Science Advances
AAAS

Supplementary Materials for

The genomic history and global migration of a windborne pest

Qing-Ling Hu *et al.*

Corresponding author: Chuan-Xi Zhang, chxzhang@zju.edu.cn; Shuai Zhan, szhan@sibs.ac.cn

Sci. Adv. **10**, eadk3852 (2024)
DOI: 10.1126/sciadv.adk3852

The PDF file includes:

Figs. S1 to S18
Legends for data S1 to S5

Other Supplementary Material for this manuscript includes the following:

Data S1 to S5

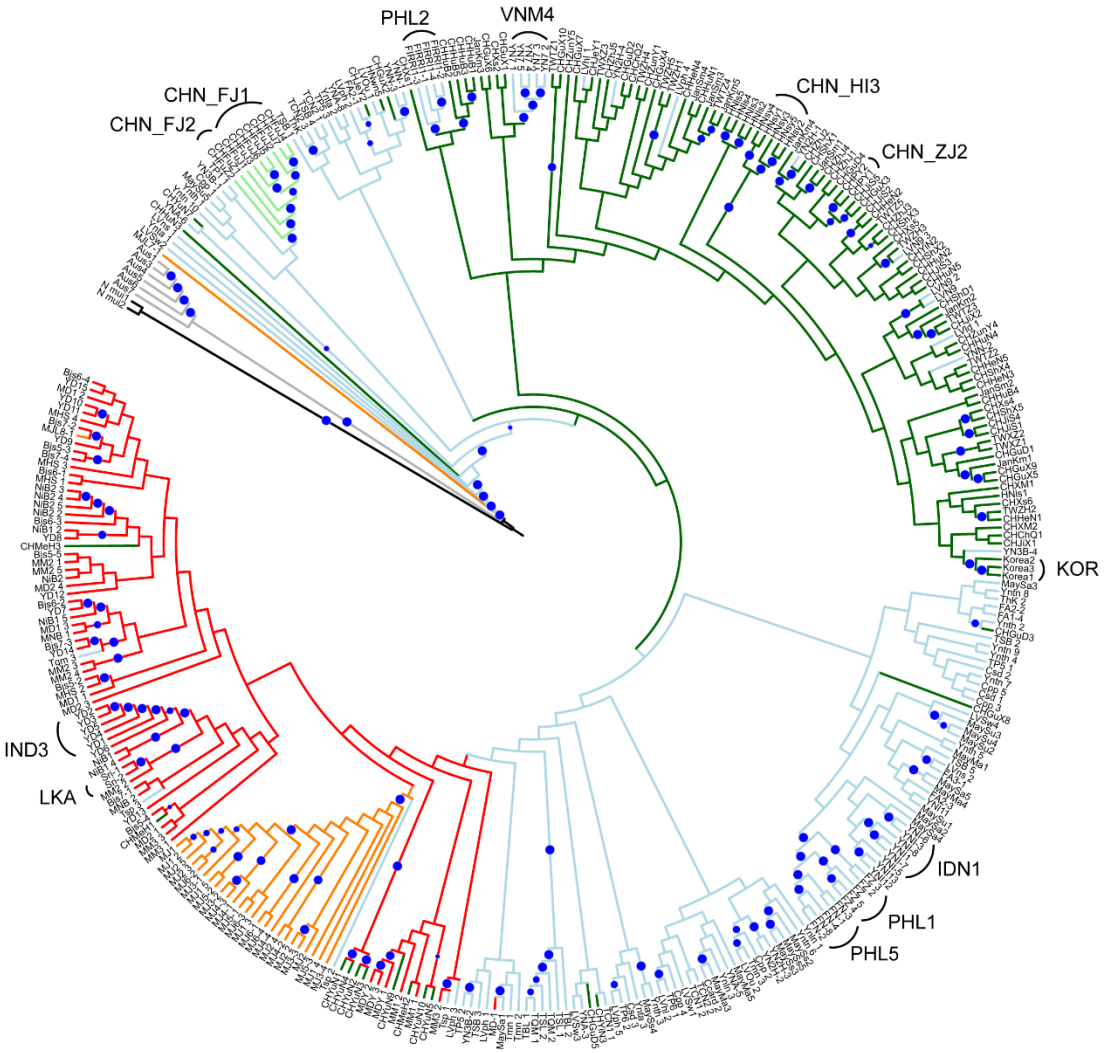

Fig. S1. The phylogenetic relationship of worldwide BPHs. (A) Phylogenomic tree in cladogram format showing the topology. The maximum likelihood tree with branch length is shown in Fig. 1. Blue circles represent the nodes with high level of bootstrap supports (>90%). The color of each sample branch was based on their geographic distribution: green, East Asia; blue, Southeast Asia; red, South Asia; orange, Bangladesh; light green, Fujian, China; grey, Australia; black, *N. muiri*. Lineages consisting of all individuals from a geographic population are outlined with the corresponding population IDs.

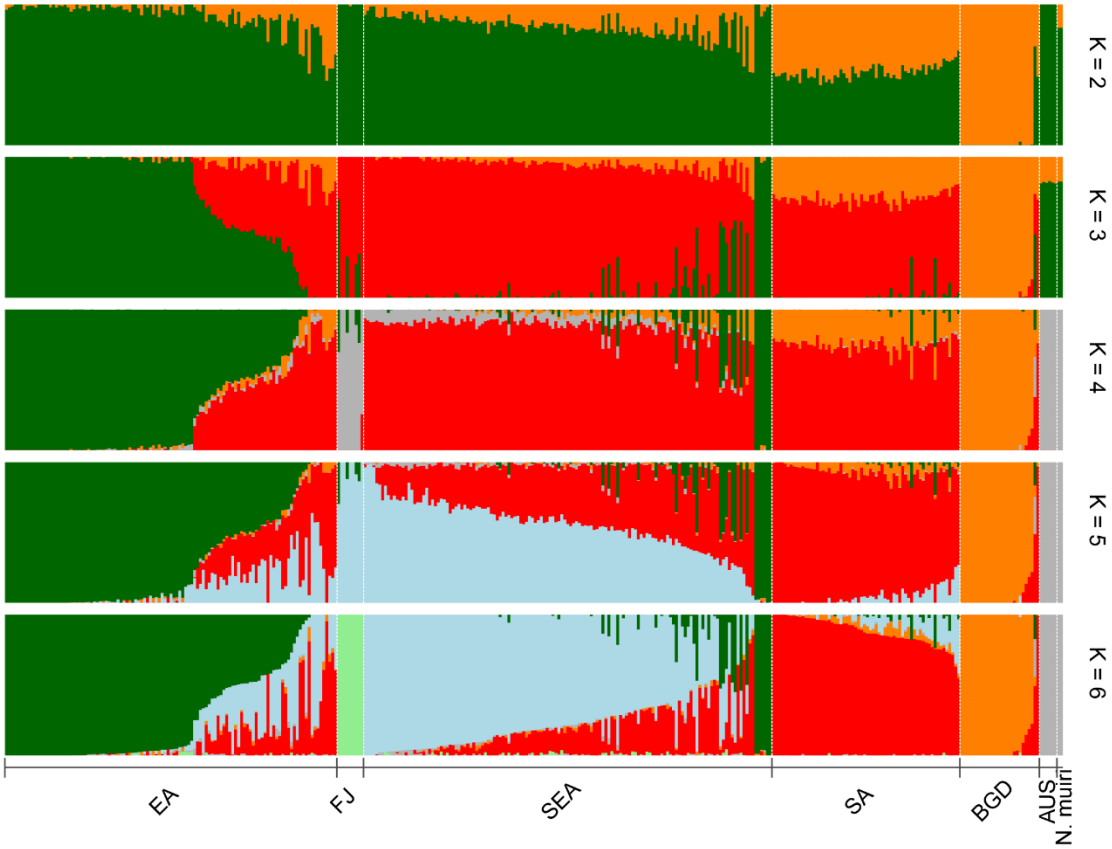

Fig. S2. The genetic strucutre of worldwide BPHs based on ancestry analysis. *K*, presumed ancestry number. See alternative results based on sNMF in Fig. 1D. EA, East Asia; FJ, Fujian, China; SEA, Southeast Asia; SA, South Asia; BGD, Bangladesh; AUS, Australia.

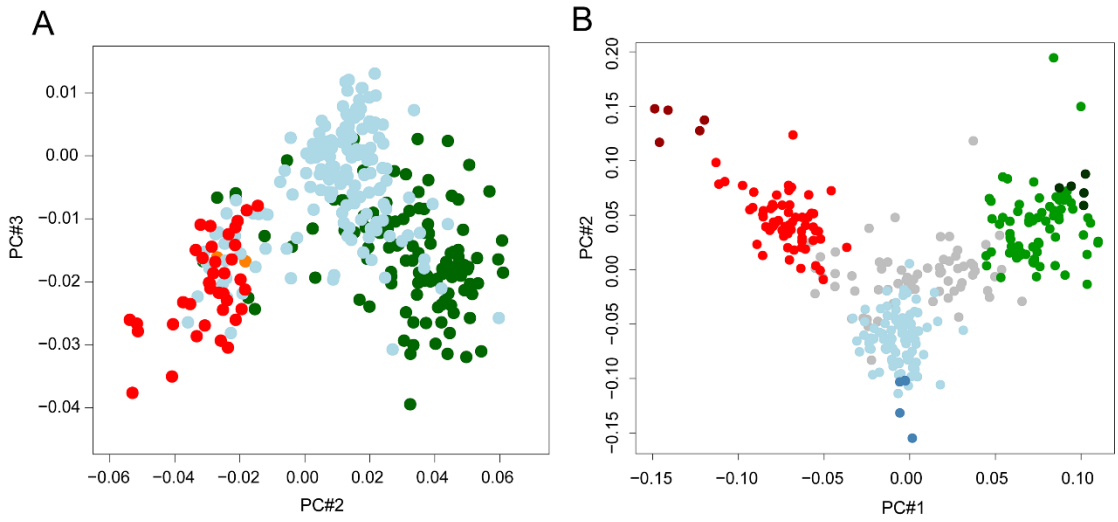

Fig. S3. Principal component analysis for BPHs of EA, SEA, and SA. (A) Enlarged PCA results on all BPH samples focusing on samples from East Asia (green dots), Southeast Asia (blue dots), and South Asia (red dots). See the full results in Fig. 1B. (B) Additional PCA analysis on samples only from the three defined migratory groups (colored as above). The dots in darker color indicate respectively defined core samples for each main groups (see Methods). Grey dots indicate samples defined as admixture among groups. See defined details for samples in Data S1.

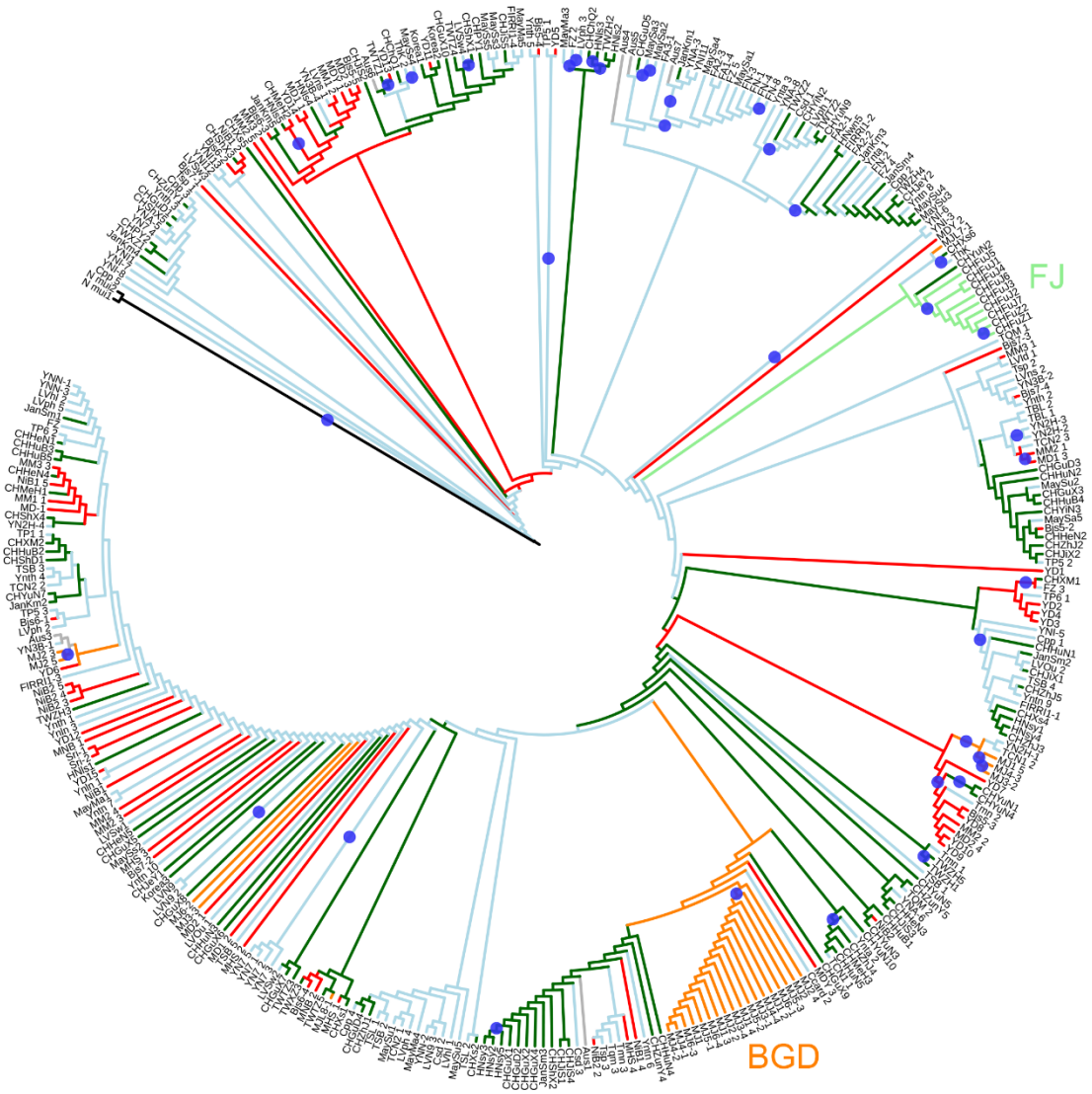

Fig. S4. **The maximum likelihood tree of all sampled BPHs based on mitochondrial sequences.** The color of each sample branch was based on their geographic distribution. Green, East Asia; blue, Southeast Asia; red, South Asia; orange, Bangladesh; lightgreen, Fujian, China; grey, Australia; black, *N. muiri.* Blue circles represent the nodes with high level of bootstrap supports (>75%).

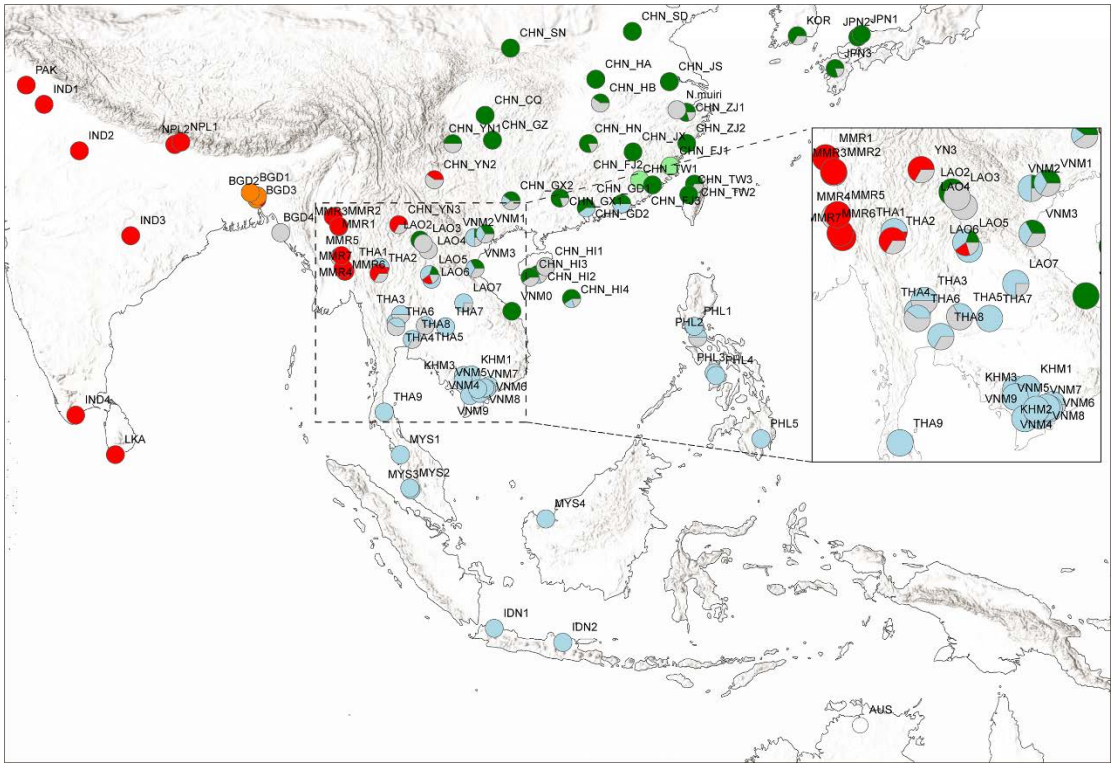

Fig. S5. Inferred ancestry proportion across all Asia BPH populations. The map was generated using ArcMap. Based on the ancestry proportion, each sample was assigned either to one of the six main group (East Asia, Southeast Asia, South Asia, Bangladesh, Fujian, and Australia) or admixed group (see methods). The proportion of each genetic group at each geographic population (labeled with IDs as Data S1) was calculated and plotted as a piechart. Different colors indicate genetic sources from corresponding main groups (Green, East Asia; blue, Southeast Asia; red, South Asia; orange, Bangladesh; light green, Fujian, China; grey, admixed). Inset shows enlarged details of Indochina Peninsula.

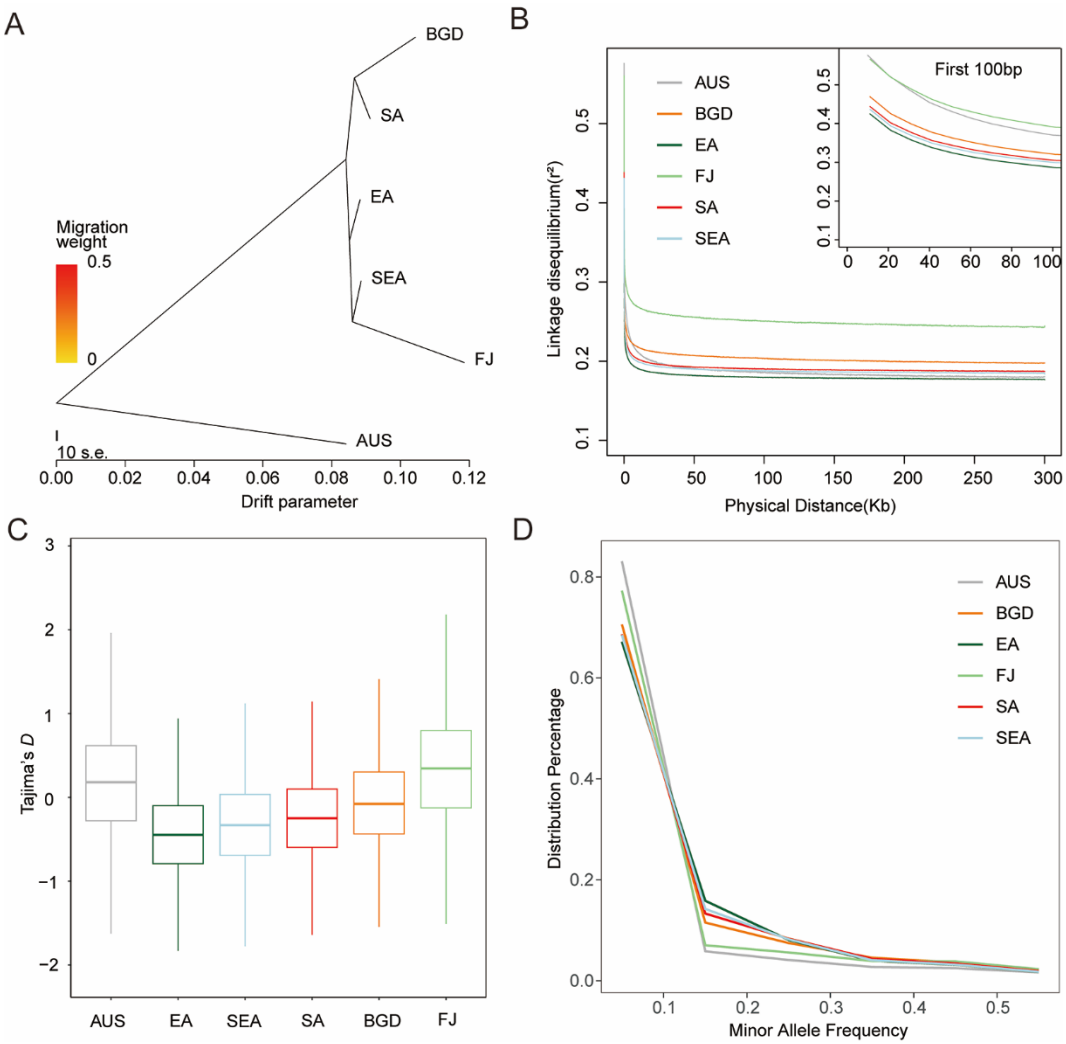

Fig. S6. The phylogenetic relationships and genetic features of six main BPH groups. (A) Treemix analysis across main groups using core samples only. The phylogenetic relationships across six main BPH groups were inferred using a maximum-likelihood approach by Treemix. Core samples (see Fig. S3B, Data S1) were used to represent each group. The tree was rooted using Australia samples. (B) Distribution of Linkage disquilibrium (LD) across all main groups. LD was calculated based on squared correlation coefficient (r^2) among any of two adjacent SNPs shorter than 300 Kb. Core samples (see Data S1) were used to calculate the LD decay for each main group. Inset shows the decay of LD within the first 100 bp. (C) Distribution of Tajima's *D* across all main groups. Box-plot showed the whole-genome distribution of Tajima's *D* of all six main groups. The horizontal line in the middle of the box represented the median, whereas the two whiskers represented the 1.5x interquartile range. (D) Distribution of minor allele frequency across all main groups. The minor allele frequency was estimated for all SNPs across the genome. To avoid the influence of population size, core samples of each group were used as representatives.

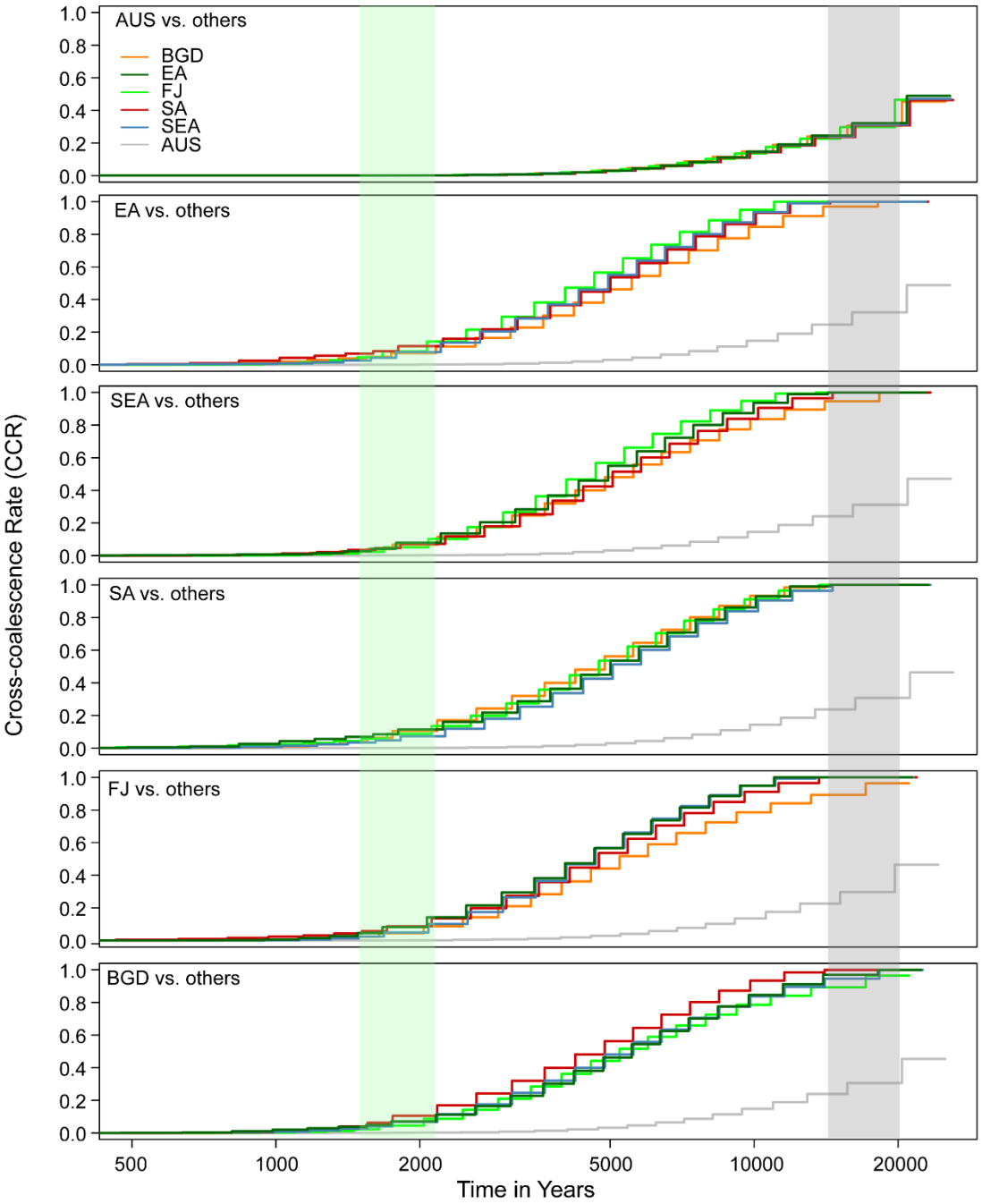

Fig. S7. Inferred population split events across BPH populations. In each panel, the cross-coalescence rates (CCR) between a main group with all the other five groups were calculated by MSMC and plotted to show the diverged pattern between groups. CCR values close to 1 represent not yet separated between populations, while values close to 0 represent full separation. The mutation rate was set as 8.4×10^{-9} based on that of *Drosophil*a. Core samples (Data S1) from each group were used as representatives. Corresponding to those in Fig. 3B, grey shadow indicates the period when Asia BPHs began to split, while the green shadow indicates the period when Asian groups fully diverged.

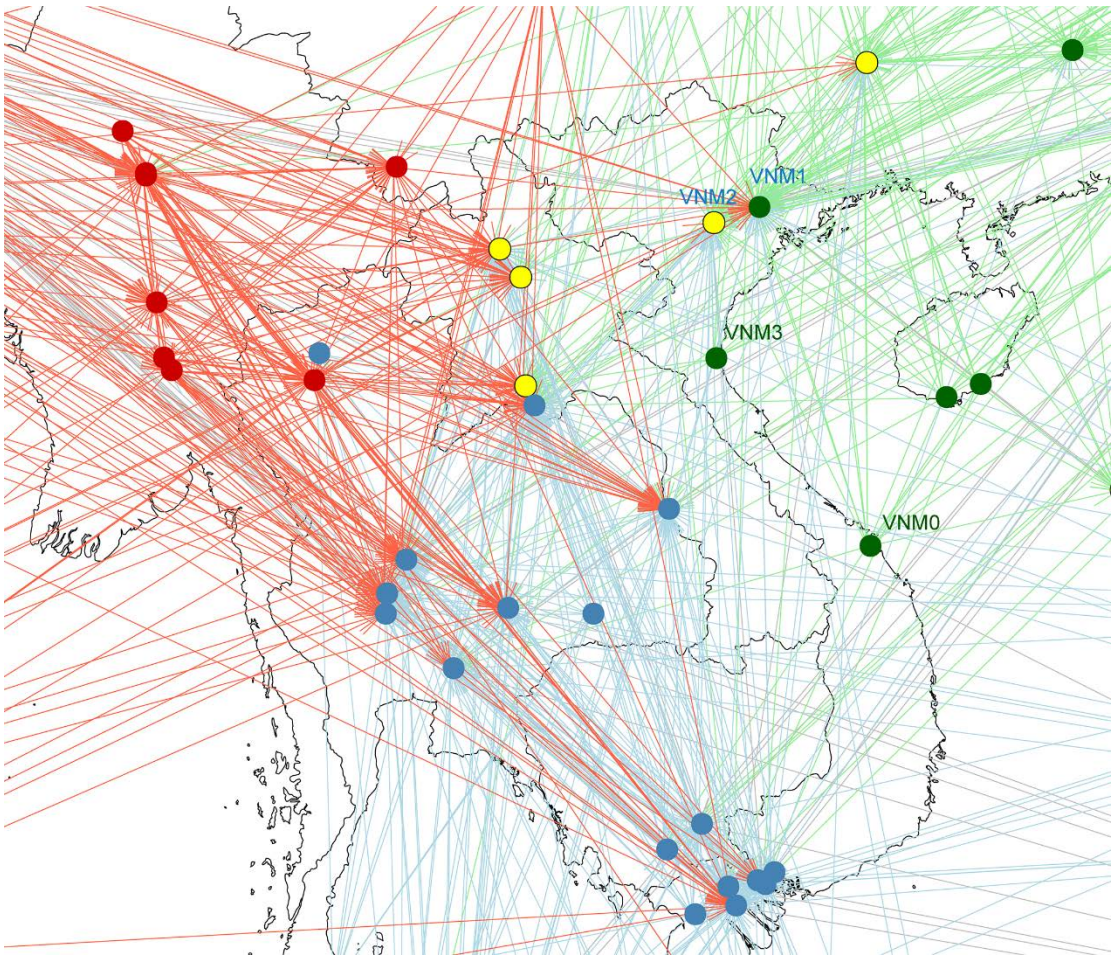

Fig. S8. The enlarged details of BPH population gene flows in Indochina peninsula. Genetic sources of BPH populations in Indochina peninsula along with inferred gene flows**.** Lines with colors indicate gene flows started from the corresponding population. Green, from an East Asia population; blue, from a Southeast Asia population; red, from a South Asia population. Color filled circles indicate the population receiving gene flows. Green, blue, red, and yellow ones indicate East Asia, Southeast Asia, South Asia, and admixed populations, respectively. This figure is the enlarged part of the Indochina peninsula of Fig. 4, additionally including gene flows within short-range. Populations indicated by green text were sampled during the early spring, while those in blue were sampled during late summer.

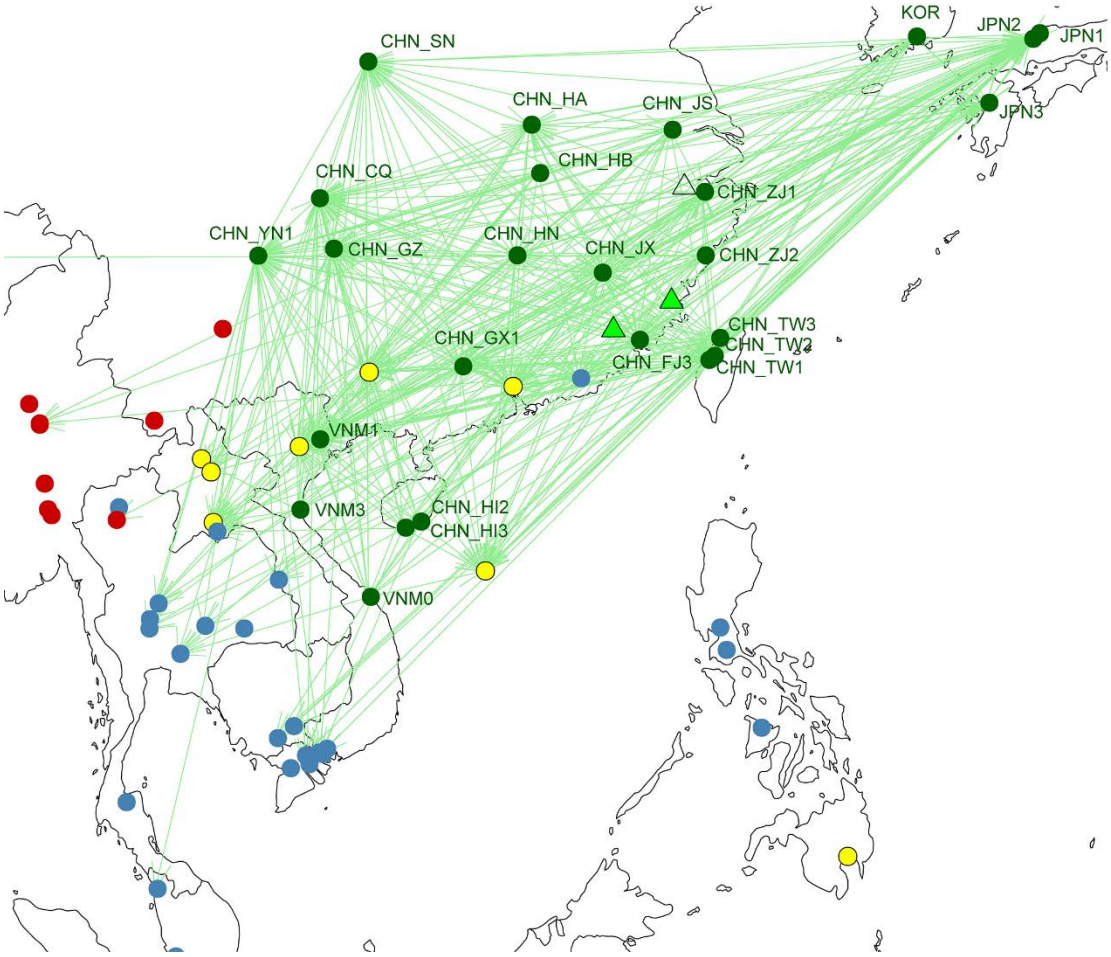

Fig. S9. The enlarged details of BPH population gene flows in East Asian regions. Arrows in green indicate inferred gene flows from an East Asia population (green circles) to the indicated population (green circles, East Asia populations; blue circles, Southeast Asia populations; red circles, South Asia populations; yellow circles, adimixed populations; green triangles, Fujian local populations; blank triangle, *N. muiri*). This figure is the enlarged part focusing on East Asia populations of Fig. 4. Genetically defined East Asia populations are indicated with population IDs.

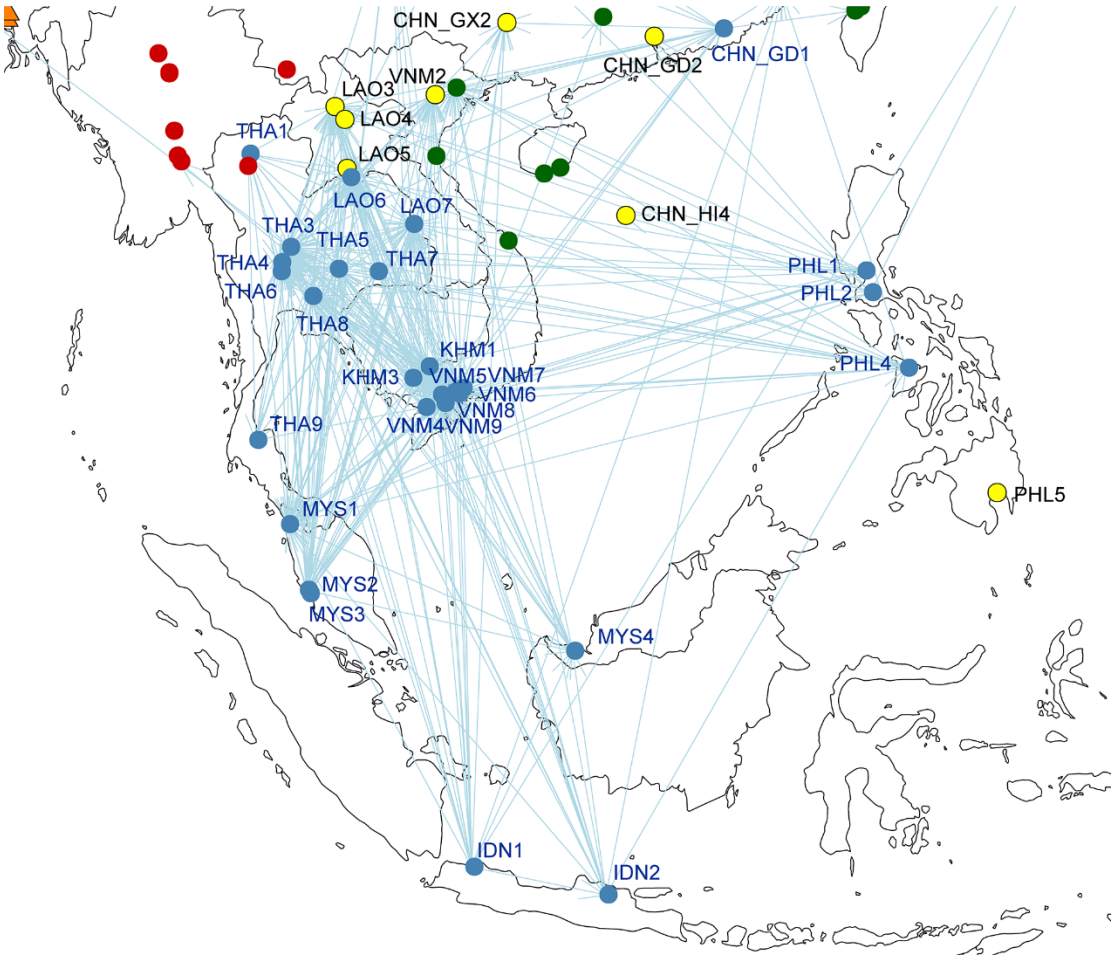

Fig. S10. The enlarged details of BPH population gene flows in Southeast Asian regions. Arrows in blue indicate inferred gene flows from a Southeast Asia population (blue circles) to the indicated population (green circles, East Asia populations; blue circles, Southeast Asia populations; red circles, South Asia populations; yellow circles, admixed populations). This figure is the enlarged part focusing on Southeast Asia populations of Fig. 4. Genetically defined Southeast Asia populations and admixed populations are indicated with population IDs.

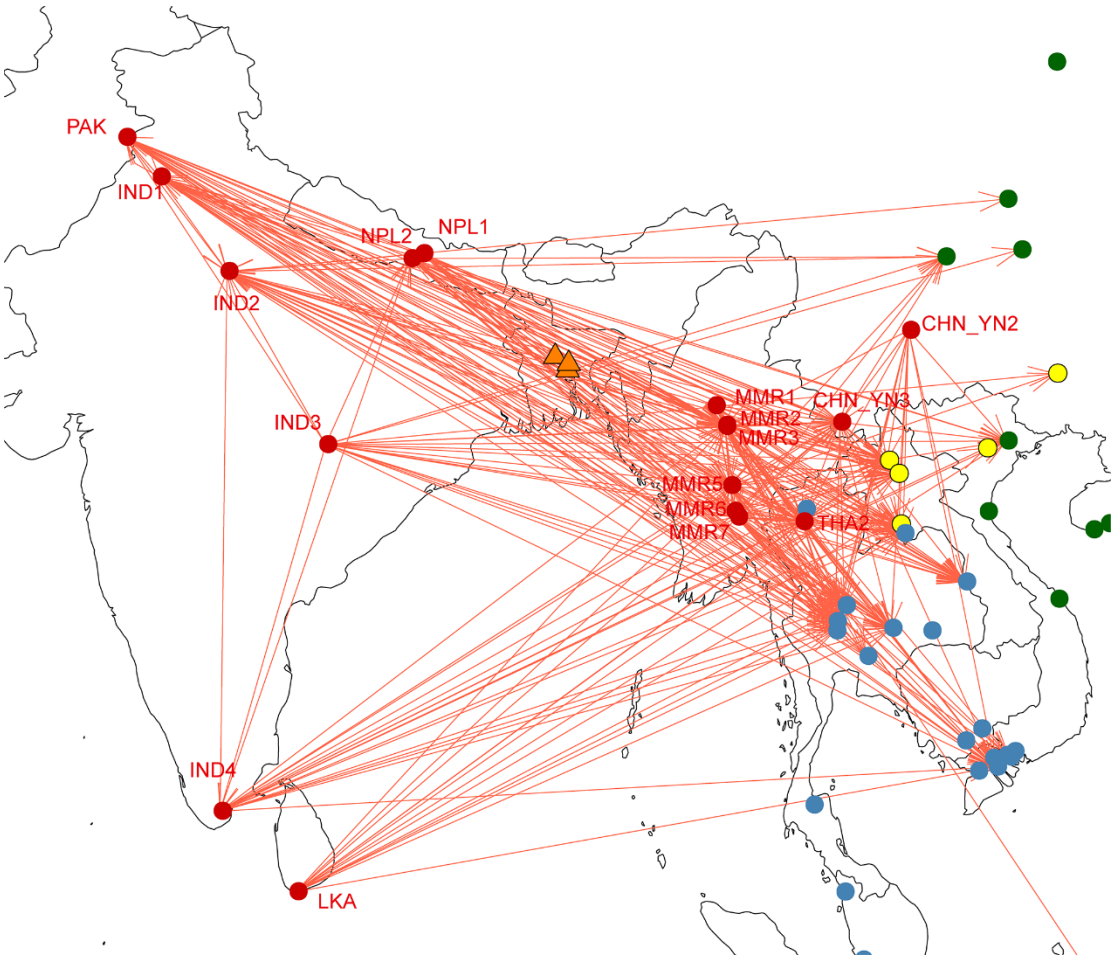

Fig. S11. The enlarged details of BPH population gene flows in SouthAsian regions. Arrows in red indicate inferred gene flows from a South Asia population (red circles) to the indicated population (green circles, East Asia populations; blue circles, Southeast Asia populations; red circles, South Asia populations; yellow circles, admixed populations; orange triangles, Bangladesh local populations). This figure is the enlarged part focusing on South Asia populations of Fig. 4. Genetically defined South Asia populations are indicated with population IDs.

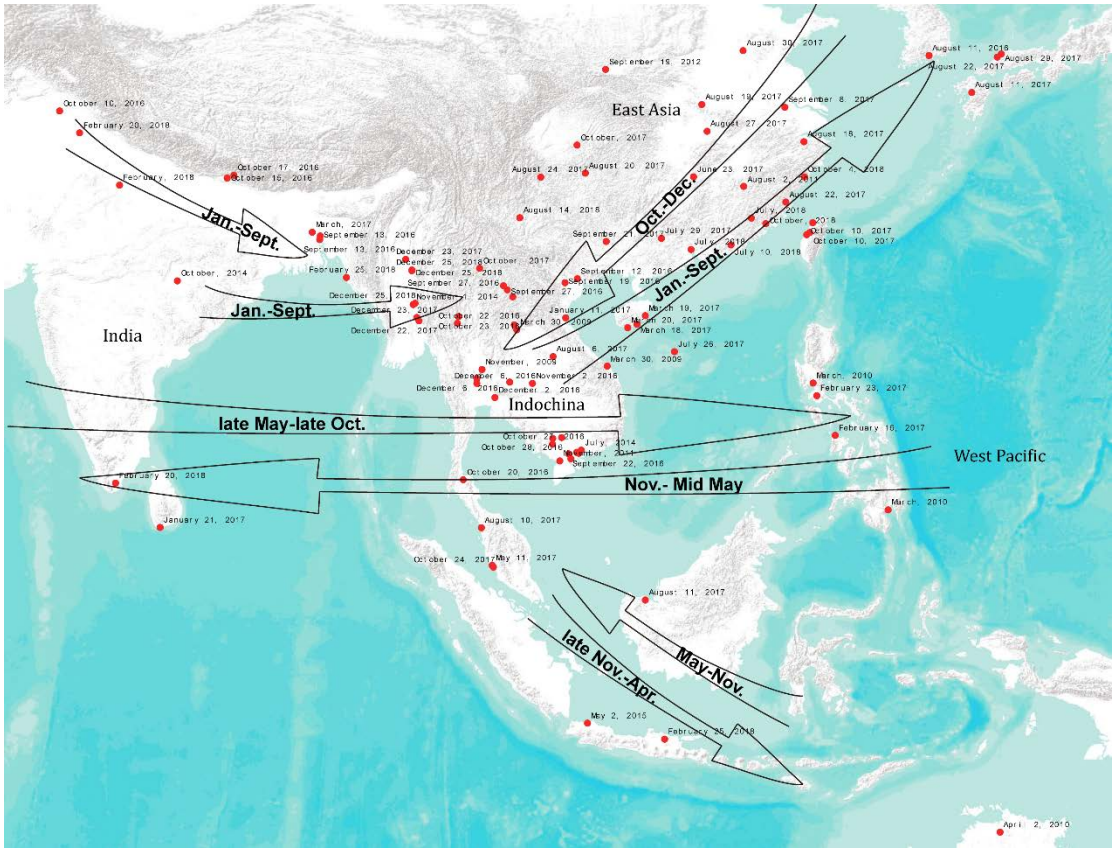

Fig. S12. Wind patterns in the context of sampling locations. Red points indicate sampling sites of this study, along with the collection date being labeled. The wind patterns (arrows) were summarized based on the air flow directions at 850hPa (usual flight altitude of BPH) from Feb, 2017 to Feb, 2018 (https://earth.nullschool.net/). Labeled time periods within arrows represent the months when the corresponding wind direction is prevailing. The southwest monsoon in East Asia occurs from January to September which supports the outbound migration from Indochina to East Asia, and the northeast winter monsoon occurs from October to December which supports the inbound migration from East Asia to Indochina (Fig. 4, Fig. S9). The air flows on West Pacific monsoon supports the gene flows between Malay Archipelago and Indochina (Fig. 4, Fig. S10). The India monsoon occurs from late May to late October supports the migration from South Asia to Indochina (Fig. 4, Fig. S11). The map is generated using ArcMap.

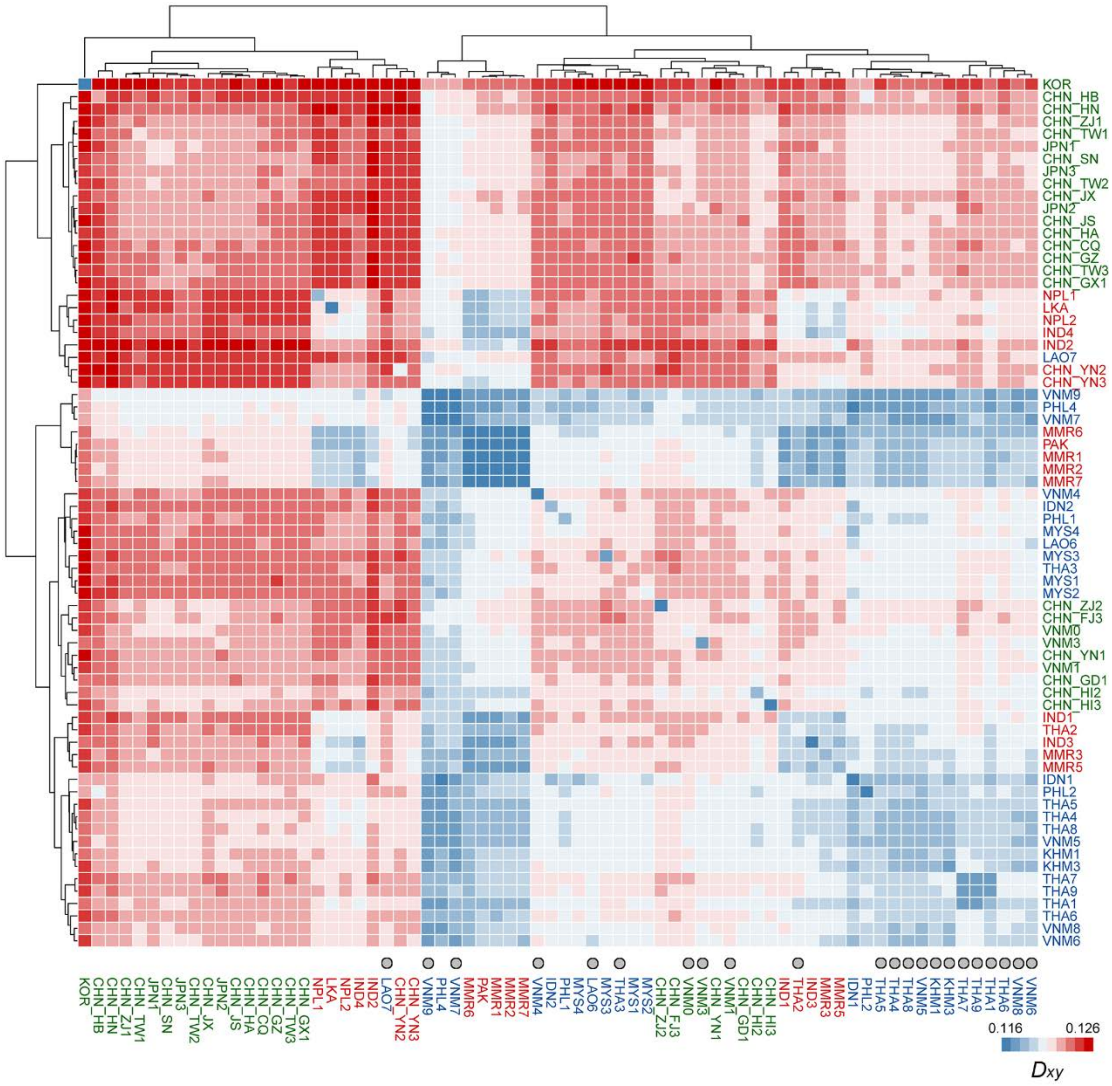

Fig. S13. Cross-population genetic differentiation based on *Dxy* between samples. Heatmap indicates the average *Dxy* between the pairwise individual comparisons of the corresponding two populations. The genetic source of each population is indicated in color. Green, East Asia populations, blue, Southeast Asia populations, red, South Asia populations. Grey circles indicate populations within the IndoChina Peninsula.

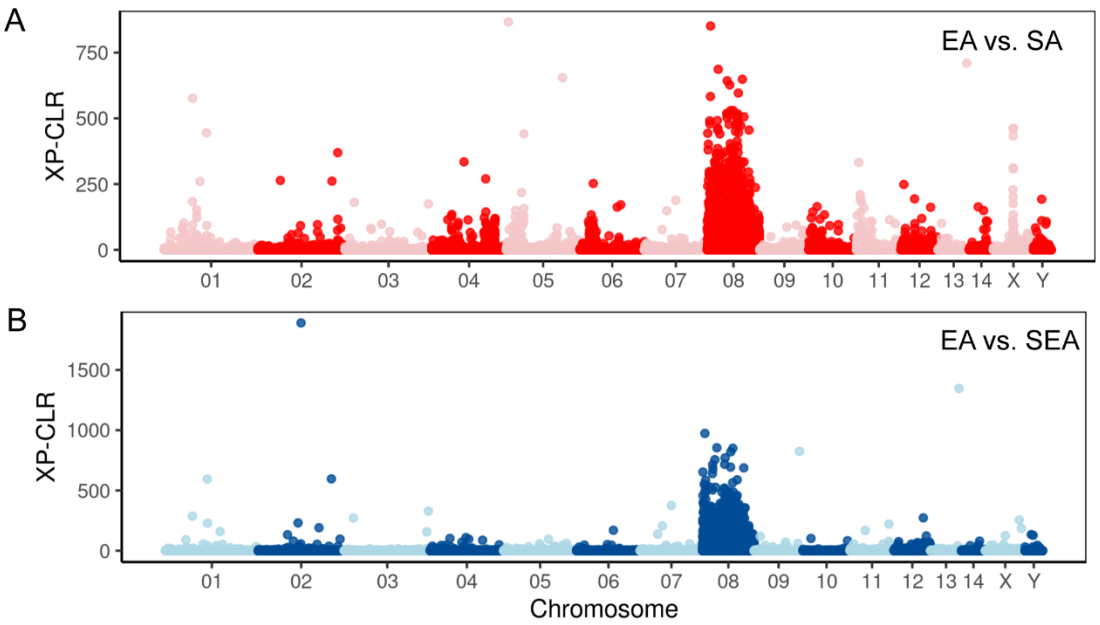

Fig. S14. Alternative genome-wide divergence between East Asia and other main groups. Genomic divergence signatures are based on cross-population composite likelihood ratio (XP-CLR). (A) XP-CLR between East Asia and South Asia groups. (B) XP-CLR between East Asia and Southeast Asia groups.

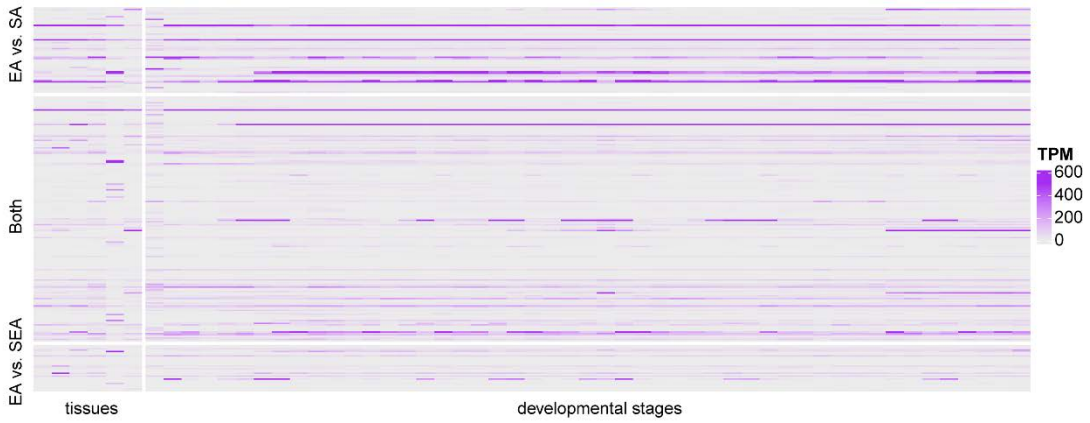

Fig. S15. Spatiotemporal expression profiles of genes under highest divergence. The heatmap indicates the expression value (estimated as transcripts per million, TPM) for genes. From above to bottom, genes under highly diverged between East Asia and South Asia groups, between both comparsions, and between East Asia and Southeast Asia groups.

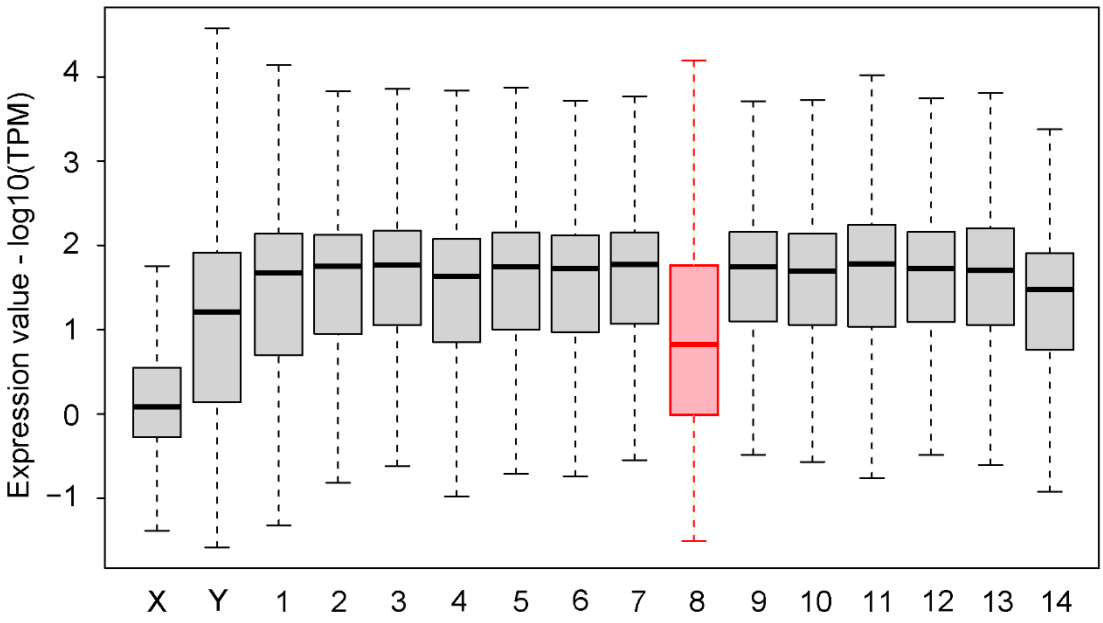

Fig. S16. The overall expression level of protein-coding gene across different BPH chromosomes. The expression profile is based on 55 transcriptomes of BPH developments stages and tissues (Methods). Gene expression is represented as Transcripts Per Kilobase of exonmodel per Million mapped reads (TPM) and log10-transformed. Distribution of chromosome-wide gene expression is represented by box-whisker plot, in which the horizontal line in the middle represent the median, the two horizontal lines above and below the box represent the quartiles, and the two whiskers show the 1.5x interquartile expression. Chromosome 8 is highlighted in red.

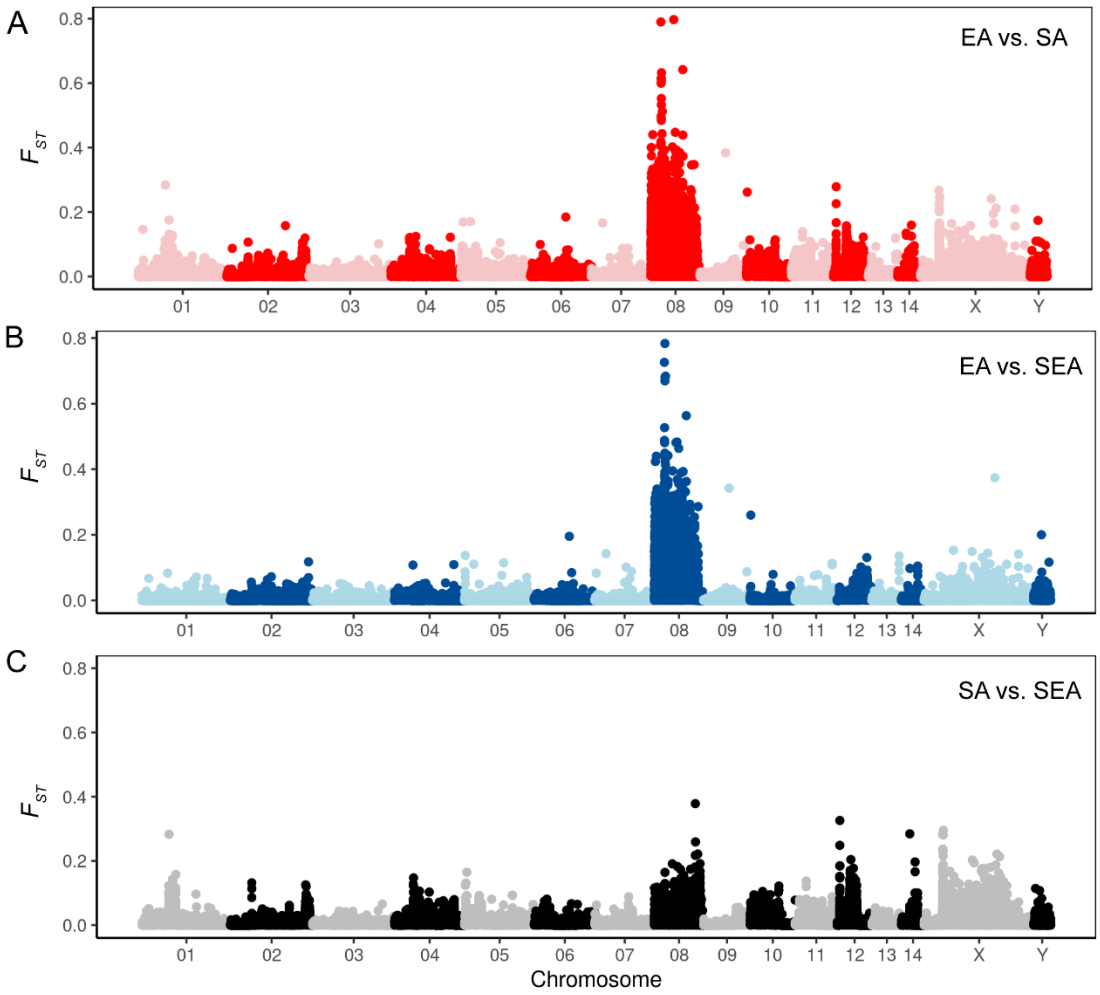

Fig. S17. Genome-wide divergence among main group. In comparison to Fig. 5B, F_{ST} values among groups were calculated by excluding variations within repetive regions, as a quality control for the potential bias by highly repetitive content on chromosome 8. (A) F_{ST} between East Asia and South Asia groups. (B) F_{ST} between East Asia and Southeast Asia groups. (C) F_{ST} between Southeast Asia and South Asia groups.

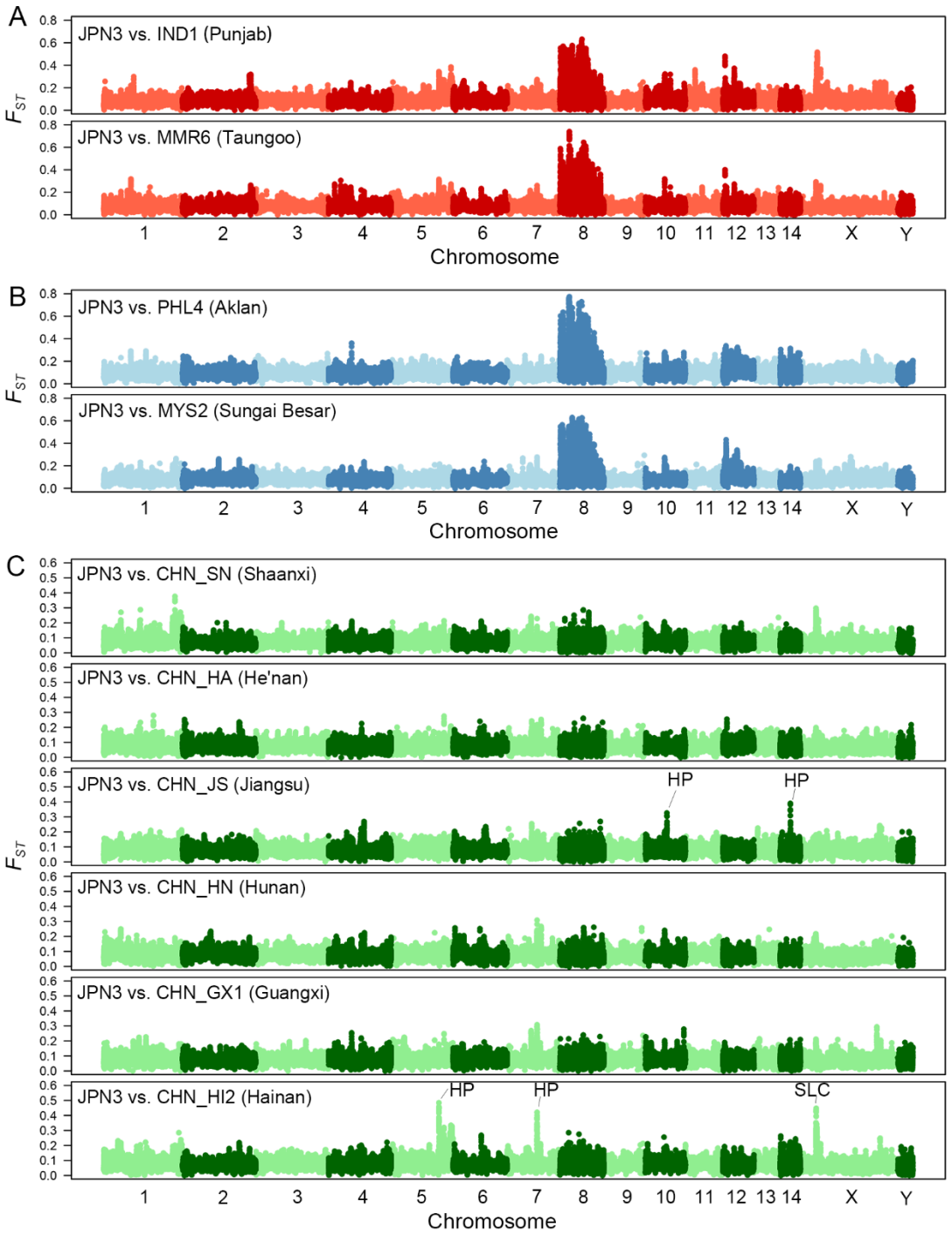

Fig. S18. Genome-wide divergence between the most northeast East Asia population with other representative populations. Divergence signatures (F_{ST}) are plotted along chromosomes between a Japan population (JPN3) and representative populations from South Asia (A), Southeast Asia (B), and East Asia (C). F_{ST} values were calculated on 50-Kb sliding windows with an interval of 5-Kb. Several divergence peaks across different comparisons are indicated by the encompassing gene. HP, hypothetical protein; SLC, solute carrier family 26 member 10.

Data S1. (separate file)

Metadata of sampling, sequencing, and genetic grouping for all samples subject for this study.

Data S2.

Statistic tests for various genomic features between chromosome 8 and other chromosomes.

Data S3.

Genomic regions with the highest 1% divergence signature between EA versus SEA and EA versus SA.

Data S4.

Genes contained in the highest 1% diverged regions.

Data S5.

Kinship analysis for each pair of individuals.