Mapping Human Genetic Diversity in Asia

The HUGO Pan-Asian SNP Consortium

Science **326**, 1541 (2009);
DOI: 10.1126/science.1177074

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

hybrid sterility involves both the unusual abundance and retention of OdsHmau protein in the *D. simulans* testis, as well as an unusual localization and possibly decondensation of the *D. simulans* Y chromosome. We conclude on the basis of these data that hybrid male sterility is caused by a gain-of-function interaction between OdsHmau and some component of the *D. simulans* Y chromosome heterochromatin, with this protein-DNA interaction representing the Dobzhansky-Muller incompatibility.

OdsH shares similarities with the hybrid sterility genes *Prdm9* (or *Meisetz*) in mouse (*23*) and *Overdrive* (*Ovd*) in *Drosophila* (*24*), all of which encode proteins with putative DNA-binding domains. Satellite DNAs have also been implicated in hybrid inviability, including a pericentric satellite locus (*Zhr*) (*25*, *26*) and a gene encoding a heterochromatin-binding protein (*Lhr*) (*27*). Thus, rapidly evolving repetitive DNA elements driven by genetic conflict may represent a major evolutionary force driving sequence divergence of speciation genes that would ultimately result in hybrid incompatibilities (*13*, *14*, *28*).

References and Notes

1. E. Mayr, *Systematics and the Origin of Species from the Viewpoint of a Zoologist* (Columbia Univ. Press, New York, 1942).
2. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer Associates, Sunderland, MA, 2004).
3. C. C. Laurie, *Genetics* **147**, 937 (1997).
4. R. M. Kliman *et al*., *Genetics* **156**, 1913 (2000).
5. C. T. Ting, S. C. Tsaur, M. L. Wu, C. I. Wu, *Science* **282**, 1501 (1998).
6. S. Sun, C. T. Ting, C. I. Wu, *Science* **305**, 81 (2004).
7. D. E. Perez, C. I. Wu, *Genetics* **140**, 201 (1995).
8. D. E. Perez, C. I. Wu, N. A. Johnson, M. L. Wu, *Genetics* **134**, 261 (1993).
9. S. D. Hueber, I. Lohmann, *Bioessays* **30**, 965 (2008).
10. C. T. Ting *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12232 (2004).
11. K. Tabuchi, S. Yoshikawa, Y. Yuasa, K. Sawamoto, H. Okano, *Neurosci. Lett.* **257**, 49 (1998).
12. M. Nei, J. Zhang, *Science* **282**, 1428 (1998).
13. S. Henikoff, K. Ahmad, H. S. Malik, *Science* **293**, 1098 (2001).
14. S. Henikoff, H. S. Malik, *Nature* **417**, 227 (2002).
15. L. Fishman, A. Saunders, *Science* **322**, 1559 (2008).
16. A. Daniel, *Am. J. Med. Genet.* **111**, 450 (2002).
17. N. Aulner *et al*., *Mol. Cell. Biol.* **22**, 1218 (2002).
18. M. Ashburner, K. G. Golic, R. S. Hawley, *Drosophila: A Laboratory Handbook* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 2005).
19. G. Cenci, S. Bonaccorsi, C. Pisano, F. Verni, M. Gatti, *J. Cell Sci.* **107**, 3521 (1994).
20. B. D. McKee, *Curr. Top. Dev. Biol.* **37**, 77 (1998).
21. J. E. Tomkiel, *Genetica* **109**, 95 (2000).
22. J. Forejt, *Trends Genet.* **12**, 412 (1996).
23. O. Mihola, Z. Trachtulec, C. Vlcek, J. C. Schimenti, J. Forejt, *Science* **323**, 373 (2009).
24. N. Phadnis, H. A. Orr, *Science* **323**, 376 (2009).
25. K. Sawamura, M. T. Yamamoto, T. K. Watanabe, *Genetics* **133**, 307 (1993).
26. P. M. Ferree, D. A. Barbash, *PLoS Biol.* **7**, e1000234 (2009).
27. N. J. Brideau *et al*., *Science* **314**, 1292 (2006).
28. H. S. Malik, S. Henikoff, *Cell* **138**, 1067 (2009).

29. We thank C-I. Wu for the *D. simulans* fertile and sterile introgression lines; C. Ting for scientific discussions and sharing data; G. Findlay for initial observations on OdsH cytology; and K. Ahmad, S. Biggins, N. Elde, S. Henikoff, N. Phadnis, T. Tsukiyama, and D. Vermaak for comments on the manuscript. Supported by NIH training grant PHS NRSA T32 GM07270 (J.J.B.), and grants from the Mathers foundation and NIH R01-GM74108 (H.S.M.). H.S.M. is an Early-Career Scientist of the Howard Hughes Medical Institute.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1181756/DC1
Materials and Methods
Figs. S1 to S8
References

10 September 2009; accepted 13 October 2009
Published online 22 October 2009;
10.1126/science.1181756
Include this information when citing this paper.

Mapping Human Genetic Diversity in Asia

The HUGO Pan-Asian SNP Consortium*†

Asia harbors substantial cultural and linguistic diversity, but the geographic structure of genetic variation across the continent remains enigmatic. Here we report a large-scale survey of autosomal variation from a broad geographic sample of Asian human populations. Our results show that genetic ancestry is strongly correlated with linguistic affiliations as well as geography. Most populations show relatedness within ethnic/linguistic groups, despite prevalent gene flow among populations. More than 90% of East Asian (EA) haplotypes could be found in either Southeast Asian (SEA) or Central-South Asian (CSA) populations and show clinal structure with haplotype diversity decreasing from south to north. Furthermore, 50% of EA haplotypes were found in SEA only and 5% were found in CSA only, indicating that SEA was a major geographic source of EA populations.

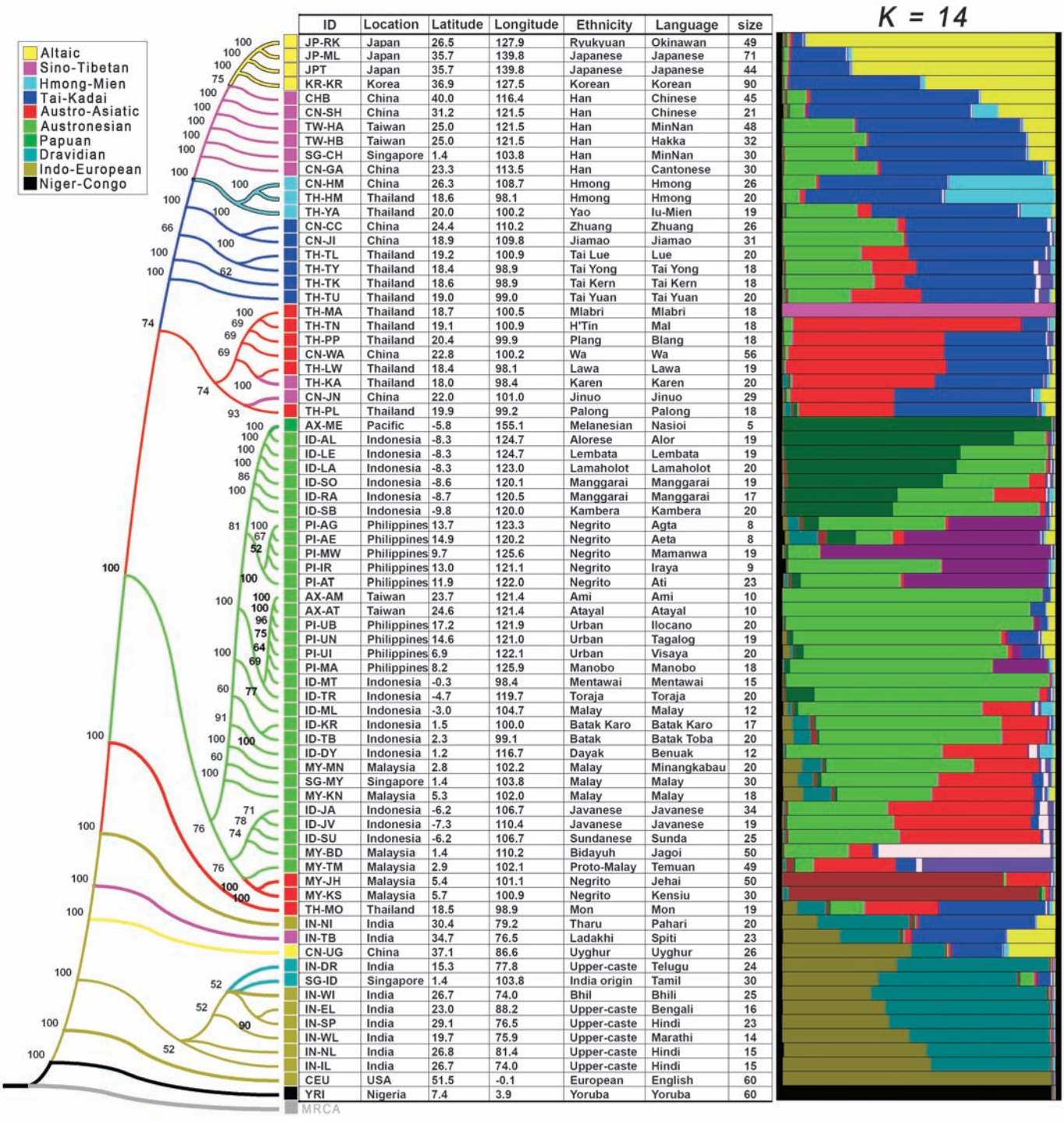
Several genome-wide studies of human genetic diversity focusing primarily on broad continental relationships, or fine-scale structure in Europe, have been published recently (*1*–*8*). We have extended this approach to Southeast Asian (SEA) and East Asian (EA) populations by using the Affymetrix GeneChip Human Mapping 50K Xba Array. Stringently quality-controlled genotypes were obtained at 54,794 autosomal single-nucleotide polymorphisms (SNPs) in 1928 individuals representing 73 Asian and two non-Asian HapMap populations (*9*). Apart from developing a general description of Asian population structure and its relation to geography, language, and demographic history, we concentrated on uncovering the geographic source(s) of EA and SEA populations.

We first performed a Bayesian clustering procedure using the STRUCTURE algorithm (*10*) to examine the ancestry of each individual. Each person is posited to derive from an arbitrary number of ancestral populations, denoted by K. We ran STRUCTURE from $K = 2$ to $K = 14$ using both the complete data set and SNP subsets to exclude those in strong linkage disequilibrium (Fig. 1 and figs. S1 to S13). At $K = 2$ and $K = 3$, all SEA and EA samples are united by predominant membership in a common cluster, with the other cluster(s) corresponding largely to Indo-European (IE) and African (AF) ancestries. At $K = 4$, a component most frequently found in Negrito populations that is also shared by all SEA populations emerges, suggesting a common SEA ancestry. Each value of K beyond 4 introduces a new component that tends to be associated with a group of populations united by membership in a linguistic family, by geographic proximity, by a known history of admixture, or, especially at higher Ks, by membership in a small population isolate. The results obtained using *frappe* (*11*), a maximum-likelihood–based clustering analysis, showed a general concordance with those of STRUCTURE (figs. S14 to S26). These analyses show that most individuals within a population share very similar ancestry estimates at all Ks, an observation that is consistent also with a phylogeny relating individuals (fig. S27) based on an allele-sharing distance (*12*). Therefore, we proceeded to evaluate the relationships among populations. A maximum-likelihood tree of populations, based on 42,793 SNPs whose ancestral states were known (Fig. 1), showed that all the SEA and EA populations make up a monophyletic clade that is supported by 100% of bootstrap replicates. This pattern remained even after data from 51 additional populations and 19,934 commonly typed SNPs from a recent study were integrated into the tree (fig. S28). These observations suggest that SEA and EA populations share a common origin.

STRUCTURE/*frappe* and principal components analyses (PCA) (*13*) (Figs. 1 and 2 and figs. S1 to S26) identify as many as 10 main population components. Each component corresponds largely to one of the five major linguistic groups (Altaic, Sino-Tibetan/Tai-Kadai, Hmong-Mien, Austro-Asiatic, and Austronesian), three ethnic categories (Philippine Negritos, Malaysian Negritos, and East Indonesians/Melanesians) and two small population isolates (the Bidayuh of Borneo and the hunter-gatherer Mlabri population of central and northern Thailand). The STRUCTURE results

*All authors with their affiliations appear at the end of this paper.
†To whom correspondence should be addressed. E-mail: ljin007@gmail.com (L.J.); liue@gis.a-star.edu.sg (E.T.L.); seielstadm@gis.a-star.edu.sg (M.S.); xushua@picb.ac.cn (S.X.)

(Fig. 1 and figs. S1 to S13), population phylogenies (Fig. 1 and figs. S27 and S28), and PCA results (Fig. 2) all show that populations from the same linguistic group tend to cluster together. A Mantel test confirms the correlation between linguistic and genetic affinities ($R^2 = 0.253$; $P < 0.0001$ with 10,000 permutations), even after controlling for geography (partial correlation = 0.136; $P <$ 0.005 with 10,000 permutations). Nevertheless, we identified eight population outliers whose linguistic and genetic affinities are inconsistent [Affymetrix-Melanesian (AX-ME), Malaysia-Jehai (MY-JH)



ID	Location	Latitude	Longitude	Ethnicity	Language	size
JP-RK	Japan	26.5	127.9	Ryukyuan	Okinawan	49
JP-ML	Japan	35.7	139.8	Japanese	Japanese	71
JPT	Japan	35.7	139.8	Japanese	Japanese	44
KR-KR	Korea	36.9	127.5	Korean	Korean	90
CHB	China	40.0	116.4	Han	Chinese	45
CN-SH	China	31.2	121.5	Han	Chinese	21
TW-HA	Taiwan	25.0	121.5	Han	MinNan	48
TW-HB	Taiwan	25.0	121.5	Han	Hakka	32
SG-CH	Singapore	1.4	103.8	Han	MinNan	30
CN-GA	China	23.3	113.5	Han	Cantonese	30
CN-HM	China	26.3	108.7	Hmong	Hmong	26
TH-HM	Thailand	18.6	98.1	Hmong	Hmong	20
TH-YA	Thailand	20.0	100.2	Yao	Iu-Mien	19
CN-CC	China	24.4	110.2	Zhuang	Zhuang	26
CN-JI	China	18.9	109.8	Jiamao	Jiamao	31
TH-TL	Thailand	19.2	100.9	Tai Lue	Lue	20
TH-TY	Thailand	18.4	98.9	Tai Yong	Tai Yong	18
TH-TK	Thailand	18.6	98.9	Tai Kern	Tai Kern	18
TH-TU	Thailand	19.0	99.0	Tai Yuan	Tai Yuan	20
TH-MA	Thailand	18.7	100.5	Mlabri	Mlabri	18
TH-TN	Thailand	19.1	100.9	H'Tin	Mal	18
TH-PP	Thailand	20.4	99.9	Plang	Blang	18
CN-WA	China	22.8	100.2	Wa	Wa	56
TH-LW	Thailand	18.4	98.1	Lawa	Lawa	19
TH-KA	Thailand	18.0	98.4	Karen	Karen	20
CN-JN	China	22.0	101.0	Jinuo	Jinuo	29
TH-PL	Thailand	19.9	99.2	Palong	Palong	18
AX-ME	Pacific	-5.8	155.1	Melanesian	Nasioi	5
ID-AL	Indonesia	-8.3	124.7	Alorese	Alor	19
ID-LE	Indonesia	-8.3	124.7	Lembata	Lembata	19
ID-LA	Indonesia	-8.3	123.0	Lamaholot	Lamaholot	20
ID-SO	Indonesia	-8.6	120.1	Manggarai	Manggarai	19
ID-RA	Indonesia	-8.7	120.5	Manggarai	Manggarai	17
ID-SB	Indonesia	-9.8	120.0	Kambera	Kambera	20
PI-AG	Philippines	13.7	123.3	Negrito	Agta	8
PI-AE	Philippines	14.9	120.2	Negrito	Aeta	8
PI-MW	Philippines	9.7	125.6	Negrito	Mamanwa	19
PI-IR	Philippines	13.0	121.1	Negrito	Iraya	9
PI-AT	Philippines	11.9	122.0	Negrito	Ati	23
AX-AM	Taiwan	23.7	121.4	Ami	Ami	10
AX-AT	Taiwan	24.6	121.4	Atayal	Atayal	10
PI-UB	Philippines	17.2	121.9	Urban	Ilocano	20
PI-UN	Philippines	14.6	121.0	Urban	Tagalog	19
PI-UI	Philippines	6.9	122.1	Urban	Visaya	20
PI-MA	Philippines	8.2	125.9	Manobo	Manobo	18
ID-MT	Indonesia	-0.3	98.4	Mentawai	Mentawai	15
ID-TR	Indonesia	-4.7	119.7	Toraja	Toraja	20
ID-ML	Indonesia	-3.0	104.7	Malay	Malay	12
ID-KR	Indonesia	1.5	100.0	Batak Karo	Batak Karo	17
ID-TB	Indonesia	2.3	99.1	Batak	Batak Toba	20
ID-DY	Indonesia	1.2	116.7	Dayak	Benuak	12
MY-MN	Malaysia	2.8	102.2	Malay	Minangkabau	20
SG-MY	Singapore	1.4	103.8	Malay	Malay	30
MY-KN	Malaysia	5.3	102.0	Malay	Malay	18
ID-JA	Indonesia	-6.2	106.7	Javanese	Javanese	34
ID-JV	Indonesia	-7.3	110.4	Javanese	Javanese	19
ID-SU	Indonesia	-6.2	106.7	Sundanese	Sunda	25
MY-BD	Malaysia	1.4	110.2	Bidayuh	Jagoi	50
MY-TM	Malaysia	2.9	102.1	Proto-Malay	Temuan	49
MY-JH	Malaysia	5.4	101.1	Negrito	Jehai	50
MY-KS	Malaysia	5.7	100.9	Negrito	Kensiu	30
TH-MO	Thailand	18.5	98.9	Mon	Mon	19
IN-NI	India	30.4	79.2	Tharu	Pahari	20
IN-TB	India	34.7	76.5	Ladakhi	Spiti	23
CN-UG	China	37.1	86.6	Uyghur	Uyghur	26
IN-DR	India	15.3	77.8	Upper-caste	Telugu	24
SG-ID	Singapore	1.4	103.8	India origin	Tamil	30
IN-WI	India	26.7	74.0	Bhil	Bhili	25
IN-EL	India	23.0	88.2	Upper-caste	Bengali	16
IN-SP	India	29.1	76.5	Upper-caste	Hindi	23
IN-WL	India	19.7	75.9	Upper-caste	Marathi	14
IN-NL	India	26.8	81.4	Upper-caste	Hindi	15
IN-IL	India	26.7	74.0	Upper-caste	Hindi	15
CEU	USA	51.5	-0.1	European	English	60
YRI	Nigeria	7.4	3.9	Yoruba	Yoruba	60

Fig. 1. Maximum-likelihood tree of 75 populations. A hypothetical most-recent common ancestor (MRCA) composed of ancestral alleles as inferred from the genotypes of one gorilla and 21 chimpanzees was used to root the tree. Branches with bootstrap values less than 50% were condensed. Population identification numbers (IDs), sample collection locations with latitudes and longitudes, ethnicities, language spoken, and size of population samples are shown in the table adjacent to each branch in the tree. Linguistic groups are indicated with colors as shown in the legend. All population IDs except the four HapMap samples are denoted by four characters. The first two letters indicate the country where the samples were collected or (in the case of Affymetrix) genotyped, according to the following convention: AX, Affymetrix; CN, China; ID, Indonesia; IN, India; JP, Japan; KR, Korea; MY, Malaysia; PI, the Philippines; SG, Singapore; TH, Thailand; and TW, Taiwan. The last two letters are unique IDs for the population. To the right of the table, an averaged graph of results from STRUCTURE is shown for $K = 14$.

(Negrito), Malaysia-Kensiu (MY-KS) (Negrito), Thailand-Mon (TH-MO), Thailand-Karen (TH-KA), China-Jinuo (CN-JN), India-Spiti (IN-TB), and China-Uyghur (CN-UG); see table S3]. These linguistic outliers tend to cluster with their geographic neighbors or [especially evident in the principal component (PC) plots of Fig. 2] occupy an intermediate position between their geographic neighbors and the more-distant members of their linguistic group. These patterns are consistent either with substantial recent admixture among the populations (*14*–*16*), a history of language replacement (*17*), or uncertainties in the linguistic classifications themselves (for example, the controversial Altaic family, which groups Korean and Japanese with Uyghur).

Considerable gene flow among Asian populations was observed among subpopulations in these clusters, including those groups believed to practice endogamy based on linguistic, cultural, and ethnic information. In fact, most populations studied, even at lower *K*s, show evidence of admixture in the STRUCTURE analyses. For example, the Han Chinese have grown to become the largest ethnic group today in a demographic expansion that has occurred mostly within historical times. STRUCTURE reveals that the six Han Chinese population samples in our study show varying degrees of admixture (Fig. 1 and figs. S1 to S26) between a northern Altaic cluster and a Sino-Tibetan/Tai-Kadai cluster, which most frequently appears in the ethnic groups sampled from southern China and northern Thailand. Finally, most of the Indian populations showed evidence of shared ancestry with European populations, which is consistent with the recent observations (*18*) and our understanding of the expansion of Indo-European–speaking populations (Fig. 1 and figs. S1 to S26).

The geographic source(s) contributing to EA populations have long been debated. One hypothesis suggests that all SEA and EA populations derive primarily from a single initial migration, which entered the continent along a southern, largely coastal route (*19*, *20*). Another hypothesis argues for at least two independent migrations into East Asia, first along a southern route, followed later by a series of migrations along a more northern route that served to bridge European and EA populations, but with little contribution to populations in Southeast Asia (*20*). The topology of a maximum-likelihood tree (Fig. 1 and fig. S28) displays a largely south-to-north ordering of the populations, and a plot of the first two PCs (Fig. 2) similarly orients most populations according to their geographic coordinates. The average

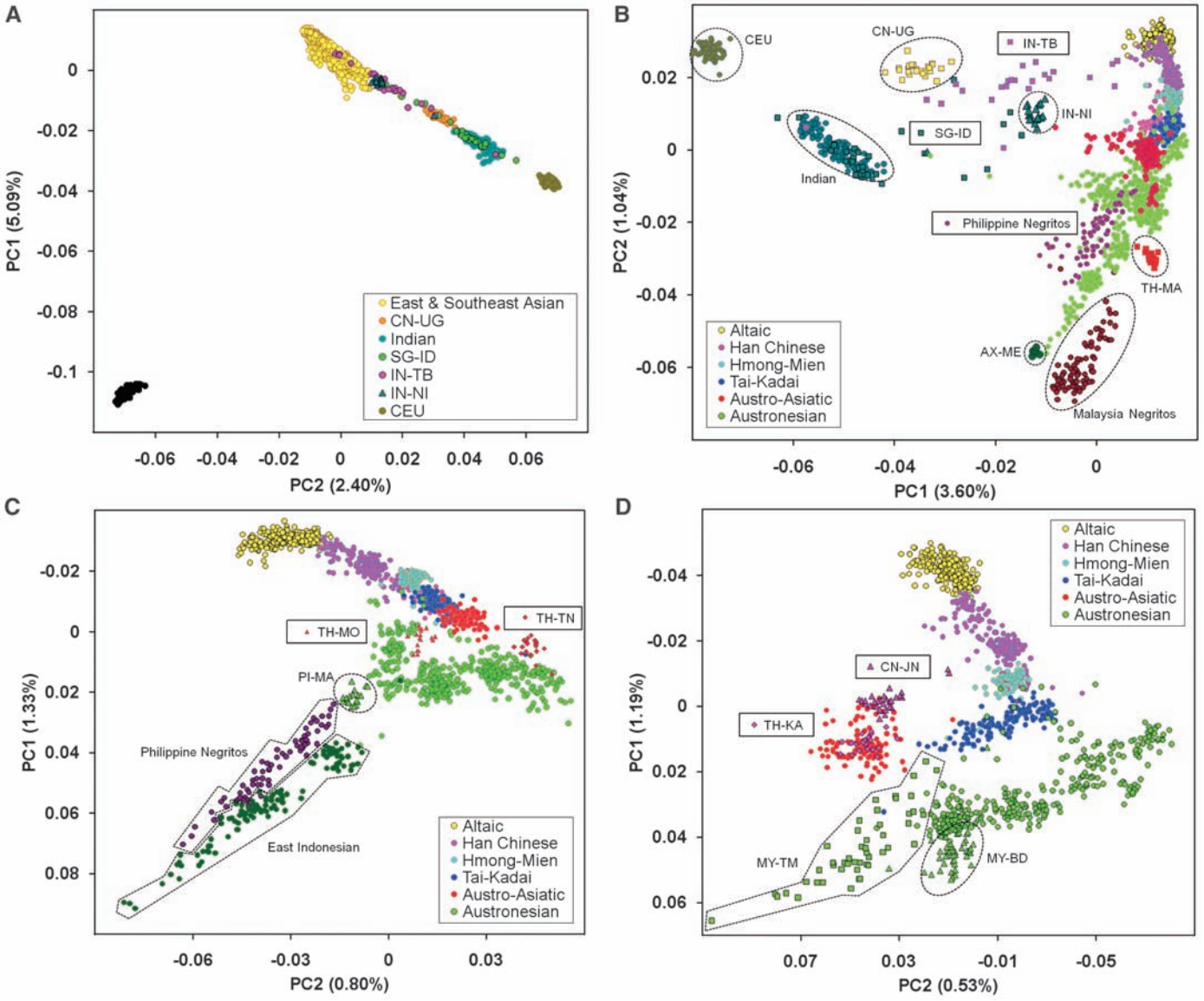


Fig. 2. Analysis of the first two PCs. (**A**) 1928 individuals representing all 75 populations. (**B**) 1868 individuals representing 74 populations (excluding YRI). (**C**) 1471 individuals representing 58 populations (excluding all Indians, CN-UG, TH-MA, AX-ME, and Negritos from Malaysia). (**D**) 1235 individuals representing 44 populations (excluding Philippine Negritos, PI-MA, and East Indonesians).

value of the first PC is highly correlated with the latitude at which the populations were sampled (R^2 = 0.79, P < 0.0001). Such a pattern could result simply from isolation-by-distance (IBD), as suggested by Ding *et al*. (*21*), although a recent study failed to detect IBD in East Asia with data from the Human Genome Diversity Project (*22*).

In an effort to distinguish between long-term historical divergence and the effects of IBD, we applied partial and multiple Mantel tests to the data (*23*) [see supporting online material (SOM) text for details]. The primary approach was to ascertain the differential correlation between genetic distance, geographical distance, and a group indicator matrix as an indication of prehistoric population divergence. The partial correlation coefficient of genetic and geographic distances was 0.228 (P < 0.0006), after controlling for the group indicator matrix (inferred from STRUCTURE/*frappe* analyses), whereas the partial correlation of the genetic and group indicator matrices was 0.403 (P < 0.0001) after controlling for geography. The superior association between genetic distance and the group indicator matrix as measured by the correlation coefficients suggests that prehistorical population divergence is the favored model over IBD in explaining the data (*24*). This conclusion is supported by simulation studies that also suggest that the observed patterns cannot be explained by simple IBD effects alone (see SOM text for details).

To further refine the analysis, we looked to haplotype organization to limit the effect of fluctuations in single-nucleotide determinations and to increase the resolution around genetic diversity. The IBD model predicts a correlation of genetic distance with geographical distance but not genetic diversity and geographic distance (*24*). By contrast, we found (Fig. 3A) that haplotype diversity is strongly correlated with latitude (R^2 = 0.91, P < 0.0001), with diversity decreasing from south to north, which is consistent with a loss of diversity as populations moved to higher latitudes. In estimating the contribution of SEA and Central-South Asian (CSA) haplotypes to the EA gene pool by haplotype sharing analyses (*16*), we found that more than 90% of haplotypes in EA populations could be found in SEA and CSA populations, of which about 50% were found in SEA and EA only and 5% found in CSA only (Fig. 3B, see also SOM text). Phylogenetic analysis of private haplotypes indicates greater similarity between EA and SEA populations relative to EA and CSA populations (Fig. 3C). These observations suggest that the geographic source(s) contributing to EA populations were mainly from SEA populations, with rather minor contributions from CSA,

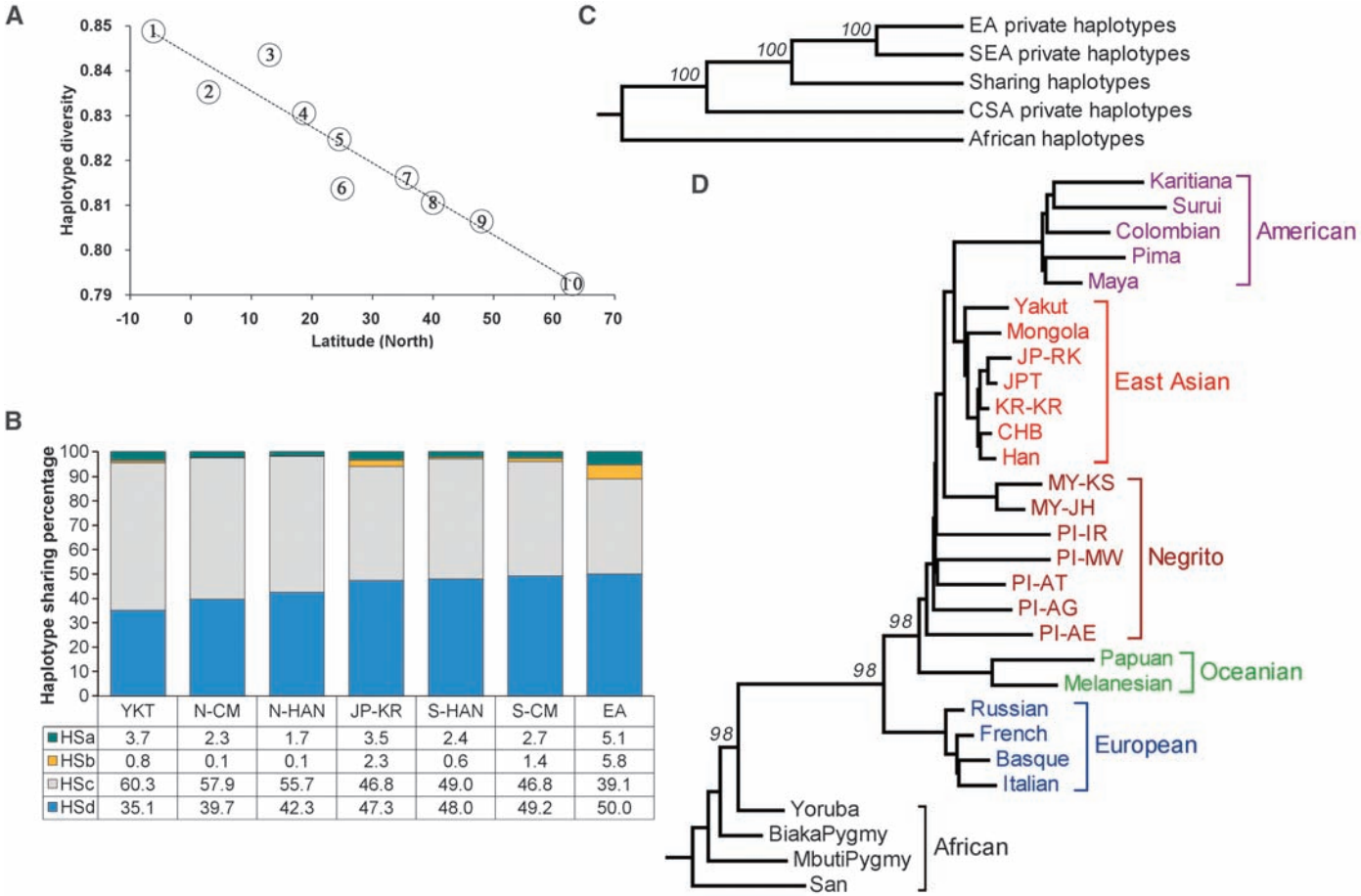


Fig. 3. Analysis of haplotype diversity, haplotype sharing, and population phylogeny. (**A**) Haplotype diversity versus latitudes. Haplotypes were estimated from combined data, and diversity was measured by heterozygosity of haplotypes. HSa, b, c, and d and the corresponding colors show the percentages of EA group haplotypes in each class: HSa, found in CSA only; HSb, found in neither CSA nor SEA; HSc, found in both CSA and SEA; HSd, found in SEA only. Latitudes (*y* axis) for groups were obtained from the center of sample collection locations. Circled numbers are as follows: 1, Indonesian; 2, Malay; 3, Philippine; 4, Thai; 5, Southern Chinese minorities; 6, Southern Han Chinese; 7, Japanese and Korean; 8, Northern Han Chinese; 9, Northern Chinese minorities; and 10, Yakut. Haplotype heterozygosity of each group was estimated from 100-kb bins and taking together all haplotypes within each group. R^2 for the regression line is 0.91 (P < 0.0001). (**B**) Haplotype sharing analysis for EA populations and groups. YKT, Yakut; N-CM, Northern Chinese minorities; N-HAN, Northern Han Chinese; JP-KR, Japanese and Korean; S-HAN, Southern Han Chinese; S-CM, Southern Chinese minorities; EA, East Asian. (**C**) Phylogeny of group private haplotypes. EA private haplotypes: haplotypes found only in EA samples; SEA private haplotypes: haplotypes found only in SEA samples; CSA private haplotypes: haplotypes found only in CSA samples; Shared haplotypes: haplotypes found in all EA, SEA, and CSA samples; African haplotypes were used as outgroup. (**D**) Maximum-likelihood tree of 29 populations. The tree is based on data from 19,934 SNPs. Bootstrap values were based on 100 replicates. Only values on splitting of African and non-African, European and Oceanian and Asian, and Oceanian and Asian are shown.

and that this clinal structure of EA populations arose from prehistoric population divergence rather than IBD or gene flow from CSA populations.

On the basis of increased cultural, linguistic, and genetic diversity, the origins of SEA populations are thought to be more complex than the origins of those to their north. Notably, the Negritos of the Philippines and Malaysia differ from neighboring populations in aspects of their physical appearance, prompting intense speculation about models of human settlement in Southeast Asia. The two-wave hypothesis, which suggests that ancestral Negrito populations settled in Southeast Asia, Australia, and Oceania before a more northerly migration originating in or near the Middle East, and spreading both toward Europe and Northeast Asia via Central Asia (*25*), has been supported by phylogenetic trees constructed from data on a limited number of protein markers (*24*, *25*). The topology of our population trees, both with and without the data from additional European and Asian populations discussed in (*1*), is inconsistent with regard to this genetic similarity of European and EA populations (Figs. 1 and 3D). Instead, on the basis of variation at a large number of independent SNPs, we observed that there is substantial genetic proximity of SEA and EA populations (fig. S28). An identical pattern is seen in the population tree of Li *et al*. (*1*) based on all of their 642,690 SNPs. Our forward-time simulation results under extreme ascertainment scenarios (SOM text) show that the observed phylogeny is not the result of ascertainment bias. Simulation studies also suggest that substantial levels of migration between populations after their initial separation are unlikely to distort the topology of the phylogeny (SOM text).

To unambiguously infer population histories represents a considerable challenge (*26*). Although this study does not disprove a two-wave model of migration, the evidence from our autosomal data and the accompanying simulation studies (figs. S29 and S30) point toward a history that unites the Negrito and non-Negrito populations of Southeast and East Asia via a single primary wave of entry of humans into the continent.

References and Notes

1. J. Z. Li *et al*., *Science* **319**, 1100 (2008).
2. M. Kayser *et al*., *Am. J. Hum. Genet.* **82**, 194 (2008).
3. N. A. Rosenberg *et al*., *PLoS Genet.* **1**, e70 (2005).
4. N. A. Rosenberg *et al*., *Science* **298**, 2381 (2002).
5. J. Novembre *et al*., *Nature* **456**, 98 (2008).
6. M. Nelis *et al*., *PLoS One* **4**, e5472 (2009).
7. C. Tian *et al*., *PLoS Genet.* **4**, e4 (2008).
8. O. Lao *et al*., *Curr. Biol.* **18**, 1241 (2008).
9. The International HapMap Consortium, *Nature* **426**, 789 (2003).
10. J. K. Pritchard, M. Stephens, P. Donnelly, *Genetics* **155**, 945 (2000).
11. H. Tang, J. Peng, P. Wang, N. J. Risch, *Genet. Epidemiol.* **28**, 289 (2005).
12. J. L. Mountain, L. L. Cavalli-Sforza, *Am. J. Hum. Genet.* **61**, 705 (1997).
13. N. Patterson, A. L. Price, D. Reich, *PLoS Genet.* **2**, e190 (2006).
14. S. Xu, L. Jin, *Am. J. Hum. Genet.* **83**, 322 (2008).
15. S. Xu, W. Huang, J. Qian, L. Jin, *Am. J. Hum. Genet.* **82**, 883 (2008).
16. S. Xu, W. Jin, L. Jin, *Mol. Biol. Evol.* **26**, 2197 (2009).
17. L. Reid, in *Language Contact and Change in the Austronesian World.* T. Dutton, T. Tryon, Eds. (Mouton de Gruyter, Berlin, 1994) pp. 443–475.
18. Indian Genome Variation Consortium, *J. Genet.* **87**, 3 (2008).
19. J. Y. Chu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **95**, 11763 (1998).
20. B. Su *et al*., *Am. J. Hum. Genet.* **65**, 1718 (1999).
21. Y. C. Ding *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14003 (2000).
22. A. Manica, F. Prugnolle, F. Balloux, *Hum. Genet.* **118**, 366 (2005).
23. M. P. Telles, J. A. Diniz-Filho, *Genet. Mol. Res.* **4**, 742 (2005).
24. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ, 1993).
25. L. L. Cavalli-Sforza, M. W. Feldman, *Nat. Genet.* **33**, 266 (2003).
26. G. Hellenthal, A. Auton, D. Falush, *PLoS Genet.* **4**, e1000078 (2008).

27. The entire consortium thanks all individuals who volunteered their DNA for this project. It is this collaboration between scientists and the public that is essential to progress in our field. All SNP data have been submitted to dbSNP with the submission handle PASNPI and will become accessible in dbSNP Build 131. See SOM text for a complete listing of all acknowledgments.

The HUGO Pan-Asian SNP Consortium

Mahmood Ameen Abdulla,[1] Ikhlak Ahmed,[2] Anunchai Assawamakin,[3,4] Jong Bhak,[5] Samir K. Brahmachari,[2] Gayvelline C. Calacal,[6] Amit Chaurasia,[2] Chien-Hsiun Chen,[7] Jieming Chen,[8] Yuan-Tsong Chen,[7] Jiayou Chu,[9] Eva Maria C. Cutiongco-de la Paz,[10] Maria Corazon A. De Ungria,[6] Frederick C. Delfin,[6] Juli Edo,[1] Suthat Fucharoen,[3] Ho Ghang,[5] Takashi Gojobori,[11,12] Junsong Han,[13] Sheng-Feng Ho,[7] Boon Peng Hoh,[14] Wei Huang,[15] Hidetoshi Inoko,[16] Pankaj Jha,[2] Timothy A. Jinam,[1] Li Jin,[17,38]† Jongsun Jung,[18] Daoroong Kangwanpong,[19] Jatupol Kampuansai,[19] Giulia C. Kennedy,[20,21] Preeti Khurana,[22] Hyung-Lae Kim,[18] Kwangjoong Kim,[18] Sangsoo Kim,[23] Woo-Yeon Kim,[5] Kuchan Kimm,[24] Ryosuke Kimura,[25] Tomohiro Koike,[11] Supasak Kulawonganunchai,[4] Vikrant Kumar,[8] Poh San Lai,[26,27] Jong-Young Lee,[18] Sunghoon Lee,[5] Edison T. Liu,[8]† Partha P. Majumder,[28] Kiran Kumar Mandapati,[22] Sangkot Marzuki,[29] Wayne Mitchell,[30,31] Mitali Mukerji,[2] Kenji Naritomi,[32] Chumpol Ngamphiw,[4] Norio Niikawa,[40] Nao Nishida,[25] Bermseok Oh,[18] Sangho Oh,[5] Jun Ohashi,[25] Akira Oka,[16] Rick Ong,[8] Carmencita D. Padilla,[10] Prasit Palittapongarnpim,[33] Henry B. Perdigon,[6] Maude Elvira Phipps,[1,34] Eileen Png,[8] Yoshiyuki Sakaki,[35] Jazelyn M. Salvador,[6] Yuliana Sandraling,[29] Vinod Scaria,[2] Mark Seielstad,[8]† Mohd Ros Sidek,[14] Amit Sinha,[2] Metawee Srikummool,[19] Herawati Sudoyo,[29] Sumio Sugano,[37] Helena Suryadi,[29] Yoshiyuki Suzuki,[11] Kristina A. Tabbada,[6] Adrian Tan,[8] Katsushi Tokunaga,[25] Sissades Tongsima,[4] Lilian P. Villamor,[6] Eric Wang,[20,21] Ying Wang,[15] Haifeng Wang,[15] Jer-Yuarn Wu,[7] Huasheng Xiao,[13] Shuhua Xu,[38]† Jin Ok Yang,[5] Yin Yao Shugart,[39] Hyang-Sook Yoo,[5] Wentao Yuan,[15] Guoping Zhao,[15] Bin Alwi Zilfalil,[14] Indian Genome Variation Consortium[2]

[1]Department of Molecular Medicine, Faculty of Medicine, and the Department of Anthropology, Faculty of Arts and Social Sciences, University of Malaya, Kuala Lumpur, 50603, Malaysia. [2]Institute of Genomics and Integrative Biology, Council for Scientific and Industrial Research, Mall Road, Delhi 110007, India. [3]Mahidol University, Salaya Campus, 25/25 M. 3, Puttamonthon 4 Road, Puttamonthon, Nakornpathom 73170, Thailand. [4]Biostatistics and Informatics Laboratory, Genome Institute, National Center for Genetic Engineering and Biotechnology, Thailand Science Park, Pathumtani 12120, Thailand. [5]Korean BioInformation Center (KOBIC), Korea Research Institute of Bioscience and Biotechnology (KRIBB), 111 Gwahangno, Yuseong-gu, Deajeon 305-806, Korea. [6]DNA Analysis Laboratory, Natural Sciences Research Institute, University of the Philippines, Diliman, Quezon City 1101, Philippines. [7]Institute of Biomedical Sciences, Academia Sinica, 128 Sec 2 Academia Road Nangang, Taipei City 115, Taiwan. [8]Genome Institute of Singapore, 60 Biopolis Street 02-01, 138672, Singapore. [9]Institute of Medical Biology, Chinese Academy of Medical Science, Kunming, China. [10]Institute of Human Genetics, National Institutes of Health, University of the Philippines Manila, 625 Pedro Gil Street, Ermita Manila 1000, Philippines. [11]Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization of Information and Systems, 1111 Yata, Mishima, Shizuoka 411-8540, Japan. [12]Biomedicinal Information Research Center, National Institute of Advanced Industrial Science and Technology, 2-42 Aomi, Koto-ku, Tokyo 135-0064, Japan. [13]National Engineering Center for Biochip at Shanghai, 151 Li Bing Road, Shanghai 201203, China. [14]Human Genome Center, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia. [15]MOST-Shanghai Laboratory of Disease and Health Genomics, Chinese National Human Genome Center Shanghai, 250 Bi Bo Road, Shanghai 201203, China. [16]Department of Molecular Life Science Division of Molecular Medical Science and Molecular Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara-A Kanagawa-Pref A259-1193, Japan. [17]State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, China. [18]Korea National Institute of Health, 194, Tongil-Lo, Eunpyung-Gu, Seoul, 122-701, Korea. [19]Department of Biology, Faculty of Science, Chiang Mai University, 239 Huay Kaew Road, Chiang Mai 50202, Thailand. [20]Genomics Collaborations, Affymetrix, 3420 Central Expressway, Santa Clara, CA 95051, USA. [21]Veracyte, 7000 Shoreline Court, Suite 250, South San Francisco, CA 94080, USA. [22]The Centre for Genomic Applications (an IGIB-IMM Collaboration), 254 Ground Floor, Phase III Okhla Industrial Estate, New Delhi 110020, India. [23]Soongsil University, Sangdo-5-dong 1-1, Dongjak-gu, Seoul 156-743, Korea. [24]Eulji University College of Medicine, 143-5 Yong-dudong Jung-gu, Dae-jeon City 301-832, Korea. [25]Department of Human Genetics, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [26]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, National University Hospital, 5 Lower Kent Ridge Road, 119074, Singapore. [27]Population Genetics Lab, Defence Medical and Environmental Research Institute, DSO National Laboratories, 27 Medical Drive, 117510, Singapore. [28]Indian Statistical Institute (Kolkata) 203 Barrackpore Trunk Road, Kolkata 700108, India. [29]Eijkman Institute for Molecular Biology, Jl. Diponegoro 69, Jakarta 10430, Indonesia. [30]Informatics Experimental Therapeutic Centre, 31 Biopolis Way, 03-01 Nanos, 138669, Singapore. [31]Division of Information Sciences, School of Computer Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore. [32]Department of Medical Genetics, University of the Ryukyus Faculty of Medicine, Nishihara, 207 Uehara, Okinawa 903-0215, Japan. [33]National Science and Technology Development Agency, 111 Thailand Science Park, Pathumtani 12120, Thailand. [34]Monash University (Sunway Campus), Jalan Lagoon Selatan, 46150 Bandar Sunway, Selangor, Malaysia. [35]RIKEN Genomic Sciences Center, W502, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan. [36]Department of Biochemistry, University of Hong Kong, 3/F Laboratory Block, Faculty of Medicine Building, 21 Sasson Road, Pokfulam, Hong Kong. [37]Laboratory of Functional Genomics, Department of Medical Genome Sciences Graduate School of Frontier Sciences, University of Tokyo (Shirokanedai Laboratory), 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. [38]Chinese Academy of Sciences-Max Planck Society Partner Institute for Computational Biology, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, 320 Yueyang Rd., Shanghai 200031, China. [39]Genomic Research Branch, National Institute of Mental Health, National Institutes of Health, 6001 Executive Boulevard, Bethesda, MD 20892 USA. [40]Research Institute of Personalized Health Sciences, Health Sciences University of Hokkaido, Tobetsu 061-0293, Japan.

Supporting Online Material

www.sciencemag.org/cgi/content/full/326/5959/1541/DC1
Materials and Methods
SOM Text
Figs. S1 to S38
Tables S1 to S4

1 June 2009; accepted 13 October 2009
10.1126/science.1177074