

Figure S1. Maps of annual H5N1 outbreak records. In each map, outbreak records are assigned to and reported by admin-2 polygons.

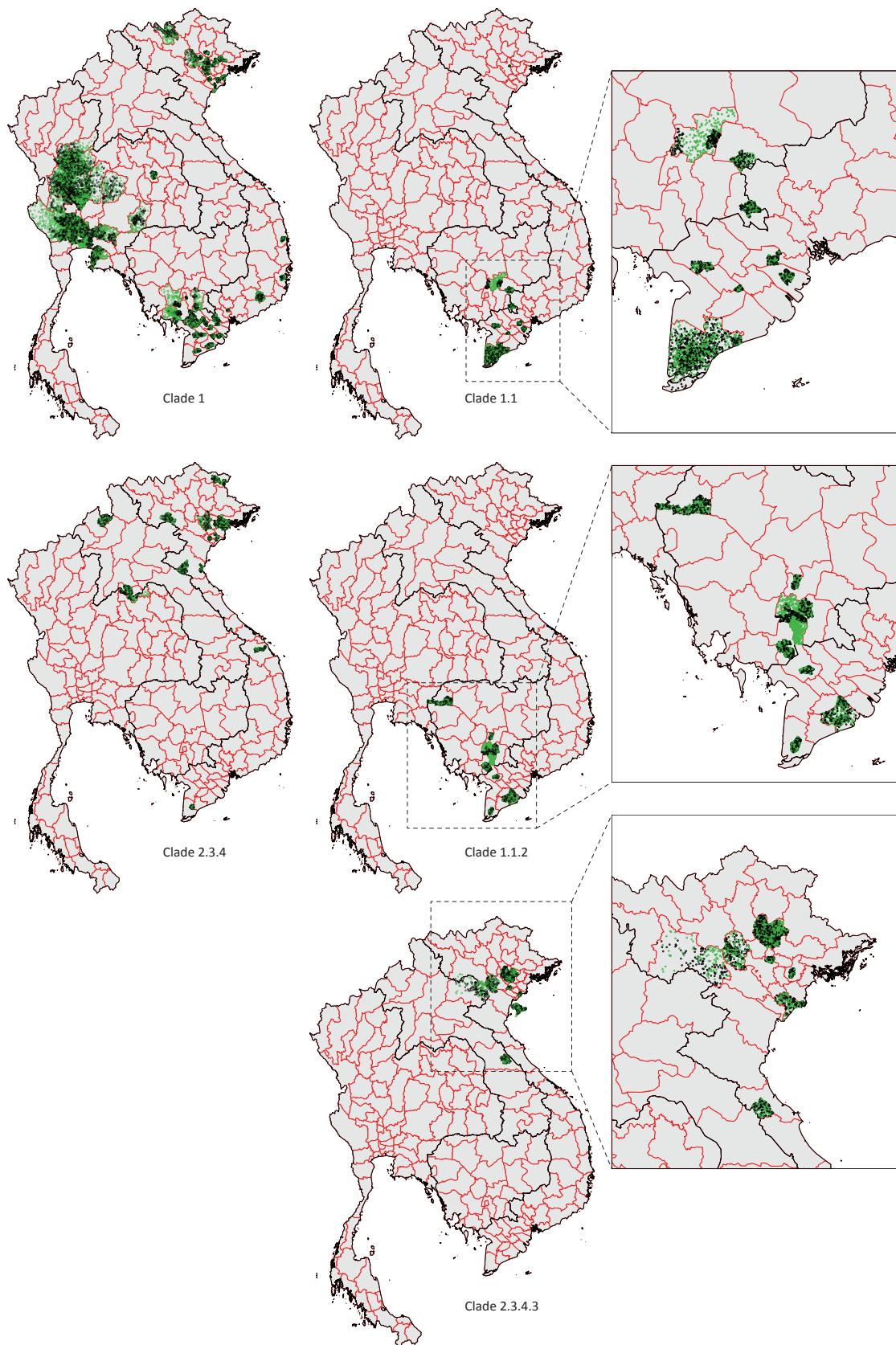


Figure S2. Comparison between unconstrained and constrained sampling location priors used for each clade-specific continuous phylogeographic analysis. Sampling locations drawn from unconstrained and constrained priors are represented by green and black dots, respectively. Similar to the other maps, international (admin-0) borders are displayed in black and admin-1 borders are displayed in red. As detailed in the text, external data associated with admin-2 polygons can be used to constrain sampling position priors for sequences originally assigned to a broad admin-1 polygon. While in the case of unconstrained priors, the entire admin-1 polygon is used to define the sampling position prior for these admin-1 sequences.

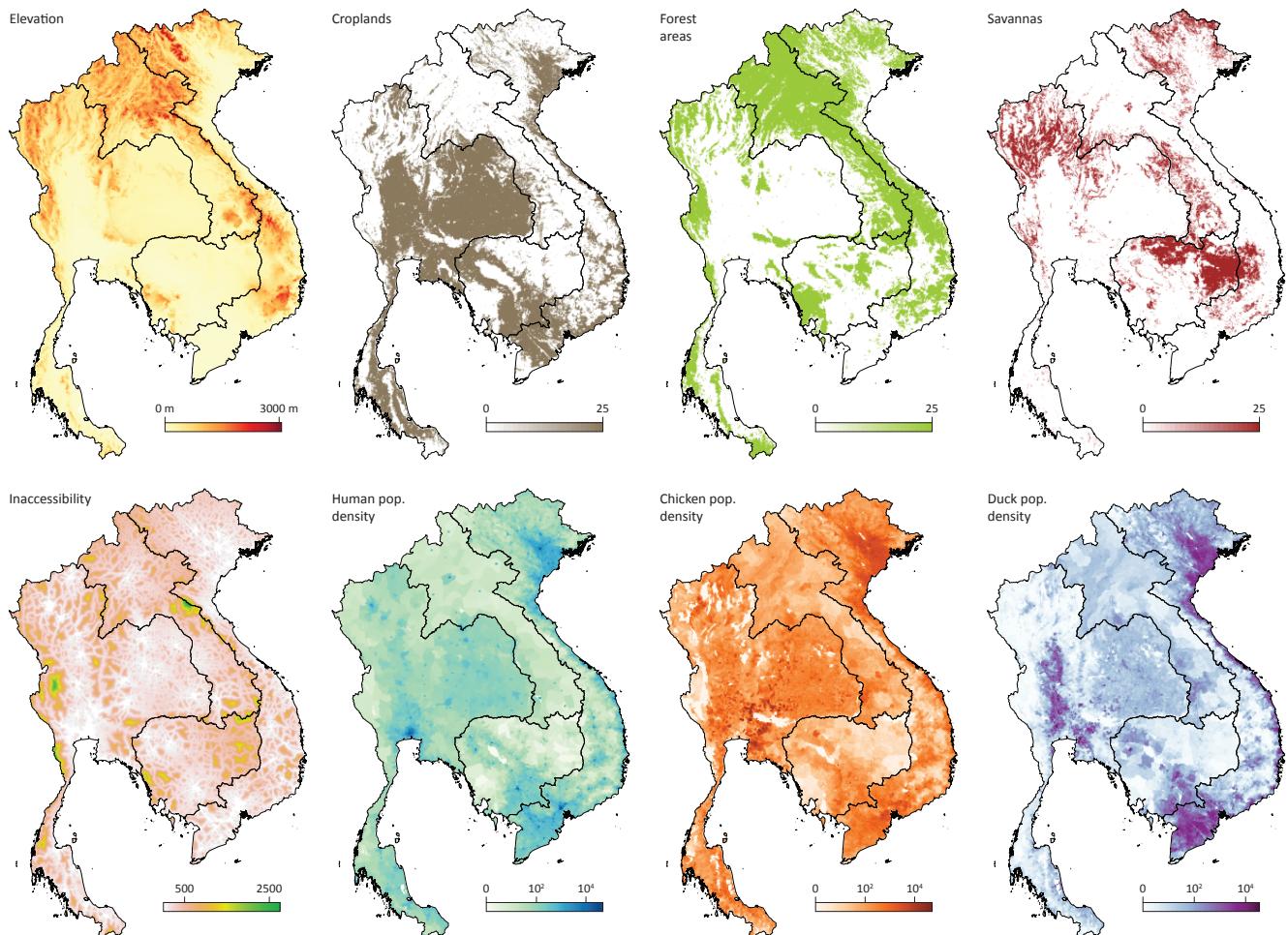


Figure S3. Environmental variables that were tested as factors that could have impacted lineage dispersal velocity. The elevation raster was obtained from the SRTM (Shuttle Radar Topography Mission) database (webmap.ornl.gov), and the “inaccessibility” raster from the Global Environment Monitoring Unit of the Joint Research Centre of the European Commission (bioval.jrc.ec.europa.eu). As for the human, poultry and duck population density rasters, they were downloaded from the MAP (Malaria Atlas Project, www.map.ox.ac.uk) and Livestock Geo-Wiki (livestock.geo-wiki.org) databases. These population density rasters were log-transformed in order to avoid providing an excessive weight to a few areas associated with high values. Finally, the original land cover raster was obtained from IGBP (International Geosphere Biosphere Programme) database (www.igbp.net). We then generated distinct land cover rasters from the original raster (~0.5 arcmin) by creating lower resolution rasters (2.4 arcmin) whose cell values equalled the number of occurrences of each land cover category within the 2.4 arcmin cells. The resolution of all the other original rasters (~0.5 arcmin) was also decreased to 2.4 arcmin for tractability.

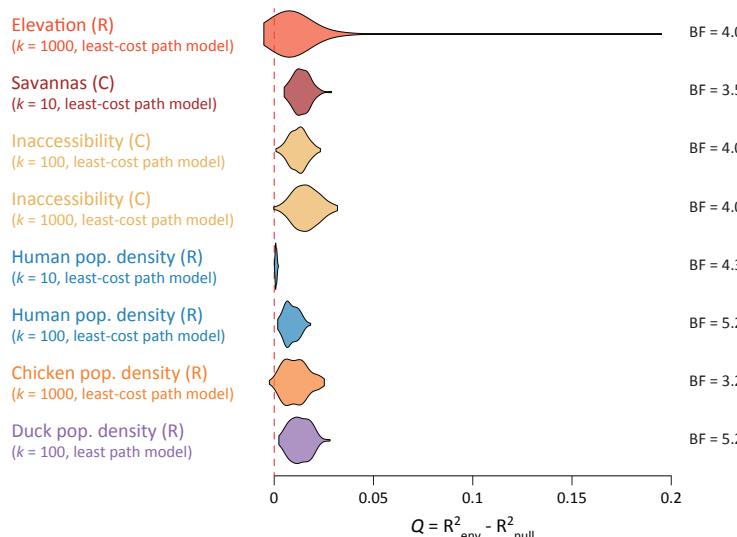


Figure S4. Analysis of the impact of several environmental factors on lineage dispersal velocity. These violin plots report the increase Q of the determination coefficient R^2 when performing a linear regression between dispersal durations and environmental distances computed on the environmental raster rather than on the “null” raster. Reported distributions of Q values are based on 100 posterior trees. Here, results are only reported for the combinations of environmental factor, parameter Q value and path model leading to a Q distribution significantly higher than 0 and associated with a Bayes factor value > 3 (see the text for further details).