

Supplemental information

Multiple introductions and recombination events underlie the emergence of a hyper-transmissible *Cryptosporidium hominis* subtype in the USA

Wanyi Huang, Yaqiong Guo, Colleen Lysen, Yuanfei Wang, Kevin Tang, Matthew H. Seabolt, Fengkun Yang, Elizabeth Cebelinski, Olga Gonzalez-Moreno, Tianyi Hou, Chengyi Chen, Ming Chen, Muchun Wan, Na Li, Michele C. Hlavsa, Dawn M. Roellig, Yaoyu Feng, and Lihua Xiao

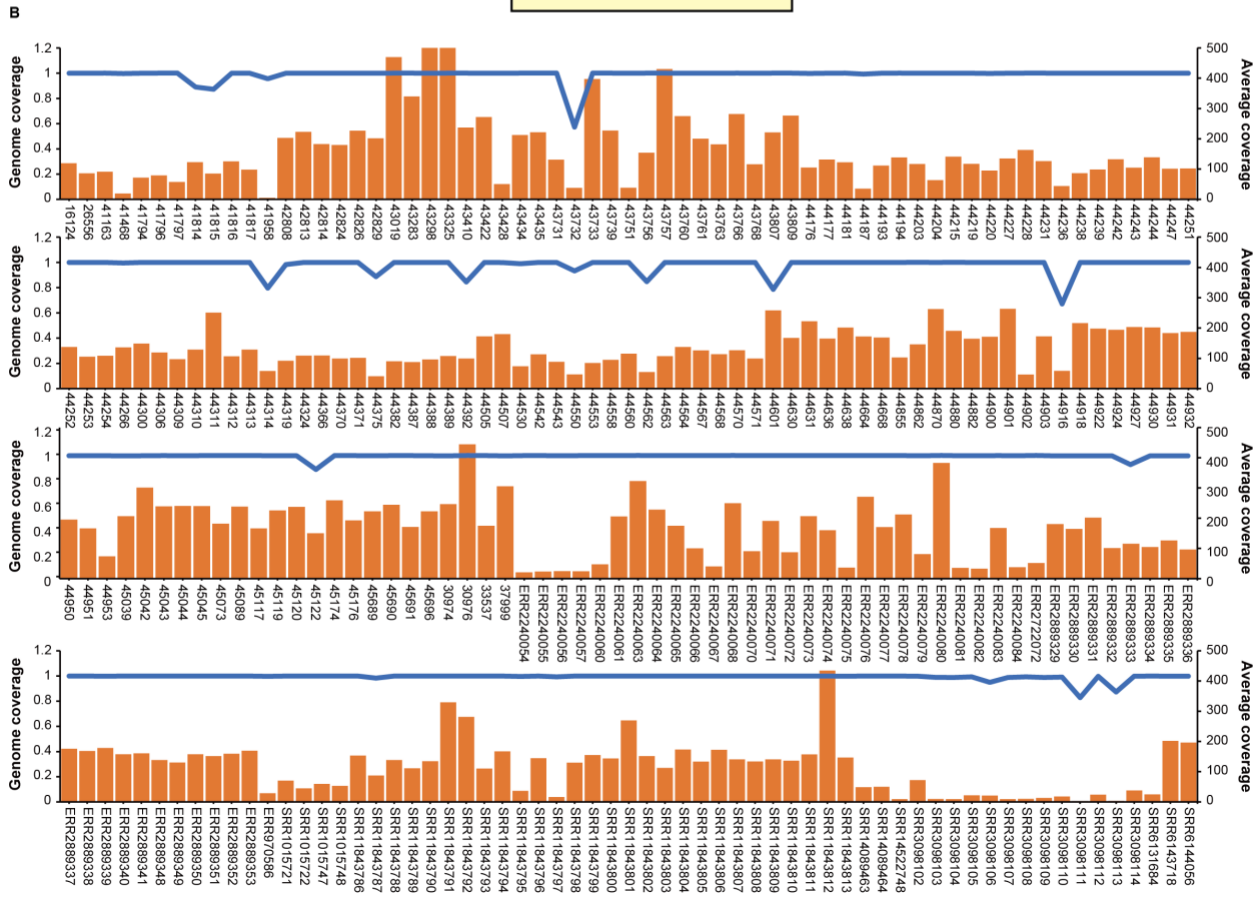
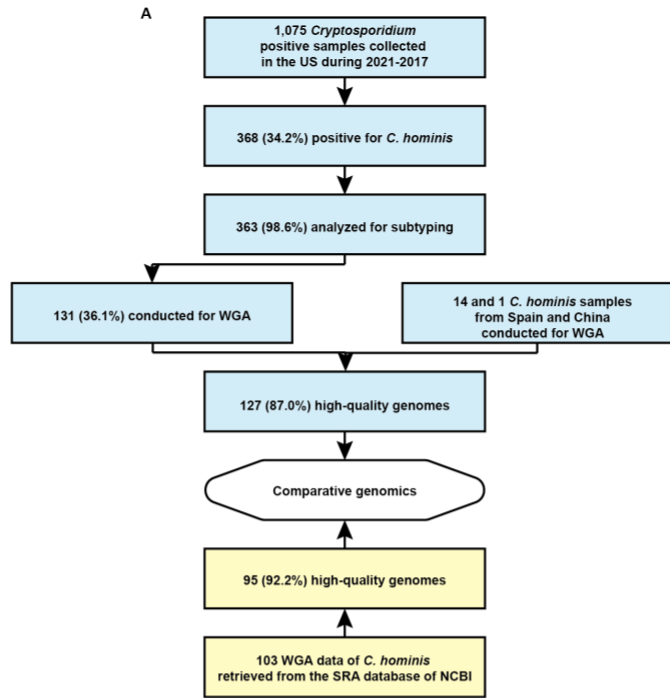
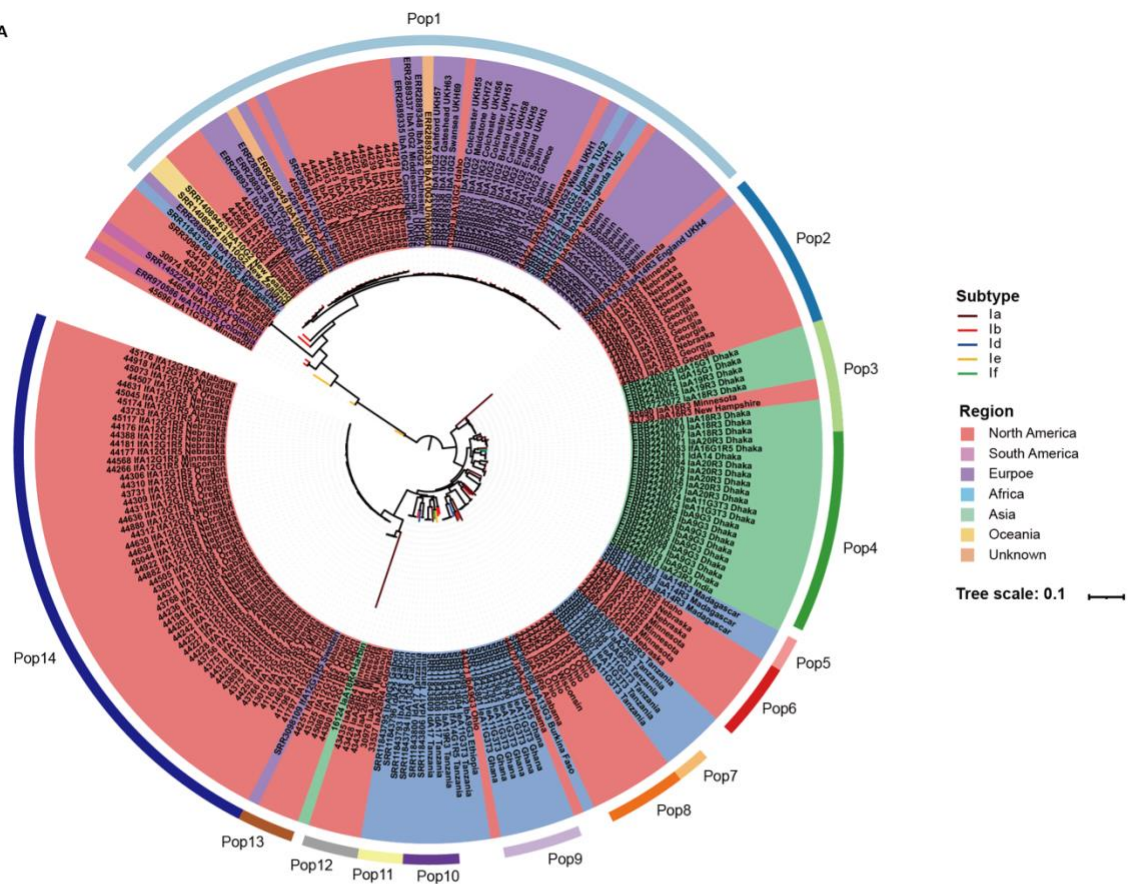


Figure S1. Acquisition of new whole genome sequence data from *Cryptosporidium hominis* for this study, Related to Figure 1 and Table S1.

(A) A workflow of *C. hominis* sampling, subtyping, sequencing, and data retrieving. (B) Quality of sequencing data based on the coverage of reference *C. hominis* genome 30976 (blue line) and the sequencing depth (orange bar). Genomes with coverage below 90% and sequencing depth below 5 were excluded from further analyses.

A



B

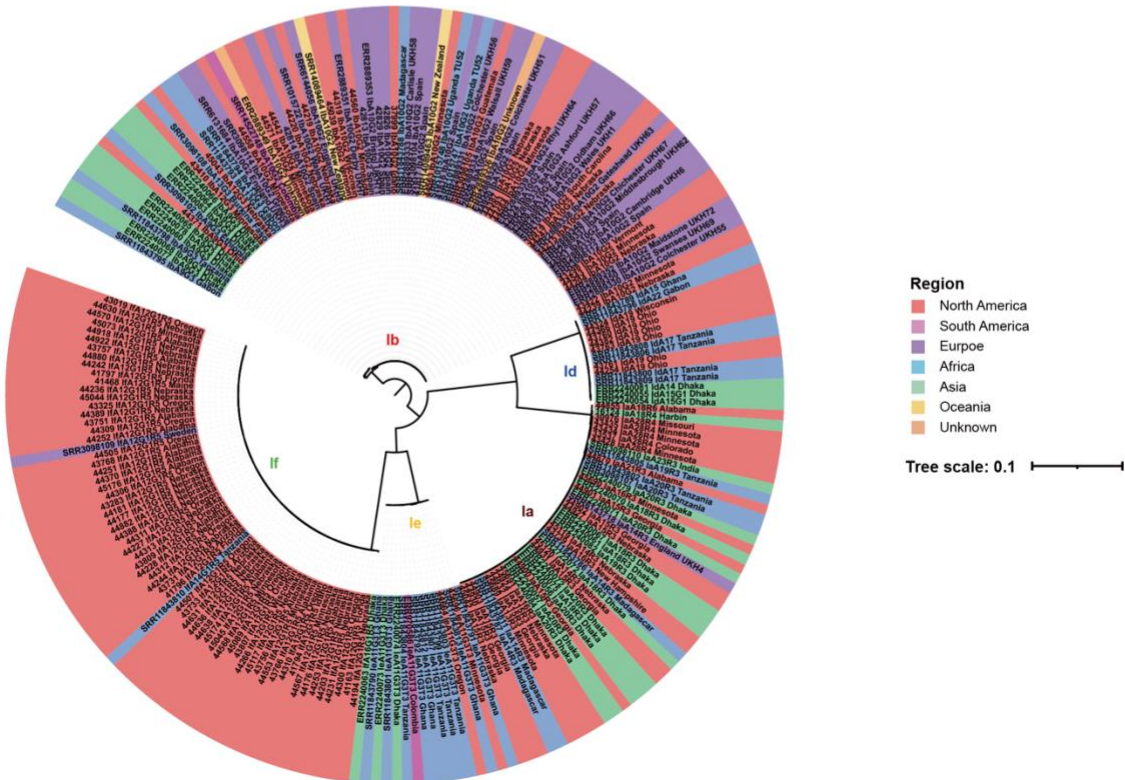


Figure S2. Population subdivision within *C. hominis*, Related to Figure 2.

(A) Phylogenetic relationship inferred by ML analysis of 12,736 wgSNPs. The branch colors represent different subtypes, while the background colors of the sample names represent the sample sources. The *C. hominis* isolates form 14 clades as in Figure 2C. (B) Phylogenetic analysis of the *gp60* gene among 222 isolates, based ML analysis with the GTR+I model. The background colors of the sample names represent the sample sources (colored the same as in A).

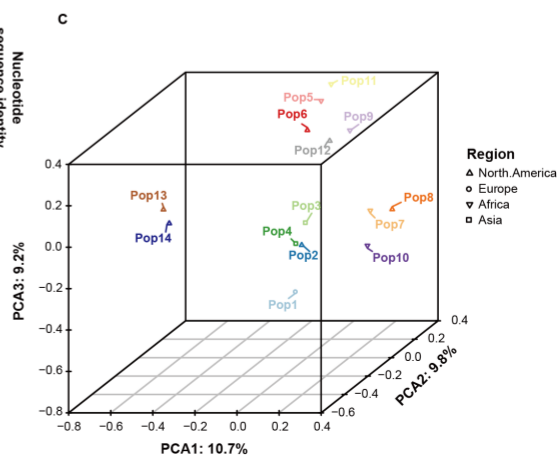
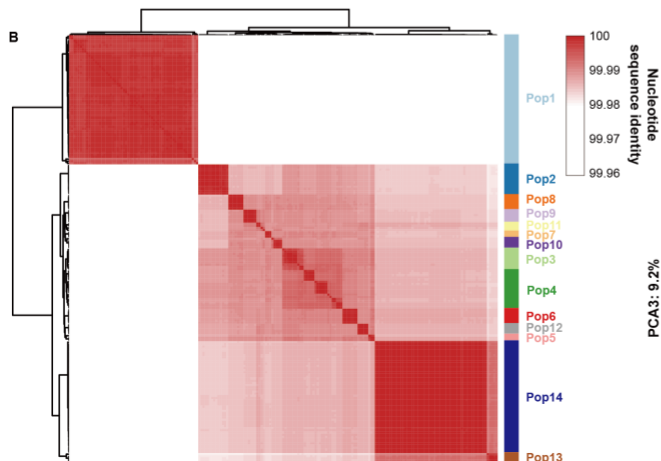
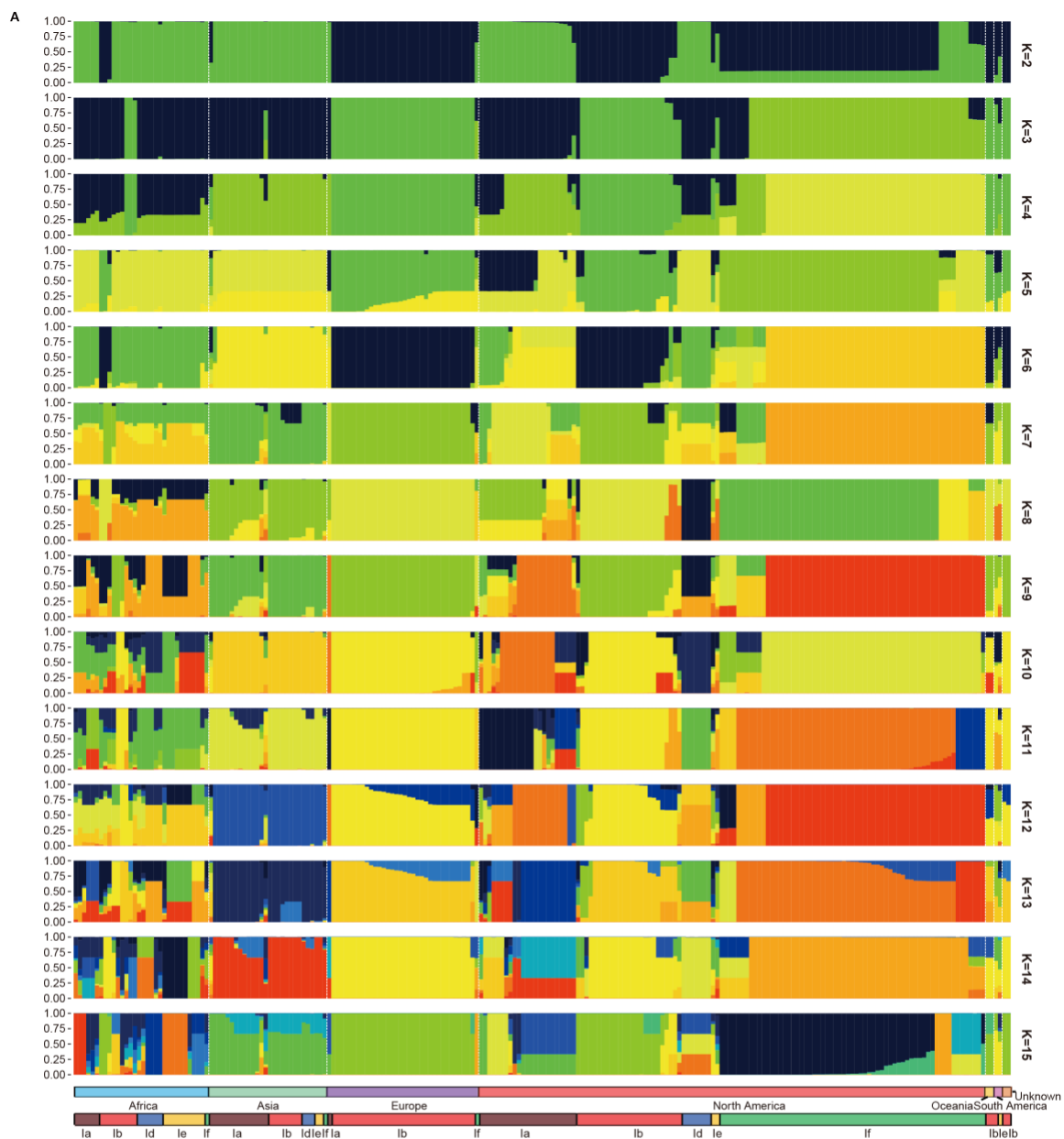


Figure S3. Genetic structure of 14 *C. hominis* populations, Related to Figure 2.

(A) STRUCTURE plots constructed using $K = 2$ to 15 indicate the representing the percentage of shared ancestry among the *C. hominis* metapopulation. (B) Nucleotide sequence identity between each pair of *C. hominis* genomes (colored them the same as in Figure 2C). (C) Outcome of the PCA analysis of 7,823 SNPs among 14 representative isolates of the 14 populations (colored the same as in Figure 2C).

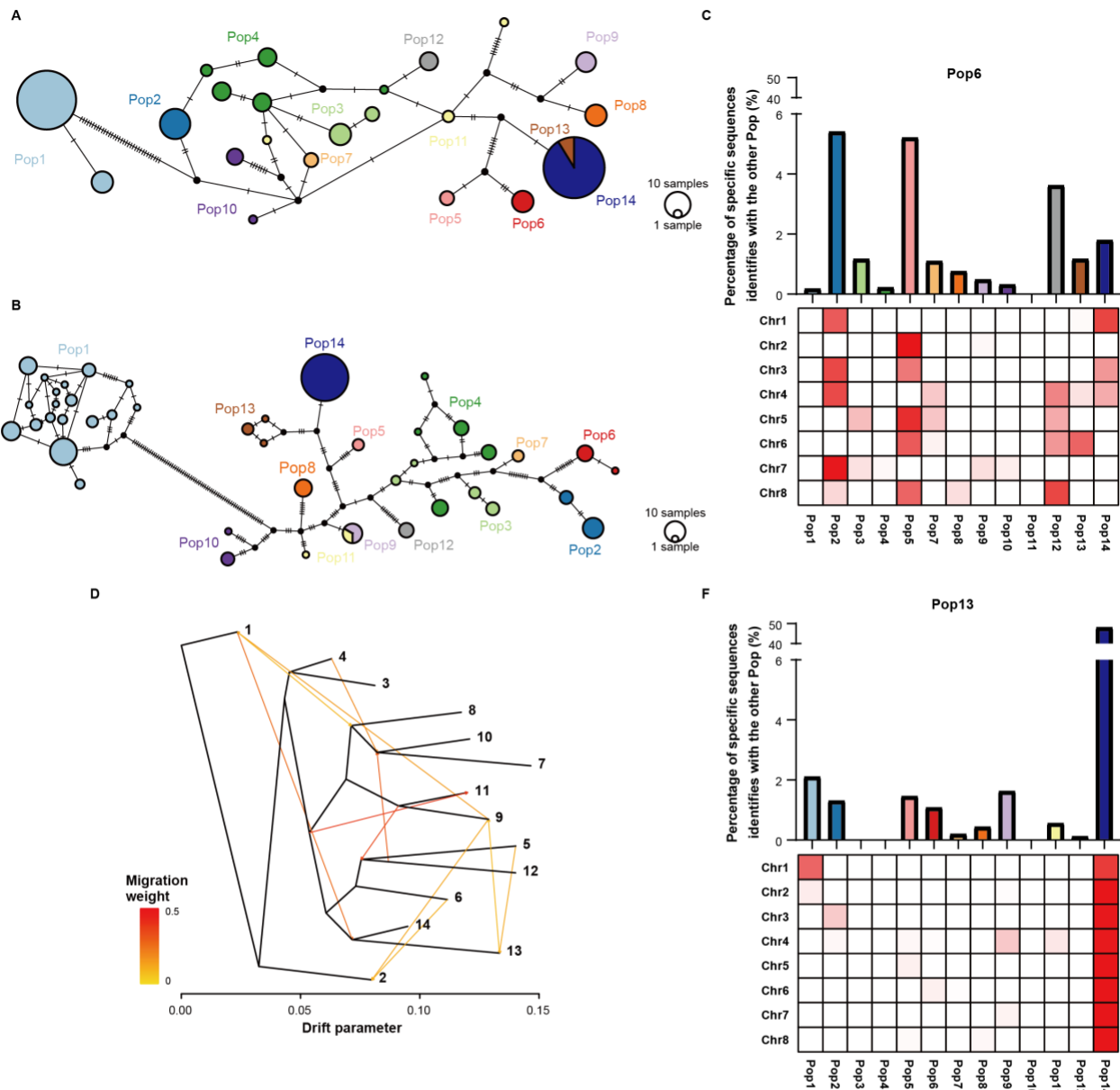


Figure S4. Multiple sequence introgression in the formation of Pop6 and Pop13, Related to Figures 3 and 4.

(A and B) Haplotype networks of linkage disequilibrium blocks in chromosomes 2 and 7 (Chr2: 199,090-303,437 and Chr7: 337,784-537,673). They illustrate different genetic relationships among populations between the blocks. (C) Introgressed regions in Pop 6 by chromosome. Windows with *Fst* value of 0 between Pop6 and the other populations were considered the introgressed regions. The percentages of introgressed regions are displayed in the top panel, while their distribution by chromosome is shown in the bottom panel. (D) Migration events between populations based on TreeMix inference with migration number (*m*) of 11. The colored lines indicate possible migrations. (E) Relationship of representative isolates of the 14 populations in chromosome 1 based on phylogenetic analysis (left panel) and alignment (right panel) of 723 high-quality SNPs. In the partial alignment of SNPs, identical nucleotides are colored as the same. (F) Introgressed regions in Pop13 by chromosome. Windows with *Fst* value of 0 between Pop13 and the other populations were considered the introgressed regions. The percentages of introgressed regions are displayed in the top panel, while their distribution by chromosome is shown in the bottom panel.

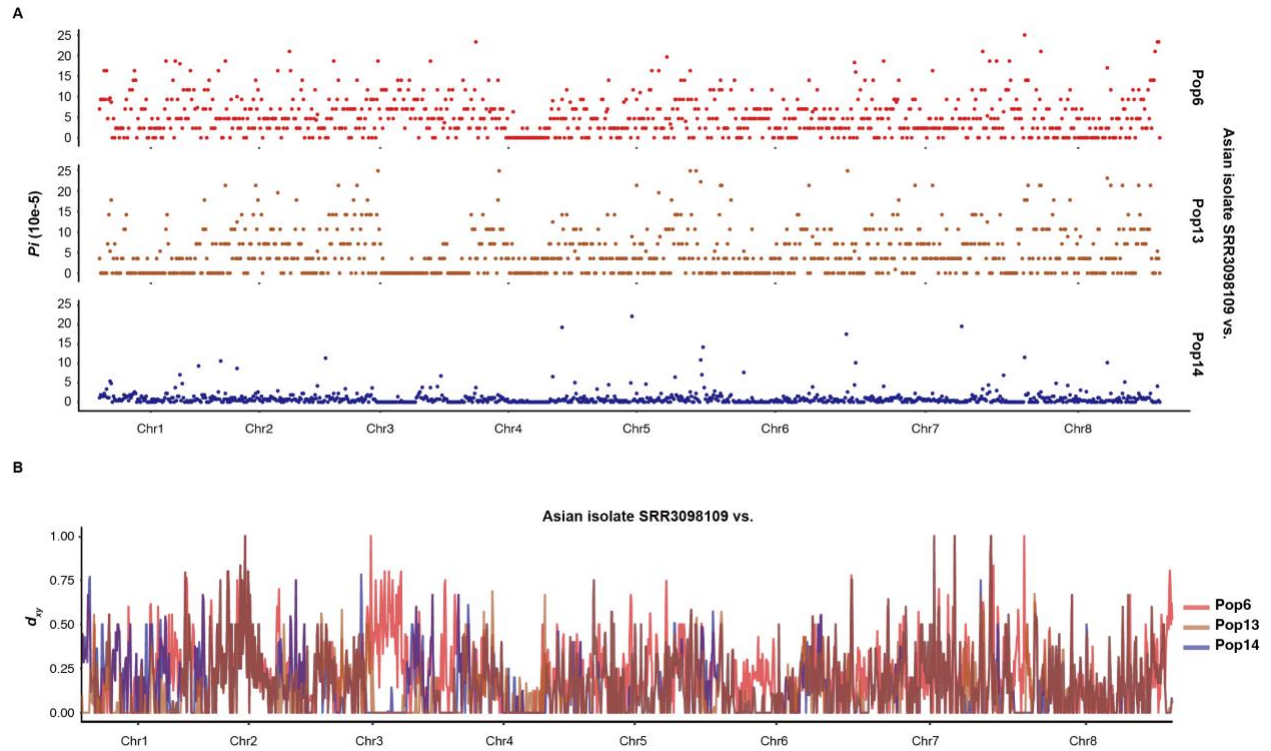


Figure S5. Genetic uniqueness of the Chinese IaA18R4 isolate compared to IfA12G1R5 isolates, Related to Figure 4.

(A) Nucleotide diversity (P_i) between the Asian and U.S. isolates of the IfA12G1R5 using a 10-kb window. (B) Absolute divergence (d_{xy}) between the Chinese IaA18R4 isolate and the three IfA12G1R5 populations (Pop6, Pop13, and Pop14) across the eight chromosomes.

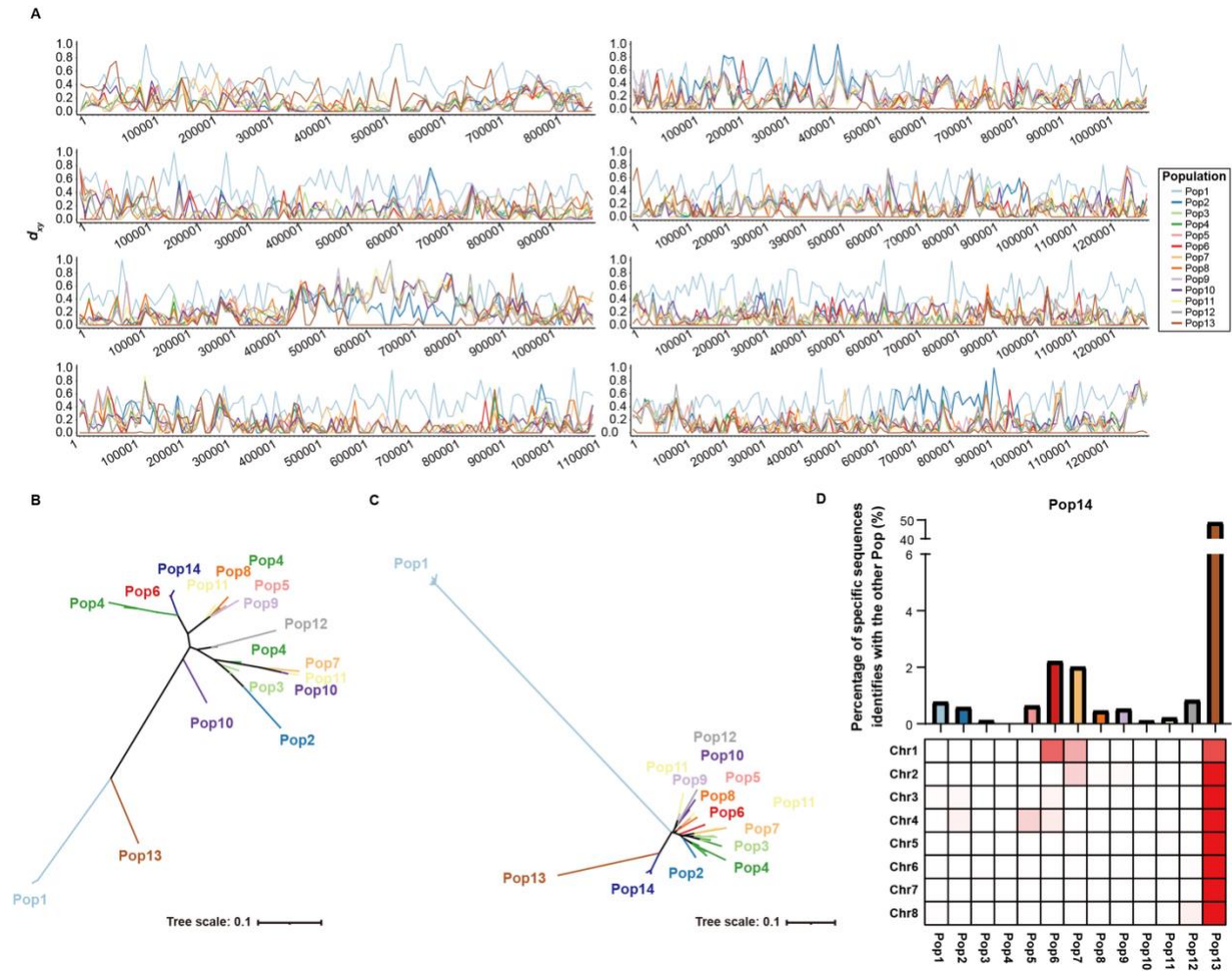


Figure S6. Occurrence sequence introgression in Pop14, Related to Figure 5.

(A) Absolute divergence (d_{xy}) between Pop14 and other populations across the eight chromosomes. (B and C) Phylogenetic relationship of SNPs in the region in the first box of Figure 5C and the entire chromosome 1 based on ML analysis using the TPM1uf and GTR+I+G model. (D) Windows with F_{st} value of 0 between Pop14 and other populations were considered introgressed regions. The percentages of introgressed regions from the other populations to Pop14 are displayed in the top panel, while their distribution by chromosome is shown in the bottom panel.