**Supplementary Figure 1.** PCA plots generated with EIGENSOFT (42) for AN-speaking groups from (A) the Philippines, (B) eastern Indonesia, and (C) western ISEA, along with reference populations. The circled groupings indicate subsets of populations consistent with simple histories according to our f_4-based test: (A) Agta, Ati, Ayta, Ilocano, Iraya, and Manobo (one wave of admixture), (B) Alorese, Kambera, Lamaholot, and Lembata (one wave), and (C) Bidayuh, Dayak, Mentawai, Javanese Jakarta, Javanese Java, and Sunda (two waves).
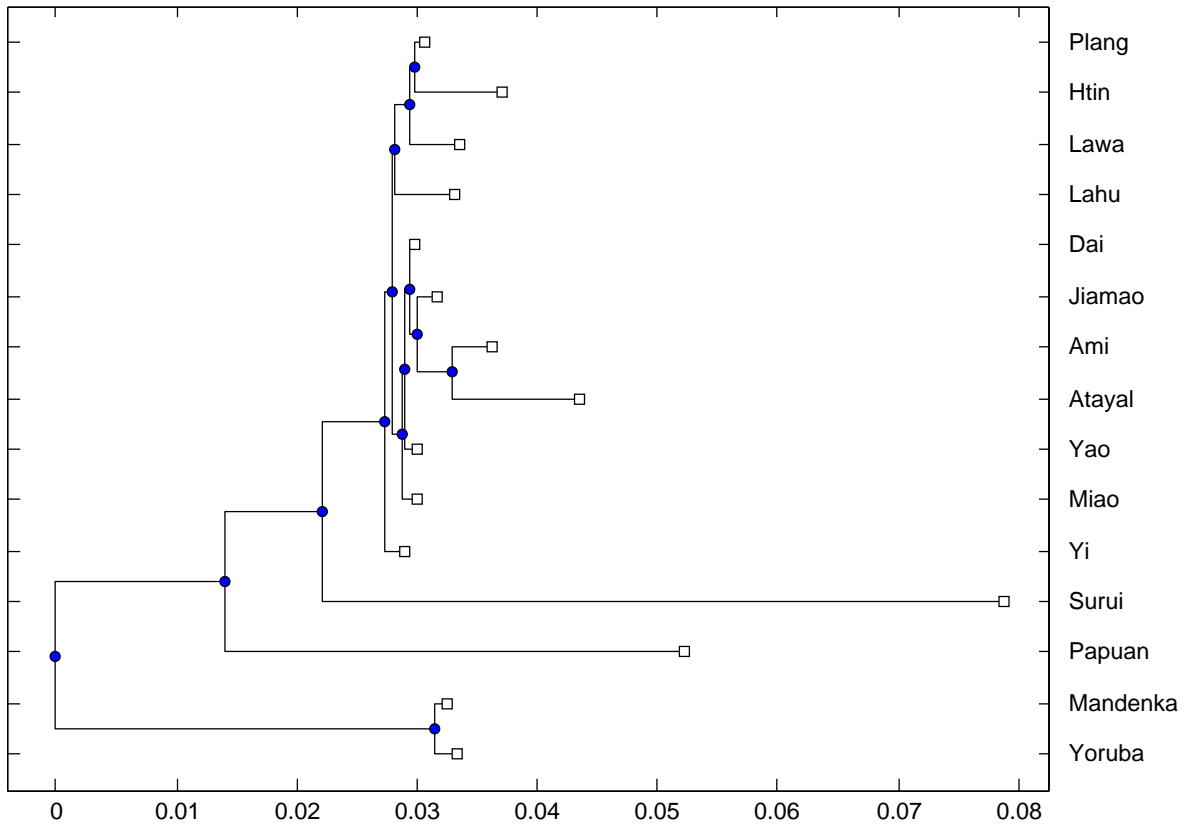
Supplementary Figure 2. Alternative 15-population scaffold tree. See Supplementary Tables 6 and 7 for full *MixMapper* results from fitting admixed Austronesian-speaking populations using this scaffold. Distances are in F_2 units.
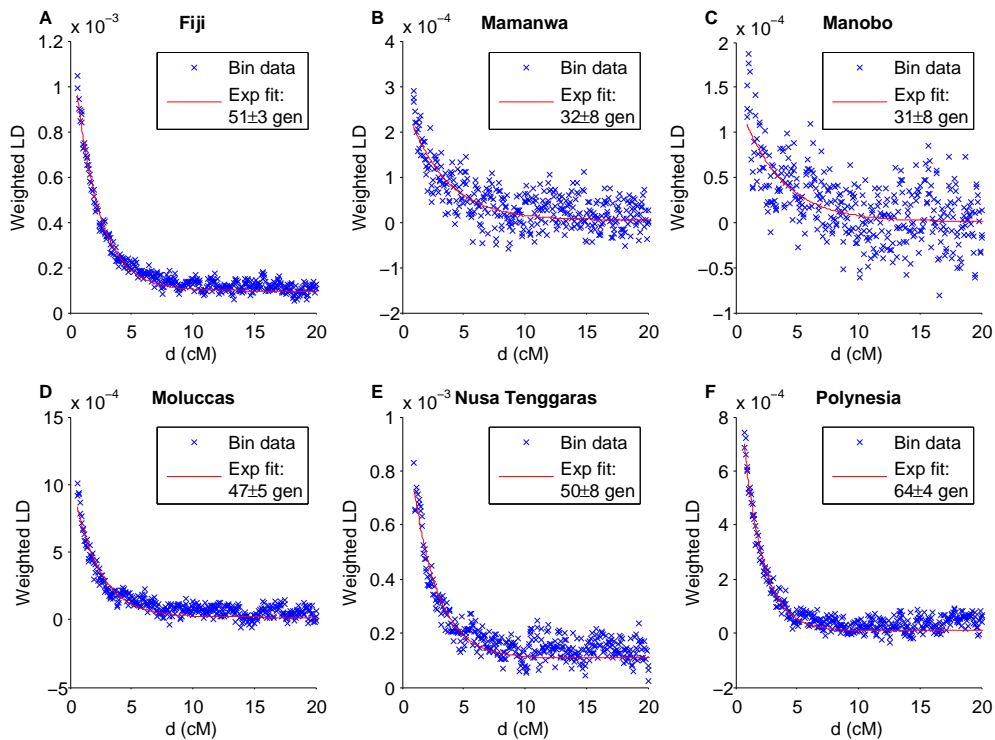
Supplementary Figure 3. Weighted LD curves and estimated dates of admixture for (A) Fiji, (B) Mamanwa, (C) Manobo, (D) Moluccas, (E) Nusa Tenggaras, and (F) Polynesia, obtained using *ALDER* (35) with Papuan and Taiwanese reference populations. Admixture dates are inferred as time constants of the exponential decay of weighted covariance with genetic distance. LD analysis requires a higher SNP density than is available with our full data set, so these inferences are restricted to samples from ref. (27). We note that our dates are much more recent than those reported in ref. (24); we hypothesize that the initial admixtures were followed by more recent mixing between groups with different proportions of Taiwan-related ancestry, in which case the date from *ALDER* is an intermediate one over the entire process. This would be consistent with the fact that the curves appear to have some deviations from a pure exponential decay shape.

4
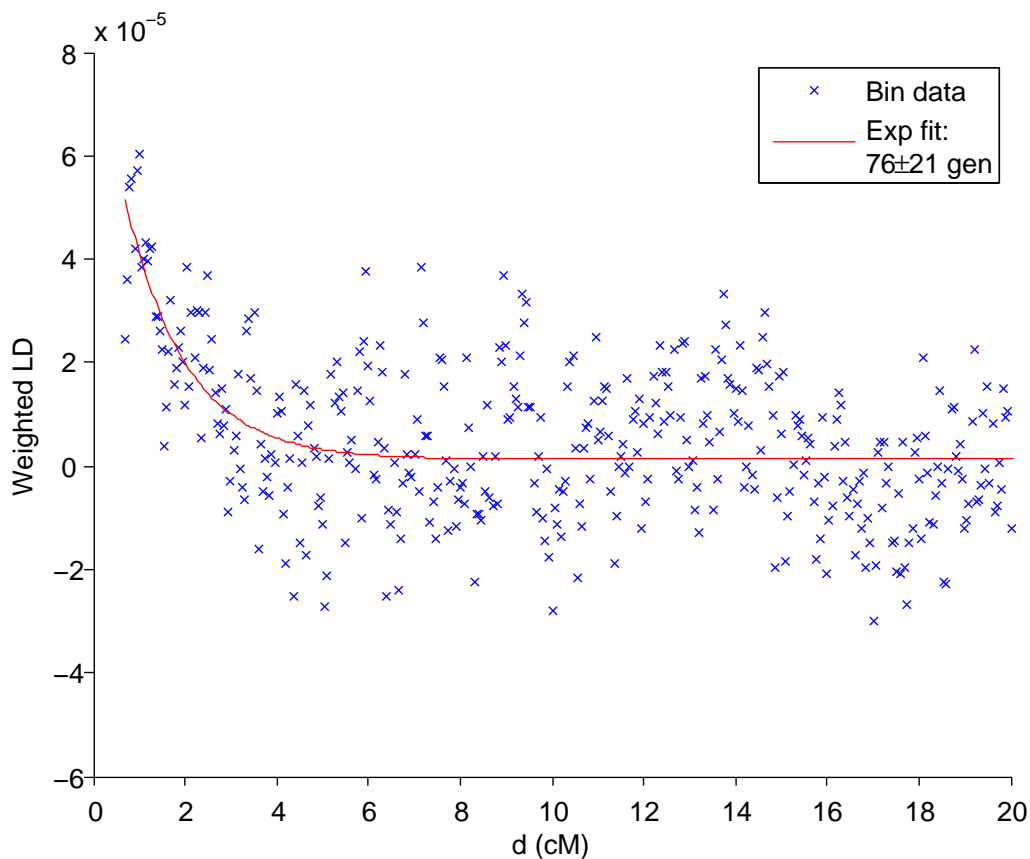
Supplementary Figure 4. Weighted LD curve and estimated date of admixture for western ISEA, obtained using *ALDER* (35) with Papuan and CHB (HapMap Chinese from Beijing (44)) reference populations. The admixture date is inferred as the time constant of the exponential decay of weighted covariance with genetic distance. LD analysis requires a higher SNP density than is available with our full data set, so these inferences are restricted to samples from ref. (27). In order to enhance the signal-to-noise ratio, we pooled samples from four populations, two each from Borneo (Bidayuh and Dayak) and Sumatra (Besemah and Semende), into a single test set, under the assumption that all four have similar admixture histories.

Supplementary Table 1. Summary of populations used in this study

Population	Country	Data set	Pan-Asia ID	# samples	Model status
Ami	Taiwan	Pan-Asia	AX-AM	10	Scaffold
Atayal	Taiwan	Pan-Asia	AX-AT	10	Scaffold
Miao	China	HGDP		10	Scaffold
She	China	HGDP		10	Scaffold
Jiamao	China	Pan-Asia	CN-JI	31	Scaffold
Lahu	China	HGDP		8	Scaffold
Wa	China	Pan-Asia	CN-WA	50	Scaffold
Yi	China	HGDP		10	Scaffold
Naxi	China	HGDP		8	Scaffold
Hmong	Thailand	Pan-Asia	TN-HM	20	Scaffold
Plang	Thailand	Pan-Asia	TH-PP	18	Scaffold
H'tin	Thailand	Pan-Asia	TH-TN	15	Scaffold
Palaung	Thailand	Pan-Asia	TH-PL	18	Scaffold
Karitiana	Brazil	HGDP		14	Scaffold
Suruí	Brazil	HGDP		8	Scaffold
Papuan	Papua New Guinea	HGDP		17	Scaffold
Mandenka	Senegal	HGDP		22	Scaffold
Yoruba	Nigeria	HGDP		21	Scaffold
Aboriginal Taiwanese	Taiwan	Reich et al. (2011)		10	*ALDER* reference
CHB	China	HapMap Phase 3 (44)		88	*ALDER* reference
Papuan[1]	Papua New Guinea	Reich et al. (2011)		24	*ALDER* reference
Agta	Philippines	Pan-Asia	PI-AG	8	Three-way admixed
Ati	Philippines	Pan-Asia	PI-AT	23	Three-way admixed
Ayta	Philippines	Pan-Asia	PI-AE	8	Two-way admixed
Iraya	Philippines	Pan-Asia	PI-IR	9	Two-way admixed
Mamanwa	Philippines	Pan-Asia	PI-MW	17	Two-way admixed
Mamanwa[1]	Philippines	Reich et al. (2011)		11	Two-way admixed
Manobo	Philippines	Pan-Asia	PI-MA	18	Two-way admixed
Manobo[1]	Philippines	Reich et al. (2011)		16	Two-way admixed
Tagalog	Philippines	Pan-Asia	PI-UN	19	Two-way admixed
Visaya	Philippines	Pan-Asia	PI-UI	20	Three-way admixed
Alorese	Indonesia	Pan-Asia	ID-AL	19	Two-way admixed
Kambera	Indonesia	Pan-Asia	ID-SB	20	Two-way admixed
Lamaholot	Indonesia	Pan-Asia	ID-LA	20	Three-way admixed
Lembata	Indonesia	Pan-Asia	ID-LE	19	Three-way admixed
Manggarai Ngada	Indonesia	Pan-Asia	ID-SO	19	Three-way admixed
Manggarai Rampasasa	Indonesia	Pan-Asia	ID-RA	16	Three-way admixed
Fiji	Fiji	Reich et al. (2011)		25	Two-way admixed
Polynesia	Multiple[2]	Reich et al. (2011)		19	Two-way admixed
Toraja	Indonesia	Pan-Asia	ID-TR	20	Three-way admixed
Moluccas	Indonesia	Reich et al. (2011)		10	Two-way admixed
Nusa Tenggaras	Indonesia	Reich et al. (2011)		10	Two-way admixed
Batak Toba	Indonesia	Pan-Asia	ID-TB	20	Three-way admixed
Bidayuh	Malaysia	Pan-Asia	MY-BD	47	Three-way admixed
Bidayuh[1]	Malaysia	Reich et al. (2011)		10	Three-way admixed
Dayak	Indonesia	Pan-Asia	ID-DY	12	Three-way admixed
Dayak[1]	Indonesia	Reich et al. (2011)		16	Three-way admixed
Javanese Jakarta	Indonesia	Pan-Asia	ID-JA	34	Three-way admixed
Javanese Java	Indonesia	Pan-Asia	ID-JV	19	Three-way admixed
Malay Indonesia	Indonesia	Pan-Asia	ID-ML	12	Three-way admixed
Malay Singapore	Singapore	Pan-Asia	SG-MY	28	Three-way admixed
Sunda	Indonesia	Pan-Asia	ID-SU	25	Three-way admixed
Besemah	Indonesia	Reich et al. (2011)		8	Three-way admixed
Semende	Indonesia	Reich et al. (2011)		9	Three-way admixed
Batak Karo	Indonesia	Pan-Asia	ID-KR	17	Uncertain admixed
Malay	Malaysia	Pan-Asia	MY-KN	18	Uncertain admixed
Malay Minangkabau	Malaysia	Pan-Asia	MY-MN	19	Uncertain admixed
Mentawai	Indonesia	Pan-Asia	ID-MT	15	Uncertain admixed
Ilocano	Philippines	Pan-Asia	PI-UB	20	Uncertain admixed
Temuan	Malaysia	Pan-Asia	MY-TM	37	Uncertain admixed
Melanesian	Papua New Guinea	HGDP		10	Uncertain admixed
Jehai	Malaysia	Pan-Asia	MY-JH	42	Two-way admixed
Kensiu	Malaysia	Pan-Asia	MY-KS	25	Two-way admixed
Zhuang	China	Pan-Asia	CN-CC	24	Other mainland
Jinuo	China	Pan-Asia	CN-JN	29	Other mainland
Han Cantonese	China	Pan-Asia	CN-GA	28	Other mainland
Hmong	China	Pan-Asia	CN-HM	20	Other mainland
Tai Lue	Thailand	Pan-Asia	TH-TL	18	Other mainland
Tai Yuan	Thailand	Pan-Asia	TH-TU	20	Other mainland

[Caption on next page.]

Summary of population samples used in this study. The first group of populations are references used in the 18-population scaffold tree and for admixture date estimation, the second group are Austronesian-speaking populations fit as admixtures, and the third group are other populations used for comparison.

[1]Samples used for admixture date inference with *ALDER* were taken from Reich et al. (2011) rather than from Pan-Asia or HGDP for the main *MixMapper* analysis.

[2]The Polynesian samples are from the Cook Islands (2), Futuna (4), Niue (1), Samoa (5), Tokelau (2), Tonga (2), and Tuvalu (3).

Supplementary Table 2. Inferred mixture parameters for two-way admixed populations with alternative SNP ascertainment

Philippine admixed population	Mixing branch 1 bootstrap distribution		Mixing branch 2 bootstrap distribution		Branch 1 ancestry (Austronesian)
Agta	(Ami,Atayal) Ami Atayal	51% 38% 11%	Papuan	62%	51–66%
Ati	(Ami,Atayal)	93%	Papuan	100%	53–68%
Ayta	(Ami,Atayal) Ami Atayal	31% 17% 48%	Papuan	89%	23–45%
Iraya	(Ami,Atayal) Ami Atayal	29% 59% 12%	Papuan Papuan opp. African	35% 6%	60–86%
Mamanwa	(Ami,Atayal) Ami Atayal	41% 42% 17%	Papuan	100%	49–66%
Manobo	(Ami,Atayal) Ami	33% 66%	Papuan	100%	77–87%
Tagalog	(Ami,Atayal)	98%	Papuan	78%	85–93%
Visaya	(Ami,Atayal) Ami Atayal	87% 5% 8%	Papuan	99%	82–91%
E. Indonesian / Oceanian admixed population	**Mixing branch 1 bootstrap distribution**		**Mixing branch 2 bootstrap distribution**		**Branch 1 ancestry (Austronesian)**
Alorese	(Ami,Atayal) Atayal	72% 20%	Papuan	100%	38–47%
Kambera	(Ami,Atayal)	95%	Papuan	100%	65–75%
Lamaholot	(Ami,Atayal)	93%	Papuan	100%	51–62%
Lembata	(Ami,Atayal) Ami Atayal	55% 15% 30%	Papuan	100%	48–57%

Sources of ancestry and mixture proportions (95% confidence intervals) from *MixMapper* for two-way admixed populations, using SNPs selected by merging the Pan-Asia data with HGDP samples typed on the Affymetrix Human Origins array (30). "Papuan opp. African" refers to the common ancestral branch of all populations in the scaffold other than Papuan and Africans, while (Ami, Atayal) designates the common ancestral branch of Ami and Atayal (see Fig. 1). Branch topologies are shown that occur for at least 5% of 500 bootstrap replicates. The results are very similar to those obtained with the original scaffold (see Supplementary Table 10).

Supplementary Table 3. Inferred mixture parameters for three-way admixed populations with alternative SNP ascertainment

E. Indonesian / Oceanian admixed population	Percent bootstrap reps with Branch 3 = H'tin	Branch 3 ancestry (Austro-Asiatic)	Branch 1 ancestry (Austronesian)
Manggarai Ngada	66%	20–31%	29–42%
Manggarai Rampasasa	27%	29–38%	33–44%
Toraja	85%	6–14%	70–79%
W. Indonesian admixed population	**Percent bootstrap reps with Branch 3 = H'tin**	**Branch 3 ancestry (Austro-Asiatic)**	**Branch 1 ancestry (Austronesian)**
Batak Toba	28%	19–35%	49–60%
Bidayuh	99%	42–58%	36–50%
Dayak	98%	27–44%	46–59%
Javanese Jakarta	100%	49–64%	28–40%
Javanese Java	100%	52–70%	24–38%
Malay Indonesia	76%	18–33%	58–73%
Malay Singapore	74%	29–49%	35–51%
Sunda	100%	50–65%	27–41%

Mixture parameters from *MixMapper* for three-way admixed populations, using SNPs selected by merging the Pan-Asia data with HGDP samples typed on the Affymetrix Human Origins array (30). Mixture proportions shown are 95% confidence intervals for re-optimized values (see Methods), using the bootstrap replicates (percentages given, out of 500) assigning the third ancestry component to the H'tin branch. The results are very similar to those obtained with the original scaffold (see Supplementary Table 11), with slightly lower but still substantial bootstrap support for the H'tin-related ancestry component.

Supplementary Table 4. Populations with negative f_3 statistics

Test population C	Reference population A	Reference population B	$f_3(C; A, B)$	Std error	Z-score
Alorese	Tagalog	Papuan	-0.0106	0.00023	-45.36
Batak Karo	Mentawai	Kalash	-0.00179	0.00025	-7.21
Batak Toba	Mentawai	Tuscan	-0.00238	0.00024	-9.96
Han Cantonese	Korean	Jiamao	-0.00079	0.00007	-12.04
Hmong China	Hmong Thailand	Mbuti Pygmy	-0.00132	0.0002	-6.55
Ilocano	Ami	Bengali	-0.00098	0.00024	-4.14
Javanese Jakarta	Ilocano	Jehai	-0.00113	0.00015	-7.43
Javanese Java	Ami	Jehai	-0.00133	0.00017	-7.69
Kambera	Tagalog	Papuan	-0.00719	0.00025	-29.26
Lamaholot	Toraja	Papuan	-0.0091	0.00022	-41.2
Lembata	Toraja	Papuan	-0.00961	0.00022	-43.26
Malay	Zhuang	GIH	-0.00322	0.00013	-24.5
Malay Indonesia	Ami	Bengali	-0.00201	0.00028	-7.19
Malay Minangkabau	Ami	Hindi Haryana	-0.00262	0.00026	-9.97
Malay Singapore	Hindi Haryana	Jiamao	-0.00209	0.00011	-18.66
Manggarai Ngada	Tagalog	Papuan	-0.00883	0.00025	-35.49
Manggarai Rampasasa	Ilocano	Papuan	-0.00682	0.00029	-23.88
Manobo	Ami	Papuan	-0.0006	0.00035	-1.7
Miao	Hmong Thailand	Colombian	-0.0004	0.00028	-1.41
Plang	Mlabri	Han-NChina	-0.00021	0.00027	-0.78
Sunda	Ilocano	Jehai	-0.00113	0.00014	-8.15
Tagalog	Ami	Hindi Rajasthan	-0.00214	0.00019	-11.11
Tai Yuan	Htin	CHB	-0.00082	0.00008	-9.76
Toraja	Ilocano	Papuan	-0.00213	0.00026	-8.18
Visaya	Ami	Hindi Rajasthan	-0.00298	0.0002	-14.79
Wa	Mlabri	Naxi	-0.00005	0.00028	-0.16
Yi	Mlabri	Naxi	-0.00006	0.00034	-0.18
Zhuang	Jiamao	Lahu	-0.0002	0.0001	-1.91

Asian populations from Supplementary Table 1 having at least one negative f_3 value. For each test population C, we show the two reference populations A and B in the data set giving the lowest Z-score for $f_3(C; A, B)$. We note that all populations on this list that are used in the scaffold have $Z > -2$, which indicates a non-significant result (especially given the presence of many hypotheses). While a significantly negative f_3 value demonstrates that the test population must be admixed, a lack of a negative value does not prove a lack of admixture.

Supplementary Table 5. Populations with no negative f_3 statistics

Agta
Ami
Atayal
Ati
Ayta
Bidayuh
Dayak
Hmong Thailand
Htin
Iraya
Jehai
Jiamao
Jinuo
Kensiu
Lahu
Mamanwa
Mentawai
Mlabri
Naxi
Paluang
She
Tai Lue
Temuan

Asian populations from Supplementary Table 1 having no negative f_3 value for any pair of reference populations in the data set. While a significantly negative f_3 value demonstrates that the test population must be admixed, a lack of a negative value does not prove a lack of admixture.

Supplementary Table 6. Inferred mixture parameters for two-way admixed populations on a 15-population alternative scaffold

Philippine admixed population	Mixing branch 1 bootstrap distribution		Mixing branch 2 bootstrap distribution		Branch 1 ancestry (Austronesian)
Agta	(Ami,Atayal) Ami	44% 56%	Papuan	100%	50–60%
Ati	(Ami,Atayal) Ami	12% 88%	Papuan	100%	49–58%
Ayta	(Ami,Atayal) Ami Atayal	21% 7% 72%	Papuan	100%	24–37%
Iraya	(Ami,Atayal) Ami	16% 84%	Papuan Papuan opp. African	39% 60%	56–78%
Mamanwa	(Ami,Atayal) Ami Atayal	40% 53% 7%	Papuan	100%	51–61%
Manobo	(Ami,Atayal) Ami	9% 91%	Papuan	100%	78–83%
Tagalog	(Ami,Atayal)	100%	Papuan Papuan opp. African	64% 34%	83–92%
Visaya	(Ami,Atayal) Ami	82% 18%	Papuan Papuan opp. African	78% 22%	72–85%
E. Indonesian / Oceanian admixed population	**Mixing branch 1 bootstrap distribution**		**Mixing branch 2 bootstrap distribution**		**Branch 1 ancestry (Austronesian)**
Alorese	(Ami,Atayal) Ami	84% 14%	Papuan	100%	37–43%
Fiji	(Ami,Atayal) Ami Atayal	16% 66% 18%	Papuan	100%	30–40%
Kambera	(Ami,Atayal)	100%	Papuan	100%	68–72%
Lamaholot	(Ami,Atayal)	96%	Papuan	100%	49–56%
Lembata	(Ami,Atayal)	98%	Papuan	100%	47–53%
Polynesia	(Ami,Atayal) Ami Atayal	23% 52% 25%	Papuan	100%	61–72%

Sources of ancestry and mixture proportions (95% confidence intervals) from *MixMapper* for two-way admixed populations using a 15-population alternative scaffold tree. The results are very similar to those obtained with the original scaffold (see Supplementary Table 10). "Papuan opp. African" refers to the common ancestral branch of all populations in the scaffold other than Papuan and Africans, while (Ami, Atayal) designates the common ancestral branch of Ami and Atayal (see Fig. 1). Branch topologies are shown that occur for at least 5% of 500 bootstrap replicates.

Supplementary Table 7. Inferred mixture parameters for three-way admixed populations on a 15-population alternative scaffold

E. Indonesian / Oceanian admixed population	Percent bootstrap reps with Branch 3 = H'tin	Branch 3 ancestry (Austro-Asiatic)	Branch 1 ancestry (Austronesian)
Manggarai Ngada	83%	24–30%	30–36%
Manggarai Rampasasa	81%	35–43%	28–36%
Toraja	90%	7–17%	68–77%
W. Indonesian admixed population	**Percent bootstrap reps with Branch 3 = H'tin**	**Branch 3 ancestry (Austro-Asiatic)**	**Branch 1 ancestry (Austronesian)**
Batak Toba	52%	23–33%	49–57%
Bidayuh	100%	52–62%	33–43%
Dayak	100%	35–44%	46–56%
Javanese Jakarta	100%	59–66%	27–33%
Javanese Java	100%	60–69%	25–33%
Malay Indonesia	87%	26–36%	54–65%
Malay Singapore	68%	40–47%	35–42%
Sunda	100%	58–65%	26–33%

Mixture parameters from *MixMapper* for three-way admixed populations using a 15-population alternative scaffold tree. The results are very similar to those obtained with the original scaffold (see Supplementary Table 11), with slightly lower but still substantial bootstrap support for the H'tin-related ancestry component. Mixture proportions shown are 95% confidence intervals for re-optimized values (see Methods), using the bootstrap replicates (percentages given, out of 500) assigning the third ancestry component to the H'tin branch.

Supplementary Table 8. Consistency of mixture parameters for two-way admixed populations on 17-population alternative scaffolds

Philippine admixed population	Taiwan bootstrap support	Papuan bootstrap support	Taiwan ancestry fraction
Agta	100 ± 0%	99 ± 0%	56 ± 1%
Ati	100 ± 0%	100 ± 0%	55 ± 0%
Ayta	99 ± 1%	100 ± 0%	32 ± 1%
Iraya	100 ± 0%	79 ± 8%	73 ± 2%
Mamanwa	100 ± 0%	100 ± 0%	56 ± 0%
Manobo	100 ± 0%	100 ± 0%	81 ± 0%
Tagalog	100 ± 0%	69 ± 11%	89 ± 0%
Visaya	100 ± 0%	83 ± 6%	83 ± 0%
E. Indonesian / Oceanian admixed population	**Taiwan bootstrap support**	**Papuan bootstrap support**	**Taiwan ancestry fraction**
Alorese	100 ± 0%	100 ± 0%	40 ± 0%
Fiji	100 ± 1%	100 ± 0%	36 ± 0%
Kambera	100 ± 0%	100 ± 0%	70 ± 1%
Lamaholot	100 ± 0%	100 ± 0%	53 ± 1%
Lembata	100 ± 0%	100 ± 0%	50 ± 1%
Polynesia	100 ± 0%	100 ± 0%	66 ± 0%

Sources of ancestry and mixture proportions (95% confidence intervals) from *MixMapper* for two-way admixed populations, removing one population at a time (other than Papuan) from the 18-population scaffold tree (Fig. 1). Values are means ± standard errors over the 17 different perturbed scaffolds. Austronesian ancestry refers to splits from the Ami and Atayal branches and their common ancestor, while Papuan support only includes splits from the Papuan branch. The results are very similar to those obtained with the original scaffold (Supplementary Table 10). Note that the branch support values are over 100 replicates, while the mixture proportions are point-estimates using all data rather than bootstraps.

Supplementary Table 9. Consistency of mixture parameters for three-way admixed populations on 17-population alternative scaffolds

E. Indonesian / Oceanian admixed population	Percent bootstrap reps with Branch 3 = H'tin, Plang, Wa	Branch 3 ancestry (Austro-Asiatic)	Branch 1 ancestry (Austronesian)
Manggarai Ngada	100 ± 1%	26 ± 2%	34 ± 2%
Manggarai Rampasasa	96 ± 15%	37 ± 2%	34 ± 2%
Toraja	99 ± 4%	12 ± 1%	72 ± 1%
W. Indonesian admixed population	**Percent bootstrap reps with Branch 3 = H'tin, Plang, Wa**	**Branch 3 ancestry (Austro-Asiatic)**	**Branch 1 ancestry (Austronesian)**
Batak Toba	93 ± 12%	27 ± 2%	53 ± 2%
Bidayuh	100 ± 0%	54 ± 2%	40 ± 1%
Dayak	100 ± 0%	39 ± 2%	52 ± 1%
Javanese Jakarta	100 ± 0%	59 ± 2%	32 ± 2%
Javanese Java	100 ± 0%	60 ± 2%	31 ± 2%
Malay Indonesia	97 ± 9%	31 ± 1%	60 ± 1%
Malay Singapore	98 ± 7%	41 ± 3%	40 ± 2%
Sunda	100 ± 0%	57 ± 2%	33 ± 1%

Mixture parameters from *MixMapper* for three-way admixed populations, removing one population at a time (other than Papuan) from the 18-population scaffold tree (Fig. 1). Values are means ± standard errors over the 17 different perturbed scaffolds. The results are very similar to those obtained with the original scaffold (see Supplementary Table 11). Mixture proportions shown are re-optimized values (see Methods), using the 17-population trees in which the third ancestry component is Austro-Asiatic (H'tin, Plang, or Wa), which were 16 of 17 for Batak Toba and Manggarai Rampasasa and all 17 trees for the other populations. Note that the branch support values are over 100 replicates, while the mixture proportions are point-estimates using all data rather than bootstraps.

Supplementary Table 10. Inferred mixture parameters for two-way admixed populations

Philippine admixed population	Mixing branch 1 bootstrap distribution		Mixing branch 2 bootstrap distribution		Branch 1 ancestry (Austronesian)
Agta	(Ami,Atayal) Ami	44% 56%	Papuan	100%	51–62%
Ati	(Ami,Atayal) Ami	15% 85%	Papuan	100%	50–59%
Ayta	(Ami,Atayal) Ami Atayal	20% 7% 73%	Papuan	100%	25–38%
Iraya	(Ami,Atayal) Ami	28% 72%	Papuan Papuan opp. African	76% 20%	61–80%
Mamanwa	(Ami,Atayal) Ami Atayal	25% 62% 13%	Papuan	100%	51–61%
Manobo	(Ami,Atayal) Ami	11% 89%	Papuan	100%	78–83%
Tagalog	(Ami,Atayal)	99%	Papuan Papuan opp. African	71% 28%	83–92%
Visaya	(Ami,Atayal) Ami	88% 11%	Papuan Papuan opp. African	85% 15%	74–85%
E. Indonesian / Oceanian admixed population	**Mixing branch 1 bootstrap distribution**		**Mixing branch 2 bootstrap distribution**		**Branch 1 ancestry (Austronesian)**
Alorese	(Ami,Atayal) Ami Atayal	77% 17% 6%	Papuan	100%	37–44%
Fiji	(Ami,Atayal) Ami Atayal	19% 64% 17%	Papuan	100%	30–41%
Kambera	(Ami,Atayal)	100%	Papuan	100%	67–73%
Lamaholot	(Ami,Atayal) Ami	93% 6%	Papuan	100%	50–56%
Lembata	(Ami,Atayal)	94%	Papuan	100%	47–53%
Polynesia	(Ami,Atayal) Ami Atayal	20% 54% 26%	Papuan	100%	61–72%

Sources of ancestry and mixture proportions (95% confidence intervals) from *MixMapper* for two-way admixed populations. "Papuan opp. African" refers to the common ancestral branch of all populations in the scaffold other than Papuan and Africans, while (Ami, Atayal) designates the common ancestral branch of Ami and Atayal (see Fig. 1). Branch topologies are shown that occur for at least 5% of 500 bootstrap replicates.

Supplementary Table 11. Inferred mixture parameters for three-way admixed populations

E. Indonesian / Oceanian admixed population	Percent bootstrap reps with Branch 3 = H'tin	Branch 3 ancestry (Austro-Asiatic)	Branch 1 ancestry (Austronesian)
Manggarai Ngada	100%	24–29%	31–37%
Manggarai Rampasasa	100%	34–41%	29–37%
Toraja	100%	10–17%	68–75%
W. Indonesian admixed population	**Percent bootstrap reps with Branch 3 = H'tin**	**Branch 3 ancestry (Austro-Asiatic)**	**Branch 1 ancestry (Austronesian)**
Batak Toba	92%	22–32%	50–57%
Bidayuh	100%	50–57%	37–44%
Dayak	100%	35–42%	48–56%
Javanese Jakarta	100%	57–63%	29–35%
Javanese Java	100%	57–64%	28–34%
Malay Indonesia	100%	26–34%	56–64%
Malay Singapore	100%	38–45%	37–43%
Sunda	100%	54–61%	30–36%

Mixture parameters from *MixMapper* for three-way admixed populations. Mixture proportions shown are 95% confidence intervals for re-optimized values (see Methods), using the bootstrap replicates (percentages given, out of 500) assigning the third ancestry component to the H'tin branch.

Supplementary Table 12. Admixture model selection for three-way admixed populations

E. Indonesian / Oceanian admixed population	Residual norm from 2-way fit	Residual norm from 3-way fit	Difference (95% CI)
Manggarai Ngada	27.0	22.7	(-1.4, 9.8)
Manggarai Rampasasa	31.2	25.1	(-1.4, 14.5)
Toraja	11.3	7.9	(-0.8, 7.0)
W. Indonesian admixed population	**Residual norm from 2-way fit**	**Residual norm from 3-way fit**	**Difference (95% CI)**
Batak Toba	22.2	16.5	(-5.5, 15.2)
Bidayuh	23.1	15.5	(-1.6, 16.5)
Dayak	32.8	11.1	(11.4, 28.1)
Javanese Jakarta	34.3	15.3	(12.4, 23.8)
Javanese Java	32.8	15.0	(10.8, 24.0)
Malay Indonesia	18.8	10.1	(0.8, 14.9)
Malay Singapore	38.8	27.0	(0.6, 21.1)
Sunda	39.1	16.8	(15.8, 27.8)

Quality of fit for alternative models for three-way admixed populations. Shown are the median norms of the vectors of residual errors for all pairwise distances $f_2(C, X)$ (see Methods for details), along with 95% confidence intervals for the differences (all multiplied by 10^6). Smaller norms indicate more accurate model fits.

Supplementary Table 13. Two-way mixture fits for East and Mainland Southeast Asian populations

Admixed population	Mixing branch 1 + branch 2	% reps	Branch 1 ancestry
Chinese Singapore	(Ami,Atayal,Jiamao) + Karitiana	56%	98–99%
	(Ami,Atayal,Jiamao) + Naxi	21%	85–93%
	(Ami,Atayal,Jiamao) + Surui	15%	98–100%
Han Hakka	(Ami,Atayal,Jiamao) + Naxi	75%	83–91%
	(Ami,Atayal,Jiamao) + She	9%	58–89%
Han Minnan	(Ami,Atayal,Jiamao) + Naxi	63%	84–91%
	(Ami,Atayal,Jiamao) + Surui	13%	99–99%
	(Ami,Atayal,Jiamao) + Karitiana	13%	99–99%
	(Ami,Atayal,Jiamao) + She	8%	60–88%
Hmong China	Hmong Thailand + Jiamao	40%	71–89%
	Hmong Thailand + (Ami,Atayal,Jiamao)	34%	57–74%
	Hmong Thailand + She	20%	56–80%
Jinuo	(H'tin,Plang,Wa) + Yi	16%	77–91%
	(Naxi,Yi) + Wa	12%	52–80%
	(Karitiana,Mandenka,Naxi,Papuan,Surui,Yi,Yoruba,root) + Wa	11%	65–88%
	(H'tin,Plang,Wa) + (Naxi,Yi)	8%	41–83%
	(H'tin,Plang,Wa) + Hmong Thailand	7%	82–97%
	(H'tin,Plang,Wa) + Papuan	7%	97–99%
	(H'tin,Plang,Wa) + Naxi	6%	74–93%
Karen	(H'tin,Plang,Wa) + Papuan	93%	92–98%
	(H'tin,Plang) + Papuan	7%	90–96%
Lawa	(H'tin,Plang) + Papuan	82%	93–98%
	(H'tin,Plang,Wa) + Papuan	5%	95–98%
	H'tin + Papuan	5%	93–98%
Mlabri	H'tin + Papuan	70%	86–97%
	H'tin + (Mandenka,Yoruba,root)	18%	85–95%
	H'tin + (Mandenka,Yoruba)	9%	92–98%
Mon	(H'tin,Plang,Wa) + (Mandenka,Yoruba,root)	90%	80–86%
Tai Khuen	Jiamao + H'tin	99%	65–75%
Tai Lue	Jiamao + H'tin	97%	68–81%
Tai Yong	Jiamao + H'tin	95%	66–76%
Tai Yuan	Jiamao + H'tin	86%	48–60%
	(Ami,Atayal,Jiamao) + H'tin	10%	56–66%
Yao	(Ami,Atayal,Jiamao) + Hmong Thailand	79%	60–86%
	Hmong Thailand + H'tin	6%	87–94%
	(Ami,Atayal,Jiamao,She) + Hmong Thailand	6%	81–89%
Zhuang	Jiamao + H'tin	99%	87–92%

Inferred sources of ancestry (with bootstrap support) and mixture proportions (95% confidence intervals) from *MixMapper* for East and Mainland Southeast Asian populations. Names with parentheses refer to the common ancestral branches of the specified nodes (see Fig. 1). Branch topologies are shown that occur for at least 5% of 500 bootstrap replicates. We see essentially no evidence of the four ancestry components found in Austronesian-speaking groups, aside from H'tin-related (Austro-Asiatic) ancestry in several populations. We note that some of the populations here may not truly be admixed, but we show all of the fits for completeness.

Supplementary Table 14. Formal test for numbers of sources of admixture

Test subset	p-value for 2 sources	p-value for 3 sources	p-value for 4 sources
Agta, Ati, Ayta, Ilocano, Iraya, Manobo	**0.000**	0.110	0.156
Alorese, Kambera, Lamaholot, Lembata	**0.000**	0.486	0.428
Alorese, Kambera, Lamaholot, Lembata, Manggarai Ngada, Manggarai Rampasasa	0.000	**0.000**	0.366
Bidayuh, Dayak, Javanese Jakarta, Javanese Java, Mentawai, Sunda	0.000	**0.000**	0.068
Bidayuh, Dayak, Javanese Jakarta	0.000	**0.018**	NA

We applied a formal test based on f_4 statistics, as described in refs. (33) and (34), to estimate how many sources of admixture are necessary to explain the observed relationships among a collection of admixed populations. Briefly, we estimate the rank of a matrix of values $f_4(A, B; C, D)$, where A and B are populations in a test set and C and D are populations in a reference set. To remove trivially linearly dependent rows and columns, we fix A and C to be the first populations in each list (without loss of generality) and let B and D vary. In order to maximize sensitivity for separate sources of Asian ancestry, we used a reference set consisting of Yoruba as the fixed outgroup C and 31 East and Southeast Asian populations as the other references D. We used a p-value threshold of 0.05; a score below this threshold implies that at least that many sources are necessary to explain the relationships among the test set. In bold are the maximal significant values, indicating the estimated number of sources for each set.

Supplementary Table 15. Mixture fits for Austronesian-speaking populations with no Taiwanese in the scaffold tree

Admixed population	Mixing branch 1 + branch 2	% reps	Branch 1 ancestry
Agta	Jiamao + Papuan	96%	51–62%
Alorese	Papuan + H'tin	81%	55–62%
	Papuan + (H'tin,Plang)	9%	55–62%
Ami	Jiamao + H'tin	43%	85–95%
	Jiamao + Karitiana	36%	98–99%
	Jiamao + Surui	15%	98–99%
Atayal	Jiamao + (Karitiana,Surui)	30%	88–97%
	Jiamao + Papuan	14%	93–99%
	Jiamao + (Mandenka,Papuan,Yoruba,root)	12%	88–97%
	Jiamao + (Karitiana,Mandenka,Papuan,Surui,Yoruba,root)	10%	71–88%
	Jiamao + (Mandenka,Yoruba,root)	8%	94–98%
	Jiamao + H'tin	7%	78–95%
	Jiamao + Surui	7%	97–99%
Ati	Jiamao + Papuan	95%	50–59%
Ayta	Papuan + H'tin	92%	61–75%
	Papuan + Jiamao	5%	63–80%
Bidayuh	Manobo + H'tin	100%	30–42%
Dayak	Manobo + H'tin	100%	46–59%
Iraya	Jiamao + Papuan	77%	68–80%
	Jiamao + (Mandenka,Papuan,Yoruba,root)	21%	57–71%
JavaneseJakarta	Manobo + H'tin	100%	34–46%
Kambera	(H'tin,Plang,Wa) + Papuan	42%	71–76%
	Jiamao + Papuan	34%	67–74%
	(HmongThailand,Jiamao,Miao,She) + Papuan	21%	69–75%
Mamanwa	Jiamao + Papuan	76%	50–60%
	H'tin + Papuan	12%	52–59%
	(H'tin,Plang,Wa) + Papuan	9%	53–64%
Manobo	Jiamao + Papuan	100%	77–84%
Sunda	Manobo + H'tin	100%	40–51%

Inferred sources of ancestry (with bootstrap support) and mixture proportions (95% confidence intervals) from *MixMapper* for selected Austronesian-speaking populations, using a 16-population scaffold tree formed by removing Ami and Atayal from the original scaffold (i.e., Miao, She, Jiamao, Lahu, Wa, Yi, Naxi, Hmong, Plang, H'tin, Palaung, Karitiana, Suruí, Papuan, Mandenka, and Yoruba). Names with parentheses refer to the common ancestral branches of the specified nodes (see Fig. 1). Branch topologies are shown that occur for at least 5% of 500 bootstrap replicates. We report admixture fits for Ami and Atayal as test populations, as well as all other Austronesian-speaking populations with no negative f_3 statistics (Supplementary Table 5) and selected others to fill in geographic coverage gaps. For both Ami and Atayal, more than half of the bootstrap replicates yield fits with 90% or more Jiamao ancestry and a very small proportion of a seemingly implausible second ancestry component (e.g., Native American). In our experience, such results indicate that the test populations should in fact be modeled as unadmixed relative to the scaffold, in this case adjacent to Jiamao (31). For other populations, meanwhile, the fits appear to be reasonable and are very similar (both in topology and mixture proportions) to those obtained with the original scaffold (with the difference that Jiamao is now the closest population to the previous location of the Taiwanese). Fits with Manobo reported as one mixing branch are three-way admixtures (proportions are not re-optimized).

Supplementary Table 16. Robustness of Austro-Asiatic ancestry with modified scaffolds

E. Indonesian / Oceanian admixed population	Percent bootstrap support with H'tin removed	Percent bootstrap support with H'tin and Plang removed
Manggarai Ngada	95%	16%
Manggarai Rampasasa	44%	0%
Toraja	84%	36%
W. Indonesian admixed population	**Percent bootstrap support with H'tin removed**	**Percent bootstrap support with H'tin and Plang removed**
Batak Toba	45%	24%
Bidayuh	100%	98%
Dayak	100%	93%
Javanese Jakarta	100%	100%
Javanese Java	100%	100%
Malay Indonesia	69%	29%
Malay Singapore	63%	31%
Sunda	100%	100%

Robustness of the Austro-Asiatic ancestry component from *MixMapper* for three-way admixed populations with either H'tin or H'tin and Plang removed from the 18-population scaffold tree. Shown are the percentages of bootstrap replicates (out of 500) assigning the third ancestry component in a three-way admixture model to an Austro-Asiatic branch in the scaffold (Plang or Wa in the first column and Wa in the second column). The fits on the reduced scaffolds are not as robust for the eastern Indonesian populations, while the lower confidences for Batak Toba and the Malay populations may be due to a small proportion of Indian ancestry (20, 25) that is picked up more often with fewer Austro-Asiatic references present.