PNAS BRIEF REPORT | EVOLUTION OPEN ACCESS

# Solving the 250-year-old mystery of the origin and global spread of the German cockroach, *Blattella germanica*

Qian Tang[a,1], Edward L. Vargo[b], Intan Ahmad[c], Hong Jiang[d], Zuzana Kotyková Varadínová[e,f], Pilot Dovih[g], Dongmin Kim[h], Thomas Bourguignon[i,j], Warren Booth[k], Coby Schal[l], Dmitry V. Mukha[m], Frank E. Rheindt[a,1], and Theodore A. Evans[n,1]

Edited by Marcus Feldman, Stanford University, Stanford, CA; received February 5, 2024; accepted April 5, 2024

**The origin of the German cockroach, *Blattella germanica*, is enigmatic, in part because it is ubiquitous worldwide in human-built structures but absent from any natural habitats. The first historical records of this species are from ca. 250 years ago (ya) from central Europe (hence its name). However, recent research suggests that the center of diversity of the genus is Asian, where its closest relatives are found. To solve this paradox, we sampled genome-wide markers of 281 cockroaches from 17 countries across six continents. We confirm that *B. germanica* evolved from the Asian cockroach *Blattella asahinai* approximately 2,100 ya, probably by adapting to human settlements in India or Myanmar. Our genomic analyses reconstructed two primary global spread routes, one older, westward route to the Middle East coinciding with various Islamic dynasties (~1,200 ya), and another younger eastward route coinciding with the European colonial period (~390 ya). While Europe was not central to the early domestication and spread of the German cockroach, European advances in long-distance transportation and temperature-controlled housing were likely important for the more recent global spread, increasing chances of successful dispersal to and establishment in new regions. The global genetic structure of German cockroaches further supports our model, as it generally aligns with geopolitical boundaries, suggesting regional bridgehead populations established following the advent of international commerce.**

domestication | globalization | insecticide resistance | integrated pest management | invasive species

The rise of human civilization has triggered the evolution and spread of commensal species adapted to urban environments. Some species have become invasive pests, with serious implications for human well-being and economic prosperity (1). Many of these pest species have spread globally, facilitated by human technological innovations, which include advancements in transportation and housing, notably plumbing and heating (2). A deeper understanding of the factors that facilitate the spread of pest species can help prevent further spread and mitigate future economic losses (3). However, tracing the spread of pests based on historical records is often obscured by a dearth of accurate record-keeping, especially prior to the 1950s (4). Advances in population genomics help trace origins and reconstruct spread routes when historical records are limited and geographically biased (4). The German cockroach, *Blattella germanica* L., the world's most prevalent cockroach pest, is ubiquitous in human buildings globally but not outdoors; it imposes significant social, medical, and economic costs (5) due to prevalent insecticide resistance allowing it to outcompete ~40 known pest cockroach species in buildings (6).

The origin and spread of the German cockroach are shrouded in mystery. Described by Linnaeus in 1776 about a decade after the Seven Years' War, historical records have suggested a global spread of German cockroaches from Europe between the late 19th to early 20th centuries (6). However, the German cockroach has no close relatives in Europe; those are in Africa and Asia. The ancestral species was suggested to be the Asian cockroach, *Blattella asahinai* Mizukubo (7, 8), native to the Bay of Bengal in Asia (east India, Myanmar, and nearby islands), and invasive in agricultural landscapes in the southern United States (7). The paradox of a European beginning but Asian phylogenetic affinity is likely due to the almost complete lack of systematic entomological knowledge across the world prior to the 20th century. To help fill the knowledge gap and solve this paradox, we used genome-wide markers from 281 samples from 17 countries around the world (Fig. 1), from which we described the genetic structure and reconstructed spread routes of the German cockroach.

## Results and Discussion

Based on the 1,536 bp mitochondrial COI gene, we found a shallow divergence between Asian and German cockroaches (~0.59%; 9 bp), in stark contrast with the 10 times higher divergences with other congenerics, e.g., *Blattella bisignata* and *Blattella lituricollis*, (>5%; 92 bp and 84 bp respectively). German cockroaches from 83% (44/53) of sites had

Author affiliations: [a]Department of Biological Sciences, National University of Singapore, Singapore 117558, Singapore; [b]Department of Entomology, Texas Agricultural and Mechanical University, College Station 77843-2475, Texas; [c]School of Life Sciences and Technology, Bandung Institute of Technology, Bandung 40132, Indonesia; [d]Department of Human Parasitology, TaiKang Medical School, Wuhan University, Wuhan 430071, China; [e]Department of Zoology, Charles University, Prague 12800, Czech Republic; [f]Department of Zoology, National Museum, Prague 11579, Czech Republic; [g]Trivedi School of Biosciences, Ashoka University, Sonipat 131029, India; [h]Florida Medical Entomology Laboratory, University of Florida, Vero Beach, FL 32962; [i]Evolutionary Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna-son 904-0495, Japan; [j]Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice 370 05, Czech Republic; [k]Department of Entomology, Virginia Polytechnic Institute and State University, Virginia, VA 24061; [l]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC 27695-7613; [m]Department of Animal Genetics, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow 117971, Russia; and [n]School of Biological Sciences, University of Western Australia, Perth 6009, Western Australia

Author contributions: Q.T. and T.A.E. designed research; Q.T., E.L.V., I.A., H.J., Z.K.V., P.D., D.K., T.B., W.B., C.S., D.V.M., F.E.R., and T.A.E. performed research; Q.T. analyzed data; and Q.T., F.E.R., and T.A.E. wrote the paper.

The authors declare no competing interest.



[1]To whom correspondence may be addressed. Email: tangbenjamin@hotmail.com, dbsrfe@nus.edu.sg, or theo.evans@uwa.edu.au.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2401185121/-/DCSupplemental.

Published May 20, 2024.

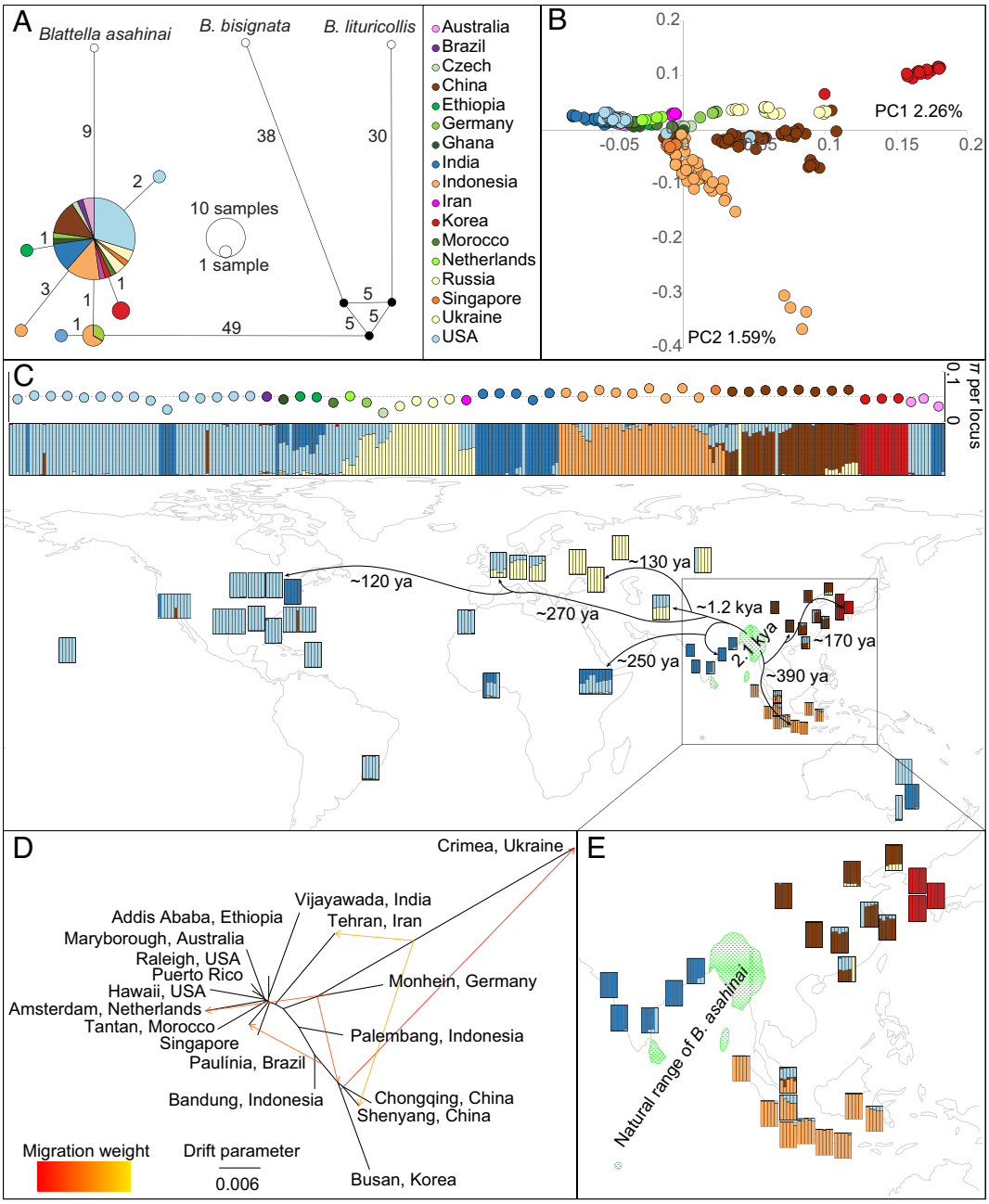


**Fig. 1.** Population genetic structure and global spread routes of the German cockroach, *Blattella germanica*. (*A*) Median joining haplotype network based on the COI gene (1,536 bp). (*B*) Principal component analysis based on 158,216 genome-wide single nucleotide polymorphisms. (*C*) Bar plot (*Top*), split into region-specific bar plots mapped to sampling sites, derived from maximum likelihood estimation of ancestry (ADMIXTURE) at the optimal number of ancestral clusters (K = 6). The diagram on top of the bar plot indicates nucleotide diversity ($\pi$) per locus at each sampling site. The inferred timing and routes of global spread are mapped with black arrows. (*D*) Divergence of key populations presented in the form of an allele-frequency-based unrooted tree with residuals of drift parameters mapped as migration edges. (*E*) A zoomed-in view of (*C*) focused on Asia. The natural range of the ancestor, *B. asahinai*, is shown in green.

identical haplotypes (Fig. 1*A*); the remainder had very low levels of differentiation (0.07 to 0.2%; 1 to 3 bp) between the minor and major COI haplotypes, indicating a recent origin and spread.

Further evidence for a recent global spread came from the limited population subdivision across the globe based on 158,216 single nucleotide polymorphisms (Fig. 1*B*). Maximum likelihood estimation of ancestry, implemented in ADMIXTURE v1.3 (9), suggested six major ancestral clusters of German cockroaches (Korea, China, Indonesia, India, Eastern Europe, and United States), which would have served as sources for modern global populations (Fig. 1*C*). Curiously, the highest divergence was among populations from Korea, Indonesia, and India in Asia even though these are close to the native range of their ancestor *B. asahinai* (7). Moreover, Asian populations (China, India, and Indonesia) exhibited a higher level of nucleotide diversity than other populations (Fig. 1*C*). Most samples from the same country/region shared ancestry, with a few exceptions revealing possible secondary introductions after the establishment of a regional bridgehead population (Fig. 1*C*). This is supported by the relatively high inbreeding within sampling sites (inbreeding coefficient, $F_{IS}$, from −0.12 to 0.65), along with low genetic differentiation between pairs of sampling sites (pairwise $F_{ST}$), ranging from 0 to 0.35. It is also apparent that regional ancestry may correspond to human commercial links. For example, German cockroaches in Singapore and Australia are more closely related to those in the United States than with geographically adjacent populations in Indonesia (Fig. 1 *C* and *D*).

We tested different scenarios of divergence among populations and reconstructed the spread routes of the German cockroach with individuals from nine selected sampling sites by comparing the observed site frequency spectrum with coalescence-based simulations [implemented using fastsimcoal2 v.2.7 (10)] (*SI Appendix*). Six of the nine selected sites (China, Indonesia, India, Korea, Ukraine, and USA) represented the six pure ancestries identified with ADMIXTURE (Fig. 1*C*); three sites represented wider regions (Ethiopia for Africa, Iran for Western Asia, and the Netherlands for Western Europe) that likely acted as steppingstones in the global spread of the German cockroach. To examine the nature of population admixture during this spread, we performed allele-frequency-based tree reconstruction as implemented in TREEMIX v.1.13 (11). The tree of population divergence produced by TREEMIX was consistent with our ancestry estimation and spread route reconstruction (Fig. 1*D*).

We found the ancestor of the German cockroach to be the Asian cockroach, probably living in human settlements, with two domesticated lineages (agricultural/peridomestic and building environments), around 2,100 years ago (ya) when human civilizations were thriving in South Asia. We determined that the global spread of German cockroaches was initiated along two routes, west and east of the origin in India or Myanmar (Fig. 1*C*). As German cockroaches were known to hitchhike in soldiers' bread baskets (12), the expansion (~1.2 kya) westward was probably due to intensifying commercial and military activities of the Islamic Umayyad or Abbasid Caliphates. The expansion (~390 ya) eastward was likely facilitated by European colonial commercial activities between South and Southeast Asia (perhaps the Dutch and British East India Companies).

As recently as the 18$^{th}$ century, the German cockroach was still mostly contained within Asia. Our estimated time for their entry into Europe (~270 ya) matches the earliest historical records in the 1760s (6). The German cockroach then spread to the rest of the world between the late 19th and early 20th century, consistent with the highest volume of first records (6). Advances that accelerated transportation (e.g. steam engines) and thus globalization of trade, and increased comfort in housing (plumbing and indoor heating), allowed German cockroach populations to colonize regions that had been previously inaccessible due to high mortality during long-distance travel and poor cold tolerance (6).

We identified six migratory events in the TREEMIX results, most of them at locations where human commercial activities would lead to an overlap of cockroaches from different ancestries. For example, the population in Singapore was introduced through the West route but received substantial migration from neighboring Indonesian populations (Fig. 1*D*). As admixture during biological invasions may facilitate adaptation to new environments (13), future studies could focus on the functional genomic aspects of this admixture to understand German cockroaches' rapid spread, evolution of insecticide resistance, and so inform better integrated pest management (5, 6).

## Materials and Methods

We collected 281 German cockroach samples from 57 sites in 17 countries across all human-inhabited continents. To acquire the COI gene, we sequenced mitochondrial genomes for 53 individuals, one individual per site, following published protocols (14). To acquire genome-wide SNPs, we followed published double digest restriction-site associated DNA library preparation protocols and associated bioinformatic pipelines (15). The COI dataset with 53 samples of the German cockroach and three samples of other *Blattella* species was used to create the haplotype network, whereas the SNP dataset of 281 samples was used to investigate population genetic structure and reconstruct the global spread of the German cockroach. For a detailed description of analyses, please see *SI Appendix*.

**Data, Materials, and Software Availability.** Raw sequence data for COI (NCBI accession PP692240–PP692292) (16) and genome-wide SNPs (PRJNA1099617) (17) are publicly available in NCBI. Tissue and DNA samples are stored in the Department of Entomology at Texas A&M University, USA. Materials are available upon request. Codes and scripts used for analyses are available at: https://github.com/qt37t247/German_cockroach_ddRAD (18).

**ACKNOWLEDGMENTS.** This work was supported by a National University of Singapore Research Scholarship from the Singapore Ministry of Education, by the Lee Hiok Kwee Endowed Fund of the Department of Biological Sciences, National University of Singapore, Singapore, by the Urban Entomology Endowment at Texas A&M University, by the Joseph R. and Mary W. Wilson Urban Entomology Endowment, by a grant from the National Research Initiative of the USDA Cooperative State Research, Education, and Extension Service (2004-35302-14880) to C.S. and E.L.V., and by a grant from the U.S. Department of Housing and Urban Development Healthy Homes program (NCHHU0053-19) to C.S. We thank Jayanth Kambil, Shakeela V., Fasil Adugna, Nega Tassie, and Eng-Kooi Tan for providing samples.

1. C. Diagne *et al.*, High and rising economic costs of biological invasions worldwide. *Nature* **592**, 571–576 (2021).
2. A. Bonnamour, J. M. W. Gippet, C. Bertelsmeier, Insect and plant invasions follow two waves of globalisation. *Ecol. Lett.* **24**, 2418–2426 (2021).
3. C. Capinha, F. Essl, M. Porto, H. Seebens, The worldwide networks of spread of recorded alien species. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2201911120 (2023).
4. A. Estoup, T. Guillemaud, Reconstructing routes of invasion using genetic data: Why, how and so what? *Mol. Ecol.* **19**, 4113–4130 (2010).
5. C.-Y. Lee, C. Wang, M. K. Rust, "German cockroach infestations in the world and their social and economic impacts" in *Biology and Management of the German Cockroach*, C. Wang, C.-Y. Lee, M. K. Rust, Eds. (CSIRO Publishing, Clayton South, Victoria, Australia, 2021), pp. 1–16.
6. Q. Tang, T. Bourguignon, L. Willenmse, E. De Coninck, T. Evans, Global spread of the German cockroach, *Blattella germanica*. *Biol. Invasions* **21**, 693–707 (2019).
7. L. M. Roth, "A taxonomic revision of the genus *Blattella* Caudell (Dictyoptera, Blattaria: Blattellidae)" in *Entomologica Scandinavica, Supplement* **22**, 1–221 (1985).
8. D. Mukha, B. M. Wiegmann, C. Schal, Evolution and phylogenetic information content of the ribosomal DNA repeat unit in the Blattodea (Insecta). *Insect Biochem. Mol. Biol.* **32**, 951–960 (2002).
9. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
10. L. Excoffier *et al.*, fastsimcoal2: Demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**, 4882–4885 (2021).
11. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
12. J. A. Rehn, Man's uninvited fellow traveler–the cockroach. *Scientific Monthly* **61**, 265–276 (1945).
13. T. Guillemaud, M. Ciosi, E. Lombaert, A. Estoup, Biological invasions in agricultural settings: Insights from evolutionary biology and population genetics. *C. R. Biol.* **334**, 237–246 (2011).
14. T. Bourguignon *et al.*, Transoceanic dispersal and plate tectonics shaped global cockroach distributions: Evidence from mitochondrial phylogenomics. *Mol. Biol. Evol.* **35**, 970–983 (2018).
15. Q. Tang *et al.*, The different fates of two Asian horseshoe crab species with different dispersal abilities. *Evol. Appl.* **14**, 2124–2133 (2021).
16. Q. Tang *et al.*, Data from "Solving the 250-year-old mystery of the origin and global spread of the German cockroach, Blattella germanica." NCBI. https://www.ncbi.nlm.nih.gov/nuccore/?term=PP692240:PP692292[accn]. Deposited 18 April 2024.
17. Q. Tang *et al.*, Data from "Solving the 250-year-old mystery of the origin and global spread of the German cockroach, Blattella germanica." *NCBI*. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1099617. Deposited 12 April 2024.
18. Q. Tang *et al.*, Scripts used for the ddRADSeq data of the German cockroaches. GitHub. https://github.com/qt37t247/German_cockroach_ddRAD. Deposited 18 April 2024.

PNAS

# Supporting Information for

Solving the 250-year-old mystery of the origin and global spread of the German cockroach, *Blattella germanica*.

Qian Tang[1,*], Edward L. Vargo[2], Intan Ahmad[3], Hong Jiang[4], Zuzana Kotyková Varadínová[5,6], Pilot Dovih[7], Dongmin Kim[8], Thomas Bourguignon[9,10], Warren Booth[11], Coby Schal[12], Dmitry V. Mukha[13], Frank E. Rheindt[1,*], Theodore A. Evans[14,*]

[1]Department of Biological Sciences, National University of Singapore, Singapore
[2]Department of Entomology, Texas A&M University, USA
[3]Bandung Institute of Technology, Indonesia
[4]TaiKang Medical School, Wuhan University, China
[5]Department of Zoology, Charles University, Czech Republic
[6]Department of Zoology, National Museum, Czech Republic
[7]Trivedi School of Biosciences, Ashoka University, India
[8]Florida Medical Entomology Laboratory, University of Florida, USA
[9]Okinawa Institute of Science and Technology Graduate University, Japan
[10]Biology Centre of the Czech Academy of Sciences, Institute of Entomology, České Budějovice, Czech Republic
[11]Department of Entomology, Virginia Polytechnic Institute and State University, USA
[12]Department of Entomology and Plant Pathology, North Carolina State University, USA
[13]Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia
[14] School of Biological Sciences, University of Western Australia, Australia

*Correspondence: Qian Tang, Frank E. Rheindt, and Theodore A. Evans
tangbenjamin@hotmail.com, dbsrfe@nus.edu.sg, and theo.evans@uwa.edu.au

**This PDF file includes:**

***Sample collection and variant calling***

In this study, we collected 281 German cockroach samples from 57 localities in 17 countries across all continents with permanent human residence (for details, refer to Dataset S1). Samples from China (1), Iran, Russia, Ukraine, and USA (2) were subsets of previous population genetic studies using microsatellites. To avoid DNA contamination from microbes and symbionts, which are abundant in the abdomen of cockroaches, two legs of each sample were ground and used as raw material for DNA extraction. Samples from India were acquired from lab strains and extracted with an DNeasy Blood & Tissue Kit (Qiagen, Germany). All other samples were captured alive and preserved in ethanol before DNA extraction using a phenol-chloroform method as described in a previous study (1). We enzymatically digested the genomic DNA using *EcoRI* and *MspI*. Subsequently, the digested DNA fragments were ligated with unique adapters. Fragments of ~350 bp were then selectively isolated using Pippin Prep (Sage Science, USA) and subjected to eight cycles of PCR amplification to produce the ultimate double digest restriction-site associated DNA (ddRAD) libraries. These libraries were sequenced using Illumina Novaseq (USA) to produce 150 bp pair-end reads.

Raw reads received from the sequence provider were confirmed to be of high quality and with low adapter content using *FastQC* (Babraham Bioinformatics, USA). We proceeded with no reads discarded or truncated. Raw reads were demultiplexed into samples using the function *process_rad_tag* as implemented in *Stacks* v2.4 (3). We aligned reads of each sample to the reference genome of the German cockroach (4) using function *BWA-MEM* as implemented in BWA (5), with all technical parameters set as default. Alignments, discarding those with MAPQ scores below 20 (-q 20), were sorted into bam files using samtools v1.9 (6).

We called variants of single nucleotide polymorphisms (SNPs) using the function *ref_map.pl* as implemented in *Stacks*, regarding all samples as a single population. Our raw SNP yield amounted to 2,050,828 (mean per-sample coverage: 30.1×; SD: 23.6×). Using *PLINK* v1.9 (7),

we removed loci with >10% missing data (--*geno* 0.1) and linked (unphased hardcall $r^2$ > 0.95) to neighboring loci (within a 25-SNP window sliding 10 SNPs per step). In the end, we converged on a SNP dataset of 158,216 biallelic loci for all 281 individuals.

We selected one sample per sampling site to sequence (Illumina HiSeq2000) the mitochondrial genome from long PCRs (see second sequencing strategy in Bourguignon *et al.* (8)). Mitochondrial genomes were *de novo* assembled as described in Bourguignon *et al* (9). In addition, to make comparisons between intra- and interspecies levels of divergence, we collected published mitochondrial genomes of three other *Blattella* species (*B. asahinai* MG882167, *B. bisignata* NC_018549.1, and *B. lituricollis* MG882212) (8, 10). We spliced the mitochondrial genomes using the *MITOS* Webserver (11) to create gene-based alignments. We aligned the COI sequences (1,536bp) based on codons using the Muscle algorithm implemented in *MEGA* v11. We converted the alignment to haplotype data and generated a haplotype network using the median-joining algorithm as implemented in *POPART* v1.7 (12).

***Population structure and ancestry estimation***

To determine genetic differentiation among individuals, we performed PCA using the *R* package *SNPRelate* v1.28.0 (13). To estimate the ancestry of our global German cockroach samples, we used maximum-likelihood ancestry estimation as implemented in *ADMIXTURE* v1.3.0 (14), testing population assignment and likelihoods across 1–15 inferred populations for each species. To understand genetic diversity within and among sampling sites, we calculated the inbreeding coefficient ($F_{IS}$) for each sampling site and the fixation index among pairs of sampling sites (pairwise $F_{ST}$), respectively, using *SNPRelate*.

We considered a population to exhibit ancestry-sharing when the minor ancestry contributed more than overall 20% in a consistent manner across individuals. To verify whether an admixture event may have led to an ancestry-sharing pattern at a particular sampling site, we compared

nucleotide diversity with the number of private alleles across all 57 sampling sites. We excluded seven individuals exhibiting signs of secondary introduction. If individuals at a sampling site that displayed a consistent ancestry-sharing pattern exhibited a relatively high nucleotide diversity but low number of private alleles, we considered it likely that these individuals may be offspring of admixture events from different ancestries. For each sampling site, nucleotide diversity was calculated as an average of per-locus nucleotide diversity using *VCFtools* v0.1.16 (15).

***Reconstruction of spread routes***

To reconstruct the demographic history of the global spread of the German cockroach, we adopted coalescence-based analyses using the site frequency spectrum (SFS) as implemented in *fastsimcoal2* v.2.7. We selected nine sampling sites (see main paper for the justification) to be involved in the spread route reconstruction based on population genetic structure and ancestry estimation. As the reference genome is only available for the German cockroach, we computed a folded SFS excluding missing data, linked loci, and monomorphic sites using the *python* script *easySFS* (16, 17). Given that *fastsimcoal2* allows a maximum of six populations in each scenario, we adopted a stepwise strategy, comprising four stages. The topology of divergence events and estimated parameters of each stage were used as priors for the subsequent stage (18). Each stage comprised one to four panels depending on the combinations of individuals involved (see Dataset S2). Each panel contained three to six scenarios with the same individuals but different divergence topologies to be simulated. We performed 50 runs for each scenario. We assumed three generations per year (19) and a mutation rate of $2.8 \times 10^{-9}$ per site per generation as estimated from *Drosophila melanogaster* (20). Considering the relative homogeneity in the ancestry patterns of individuals at the selected sampling sites, we assumed no gene flow after divergence throughout the demographic models. All the procedures and related files are documented in the online repository (https://github.com/qt37t247/German_cockroach_ddRAD).

Demographic modelling started with the inclusion of individuals from only three sampling sites (panel PRE, stage I) exhibiting the highest divergence in the PCA plot (Busan-Korea, Bandung-Indonesia, and Vijayawada-India). All three possible topologies were tested: #A (India (Korea, Indonesia)) – delta likelihood 803; #B (Indonesia (Korea, India)) – delta likelihood 868; #C (Korea (India, Indonesia)) – delta likelihood 954. According to delta likelihood, the most plausible scenario envisaged an initial divergence of India, then Korea and Indonesia.

In stage II, we aimed to investigate the relationship among the six identified ancestries. In addition to the individuals representing the three ancestries involved in stage I, we added individuals from China (CHN) and Iran (IRA); the latter constitutes the co-ancestor of the Eastern European and USA ancestries. Therefore, stage II comprised two panels: CHN and IRA, with each panel comprising five scenarios (scenarios A, B, C, D, and E).

For panel CHN, in which we added individuals from China to those from stage I, five topologies of divergence were tested: #A ((India, China), (Korea, Indonesia)) – delta likelihood 2874; #B (India, ((China, Korea), Indonesia))) – delta likelihood 2772; #C (India, ((China, Indonesia), Korea))) – delta likelihood 2848; #D (India, (China, (Indonesia, Korea))) – delta likelihood 2843; #E (China, (India, (Indonesia, Korea))) – delta likelihood 2926. The most plausible scenario, i.e., the one with the lowest delta likelihood, #B, had India diverging first, then Indonesia, then Korea and China.

For panel IRA, in which we added individuals from Iran to those from stage I, five topologies of divergence were tested: #A ((India, Iran), (Korea, Indonesia)) – delta likelihood 2909; #B (India, ((Indonesia, (China, Korea))) – delta likelihood 2976; #C (India, (Korea (Indonesia, Iran))) – delta likelihood 3036; #D (India, (Iran, (Indonesia, Korea))) – delta likelihood 3035; #E (Iran, (India, (Indonesia, Korea))) – delta likelihood 3194. The most plausible scenario, #A, envisaged an initial divergence between Korea + Indonesia on the one hand and India + Iran on the other hand.

In stage III, we placed individuals from each sampling site that we had not used in the previous stages onto the topology that we had obtained from the previous stage. This stage comprised four panels: ETP, COH, UKR, and NDL.

For panel ETP, in which we added individuals from Ethiopia to those from panel IRA (stage II), five topologies of divergence were tested: #A ((Korea, Indonesia), (India, (Iran, Ethiopia))) – delta likelihood 6038; #B ((Korea, Indonesia), (Ethiopia, (India, Iran))) – delta likelihood 6090; #C ((Korea, Indonesia), (Iran, (Ethiopia, India))) – delta likelihood 6013; #D ((Ethiopia, (Korea, Indonesia)), (India, (Iran, Ethiopia)) – delta likelihood 6908; #E (Ethiopia, ((Korea, Indonesia), (India, (Iran, Ethiopia))) – delta likelihood 9517. The most plausible scenario, #C, exhibited a divergence between (Korea + Indonesia) and (Ethiopia + India + Iran). Within the latter subcluster, Iran split off first, then Ethiopia and India.

For panel COH, in which we added individuals from the USA to those from panel IRA (stage II), four topologies of divergence were tested: #A ((Korea, Indonesia), (India, (Iran, USA))) – delta likelihood 6071; #B ((Korea, Indonesia), (USA, (India, Iran))) – delta likelihood 6109; #C ((Korea, Indonesia), (Iran, (USA, India))) – delta likelihood 6013; #D ((USA, (Korea, Indonesia)), (India, (Iran, Ethiopia)) – delta likelihood 6797. The most plausible scenario, #A, resulted in a divergence between (Korea + Indonesia) and (India + Iran + USA). Within the latter subcluster, India split off first, then Iran and USA.

For panel UKR, in which we added individuals from Ukraine to those from panel IRA (stage II), four topologies of divergence were tested: #A ((Korea, Indonesia), (India, (Iran, Ukraine))) – delta likelihood 5912; #B ((Korea, Indonesia), (Ukraine, (India, Iran))) – delta likelihood 6027; #C ((Korea, Indonesia), (Iran, (Ukraine, India))) – delta likelihood 6036; #D ((Ukraine, (Korea, Indonesia)),

(India, (Iran, Ethiopia)) – delta likelihood 6141. The most plausible scenario, #A, exhibited an initial divergence between (Korea + Indonesia) and (India + Iran + Ukraine). Within the latter subcluster, India split off first, then Iran and Ukraine.

For panel NDL, in which we added individuals from the Netherlands to those from panel IRA (stage II), four topologies of divergence were tested: #A ((Korea, Indonesia), (India, (Iran, Netherlands))) – delta likelihood 6079; #B ((Korea, Indonesia), (Netherlands, (India, Iran))) – delta likelihood 6130; #C ((Korea, Indonesia), (Iran, (Netherlands, India))) – delta likelihood 6103; #D ((Netherlands, (Korea, Indonesia)), (India, (Iran, Ethiopia)) – delta likelihood 35641. The most plausible scenario, #A, envisaged a divergence between (Korea + Indonesia) and (India + Iran + Netherlands). Within the latter subcluster, India split off first, then Iran and the Netherlands.

Throughout the first three stages, we examined the sequence of divergence events across nine geographical areas, showing that the global spread of the German cockroach occurred along two primary routes, one to the east involving ancestries from China, Indonesia, and Korea, the other to the west involving the remaining ancestries. Stage IV aimed to investigate the detailed placement of populations along the western route with a more comprehensive incorporation of samples. Therefore, stage IV only included one panel with six scenarios.

For panel WST, we used a fixed topology and demographic parameters for the eastern route and the first divergence in the western route: ((Korea, Indonesia), (India, all other western individuals)). Six topologies of divergence were tested for the non-Indian western individuals: #A (Iran, (USA, (Netherlands, Ukraine))) – delta likelihood 4964; #B (Iran, (Netherlands, (USA, Ukraine))) – delta likelihood 4965; #C (Iran, (Ukraine, (USA, Netherlands))) – delta likelihood 4958; #D ((Iran, USA), (Netherlands, Ukraine)) – delta likelihood 4965. #E ((Iran, Netherlands), (Ukraine, USA)) – delta likelihood 4964; #F ((Iran, Ukraine), (Netherlands, USA)) – delta likelihood 4872. The most plausible scenario, #F, envisaged a divergence between (Iran + Ukraine) and (Netherlands + USA).

At the final stage, with the consolidated topology of population divergences reconstructed, we carried out demographic parameter estimations and parametric bootstrapping. The final stage comprised two panels: EST and WLD. Each panel included only one scenario. Panel EST involved individuals from China, Korea, Indonesia, India, and Iran with divergences in the order of ((India, Iran), (Indonesia, (China, Korea))). Panel WLD involved all individuals, with th following topology: ((Indonesia, (China, Korea)), ((Ethiopia, India), ((Iran, Ukraine), (Netherlands, USA)))). We performed 50 runs for each scenario for the initial parameter estimation. The best run for 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.

***Population divergence with migration***

To examine the nature of population admixture during the global spread of German cockroaches, we performed allele-frequency-based tree reconstruction as implemented in TREEMIX v.1.13 (21). The input allele frequency was calculated with PLINK based on a SNP dataset that excluded loci with >20% missing data (--*geno* 0.2) and/or close linkage (unphased hardcall $r^2$ < 0.1) to neighboring loci (within a 25-SNP window sliding 10 SNPs per step) as suggested by the user manual. We first ran TREEMIX on 49 populations, excluding those with signs of secondary introductions (see Dataset S1). We ran migration edges from 0 to 24, with 10 independent runs for each migration edge. Based on the Evanno method as implemented in OptM v0.1.6 (22), the optimal number of migration edges was 13. However, the variance explained did not reach the 99.8% threshold, and further increases to the number of migration edges made the analysis too computationally intensive.

In the first batch of TREEMIX runs with 49 populations at 13 migration edges, most populations collapsed and most migratory events were identified within the Indonesian population only. The second batch of TREEMIX runs with 18 populations, each representing a geographic region or a population with prominent ancestry-sharing (see Dataset S1), balanced numbers between admixed

and non-admixed populations to prevent over-estimation of migration edges. In the second batch, we ran migration edges from 0 – 19 (99.8% variance explained threshold reached at 15), with 10 independent runs for each migration edge. For both batches of TREEMIX runs, we grouped 500 SNPs per linkage block (-k 500) with a round of global rearrangements after adding all populations (-global).

**SI References**


1. Q. Tang, H. Jiang, Y. Li, T. Bourguignon, T. A. Evans, Population structure of the German cockroach, Blattella germanica, shows two expansions across China. *Biol Invasions* **18**, 2391–2402, (2016).

2. E. L. Vargo, *et al.*, Hierarchical Genetic Analysis of German Cockroach (Blattella germanica) Populations from within Buildings to across Continents. *PLOS ONE* **9**, e102321, (2014).

3. N. C. Rochette, A. G. Rivera-Colón, J. M. Catchen, Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular ecology* **28**, 4737–4754, (2019).

4. M. C. Harrison, *et al.*, Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature ecology & evolution* **2**, 557–566, (2018).

5. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio]*, (2013).

6. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, (2009).

7. C. C. Chang, *et al.*, Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742-015-0047–8, (2015).

8. T. Bourguignon, *et al.*, Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. *Molecular biology and evolution* **35**, 970–983, (2018).

9. T. Bourguignon, *et al.*, The evolutionary history of termites as inferred from 66 mitochondrial genomes. *Molecular biology and evolution* **32**, 406–421, (2014).

10. A.-H. Chen, Complete mitochondrial genome of the double-striped cockroach *Blattella bisignata* (Insecta: Blattaria: Blaberoidea). *Mitochondrial DNA* **24**, 14–16, (2013).

11. M. Bernt, *et al.*, MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular phylogenetics and evolution* **69**, 313–319, (2013).

12. J. W. Leigh, D. Bryant, POPART: full-feature software for haplotype network construction. *Methods in ecology and evolution* **6**, 1110–1116, (2015).

13. X. Zheng, *et al.*, A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328, (2012).

14. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664, (2009).

15. P. Danecek, *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158, (2011).

16. A. J. Coffman, P. H. Hsieh, S. Gravel, R. N. Gutenkunst, Computationally efficient composite likelihood statistics for demographic inference. *Molecular biology and evolution* **33**, 591–593, (2016).

17. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS genetics* **5**, e1000695, (2009).

18. N. Marchi, *et al.*, The genomic origins of the world's first farmers. *Cell* **185**, 1842–1859, (2022).

19. G. E. Gould, The effects of temperature upon the development of cockroaches in *Proceedings of the Indiana Academy of Science*, (1940), pp. 242–248.

20. P. D. Keightley, R. W. Ness, D. L. Halligan, P. R. Haddrill, Estimation of the spontaneous mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. *Genetics* **196**, 313–320, (2014).

21. J. K. Pickrell, J. K. Pritchard, Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLOS Genetics* **8**, e1002967, (2012).

22. R. R. Fitak, OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols* **6**, bpab017, (2021).