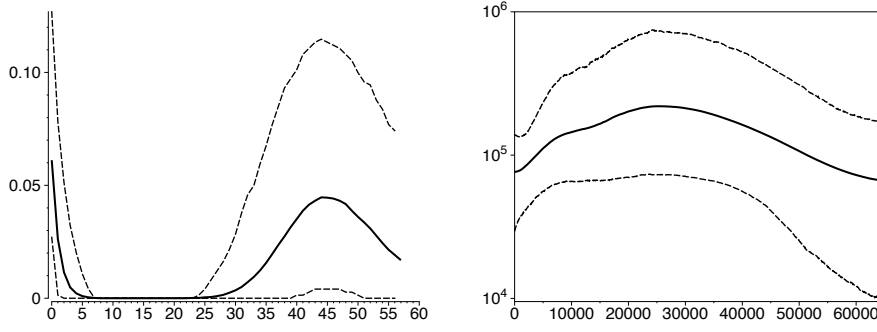
Rukai (Southern Highland Formosan)
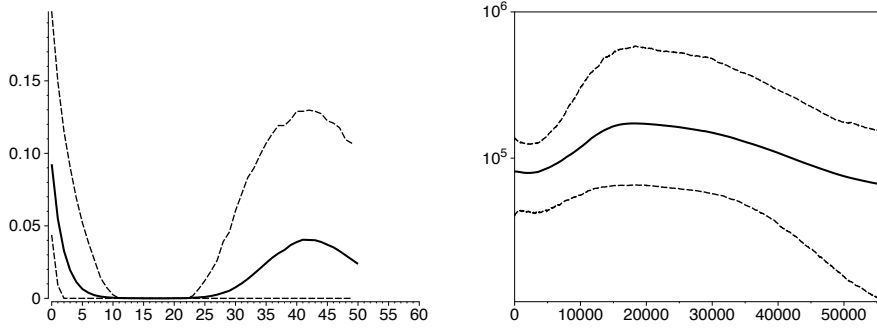
Paiwan (Southern Highland Formosan)
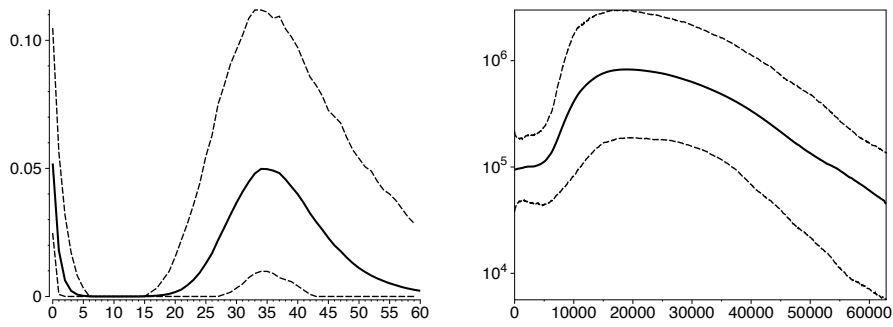
Ami (Southern Highland Formosan)

21
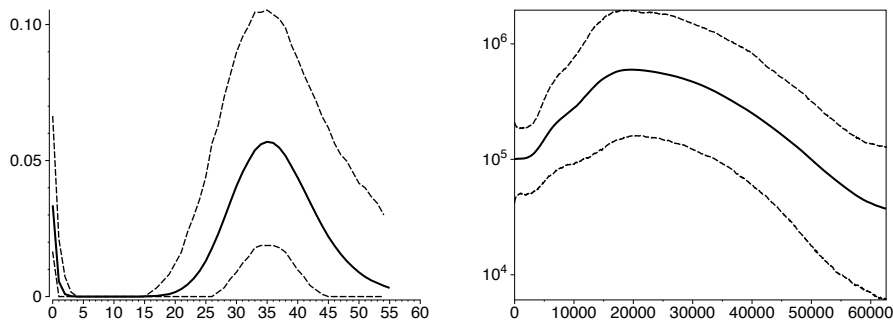
Makatao (Lowland Formosan)
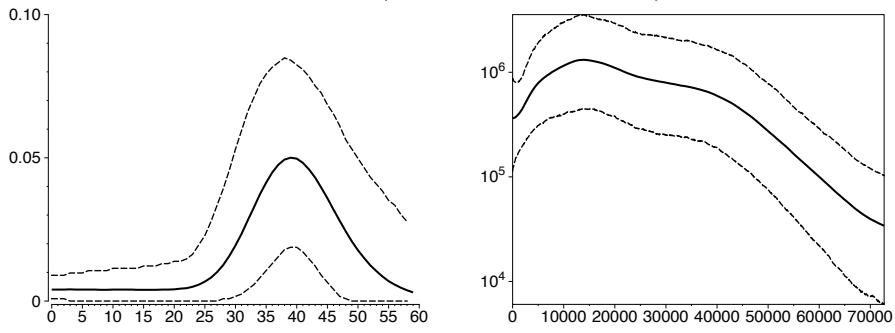
Tao (Taiwanese group)
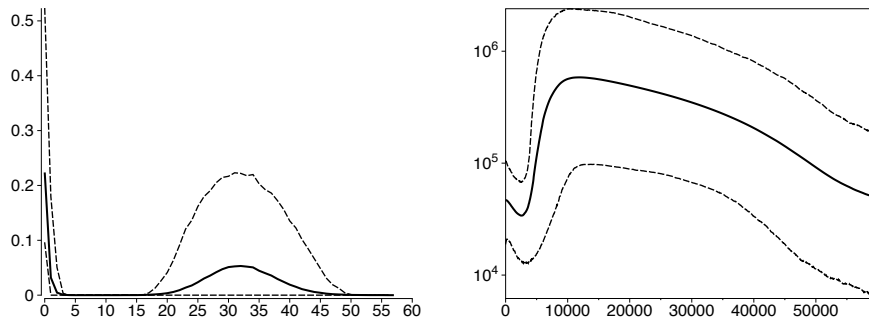
Hakka (Taiwanese group)
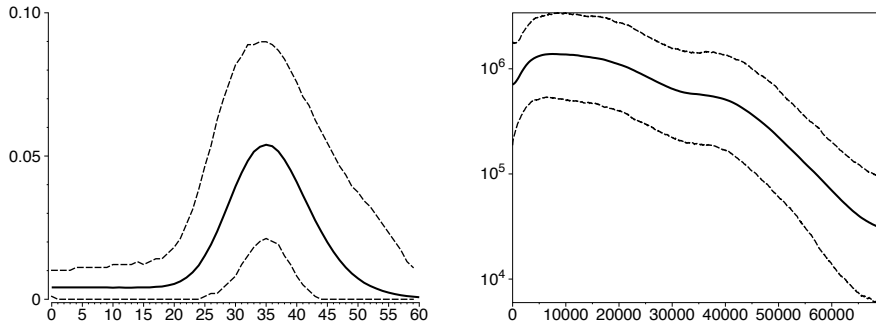
Minnan (Taiwanese group)
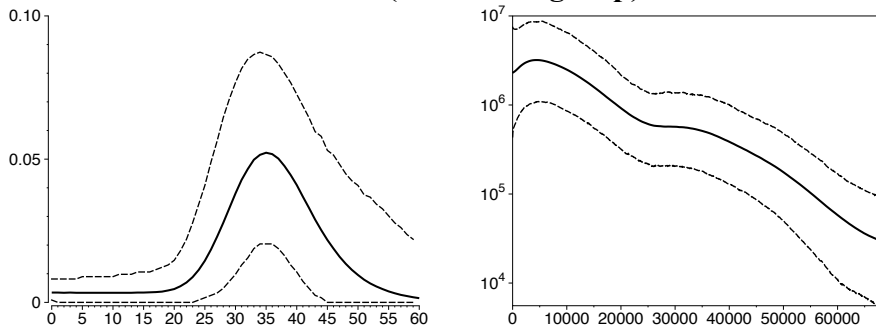
Table S1. Demographic information of the 12 Taiwanese groups

Linguistic affiliation	Group	Code	Sampling coordinates			
			County	Township	Latitude	Longitude
Formosan (highland) †	Saisiat	SAI	Miaoli	Nanjhuang	24°36'	121°0'
	Atayal	ATA	Hsinchu	Jianshih	24°35'	121°17'
	Tsou	TSO	Chiayi	Alishan	23°27'	120°44'
	Bunun	BUN	Kaohsiung	Sanmin	23°13'	120°41'
	Puyuma	PUY	Taitung	Beinan	22°46'	121°5'
	Rukai	RUK	Pingtung	Wutai	22°44'	120°43'
	Paiwan	PAI	Pingtung	Majia	22°40'	120°41'
	Ami	AMI	Taitung	Taimali	22°32'	120°57'
Formosan (lowland)	Makatao	MAK	Pingtung	Wanluan	22°35'	120°36'
Malayo-Polynesian	Tao	TAO	Taitung	Lanyu	22°1'	121°33'
Sinitic	Hakka	HAK	Pingtung	Neipu	22°40'	120°35'
	Minnan	MIN	Kaohsiung		22°37'	120°18'

†highland Formosan tribes ordered by decreasing latitude

Table S2. Parameter estimates of the Into/Out-of-Taiwan

Parameters	Prior			Distribution	Mean	Mode	95% HPD		Pseudo-observed				
									R^2	Bias	RMSE	Coverage	Factor 2
Into Taiwan													
T	1	-	12,000	Uniform	9491	10,066	8093	10,306	0.96	0.03	0.13	91	1
Han	1	-	330,000	Uniform	19,075	19,555	17,902	20,060	0.92	0.01	0.03	85	1
Formosan	1	-	80,000	Uniform	4401	4128	3276	5544	0.98	0.05	0.11	82	1
Ancestor	1	-	330,000	Uniform	2358	2464	2125	2538	0.92	0.08	0.18	70	0.99
μ	$2.16×10^{-8}$	-	$3.16×10^{-8}$	Uniform	$3.07×10^{-8}$	$3.11×10^{-8}$	$2.98×10^{-8}$	$3.12×10^{-8}$	0.99	-0.02	0.06	95	1
Out-of-Taiwan													
T1	6000	-	30,000	Uniform	23,595	28,489	13,725	29,601	0.75	0.05	0.27	91	0.98
T2	1	-	6000	Uniform	4138	4286	3825	4450	0.81	0.02	0.05	70	1
Formosan	1	-	80,000	Uniform	6015	6317	5129	6888	0.93	0.02	0.09	91	1
Philippines	1	-	160,000	Uniform	4337	4814	3368	4951	0.91	0.05	0.11	84	1
Ancestor	1	-	160,000	Uniform	1283	1785	481	2021	0.95	0.3	0.36	63	0.93
m	0	-	1	Uniform	0.0473	0.0154	0.0121	0.093	0.56	0.15	1.02	75	0.53
μ	$2.16×10^{-8}$	-	$3.16×10^{-8}$	Uniform	$2.90×10^{-8}$	$3.10×10^{-8}$	$2.50×10^{-8}$	$3.14×10^{-8}$	0.87	0.02	0.06	74	1

Parameters in this table refers to Figure S13

T, divergence time between Han and Formosan, in years before present

T1, divergence time between the ancestors of Formosan and indigenous of Philippines, in years before present

T2, migration time of the Formosan into northern Philippines, in years before present

Han, Formosan, and Philippines are present-day effective sizes

Ancestor, effective size of ancestral population from which the two populations diverged

m, migration rate per generation

μ, whole mtDNA substitution rate per site per year

Table S3. Improved fit of the posterior to observed than prior

(a) *Into Taiwan*

Statistics	Observed	Prior		Posterior	
		Bias	RMSE	Bias	RMSE
K_1	86	0.07	0.1	0.01	0.02
K_2	164	0.7	0.79	0.09	0.1
H_1	0.998	0	0.02	0	0
H_2	0.991	0.01	0.02	0	0
S_1	584	0.13	0.3	-0.31	0.31
S_2	512	0.41	0.47	-0.22	0.22
D_1	-2.34	-1.44	-1.63	-0.3	-0.32
D_2	-1.63	-1.97	-2.4	-0.3	-0.36
Pi_1	35.7	4.05	4.87	0.14	0.25
Pi_2	37.3	3.69	4.49	0.04	0.21
Φst	0.0228	-0.03	2.69	0.07	0.41
Va	2.28	-0.03	2.69	0.07	0.41
Vb	97.7	0	0.06	0	0.01

(b) *Out-of-Taiwan*

Statistics	Observed	Prior		Posterior	
		Bias	RMSE	Bias	RMSE
K_1	164	0.72	0.8	0.06	0.08
K_2	49	0.31	0.33	0.05	0.06
H_1	0.991	0.01	0.03	0	0
H_2	0.988	0.01	0.02	0	0
S_1	513	0.02	0.19	-0.25	0.26
S_2	270	0.47	0.58	-0.06	0.1
D_1	-1.63	-0.63	-0.96	-0.31	-0.34
D_2	-1.30	-0.4	-0.98	-0.2	-0.29
Pi_1	37.3	0.8	1.28	0.01	0.11
Pi_2	35.5	0.89	1.38	0.05	0.17
Φst	0.0296	-0.75	1.36	-0.15	0.31
Va	2.96	-0.75	1.36	-0.15	0.31
Vb	97.0	0.02	0.04	0	0.01
hs1_1	0.988	0	0.02	0	0
hs1_2	0.0122	-0.22	1.66	-0.17	0.35
hs2_1	0.0408	0.03	2	-0.17	0.34
hs2_2	0.959	0	0.09	0.01	0.01

Abbrev: haplotypes (K), haplotype diversity (H), polymorphic sites (S), Tajima's D (D), pairwise differences (Pi), pairwise Φbb, %variation among groups (Va) and within populations (Vb), haplotype sharing (hs); 1 and 2 is Han/Formosan (a), and Formosan/Philippines (b), respectively.

Table S4. Parameter estimates of the Formosan phylogeny – re-simulations of the best tree 149 (#62)

Parameters	Prior			Distribution	Mean	Mode	95% HPD		Pseudo-observed				
									R^2	Bias	RMSE	Coverage	Factor 2
T1	1	-	6000	Uniform	5386	5243	5035	5814	0.77	0.02	0.07	100	1
T2	1	-	6000	Uniform	1248	1364	485	2050	0.57	0.11	0.67	100	0.72
T3	1	-	6000	Uniform	4226	4271	4049	4452	0.66	0.05	0.16	100	0.99
T4	1	-	6000	Uniform	2281	2437	1807	2654	0.73	0.12	0.72	100	0.67
T5	1	-	6000	Uniform	3376	3423	3010	3794	0.68	0.02	0.32	97	0.96
T6	1	-	6000	Uniform	2266	2355	1600	2687	0.61	0.11	0.55	98	0.77
T7	1	-	6000	Uniform	1383	1098	668	2624	0.73	-0.02	0.68	92	0.74
AMI	1	-	10,000	Uniform	1681	1717	1506	1829	0.67	0.08	0.21	100	1
ATA	1	-	10,000	Uniform	1168	1068	844	1911	0.86	0.11	0.52	100	0.89
BUN	1	-	10,000	Uniform	1902	1898	1871	1937	0.84	-0.02	0.05	100	1
PAI	1	-	10,000	Uniform	1356	1454	957	1796	0.77	0.07	0.35	90	0.96
PUY	1	-	10,000	Uniform	770	798	557	981	0.75	0.15	0.4	99	0.89
RUK	1	-	10,000	Uniform	803	540	367	1419	0.86	0.14	0.74	95	0.87
SAI	1	-	10,000	Uniform	1058	1017	676	1258	0.51	0.14	0.56	99	0.9
TSO	1	-	10,000	Uniform	1277	1511	351	1823	0.89	0.19	0.38	95	0.91
m	0	-	0.05	Uniform	0.00898	0.0103	0.00519	0.0124	0.38	0.48	0.63	95	0.78
μ	2.16×10^{-8}	-	3.16×10^{-8}	Uniform	2.46×10^{-8}	2.4×10^{-8}	2.3×10^{-8}	2.72×10^{-8}	0.79	-0.01	0.11	92	1

Parameters in this table refers to Figure S16
T1-T7, in years before present
AMI-TSO, are present-day effective sizes
m, is 'stepping-stone' migration rate between tribes per generation
μ, whole mtDNA substitution rate per site per year

Table S5. Posterior of the best tree 149 (#62) shows improved fit to observed than prior

Statistics	Observed	Prior		Posterior	
		Bias	RMSE	Bias	RMSE
K_1	30	0.08	0.32	0.01	0.05
K_2	25	0.26	0.44	0.01	0.07
K_3	31	0.03	0.3	0.02	0.05
K_4	27	0.18	0.38	0	0.07
K_5	19	0.39	0.53	-0.03	0.12
K_6	25	0.3	0.47	-0.14	0.18
K_7	14	0.31	0.44	0.07	0.12
K_8	21	0.47	0.63	0.1	0.16
H_1	0.976	-0.02	0.1	0	0.01
H_2	0.958	0	0.1	0	0.01
H_3	0.976	-0.02	0.09	0	0
H_4	0.962	-0.01	0.09	0	0.01
H_5	0.934	0.02	0.1	0.01	0.01
H_6	0.915	0.05	0.11	0.02	0.02
H_7	0.953	0	0.1	0	0.01
H_8	0.946	0.01	0.1	0	0.01
S_1	200	1.63	2.35	-0.04	0.12
S_2	170	1.79	2.61	-0.06	0.14
S_3	204	1.42	2.13	-0.05	0.12
S_4	200	1.61	2.32	-0.11	0.17
S_5	164	2.02	2.83	-0.06	0.16
S_6	151	2.53	3.39	0	0.16
S_7	149	1.78	2.69	-0.08	0.17
S_8	193	1.26	2.03	-0.2	0.25
Pi_1	32.3	2.76	4.09	0.12	0.22
Pi_2	34.9	2.24	3.55	-0.02	0.21
Pi_3	30.9	2.67	4.08	0.14	0.26
Pi_4	33.2	2.63	3.92	0.11	0.22
Pi_5	35.9	2.36	3.58	-0.02	0.15
Pi_6	29.7	3.1	4.49	0.06	0.24
Pi_7	38.6	1.97	3.22	-0.07	0.19
Pi_8	31.4	2.34	3.79	0.06	0.24

Abbrev: haplotypes (K), haplotype diversity (H), polymorphic positions (S), pairwise differences (Pi), the numbers 1-8 indicates tribal codes by alphabetical order, AMI-TSO, respectively. For example, K_4 is number of haplotypes observed in the Paiwan and it is 27.

Table S6. Summary statistics and mismatch analysis of the 12 Taiwanese groups

							Mismatch analysis							
							Demographic expansion model				Spatial expansion model			
	n	**S**	**h**	**hd**	**nd**	**π**	**τ**	**θ$_0$**	**θ$_1$**	**P(SSD)**	**τ**	**θ**	**M**	**P(SSD)**
Highlanders														
Saisiat	24	149	14	0.9529	0.002333	38.65	50.26	0.48	134.65	0.04	43.75	7.95	10.76	**0.27**
Atayal	50	170	25	0.9576	0.002108	34.93	45.54	1.75	102.70	0.12	39.33	11.26	7.49	**0.42**
Tsou	48	193	21	0.9459	0.001894	31.37	30.33	5.65	174.02	0.01	29.43	6.61	10.40	**0.55**
Bunun	50	204	31	0.9763	0.001866	30.92	33.32	3.35	147.60	0.11	31.05	6.52	18.06	**0.45**
Puyuma	39	164	19	0.9339	0.002167	35.91	49.42	0.00	103.64	0.07	43.37	6.98	8.25	**0.46**
Rukai	50	151	25	0.9151	0.001794	29.73	45.56	0.00	113.32	0.00	40.31	6.38	3.45	**0.32**
Paiwan	50	200	27	0.9616	0.002006	33.23	35.32	3.54	159.65	0.03	32.80	6.63	11.65	**0.60**
Ami	50	200	30	0.9763	0.001947	32.26	38.07	0.01	132.10	0.43	33.52	4.97	24.11	**0.51**
Outgroups														
Makatao	50	379	41	0.9910	0.002336	38.70	38.23	4.91	246.99	**0.39**	35.88	7.97	94.36	0.27
Tao	44	127	11	0.7801	0.001617	26.79	34.79	0.00	50.46	0.00	33.04	0.64	2.82	**0.27**
Hakka	45	372	39	0.9939	0.002128	35.26	34.07	4.25	239.35	**0.52**	32.25	6.42	129.92	0.32
Minnan	50	398	48	0.9984	0.002183	36.17	33.62	5.43	294.28	**0.75**	32.05	7.41	430.10	0.41
Combined														
Highlanders†	361	512	164	0.9905	0.002253	37.32	34.65	7.84	246.46	0.20	32.70	10.39	79.04	**0.23**
Han‡	95	584	86	0.9978	0.002156	35.72	33.76	5.06	255.99	**0.59**	31.98	7.27	308.82	0.25

Mismatch analysis, based on 10,000 coalescent simulations in Arlequin 3.5.1.3[6,7]; boldface indicates higher model probability

† Highlanders, 8 highland Formosan tribes combined

‡ Han, Hakka and Minnan combined

Abbrev: n, sample size, S, polymorphic sites, h, haplotypes, hd, haplotype diversity, nd, nucleotide diversity, π, observed mean number of pairwise differences, τ, tau value, θ, theta, M, migration, P(SSD), probability of simulated exceeds observed sum of squares differences

Table S7. Analysis of molecular variance of the fit between genetic and linguistic models

Linguistic model	Source of variation	% Variance	*P*-value*
Blust†			
	Among groups	13.29	<0.001
	Within populations	86.71	
Li			
	Among groups	6.36	0.030
	Among population within groups	7.66	
	Within populations	85.98	
Ross			
	Among groups	3.59	0.128
	Among population within groups	10.78	
	Within populations	85.63	
Sagart			
	Among groups	4.47	0.037
	Among population within groups	9.83	
	Within populations	85.70	

* Φ_{ct}, variance among groups, after 10,000 permutations in Arlequin 3.5.1.3
† No population within groups as each is a linguistic branch

Table S8. Haplogroup frequencies of the 12 Taiwanese groups

(a) Highland Formosan

	Ami	Atayal	Bunun	Paiwan	Puyuma	Rukai	Saisiat	Tsou
A5b1			2	2				
B4a1a	10			8		4		8.3
B4a1a3	12							
B4a2a		4		24		4	4.2	
B4b1a2	8	2	34					10.4
B5a2a	2		12	8	2.6	8	16.7	2.1
C7a				4				
D4a								2.1
D4i		2						
D5a2						2		
D5b			2	12	2.6	48		
D5b1c1	8				2.6			
D6a2		4						
E1a	10							
E1a1		2						
E1a1a	4	6	4		5.1		12.5	4.2
E1a1a2					20.5			
E2b1	2	8	2	2	7.7		4.2	2.1
F1a1d			4	2		2		8.3
F1a3						2		
F1a3a			4		17.9			8.3
F1a4a1	2		2					10.4
F2				2				
F3b1	2		8	12	28.2	20		4.2
F4b	2	22	18				12.5	2.1
M7b1'2'4'5	22							
M7b3a		40		10			37.5	
M7c'e'f	4							
M7c1								10.4
M7c3a	8		4		2.6			
M7c3b				2				
M7c3c	2		2	12	7.7	6		
M8a2	2							
N9a10a		6						
Q1							8.3	
R9b1a2								4.2
R9c1a			2		2.6	4		22.9
Y2a		4					4.2	

(b) Lowland Formosan and other Taiwanese groups

	Makatao	Tao	Hakka	Minnan
A4e1	2			
A5b1			2.2	
B4	2			4
B4a			4.4	
B4a1	2			
B4a1a			2.2	
B4a1c			6.7	4
B4a2a	2	15.9	4.4	
B4b1a			2.2	
B4b1a2	6			
B4b1a2a	2			
B4b1c			2.2	
B4c1b2				2
B4c1b2a2		36.4	2.2	
B4c1b2c			2.2	2
B4c2				2
B4g		2.3		
B4h	2			
B5a			4.4	
B5a1a				2
B5a1c				6
B5a1c1			2.2	
B5a2a	4		2.2	
B5b			2.2	
B5b2a				2
B5b2c	2			
C7			2.2	
C7a	2		2.2	2
C7a1	2			
D4			2.2	
D4a			6.7	6
D4a1	2			
D4a1e				2
D4a3b2				2
D4b2b			2.2	
D4b2b2b			2.2	
D4j				2
D5a2a1b			2.2	2
D5b	2			4
D5b1b				2
D5b1c1				2
D5c1				2
D6c				2
E1a1a	2			
E1a1a2	6		2.2	

	Makatao	Tao	Hakka	Minnan
E2b1	6	2.3		
F1a1a			2.2	
F1a1c				2
F1a1d		34.1		
F1a3	2			
F1a3a	2			
F1a4a1	4			
F1c1a				2
F1d				2
F1g1				2
F2	2			
F2a	2			
F2b1			2.2	
F3a1			2.2	
F3b1	4			
F4b	2			
G1a1	2			
G2b1a			4.4	
G3a1'2				2
M10a1a2			2.2	
M10a2	4			
M11b				2
M7b1'2'4'5'	4		2.2	6
M7b3a	8	4.5		
M7b4	2			2
M7c1			2.2	4
M7c3				2
M7c3a			6.7	
M7c3b				2
M7c3c	2			
M7c3c1		4.5		
M7e	2			2
M8a2			2.2	
M9a			2.2	
M9a1a				4
N11a	2			
N9a1			2.2	
N9a10				2
N9a2'4'5			4.4	2
N9a4				2
R9b1a2			2.2	4
R9c1	2			
R9c1a	8			
Y1				2
Z3				2

References:

1. Trejaut, J.A., Kivisild, T., Loo, J.H., Lee, C.L., He, C.L., Hsu, C.J., Lee, Z.Y., and Lin, M. (2005). Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. PLoS Biol 3, e247.
2. Blust, R. (1999). Selected papers from the Eighth International Conference on Austronesian Linguistics. In Symposium series of the Institute of Linguistics, Academia Sinica, Taipei, Taiwan, E. Zeitoun and P. Li, eds., pp 31-94.
3. Li, P.J.K. (2008). Time perspective of Formosan Aborigines. In Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics R.B. A Sanchez-Mazas, M Ross, I Peiros, M Lin, ed. (Routledge), pp 211-218.
4. Ross, M. (2009). Proto Austronesian verbal morphology: A reappraisal. In Austronesian historical linguistics and culture history: a festschrift for Robert Blust, A.P. A Adelaar, ed. (Canberra: Pacific Linguistics), pp 295-326.
5. Sagart, L. (2008). The expansion of Setaria farmers in East Asia: A linguistic and archaeological model. In Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics R.B. A Sanchez-Mazas, M Ross, I Peiros, M Lin, ed. (Routledge), pp 133-167.
6. Excoffier, L., and Lischer, H.E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10, 564-567.
7. Excoffier, L. (2004). Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Mol Ecol 13, 853-864.