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

Timothy A. Crombie[1], Stefan Zdraljevic[1,2], Daniel E. Cook[1,2], Robyn E. Tanny[1], Shannon C. Brady[1,2], Ye Wang[1], Kathryn S. Evans[1,2], Steffen Hahnel[1], Daehan Lee[1], Briana C. Rodriguez[1], Gaotian Zhang[1], Joost van der Zwaag[1], Karin C. Kiontke[3], and Erik C. Andersen[1,*]

1. Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA
2. Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL 60208, USA
3. Department of Biology, New York University, New York, NY 10003, USA
* Corresponding author

Erik C. Andersen
Associate Professor of Molecular Biosciences
Northwestern University
Evanston, IL 60208, USA
Tel: (847) 467-4382
Email: erik.andersen@northwestern.edu

Timothy A. Crombie, tcrombie@northwestern.edu, ORCID 0000-0002-5645-4154
Stefan Zdraljevic, stefanzdraljevic2018@u.northwestern.edu, ORCID 0000-0003-2883-4616
Daniel E. Cook, daniel.cook@crick.ac.uk, ORCID 0000-0003-3347-562X
Robyn E. Tanny, robyn.tanny@northwestern.edu, ORCID 0000-0002-0611-3909
Shannon C. Brady, shannonbrady2014@u.northwestern.edu, ORCID 0000-0002-3043-1544
Ye Wang, ye.wang@northwestern.edu, ORCID 0000-0002-5423-6196
Kathryn S. Evans, kathryn.evans@u.northwestern.edu, ORCID 0000-0002-1388-8155
Steffen Hahnel, steffenhahnel@directbox.com, ORCID 0000-0001-8848-0691
Daehan Lee, daehan.lee@northwestern.edu, ORCID 0000-0002-0546-8484
Briana C. Rodriguez, briana.rodriguez@northwestern.edu, ORCID 0000-0002-5282-0815
Gaotian Zhang, gaotian.zhang@northwestern.edu, ORCID 0000-0001-6468-1341
J van der Zwaag, joost.vanderzwaag@wur.nl
Karin C. Kiontke, kk52@nyu.edu, ORCID 0000-0003-1588-4884
Erik Andersen, erik.andersen@northwestern.edu, ORCID 0000-0003-0229-9651

Running title: Hawaiian *C. elegans* diversity

Keywords:
Caenorhabditis, C. elegans, genetic diversity, niche, admixture

Abstract

Hawaiian isolates of the nematode species *Caenorhabditis elegans* have long been known to harbor genetic diversity greater than the rest of the worldwide population, but this observation was supported by only a small number of wild strains. To better characterize the niche and genetic diversity of Hawaiian *C. elegans* and other *Caenorhabditis* species, we sampled different substrates and niches across the Hawaiian islands. We identified hundreds of new *Caenorhabditis* strains from known species and a new species, *Caenorhabditis oiwi*. Hawaiian *C. elegans* are found in cooler climates at high elevations but are not associated with any specific substrate, as compared to other *Caenorhabditis* species. Surprisingly, admixture analysis revealed evidence of shared ancestry between some Hawaiian and non-Hawaiian *C. elegans* isolates. We suggest that the deep diversity we observed in Hawaii might represent patterns of ancestral genetic diversity in the *C. elegans* species before human influence.

Introduction

Over the last 50 years, the nematode *Caenorhabditis elegans* has been central to many important discoveries in the fields of developmental, cellular, and molecular biology. The vast majority of these insights came from the study of a single laboratory-adapted strain collected in Bristol, England known as N2 (Brenner, 1974; Chalfie et al., 1994; Consortium, 1998; Fire et al., 1998; Grishok et al., 2000; Hodgkin and Brenner, 1977; Lee et al., 1993; Sulston et al., 1983). Recent sampling efforts have led to the identification of numerous wild *C. elegans* strains and enabled the study of genetic diversity and ecology of the species (Andersen et al., 2012; Barrière and Félix, 2014; Cook et al., 2016; Félix and Duveau, 2012; Ferrari et al., 2017; Hahnel et al., 2018; Lee et al., 2019; Richaud et al., 2018). The earliest studies of *C. elegans* genetic variation showed that patterns of single-nucleotide variant (SNV) diversity were shared among most wild strains, with the exception of a Hawaiian strain, CB4856, which has distinct and high levels of variation relative to other strains (Cutter, 2006; Koch et al., 2000; Rockman and Kruglyak, 2009). Subsequent analyses revealed that *C. elegans* has reduced levels of diversity relative to the obligate outcrossing *Caenorhabditis* species and the facultative selfer *C. briggsae* (Dey et al., 2013; Thomas et al., 2015). The most comprehensive analysis of *C. elegans* genetic diversity to date used data from thousands of genome fragments across a globally distributed collection of 97 genetically distinct strains to show that recent selective sweeps have largely homogenized the genome (Andersen et al., 2012). The authors hypothesized that these selective sweeps might contain alleles that facilitate human-assisted dispersal and/or increase fitness in human-associated habitats. Consistent with the previous analyses, two Hawaiian strains, CB4856 and DL238, did not share patterns of reduced genetic diversity caused by the selective sweeps that affected the rest of the *C. elegans* population – a trend that has held true as the number of Hawaiian strains has increased (Cook et al., 2017, 2016; Hahnel et al., 2018; Lee et al., 2019). Taken together, these studies suggest that the Hawaiian *C. elegans* population might be more representative of ancestral genetic diversity that existed prior to the selective pressures associated with recent human influence.

To better characterize the genetic diversity of the *C. elegans* species on the Hawaiian Islands, we performed deep sampling across five Hawaiian islands: Kauai, Oahu, Molokai, Maui, and the Big Island. Because incomplete data on locations and environmental parameters are common issues for some field studies of *C. elegans* (Andersen et al., 2012; McGrath et al., 2009; Rockman and Kruglyak, 2009), we developed a standardized collection procedure with the Fulcrum® mobile data collection application. This streamlined procedure enabled us to rapidly record GPS coordinates and environmental niche parameters at each collection site, and accurately link these data with the nematodes we isolated. The Hawaiian Islands are an ideal location to study characteristics of the *C. elegans* niche because the Islands contain many steep, wide-ranging gradients of temperature, humidity, elevation, and landscape usage. In total, we collected samples from 2,263 sites across the islands and isolated 2,531 nematodes, including 309 individuals from the *Caenorhabditis* genus. Among these isolates, we identified 100 new *C. elegans* strains, 95 of which proliferated in the lab and were whole-genome sequenced. Analysis of genomic variation revealed that these

strains represent 26 distinct genome-wide haplotypes not sampled previously. We refer to these genome-wide haplotypes as isotypes. We grouped these 26 Hawaiian isotypes with the 17 previously isolated Hawaiian isotypes and compared their genetic variation to 233 non-Hawaiian isotypes from around the globe. Consistent with previous observations, we found that the Hawaiian isotypes had approximately three times more diversity than the non-Hawaiian isotypes. However, we were surprised to find that, in a subset of Hawaiian isotypes, some genomic regions appear to be shared with non-Hawaiian isotypes from around the globe. These results provide the first evidence of gene flow between these populations and suggest that future sampling efforts in the Hawaiian Islands and the surrounding Pacific region will help elucidate the evolutionary processes that have shaped the genetic diversity in the *C. elegans* species.

Results

Hawaiian nematode diversity

In August 2017, we collected a total of 2,263 samples across five Hawaiian islands and ascertained the presence of nematodes in each sample (**Figure 1, Supplementary File 1, Supplementary Data 1**). We isolated one or more nematodes from 1,120 of 2,263 (49%) samples, and an additional 431 of 2,263 (19%) samples had circumstantial evidence of nematodes (tracks were observed but no nematodes could be isolated). Altogether, we isolated 2,531 nematodes from 1,120 samples and genotyped them by analysis of the Internal Transcribed Spacer (ITS2) region between the 5.8S and 28S rDNA genes (Barrière and Félix, 2014; Kiontke et al., 2011). We refer to isolates where the ITS2 region was amplified by PCR as 'PCR-positive', isolates with no amplification as 'PCR-negative', and isolates from which we could not extract high-quality genomic DNA for PCR as 'Not genotyped' (see Methods). The PCR-positive category comprises *Caenorhabditis* isolates that we identified to the species level and isolates from genera other than *Caenorhabditis* that we identified to the genus level. Using this categorization strategy, we found that 427 of 2,531 isolates (17%) were PCR-positive and belonged to 13 distinct taxa. Among all isolates, we identified five *Caenorhabditis* species at different frequencies across the 2,263 samples: *C. briggsae* (4.2%), *C. elegans* (1.7%), *C. tropicalis* (0.57%), *C. kamaaina* (0.088%), and a new species *C. oiwi* (0.53%) (**Supplementary Data 1**). We formally describe a new species (**Supplementary File 2**), which we named *Caenorhabditis oiwi* for the Hawaiian word meaning "native" in reference to its endemic status on the Hawaiian Islands. This species was found to be distinct based on molecular barcodes (Kiontke et al., 2011) and on biological species inference from crosses (Félix et al., 2014). *C. briggsae* was the most common *Caenorhabditis* species we isolated, which is consistent with nematode collection efforts by other groups that suggest *C. briggsae* is a ubiquitous species in many regions of the world (Félix et al., 2013). We found that two *Caenorhabditis* species, *C. elegans* and *C. oiwi*, were enriched on certain islands. *C. elegans* was enriched on the Big Island relative to Kauai and Maui (Fisher's Exact Test, $p < 0.01$), and *C. oiwi* was enriched on Oahu relative to Kauai, Maui, and the Big Island (Fisher's Exact Test, $p < 0.00001$).

The *C. elegans* niche is distinct from other *Caenorhabditis* species on Hawaii

To characterize more about a nematode niche on the Hawaiian Islands, we classified the substrate for each distinct collection and measured various environmental parameters at the collection sites. Of the six major classes of substrate, we found nematodes most often on leaf litter (55%) (**Figure 2A**). When we account for collections with nematode-like tracks on the collection plate, we estimated that greater than 80% of leaf litter substrates contained nematodes (**Figure 2A**). The isolation success rate for the other classes of substrate ranged from 14% to 49% (**Figure 2A**). In comparison to overall nematode isolation rates, *Caenorhabditis* nematodes were isolated more frequently from flower substrates (40 of 202 collections) than any other substrate category (Fisher's Exact Test, $p < 0.026$) (**Figure 2A**). We also found that *Caenorhabditis* nematodes were enriched on fruit (33 of 327 collections) relative to leaf litter substrates (76 of 1493 collections) (Fisher's Exact Test, $p < 0.011$) but not other substrate classes (**Figure 2A**). These findings are

consistent with other collection surveys that have shown leaf litter substrates harbor fewer *Caenorhabditis* nematodes than rotting flowers and fruits (Félix et al., 2013; Ferrari et al., 2017). We observed similar trends of flower-substrate enrichment relative to leaf litter for *C. briggsae* (Fisher's Exact Test, $p = 0.00044$; flower, 21 of 202 collections and leaf litter 51 of 1493 collections) and *C. tropicalis* (Fisher's Exact Test, $p = 0.0056$; flower, five of 202 collections and leaf litter, three of 1493 collections) but not for *C. elegans* (Fisher's Exact Test, $p = 1$), which exhibited no substrate enrichment (**Figure 2B-C**). Interestingly, the new species, *C. oiwi*, was only isolated from flower and fruit and was enriched on flower substrates (Fisher's Exact Test, $p = 0.013$; flower, nine of 202 collections; and fruit, three of 327 collections) (**Figure 2B-C**).

The enrichment of *C. briggsae*, *C. tropicalis*, and *C. oiwi* on flowers might indicate that this substrate class has a higher nutrient quality for these species. If this hypothesis is correct, we might expect to see a greater incidence of proliferating populations on flower substrates than other substrates. However, we saw no observable association between large population size (approximate number of nematodes on collection plate) and substrate class for *C. briggsae* (Spearman's *rho* = -0.0198, $p = 0.566$ flower vs. leaf litter), *C. tropicalis* (Spearman's *rho* = -0.258, $p = 0.732$ flower vs. leaf litter), nor *C. oiwi* (Spearman's *rho* = 0.258, $p = 0.209$ flower vs. fruit), which suggests that other factors might drive the observed flower enrichment or that we are limited by the small sample size. Taken together, these data suggest that the *Caenorhabditis* species we isolated do not exhibit substrate specificity, despite flower-substrate preferences of *C. briggsae*, *C. tropicalis*, and *C. oiwi*, which is different from some other species in the genus that demonstrate substrate specificity (*e.g.*, *C. astrocarya* and *C. inopinata*) (Ferrari et al., 2017; Kanzaki et al., 2018).

In addition to recording substrate classes, we measured elevation, ambient temperature and humidity, and substrate temperature and moisture to determine if these niche parameters were important for individual *Caenorhabditis* species (**Figure 3**; see Methods). Consistent with previous *C. elegans* collections in tropical regions (Andersen et al., 2012; Dolgin et al., 2008), all *C. elegans* isolates were collected from elevations greater than 500 meters and were generally found at higher elevations than other *Caenorhabditis* species (**Figure 3E;** mean = 867 m; elevation: Dunn test, $p < 0.00001$). We also found that *C. elegans*-positive collections tended to be at cooler ambient and substrate temperatures than other *Caenorhabditis* species (ambient temperature: Dunn test, $p < 0.005$; substrate temperature: Dunn test, $p < 0.00001$), although these two environmental parameters were correlated with elevation (**Figure 3F**). Notably, the average substrate temperatures for *C. elegans* (19.4 ºC), *C. tropicalis* (26.0 ºC), and *C. briggsae* (23.7ºC) positive collections are close to the optimal growth temperatures for these species in the laboratory setting (**Figure 3B**) (Poullet et al., 2015). Our collections also indicate that *C. oiwi* tends to be found on drier substrates than *C. elegans* (**Figure 3D**; Dunn test, $p = 0.0021$), but we observed no differences among species for ambient humidity (**Figure 3C**). Given the similar substrate and environmental parameter preferences of *C. tropicalis*, *C. briggsae*, and *C. oiwi*, we next asked if these species colocalized at either the local (< 30 m^2) or substrate (< 10 cm^2) scales. To sample at the local scale, we collected samples from 20 gridsects (see Methods; **Figure 3-figure supplement 1**) and observed no colocalization of these three species, although only 16% of the total collections were a part of a gridsect. At the substrate scale, we found *C. tropicalis* and *C. briggsae* cohabitating on two of 108 substrates with either species present and *C. oiwi* and *C. briggsae* cohabitating on one of 107 substrates with either species present (**Figure 3-figure supplement 2**). Among 95 substrates with *C. briggsae,* we observed nine instances of *C. briggsae* cohabitating with other PCR-positive species. We did not collect any samples that harbored *C. elegans* and any other *Caenorhabditis* species. We emphasize that these co-occurrence frequencies represent a lower bound estimate of the real co-occurrence frequencies because in many cases we only isolated a small fraction of the nematodes present on a particular sample. Taken together, these results highlight the ubiquitous nature of *C. briggsae* on the Hawaiian Islands and further suggests that the niche of *C. elegans* might be distinct from *C. tropicalis*, *C. briggsae*, and *C. oiwi* on the Hawaiian Islands.

Hawaiian *C. elegans* are divergent from most strains sampled across the globe

We previously showed that two *C. elegans* isolates from Hawaii are highly divergent relative to wild isolates from other regions of the world and represent a large portion of the genetic diversity found within the species (Andersen et al., 2012). Since this analysis, an additional 15 isolates have been collected from the islands and show similarly high levels of genetic diversity (Cook et al., 2016; Hahnel et al., 2018). To better characterize the genetic diversity in Hawaii, we acquired whole-genome sequence data from 95 *C. elegans* isolates that we collected in this study. By analyzing the variant composition of these 95 isolates, we identified 26 distinct genome-wide haplotypes that we refer to as isotypes (see Methods). Within these 26 isotypes, we identified approximately 1.54 million single nucleotide variants (SNVs) that passed our filtering strategy (see Methods; hard-filter VCF; **Supplementary File 3**), which is 27.6% greater than the total number of SNVs identified in all of the 233 non-Hawaiian isotypes included in this study. We found that distinct isotypes are frequently isolated within close proximity to one another in Hawaii. We identified up to seven unique isotypes colocalized within a single gridsect (less than 30 m^2) (**Figure 3-figure supplement 3A**). We also found that colocalization occurred at the substrate level; among the 38 substrates from which we isolated *C. elegans*, 12 contained two or more isotypes (**Figure 3-figure supplement 3B**).

The variant data from all 43 Hawaiian isotypes (26 new with 17 previously described Hawaiian isotypes) allowed us to perform detailed analyses of Hawaiian genetic diversity. Consistent with what is known about *C. elegans* genetic diversity (Andersen et al., 2012), we observed a high degree of genome-wide relatedness among the majority of non-Hawaiian isotypes (**Figure 4-figure supplement 1**). By contrast, isotypes sampled from Hawaii are divergent from isolates sampled outside of Hawaii with the exception of five non-Hawaiian isotypes. Among these exceptions, ECA36 and QX1211 were collected from urban gardens in New Zealand and San Francisco, CA, respectively, and grouped with some of the most divergent isotypes from Hawaii. The other exceptions were the isotypes JU2879, MY23, and MY16, which were isolated across the Pacific region, (*i.e*., Mexico City, Mexico; Concepción, Chile; Nantou County, Taiwan) as well as near the city of Münster, Germany. This result suggests that high genetic diversity is not specific to the Hawaiian Islands but might also be common throughout the Pacific region. Within the sampled Hawaiian isotypes, genome-wide relatedness revealed a high degree of divergence (**Figure 4-figure supplement 1**). This trend is further supported by elevated levels of genome-wide average nucleotide diversity (π) in the Hawaiian sample relative to the non-Hawaiian sample, which we found to be three-fold higher (Hawaii π = 0.00109; non-Hawaiian π = 0.000368; **Figure 4A**). The genomic distribution of diversity followed a similar pattern across chromosomes for both samples, wherein chromosome centers and tips exhibited lower diversity on average than chromosome arms (**Figure 4A**). This pattern is likely explained by lower recombination rates, higher gene densities, and elevated levels of background selection on chromosome centers (Consortium, 1998; Cutter and Payseur, 2003; Rockman et al., 2010). Interestingly, we observed discrete peaks of diversity in specific genomic regions (*e.g.*, chr IV center), which suggests that balancing selection might maintain diversity at these loci in both samples (**Figure 4A**). This hypothesis is supported by corresponding spikes in Tajima's *D* (**Figure 4B**) (Tajima, 1989). Alternatively, higher values of Tajima's *D* might indicate a population contraction, but the discrete nature of these peaks makes this possibility less likely. A third possible explanation is that uncharacterized structural variation (*e.g.*, duplication and divergence) exists in these regions. Nevertheless, the variant sites within these discrete peaks in π and Tajima's *D* are unlikely the result of sequencing errors because they are identified across multiple isotypes (see Methods). Our previous analysis showed that 70–90% of isotypes contain reduced levels of diversity across several megabases (Mb) on chromosomes I, IV, V, and X (Andersen et al., 2012). This reduced diversity was hypothesized to be caused by selective sweeps that occurred within the last few hundred years, potentially through drastic alterations of global environments by humans. The two Hawaiian isotypes, CB4856 and DL238, did not share this pattern of reduced diversity, suggesting that they avoided the selective pressure. Consistent with this previous analysis, we did not observe signatures of selection in the sampled Hawaiian isotypes on chromosomes I, IV, V, and X, as measured by Tajima's *D* (**Figure 4B**), which suggests that the Hawaiian and non-Hawaiian samples have distinct evolutionary histories. This distinction is also

captured in genome-wide Hudson's F_{ST}, where the differentiation between the two samples is highest in regions of the genome impacted by the selective sweeps (**Figure 4C**) (Bhatia et al., 2013; Hudson et al., 1992). Taken together, these data suggest that the isotypes sampled from Hawaii have largely been insulated from the selective pressures thought to be associated with human activity in many regions of the world.

C. elegans population structure

To assess population structure among all 276 isotypes sampled, we performed admixture analysis (see Methods). This analysis identified at least 11 genetic groups for this set of *C. elegans* isotypes, each of which we refer to as a population for the purposes of further analyses (**Figure 5**). Support for at least 11 distinct populations (K) is shown by the minimization of cross-validation (CV) error between Ks 11-15 (**Figure 5-figure supplement 1**). Furthermore, the population assignments for K=11 closely aligned to the relatedness clusters we observed in a neighbor-joining network of all Hawaiian isotypes and the species-wide tree (**Figure 5, Figure 4-figure supplement 1**). For Ks 11-15, the majority of Hawaiian isotypes consistently exhibit no admixture with non-Hawaiian populations. However, a minority of Hawaiian isotypes are consistently either admixed with non-Hawaiian populations (e.g. K=11, 14, and 15) or assigned to populations that contain non-Hawaiian isotypes (e.g. K=12 and 13) (**Figure 5-figure supplement 1**). These data support that a subset of Hawaiian isotypes consistently exhibit a greater degree of genetic relatedness with non-Hawaiian isotypes across different numbers of population subdivisions (Ks). Together, we found at least four distinct populations sampled from the Hawaiian Islands and at least seven additional non-Hawaiian populations sampled from around the globe (**Figure 5-figure supplement 2**).

The majority of isotypes assigned to the seven non-Hawaiian populations exhibit a high degree of admixture with one another, indicating that these populations are not well differentiated. By contrast, isotypes assigned to three of the four Hawaiian populations showed little admixture with other populations (**Figure 5-figure supplement 2A**). We refer to the four Hawaiian populations as Volcano, Hawaiian Divergent, Hawaiian Invaded, and Hawaiian Low for the following reasons. All eight isotypes in the Volcano population were isolated on the Big Island of Hawaii from the town of Volcano at high elevation in wet rainforests primarily composed of ferns, ʻŌhiʻa lehua, and koa trees. The Hawaiian Divergent population is named for the two highly divergent isotypes, XZ1516 and ECA701, which were isolated from Kauai, the oldest Hawaiian island sampled. However, we emphasize that the population assignment of these two highly divergent isotypes might not be correct given that they each contain many strain-specific variants that were filtered from the admixture analysis. The Hawaiian Invaded population is named because many of the isotypes assigned to this population exhibited admixture with non-Hawaiian populations, which is suggestive of an invasion of non-Hawaiian alleles into Hawaii (**Figure 6A, Figure 6-figure supplement 1, Figure 6-figure supplement 2**). The Hawaiian Low population is named because isotypes assigned to this population tended to be isolated at lower elevations than those assigned to the other Hawaiian populations (See Methods, **Figure 6B**). Together, our data suggest that the population structure of the Hawaiian isotypes is associated with either Island, elevation, or both.

Within the Hawaiian Invaded population, one of the 19 isotypes was isolated from outside of Hawaii (MY23), and 11 of 18 Hawaiian isotypes showed admixture with various non-Hawaiian populations, particularly the non-Hawaiian population C (**Figure 6, Figure 5-figure supplement 2**). By contrast, just one isotype (QG556, from the C population) was admixed with the Hawaiian Invaded population and it was isolated from California (**Figure 6-figure supplement 1**). This result suggested that these populations either share ancestry or recent gene flow occurred between them. To distinguish between these possibilities, we explicitly tested for the presence of gene flow among all populations using TreeMix (Pickrell and Pritchard, 2012), which estimates the historical relationships among populations accounting for both population splits and migration events. We found evidence of gene flow between the Hawaiian Invaded population and the non-Hawaiian population C (**Figure 7; Figure 7-figure supplement 1**). The topological position of the fourth highest-weight migration event identified by TreeMix (*i.e.*, C→Hawaiian Invaded) suggested that the

evidence of gene flow is not caused by incomplete assortment of ancestral alleles (*i.e.*, the migration arrows connect the 'C' and Hawaiian Invaded linages at the branch tips) (**Figure 7C; Figure 7-figure supplement 1**). TreeMix also detected evidence of gene flow from a population related to the Hawaiian Low and Hawaiian Invaded populations to the non-Hawaiian G population (**Figure 7C; Figure 7-figure supplement 1**). The G population comprises 15 isotypes typically isolated from Portugal and its two archipelagic autonomous regions, the Azores and Madeira. Interestingly, all of the G isotypes collected from the Azores and Madeira (JU258, NIC199, NIC251) are greater than 13% admixed with the Hawaii Low population (**Figure 5-figure supplement 2A)**. Although we do not know exactly where the source population for this migration was located, the relatively high genetic diversity sampled within the Pacific region suggests that it was likely located there and possibly within Hawaii. Importantly, TreeMix cannot definitively distinguish the direction of migration between populations.

To further assess evidence of gene flow between the Hawaiian and non-Hawaiian populations, we analyzed the haplotype structure across the genomes of all 276 *C. elegans* isotypes (Browning and Browning, 2016). Within the Hawaiian Divergent, Volcano, and Hawaiian Low populations, we observed haplotypes that were largely absent from the non-Hawaiian isotypes. By contrast, the Hawaiian Invaded population shared portions of haplotypes that were commonly found in non-Hawaiian populations. For example, the isotypes in the Hawaiian Invaded population exhibiting admixture with the non-Hawaiian C population share haplotype arrangements on the left and center of Chr III (red and orange Chr III, **Figure 7A**). We also found evidence of the globally swept haplotype in all of the isotypes from the Hawaiian Invaded population, particularly on chromosomes I, V, and X, but less so on chromosome IV (**Figure 7A-B, Figure 7-figure supplement 2**). By contrast, greater than 50% of chromosome IV contained the swept haplotype in all of the isotypes from the C population (**Figure 7-figure supplement 2)**. The Hawaiian Low population generally contained smaller portions of the globally swept haplotypes than the Hawaii Invaded population with the exception of the QX1791 isotype, which is likely because of its admixture with non-Hawaiian populations 'C' and 'K' (**Figure 7A-B, Figure 5-figure supplement 2A)**. Taken together, our data showed that the Hawaiian isotypes from the Volcano and Hawaiian Divergent populations have avoided the selective sweeps that are found in isotypes across most regions of the globe, and individuals within the Hawaiian Invaded subpopulation have likely been outcrossed with these swept haplotypes.

Discussion

We sought to deeply sample the natural genetic variation within the *C. elegans* species to better understand the evolutionary history and driving forces of genome evolution in this powerful model system. Because the Hawaiian Islands have been shown to harbor highly divergent strains relative to most regions of the world, we choose to sample extensively on these islands. We developed a streamlined collection procedure that facilitated our collection of over 2,000 samples across five Hawaiian islands. From these collections, we isolated over 2,500 nematodes and used molecular data to partition 427 of these isolates into 13 distinct taxa, mostly from the Rhabditidae family. In total, we identified and cryogenically preserved 95 new *C. elegans* isolates that represent 26 genetically distinct isotypes. These isotypes represent the largest single *C. elegans* collection effort on any island system and contain 27% more SNVs than all 233 non-Hawaiian isotypes combined. Our findings confirm previous results where high diversity is found on the Hawaiian Islands (Andersen et al., 2012), and document the first evidence of gene flow between Hawaiian and non-Hawaiian populations. Additionally, the wealth of genetic variation within these new Hawaiian isotypes can be used as a source of novel variants that might impact phenotypes of interest. It is important to note that future genome-wide association mappings using these highly diverged isotypes, which contain population-specific SNVs, might detect spurious associations that cannot be mitigated easily by corrections of population stratification (Zhou and Stephens, 2012). Like other diverse population samples, further sampling and a larger number of individuals will enable these types of mappings in the future.

The origins of *C. elegans*

The origin of many species can be traced to the location where genetic diversity is highest (Li and Stephan 2006; Nielsen et al., 2017; Peter et al. 2018). The high genetic diversity we sampled from Hawaii suggests that the Hawaiian Islands could be the origin of *C. elegans*. The extant Hawaii Islands range in age from the still-forming Big Island to the 5.1 million year old Kauai, but the now submerged Emperor Seamounts represent approximately 70 million years of stable land masses over the Pacific Hotspot (Neall and Trewick, 2008). These older landmasses predate the split between *C. elegans* and its closest know relative *C. inopinata,* which is estimated to be 10.5 million years (Kanzaki et al., 2018). However, divergent isolates are also found throughout the Pacific region and sampling efforts historically are biased towards Europe. Notably, *C. inopinata*, was isolated in South Japan, suggesting that these species could have originated from a common ancestor located outside of Hawaii but within the Pacific region (Kanzaki et al., 2018). If *C. elegans* did originate from the Pacific region or the Hawaiian Islands, it is possible that the high diversity sampled from these locations could have been maintained after the transition from a genetically divergent outcrossing species to a selfing species (Thompson et al. 2015). Outcrossing *Caenorhabditis* species generally harbor much higher levels of genetic diversity as compared to selfing *Caenorhabditis* species (Graustein et al. 2002; Jovelin et al. 2003; Li et al. 2014; Cutter et al. 2019). Nevertheless, the deep diversity among the Hawaiian isotypes, all of which are selfing hermaphroditic strains, implies that the origin of selfing might have occurred earlier in the *C. elegans* evolutionary history than previously hypothesized (Cutter et al. 2008).

The higher genetic diversity sampled from the Hawaii Islands could have at least two possible explanations other than residual ancestral diversity present since the transition to selfing. First, the higher genetic diversity in the Hawaiian Islands might be driven by population demography on the Islands. It is possible that Hawaii harbors larger, more temporally stable effective population sizes than other regions of the world that have been sampled. Under a neutral model, populations with a larger effective population size are expected to have a greater number of neutral polymorphisms (Kimura, 1991). These larger, more stable effective population sizes are plausible in Hawaii given the abundant supply of available habitats (*e.g.* rotting fruits and vegetation) and stable temperatures throughout the year. The Hawaiian climate is particularly less variable than many temperate regions where *C. elegans* populations are known to exhibit seasonal population expansions and contractions (Frézal and Félix, 2015; Richaud et al., 2018). Furthermore, population bottlenecks associated with long-range dispersal to new habitats might also contribute to reduced diversity in many parts of the globe. Second, high genetic diversity might not have been restricted to Hawaii but present throughout the globe until recently. If true, then recent selective sweeps associated with human activity might have purged diversity from most regions except the Hawaiian Islands. This explanation is less plausible because genetically diverse isotypes are rarely found outside of the Pacific region, despite massive sampling efforts (especially in Europe). Ultimately, the pattern of genetic variation in Hawaiian populations is likely influenced by a combination of demographic history (*e.g.*, reproductive mode, changes in population size, short- and long-range migration events, and admixture) as well as evolutionary processes such as natural selection, recombination, and mutation. To further untangle the evolutionary history and origin of this species, additional samples from natural areas around the globe and in particular the Pacific region will be required.

Out of Hawaii or invasion of Hawaii?

Our sampling efforts within the Hawaiian Islands allowed us to explore how gene flow into and out of the Hawaiian Islands contributes to the patterns of genetic diversity we observe within the current global sample of *C. elegans*. In particular, we asked whether haplotype sharing among Hawaiian and non-Hawaiian isotypes could be explained by emigration of nematodes from Hawaii (out of Hawaii) or immigration of nematodes to Hawaii (invasion of Hawaii).

Our data support the immigration of alleles into the Hawaiian Invaded population from the less-diverse, non-Hawaiian C population. Moreover, most isotypes from the Hawaiian Invaded and non-Hawaiian

populations share portions of swept haplotypes. Although, isotypes within the Hawaiian Invaded population contain smaller portions of the swept haplotypes than those sampled outside of Hawaii. This pattern of haplotype sharing could be explained by an 'invasion of Hawaii' scenario, wherein swept haplotypes have invaded Hawaii. The genetic diversity of the Hawaiian populations might be threatened if the invading alleles confer strong fitness advantages as is expected for swept haplotypes (Andersen et al., 2012). However, if an invasion of Hawaii is currently underway, we have little evidence to support the selection of the swept haplotypes in Hawaii. First, the Hawaiian Invaded population only contains small fractions of the swept haplotypes on chromosomes I, V, X, and even smaller fractions on chromosome IV. Second, it would take a large number of generations to create these small swept haplotypes because of the low outcrossing rates and high incidence of outbreeding depression in *C. elegans* (Dolgin et al., 2007). Therefore, we speculate that the swept haplotypes introgressed into the Hawaiian Invaded population do not confer the same fitness advantages in the high elevation, cool temperature Hawaiian niche as they do in other parts of the world.

We also discovered evidence suggesting that alleles have emigrated from Hawaii to other regions of the globe. TreeMix analyses suggests that gene flow occurred from a population related to the Hawaiian Invaded and Hawaiian Low populations into the non-Hawaiian G population. Although we cannot be certain if this source population was located within Hawaii, its relation to the two Hawaiian populations suggests that it was. This 'Out of Hawaii' scenario might have occurred through long-range dispersal of genotypes similar to those found in the Hawaiian Invaded and Hawaiian Low populations to other regions of the globe. Recent migration out of Hawaii could have been aided by the transition of the Hawaiian economy towards large-scale production and export of sugarcane and tropical fruits, which began in the late nineteenth century (Bartholomew et al., 2012). If correct, then this situation is similar to what is thought to have occurred within *Drosophila melanogaster* where the fruit trade might have facilitated recent migrations from native regions to oceanic islands (David and Capy, 1988; Hales et al., 2015). Taken together, our results suggest that gene flow into and out of the Hawaiian Islands contributes greatly to the patterns of genetic diversity we observe in *C. elegans* today.

The ancestral niche of *C. elegans* might be similar to the Hawaiian niche

We used publicly available weather data from the National Oceanic and Atmospheric Administration and the National Climatic Data Center to measure the variation in seasonal temperatures for locations close to the sites where isotypes were collected (Evans et al., 2017). We found that the Hawaiian sampling locations experienced less seasonal variability in temperature than any of the non-Hawaiian locations (**Figure 5-figure supplement 3**). These findings raise the possibility that the ancestral niche of *C. elegans* might be similar to the thermally stable Hawaiian habitats where genetic diversity is highest. However, factors other than seasonal temperature variation might also characterize the ancestral niche of *C. elegans*. The Hawaiian Divergent population was enriched at higher elevations, which have been less impacted by human activities in Hawaii since the time of Polynesian colonization (Alison Kay, 1994). By contrast, the Hawaiian Invaded and the Hawaiian Low populations are found at lower elevations, each of which show evidence of the swept haplotypes. Although it remains unclear what factors restrict the flow of swept haplotypes to the Hawaiian Divergent and Volcano populations, selective pressures connected to reduced human influence might contribute to the lack of genetic exchange. This possibility would be consistent with the hypothesis that the global sweeps, present in the Hawaiian Invaded and Hawaiian Low populations, originated through positive selection acting on loci that confer fitness advantages in human-associated habitats (Andersen et al., 2012). Alternatively, the presence of swept haplotypes in the Hawaiian Invaded and Hawaiian Low populations may simply reflect that they are found at lower elevations closer to the coasts where they might be more likely to mate with immigrating nematodes. Nevertheless, we suspect that the ancestral niche of *C. elegans* is most likely to be similar to the thermally stable, high elevation Hawaiian habitats where human impacts are less prevalent.

Unravelling the evolutionary history of *C. elegans*

More accurate models of *C. elegans* niche preferences will facilitate our ability to unravel the evolutionary history of this species by directing researchers to areas most likely to harbor *C. elegans* populations. In order to build more accurate niche models, future sampling efforts should include unbiased sampling across environmental gradients in multiple locations over time because data on niche parameters where *C. elegans* is not found is as important as data where *C. elegans* is found (Félix and Duveau, 2012; Ferrari et al., 2017; Petersen et al., 2014; Richaud et al., 2018). Additionally, we must identify and quantify important biotic niche factors, including associated bacteria, fungi, and invertebrates. These types of data will help facilitate the identification of genes and molecular processes that are under selection in different subpopulations across the species range. *C. elegans* offers a tractable and powerful animal model system to connect environmental parameters to functional genomic variation. These data will deepen our understanding of the evolutionary history of *C. elegans* by revealing how selection and demographic forces have shaped the genome of this important model system.

Methods

Key Resources Table:

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
sequence-based reagent	oECA305; ITS2 F primer	This paper	Andersen Lab: oECA305 F primer	GCTGCGTTATTTACCACGAATTGCARAC
sequence-based reagent	oECA202; ITS2 R primer	10.1186/1471-2148-11-339	Andersen Lab: oECA202 R primer	GCGGTATTTGCTACTACCAYYAMGATCTG C
sequence-based reagent	RHAB1350F; Rhabditid F primer	10.1093/molbev/msh264	Andersen Lab: oECA1271	TACAATGGAAGGCAGCAGGC
sequence-based reagent	RHAB1868R; Rhabditid R primer	10.1093/molbev/msh264	Andersen Lab: oECA1272	CCTCTGACTTTCGTTCTTGATTAA
commercial assay or kit	Blood and Tissue DNA isolation kit	QIAGEN	cat# 69506	
commercial assay or kit	Qubit dsDNA Broad Range Assay Kit	Invitrogen	cat# Q32850	
software, algorithm	Fulcrum	Spatial Networks	Spatial Networks: Fulcrum	https://www.fulcrumapp.com/
software, algorithm	UGENE	10.1093/bioinformatics/bts091	Unipro: UGENE v.1.27.0	http://ugene.net/
software, algorithm	Nextflow	10.1038/nbt.3820	Nexflow	https://github.com/nextflow-io/nextflow

Strains

Nematodes were reared at 20°C using *Escherichia coli* OP50 bacteria grown on modified nematode growth medium (NGMA), containing 1% agar and 0.7% agarose to prevent animals from burrowing (Andersen et al., 2014). In total, 169 *C. briggsae*, 100 *C. elegans*, 21 *C. tropicalis*, 15 *C. oiwi*, and four *C. kamaaina* wild isolates were collected. Of these strains, 95 *C. elegans,* 19 *C. tropicalis*, and 12 *C. oiwi* wild isolates were cryopreserved and are available upon request along with the other *C. elegans* strains

included in our analysis (**Supplementary Data 2)**. The type specimen for *C. oiwi* (ECA1100) is also deposited at the *Caenorhabditis* Genetics Center (**Supplementary File 2**).

Sampling strategy

We sampled nematodes at 2,263 sites across five Hawaiian Islands during August 2017. Before travelling to Hawaii, general sampling locations were selected based on accessibility via hiking trails and by proximity to where *C. elegans* had been collected previously (Andersen et al., 2012; Cook et al., 2016; Hahnel et al., 2018; Hodgkin and Doniach, 1997). Sampling hikes with large elevation changes were prioritized to ensure that we sampled across a broad range of environmental parameters. On these hikes, we opportunistically sampled substrates known to harbor *C. elegans*, including fruits, flowers, stems, leaf litter, fungus, compost, wood, live arthropods, and molluscs (Ferrari et al., 2017; Frézal and Félix, 2015; Schulenburg and Félix, 2017). We chose to group these substrates into six major categories: leaf litter, fruit, flower, fungus, invertebrate, and vegetation. The vegetation category contains all substrates that do not fall into any of the other five categories. We chose to distinguish the leaf litter, fruit, and flower substrates from the vegetation category because we suspect these substrates provide distinct microhabitats and each substrate was sampled extensively. In 20 locations, we performed extensive local sampling in an approximately 30 square meter area that we refer to as a 'gridsect'. The gridsect comprised a center sampling point with additional sampling sites at one, two, and three meters away from the center in six directions with each direction 60º apart from each other (**Figure 3-figure supplement 1**).

Field sampling and environmental data collection

To characterize the *Caenorhabditis* abiotic niche, we collected and organized data for several environmental parameters at each sampling site using a customizable geographic data-collection application called Fulcrum®. We named our customized Fulcrum® application 'Nematode field sampling' and used the following workflow to enter the environmental data into the application while in the field. First, we used a mobile device camera to scan a unique collection barcode from a pre-labelled plastic collection bag. This barcode is referred to as a collection label or ‘C-label’ in the application and is used to associate a particular sample with its environmental and nematode isolation data. Next, we entered the substrate type, landscape, and sky view data into the application using drop down menus and photographed the sample in place using a mobile device camera. The GPS coordinates for the sample are automatically recorded in the photo metadata. We then measured the surface temperature of the sample using an infrared thermometer Lasergrip 1080 (Etekcity, Anaheim, CA), its moisture content using a handheld pin-type wood moisture meter MD912 (Dr. Meter, Los Angeles, CA), and the ambient temperature and humidity near the sample using a combined thermometer and hygrometer device GM1362 (GoerTek, Weifang, China). These measurements were entered into the appropriate fields in the application (**Supplementary File 4**). Finally, we transferred the sample into a collection bag and stored it in a cool location before we attempted to isolate nematodes. Seventy samples in our raw data had missing GPS coordinates or GPS coordinates that were distant from actual sampling locations after visual inspection using satellite imagery. The positions for these samples were corrected using the average position of the two samples collected before and after the errant data point or by manually assigning estimated positions.

Nematode isolation

Following each collection, the substrate sample was transferred from the barcoded collection bag to an identically barcoded 10 cm NGMA plate seeded with OP50 bacteria. For 1,989 of the 2,263 samples collected, we isolated nematodes that crawled off the substrates onto the collection plates approximately 47 hours after the samples were collected from the field (mean = 46.9 h, std. dev. = 19.5 h). The remaining 274 samples were shipped overnight from Hawaii to Northwestern University in collection bags, and the nematodes were isolated approximately 172 hours after sample collection (mean = 172.5 h, std. dev. = 17.9

h). For each collection plate, up to seven gravid nematodes were isolated by transferring them individually to pre-labeled 3.5 cm NGMA isolation plates seeded with OP50 bacteria. We refer to these isolation plates as 'S-plates' in the Fulcrum® application we called 'Nematode isolation' (**Supplementary File 5)**. At the time of isolation, we recorded the approximate number of nematodes on the collection plate. We merged the collection, isolation, and environmental data together into a single data file with the 'process_fulcrum_data.R' script that can be found in the scripts folder of the GitHub repository (https://github.com/AndersenLab/HawaiiMS) (**Supplementary Data 1**).

Nematode identification

The isolated nematodes were stored at 20°C for approximately 14 days (mean = 14.3 d, std. Dev. = 4.9 d) but were not passaged during this time to avoid multiple generations of proliferation. For initial genotyping, whenever possible five to ten nematodes were lysed in 8 µl of lysis solution (100 mM KCl, 20 mM Tris pH 8.2, 5 mM $MgCl_2$, 0.9% IGEPAL, 0.9% Tween 20, 0.02% gelatin with proteinase K added to a final concentration of 0.4 mg/ml) then frozen at -80°C for up to 12 hours. If isolated nematodes were not found on the isolation plates prior to genotyping, they were categorized as 'Not genotyped'. If the isolation plate only contained dead or sterile nematodes, we attempted to lyse five to ten carcasses or sterile individuals. The lysed material was thawed on ice, and 1 µl was loaded directly into 40 µl reactions with primers spanning a portion of the ITS2 region (Internal Transcribed Spacer) between the 5.8S and 28S rDNA genes with forward primer oECA305 (GCTGCGTTATTTACCACGAATTGCARAC) and reverse primer oECA202 (GCGGTATTTGCTACTACCAYYAMGATCTGC) (Kiontke et al., 2011). The ITS2 PCR used the following conditions: 180 seconds denaturation step at 95°C; then 34 cycles of 95°C for 15 seconds, 60°C for 15 seconds, and 72°C for two minutes; followed by a five-minute elongation step at 72°C. We also loaded 1 µl of the lysed material into 40 µl reactions with second set of primers RHAB1350F (TACAATGGAAGGCAGCAGGC) and RHAB1868R (CCTCTGACTTTCGTTCTTGATTAA), which amplify about 500 bp of 18S rDNA of Rhabditid nematodes (Haber et al. 2005). The Rhabditid PCR used the following conditions: 120 second denaturation step at 95°C; then 35 cycles of 95°C for 20 seconds, 55°C for 60 seconds, and 72°C for 30 seconds; followed by a five-minute elongation step at 72°C. The presence of PCR products was visualized on a 2% agarose gel in 1X TAE buffer. Because we could not be certain of the genomic DNA template quality, isolates that did not produce the expected 500 bp Rhabditid PCR product were labelled as 'Not genotyped'. Those isolates that did not yield an ITS2 PCR product were labelled as 'PCR-negative', and those isolates that did yield the expected 2 kb ITS2 PCR product were labelled as 'PCR-positive'. To identify the genus of the isolates labeled as 'PCR-positive', we then used Sanger sequencing of the ITS2 PCR products with forward primer oECA305. We classified *Caenorhabditis* species by comparing the ITS2 sequences to the National Center for Biotechnology Information (NCBI) database using the BLAST algorithm. Isolates with sequences that aligned best to genera other than *Caenorhabditis* were only classified to the genus level.

For every isolate where the BLAST results either aligned to *C. elegans*, had an unexpectedly high number of mismatches in the center of the read, or did not match any known sequences because of poor sequence quality, we performed another independent lysis and PCR using high-quality Taq polymerase (cat# RR001C, TaKaRa) to confirm our original results. For this confirmation, we used the forward primer oECA305 and the reverse primer oECA306 (CACTTTCAAGCAACCCGAC) to sequence the confirmation ITS2 amplicon in both directions. The sequence chromatograms were then quality trimmed by eye with Unipro UGENE software (version 1.27.0) and compared to known nematode species in the NCBI sequence database using the BLAST algorithm. We used the consensus alignment of the forward and reverse reads to confirm our original results. For *C. elegans*, five of the 100 strains perished before we could confirm their identity. We also confirmed that several strains that best aligned to *C. kamaaina* shared a large number of mismatches in the center of the ITS2 amplicon, suggesting they belonged to a new species. For these strains, we performed reciprocal mating tests with *C. kamaaina* to infer the new species by the biological species

concept (Félix et al., 2014). None of these crosses produced viable progeny, suggesting that these isolates represent a new *Caenorhabditis* species (**Supplementary File 2**).

Illumina library construction and whole-genome sequencing

To extract DNA, we transferred nematodes from two 10 cm NGMA plates spotted with OP50 *E. coli* into a 15 ml conical tube by washing with 10 mL of M9. We then used gravity to settle animals on the bottom of the conical tube, removed the supernatant, and added 10 mL of fresh M9. We repeated this wash method three times over the course of one hour to serially dilute the *E. coli* in the M9 and allow the animals time to purge ingested *E. coli*. Genomic DNA was isolated from 100-300 µl nematode pellets using the Blood and Tissue DNA isolation kit cat# 69506 (QIAGEN, Valencia, CA) following established protocols (Cook et al., 2016). The DNA concentration was determined for each sample with the Qubit dsDNA Broad Range Assay Kit cat# Q32850 (Invitrogen, Carlsbad, CA). The DNA samples were then submitted to the Duke Center for Genomic and Computational Biology per their requirements. The Illumina library construction and sequencing were performed at Duke University using KAPA Hyper Prep kits (Kapa Biosystems, Wilmington, MA) and the Illumina NovaSeq 6000 platform (paired-end 150 bp reads). The raw sequencing reads for strains used in this project are available from the NCBI Sequence Read Archive (Project PRJNA549503).

Variant calling

To ensure reproducible data analysis, all genomic analyses were performed using pipelines generated in the Nextflow workflow management system framework (Di Tommaso et al., 2017). Each Nextflow pipeline used in this study is briefly described below (**Supplementary File 6**). All pipelines follow the "*pipeline name-nf*" naming convention and full descriptions can be found on the Andersen lab dry-guide website: (http://andersenlab.org/dry-guide/pipeline-overview/).

Raw sequencing reads were trimmed using *trimmomatic-nf*, which uses trimmomatic (v0.36) (Bolger et al., 2014) to remove low-quality bases and adapter sequences. Following trimming, we used the *concordance-nf* pipeline to characterize *C. elegans* strains isolated in this study and previously described strains (Cook et al., 2017, 2016; Hahnel et al., 2018). The *concordance-nf* pipeline calls SNVs using the BCFtools (v.1.9) (Danecek et al., 2014) variant calling software. The variants are filtered by: Depth (FORMAT/DP) ≥ 3; Mapping Quality (INFO/MQ) > 40; Variant quality (QUAL) > 30; (Allelic Depth (FORMAT/AD) / Num of high quality bases (FORMAT/DP)) ratio > 0.5. We determined the pairwise similarity of all strains by calculating the fraction of shared SNVs. Finally, we classified two or more strains as the same isotype if they shared >99.9% SNVs. If a strain did not meet this criterion, we considered it as a unique isotype. Newly assigned isotypes were added to CeNDR (Cook et al., 2017).

After isotypes are assigned, we used *alignment-nf* with BWA (v0.7.17-r1188) (Li, 2013; Li and Durbin, 2009) to align trimmed sequence data for distinct isotypes to the N2 reference genome (WS245) (Lee et al., 2018). Next, we called SNVs using *wi-nf*, which uses the BCFtools (v.1.9) (Danecek et al., 2014). The *wi-nf* pipeline generates two population-wide VCFs that we refer to as the soft-filtered and hard-filtered VCFs (**Supplementary File 3**). After variant calling, a soft-filtered VCF was generated for each sample by appending the following soft-filters to variant sites: Depth (FORMAT/DP) > 10; Mapping Quality (INFO/MQ) > 40; Variant quality (QUAL) > 10; (Allelic Depth (FORMAT/AD) / Number of high quality bases (FORMAT/DP)) ratio > 0.5. These soft-filters were appended to the FT field of the VCF using *VCF-kit* (Cook and Andersen, 2017). Next, sample VCFs were merged using the merge utility of BCFtools (v.1.9). Once the population VCF was generated, variant sites with greater than 90% missing genotypes (high_missing) or greater than 10% heterozygosity (high_heterozygosity) were flagged. We refer to this VCF as the 'soft-filtered VCF'. To construct the 'hard-filtered VCF', we removed all variants that did not pass the filters described above. Both the soft- and hard-filtered isotype-level VCFs are available to download on the CeNDR website (version 20180527) (Cook et al., 2017).

We further pruned the hard-filtered VCF to contain sites with no missing genotype calls using BCFtools (v.1.9) (Danecek et al., 2014). The predicted variant effects were appended to the VCF using SnpEff (v 4.3) (Cingolani et al., 2012). We further annotated this VCF with exons, G-quartets, transcription factor binding sites, histone binding sites, miRNA binding sites, splice sites, ancestral alleles (XZ1516 set as ancestor), the genetic map position, and repetitive elements using vcfanno (v 0.2.8) (Pedersen et al., 2016). All annotations were obtained from WS266. We removed regions that were annotated as repetitive. We named this VCF the 'PopGen VCF' (**Supplementary Data 3; Supplementary File 3**).

Phylogenetic analyses

We characterized the relatedness of the *C. elegans* isotypes using RAxML-ng with the GTR DNA substitution model and maximum likelihood estimation to find the parameter values that maximize the phylogenetic likelihood function, and thus provide the best explanation for the observed data (Kozlov et al., 2019). We pruned the 'PopGen VCF' by removing sites in high linkage disequilibrium (LD) using PLINK (v1.9) (Chang et al., 2015; Purcell et al., 2007) with the *--indep-pairwise 50 1 0.95* command. We used the vcf2phylip.py script (Ortiz, n.d.) to convert this pruned VCF to the PHYLIP format (Felsenstein, 1993) required to run RAxML-ng. To construct the unrooted tree that included 276 strains **(Figure 4-figure supplement 1)**, we used the GTR evolutionary model available in RAxML-ng (Lanave et al., 1984; Tavaré, 1986). This tree was visualized using the ggtree (v1.10.5) R package (Yu et al., 2017). To construct the neighbor-net phylogenies **(Figure 5, Figure 5-figure supplement 4)** we further pruned the 'PopGen VCF' by removing sites in linkage disequilibrium using PLINK (v1.9) with the *--indep-pairwise 50 1 0.8* command. We also removed variants only present in one isotype. To visualize the neighbor-net phylogenies we used SplitsTree4 (Huson and Bryant, 2006).

Population genetic statistics

Genome-wide π, Hudson's F_{ST}, and Tajima's D were calculated using the PopGenome package in R (Pfeifer et al., 2014). All statistics were calculated along sliding windows with a 10 kb window size and a 1 kb step size with the 'PopGen VCF' **(Supplementary Data 3; Supplementary File 3**). Importantly, our statistics are calculated among isotypes not among isolates. The standard method for calculating these statistics is among isolates. We choose to calculate these values among isotypes to avoid oversampling of highly related individuals that are often isolated from the same substrate.

Admixture analysis

We performed admixture analysis using ADMIXTURE (v1.3.0) (Alexander et al., 2009). Prior to running ADMIXTURE, we LD-pruned the 'PopGen VCF' (**Supplementary Data 3**) using PLINK (v1.9) (Chang et al., 2015; Purcell et al., 2007) with the command *--indep-pairwise 50 10 0.8*. We also removed variants only present in one isotype. We ran ADMIXTURE ten independent times for K sizes ranging from 2 to 20 for all 276 isotypes. Visualization of admixture results was performed using the pophelper (v2.2.5) R package (Francis, 2017). We chose K=11 for future analyses because the cross-validation (CV) error approached minimization at this K (**Figure 5-figure supplement 1**). Furthermore, K=11 subset the Hawaiian isotypes into four distinct populations, which exactly matched the subsets obtained from running ADMIXTURE on just the 43 Hawaiian isotypes at K=4 (K=4 minimized CV for ADMIXTURE with Hawaiian isotypes only, (**Figure 5-figure supplement 4**). We performed TreeMix analysis on K=11 for zero to five migration events (Pickrell and Pritchard, 2012).

Haplotype analysis

We determined identity-by-descent (IBD) of strains using IBDSeq (Browning and Browning, 2013) run on the 'PopGen VCF' (**Supplementary Data 3**) with the following parameters: *minalleles=0.01, ibdtrim=0, r2max=0.8*. IBD segments were then used to infer haplotype structure among isotypes as described

previously (Andersen et al., 2012). After haplotypes were identified, we defined the most common haplotype found on chromosomes I, IV, V, and X as the swept haplotype. We then retained the swept haplotypes within isotypes that passed the following per chromosome filters: total length > 1 Mb; total length / maximum population-wide swept haplotype length > 0.03. We classified chromosomes within isotypes as swept if the sum of the retained swept haplotypes for a chromosome was > 3% of the maximum population wide swept haplotype length for that chromosome.

Environmental parameter analysis

We calculated the pairwise distances among all *C. elegans*-positive collections on Hawaii and detected five distinct geographic clusters, each of which contain collections that are within 20 meters of one another. The largest of these clusters comprised 18 collections in the Kalopa State Recreation Area on the Big Island of Hawaii. This cluster contained 11 collections from gridsect-3 and seven additional collections within 20 meters from the edge of the gridsect. The other four geographic clusters contain four or fewer collections each. We used the average values of environmental parameters from geographically clustered collections to avoid biasing our results by local oversampling. We applied this strategy to the comparison of environmental parameters between the Hawaiian admixture populations and used the Kruskal-Wallis test to detect differences ($\alpha = 0.05$).

Climate data analysis

Climate data were downloaded from the Integrated Surface Data (ISD) FTP server (ftp://ftp.ncdc.noaa.gov/pub/data/noaa/) managed by the National Oceanic and Atmospheric Administration (NOAA), and the National Climatic Data Center (NCDC). The data were processed using the noaa-nf pipeline available at the GitHub repository (https://github.com/AndersenLab/noaa-nf).

Acknowledgements

We thank the members of the Andersen lab for editing the manuscript for flow and content and for making reagents used in the experiments presented. We are grateful to landowners who gave us permission to collect nematodes on their property. We also thank individuals who have helped us collect additional strains. We would also like to thank the Hawaii Department of Land and Natural Resources as well as the Natural Area Reserves System for permitting, support for these studies, and general advice about the Hawaiian Islands. Additionally, Dr. Sam Gon from The Nature Conservancy Hawai'i Program helped with the naming of *Caenorhabditis oiwi*. This research was supported by start-up funds from Weinberg College of Arts and Sciences and the Molecular Biosciences department. KK is supported by NSF DEB 0922012 to D. H. A. Fitch.

Competing interests

The authors declare no conflicts of interest.

References

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**:1655–1664.

Alison Kay E. 1994. A Natural History of the Hawaiian Islands: Selected Readings II. University of Hawaii Press.

Andersen EC, Bloom JS, Gerke JP, Kruglyak L. 2014. A variant in the neuropeptide receptor npr-1 is a major determinant of Caenorhabditis elegans growth and physiology. *PLoS Genet* **10**:e1004156.

Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, Félix M-A, Kruglyak L. 2012. Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. *Nat Genet* **44**:285–290.

Barrière A, Félix M-A. 2014. Isolation of C. elegans and related nematodes. *WormBook* 1–19.

685 Bartholomew DP, Hawkins RA, Lopez JA. 2012. Hawaii Pineapple: The Rise and Fall of an Industry.
686 *HortScience* **47**:1390–1398.
687 Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: the impact of
688 rare variants. *Genome Res* **23**:1514–1521.
689 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data.
690 *Bioinformatics* **30**:2114–2120.
691 Brenner S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* **77**:71–94.
692 Browning BL, Browning SR. 2016. Genotype Imputation with Millions of Reference Samples. *Am J Hum*
693 *Genet* **98**:116–126.
694 Browning BL, Browning SR. 2013. Detecting Identity by Descent and Estimating Genotype Error Rates in
695 Sequence Data. *Am J Hum Genet* **93**:840–851.
696 Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. 1994. Green fluorescent protein as a marker for
697 gene expression. *Science* **263**:802–805.
698 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to
699 the challenge of larger and richer datasets. *Gigascience* **4**:7.
700 Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program
701 for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the
702 genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**:80–92.
703 Consortium TCES. 1998. Genome Sequence of the Nematode C. elegans: A Platform for Investigating
704 Biology. *Science* 1–8.
705 Cook DE, Andersen EC. 2017. VCF-kit: assorted utilities for the variant call format. *Bioinformatics*
706 **33**:1581–1582.
707 Cook DE, Zdraljevic S, Roberts JP, Andersen EC. 2017. CeNDR, the Caenorhabditis elegans natural
708 diversity resource. *Nucleic Acids Res* **45**:D650–D657.
709 Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, Rockman MV, Alkema MJ, Braendle C,
710 Kammenga JE, Wang J, Kruglyak L, Félix M-A, Lee J, Andersen EC. 2016. The Genetic Basis of
711 Natural Variation in Caenorhabditis elegans Telomere Length. *Genetics* genetics.116.191148.
712 Cutter AD, Payseur BA. 2003. Selection at linked sites in the partial selfer Caenorhabditis elegans. *Mol Biol*
713 *Evol* **20**:665–673.
714 Cutter AD, 2006. Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial
715 selfer Caenorhabditis elegans. *Genetics* **172**: 171–184.
716 Danecek P, Schiffels S, Durbin R. 2014. Multiallelic calling model in bcftools (-m) 10–11.
717 David JR, Capy P. 1988. Genetic variation of Drosophila melanogaster natural populations. *Trends Genet*
718 **4**:106–111.
719 Dey A, Chan CKW, Thomas CG, Cutter AD. 2013. Molecular hyperdiversity defines populations of the
720 nematode Caenorhabditis brenneri. *Proc Natl Acad Sci U S A* **110**:11056–11060.
721 Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. 2017. Nextflow enables
722 reproducible computational workflows. *Nat Biotechnol* **35**:316–319.
723 Dolgin ES, Charlesworth B, Baird SE, Cutter AD. 2007. Inbreeding and outbreeding depression in
724 Caenorhabditis nematodes. *Evolution* **61**:1339–1352.
725 Dolgin ES, Félix M-A, Cutter AD. 2008. Hakuna Nematoda: genetic and phenotypic diversity in African
726 isolates of Caenorhabditis elegans and C. briggsae. *Heredity* **100**:304–315.
727 Evans KS, Zhao Y, Brady SC, Long L, McGrath PT, Andersen EC. 2017. Correlations of Genotype with
728 Climate Parameters Suggest Caenorhabditis elegans Niche Adaptations. *G3* **7**:289–298.
729 Félix M-A, Braendle C, Cutter AD. 2014. A streamlined system for species diagnosis in Caenorhabditis
730 (Nematoda: Rhabditidae) with name designations for 15 distinct biological species. *PLoS One*
731 **9**:e94723.
732 Félix M-A, Duveau F. 2012. Population dynamics and habitat sharing of natural populations of
733 Caenorhabditis elegans and C. briggsae. *BMC Biol* **10**:59.
734 Félix M-A, Jovelin R, Ferrari C, Han S, Cho YR, Andersen EC, Cutter AD, Braendle C. 2013. Species
735 richness, distribution and genetic diversity of Caenorhabditis nematodes in a remote tropical rainforest.
736 *BMC Evol Biol* **13**:10.
737 Felsenstein J. 1993. Phylogeny inference package. *Department of Genetics, University of Washington*.
738 Ferrari C, Salle R, Callemeyn-Torre N, Jovelin R, Cutter AD, Braendle C. 2017. Ephemeral-habitat
739 colonization and neotropical species richness of Caenorhabditis nematodes. *BMC Ecol* **17**:43.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**:806–811.
Francis RM. 2017. pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour* **17**:27–32.
Frankham R. 1997. Do island populations have less genetic variation than mainland populations? *Heredity* **78 (Pt 3)**:311–327.
Frézal L, Félix M-A. 2015. C. elegans outside the Petri dish. *Elife* **4**. doi:10.7554/eLife.05849
Graustein A, Gaspar JM, Walters JR, Palopoli MF. 2002. Levels of DNA polymorphism vary with mating system in the nematode genus caenorhabditis. *Genetics* **161**:99–107.
Grishok A, Tabara H, Mello CC. 2000. Genetic requirements for inheritance of RNAi in C. elegans. *Science* **287**:2494–2497.
Haber M, Schüngel M, Putz A, Müller S, Hasert B, Schulenburg H. 2005. Evolutionary history of Caenorhabditis elegans inferred from microsatellites: evidence for spatial and temporal genetic differentiation and the occurrence of outbreeding. *Mol Biol Evol* **22**:160–173.
Hahnel SR, Zdraljevic S, Rodriguez BC, Zhao Y, McGrath PT, Andersen EC. 2018. Extreme allelic heterogeneity at a Caenorhabditis elegans beta-tubulin locus explains natural resistance to benzimidazoles. *bioRxiv*. doi:10.1101/372623
Hales KG, Korey CA, Larracuente AM, Roberts DM. 2015. Genetics on the Fly: A Primer on the Drosophila Model System. *Genetics* **201**:815–842.
Hodgkin JA, Brenner S. 1977. Mutations causing transformation of sexual phenotype in the nematode Caenorhabditis elegans. *Genetics* **86**:275–287.
Hodgkin J, Doniach T. 1997. Natural variation and copulatory plug formation in Caenorhabditis elegans. *Genetics* **146**:149–164.
Huang R-E, Ren X, Qiu Y, Zhao Z. 2014. Description of Caenorhabditis sinica sp. n. (Nematoda: Rhabditidae), a nematode species used in comparative biology for C. elegans. *PLoS One* **9**:e110957.
Hudson RR, Boos DD, Kaplan NL. 1992. A statistical test for detecting geographic subdivision. *Mol Biol Evol* **9**:138–151.
Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**:254–267.
Jovelin R, Ajie BC, Phillips PC. 2003. Molecular evolution and quantitative variation for chemosensory behaviour in the nematode genus Caenorhabditis. *Mol Ecol* **12**:1325–1337.
Kanzaki N, Tsai IJ, Tanaka R, Hunt VL, Liu D, Tsuyama K, Maeda Y, Namai S, Kumagai R, Tracey A, Holroyd N, Doyle SR, Woodruff GC, Murase K, Kitazume H, Chai C, Akagi A, Panda O, Ke H-M, Schroeder FC, Wang J, Berriman M, Sternberg PW, Sugimoto A, Kikuchi T. 2018. Biology and genome of a newly discovered sibling species of Caenorhabditis elegans. *Nat Commun* **9**:3216.
Kimura M. 1991. The neutral theory of molecular evolution: a review of recent evidence. *Jpn J Genet* **66**:367–386.
Kiontke KC, Félix M-A, Ailion M, Rockman MV, Braendle C, Pénigault J-B, Fitch DHA. 2011. A phylogeny and molecular barcodes for Caenorhabditis, with numerous new species from rotting fruits. *BMC Evol Biol* **11**:339.
Koch R, van Luenen HG, van der Horst M, Thijssen KL, Plasterk RH. 2000. Single nucleotide polymorphisms in wild isolates of Caenorhabditis elegans. *Genome Res* **10**:1690–1696.
Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. doi:10.1093/bioinformatics/btz305
Lanave C, Preparata G, Saccone C, Serio G. 1984. A new method for calculating evolutionary substitution rates. *J Mol Evol* **20**:86–93.
Lee D, Zdraljevic S, Cook DE, Frézal L, Hsu J-C, Sterken MG, Riksen JAG, Wang J, Kammenga JE, Braendle C, Félix M-A, Schroeder FC, Andersen EC. 2019. Selection and gene flow shape niche-associated copy-number variation of pheromone receptor genes. *bioRxiv*. doi:10.1101/580803
Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**:843–854.
Lee RYN, Howe KL, Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Davis P, Gao S, Grove C, Kishore R, Muller H-M, Nakamura C, Nuin P, Paulini M, Raciti D, Rodgers F, Russell M, Schindelman G, Tuli MA, Van Auken K, Wang Q, Williams G, Wright A, Yook K, Berriman M, Kersey P, Schedl T, Stein L,

Sternberg PW. 2018. WormBase 2017: molting into a new stage. *Nucleic Acids Res* **46**:D869–D874.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bioGN]*. doi:arXiv:1303.3997 [q-bio.GN]
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760.
Li H, Stephan W. 2006. Inferring the demographic history and rate of adaptive substitution in Drosophila. *PLoS Genet* **2**:e166.
Li S, Jovelin R, Yoshiga T, Tanaka R, Cutter AD. 2014. Specialist versus generalist life histories and nucleotide diversity in Caenorhabditis nematodes. *Proc Biol Sci* **281**:20132858.
McGrath PT, Rockman MV, Zimmer M, Jang H, Macosko EZ, Kruglyak L, Bargmann CI. 2009. Quantitative mapping of a digenic behavioral trait implicates globin variation in C. elegans sensory behaviors. *Neuron* **61**:692–699.
Neall VE, Trewick SA. 2008. The age and origin of the Pacific islands: a geological overview. *Philos Trans R Soc Lond B Biol Sci* **363**:3293–3308.
Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. 2017. Tracing the peopling of the world through genomics. *Nature* **541**:302–310.
Ortiz EM. n.d. vcf2phylip. Github.
Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, Sigwalt A, Barre B, Freel K, Llored A, Cruaud C, Labadie K, Aury J-M, Istace B, Lebrigand K, Barbry P, Engelen S, Lemainque A, Wincker P, Liti G, Schacherer J. 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature* **556**:339–344.
Pedersen BS, Layer RM, Quinlan AR. 2016. Vcfanno: fast, flexible annotation of genetic variants. *Genome Biol* **17**:118.
Petersen C, Dirksen P, Prahl S, Strathmann EA, Schulenburg H. 2014. The prevalence of *Caenorhabditis elegans* across 1.5 years in selected North German locations: the importance of substrate type, abiotic parameters, and Caenorhabditis competitors. *BMC Ecol* **14**:4.
Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Mol Biol Evol* **31**:1929–1936.
Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* **8**:e1002967.
Poullet N, Vielle A, Gimond C, Ferrari C, Braendle C. 2015. Evolutionarily divergent thermal sensitivity of germline development and fertility in hermaphroditic Caenorhabditis nematodes. *Evol Dev* **17**:380–397.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**:559–575.
Richaud A, Zhang G, Lee D, Lee J, Félix M-A. 2018. The Local Coexistence Pattern of Selfing Genotypes in Caenorhabditis elegans Natural Metapopulations. *Genetics* **208**:807–821.
Rockman MV, Kruglyak L. 2009. Recombinational Landscape and Population Genomics of Caenorhabditis elegans. *PLoS Genet* **5**:e1000419.
Rockman MV, Skrovanek SS, Kruglyak L. 2010. Selection at linked sites shapes heritable phenotypic variation in C. elegans. *Science* **330**:372–376.
Schulenburg H, Félix M-A. 2017. The Natural Biotic Environment of Caenorhabditis elegans. *Genetics* **206**:55–86.
Slos D, Sudhaus W, Stevens L, Bert W, Blaxter M. 2017. Caenorhabditis monodelphis sp. n.: defining the stem morphology and genomics of the genus Caenorhabditis. *BMC Zoology* **2**:4.
Stevens L, Félix M-A, Beltran T, Braendle C, Caurcel C, Fausett S, Fitch D, Frézal L, Gosse C, Kaur T, Kiontke K, Newton MD, Noble LM, Richaud A, Rockman MV, Sudhaus W, Blaxter M. 2019. Comparative genomics of 10 new Caenorhabditis species. *Evolution Letters* **3**:217–236.
Sudhaus W, Kiontke K. 2007. Comparison of the cryptic nematode species Caenorhabditis brenneri sp. n. and C. remanei (Nematoda: Rhabditidae) with the stem species pattern of the Caenorhabditis Elegans group. *Zootaxa* **1456**:45–62.
Sulston JE, Schierenberg E, White JG, Thomson JN. 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev Biol* **100**:64–119.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.

Genetics **123**:585–595.
Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on mathematics in the life sciences* **17**:57–86.
Thompson OA, Snoek LB, Nijveen H, Sterken MG, Volkers RJM, Brenchley R, Van't Hof A, Bevers RPJ, Cossins AR, Yanai I, Hajnal A, Schmid T, Perkins JD, Spencer D, Kruglyak L, Andersen EC, Moerman DG, Hillier LW, Kammenga JE, Waterston RH. 2015. Remarkably Divergent Regions Punctuate the Genome Assembly of the Caenorhabditis elegans Hawaiian Strain CB4856. *Genetics* **200**:975–989.
Thomas CG, Wang W, Jovelin R, Ghosh R, Lomasko T, Trinh Q, Kruglyak L, Stein LD, Cutter AD. 2015. Full-genome evolutionary histories of selfing, splitting, and selection in Caenorhabditis. *Genome Res* **25**:667–678.
Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**:28–36.
Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**:821–824.

Legends:

Figure 1 - **Geographic distribution of sampling sites across five Hawaiian islands**. In total we sampled 2,263 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 bottom 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 (**Supplementary Data 1**). We focused our studies on *Caenorhabditis* nematode collections, excluding *C. kamaaina* because it was only found at two sampling sites. Maps © www.thunderforest.com, Data © www.osm.org/copyright.

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 µm and 20 µm in D-F. A formal description of *C. oiwi* is provided (**Supplementary File 2**).

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 guernaculum 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 µm. A formal description of *C. oiwi* is provided (**Supplementary File 2**).

Figure 1-source data 1: The collection categories and location data for each of the 2,263 samples collected are organized here. The c_label column indicates the unique ID for each sample collected. The collection_type column indicates which collection class into which the sample was placed. The island column indicates which island from which the sample was collected. The PCR_positive column contains a 0 for samples that had a PCR product with the Rhabditid primers but did not have PCR product with the ITS2 primers. The worms_on_sample column is 'Yes' if worms were observed on the collection plate, 'No' if they were not observed, or 'Tracks' if we only saw tracks on the plate. Longitude and Latitude columns indicate the GPS position of the sample. The multiple_type column is 'no' if only one collection_type was observed for the c_label and 'yes' if more than one collection type was observed. (Used to generate Figure 1A-F)

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 2-source data 1: The fraction of samples in each collection category organized by major substrate classes. The fixed_substrate column indicates the substrate category. The worm_per_substrate column indicates how many instances of that collection category were sampled within the substrate class. The total_susbtrates column indicates how many substrates of that class were sampled in total. The

perc_worm_sub column indicates the fraction of samples in that belong to a certain collection category within each substrate class. The plot_type column indicates the various collection categories. (These data were used to generate Figure 2A).

Figure 2-source data 2: The fraction of samples containing each *Caenorhabditis* species by substrate class are organized here. The invertebrate substrate class is not included because we did not sample *Caenorhabditis* from that class. The species_family column indicates the Caenorhabditis species under consideration for that row of data. The fixed_substrate column indicates the substrate class. The perc_worm_sub2 column indicates the fraction of samples containing each *Caenorhabditis* species by substrate class. (These data were used to generate Figure 2B).

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 (ºC) was typically cooler at the sites were *C. elegans* were isolated compared to sites for all other *Caenorhabditis* species (Dunn test, $p < 0.005$). (**B**) Substrate temperature (ºC) 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 60º 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 3-source data 1: Environmental parameter data for all *Caenorhabditis*-positive collections are organized here. The c_label column indicates the unique ID for each sample collected. The s_label column is the unique ID for each nematode isolated. The species_family column corresponds to the species isolated from that isolation plate. The env_par column indicated the environmental parameter being measured in the value column. The KM_pvalue indicates the p-value of the Kruskal-Wallis rank sum test for each environmental parameter. (Used to generate Figure 3A-E)

Figure 3-source data 2: Pearson correlation coefficients for environmental parameter pairs. The X and Y columns correspond to the environmental parameters tested for correlations. The pearson_correlation column indicates the Pearson correlation coefficients for the parameter pairs. (Used to generate Figure 3F)

Figure 3-figure supplement 1-source data 1: Collection data for all gridsect samples. (Used to generate Figure 3-figure supplement 1)

Figure 3-figure supplement 2-source data 1: Collection data for all samples where two or more PCR-positive nematodes were isolated from the same sample. (Used to generate Figure 3-figure supplement 2)

Figure 3-figure supplement 3-source data 1: Collection data for gridsects from which *C. elegans* were isolated. (Used to generate Figure 3-figure supplement 3A)

Figure 3-figure supplement 3-source data 2: Collection data for all samples from which multiple *C. elegans* isotypes were isolated. (Used to generate Figure 3-figure supplement 3B)

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 4-source data 1: Chromosomal patterns of *C. elegans* diversity and differentiation. The values of nucleotide diversity (π) for 10 kb windows in the Hawaiian and non-Hawaiian samples are found in the Diversity_Hawaiian and Diversity_non-Hawaiian columns, respectively. The snp_index, position, and CHROM columns indicate the index ID, nucleotide position, and chromosome start points for each of the 10 kb windows. The Fst column holds Hudson's F_{ST} values indicating differentiation among the two samples for the same windows. (Used to generate Figures 4A and 4C)

Figure 4-source data 2: Chromosomal patterns of Tajima's *D* for the Hawaiian and non-Hawaiian samples. The CHROM and position columns indicate the chromosome and nucleotide starting positions for the 10 kb windows used to calculate Tajima's *D*. The Population column corresponds to the sample being tested, and the value column indicates the value of the Tajima's *D* statistic calculated for each window. (Used to generate Figure 4B)

Figure 4-figure supplement 1-source data 1: Best-scoring maximum likelihood tree output from RAxML-ng. (Used to generate Figure 4-figure supplement 1)

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 2-20 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 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 3 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 1st 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 2-10 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 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 5-source data 1: VCF used to generate the nexus file with vcf2phylip.py script for the neighbor-joining network shown in Figure 5. This VCF is processed from the 'PopGen VCF' (**Supplementary Data 3**) (see Methods - Phylogenetic analyses). (Used to generate Figure 5)

Figure 5-source data 2: The nexus file for creating neighbor-joining network shown in Figure 5. This file was generated from the VCF (Figure 5-source data 1) using the vcf2phylip.py script (see Methods - Phylogenetic analyses). (Used to generate Figure 5)

Figure 5-source data 3: Admixture population proportions for all isotypes within each admixture population (K=11). The pop-assignment column indicates the admixture population. The n column is the number of isotypes assigned to that admixture population. The cluster column corresponds to each of the 11 admixture populations identified labelled as perc_a – perc_k. The cluster_perc column indicates the fraction of each cluster found within all n isotypes assigned to each admixture population. The total_perc column indicates the sum of all cluster_perc values for each admixture population. (Used to generate pie charts in Figure 5)

Figure 5-figure supplement 1-source data 1: Cross-validation error from ten Independent ADMIXTURE runs at Ks 2-20. (Used to generate Figure 5-figure supplement 1A)

Figure 5-figure supplement 1-source data 2: The inferred population proportions estimated by ADMIXTURE for all isotypes at Ks 10-15. (Used to generate Figure 5-figure supplement 1B)

Figure 5-figure supplement 2-source data 1: The sampling locations and admixture population assignment for all strains including isotype reference strains. (Used to generate Figure 5-figure supplement 2B-D)

Figure 5-figure supplement 2-source data 2: The inferred population proportions estimated by ADMIXTURE for all isotypes at K = 11. (Used to generate Figure 5-figure supplement 2A)

Figure 5-figure supplement 3-source data 1: Seasonal temperature variation for all isotypes over a 12-month period. The data are taken from weather stations nearest to the isolation locations. (Used to generate Figure 5-figure supplement 3)

Figure 5-figure supplement 4-source data 1: The inferred population proportions estimated by ADMIXTURE for all Hawaiian isotypes at Ks 3-5. (Used to generate Figure 5-figure supplement 4B)

Figure 5-figure supplement 4-source data 2: Cross-validation error from ten Independent ADMIXTURE runs with only the Hawiian isotypes across Ks 2-10. (Used to generate Figure 5-figure supplement 4A)

Figure 5-figure supplement 4-source data 3: The nexus file for creating the neighbor-joining network. (Used to generate Figure 5-figure supplement 4C)

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 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 6-source data 1: The inferred population fractions for each Hawaiian isotype as estimated by ADMIXTURE (K=11; 276 *C. elegans* isotypes). The isotype column gives the isotype name. The max_pop_frac colum indicates the largest population fraction for any of the 11 populations estimated by ADMIXTURE for that isotype. The pop_assignment column contains the admixture population (A – K) with the largest population fraction for that isotype. The Hawaiian column is 'TRUE' if the isotype was isolated in Hawaii. The cluster column indicates the admixture population under consideration for that isotype. The frac_cluster column is the fraction of each admixture population estimated for each isotype. (Used to generate Figure 6A)

Figure 6-source data 2: Environmental parameters for distinct collections of admixture populations in Hawaii. In cases where multiple collections were made within 20 meters for the same admixture population, we took the average parameter values for plotting and statistical calculations (see Methods – Environmental parameter analysis). (Used to generate Figure 6B-C)

Figure 6-source data 3: *post hoc* Dunn multiple comparison tests for differences in environmental parameters among the Hawaiian populations identified by ADMIXTURE are organized here. (Used to generate Figure 6B-C)

Figure 6-figure supplement 1-source data 1: Isolation location data for distinct isolations of all isotypes and their inferred population proportions estimated by ADMIXTURE. (Used to generate Figure 6-figure supplement 1)

Figure 6-figure supplement 2-source data 1: The inferred population proportions estimated by ADMIXTURE (K=11) for isotypes assigned to the Hawaii Invaded, Volcano, and non-Hawaiian C populations. (Used to generate Figure 6-figure supplement 2)

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 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.

Figure 7-source data 1: Haplotypes or IBD blocks across the genomes of all 276 isotypes are organized here. (Used to generate Figure 7A)

Figure 7-source data 2: Fraction of the genome with the swept haplotypes for each isotype in the four Hawaiian populations and the non-Hawaiian C population. (Used to generate Figure 7B)

Figure 7-source data 3: Ancestry fractions (.Q file) and allele frequencies of the inferred ancestral populations (.P file) from ADMIXTURE analysis. (These data are used to generate Figure 7C).

Figure 7-figure supplement 1-source data 1: Ancestry fractions (.Q file) and allele frequencies of the inferred ancestral populations (.P file) from ADMIXTURE analysis (K=11). (Used to generate Figure 7-figure supplement 1A-F)

Figure 7-figure supplement 2-source data 1: Fraction of each chromosome with the most common global haplotype (swept haplotypes for chromosomes I, IV, V, and X) for each isotype in the four Hawaiian populations and the non-Hawaiian C population.

Supplementary Data 1: Complete sampling data for the collections and isolations completed for this study as a .csv file.

Supplementary Data 2: Cryopreserved strains that are available upon request, including any known alternative names that might be used in the literature.

Supplementary Data 3: Variants among the 276 isotypes referred to as the 'PopGen VCF'.

Supplementary File 1 - Collection categories identified on each island. Multiple collection categories were found for some samples. For this reason, the total number of distinct collections (2,594) exceeds the total number of samples (2,263).

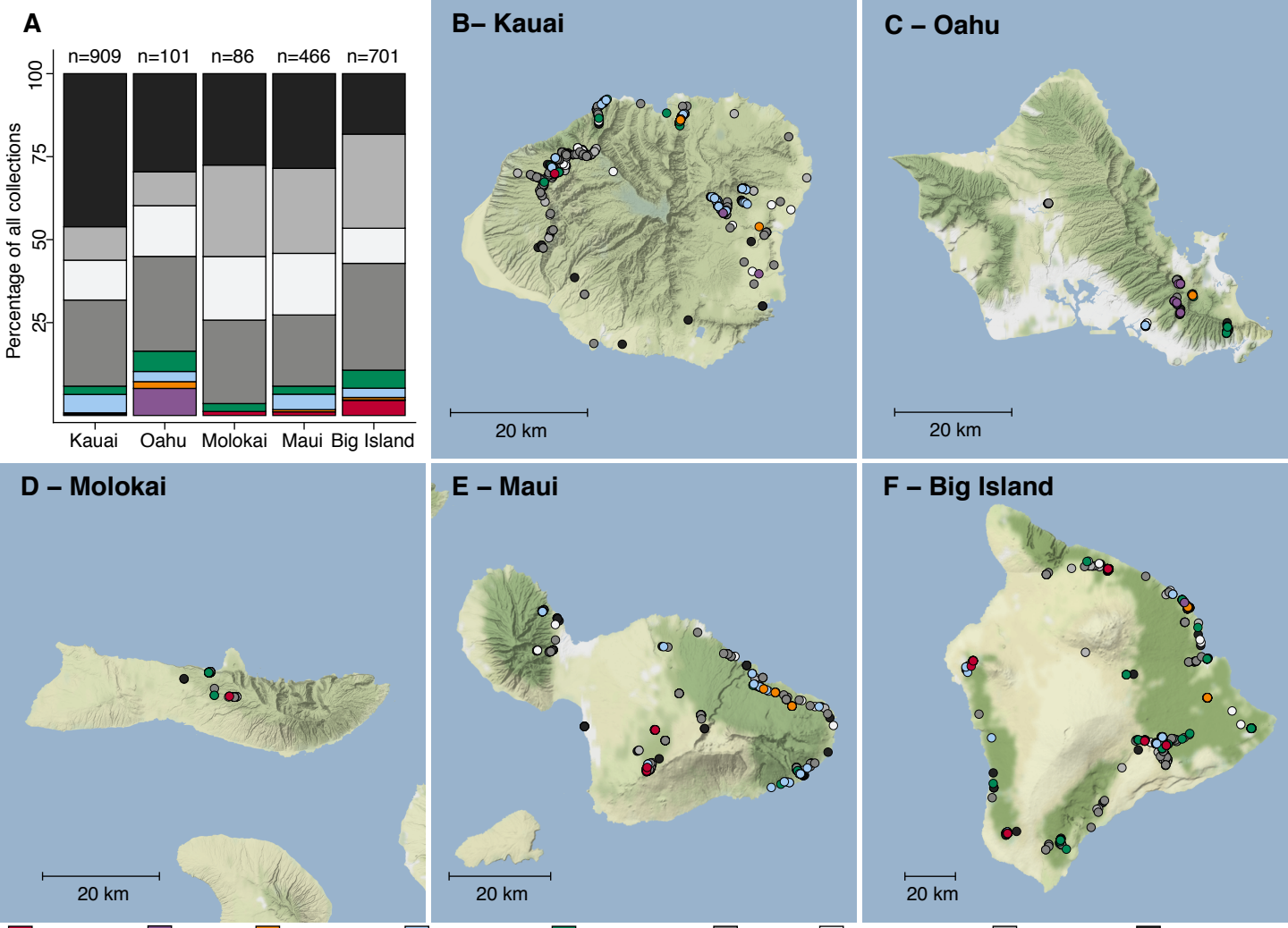
Supplementary File 2: *Caenorhabditis oiwi* new species description.

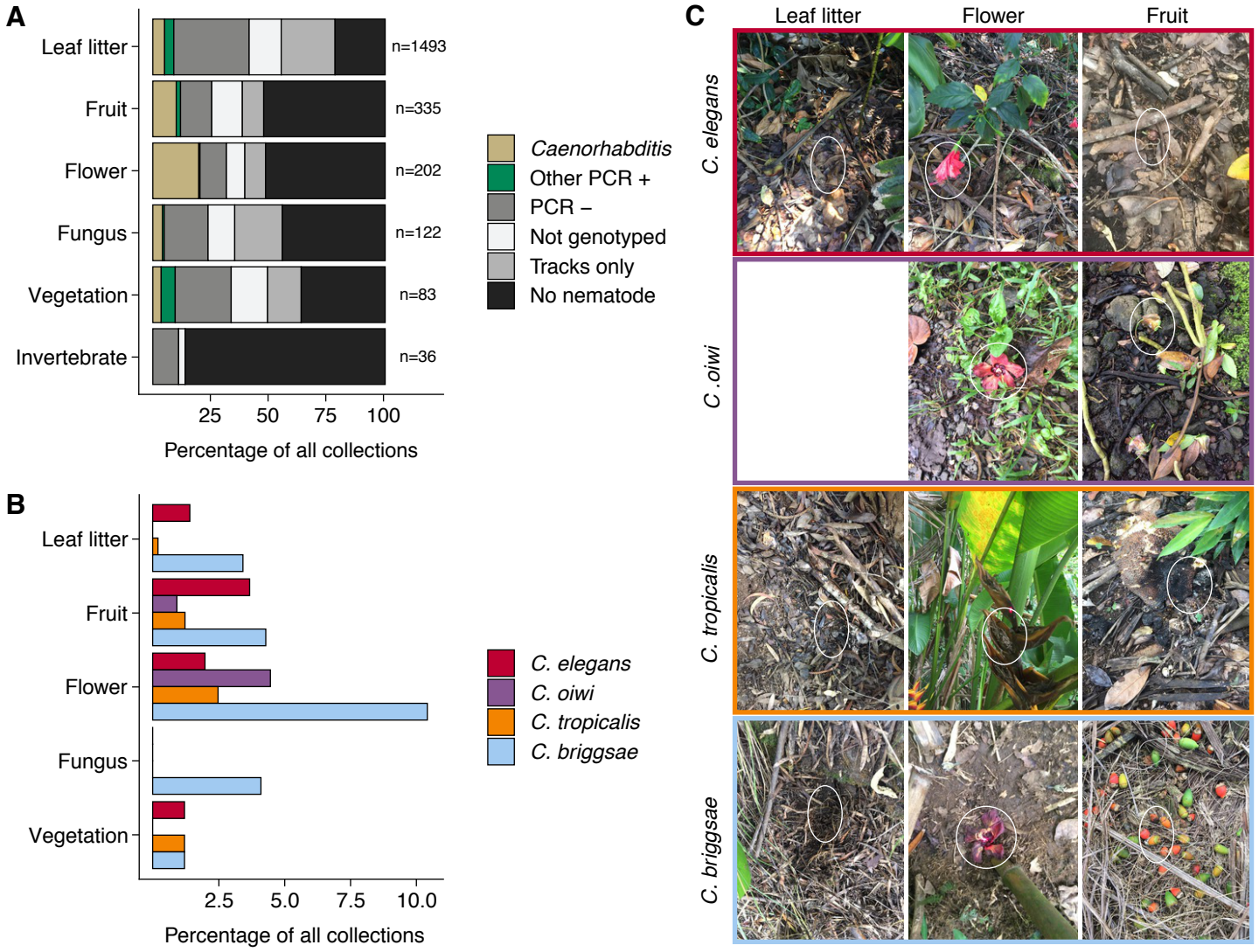
Supplementary File 3 - Description of variant sets used in this study. The additional filters applied to the PopGen VCF for specific uses are described in methods.

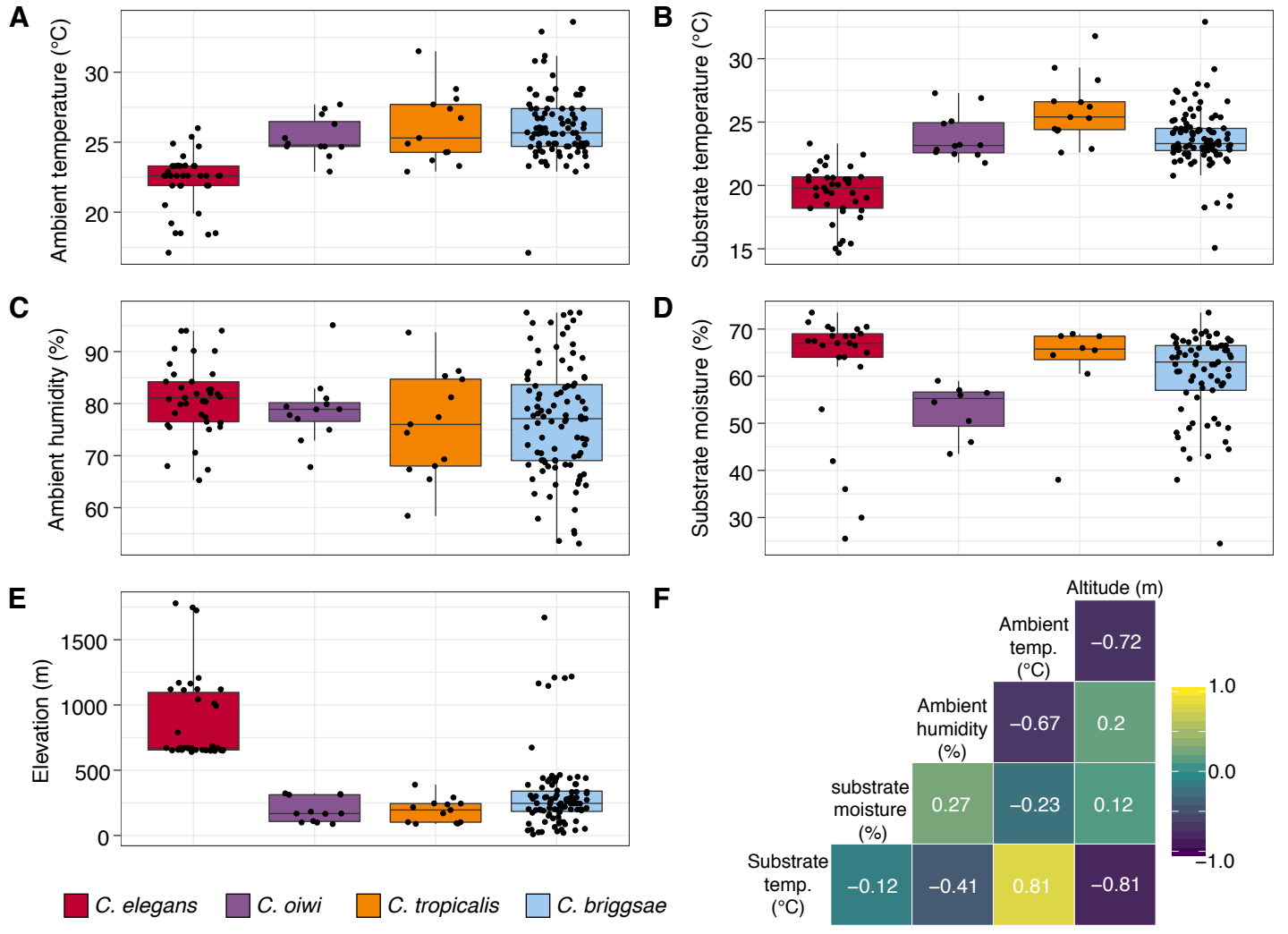
Supplementary File 4 - Nematode field sampling data form. Sampling data are entered into fields of the data form and stored in a cloud database at Fulcrum® (http://www.fulcrumapp.com).

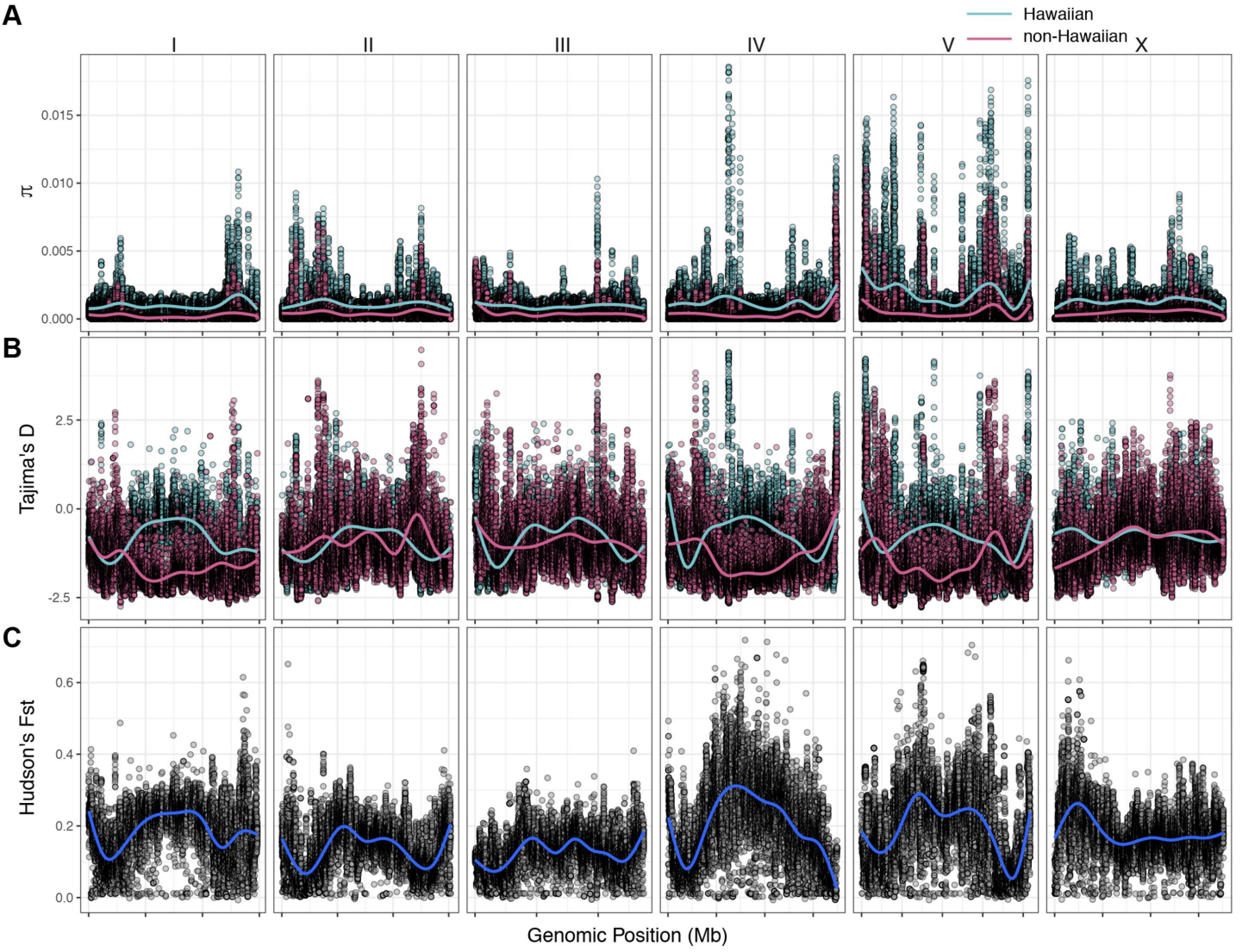
Supplementary File 5 - Nematode isolation data form. Nematode isolation data are entered into fields of the data form. All data associated with the substrate, C-label, and S-label are linked and stored in a cloud database at Fulcrum® (http://www.fulcrumapp.com).

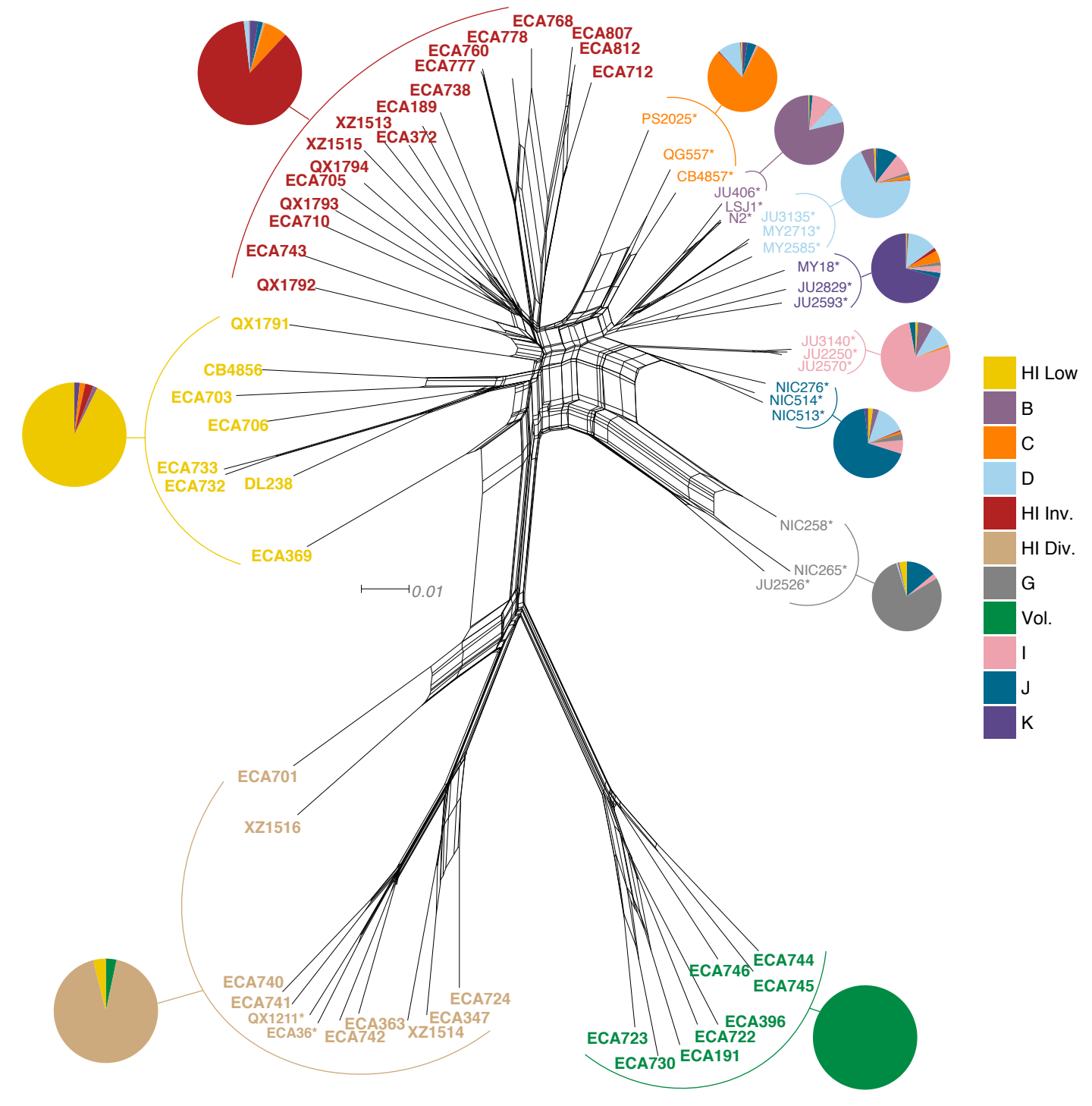
Supplementary File 6 - Nextflow pipelines used in our study.

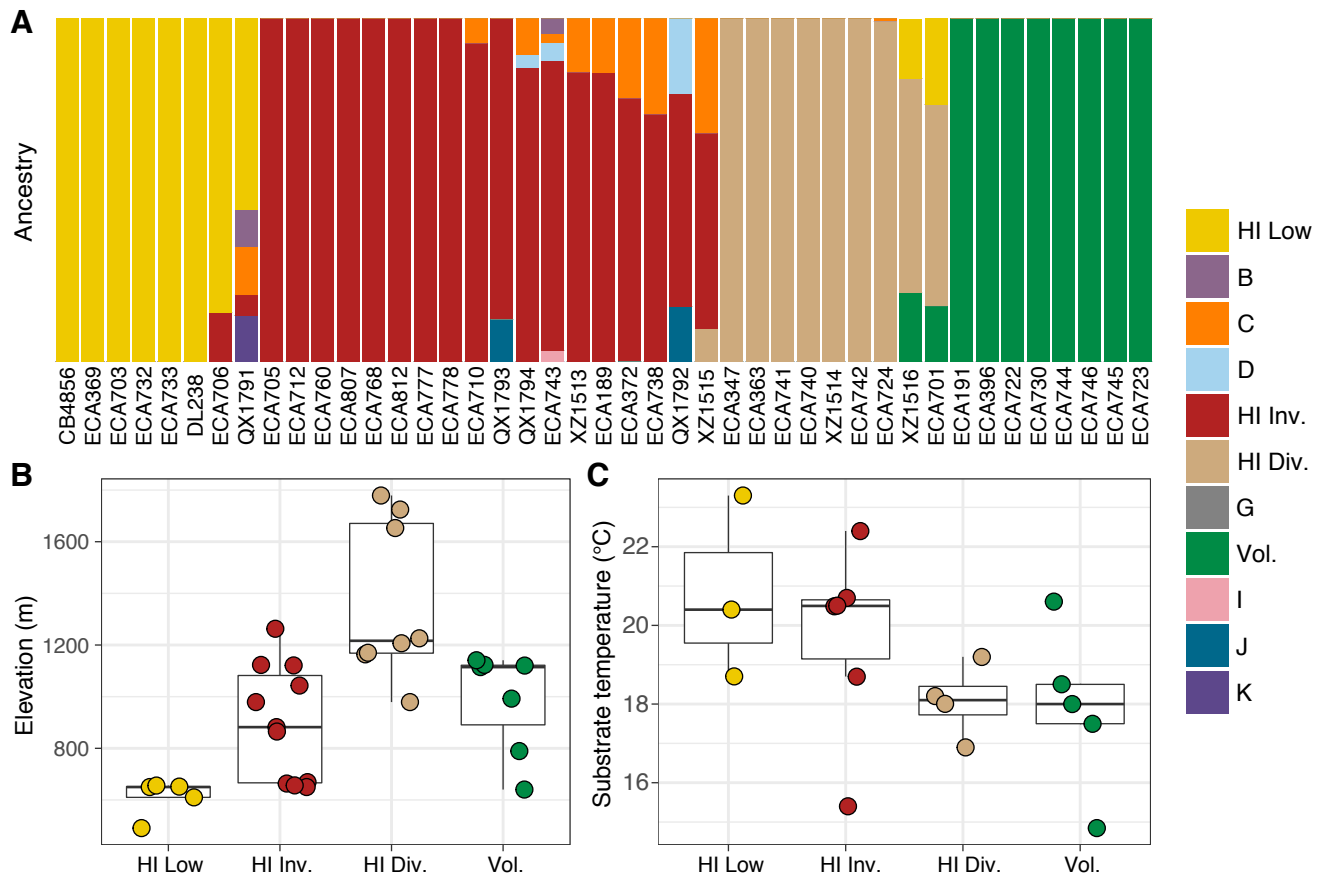

A
n=909
n=101
n=86
n=466
n=701
Percentage of all collections
100
75
50
25
Kauai
Oahu
Molokai
Maui
Big Island
B– Kauai
20 km
C – Oahu
20 km
D – Molokai
20 km
E – Maui
20 km
F – Big Island
20 km
C. elegans
C. oiwi
C. tropicalis
C. briggsae
Other PCR +
PCR –
Not genotyped
Tracks only
No nematode

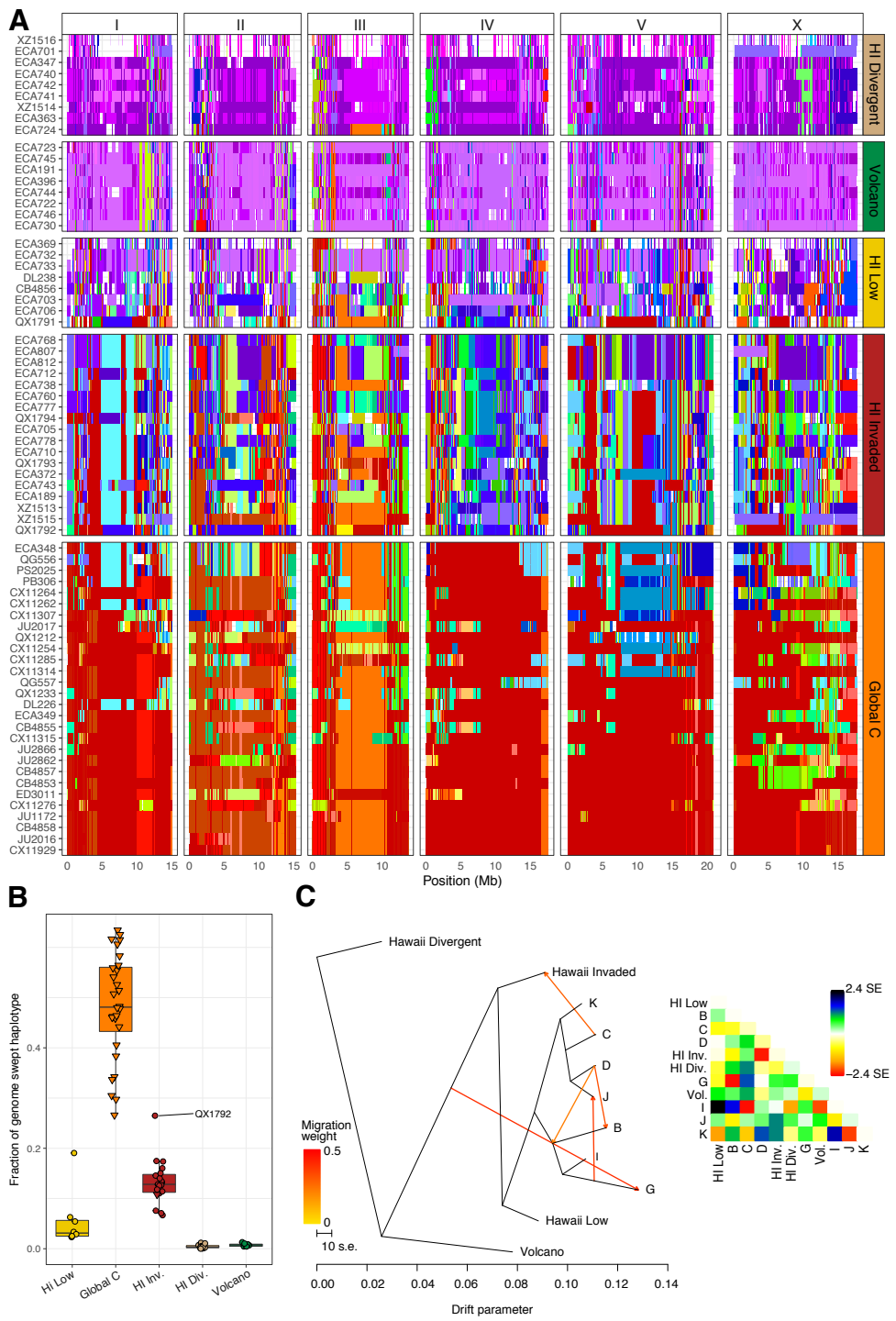

A
Leaf litter
Fruit
Flower
Fungus
Vegetation
Invertebrate
n=1493
n=335
n=202
n=122
n=83
n=36
25
50
75
100
Percentage of all collections
Caenorhabditis
Other PCR +
PCR –
Not genotyped
Tracks only
No nematode
B
Leaf litter
Fruit
Flower
Fungus
Vegetation
2.5
5.0
7.5
10.0
Percentage of all collections
C. elegans
C. oiwi
C. tropicalis
C. briggsae
C
Leaf litter
Flower
Fruit
C. elegans
C .oiwi
C. tropicalis
C. briggsae

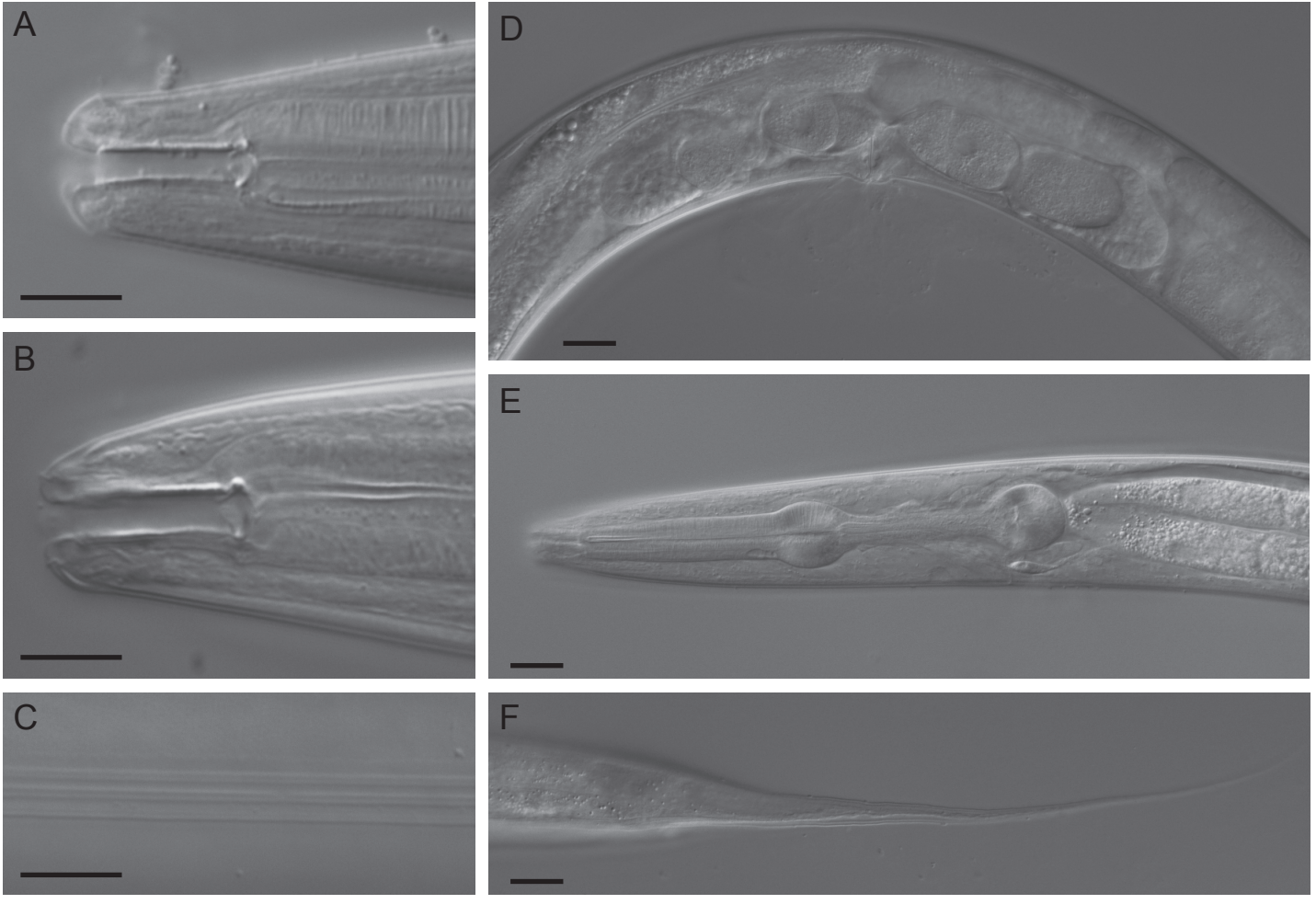

A
Ambient temperature (°C)
30
25
20
B
Substrate temperature (°C)
30
25
20
15
C
Ambient humidity (%)
90
80
70
60
D
Substrate moisture (%)
70
60
50
40
30
E
Elevation (m)
1500
1000
500
0
F
Altitude (m)
Ambient temp. (°C)
Ambient humidity (%)
substrate moisture (%)
Substrate temp. (°C)
−0.72
−0.67
0.2
0.27
−0.23
0.12
−0.12
−0.41
0.81
−0.81
1.0
0.0
−1.0
C. elegans
C. oiwi
C. tropicalis
C. briggsae

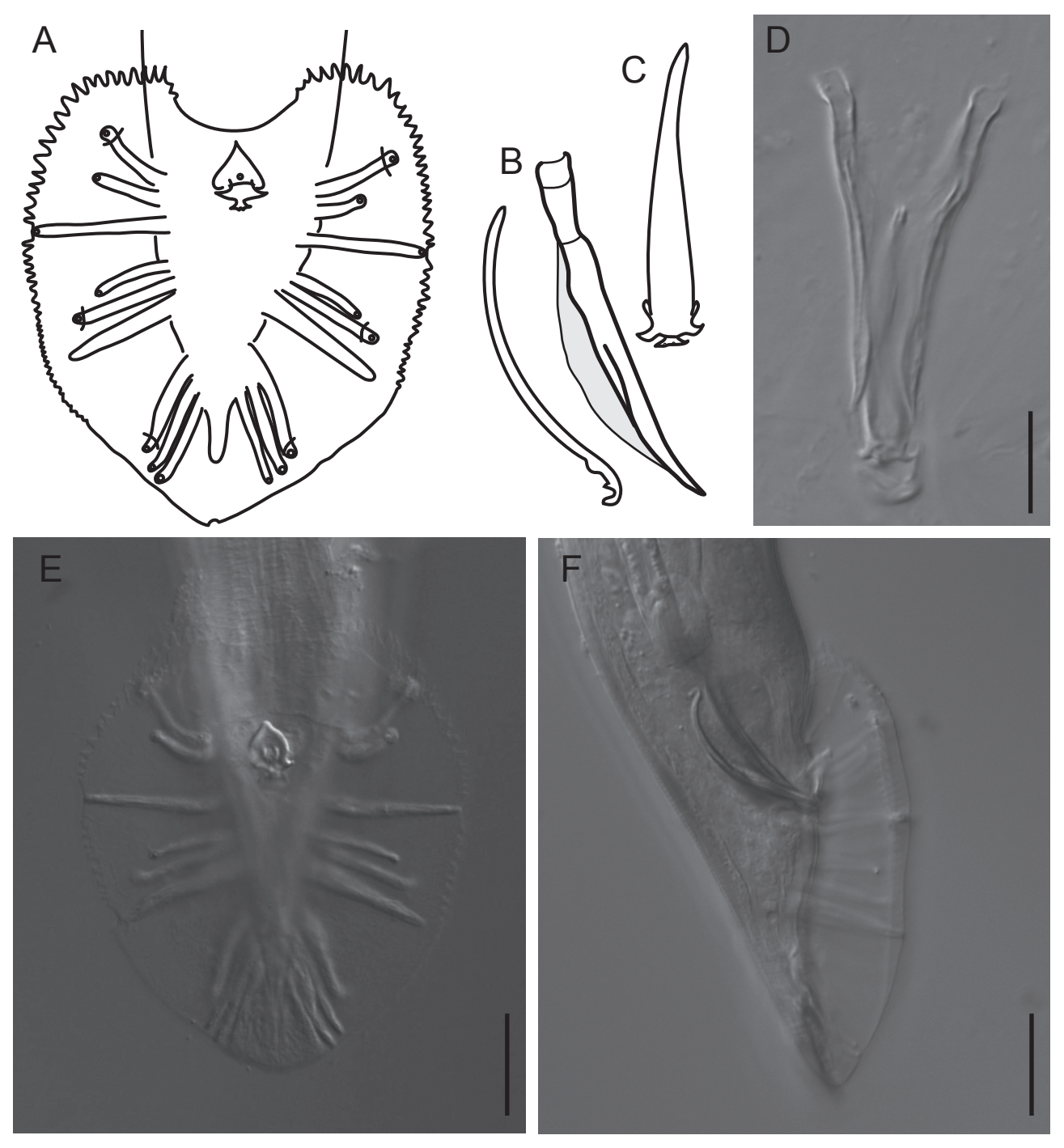

A
B
C
Hawaiian
non-Hawaiian
I
II
III
IV
V
X
π
0.015
0.010
0.005
0.000
Tajima's D
2.5
0.0
-2.5
Hudson's Fst
0.6
0.4
0.2
0.0
Genomic Position (Mb)

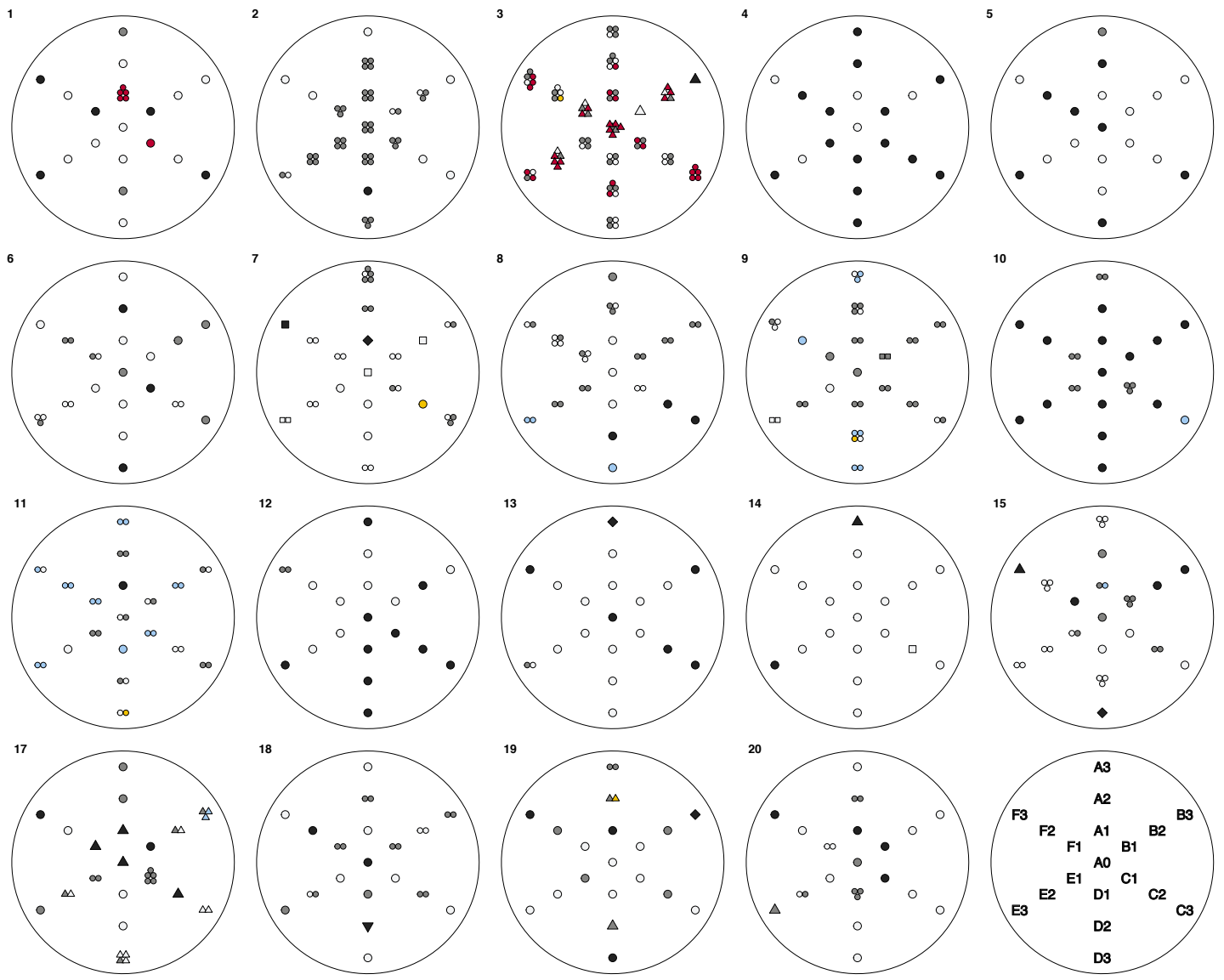

ECA768
ECA778
ECA807
ECA760
ECA812
ECA777
ECA712
ECA738
ECA189
XZ1513
ECA372
XZ1515
QX1794
ECA705
QX1793
ECA710
ECA743
QX1792
QX1791
CB4856
ECA703
ECA706
ECA733
ECA732
DL238
ECA369
PS2025*
QG557*
CB4857*
JU406*
LSJ1*
N2*
JU3135*
MY2713*
MY2585*
MY18*
JU2829*
JU2593*
JU3140*
JU2250*
JU2570*
NIC276*
NIC514*
NIC513*
NIC258*
NIC265*
JU2526*
0.01
ECA701
XZ1516
ECA740
ECA741
QX1211*
ECA36*
ECA742
ECA363
XZ1514
ECA347
ECA724
ECA744
ECA746
ECA745
ECA396
ECA722
ECA723
ECA730
ECA191
HI Low
B
C
D
HI Inv.
HI Div.
G
Vol.
I
J
K

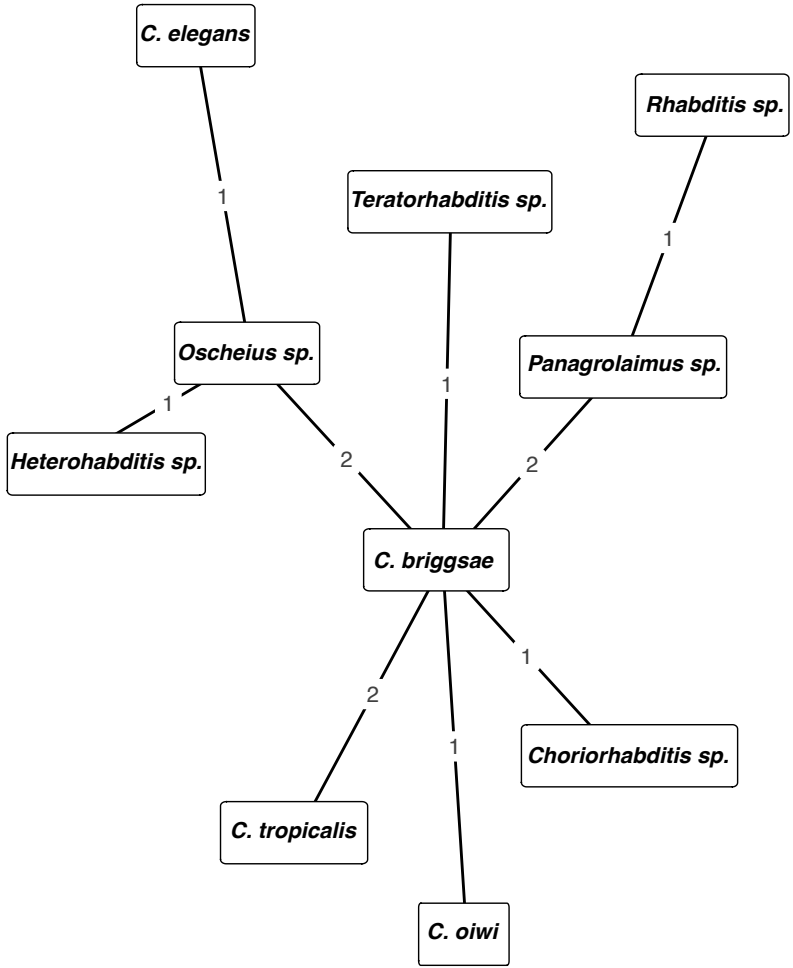

A
Ancestry
CB4856
ECA369
ECA703
ECA732
ECA733
DL238
ECA706
QX1791
ECA705
ECA712
ECA760
ECA807
ECA768
ECA812
ECA777
ECA778
ECA710
QX1793
QX1794
ECA743
XZ1513
ECA189
ECA372
ECA738
QX1792
XZ1515
ECA347
ECA363
ECA741
ECA740
XZ1514
ECA742
ECA724
XZ1516
ECA701
ECA191
ECA396
ECA722
ECA730
ECA744
ECA746
ECA745
ECA723
HI Low
B
C
D
HI Inv.
HI Div.
G
Vol.
I
J
K
B
Elevation (m)
1600
1200
800
HI Low
HI Inv.
HI Div.
Vol.
C
Substrate temperature (°C)
22
20
18
16
HI Low
HI Inv.
HI Div.
Vol.

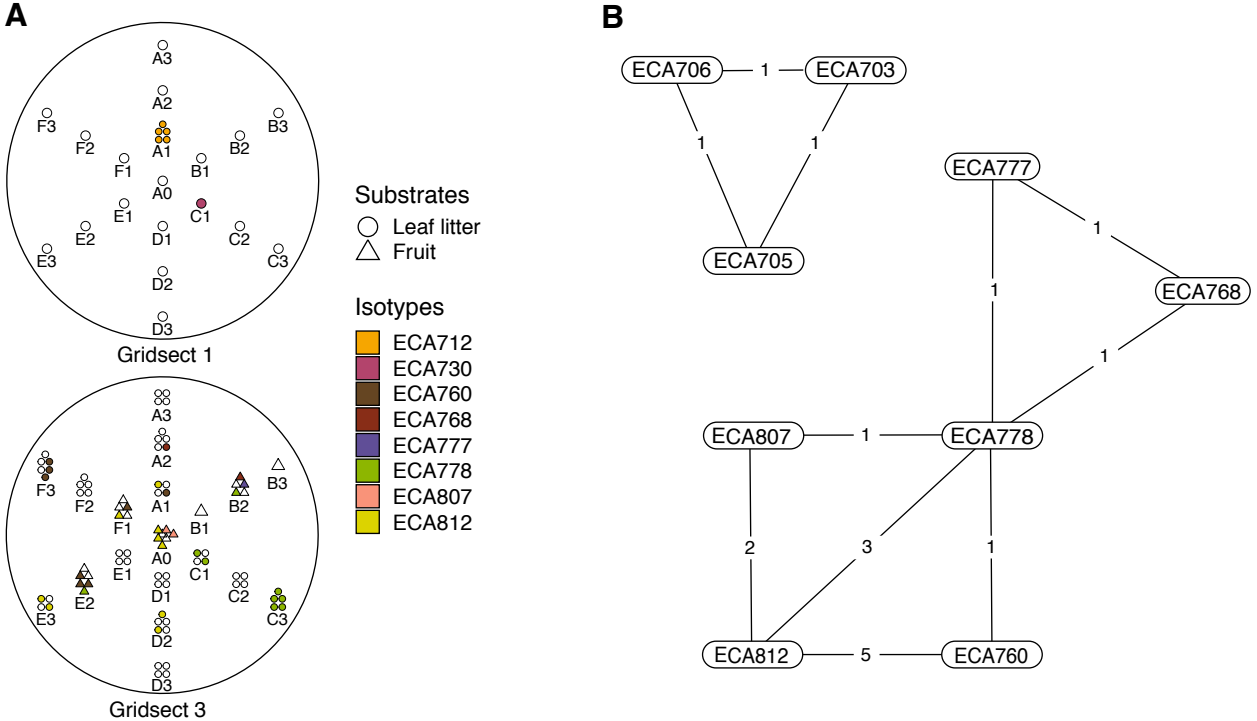

A
I
II
III
IV
V
X
XZ1516
ECA701
ECA347
ECA740
ECA742
ECA741
XZ1514
ECA363
ECA724
HI Divergent
ECA723
ECA745
ECA191
ECA396
ECA744
ECA722
ECA746
ECA730
Volcano
ECA369
ECA732
ECA733
DL238
CB4856
ECA703
ECA706
QX1791
HI Low
ECA768
ECA807
ECA812
ECA712
ECA738
ECA760
ECA777
QX1794
ECA705
ECA778
ECA710
QX1793
ECA372
ECA743
ECA189
XZ1513
XZ1515
QX1792
HI Invaded
ECA348
QG556
PS2025
PB306
CX11264
CX11262
CX11263
JU2017
QX1212
CX11254
CX11285
CX11314
QG557
QX1233
DL226
ECA349
CB4855
CX11315
JU2866
JU2862
CB4857
CB4853
ED3011
CX11276
JU1172
CB4858
JU2016
CX11929
Global C
0 5 10 15
0 5 10 15
0 5 10
0 5 10 15
0 5 10 15
0 5 10 15 20
0 5 10 15
Position (Mb)
B
Fraction of genome swept haplotype
0.0
0.2
0.4
QX1792
HI Low
Global C
HI Inv.
HI Div.
Volcano
C
Hawaii Divergent
Hawaii Invaded
K
C
D
J
B
I
G
Hawaii Low
Volcano
Migration weight
0.5
0
10 s.e.
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14
Drift parameter
2.4 SE
−2.4 SE
HI Low
B
C
D
HI Inv.
HI Div.
G
Vol.
I
J
K

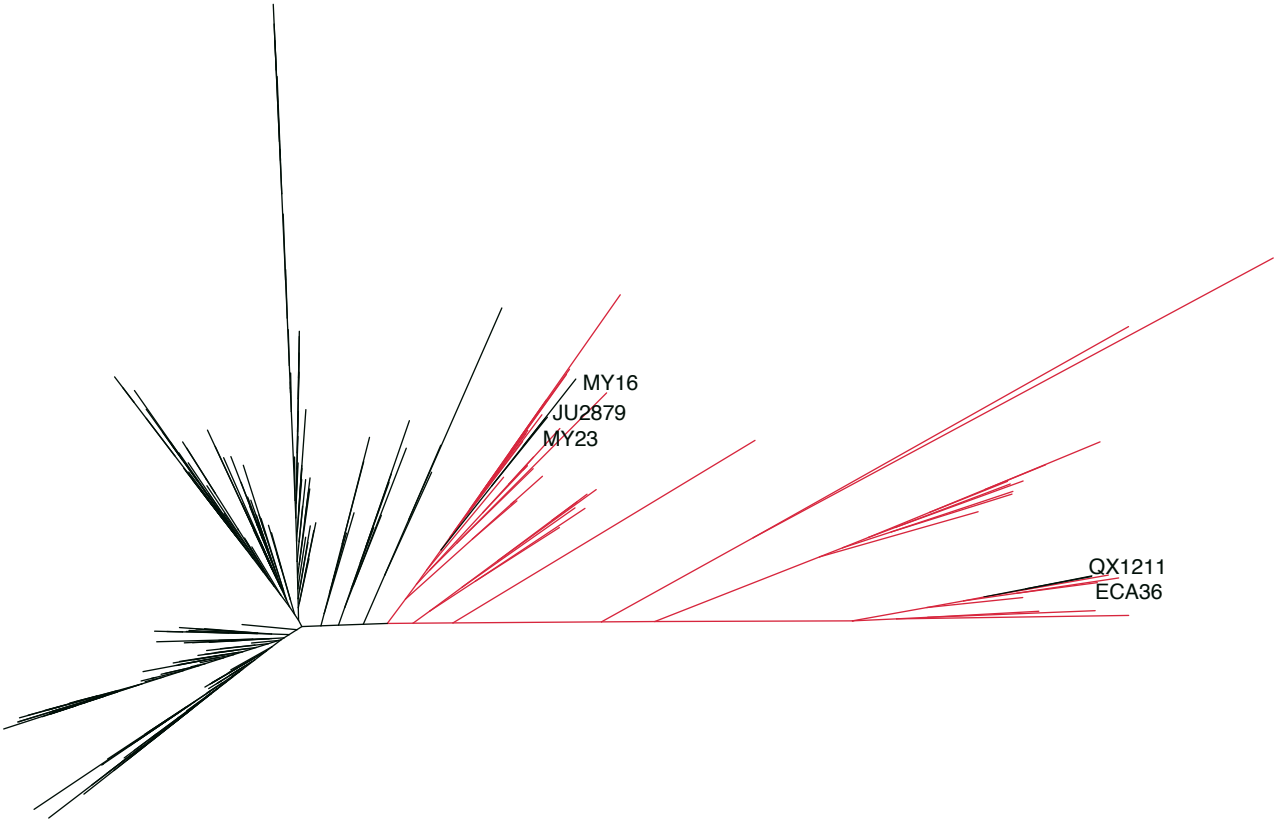

A
B
C
D
E
F

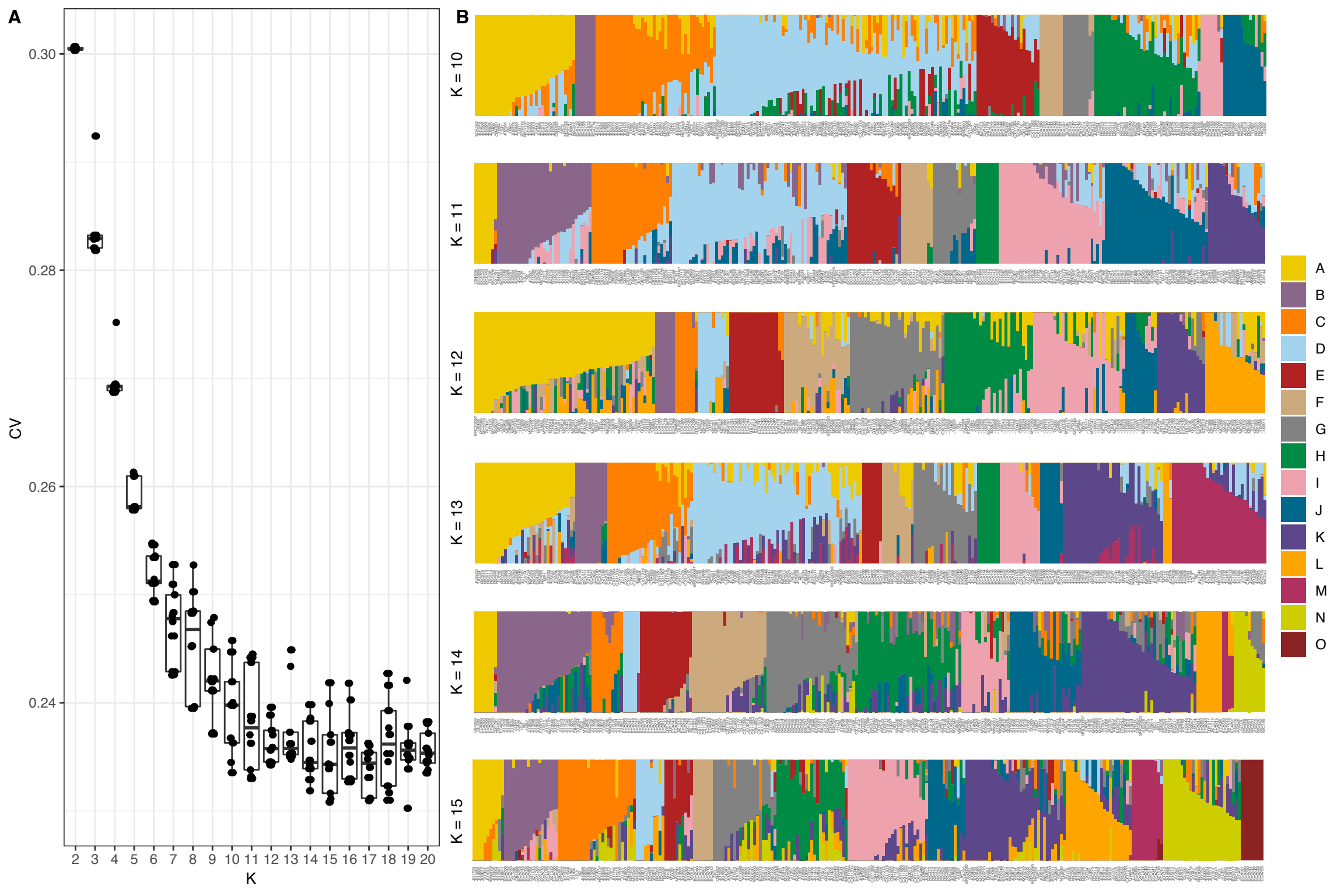

A
B
C
D
E
F

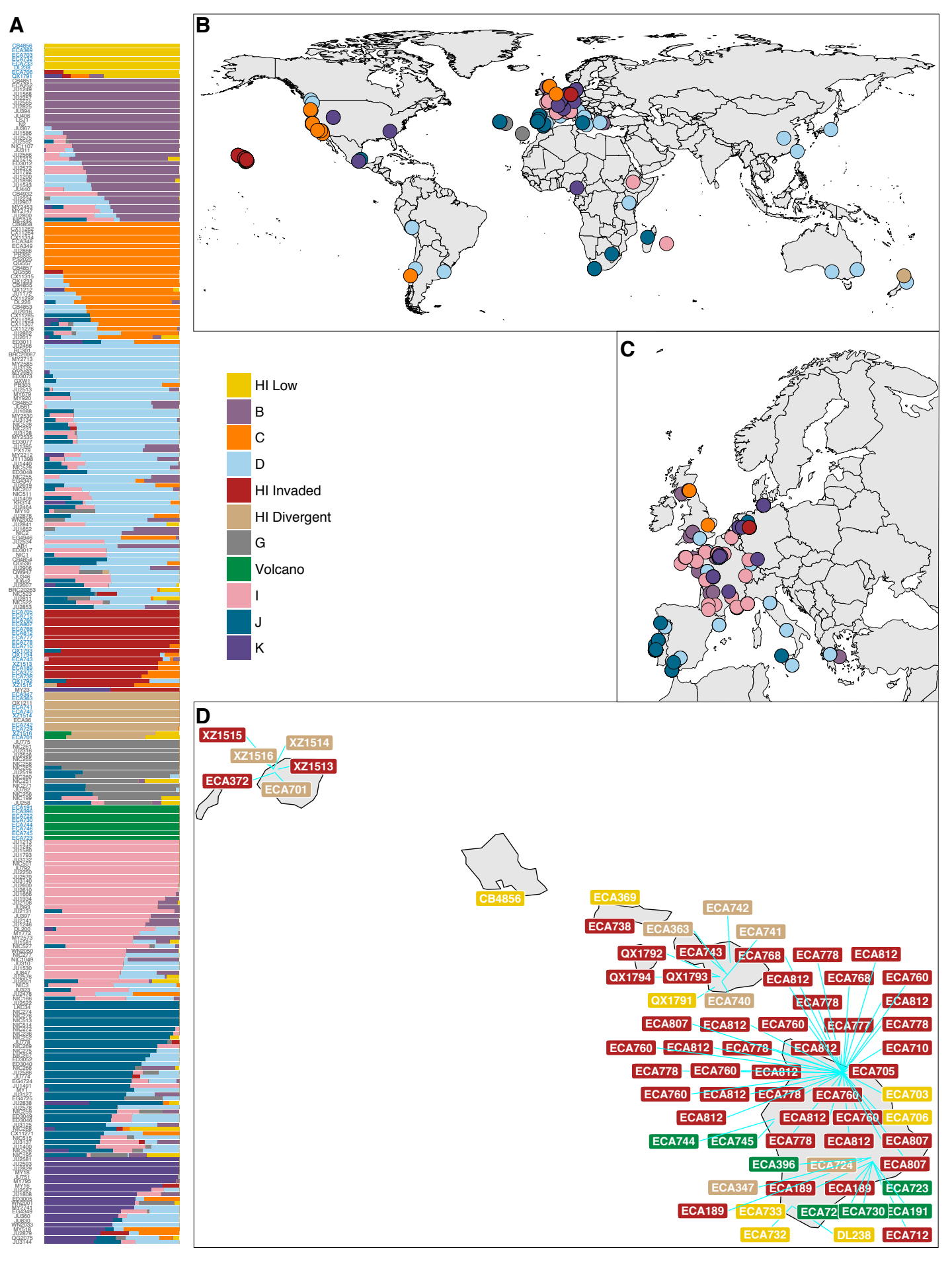

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
17
18
19
20
A3
A2
F3
B3
F2
A1
B2
F1
B1
A0
E1
C1
E2
D1
C2
E3
C3
D2
D3
species
C. elegans
C. oiwi
C. tropicalis
Panagrolaimus sp.
Oscheius sp.
C. briggsae
Other PCR +
PCR –
Not genotyped
No Worm
substrates
Leaf litter
Fruit/nut/vegetable
Flower
Fungus
Isopod

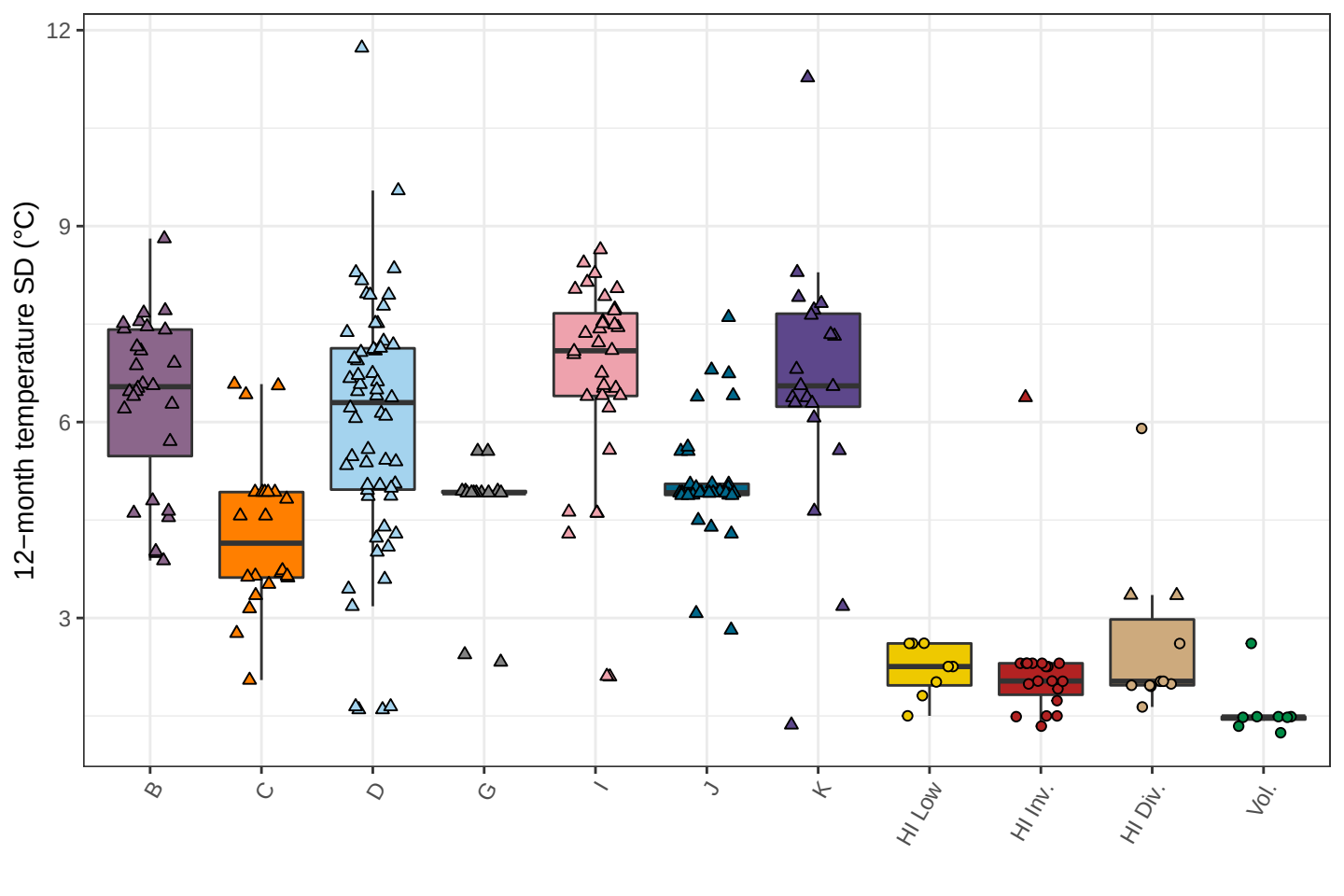

C. elegans
Rhabditis sp.
Teratorhabditis sp.
1
1
1
Oscheius sp.
Panagrolaimus sp.
1
1
2
2
Heterohabditis sp.
C. briggsae
1
2
1
Choriorhabditis sp.
C. tropicalis
C. oiwi

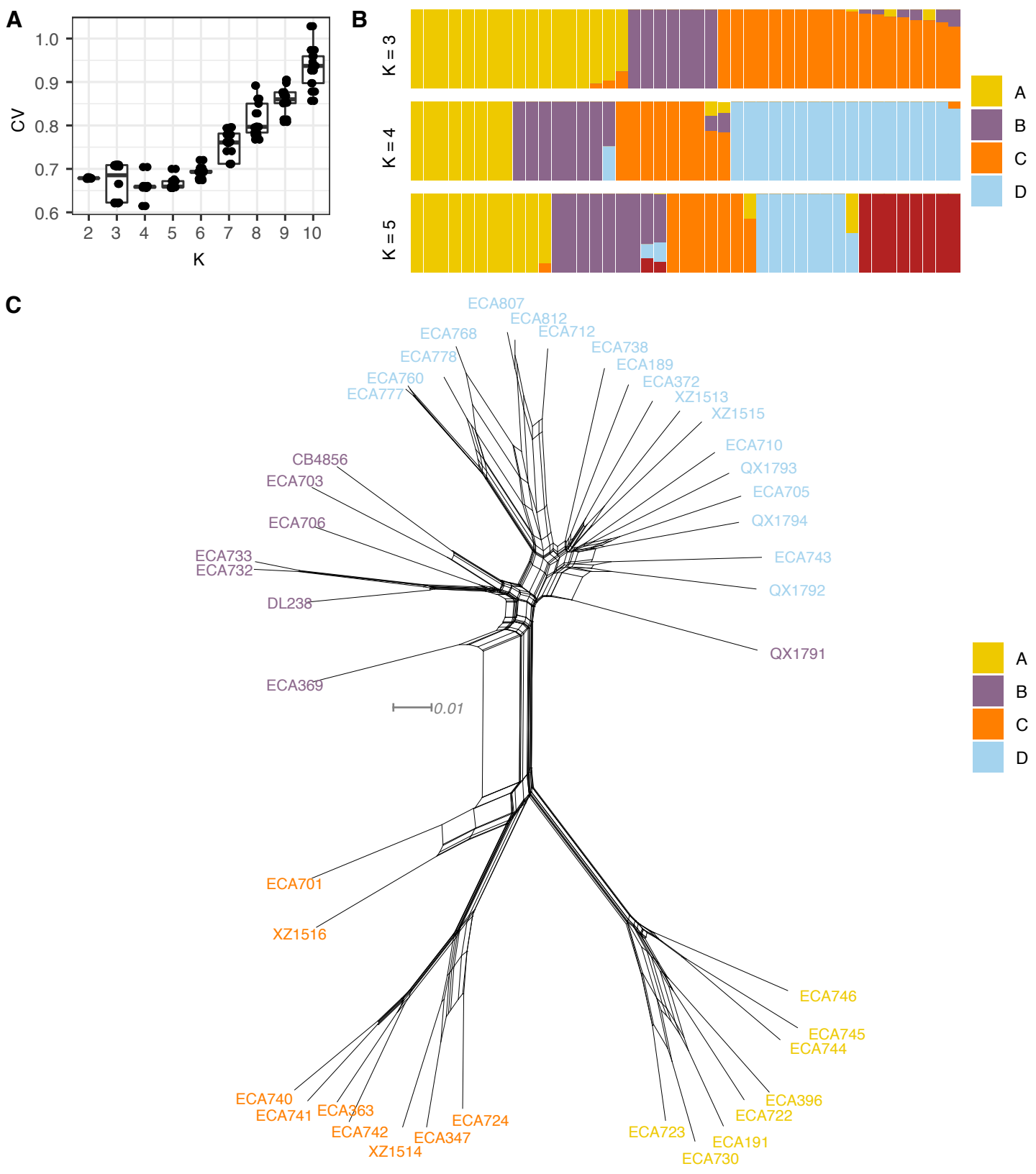

A
A3
A2
A1
A0
B3
B2
B1
C1
C2
C3
D1
D2
D3
E3
E2
E1
F3
F2
F1
Gridsect 1
Gridsect 3
Substrates
Leaf litter
Fruit
Isotypes
ECA712
ECA730
ECA760
ECA768
ECA777
ECA778
ECA807
ECA812
B
ECA706
ECA703
ECA705
ECA777
ECA768
ECA807
ECA778
ECA812
ECA760
1
2
3
5

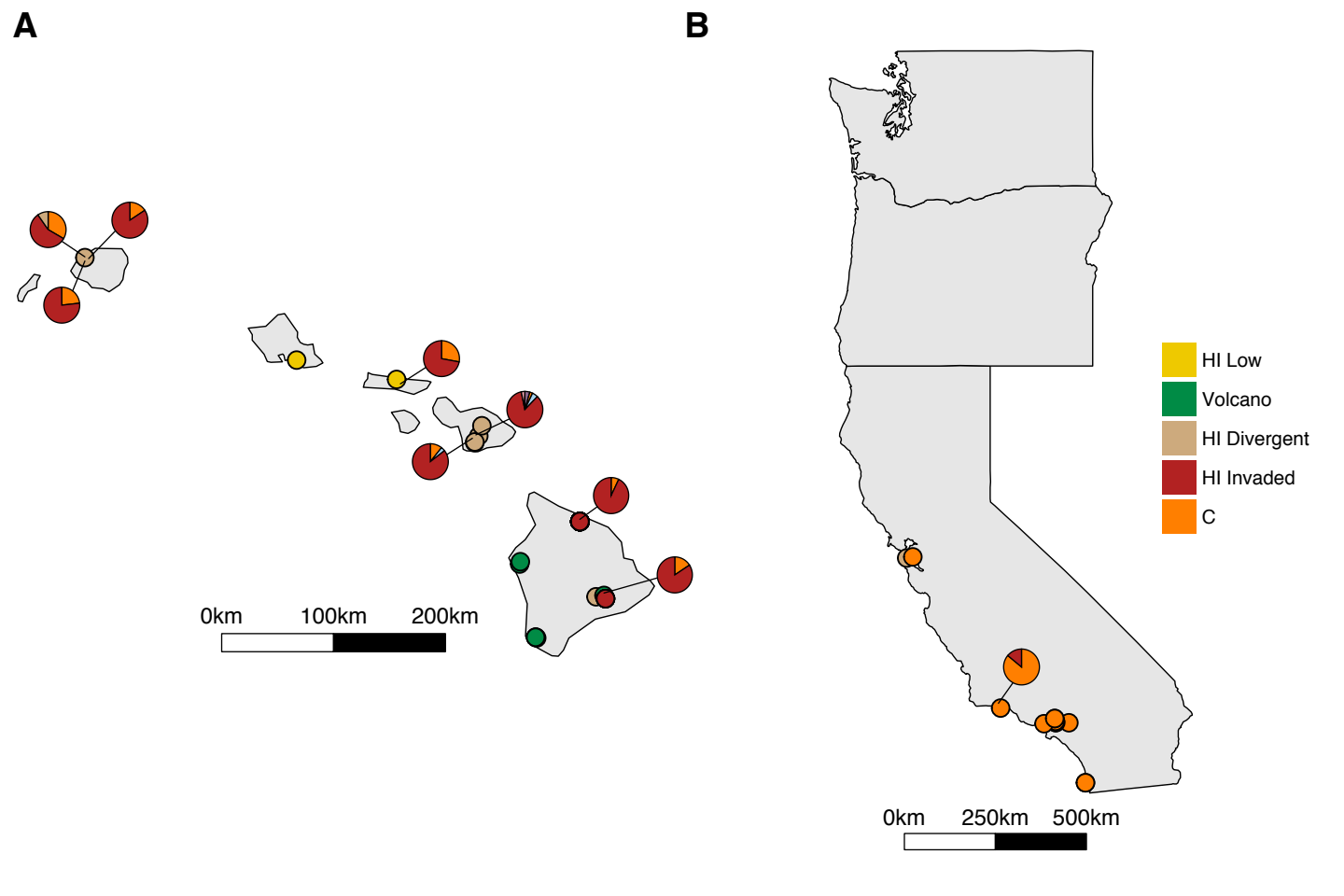

MY16
JU2879
MY23
QX1211
ECA36

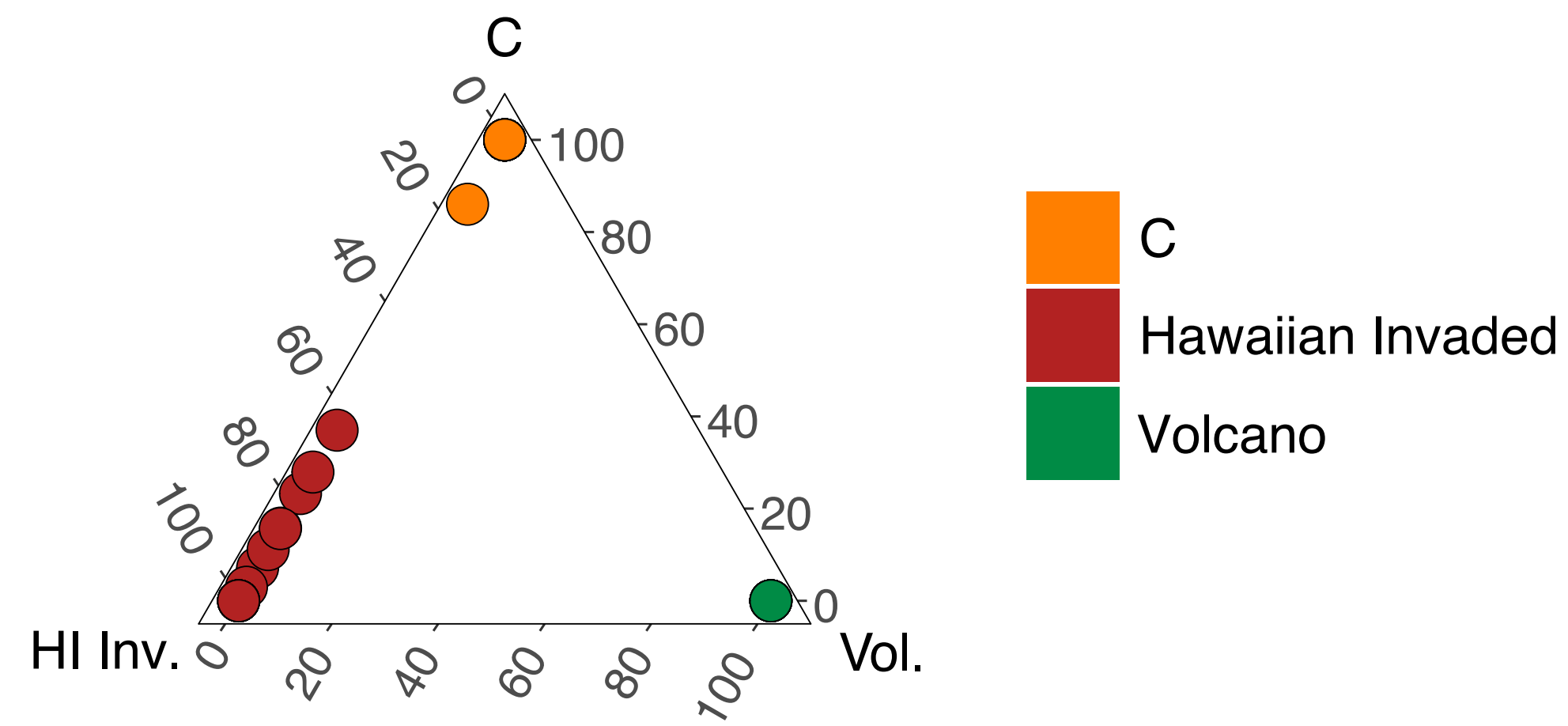

A
0.30
0.28
0.26
0.24
CV
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
K
B
K = 10
K = 11
K = 12
K = 13
K = 14
K = 15
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O

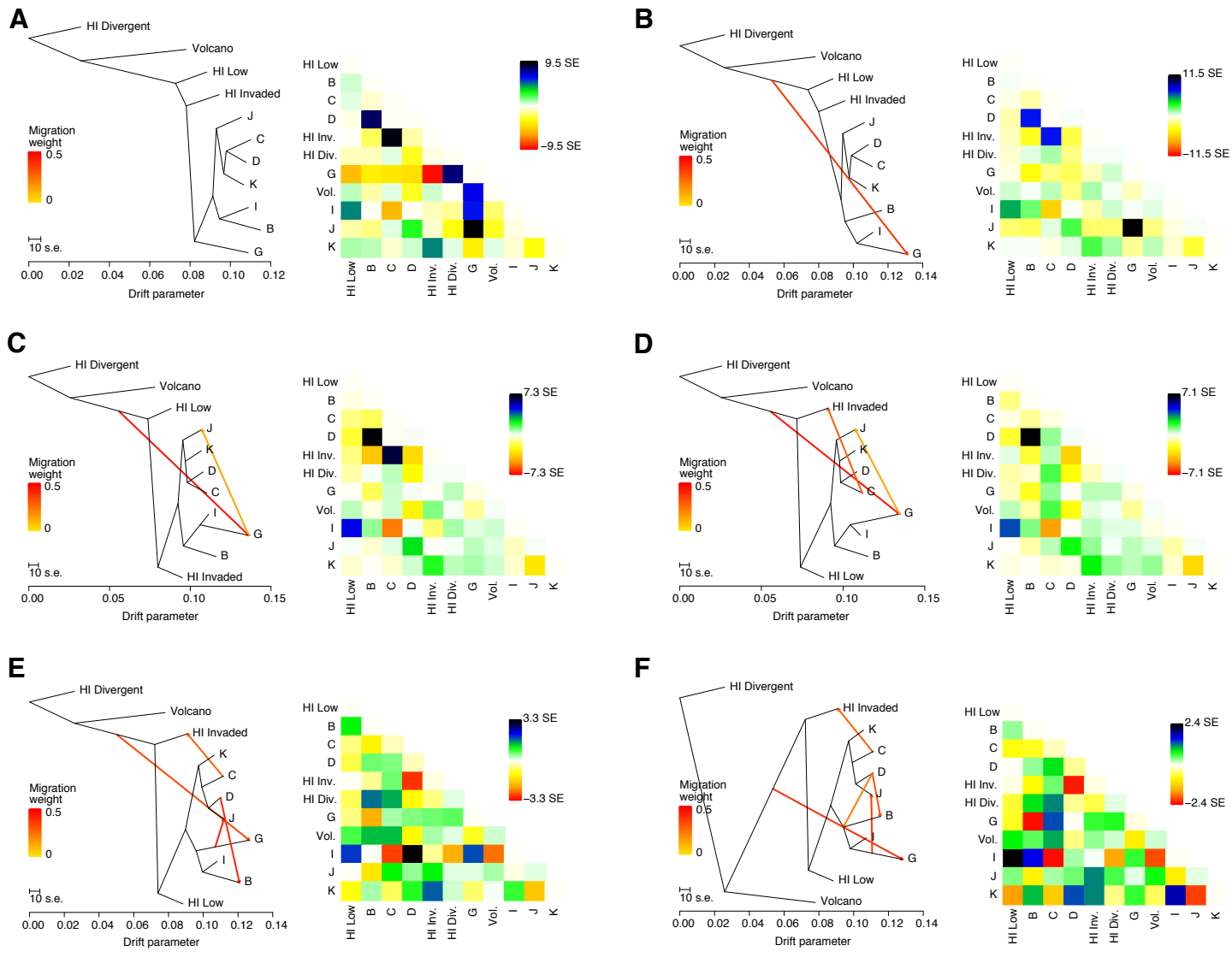

A
B
C
D
HI Low
B
C
D
HI Invaded
HI Divergent
G
Volcano
I
J
K
XZ1515
XZ1514
XZ1516
XZ1513
ECA372
ECA701
CB4856
ECA369
ECA742
ECA738
ECA363
ECA741
QX1792
ECA743
ECA768
ECA778
ECA812
QX1794
QX1793
ECA812
ECA768
ECA760
QX1791
ECA740
ECA778
ECA812
ECA807
ECA812
ECA760
ECA777
ECA778
ECA760
ECA812
ECA778
ECA812
ECA710
ECA778
ECA760
ECA812
ECA705
ECA760
ECA812
ECA778
ECA760
ECA703
ECA812
ECA812
ECA760
ECA706
ECA744
ECA745
ECA778
ECA812
ECA807
ECA396
ECA724
ECA807
ECA347
ECA189
ECA189
ECA723
ECA189
ECA733
ECA72
ECA730
ECA191
ECA732
DL238
ECA712

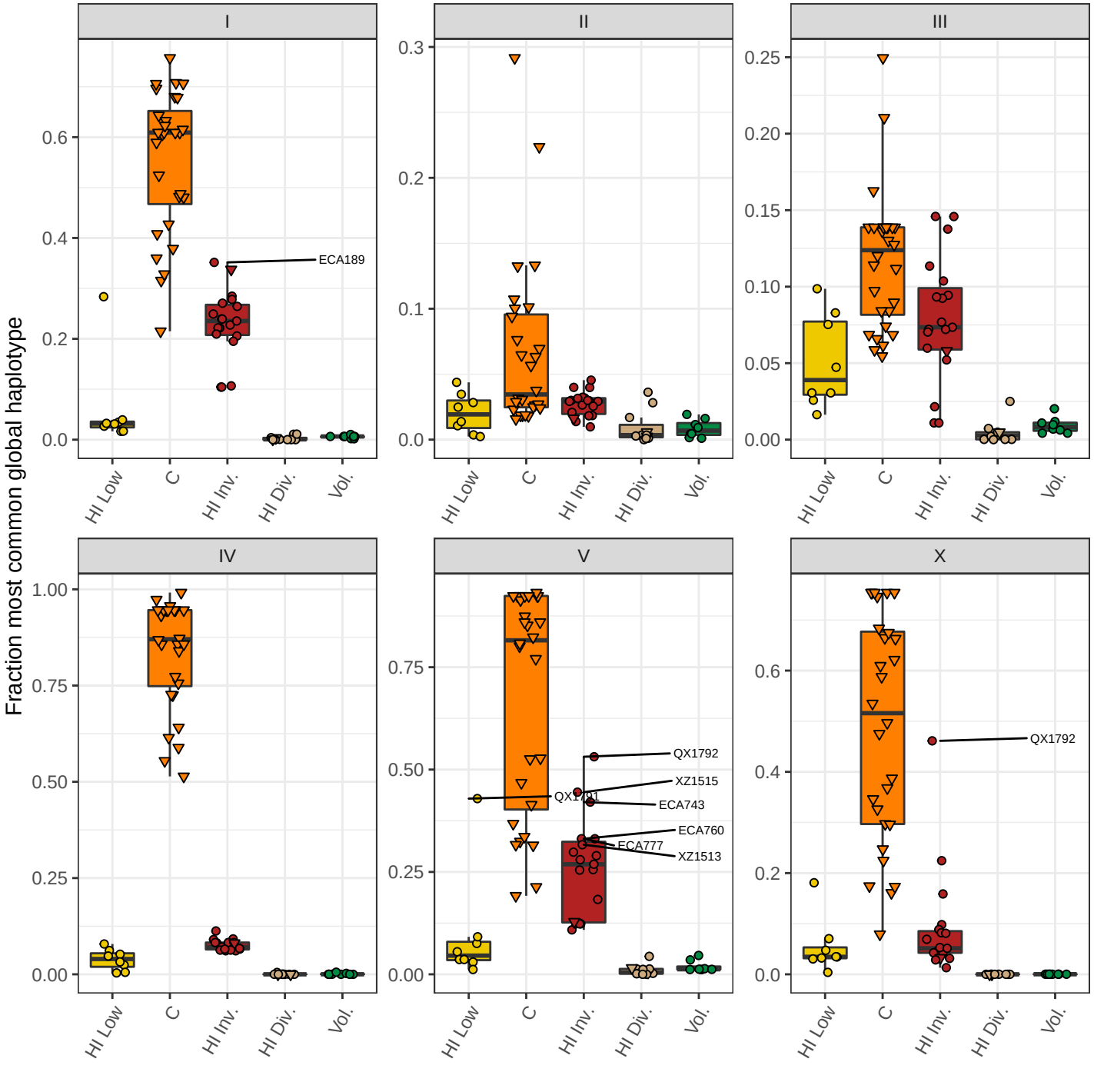

12
9
6
3
12−month temperature SD (°C)
B
C
D
G
I
J
K
HI Low
HI Inv.
HI Div.
Vol.


A
CV
K
B
K = 3
K = 4
K = 5
A
B
C
D
C
ECA807
ECA812
ECA768
ECA712
ECA778
ECA738
ECA760
ECA189
ECA777
ECA372
XZ1513
XZ1515
ECA710
CB4856
QX1793
ECA703
ECA705
ECA706
QX1794
ECA733
ECA743
ECA732
QX1792
DL238
QX1791
ECA369
0.01
ECA701
XZ1516
ECA746
ECA745
ECA744
ECA740
ECA396
ECA741
ECA363
ECA722
ECA724
ECA742
ECA723
ECA347
ECA191
XZ1514
ECA730


A
0km
100km
200km
B
HI Low
Volcano
HI Divergent
HI Invaded
C
0km
250km
500km


C
0
20
40
60
80
100
100
80
60
40
20
0
HI Inv.
0
20
40
60
80
100
Vol.
C
Hawaiian Invaded
Volcano


A
HI Divergent
Volcano
HI Low
HI Invaded
J
C
D
K
I
B
G
Migration weight
0.5
0
10 s.e.
0.00 0.02 0.04 0.06 0.08 0.10 0.12
Drift parameter
HI Low B C D HI Inv. HI Div. G Vol. I J K
9.5 SE
−9.5 SE
B
HI Divergent
Volcano
HI Low
HI Invaded
J
D
C
K
B
I
G
Migration weight
0.5
0
10 s.e.
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14
Drift parameter
HI Low B C D HI Inv. HI Div. G Vol. I J K
11.5 SE
−11.5 SE
C
HI Divergent
Volcano
HI Low
J
K
D
C
I
G
B
HI Invaded
Migration weight
0.5
0
10 s.e.
0.00 0.05 0.10 0.15
Drift parameter
HI Low B C D HI Inv. HI Div. G Vol. I J K
7.3 SE
−7.3 SE
D
HI Divergent
Volcano
HI Invaded
J
K
D
C
G
I
B
HI Low
Migration weight
0.5
0
10 s.e.
0.00 0.05 0.10 0.15
Drift parameter
HI Low B C D HI Inv. HI Div. G Vol. I J K
7.1 SE
−7.1 SE
E
HI Divergent
Volcano
HI Invaded
K
C
D
J
G
I
B
HI Low
Migration weight
0.5
0
10 s.e.
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14
Drift parameter
HI Low B C D HI Inv. HI Div. G Vol. I J K
3.3 SE
−3.3 SE
F
HI Divergent
HI Invaded
K
C
D
J
B
I
G
HI Low
Volcano
Migration weight
0.5
0
10 s.e.
0.00 0.02 0.04 0.06 0.08 0.10 0.12 0.14
Drift parameter
HI Low B C D HI Inv. HI Div. G Vol. I J K
2.4 SE
−2.4 SE


I
II
III
IV
V
X
Fraction most common global haplotype
HI Low
C
HI Inv.
HI Div.
Vol.
ECA189
QX1792
XZ1515
QX1791
ECA743
ECA760
ECA777
XZ1513