

SUPPLEMENTARY MATERIAL

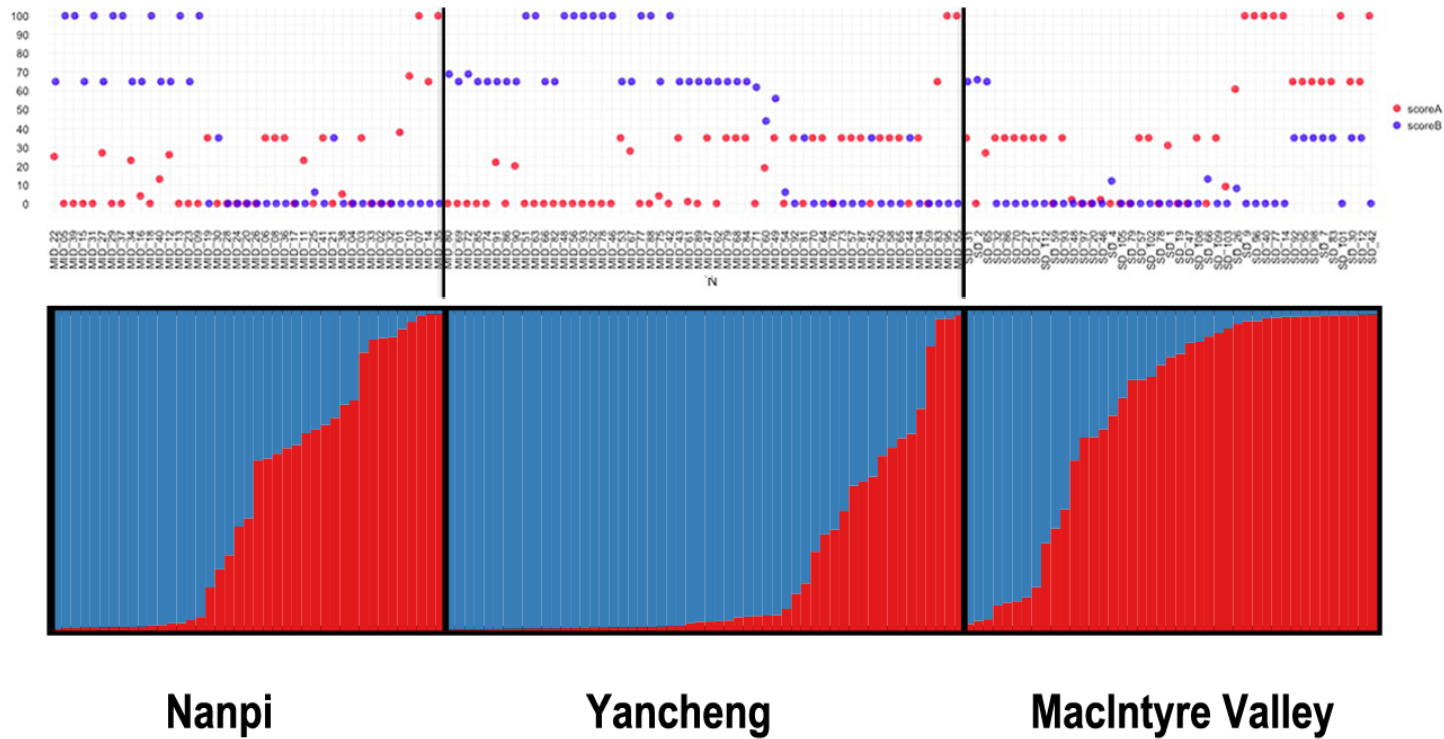


Figure S1: Correlation between a single, randomly-selected graphical representation of a STRUCTURE run and the results of multiple runs. A total of 100 runs were carried out in STRUCTURE using $K=2$. The dotplot shows the number of times each individual was strongly (probability of 0.95 or higher) assigned to a particular cluster, while the coloured barplot is a typical representation of a STRUCTURE analysis to infer population structure from genotypic data.

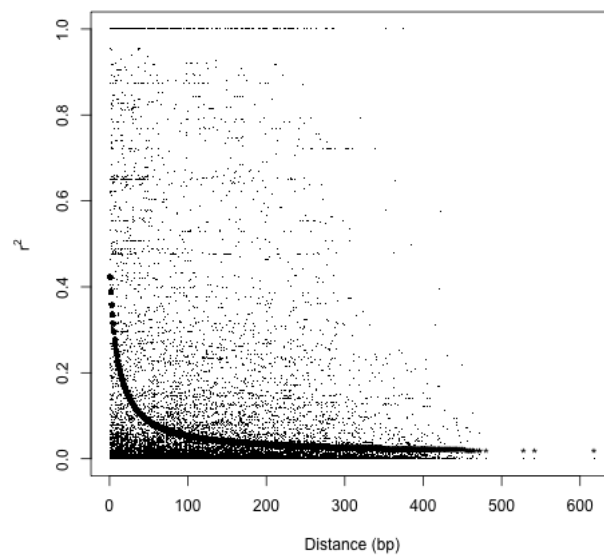


Figure S2: Decay of linkage disequilibrium across 40 loci compared in the Nanpi and Yancheng populations. Data points represent r^2 values from each locus plotted against physical distance in base pairs. The curve shows the decay of LD modelled on the expectations of Hill & Weir (1988).

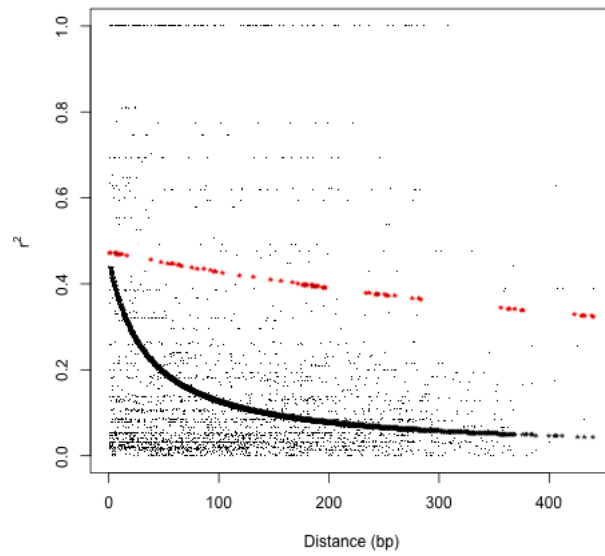


Figure S3: Decay of linkage disequilibrium across eight loci in the MacIntyre Valley, Nanpi and Yancheng populations. Data points represent r^2 values taken from *Cyp303down1*, *Cyp303down3*, *Cyp305b1*, *Period*, *Phc*, *SCAP*, *Tc* and *Tpi*, plotted against physical distance in base pairs. The curve in red models the decay of LD for *Cyp303down1* while the decay of the other seven loci (pooled) is modelled in grey. The models used are described by the formula in Hill & Weir (1988).

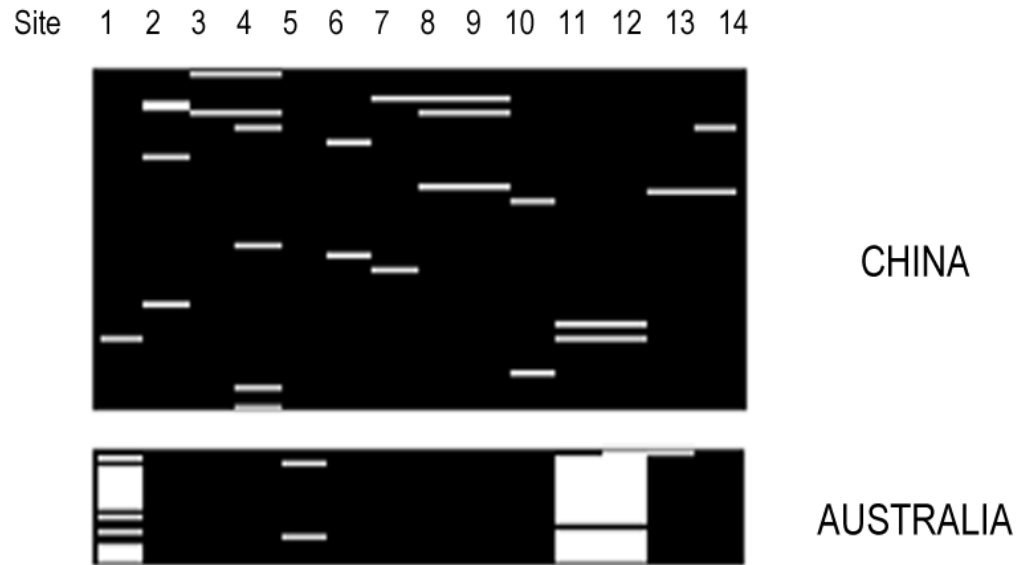


Figure S4: The *Cyp303down3* haplotypes in China ($n=69$) and Australia ($n=22$). Each row represents an individual and alleles are arbitrarily coded as black and white. The haplotype is characterised by 14 sites. At three sites (1, 11 and 12), there is a distinct difference in allele frequencies between Chinese and Australian individuals. The major Australian haplotype is primarily defined by Sites 11 and 12.

Primer	Sequence 5'–3'
BGIBMGA000508F	GCCTCCCTCGCGCCAaggttctcttaaacatcaagaaatg
BGIBMGA000508R	GCCTTGCCAGCCCGCttcaccatthaagcttatacaNcg
BGIBMGA000529F	GCCTCCCTCGCGCCAatcgtcaacgccttctactc
BGIBMGA000529R	GCCTTGCCAGCCCGCtctgcatggttctcgccttg
BGIBMGA000615F	GCCTCCCTCGCGCCAAttggaaggagattcaagagaa
BGIBMGA000615R	GCCTTGCCAGCCCGCccatcttcataatagctgccttt
BGIBMGA000615F2	GCCTCCCTCGCGCCAacgagttcgtcgccaacaa
BGIBMGA000615R2	GCCTTGCCAGCCCGCgaaggagtttgagagcggttt
BGIBMGA000686F	GCCTCCCTCGCGCCAagattgtatggacgacttcagtgc
BGIBMGA000686R	GCCTTGCCAGCCCGCcatgtccgcgagacagtta
BGIBMGA000687F	GCCTCCCTCGCGCCAatggcagctggtatattcatca
BGIBMGA000687R	GCCTTGCCAGCCCGCtctctgaatatcactccacaca
BGIBMGA000698F	GCCTCCCTCGCGCCAagacgaagagagccttttatttca
BGIBMGA000698R	GCCTTGCCAGCCCGCgcattgctcattttgaatgg
BGIBMGA002020F	GCCTCCCTCGCGCCAacattccagatgttcaactacca
BGIBMGA002020R	GCCTTGCCAGCCCGCgcgcaaagtcgggtaaataag
BGIBMGA002034F	GCCTCCCTCGCGCCAacgcttgaaggttctccatga
BGIBMGA002034R	GCCTTGCCAGCCCGCcgcggtgggtatctccactt
BGIBMGA002043F	GCCTCCCTCGCGCCAgtgccgtaccacaacttcg
BGIBMGA002043R	GCCTTGCCAGCCCGCcatggtgacNgttatggtgaagtga
BGIBMGA002140F	GCCTCCCTCGCGCCAacgttcaaccgaactatttgc
BGIBMGA002140R	GCCTTGCCAGCCCGCaccaagttgccactatcac
BGIBMGA002140F2	GCCTCCCTCGCGCCAagggtacagtcaccgcacaa
BGIBMGA002140R2	GCCTTGCCAGCCCGCacgacgacgectactatcag
BGIBMGA002144F	GCCTCCCTCGCGCCAaccgataaacgtcagtaaacag
BGIBMGA002144R	GCCTTGCCAGCCCGCttgcatacgtcttcacacactg
BGIBMGA002144F2	GCCTCCCTCGCGCCAagcgaaagataacccaagttc
BGIBMGA002144R2	GCCTTGCCAGCCCGCcttctcccagcaacagacc
BGIBMGA003858F	GCCTCCCTCGCGCCAagatacgggacacggtttacgc
BGIBMGA003858R	GCCTTGCCAGCCCGCcattggtggtcggttctac
BGIBMGA003864F	GCCTCCCTCGCGCCAaggtggtatcaatggcgtatg
BGIBMGA003864R	GCCTTGCCAGCCCGCttgtcggccatgtcgttc
BGIBMGA003866F	GCCTCCCTCGCGCCAacgcacagttctcacaggaaa
BGIBMGA003866R	GCCTTGCCAGCCCGCgtggtgtcagggtcgaactc

BGIBMGA012230F	GCCTCCCTCGCGCCAcacagttcaagaaggatttga
BGIBMGA012230R	GCCTTGCCAGCCCGCttgatgagttcatgaaggtg
BGIBMGA012230F2	GCCTCCCTCGCGCCAaggagctgcgtacaacatc
BGIBMGA012230R2	GCCTTGCCAGCCCGCccattcgcgaggagatt
BGIBMGA012240F	GCCTCCCTCGCGCCAagttgcgagagttgccagt
BGIBMGA012240R	GCCTTGCCAGCCCGCtgatecgtgcggcattcat
BGIBMGA012269F	GCCTCCCTCGCGCCAttcggactgtgtatgcctca
BGIBMGA012269R	GCCTTGCCAGCCCGCgcagtcgcgttgaagtcg
BGIBMGA012322F	GCCTCCCTCGCGCCAgtggtgtcacatacctcatgc
BGIBMGA012322R	GCCTTGCCAGCCCGCcaatcggtgcggtcttc
BGIBMGA012323F	GCCTCCCTCGCGCCAaggaacagccgatggatag
BGIBMGA012323R	GCCTTGCCAGCCCGCcaggaatggggtctcggtg
BGIBMGA013317F	GCCTCCCTCGCGCCAtcaagagcgaatggatttgt
BGIBMGA013317R	GCCTTGCCAGCCCGCtttgcgtggagacgtggaaa
BGIBMGA013317F2	GCCTCCCTCGCGCCAttccttcgcatcactcac
BGIBMGA013317R2	GCCTTGCCAGCCCGCgacctgggtacacgccatt
BGIBMGA013328F	GCCTCCCTCGCGCCAagacaagatggagaatcagagc
BGIBMGA013328R	GCCTTGCCAGCCCGCgcattgettacgtagttcgtc
<i>Cyp303down1F</i>	GCCTCCCTCGCGCCAgcactccacagcggtattt
<i>Cyp303down1R</i>	GCCTTGCCAGCCCGCtgategaaatccaaaggaagtca
<i>Cyp303down3F</i>	GCCTCCCTCGCGCCActaattgaccgacatgctaggg
<i>Cyp303down3R</i>	GCCTTGCCAGCCCGCtaactactgggttgccttctgg
<i>Cyp303F0</i>	GCCTCCCTCGCGCCAaggactttgctatgcgaaac
<i>Cyp303R0</i>	GCCTTGCCAGCCCGCtgategacggctgcatcg
<i>Cyp303F1</i>	GCCTCCCTCGCGCCAaggagaacgccgaggagtatt
<i>Cyp303R1</i>	GCCTTGCCAGCCCGCacttctgctcctcaccag
<i>Cyp303F2</i>	GCCTCCCTCGCGCCAacggcacaagagaggtttga
<i>Cyp303R2</i>	GCCTTGCCAGCCCGCttcttacctctggtataaaatcc
<i>Cyp303F3</i>	GCCTCCCTCGCGCCAtgcgtacgacaattgaataaata
<i>Cyp303R3</i>	GCCTTGCCAGCCCGCgtgtaaattcgateaacaaga
<i>Cyp303up1F</i>	GCCTCCCTCGCGCCAacgtgcctgaacttacaaca
<i>Cyp303up1R</i>	GCCTTGCCAGCCCGCtcttcggcttatgggtatcg
<i>Cyp305b1F</i>	GCCTCCCTCGCGCCAacacgtctgcgtttctccaa
<i>Cyp305b1R</i>	GCCTTGCCAGCCCGCggtgtagccaatataccaatcaac
<i>KettinF</i>	GCCTCCCTCGCGCCAaggctcatcaatggcaacac
<i>KettinR</i>	GCCTTGCCAGCCCGCggcctgtggtcaataactcc

<i>PeriodF</i>	GCCTCCCTCGCGCCAatggcaatgggcagcaac
<i>PeriodR</i>	GCCTTGCCAGCCCGCgcactggttgatgtagga
<i>PhcF</i>	GCCTCCCTCGCGCCAatcggaagatgtgtacaagg
<i>PhcR</i>	GCCTTGCCAGCCCGCgcgcacgttggtgataatcc
<i>ScapF</i>	GCCTCCCTCGCGCCAatggactggtgtcaggettat
<i>ScapR</i>	GCCTTGCCAGCCCGCcggttacatttccttcagctt
<i>TcF</i>	GCCTCCCTCGCGCCAaaatgtgtgctcaagagttgg
<i>TcR</i>	GCCTTGCCAGCCCGCttgttgatagcttcgcaagagt
<i>TpiF</i>	GCCTCCCTCGCGCCAattcggttggtggtaactgga
<i>TpiR</i>	GCCTTGCCAGCCCGCtaccgatagcccaaactggt

Table S1: List of 454 primer sequences. The first 15 bases (uppercase) are the universal tails, with a different sequence for forward and reverse primers. The remaining bases (lowercase) are the locus-specific sequence.

Locus	n	Number of sites (bp)^a	S^b		Haplotype diversity	π	
BGIBMGA000508 (22)	Nanpi (13)	472	21	29	0.974	0.010	0.009
	Yanch. (9)		13			0.009	
BGIBMGA000529 (56)	Nanpi (24)	312–367	122	115	0.999	0.087	0.082
	Yanch. (32)		123			0.078	
BGIBMGA000615 (68)	Nanpi (26)	179–188	38	52	0.788	0.043	0.026
	Yanch. (42)		37			0.025	
BGIBMGA000615F2 (47)	Nanpi (20)	347–385	99	92	0.952	0.038	0.035
	Yanch. (27)		64			0.032	
BGIBMGA000686 (61)	Nanpi (29)	392–419	92	105	0.980	0.030	0.030
	Yanch. (32)		79			0.029	
BGIBMGA000687 (66)	Nanpi (30)	224–251	45	32	0.989	0.027	0.026
	Yanch. (36)		27			0.026	
BGIBMGA000698 (61)	Nanpi (23)	130–156	40	37	0.950	0.031	0.029
	Yanch. (38)		28			0.029	
BGIBMGA002020 (19)	Nanpi (8)	188	16	26	0.994	0.034	0.037
	Yanch. (11)		23			0.042	
BGIBMGA002034 (54)	Nanpi (22)	282–325	79	104	0.859	0.046	0.044
	Yanch. (32)		92			0.044	
BGIBMGA002043 (46)	Nanpi (26)	411–432	47	52	0.976	0.021	0.019
	Yanch. (20)		25			0.016	

BGIBMGA002140 (50)	Nanpi (13)	388–391	42	98	0.979	0.029	0.032
	Yanch. (37)		90			0.035	
BGIBMGA002140F2 (90)	Nanpi (37)	424–425	68	87	0.975	0.028	0.029
	Yanch. (53)		73			0.029	
BGIBMGA002144 (28)	Nanpi (10)	343–373	65	84	0.988	0.048	0.048
	Yanch. (18)		78			0.049	
BGIBMGA002144F2 (57)	Nanpi (24)	375–399	63	95	0.978	0.030	0.032
	Yanch. (33)		90			0.033	
BGIBMGA003858 (10)	Nanpi (8)	446–487	16	18	0.867	0.010	0.009
	Yanch. (2)		3			0.006	
BGIBMGA003864 (52)	Nanpi (19)	223–247	53	56	0.878	0.052	0.046
	Yanch. (33)		51			0.044	
BGIBMGA003866 (93)	Nanpi (40)	229–241	61	87	0.881	0.063	0.067
	Yanch. (53)		81			0.073	
BGIBMGA012230 (19)	Nanpi (13)	443	7	13	0.868	0.003	0.004
	Yanch. (6)		7			0.006	
BGIBMGA012230F2 (83)	Nanpi (30)	441–453	14	31	0.752	0.004	0.005
	Yanch. (53)		27			0.006	
BGIBMGA012240 (88)	Nanpi (38)	274–287	34	55	0.888	0.027	0.027
	Yanch. (50)		50			0.027	
BGIBMGA012269 (13)	Nanpi (4)	356–366	24	49	1.000	0.034	0.032
	Yanch. (9)		42			0.031	
BGIBMGA012322 (83)	Nanpi (38)	241–249	38	58	0.994	0.023	0.027

	Yanch. (45)		53			0.031	
BGIBMGA012323 (31)	Nanpi (8)	204–295	93	115	0.987	0.111	0.125
	Yanch. (23)		113			0.130	
BGIBMGA013317 (24)	Nanpi (14)	356–374	35	69	0.867	0.027	0.036
	Yanch. (10)		56			0.046	
BGIBMGA013317F2 (84)	Nanpi (39)	289–300	13	20	0.637	0.006	0.006
	Yanch. (45)		16			0.006	
BGIBMGA013328 (68)	Nanpi (28)	302	12	14	0.772	0.006	0.006
	Yanch. (40)		11			0.006	
<i>Cyp303down1</i> (33)	Nanpi (15)	472–521	41	50	0.991	0.023	0.021
	Yanch. (18)		35			0.016	
<i>Cyp303down3</i> (69)	Nanpi (26)	446–455	33	51	0.926	0.012	0.009
	Yanch. (43)		31			0.008	
<i>Cyp303F0</i> (75)	Nanpi (34)	125–133	14	26	0.757	0.014	0.013
	Yanch. (41)		18			0.012	
<i>Cyp303F1</i> (88)	Nanpi (38)	660	18	49	0.882	0.003	0.004
	Yanch. (50)		36			0.004	
<i>Cyp303F2</i> (83)	Nanpi (35)	216	14	25	0.503	0.004	0.005
	Yanch. (48)		19			0.005	
<i>Cyp303F3</i> (25)	Nanpi (4)	439–463	19	30	0.980	0.023	0.016
	Yanch. (21)		27			0.014	
<i>Cyp303up1</i> (78)	Nanpi (34)	284–377	59	79	0.980	0.022	0.023
	Yanch. (44)		67			0.024	

<i>Cyp305b1</i> (36)	Nanpi (17)	460–470	35	48	0.940	0.020	0.020
	Yanch. (19)		36			0.021	
<i>Kettin</i> (93)	Nanpi (41)	446–447	33	48	0.931	0.007	0.007
	Yanch. (52)		34			0.007	
<i>Period</i> (87)	Nanpi (37)	248–280	27	35	0.815	0.010	0.016
	Yanch. (50)		34			0.019	
<i>Phc</i> (59)	Nanpi (25)	286–394	77	66	0.906	0.046	0.039
	Yanch. (34)		58			0.037	
<i>Scap</i> (79)	Nanpi (34)	419–446	86	111	0.929	0.037	0.034
	Yanch. (45)		74			0.032	
<i>Tc</i> (89)	Nanpi (38)	449	20	31	0.738	0.009	0.008
	Yanch. (51)		22			0.007	
<i>Tpi</i> (48)	Nanpi (17)	463	78	109	0.950	0.052	0.048
	Yanch. (31)		91			0.045	

Table S2: Nucleotide diversity and haplotype diversity for 40 loci surveyed in this study. Figures in brackets after the locus name represent the total number of sequences surveyed. Where figures are presented in two columns under a single heading, the left column reports the estimates for an individual population while the right column reports the estimates after pooling the sequences from all populations.

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

^a the number of sites is presented as a range due to the differing subsets of indel polymorphisms present in different populations. Since gapped sites are excluded from this analysis, the lower boundary represents the number of sites considered when alleles from both populations are pooled.

^b number of segregating sites including singletons.

Inferred ancestry of individuals:

Label	(% missing)	Pop	Inferred clusters	
MID-01	(52)	1	0.998	0.002
MID-02	(72)	1	0.906	0.094
MID-03	(43)	1	0.999	0.001
MID-04	(29)	1	0.955	0.045
MID-05	(25)	1	1.000	0.000
MID-06	(59)	1	1.000	0.000
MID-07	(53)	1	0.921	0.079
MID-08	(58)	1	0.998	0.002
MID-09	(38)	1	0.999	0.001
MID-10	(67)	1	0.999	0.001
MID-11	(29)	1	0.998	0.002
MID-12	(22)	1	0.999	0.001
MID-13	(40)	1	0.975	0.025
MID-14	(40)	1	0.952	0.048
MID-15	(20)	1	0.999	0.001
MID-16	(29)	1	1.000	0.000
MID-17	(33)	1	0.999	0.001
MID-18	(34)	1	0.998	0.002
MID-19	(32)	1	0.998	0.002
MID-20	(49)	1	0.954	0.046
MID-21	(57)	1	0.943	0.057
MID-22	(58)	1	0.999	0.001
MID-23	(41)	1	0.997	0.003
MID-24	(53)	1	0.998	0.002
MID-25	(24)	1	0.927	0.073
MID-26	(35)	1	0.993	0.007
MID-27	(27)	1	0.970	0.030
MID-28	(47)	1	0.991	0.009
MID-29	(40)	1	1.000	0.000
MID-30	(33)	1	1.000	0.000
MID-31	(25)	1	0.943	0.057
MID-32	(62)	1	0.999	0.001
MID-33	(31)	1	0.994	0.006
MID-34	(27)	1	1.000	0.000

MID-35	(33)	1	0.999	0.001
MID-36	(27)	1	0.993	0.007
MID-37	(65)	1	1.000	0.000
MID-38	(45)	1	0.971	0.029
MID-39	(41)	1	0.980	0.020
MID-40	(37)	1	1.000	0.000
MID-41	(16)	1	0.982	0.018
MID-42	(48)	2	0.998	0.002
MID-43	(29)	2	0.994	0.006
MID-44	(37)	2	0.993	0.007
MID-45	(41)	2	0.972	0.028
MID-46	(51)	2	0.978	0.022
MID-47	(37)	2	0.914	0.086
MID-48	(42)	2	0.999	0.001
MID-49	(36)	2	0.993	0.007
MID-50	(36)	2	0.973	0.027
MID-51	(35)	2	0.999	0.001
MID-52	(30)	2	0.988	0.012
MID-53	(51)	2	0.998	0.002
MID-54	(47)	2	0.996	0.004
MID-55	(23)	2	0.998	0.002
MID-56	(43)	2	0.998	0.002
MID-57	(14)	2	0.998	0.002
MID-58	(28)	2	0.927	0.073
MID-59	(38)	2	0.994	0.006
MID-60	(32)	2	0.999	0.001
MID-61	(38)	2	0.983	0.017
MID-62	(24)	2	1.000	0.000
MID-63	(30)	2	1.000	0.000
MID-64	(24)	2	0.997	0.003
MID-65	(42)	2	0.997	0.003
MID-66	(27)	2	1.000	0.000
MID-67	(35)	2	0.995	0.005
MID-68	(38)	2	0.998	0.002
MID-69	(46)	2	0.999	0.001
MID-70	(29)	2	0.922	0.078

MID-71	(28)	2	0.999	0.001
MID-72	(36)	2	1.000	0.000
MID-73	(28)	2	0.999	0.001
MID-74	(38)	2	1.000	0.000
MID-75	(20)	2	0.996	0.004
MID-76	(27)	2	0.999	0.001
MID-77	(41)	2	1.000	0.000
MID-78	(15)	2	1.000	0.000
MID-79	(62)	2	0.998	0.002
MID-80	(31)	2	0.999	0.001
MID-81	(26)	2	0.995	0.005
MID-82	(17)	2	0.999	0.001
MID-83	(25)	2	0.995	0.005
MID-84	(33)	2	0.996	0.004
MID-85	(28)	2	1.000	0.000
MID-86	(13)	2	1.000	0.000
MID-87	(23)	2	0.989	0.011
MID-88	(52)	2	0.992	0.008
MID-89	(71)	2	0.998	0.002
MID-90	(37)	2	0.999	0.001
MID-91	(40)	2	1.000	0.000
MID-92	(36)	2	0.999	0.001
MID-93	(25)	2	1.000	0.000
MID-94	(42)	2	0.870	0.130
MID-95	(59)	2	0.856	0.144

Table S3: Output of STRUCTURE analysis used to generate Figure 2A showing inferred ancestry of individuals. Column names refer to individual names (labels), percentage of missing data for each individual, population to which each individual belongs (1=Nanpi, 2=Yancheng) and inferred clusters. The analysis was run using $K=2$ for a total of 1352 sites under a model incorporating admixture and independent allele frequencies between populations, without using prior population information.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	 Ln''(K) 	ΔK
1	10	-8637.08	0.59	—	—	—
2	10	-7835.75	23.65	801.33	579.68	24.51
3	10	-7614.10	148.64	221.65	141.08	0.95
4	10	-7533.53	71.42	80.57	533.88	7.48
5	10	-7986.84	1391.28	-453.31	655.73	0.47
6	10	-7784.42	915.85	202.42	581.56	0.63
7	10	-8163.56	1558.53	-379.14	887.71	0.57
8	10	-7654.99	310.14	508.57	—	—

Table S4: Results of the Evanno method (Evanno *et al.*, 2005) for evaluating ΔK for the Nanpi, Yancheng and MacIntyre Valley populations using 310 sites. The most likely value of K (number of genetically defined populations) was evaluated as that which corresponds to the largest value of ΔK .

REFERENCES

- Evanno, G., Regnaut, S. & Goudet, J.** (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**, 2611–2620, 2611–2620. ISSN 1365-294X. doi:10.1111/j.1365-294X.2005.02553.x.
- Hill, W.G. & Weir, B.S.** (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology* **33**, 54–78, 54–78. ISSN 0040-5809. doi:10.1016/0040-5809(88)90004-4.