



Figures and figure supplements

Deep sampling of Hawaiian *Caenorhabditis elegans* reveals high genetic diversity and admixture with global populations

Tim A Crombie et al

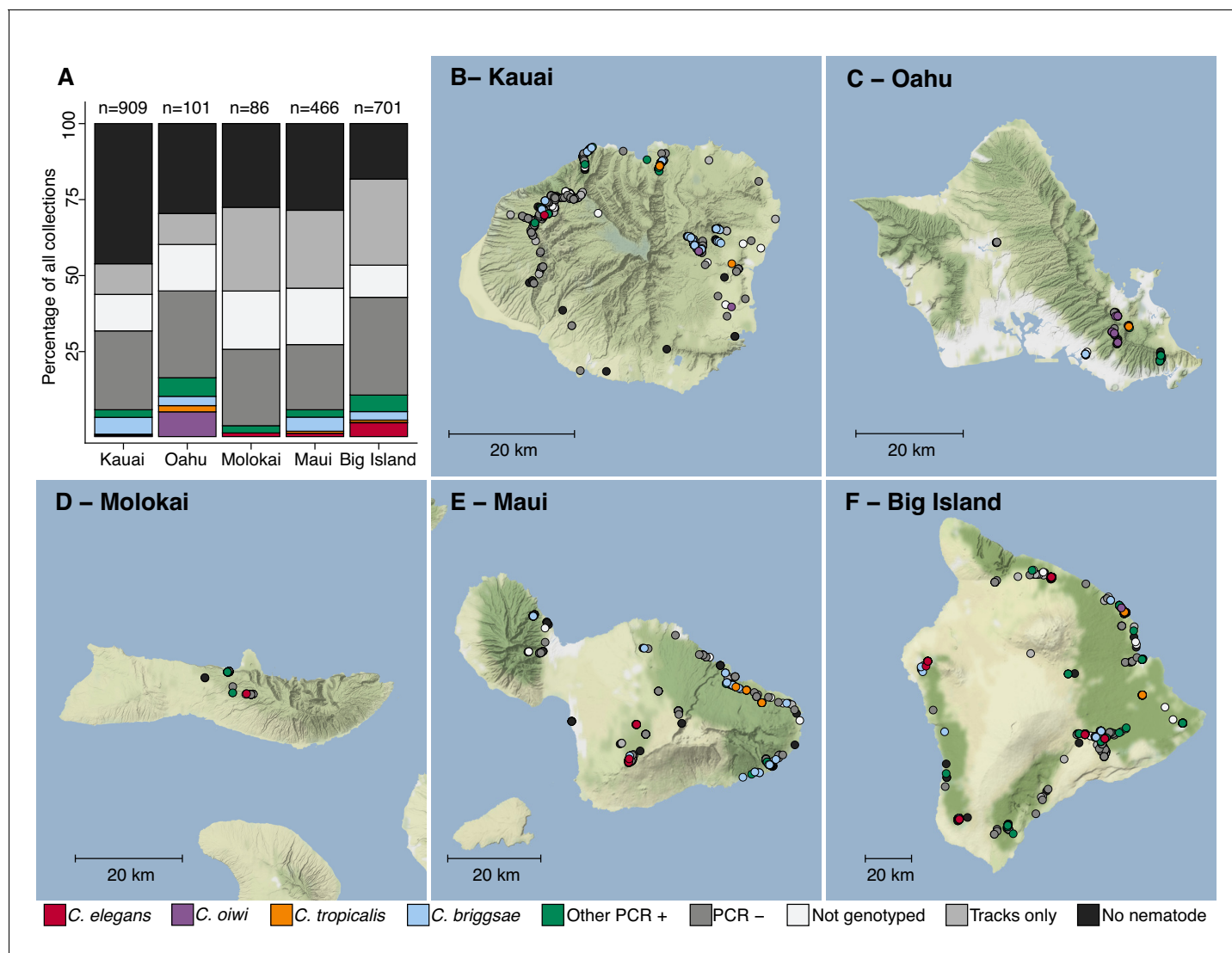


Figure 1. Geographic distribution of sampling sites across five Hawaiian islands. In total we sampled 2263 unique sites. **(A)** The percentage of each collection category is shown by island. The collection categories are colored according to the legend at the bottom of the panel, and the total number of samples for each island are shown above the bars. **(B–F)** The circles indicate unique sampling sites ($n = 2,263$) and are colored by the collection categories shown in the legend. For sampling sites where multiple collection categories apply ($n = 299$), the site is colored by the collection category shown in the legend from left to right, respectively. For all sampling sites, the GPS coordinates and collection categories found at that site are included in (**Source data 1**). We focused our studies on *Caenorhabditis* nematode collections, excluding *C. kamaaina* because it was only found at two sampling sites.

© 2019, Thunderforest. Maps in Panels B–F adapted from Thunderforest under a CC BY-NC-SA 2.0 license.

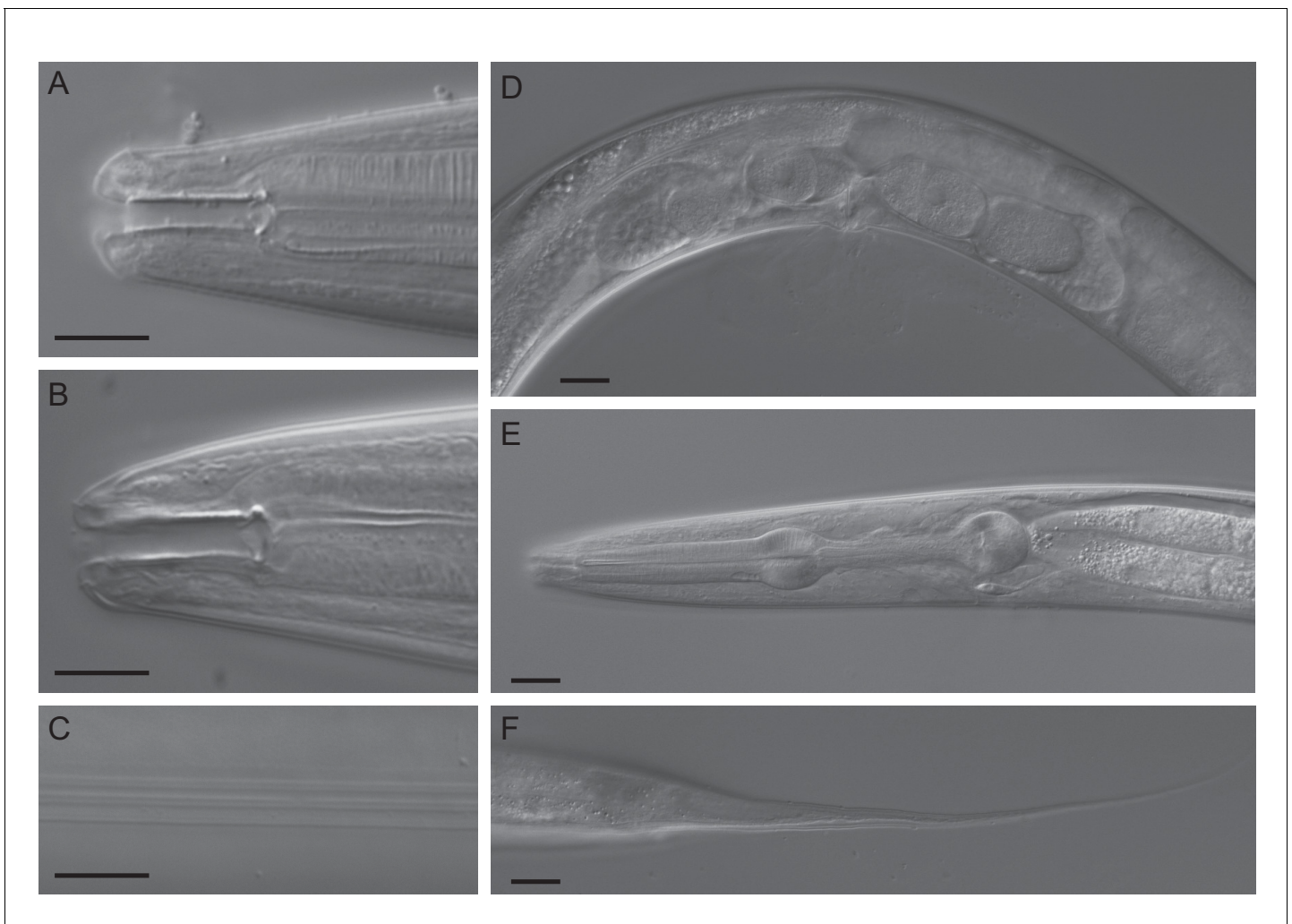


Figure 1—figure supplement 1. DIC micrographs of *C. oiwi* sp. n. (A) stoma of male (subventral right, dorsal is up); (B) stoma of female subventral right, dorsal is up); (C) female lateral field with alae; (D) female midbody region showing vulva, one embryo in each uterus, one oocyte in each spermatheca and part of the posterior ovary (left side view); (E) pharynx region of female (left side view); and (F) female tail (left side view). Scale bars in A-C are 10 μ m and 20 μ m in D-F. A formal description of *C. oiwi* is provided (Appendix 1).

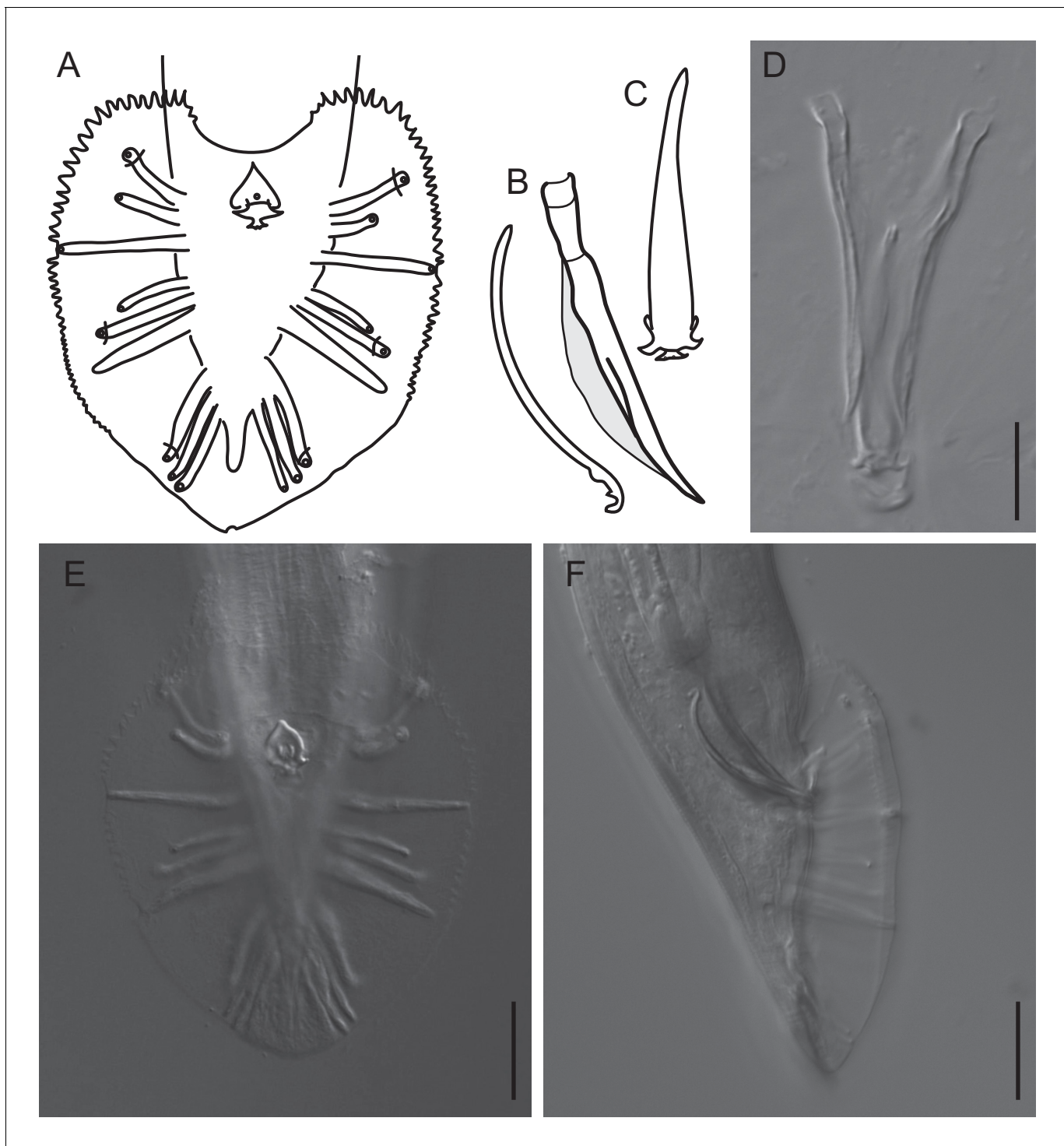


Figure 1—figure supplement 2. Features of the male tail of *C. oiwi* sp. n. (A) Drawing of the male tail in ventral view. The rays in position 1, 5 and 7 (from anterior) open to the dorsal side of the fan. (B) A drawing of the spicule and gubernaculum in right lateral view is shown. (C) A drawing of the gubernaculum in ventral view is shown. (D) DIC micrograph of the spicules and gubernaculum in ventral view is shown. (E, F) DIC micrographs of the male tail in ventral (E) and lateral right view (F) are shown. Scale bars are 20 μ m. A formal description of *C. oiwi* is provided (Appendix 1).

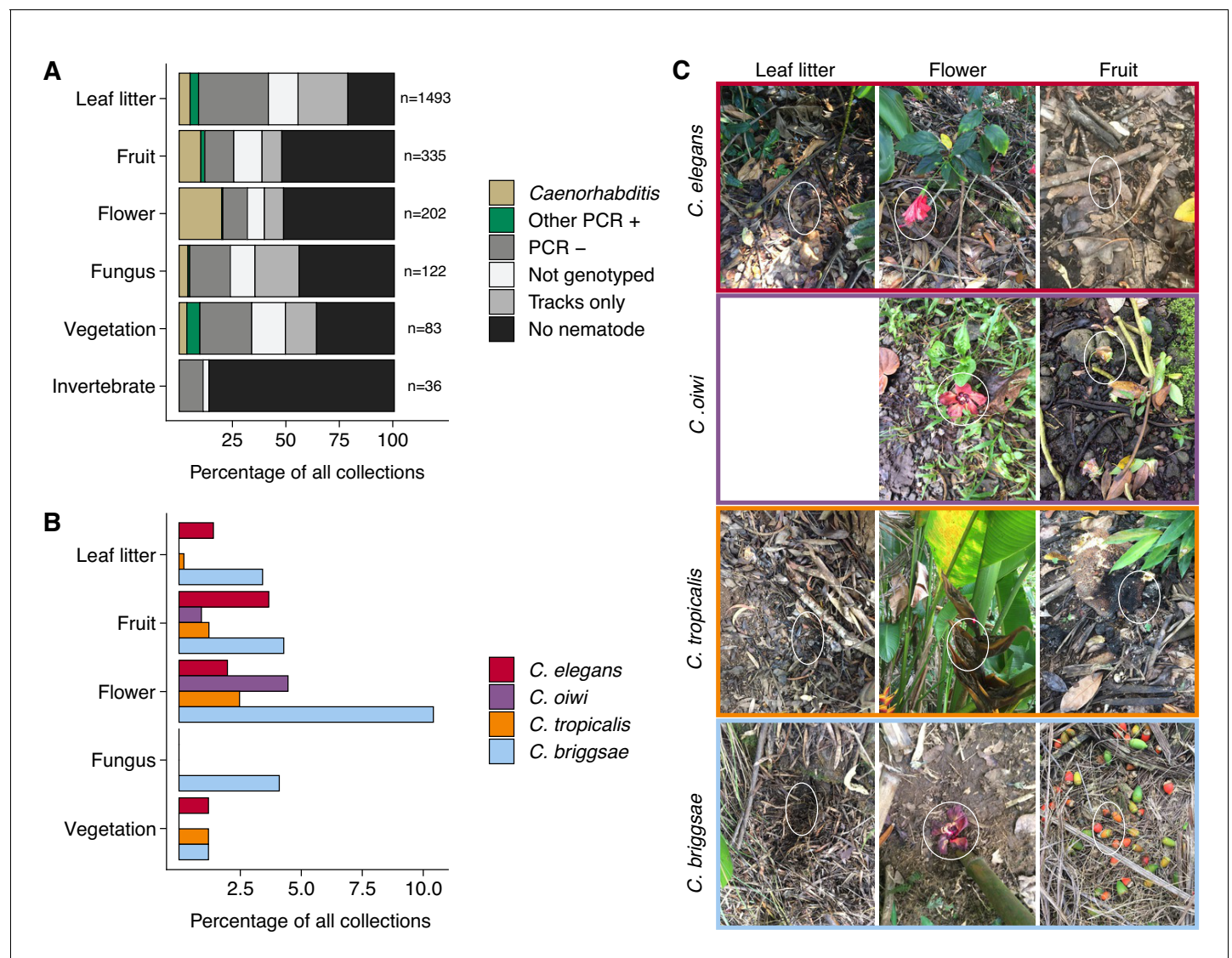


Figure 2. Collection categories by substrate type. (A) The percentage of each collection category is shown by substrate type. The collection categories are colored according to the legend at the right, and the total number of samples for each substrate are shown to the right of bars. (B) The percentage of collections is shown by substrate type for each *Caenorhabditis* species (excluding *C. kamaaina*, $n = 2$). (C) Examples of substrate photographs for *Caenorhabditis* species are shown and white ellipses indicate what was sampled. The *C. oiwi* leaf litter cell is blank because *C. oiwi* was only isolated from flowers and fruit.

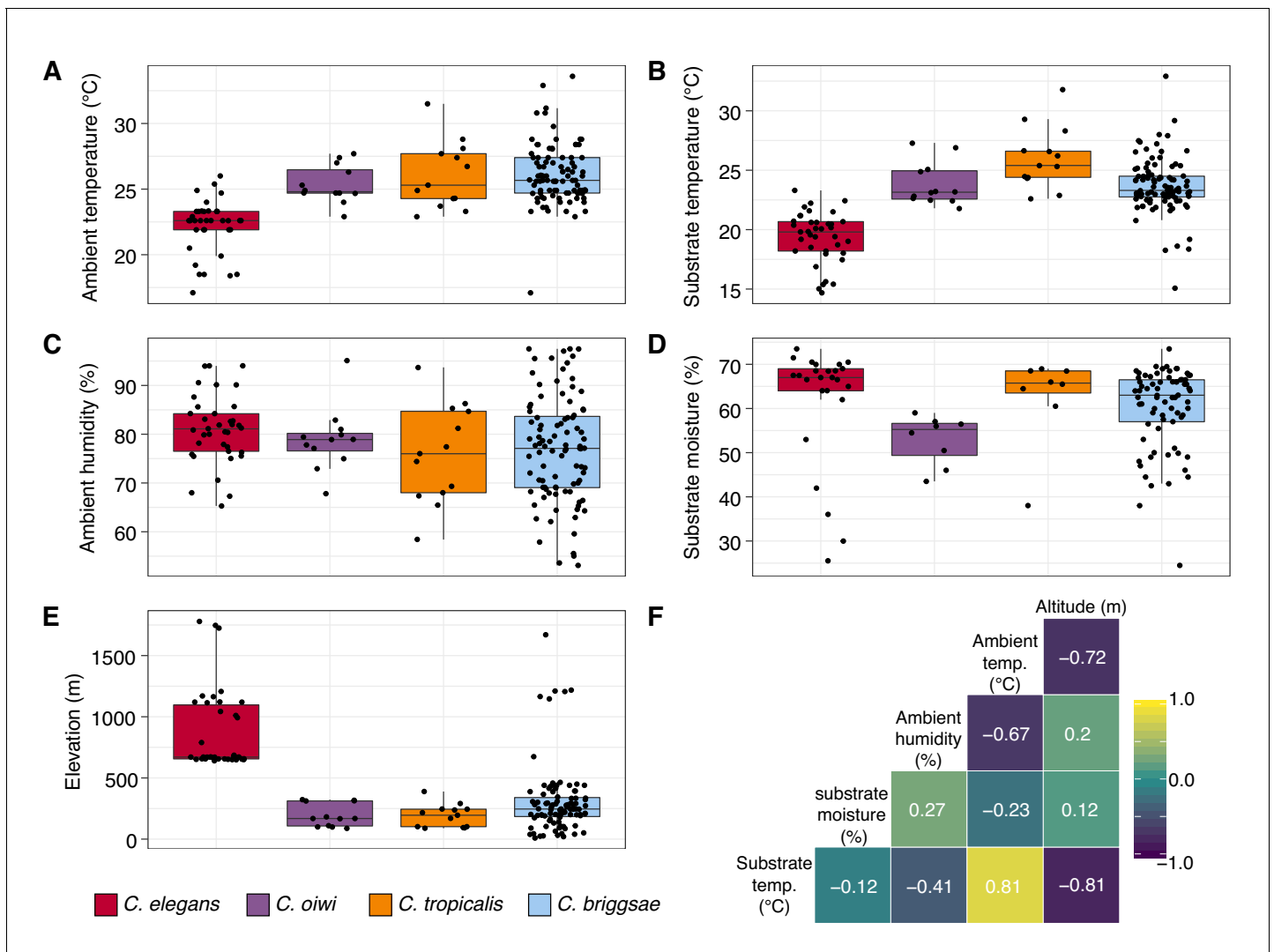


Figure 3. Environmental parameter values for sites where *Caenorhabditis* species were isolated. (A–E) Tukey box plots are plotted by species (colors) for different environmental parameters. Each dot corresponds to a unique sampling site where that species was identified. In cases where two *Caenorhabditis* species were identified from the same sample ($n = 3$), the same parameter values are plotted for both species. All p -values were calculated using Kruskal-Wallis test and Dunn test for multiple comparisons with p values adjusted using the Bonferroni method; comparisons not mentioned were not significant ($\alpha = 0.05$). (A) Ambient temperature (°C) was typically cooler at the sites where *C. elegans* were isolated compared to sites for all other *Caenorhabditis* species (Dunn test, $p < 0.005$). (B) Substrate temperature (°C) was also generally cooler for *C. elegans* than all other *Caenorhabditis* species (Dunn test, $p < 0.00001$). (C) Ambient humidity (%) did not differ significantly among the *Caenorhabditis*-positive sites. (D) Substrate moisture (%) was generally greater for *C. elegans* than *C. oiwi* (Dunn test, $p = 0.002$). (E) Elevation (meters) was typically greater at sites where *C. elegans* were isolated compared to sites for all other *Caenorhabditis* species (Dunn test, $p < 0.00001$). (F) A correlation matrix for the environmental parameters was made using sample data from the *Caenorhabditis* species shown. The parameter labels for the matrix are printed on the diagonal, and the Pearson correlation coefficients are printed in the cells. The color scale also indicates the strength and sign of the correlations shown in the matrix.

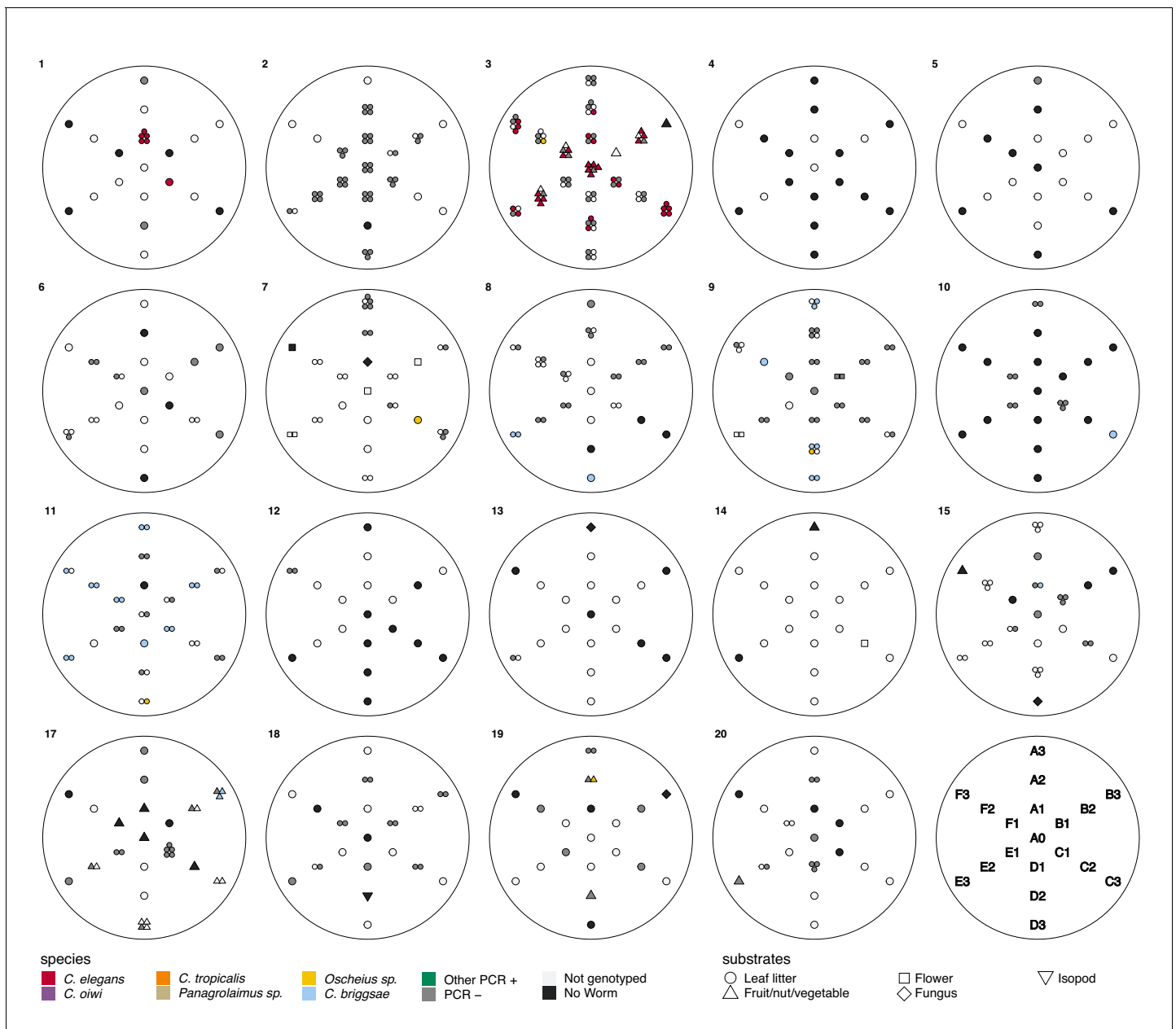


Figure 3—figure supplement 1. Local scale sampling with gridsects. Nineteen of twenty gridsects are shown. A key defining the sample positions for each gridsect is shown on the lower right. Samples were collected in six directions (A–F), each 60° apart. The center is denoted as A0, and samples were collected in each direction at one, two, and three meters from the center. Gridsect 16 is omitted because it was incomplete. The colors and shapes plotted at gridsect positions show the collection category and substrate class collected at that position as defined in the plot legend. We isolated multiple nematodes from some substrates, in these cases multiple shapes are plotted at that position to indicate the various collection categories that were found there.

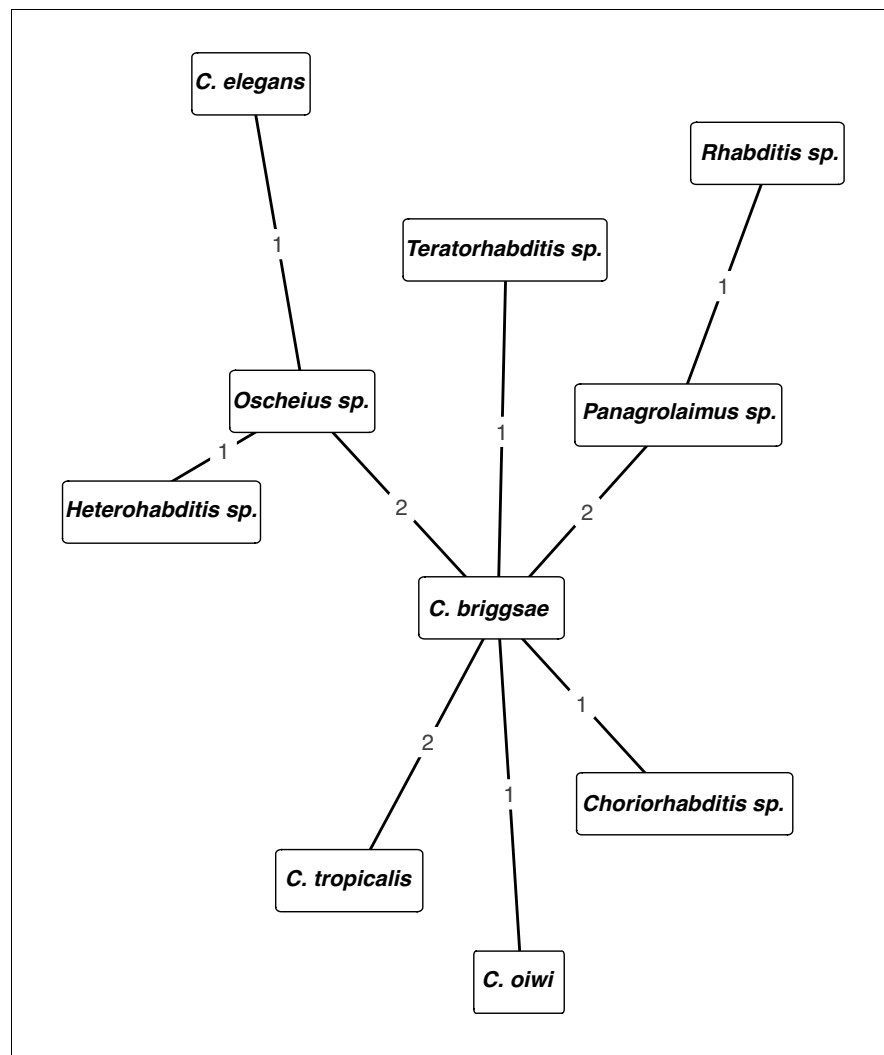


Figure 3—figure supplement 2. Network of cohabiting species isolated from samples. The nodes are labeled with the taxa. The edges are labeled with the number of times the taxa shown on the nodes were isolated from the same sample.

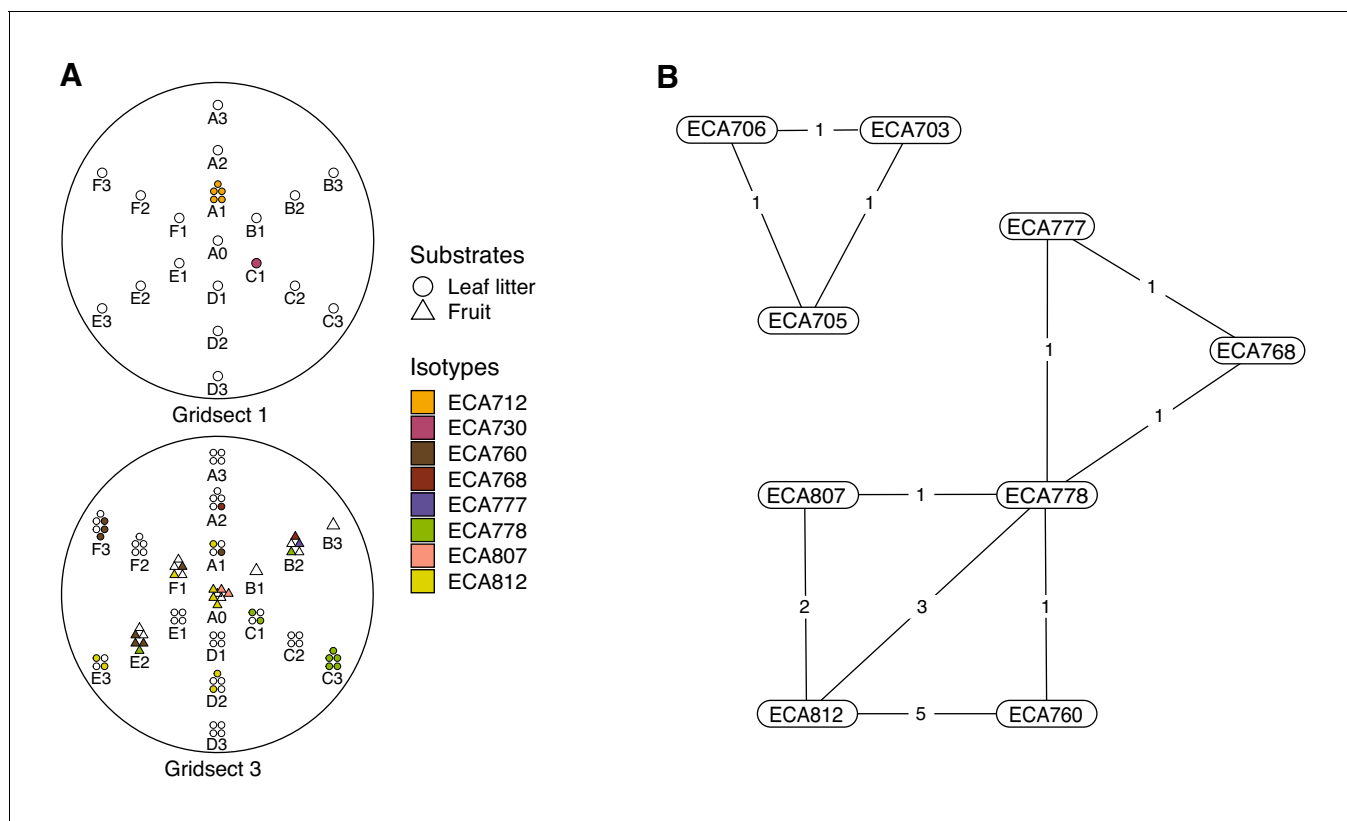


Figure 3—figure supplement 3. Local diversity and colocalization of isotypes. **(A)** Local scale gridsect sampling is shown. Each gridsect contains a total of 19 samples centered on the sample position labeled A0. The remaining sample positions are labeled by one of six transect lines (A–F) followed by the distance (in meters) from position A0. The shapes plotted above the position label show the collection category at that position as defined in the plot legend, and the colors correspond to *C. elegans* isotypes in the legend. Samples that did not contain *C. elegans* strains are colored white. **(B)** A colocalization network is shown for *C. elegans* isotypes from all Hawaiian samples. The numbers inset on the lines connecting two isotypes correspond to the number of unique samples where the two strains were isolated together.

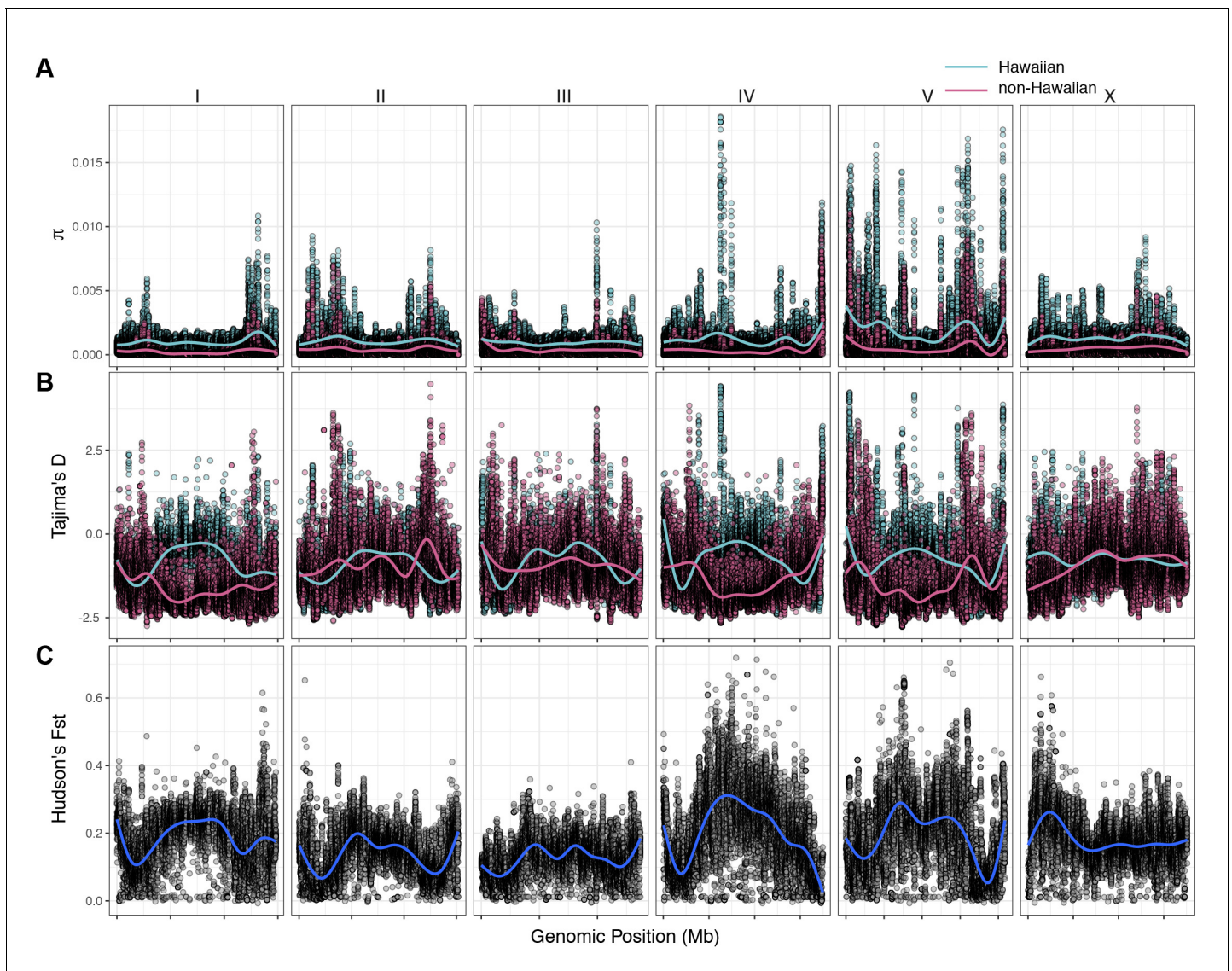


Figure 4. Chromosomal patterns of *C. elegans* diversity and differentiation. All comparisons are between the 43 Hawaiian isotypes and the 233 non-Hawaiian isotypes from the rest of the world. All statistics were calculated along a sliding window of size 10 kb with a step size of 1 kb. Each dot corresponds to the calculated value for a particular window. (A) Genome-wide π calculated for Hawaiian isotypes (light blue) and non-Hawaiian isotypes (pink) are shown. (B) Genome-wide Tajima's D statistics for Hawaiian isotypes (light blue) and non-Hawaiian isotypes (pink) are shown. (C) Genome-wide Hudson's F_{ST} comparing the Hawaiian and non-Hawaiian isotypes are shown.

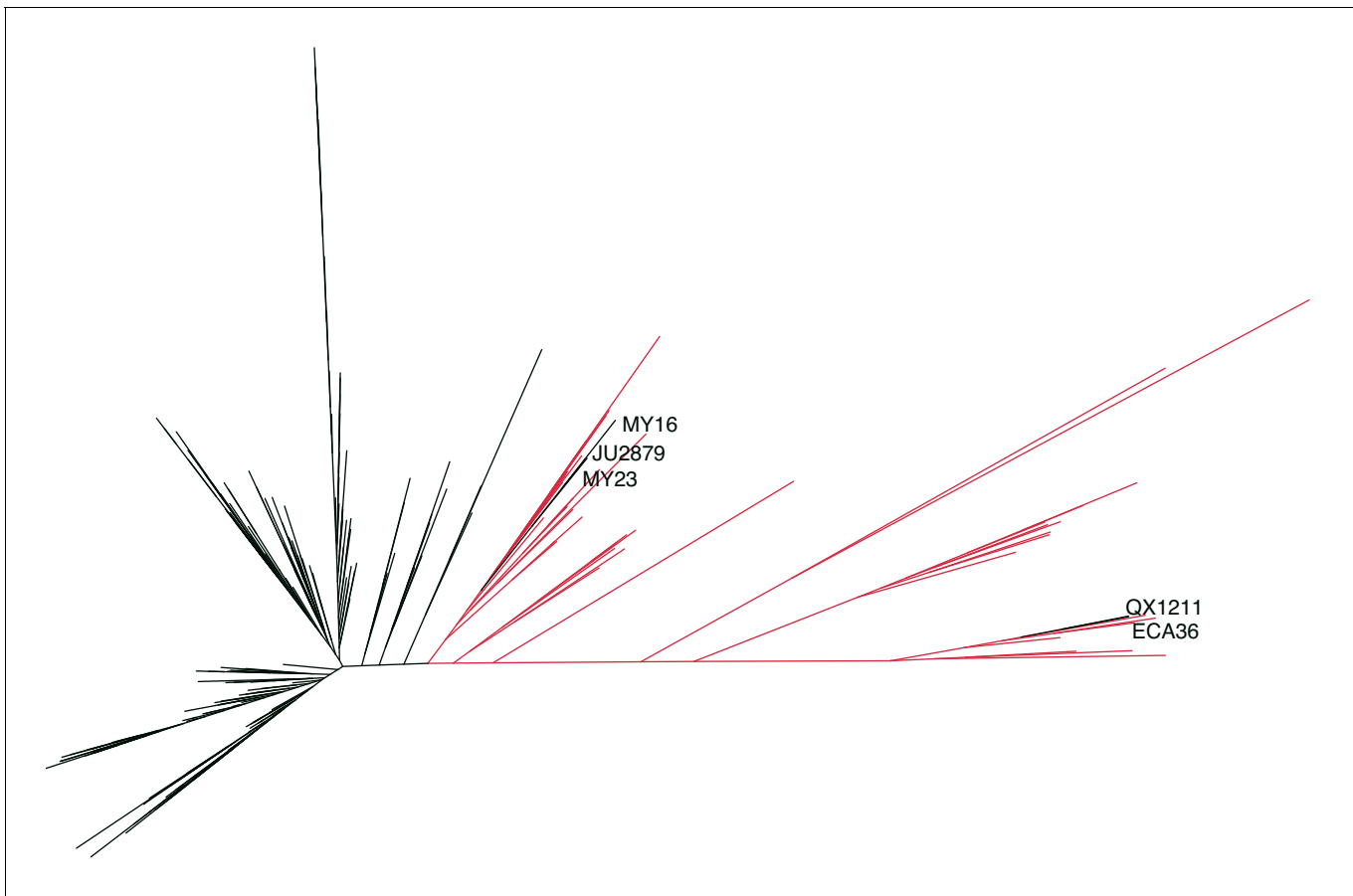


Figure 4—figure supplement 1. *Caenorhabditis elegans* unrooted tree for 276 isotypes. A maximum likelihood tree built using single nucleotide variants found in the 276 *C. elegans* isotypes sampled, including the 26 new Hawaiian isotypes. (Substitution model: GTR+FO). The isotypes labeled in red were isolated from the Hawaiian Islands (n = 43). The five isotypes labeled are non-Hawaiian isotypes that group within the Hawaiian isotypes.

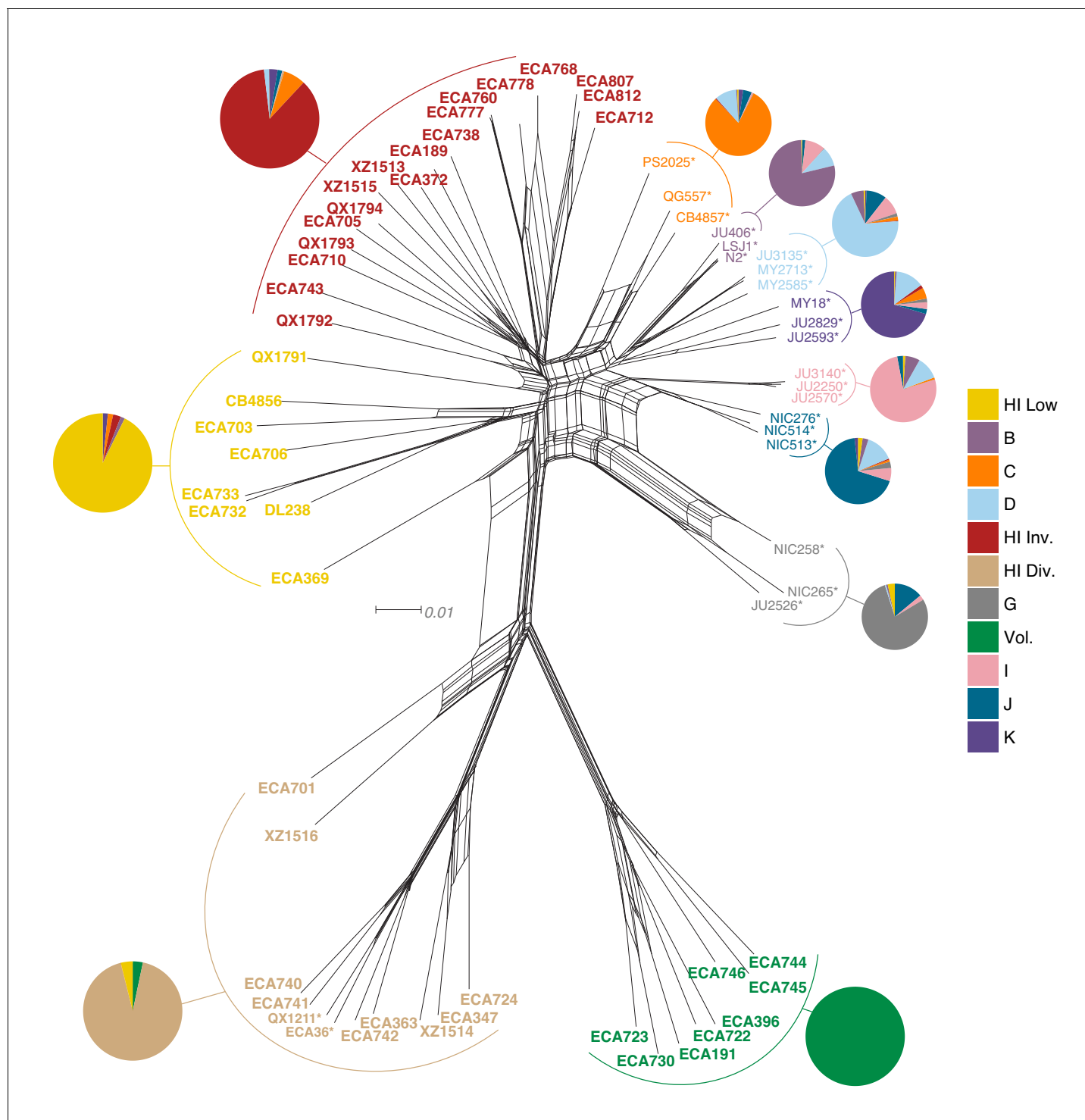


Figure 5. Relatedness of the Hawaiian *C. elegans* isotypes. Neighbor-joining net showing the genetic relatedness of the Hawaiian *C. elegans* isotypes relative to a representative set of non-admixed, non-Hawaiian isotypes from each population defined by ADMIXTURE ($K = 11$). Colors of isotype names indicate the maximum fraction of population assignment from ADMIXTURE ($K = 11$), including the seven non-Hawaiian populations (B–K) and the four Hawaiian populations (Hawaiian Invaded, Hawaiian Low, Hawaiian Divergent, and Volcano). Isotypes labeled with an asterisk are representative of non-admixed, non-Hawaiian isotypes from each population defined by ADMIXTURE ($K = 11$). Pie charts represent population proportions for all isotypes within the full admixture population.

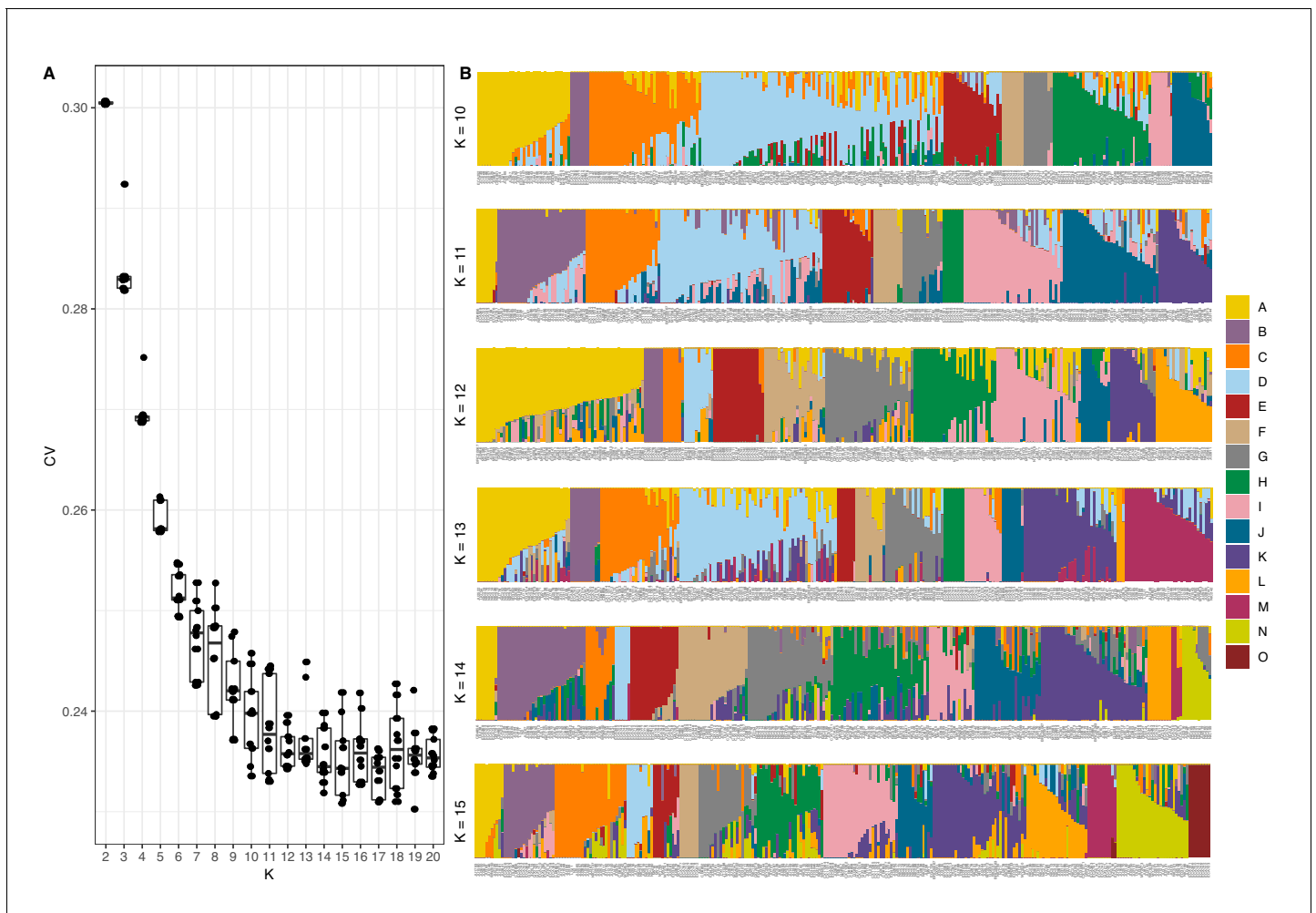


Figure 5—figure supplement 1. Summary of ADMIXTURE analysis. (A) Tukey boxplots of ten independent ADMIXTURE runs showing the cross-validation error on the y-axis for the number of populations (K) ranging from 2 to 20 on the x-axis. (B) The inferred population proportions estimated by ADMIXTURE are shown on the y-axis of the all *C. elegans* isotypes on the x-axis. Each isotype is represented by a vertical line, which is partitioned into colored segments that represent the isotype's membership fractions for the populations shown in the legend.

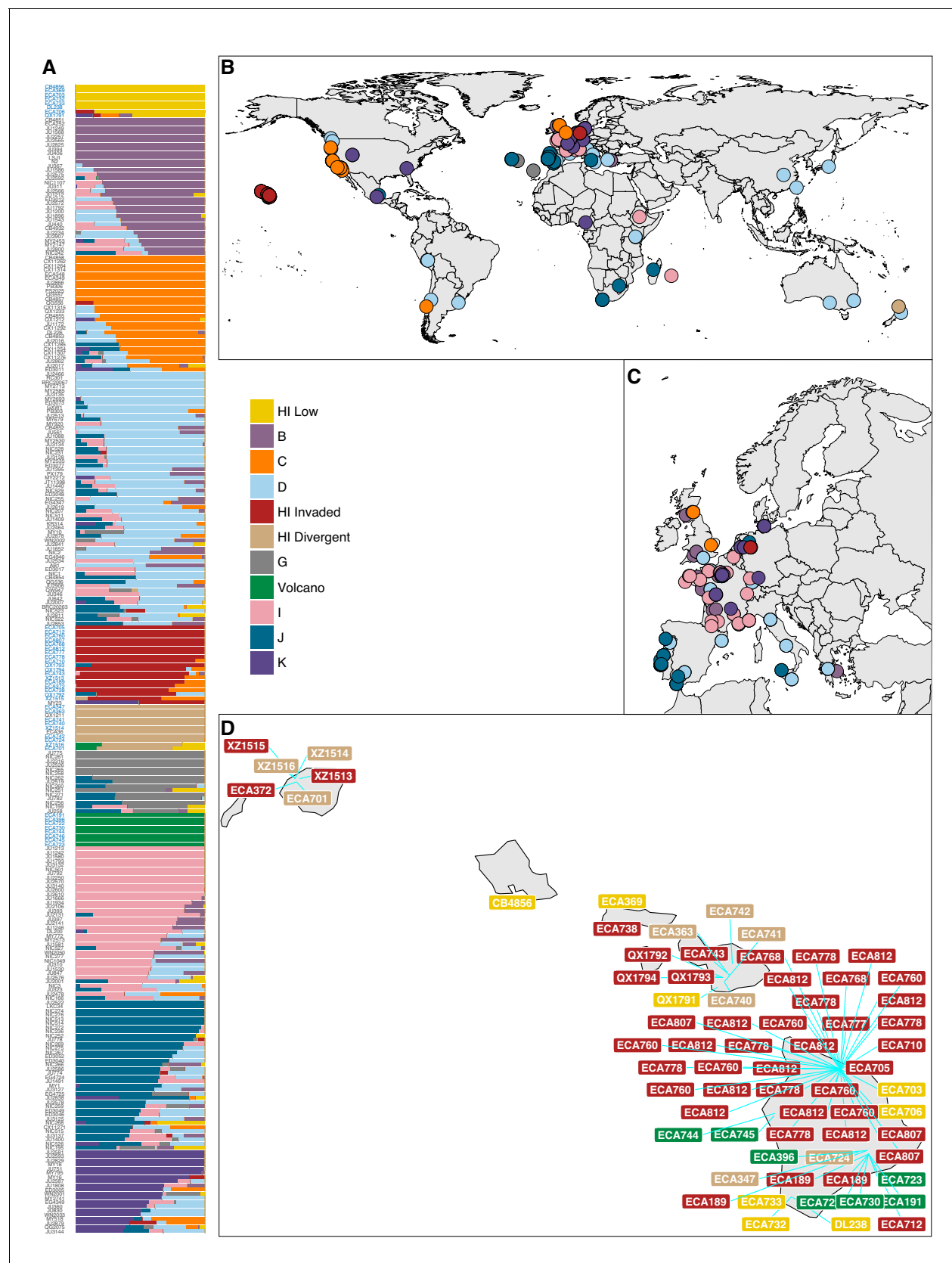


Figure 5—figure supplement 2. Global population structure. (A) The inferred population proportions estimated by ADMIXTURE (K = 11) with *C. elegans* isotype names on the y-axis. The names colored in blue represent Hawaiian isotypes. (B) The global distribution of all 276 isotypes are shown Figure 5—figure supplement 2 continued on next page

Figure 5—figure supplement 2 continued

with colors corresponding to the legend. Colors are assigned based on the largest ancestral population fraction for that isotype (e.g. the isotype MY23 from Germany was assigned as 51% 'Hawaii Invaded' and 49% global 'K' but it is colored red on the map for 'Hawaii Invaded'). (C) The same data are shown but with more resolution in Europe. (D) All distinct collections of Hawaiian isotypes are shown with labels for isotype names colored by the ancestral population assignment. The blue lines point to specific collection locations for those isotypes. The cluster of 'Hawaiian Invaded' isotypes on the Big Island correspond to gridsect three and adjacent collections from the Kalōpā state recreation area.

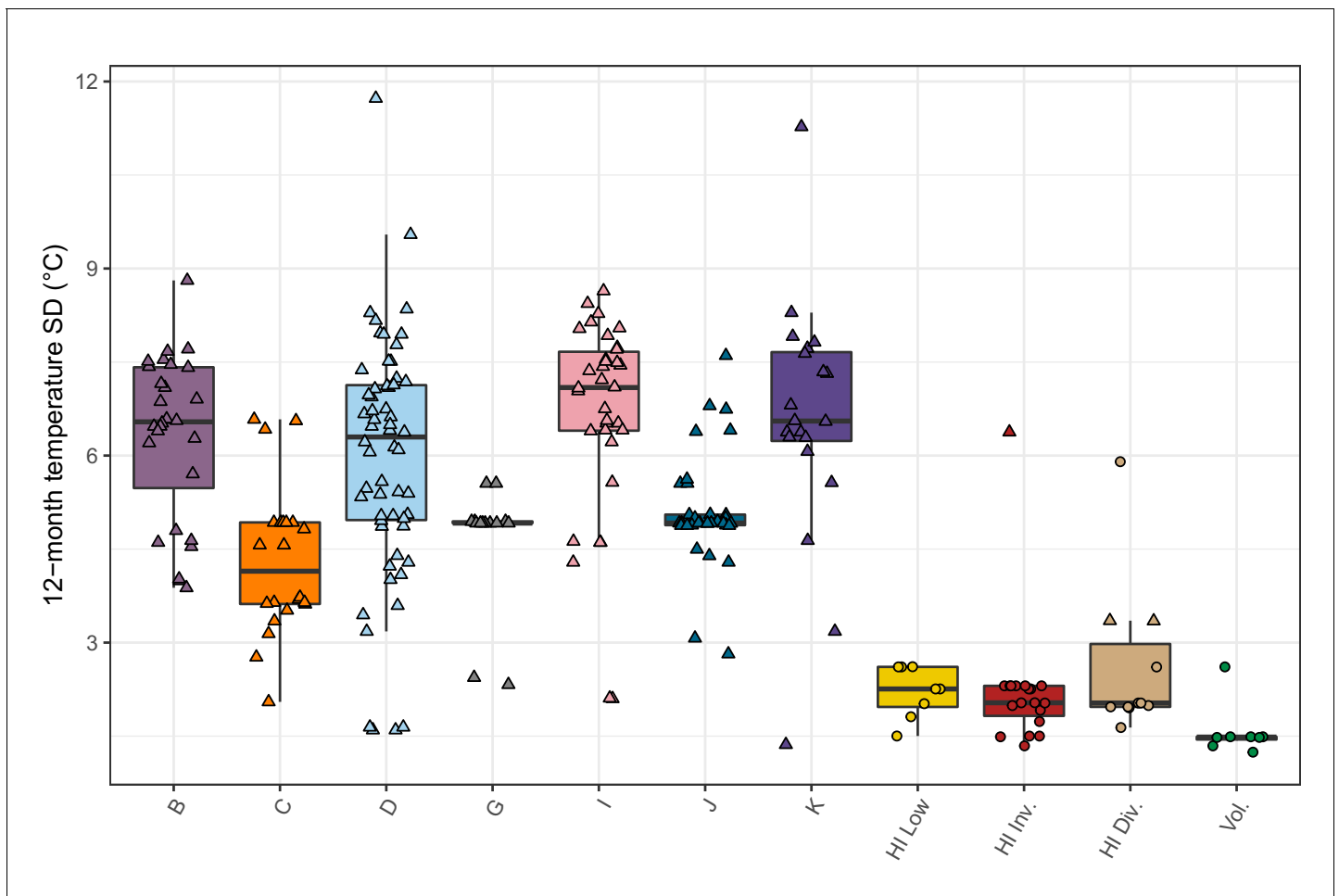


Figure 5—figure supplement 3. Seasonal temperature variation by population. The standard deviation of daily mean temperatures for a 12-month period centered on the collection date for each isolate is shown. If only the year of collection is known, then the 12-month period is centered on January 1 st of that year, and if the year and month are known but not the exact date, then the 12-month period is centered on the first of that month. The data points correspond to isotypes and are colored by their assigned populations from admixture analysis. The Hawaiian isotypes are plotted as circles and non-Hawaiian isotypes are plotted as triangles.

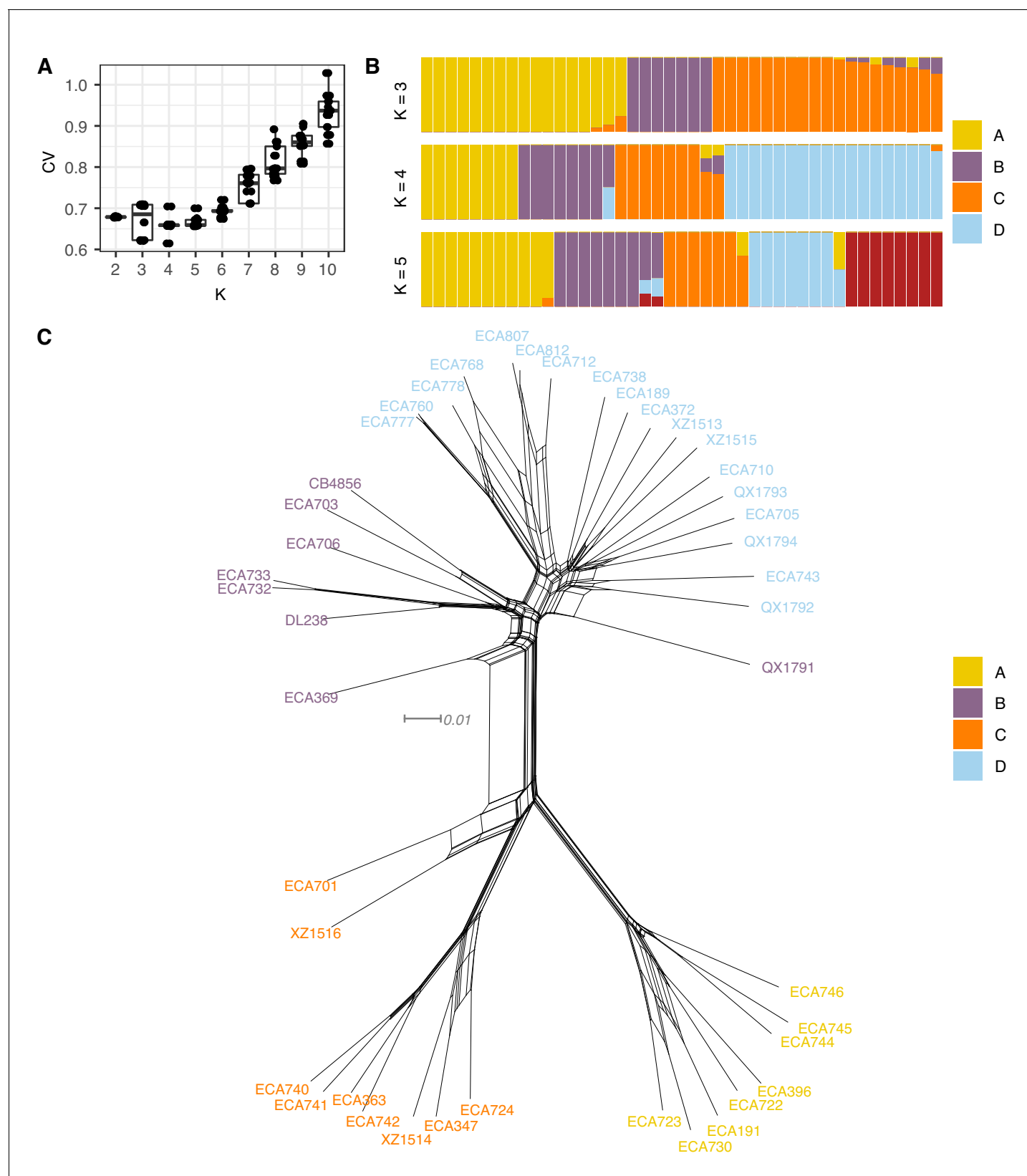


Figure 5—figure supplement 4. Summary of ADMIXTURE analysis on Hawaiian *C. elegans* isotypes. (A) Tukey boxplots of ten independent ADMIXTURE runs showing the cross-validation error on the y-axis for the population number (K) ranging from 2 to 10 on the x-axis. (B) The inferred population proportions estimated by ADMIXTURE are shown on the y-axis for the Hawaiian *C. elegans* isotypes on the x-axis. Each isotype is

Figure 5—figure supplement 4 continued on next page

Figure 5—figure supplement 4 continued

represented by a vertical line, which is partitioned into colored segments that represent the isotype's membership fractions in the clusters shown in the legend. (C) A neighbor-joining net showing the genetic relatedness of the Hawaiian isotypes is shown. Colors of labels indicate the largest fractional population assignment from ADMIXTURE ($K = 4$).

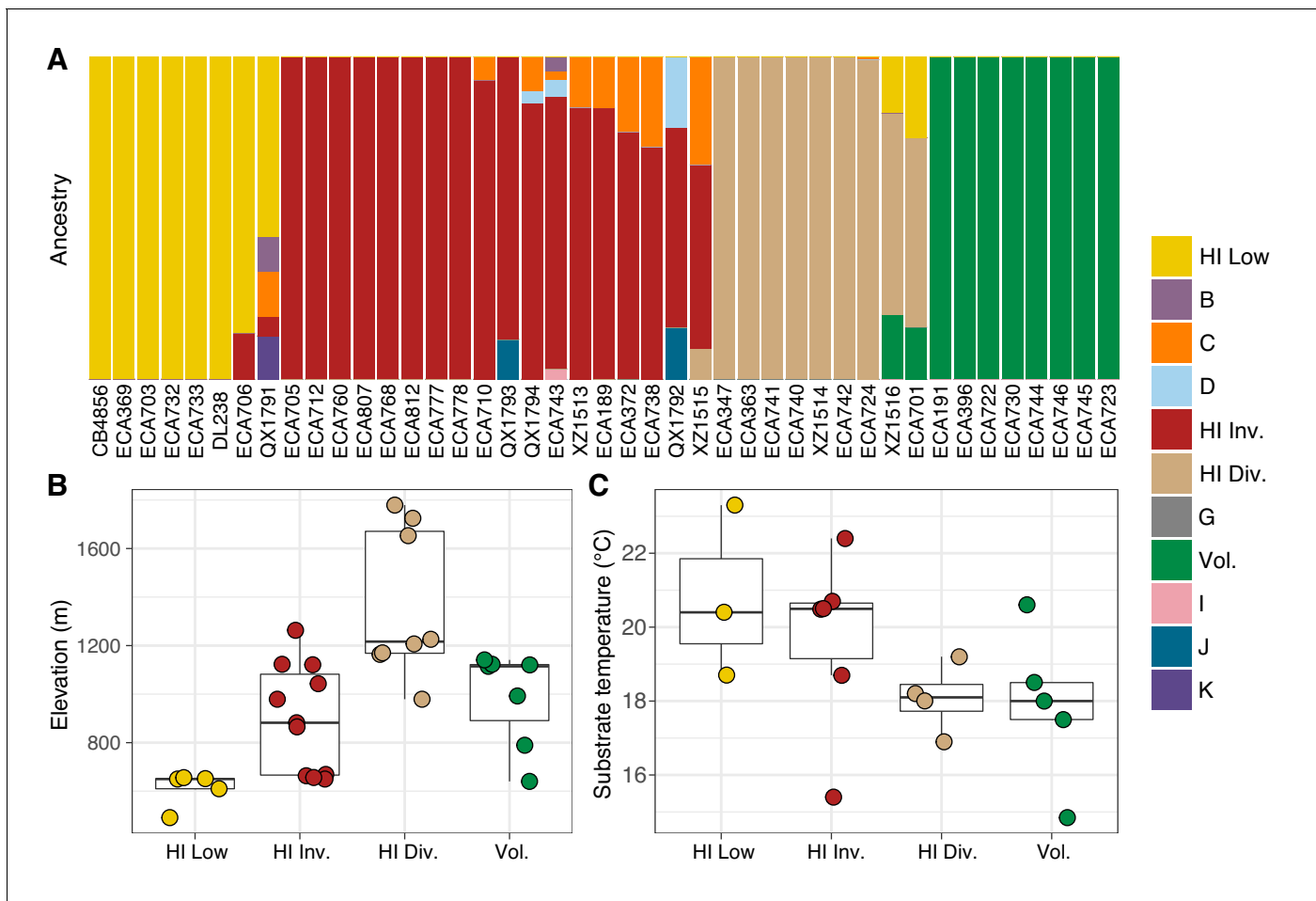


Figure 6. Environmental parameters of Hawaiian *C. elegans* populations. (A) The inferred ancestral population fractions for each Hawaiian isotype as estimated by ADMIXTURE ($K = 11$; 276 *C. elegans* isotypes) are shown. The bar colors represent the fraction of population assignments from ADMIXTURE for the isotypes named on the x-axis. (B–C) Tukey box plots are shown by population assignments (colors) for different environmental parameters. We used the average values of environmental parameters from geographically clustered collections to avoid biasing our results by local oversampling (See Materials and methods - Environmental parameter analysis). All p-values were calculated using Kruskal-Wallis test and Dunn test for multiple comparisons with p values adjusted using the Bonferroni method; comparisons not mentioned were not significant ($\alpha = 0.05$). (B) The collection site elevations for Hawaiian isotypes colored by population assignments are shown. The Hawaiian Low and the Hawaiian Invaded populations were typically found at lower elevations than the Hawaiian Divergent population (Dunn test, p-values=0.000168, and 0.037 respectively). (C) The substrate temperatures for Hawaiian isotypes colored by population assignments are shown.

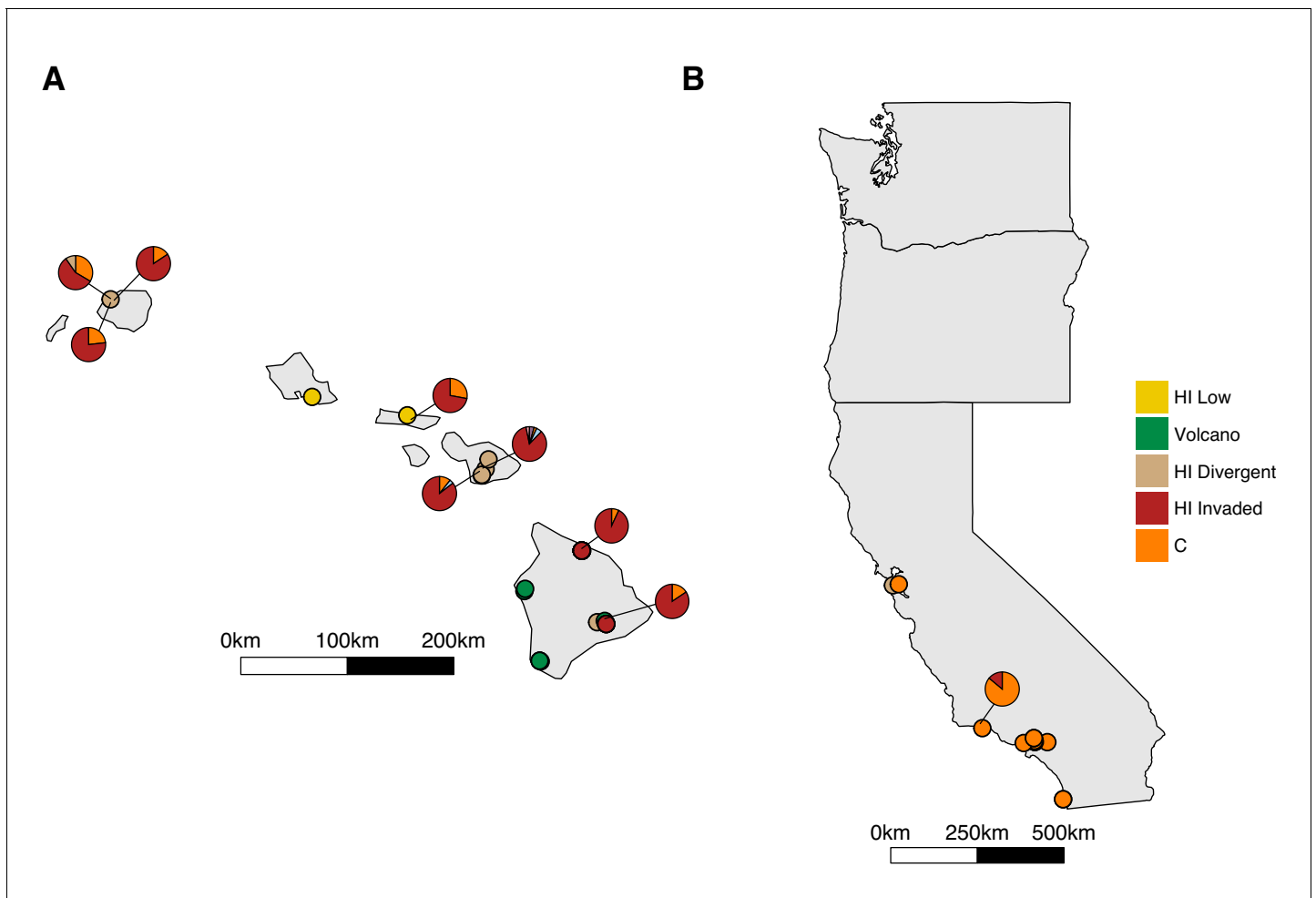


Figure 6—figure supplement 1. Admixture between the Hawaii Invaded and the global C populations. Unique isolation locations for non-admixed isotypes from the four Hawaiian populations and the global C population (solid circles) are shown for Hawaii (A) and the west coast of the United States (B). Unique isolation locations for isotypes from the Hawaiian Invaded and global C populations that are admixed with one another (pie charts) are shown for Hawaii (A) and the west coast of the United States (B). The colors correspond to the admixture populations in the legend.

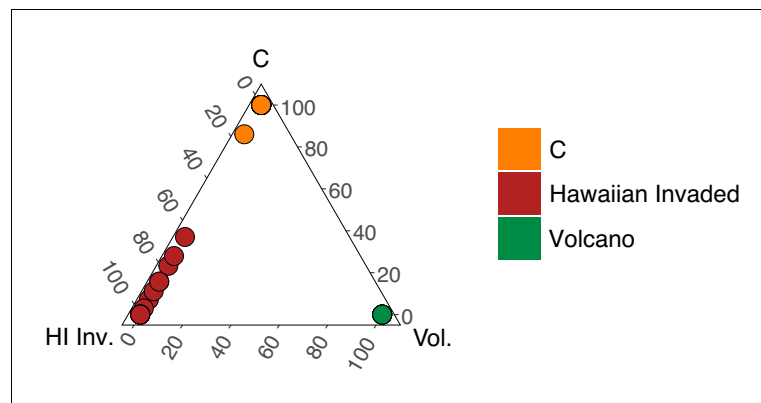


Figure 6—figure supplement 2. Admixture between the 'Hawaiian Invaded' and 'global C' populations. The fraction of admixture among the three populations is shown on a ternary plot. The data points correspond to isotypes and are colored by the highest fractional population assignment.

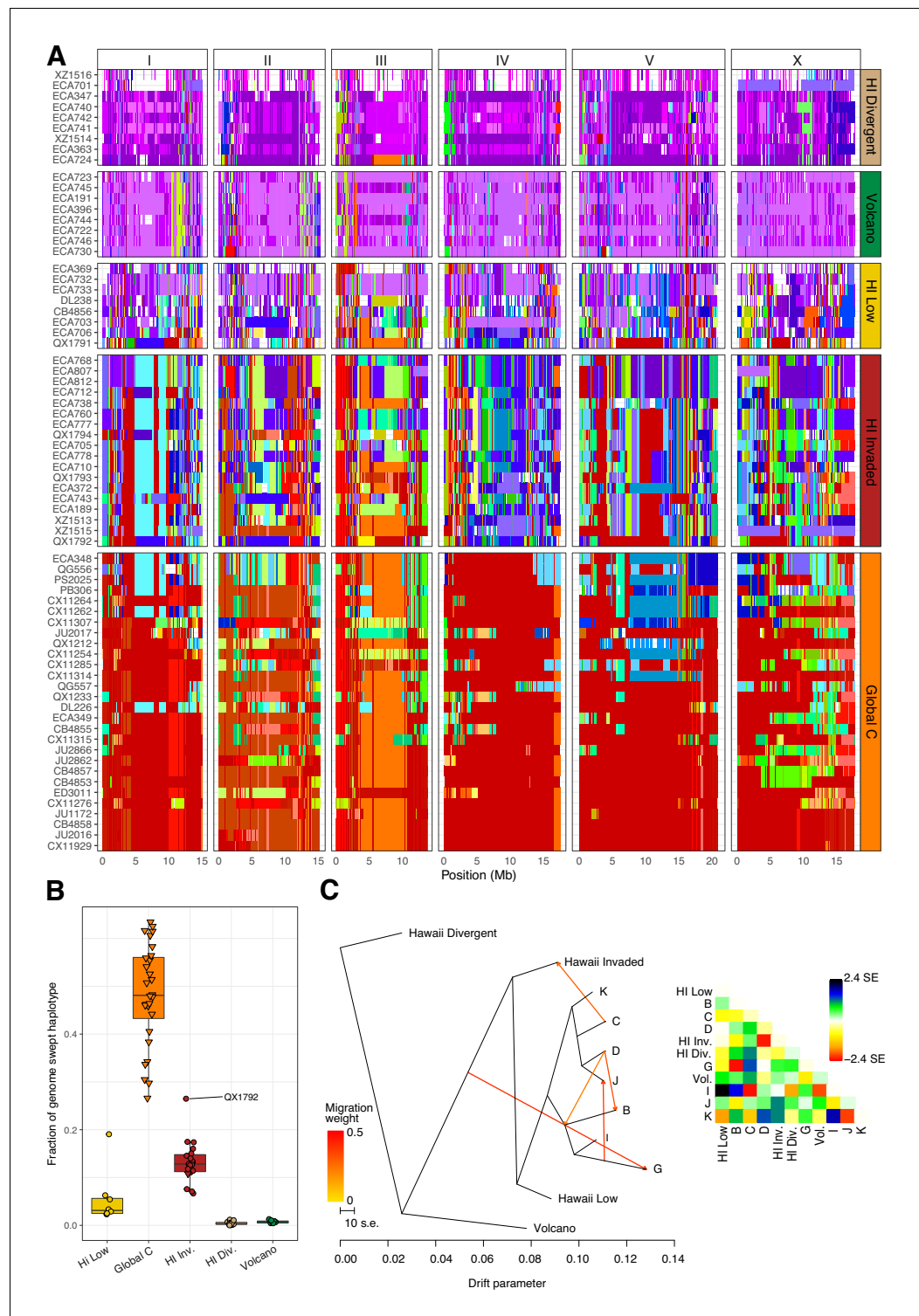


Figure 7. Evidence of migration between the Hawaiian and world populations. (A) The haplotypes or inferred blocks of identity by descent (IBD) across the genome are shown. The genomic position is plotted on the x-axis for each isotype plotted on the y-axis. The block colors correspond to a uniquely defined IBD group. The dark red blocks correspond to the most common global haplotype (*i.e.*, the swept haplotypes on chr I, IV, V, and left of X). Genomic regions with no color represent regions for which no IBD groups could be determined. The four Hawaiian populations are shown in the top four facets, excluding non-Hawaiian isotypes. The bottom facet shows the non-Hawaiian C population. (B) The total fraction of the genome with the swept haplotype is shown by Figure 7 continued on next page

Figure 7 continued

population. The data points correspond to isotypes and are colored by their assigned populations. The Hawaiian isotypes are plotted as circles and non-Hawaiian isotypes are plotted as triangles. Hawaiian isotypes with greater than 25% of their genome swept are labelled. (C) The inferred relationship among the populations allowing for five migration events (ADMIXTURE, $K = 11$). The heat map to the right represents the residual fit to the migration model.

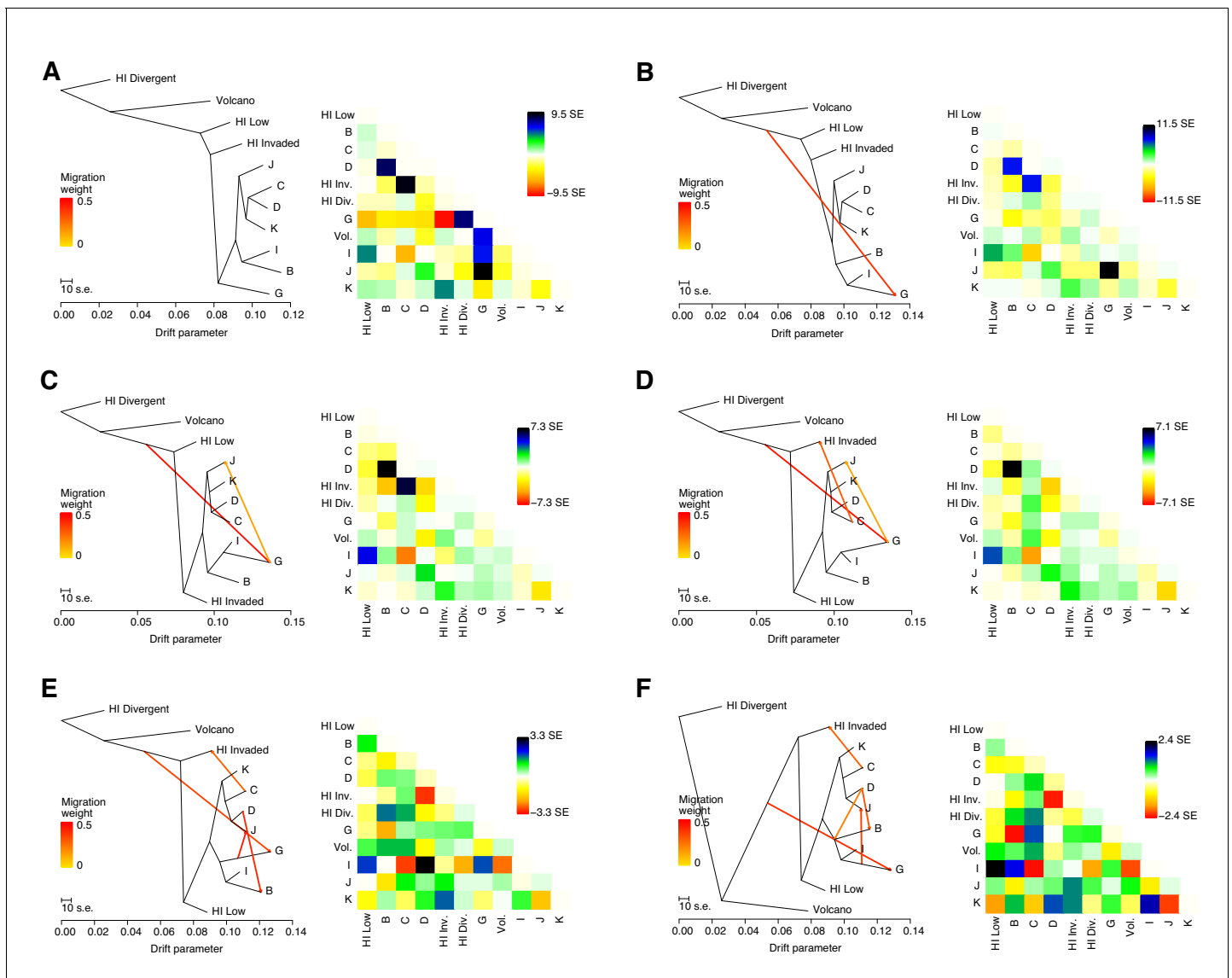


Figure 7—figure supplement 1. Evidence of migration between Hawaiian and non-Hawaiian populations. (A–D) The inferred relationships among the populations (ADMIXTURE, $K = 11$) with zero (A), one (B), two (C), three (D), four (E), or five (F) migration events. The right panels contain heat maps corresponding to the residual fit to the migration models.

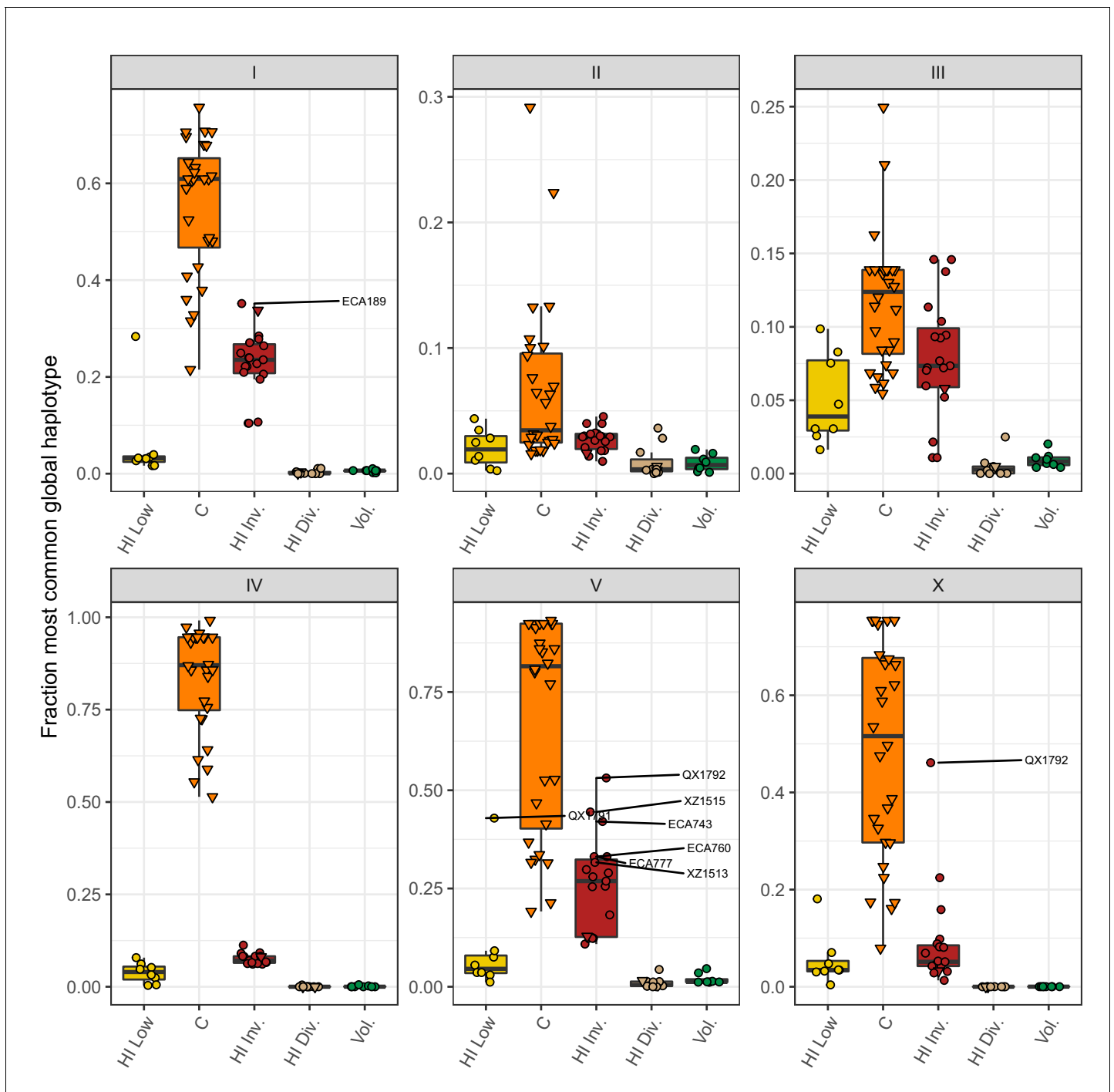


Figure 7—figure supplement 2. Most common global haplotype sharing by chromosome and population. The fraction of each chromosome that belongs to the most common global haplotype is shown on the y-axis. The data points correspond to isotypes and are colored by their assigned populations. The Hawaiian isotypes are plotted as circles and non-Hawaiian isotypes are plotted as triangles. Hawaiian isotypes with greater than 30% of a chromosome belonging to the most common global haplotype are labelled in the chromosome facet. The most common global haplotype is synonymous with the globally swept haplotype on chromosomes I, IV, V, and the left of X.