Supporting Information

Larson et al. 10.1073/pnas.0912264107

SI Text

Sample Information. For this study, we generated 151 mtDNA sequences from domestic and wild boar across East Asia and combined the results with more than 1,500 published sequences from GenBank. Details of these sequences are provided in Table S1, and GenBank sequences are provided in Table S2.

Table S3 lists all of the ancient specimens analyzed for this study and includes full provenance, context, and date information alongside the results of the genetic analysis. Dates have been determined in different manners for each of the sites. At Jiahu, radiocarbon dates of several samples were obtained and calibrated. At Wangchenggang, Taosi, and Wadian, dates were determined based on radiocarbon dating and associated stratigraphy and artifacts. Although no radiocarbon dates were obtained at either Gaohong or Guchengzhai, relative dates were calculated based upon stratigraphy and correlated cultural association determined by other sites possessing similar cultural layers.

Table S4 lists all of the names of the haplotypes and the samples included in each within the three alignments used to build the tree (Fig. 1), the network consisting of the modern samples (Fig. 2*A*), and the network that incorporates the ancient results (Fig. 2*B*). The code column corresponds to the sample identification code used in Tables S1 and S2. Table S5 lists all of the samples that correspond to the identified clades on the phylogenetic tree (Fig. 1), including the 13 clades that consist solely of wild boar, the four clades that possess both wild boar and domestic pigs, and the nine clades that are made up only of domestic pigs. In addition, this table also lists both the samples that possess specific provenance information and those that do not, all of whom are not affiliated with any of the described clades.

A larger version of the tree shown in Fig. 1 depicting the haplotype names on each of the tips is shown in Fig. S1. Fig. S2 depicts versions of the networks shown in Fig. 2 but includes the names of the haplotypes alongside the circles. Fig. S3 shows the locations of samples belong to the 13 wild clades, the four mixed clades, and the general cluster.

Ancient Samples. The 48 pig bones from the six archeological sites were selected for this analysis by J.Y. Although a multitude of different techniques have been generally applied to archeological animal bones to determine their wild or domestic status, no single approach can provide a definitive ascertainment (see supporting discussion in ref. 1 for more information). For the purposes of this study, accurate status determinations are not necessary given both the genetic similarity between the ancient specimens (many of which are likely domestic) and many of the modern domestic and wild samples, and the fact that none of the ancient samples fell into the W13 clade that consists solely of modern wild boar from Central China, a result that may have suggested the bones were derived from wild boar.

This is not to say that status calls have not been attempted on the ancient material. Several lines of evidence collected from pig remains (including morphology, M3 size, and age distribution) suggest that all of the samples excavated from Wadian (2) and Wangchenggang (3) are derived from domestic pigs. The results from the remains from Taosi, Guchengzhai, and Gaohong (not yet published) also suggest that pigs from these sites (including those included as part of this study) are also domestic. Last, the conclusion of one publication was that a definitive status determination of the samples from Jiahu was not possible using traditional techniques (4). A second publication claimed that wild and domestic pigs from Jiahu were identifiable (5), and a more recent analysis that employed a geometric morphometric approach concluded that although the pigs from Jiahu phase 1 (9,000–8,600 B.P.) were not unequivocally domestic, a clear domestication signature was obtained from pigs excavated from Jiahu phase 2 (8,600–8,200 B.P.) (6).

Ancient and Modern DNA Extraction. All pre-PCR work was carried out in a physically isolated laboratory dedicated to ancient DNA analysis at the China Agricultural University. After the surfaces of the bones and teeth were removed, 0.2–0.4 g of bone powder was generated by drilling into the material. Following digestion in a lysis buffer [100 mM Tris·Cl, 100 mM NaCl, 100 mM EDTA (pH 8.0), 0.5% SDS], the sample was incubated at room temperature overnight.

The tubes were then centrifuged at 10,000 rpm for 1 min, after which the supernatant was transferred to a tube containing 1 mL of a separate buffer (5 M KAc, 6 M LiCl), and then incubated for 1 h on ice. After centrifuging the samples at 12,000 rpm for 15 min, the supernatant was added to an equal volume of isopropanol and stored at −20 °C for at least 2 h.

Following a subsequent spin at 13,000 rpm for 1 h, the isopropanol was discarded and the precipitate was rinsed with 1 mL of 70% ethanol, dried, and dissolved in an appropriate volume of ddH_2O. Several blank extractions were processed alongside the samples in the same manner.

Finally, a total of five samples representing the haplotype variation found in the ancient samples were sent to the Ancient DNA Laboratory at the Research Center for Chinese Frontier Archeology at Jilin University for independent replication using a modified extraction technique (7). The sequences obtained through independent extraction were identical to those generated in Beijing. Modern DNA extraction from ear tissues and blood followed the standard phenol–chloroform methods described previously (8).

PCR Methods. PCRs were set up using 25-μL volumes containing 1 U AmpliTaq Gold (Applied Biosystems), 1× PCR buffer attached with Ampli aq Gold, 3 mM Mg^{2+}, 400 AmpliTaq Gold M dNTPs, 2 mg/mL BSA (Promega), 0.2 μL *Escherichia coli* UNG (Sigma), 0.5 μM forward and reverse primers, and 3–5 μL of DNA extract.

Primers used to generate the modern sequences are as follows (listed 5′ to 3′): DloopL-ACTAACTCCGCCATCAGCAC and DloopR-GTTTGGCAAGGCGTTATAG.

Primers used to generate the ancient sequences are as follows (listed 5′ to 3′): L99 – ACAAATATGCGACCCCAAAA (97 bp with R196), R196-ATGCATGGGGACTAGCAGTT, L180-TGCTAGTCCCCATGCATATAA (179 bp with R358), and R358-CCTGCCAAGCGGGTTGCTGG.

PCRs were run in a Mastercycler Personal PTC-200 (Bio-Rad) using the following cycling conditions: 37 °C for 10 min, 94 °C for 5 min, followed by 60 cycles of 94 °C for 45 s, 6 °C for 30 s, 72 °C for 30 s, and a final extension of 10 min at 72 °C. Amplifications of blank extractions and PCR blanks were performed in all experiments to monitor contamination.

Ancient PCR products were purified using QIAquik™ PCR (QIAGEN) and cloned using the Invitrogen Topo-TA cloning kit. The sequencing reaction was carried out on an ABI PRISM 3730 automated DNA sequencer (Applied Biosystems) using ABI Prism Big Dye Terminator v3.1 Cycle Sequencing kit. Five microliters of PCR product were then separated by electrophoresis on a 2% agarose gel (Biowest).

Modern PCR products were purified using a QIAEX Gel Extraction Kit (QIAGEN) and were directly sequenced. The sequencing reaction was carried out on an ABI 310 automated DNA

sequencer (Applied Biosystems) using ABI Prism Big Dye Terminator v3.1 Cycle Sequencing kit. The primers used for amplification were also used for sequencing, and each fragment was sequenced in both directions. The obtained electropherograms were assembled to examine any base pair ambiguities using Chromas 2.22 (http://www.technelysium.com.au/).

1. Larson G, et al. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proc Natl Acad Sci USA* 104:15276–15281.
2. Henan Provincial Cultural Relics and Archaeology Institute (2004) *Yuzhou Wadian* (World Publishing Corp., Beijing).
3. School of Archaeology and Museology, Peking University and the Henan Provincial Institute of Archaeology and Cultural Relics. (2007) *Archaeological Discovery and Research at the Wangchenggang Site in Dengfeng (2002-2005)* (Elephant, Beijing).
4. Jing Y, Flad R, Luo YB (2008) Meat-acquisition patterns in the Neolithic Yangzi river valley, China. *Antiquity* 82:351–366.
5. Luo Y, Zhang J (2008) Restudy of the pigs' bones from the Jiahu site in Wuyang County, Henen. *Kaogu* 1:90–96 (in Chinese).
6. Cucchi T, Yuan J, Hulme-Beaman K, Dobney K (2008) New evidence for early Neolithic pig domestication at Jiahu, Henan Province, China: Clues from tooth shape analyses using geometric morphometrics. *J Archaeol Sci*, in press.
7. Fu YQ, et al. (2007) Molecular genetic analysis of Wanggu remains, Inner Mongolia, China. *Am J Phys Anthropol* 132:285–291.
8. Sambrook JE, Fritsch F, Maniatis T (2002) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
9. Larson G, et al. (2007) Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in island southeast Asia and Oceania. *Proc Natl Acad Sci USA* 104: 4834–4839.

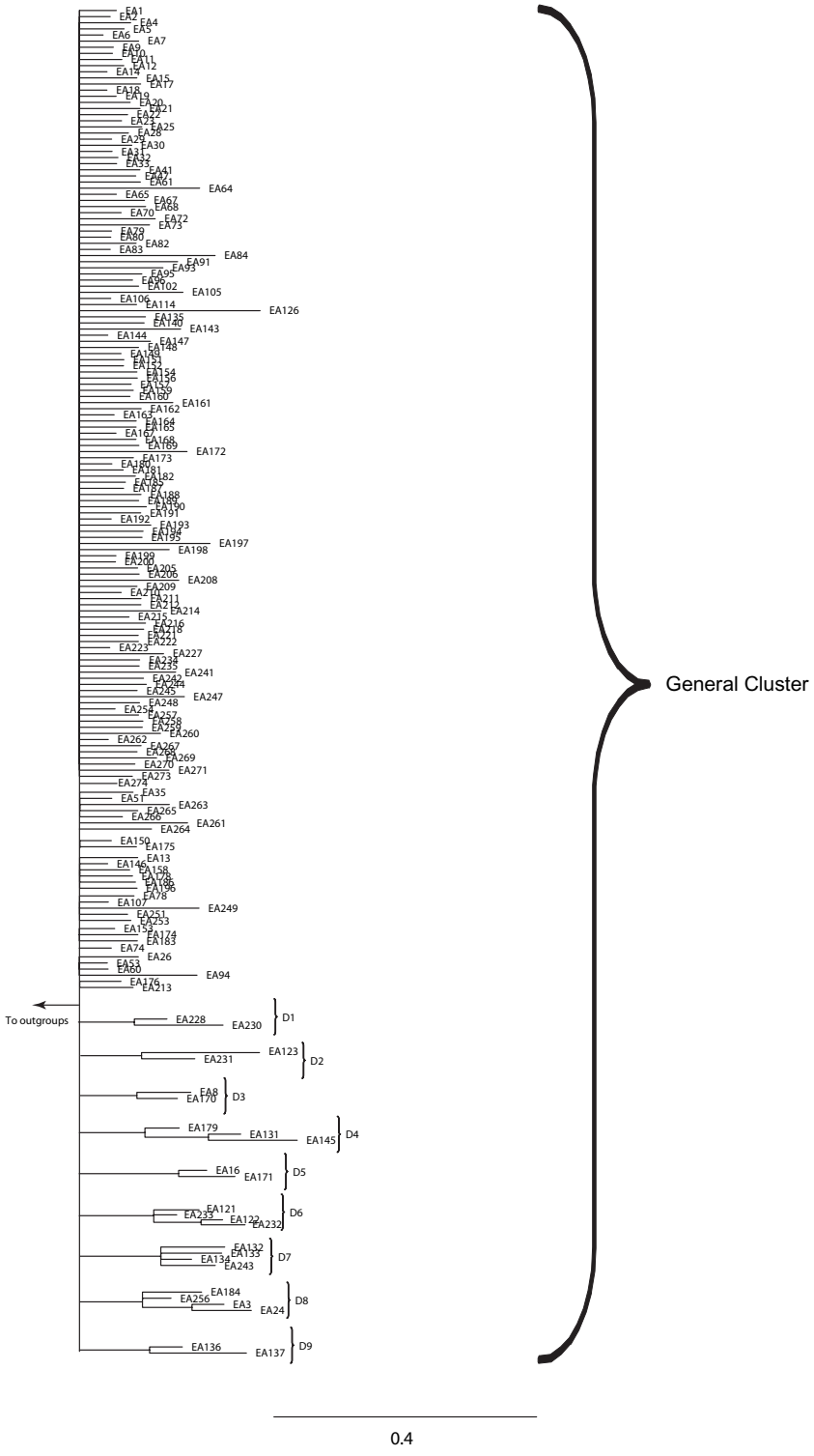

Fig. S1. (Continued)

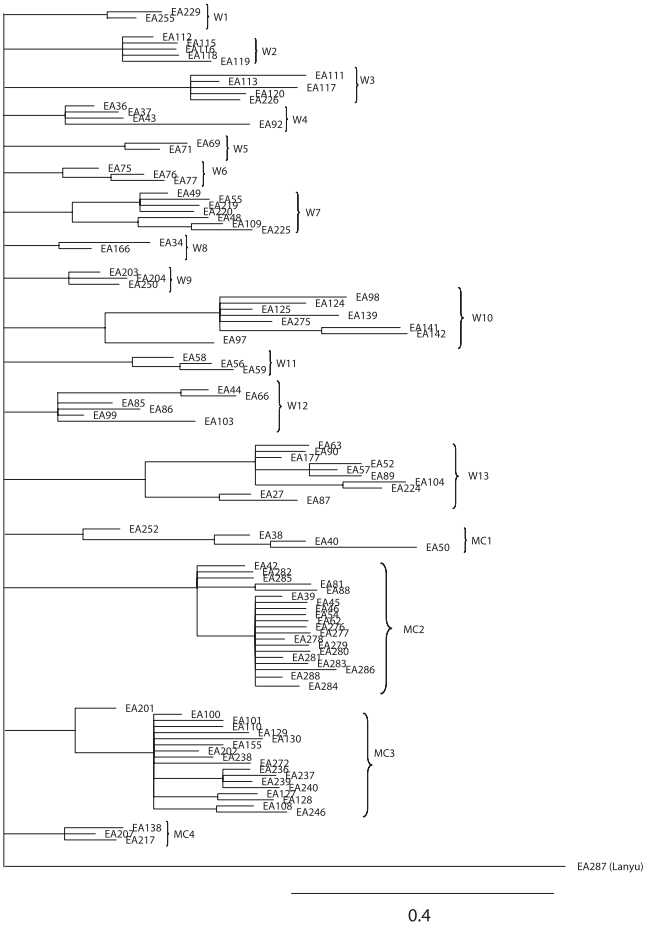

Fig. S1. A consensus phylogenetic tree spread over two pages depicting the relationships between the named clades and general cluster discussed in the text, a simplified version of which is shown in Fig. 1. Note that for space reasons, the scales of the branch lengths are different on each page. Clades made up of only domestic samples, only wild samples, and mixed clades, as well as haplotypes belonging to the general cluster, have been identified. Methods used to generate the tree are discussed in the main text.

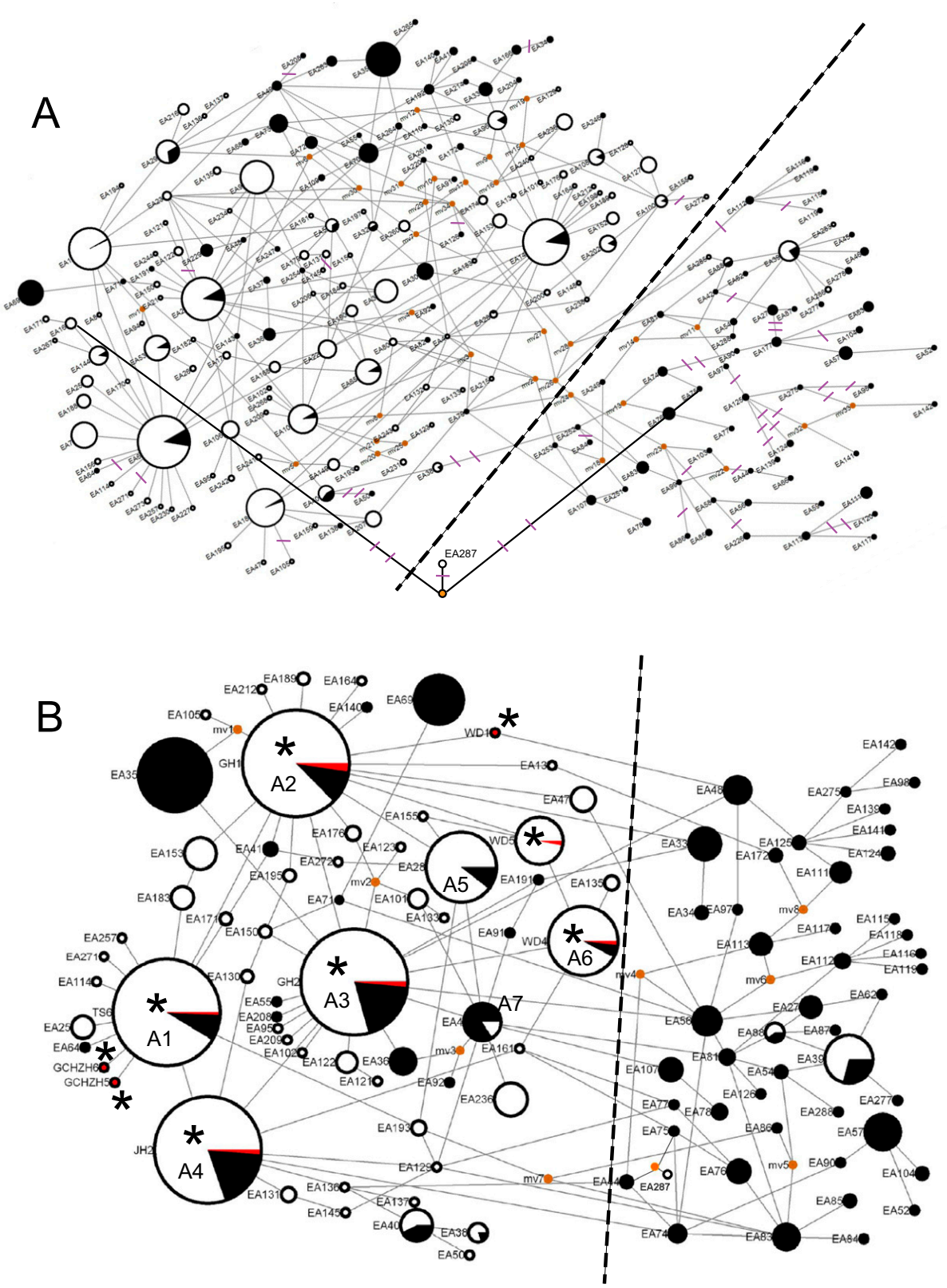

Fig. S2. Two median-joining networks (in which node sizes are proportional to haplotype frequencies) depicting first (*A*) the relationships among 1,541 modern wild and domestic samples in a 378-bp alignment and second (*B*) the relative position of the ancient haplotypes after the archeological samples were added to a reduced 185-bp alignment. Wild, domestic, and ancient samples are shown in black, white, and red respectively, and asterisks mark the locations of the ancient samples. The dashed black line dividing the networks correlates with the line shown in Fig. 1 demonstrating the consistent distinction between the general cluster and the wild and mixed haplotypes. The location of the Lanyu sequence (EA287) is shown here, as are small purple lines that indicate mutations along the branches. The numerous mutations separating the Lanyu sequence from all other East Asian sequences demonstrates its unique mitochondrial character. Inferred haplotypes are represented by small orange dots. Additional information about the haplotype names positioned to the left of the circles can be found in Tables S1–S5.

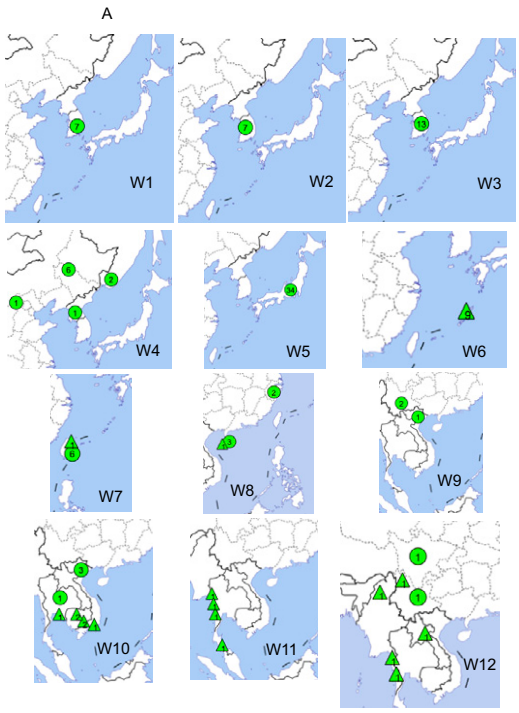

Fig. S3. (Continued)

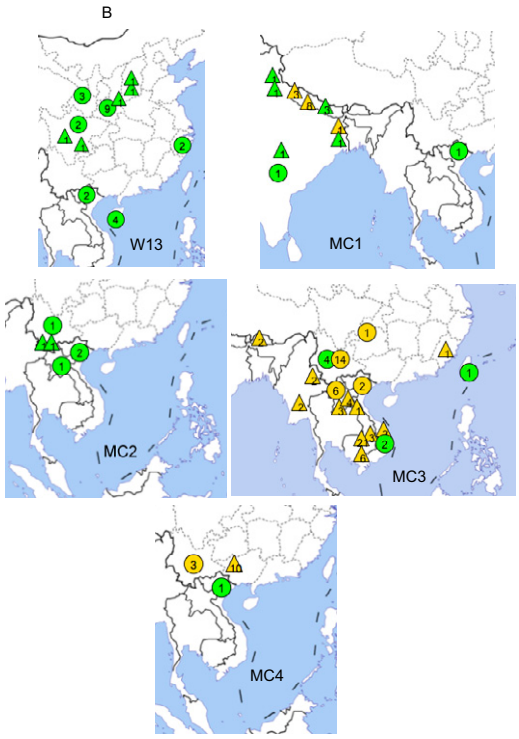

Fig. S3. (Continued)

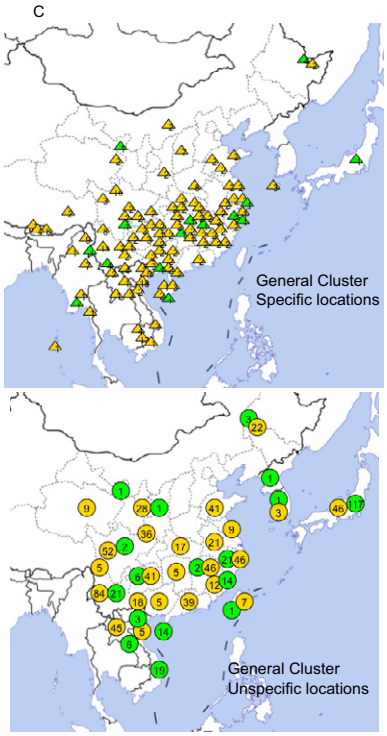

Fig. S3. (*A*) A series of 12 maps depicting the locations of the samples belonging to wild clades W1–W12, respectively. Numbers inside circles and triangles represent the total number of samples at that location. Triangles are used when the sample provenance information is at the level of city or region. Circles are used when the sample can only be assigned to a country or Chinese or Indian provenance. Green shapes represent wild samples. (*B*) A series of five maps depicting the locations of the samples belonging to wild clade W13 and mixed clades MC1–MC4, respectively. Numbers inside circles and triangles represent the total number of samples at that location. Triangles are used when the sample provenance information is at the level of city or region. Circles are used when the sample can only be assigned to a country or Chinese or Indian provenance. Green shapes represent wild samples, and yellow shapes represent domestic or known feral samples. Only the locations of the wild samples belonging to the Pacific clade (MC2) are shown in this figure. Locations of the Pacific clade domestic and feral samples can be found in ref. 9. (*C*) Two maps depicting the locations of the samples belonging to the general cluster. Numbers inside circles and triangles represent the total number of samples at that location. Green shapes represent wild samples, and yellow shapes represent domestic or known feral samples. The top map shows only those samples whose specific locations are known, and the bottom map depicts the remaining samples in the general cluster whose locations are only known down to the level of country or Chinese or Indian provenance.

Other Supporting Information Files

Table S1 (XLS)
Table S2 (XLS)
Table S3 (XLS)
Table S4 (XLS)
Table S5 (XLS)