To reconstruct the demographic history for the global spread of the German cockroach, we adopted coalescence-based analyses using site frequency spectrum (SFS) as implemented in *fastsimcoal2* v.2.7. We selected nine sampling sites to be involved in the spread route reconstruction based on the population genetic structure and ancestry estimation. As the reference genome is only available for the German cockroach, we computed folded SFS excluding missing data, linked loci, and monomorphic sites using *python* script *easySFS*. Given that *fastsimcoal2* allows a maximum of six populations in each scenario, we adopted a stepwise strategy, which comprises four stages in this study. The resulting topology of divergence and estimated parameters in the previous stage are used as prior in the next stage. Each stage may comprise one to four panels depending on the combinations of individuals involved (Table S1). Each panel contains three to six scenarios with the same individuals but different divergence topologies to be simulated. We performed 50 runs for each scenario. We assumed four generations per year and a mutation rate of 2.8×10^{-9} per site per generation as estimated from *Drosophila melanogaster*. Considering the relative homogeneity in the ancestry pattern of individuals in the selected sampling sites and the nature that established German cockroach populations are unlikely to accept newcomers, we assumed no geneflow after the divergence throughout the demographic models.

Table S1. Sample composition of panels used for demographic modelling. "○" means the individuals in the sampling site were included in the observed site frequency spectrum, whereas "×" means only demographic parameters were used to represent the individuals in the sampling site.

	PRE Stage I	IRA Stage II	CHN Stage II	ETP Stage III	UKR Stage III	COH Stage III	NDL Stage III	WST Stage IV	EST Final	WLD Final
Addis Ababa, Ethiopia				○						○
Bandung, Indonesia	○	○	○	○	○	○	○	×	○	×
Bijlmer, Netherlands							○	○		○
Busan, Korea	○	○	○	○	○	○	○	×	○	×
Chongqing, China			○						○	×
Crimea, Ukraine					○			○		○
Raleigh, USA						○		○		○
Tehran, Iran		○		○	○	○	○	○	○	○
Vijayawada, India	○	○	○	○	○	○	○	×	○	○

Demographic modelling starts with the inclusion of only individuals from three sampling sites (panel PRE, stage I) exhibiting the highest divergence in the PCA plot (Busan-Korea, Bandung-Indonesia, and Vijayawada-India). All three topologies of divergence have been tested (scenario A, B, and C). According to the delta likelihood, the most plausible scenario would be that the individuals in India diverged, then individuals in Korea and Indonesia (Fig. S1).
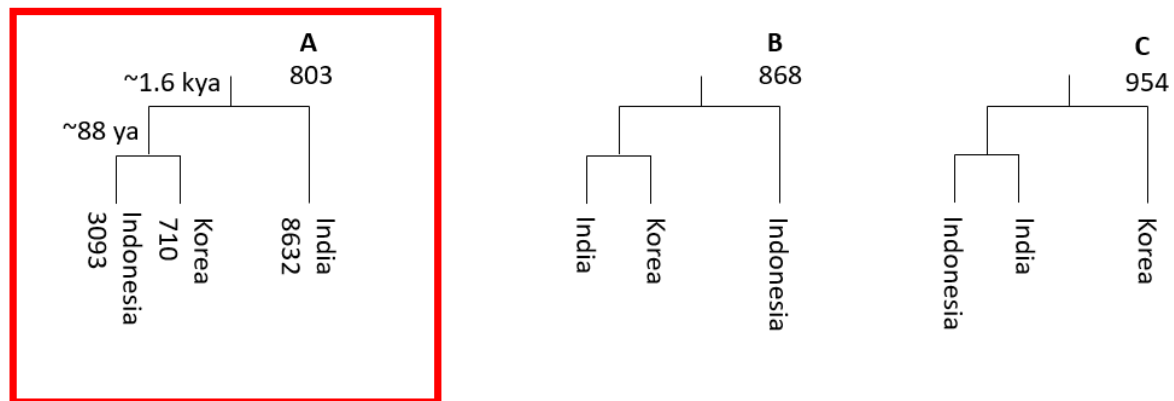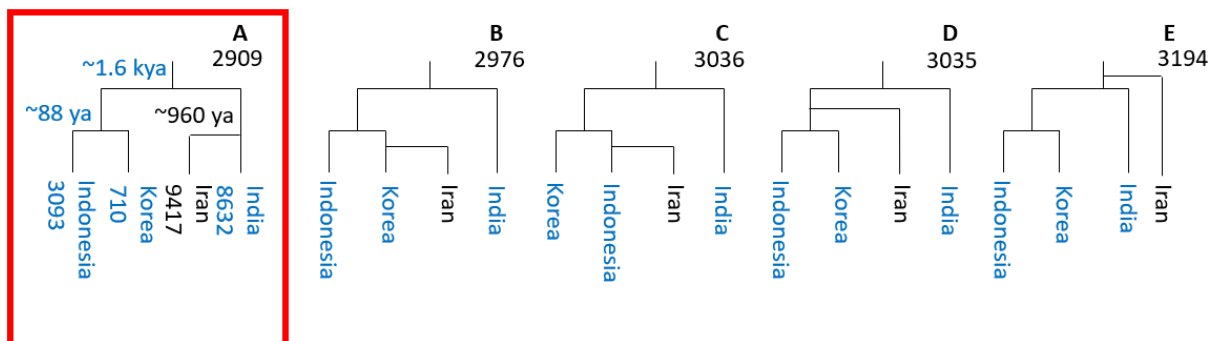

Fig. S1. Schematic description of panel PRE, stage I. Individuals from Busan-Korea, Bandung-Indonesia, and Vijayawada -India are included in this panel. Three topologies of divergence (scenario A, B, and C) were tested. The best scenario is highlighted with a red box. Delta likelihood values are below the scenario names. Estimated population effect sizes are to the left of the population name, whereas the estimated time of divergence is placed at each node.

In stage II, we aimed to investigate the relationship among the six identified ancestries. In addition to the individuals representing three ancestries involved in stage I, we added individuals in China and Iran (from the co-ancestor of the Eastern Europe and USA ancestries). Therefore, stage II comprises two panels: CHN (Fig. S2) and IRA (Fig. S3) with each panel comprising five scenarios (scenario A, B, C, D, and E).
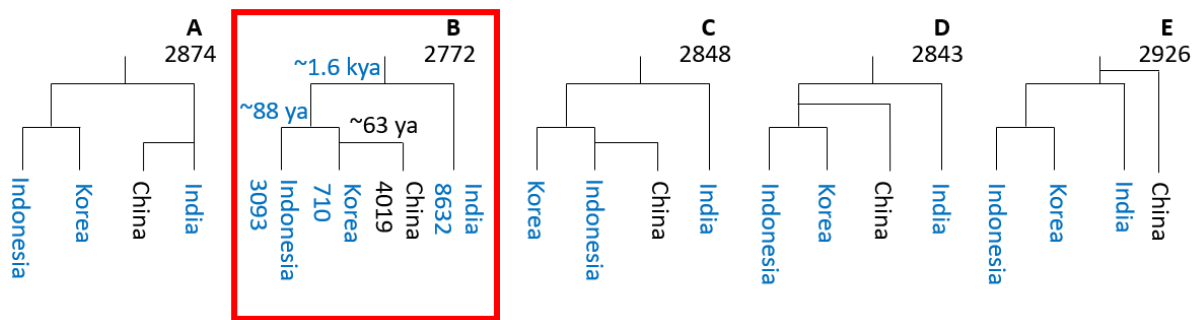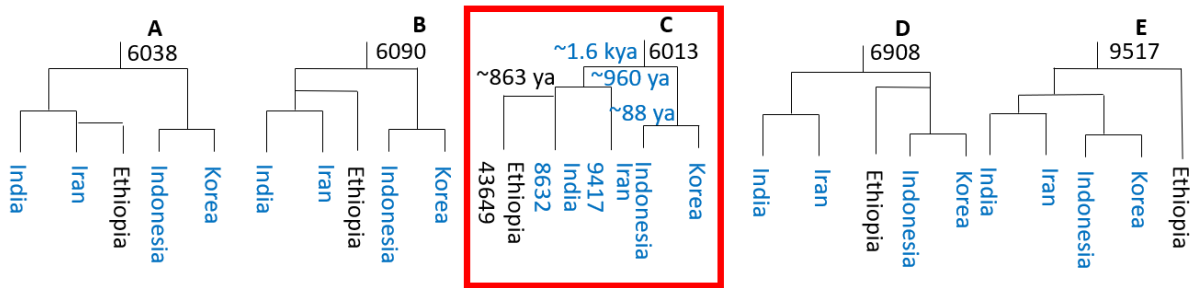
Fig. S2. Schematic description of panel CHN, stage II. Individuals from Busan-Korea, Bandung-Indonesia, Chongqing-China, and Vijayawada-India are included in this panel. Five topologies of divergence (scenario A, B, C, D, and E) were tested. The best scenario is highlighted with a red box. Delta likelihood values are below the scenario names. Estimated population effect sizes are to the left of the population name, whereas the estimated time of divergence is placed at each node. Parameters used from the estimates of the previous stage are in blue font.
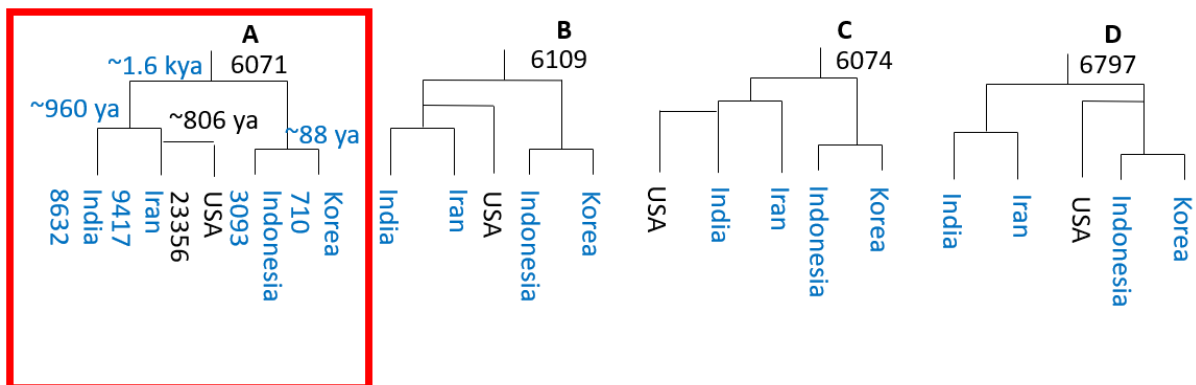

Fig. S3. Schematic description of panel IRA, stage II. Individuals from Busan-Korea, Bandung-Indonesia, Tehran-Iran, and Vijayawada-India are included in this panel. Five topologies of divergence (scenario A, B, C, D, and E) were tested. The best scenario is highlighted with a red box. Delta likelihood values are below the scenario names. Estimated population effect sizes are to the left of the population name, whereas estimated time of divergence is placed at each node. Parameters used from the estimates of the previous stage are in blue font.

In stage III, for each sampling site that we didn't use in the previous stages, we placed the individuals to the topology we got from the previous stage. This stage comprises four panels: ETP (Fig. S4), COH (Fig. S5), UKR (Fig. S6), and NDL (Fig. S7).
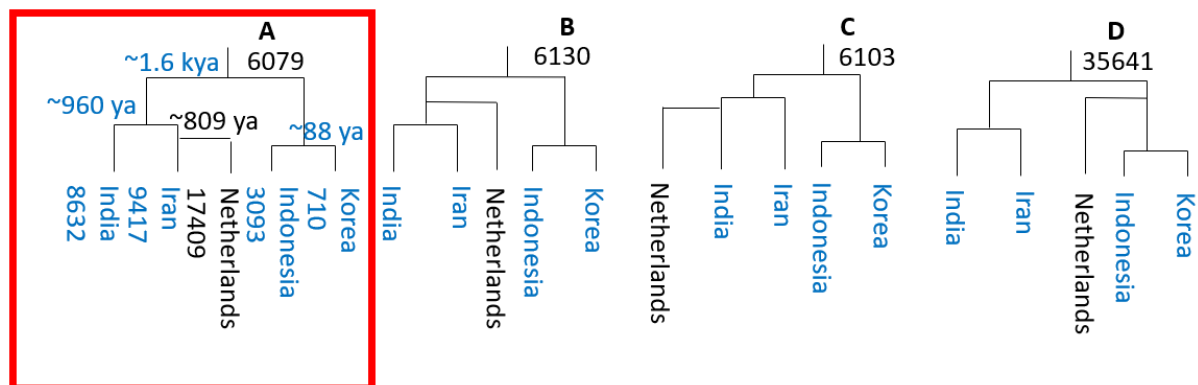
Fig. S4. Schematic description of panel ETP, stage III. Individuals from Addis Ababa-Ethiopia, Busan-Korea, Bandung-Indonesia, Tehran-Iran, and Vijayawada-India are included in this panel. Five topologies of divergence (scenario A, B, C, D, and E) were tested. The best scenario is highlighted with a red box. Delta likelihood values are below the scenario names. Estimated population effect sizes are to the left of the population name, whereas estimated time of divergence is placed at each node. Parameters used from the estimates of the previous stage are in blue font.

Fig. S5. Schematic description of panel COH, stage III. Individuals from Busan-Korea, Bandung-Indonesia, Raleigh-USA, Tehran-Iran, and Vijayawada-India are included in this panel. Four topologies of divergence (scenario A, B, C, and D) were tested. The best scenario is highlighted with a red box. Delta likelihood values are below the scenario names. Estimated population effect sizes are to the left of the population name, whereas estimated time of divergence is placed at each node. Parameters used from the estimates of the previous stage are in blue font.
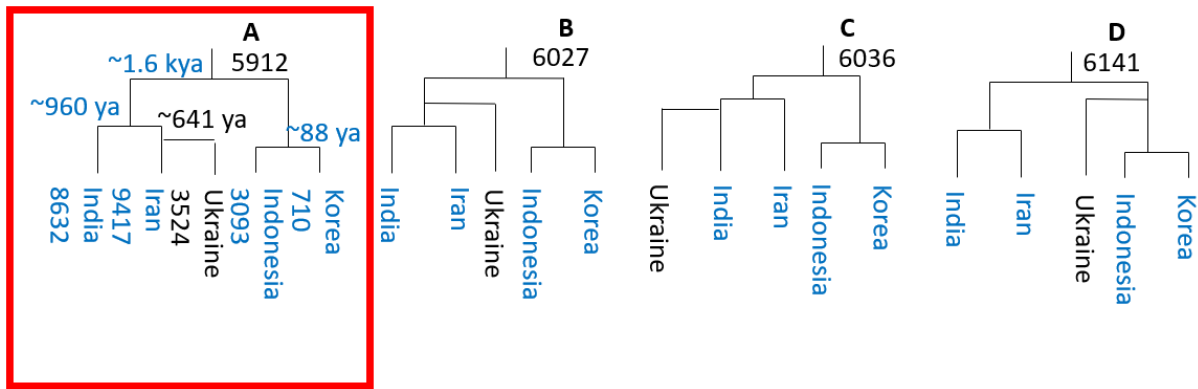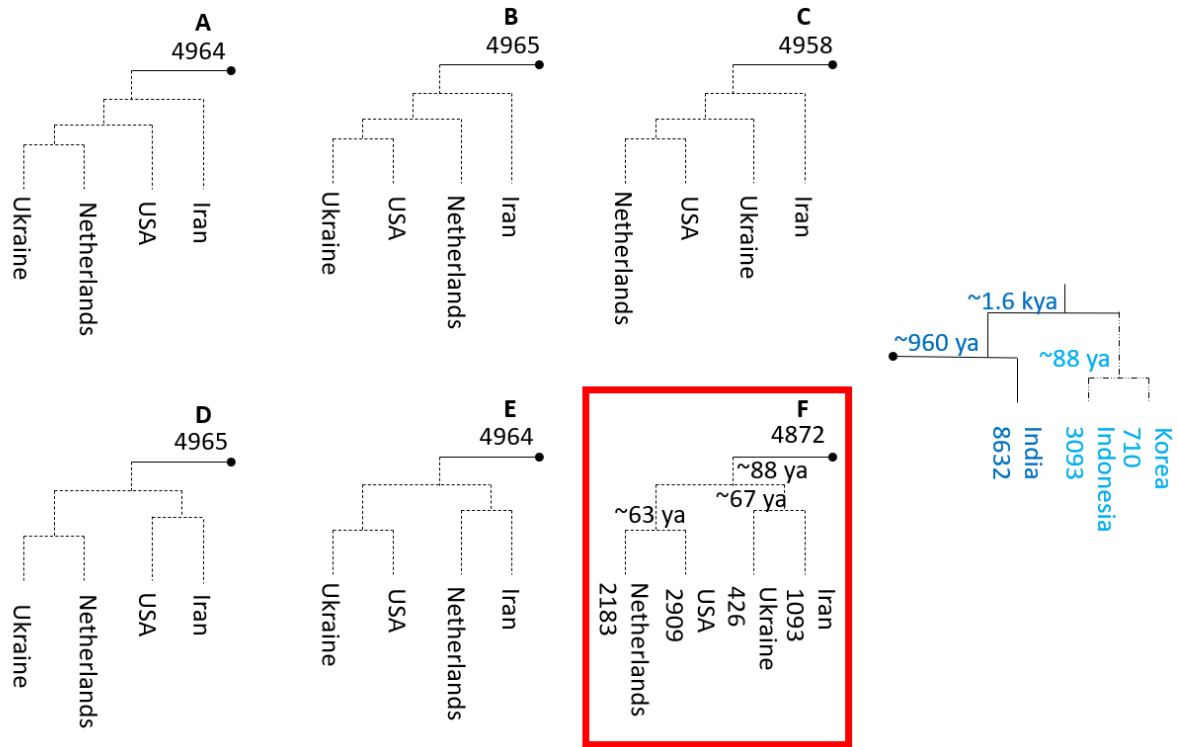
Fig. S6. Schematic description of panel UKR, stage III. Individuals from Busan-Korea, Bandung-Indonesia, Crimea-Ukraine, Tehran-Iran, and Vijayawada-India are included in this panel. Four topologies of divergence (scenario A, B, C, and D) were tested. The best scenario is highlighted with a red box. Delta likelihood values are below the scenario names. Estimated population effect sizes are to the left of the population name, whereas estimated time of divergence is placed at each node. Parameters used from the estimates of the previous stage are in blue font.

Fig. S7. Schematic description of panel NDL, stage III. Individuals from Bijlmer-Netherlands, Busan-Korea, Bandung-Indonesia, Tehran-Iran, and Vijayawada-India are included in this panel. Four topologies of divergence (scenario A, B, C, and D) were tested. The best scenario is highlighted with a red box. Delta likelihood values are below the scenario names. Estimated population effect sizes are to the left of the population name, whereas estimated time of divergence is placed at each node. Parameters used from the estimates of the previous stage are in blue font.

Through the three stages, we clarified the sequences of divergence among the tested individuals across nine geographical realms. The global spread of the German cockroach has been consolidated with two primary routes, one to the east involving China, Indonesia, and Korea ancestries, the other to the west involving the rest of the ancestries. Stage IV aims to investigate the detailed placement of populations in the west route with more comprehensive involvement of samples. Therefore, stage IV has only one panel with six scenarios (Fig. S8).

Fig. S8. Schematic description of panel WST, stage IV. Individuals from Bijlmer-Netherlands, Crimea-Ukraine, Raleigh-USA, and Tehran-Iran are included in this panel. Six topologies of divergence (scenario A, B, C, D, E, and F) were tested. The best scenario is highlighted with a red box. Delta likelihood values are below the scenario names. Estimated population effect sizes are to the left of the population name, whereas estimated time of divergence is placed at each node. Parameters used from the estimates of the previous stage are in blue font.

At the final stage, with the consolidated topology of population divergence reconstructed, we carried out demographic parameter estimations and parametric bootstrapping. The final stage comprises two panels: EST and WLD (Fig. S9). Each panel comprises only one scenario. We performed 50 runs for

each scenario for the initial parameter estimation. The best run of each scenario was used as the input file for the parametric bootstrapping, which iterated 100 times. The resulting parameters of all iterations were collected to calculate medians and confidence intervals.
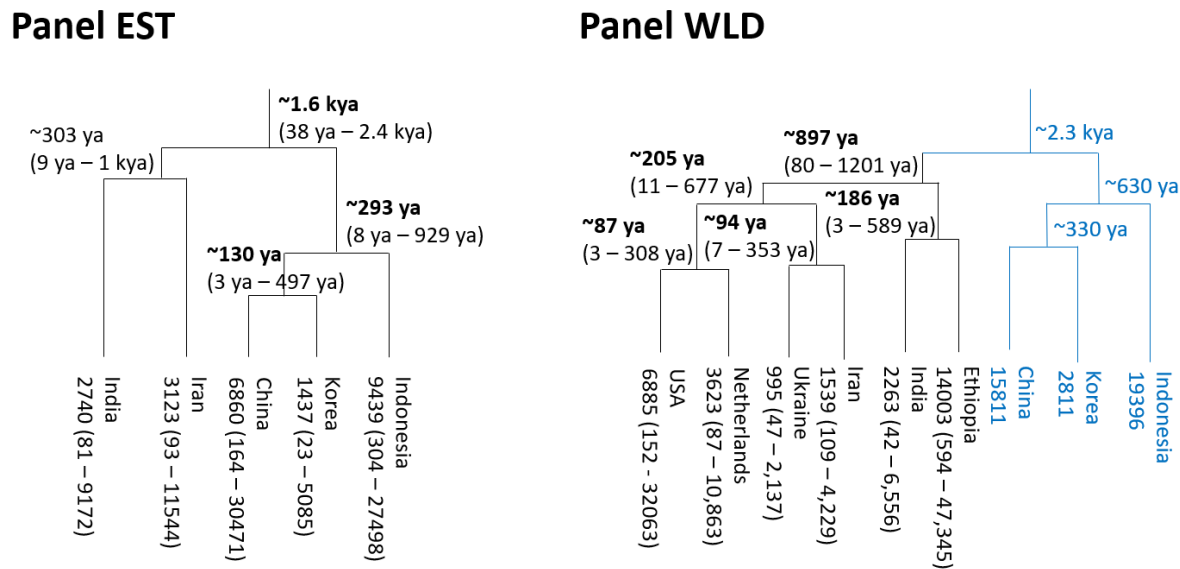
Fig. 9. Schematic description of panels EST and WLD at final stage. Estimated population effect sizes (mean value with confidence interval) are under the population name, whereas estimated time of divergence (mean value with confidence interval) is placed at each node. Parameters used from the estimates of previous stage are in blue font. Bold font is used in Figure 1 in the main text.