

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

3

4 Timothy A. Crombie¹, Stefan Zdraljevic^{1,2}, Daniel E. Cook^{1,2}, Robyn E. Tanny¹, Shannon C. Brady^{1,2}, Ye
5 Wang¹, Kathryn S. Evans^{1,2}, Steffen Hahnel¹, Daehan Lee¹, Briana C. Rodriguez¹, Gaotian Zhang¹, Joost
6 van der Zwaag¹, Karin C. Kiontke³, and Erik C. Andersen^{1,*}

7

8 1. Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

9 2. Interdisciplinary Biological Sciences Program, Northwestern University, Evanston, IL 60208, USA

10 3. Department of Biology, New York University, New York, NY 10003, USA

11 * Corresponding author

12

13 **Erik C. Andersen**

14 Associate Professor of Molecular Biosciences

15 Northwestern University

16 Evanston, IL 60208, USA

17 Tel: (847) 467-4382

18 Email: erik.andersen@northwestern.edu

19

20 Timothy A. Crombie, tcrombie@northwestern.edu, ORCID 0000-0002-5645-4154

21 Stefan Zdraljevic, stefanzdraljevic2018@u.northwestern.edu, ORCID 0000-0003-2883-4616

22 Daniel E. Cook, daniel.cook@crick.ac.uk, ORCID 0000-0003-3347-562X

23 Robyn E. Tanny, robyn.tanny@northwestern.edu, ORCID 0000-0002-0611-3909

24 Shannon C. Brady, shannonbrady2014@u.northwestern.edu, ORCID 0000-0002-3043-1544

25 Ye Wang, ye.wang@northwestern.edu, ORCID 0000-0002-5423-6196

26 Kathryn S. Evans, kathryn.evans@u.northwestern.edu, ORCID 0000-0002-1388-8155

27 Steffen Hahnel, steffenhahnel@directbox.com, ORCID 0000-0001-8848-0691

28 Daehan Lee, daehan.lee@northwestern.edu, ORCID 0000-0002-0546-8484

29 Briana C. Rodriguez, briana.rodriguez@northwestern.edu , ORCID 0000-0002-5282-0815

30 Gaotian Zhang, gaotian.zhang@northwestern.edu, ORCID 0000-0001-6468-1341

31 J van der Zwaag, joost.vanderzwaag@wur.nl

32 Karin C. Kiontke, kk52@nyu.edu, ORCID 0000-0003-1588-4884

33 Erik Andersen, erik.andersen@northwestern.edu, ORCID 0000-0003-0229-9651

34 **Running title:** Hawaiian *C. elegans* diversity

35 **Keywords:**

36 *Caenorhabditis*, *C. elegans*, genetic diversity, niche, admixture

37 **Abstract**

38 Recent efforts to understand the natural niche of the keystone model organism *Caenorhabditis elegans* have
39 suggested that this species is cosmopolitan and associated with rotting vegetation and fruits. However, most
40 of the strains isolated from nature have low genetic diversity likely because recent chromosome-scale
41 selective sweeps contain alleles that increase fitness in human-associated habitats. Strains from the Hawaii
42 Islands are highly divergent from non-Hawaiian strains. This result suggests that Hawaiian strains might
43 contain ancestral genetic diversity that was purged from most non-Hawaiian strains by the selective sweeps.
44 To characterize the genetic diversity and niche of Hawaiian *C. elegans*, we sampled across the Hawaiian
45 Islands and isolated 100 new *C. elegans* strains. We found that *C. elegans* strains are not associated with
46 any one substrate but are found in cooler climates at high elevations. These Hawaiian strains are highly

47 diverged compared to the rest of the global population. Admixture analysis identified 11 global populations,
48 four of which are from Hawaii. Surprisingly, one of the Hawaiian populations shares recent ancestry with non-
49 Hawaiian populations, including portions of globally swept haplotypes. This discovery provides the first
50 evidence of gene flow between Hawaiian and non-Hawaiian populations. Most importantly, the high levels of
51 diversity observed in Hawaiian strains might represent the complex patterns of ancestral genetic diversity in
52 the *C. elegans* species before human influence.
53

54 **Introduction**

55 Over the last 50 years, the nematode *Caenorhabditis elegans* has been central to many important
56 discoveries in the fields of developmental, cellular, and molecular biology. The vast majority of these insights
57 came from the study of a single laboratory-adapted strain collected in Bristol, England known as N2 (Brenner,
58 1974; Chalfie et al., 1994; Consortium, 1998; Fire et al., 1998; Grishok et al., 2000; Hodgkin and Brenner,
59 1977; Lee et al., 1993; Sulston et al., 1983). Recent sampling efforts have led to the identification of numerous
60 wild *C. elegans* strains and enabled the study of genetic diversity and ecology of the species (Andersen et
61 al., 2012; Barrière and Félix, 2014; Cook et al., 2016; Félix and Duveau, 2012; Ferrari et al., 2017; Hahnel
62 et al., 2018; Lee et al., 2019; Richaud et al., 2018). The earliest studies of *C. elegans* genetic variation
63 showed that patterns of single-nucleotide variant (SNV) diversity were shared among most wild strains, with
64 the exception of a Hawaiian strain, CB4856, which has distinct and high levels of variation relative to other
65 strains (Koch et al., 2000). Subsequent analyses revealed that *C. elegans* has reduced levels of diversity
66 relative to the obligate outcrossing *Caenorhabditis* species and the facultative selfer *C. briggsae* (Dey et al.,
67 2013; Thomas et al., 2015). The most comprehensive analysis of *C. elegans* genetic diversity to date used
68 data from thousands of genome fragments across a globally distributed collection of 97 genetically distinct
69 strains to show that recent selective sweeps have largely homogenized the genome (Andersen et al., 2012).
70 The authors hypothesized that these selective sweeps might contain alleles that facilitate human-assisted
71 dispersal and/or increase fitness in human-associated habitats. Consistent with the previous analyses, two
72 Hawaiian strains, CB4856 and DL238, did not share patterns of reduced genetic diversity caused by the
73 selective sweeps that affected the rest of the *C. elegans* population – a trend that has held true as the number
74 of Hawaiian strains has increased (Cook et al., 2017, 2016; Hahnel et al., 2018; Lee et al., 2019). Taken
75 together, these studies suggest that the Hawaiian *C. elegans* population might be more representative of
76 ancestral genetic diversity that existed prior to the selective pressures associated with recent human
77 influence.

78 To better characterize the genetic diversity of the *C. elegans* species on the Hawaiian Islands, we
79 performed deep sampling across five Hawaiian islands: Kauai, Oahu, Molokai, Maui, and the Big Island.
80 Because incomplete data on locations and environmental parameters are common issues for some field
81 studies of *C. elegans* (Andersen et al., 2012; McGrath et al., 2009; Rockman and Kruglyak, 2009), we
82 developed a standardized collection procedure with the Fulcrum® mobile data collection application. This
83 streamlined procedure enabled us to rapidly record GPS coordinates and environmental niche parameters
84 at each collection site, and accurately link these data with the nematodes we isolated. The Hawaiian Islands
85 are an ideal location to study characteristics of the *C. elegans* niche because the Islands contain many steep,
86 wide-ranging gradients of temperature, humidity, elevation, and landscape usage. In total, we collected
87 samples from 2,263 sites across the islands and isolated 2,532 nematodes, including 309 individuals from
88 the *Caenorhabditis* genus. Among these isolates, we identified 100 new *C. elegans* strains, 95 of which
89 proliferated in the lab and were whole-genome sequenced. Analysis of genomic variation revealed that these
90 strains represent 26 distinct genome-wide haplotypes not sampled previously. We refer to these genome-
91 wide haplotypes as isotypes. We grouped these 26 Hawaiian isotypes with the 17 previously isolated
92 Hawaiian isotypes and compared their genetic variation to 233 non-Hawaiian isotypes from around the globe.
93 Consistent with previous observations, we found that the Hawaiian population has approximately three times
94 more diversity than the non-Hawaiian population. However, we were surprised to find that, in a subset of

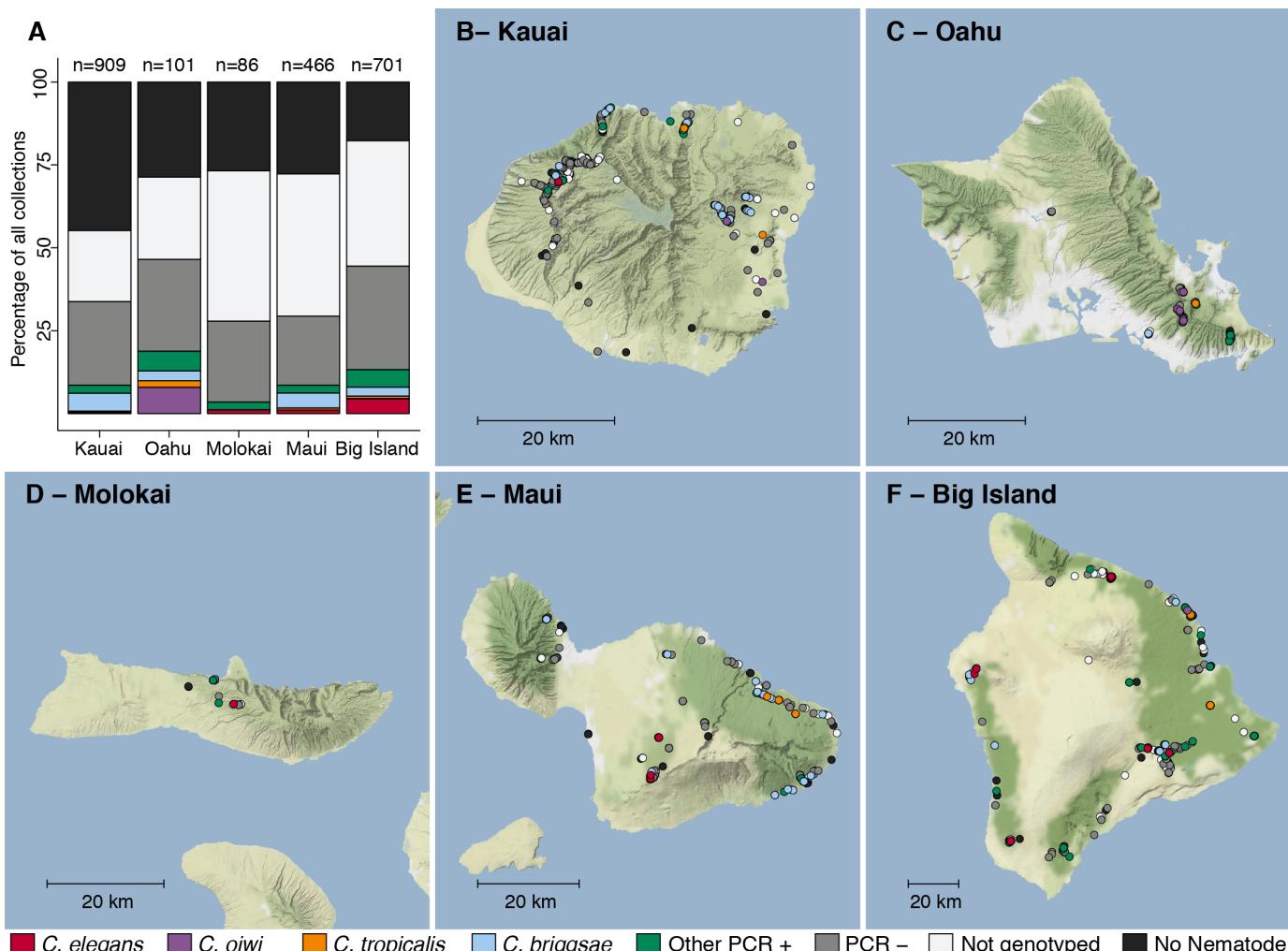
95 Hawaiian isotypes, some genomic regions appear to be shared with non-Hawaiian isotypes from around the
96 globe. These results provide the first evidence of gene flow between these populations and suggest that
97 future sampling efforts in the Hawaiian Islands will help elucidate the evolutionary processes that have
98 shaped the genetic diversity in the *C. elegans* species.
99

100 **Results**

101

102 **Hawaiian nematode diversity**

103 In August 2017, we collected a total of 2,263 samples across five Hawaiian islands and ascertained the
104 presence of nematodes in each sample (**Figure 1, Supplemental Table 5**). We isolated one or more
105 nematodes from 1,120 of 2,263 (48%) samples, and an additional 431 of 2,263 (19%) samples had
106 circumstantial evidence of nematodes (tracks but no nematodes could be found on the collection plate).
107 Altogether, we isolated 2,531 nematodes from 1,120 samples and genotyped them by analysis of the Internal
108 Transcribed Spacer (ITS2) region between the 5.8S and 28S rDNA genes (Barrière and Félix, 2014; Kiontke
109 et al., 2011). We refer to isolates where the ITS2 region was amplified by PCR as ‘PCR-positive’ and isolates
110 with no amplification as ‘PCR-negative’ (see Methods). The PCR-positive category comprises
111 *Caenorhabditis* isolates that we identified to the species level and isolates from genera other than
112 *Caenorhabditis* that we identified to the genus level. Using this categorization strategy, we found that 427 of
113 2,531 isolates (17%) were PCR-positive and belonged to 13 distinct taxa. Among all isolates, we identified
114 five *Caenorhabditis* species at different frequencies across the 2,263 samples: *C. briggsae* (4.2%),
115 *C. elegans* (1.7%), *C. tropicalis* (0.57%), *C. kamaaina* (0.088%), and a new species *C. oiwi* (0.53%)
116 (**Supplemental Table 5**). We named *Caenorhabditis oiwi* for the Hawaiian word meaning “native” in
117 reference to its endemic status on the Hawaiian Islands. This species was found to be distinct based on
118 molecular barcodes (Kiontke et al., 2011) and on biological species inference from mating crosses (Félix et
119 al., 2014) (**Supplemental File 1**). The most common *Caenorhabditis* species we isolated was *C. briggsae*,
120 which is consistent with nematode collection efforts by other groups that suggest *C. briggsae* is a ubiquitous
121 species in many regions of the world (Félix et al., 2013). We found no evidence of island enrichment for
122 *Caenorhabditis* species apart from *C. elegans*, where it was enriched on the Big Island relative to Kauai and
123 Maui (Fisher’s Exact Test, $p < 0.01$).



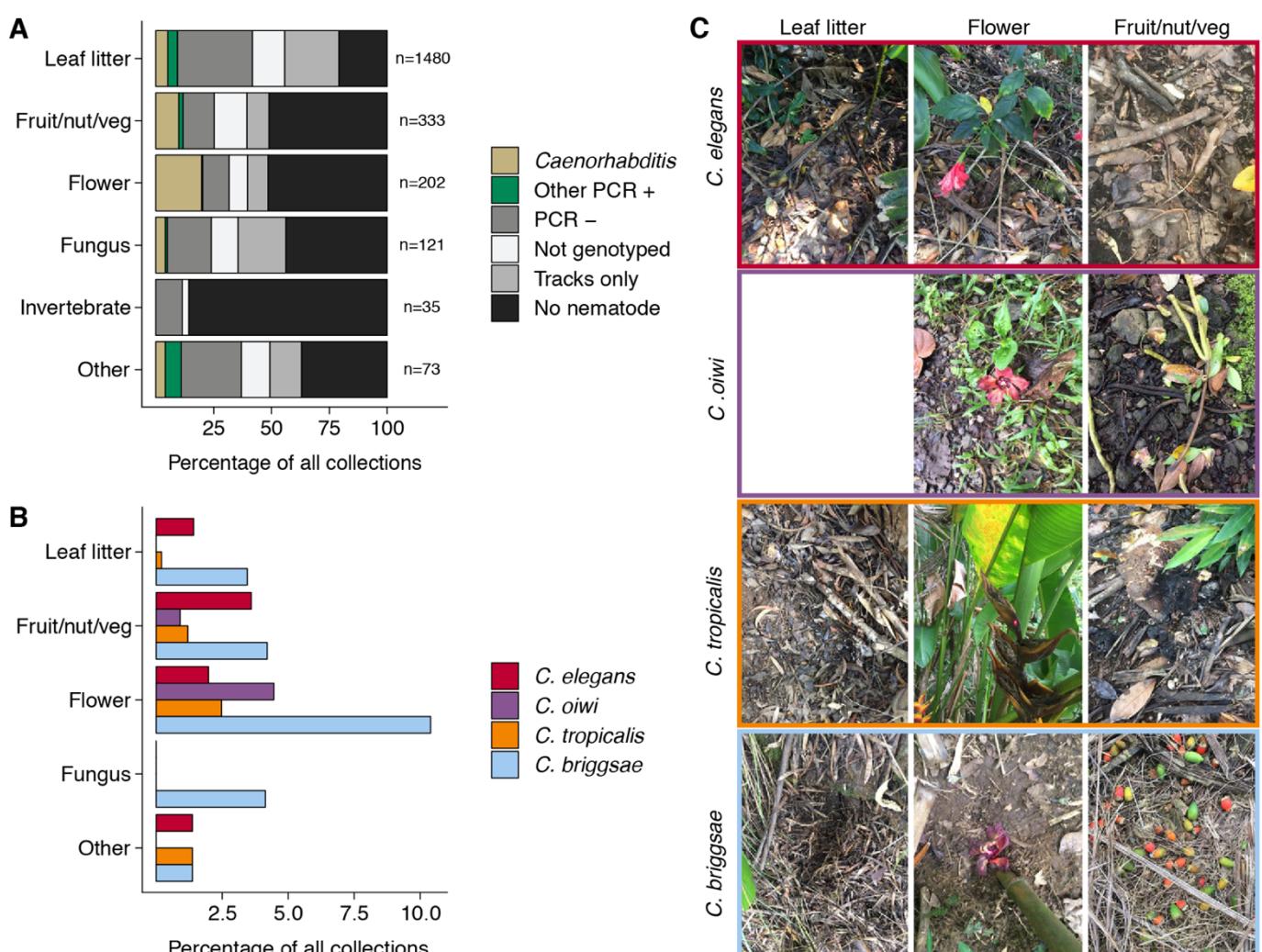
124
125 **Figure 1 - Geographic distribution of sampling sites across five Hawaiian islands.** In total we sampled 2,263
126 unique sampling sites. (A) The percentage of each collection category is shown by island. The collection categories are
127 colored according to the legend at the bottom of the panel, and the total number of samples for each island are shown
128 above the bars. (B-F) The circles indicate unique sampling sites ($n = 2,263$) and are colored by the collection categories
129 shown in the bottom legend. For sampling sites where multiple collection categories apply ($n = 299$), the site is colored
130 by the collection category shown in the legend from left to right, respectively. For all sampling sites, the GPS coordinates
131 and collection categories found at that site are included in (Supplemental Data 1). We focused our studies on
132 Caenorhabditis nematode collections, excluding *C. kamaania* because it was only found at two sampling sites. Maps ©
133 www.thunderforest.com, Data © www.osm.org/copyright.

134

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

136 To characterize more about a nematode niche on the Hawaiian Islands, we classified the substrate for each
137 distinct collection and measured various environmental parameters. Of the six major classes of substrate,
138 we found nematodes most often on leaf litter (56%). When we account for collections with nematode-like
139 tracks on the collection plate, we estimated that greater than 80% of leaf litter substrates contained
140 nematodes (Figure 2A). The isolation success rate for the other classes of substrate ranged from 35% to
141 48% (Figure 2A). In comparison to overall nematode isolation rates, *Caenorhabditis* nematodes were
142 isolated more frequently from flower substrates (40 of 202 collections) than any other substrate category
143 (Fisher's Exact Test, $p < 0.02$) (Figure 2A). We also found that *Caenorhabditis* nematodes were enriched
144 on rotting fruits, nuts, or vegetables (33 of 333 collections) relative to leaf litter substrates (76 of 1480
145 collections) (Fisher's Exact Test, $p < 0.02$) but not other substrate classes (Figure 2A). These findings are
146 consistent with other collection surveys that have shown leaf litter substrates harbor fewer *Caenorhabditis*
147 nematodes than rotting flowers and fruits (Félix et al., 2013; Ferrari et al., 2017). We observed similar trends

148 of flower-substrate enrichment relative to leaf litter for *C. briggsae* (Fisher's Exact Test, $p = 0.00049$; flower,
149 21 of 202 collections and leaf litter 51 of 1480 collections) and *C. tropicalis* (Fisher's Exact Test, $p = 0.0059$;
150 flower, five of 202 collections and leaf litter, three of 1480 collections) but not for *C. elegans* (Fisher's Exact
151 Test, $p = 1$), which exhibited no substrate enrichment (**Figure 2B-C**). Interestingly, the new species, *C. oiwi*,
152 was only isolated from flower and fruit/nut/vegetable substrates and was enriched on flower substrates
153 (Fisher's Exact Test, $p = 0.0124$; flower, nine of 202 collections and fruit/nut/vegetable, three of 333
154 collections) (**Figure 2B-C**).
155



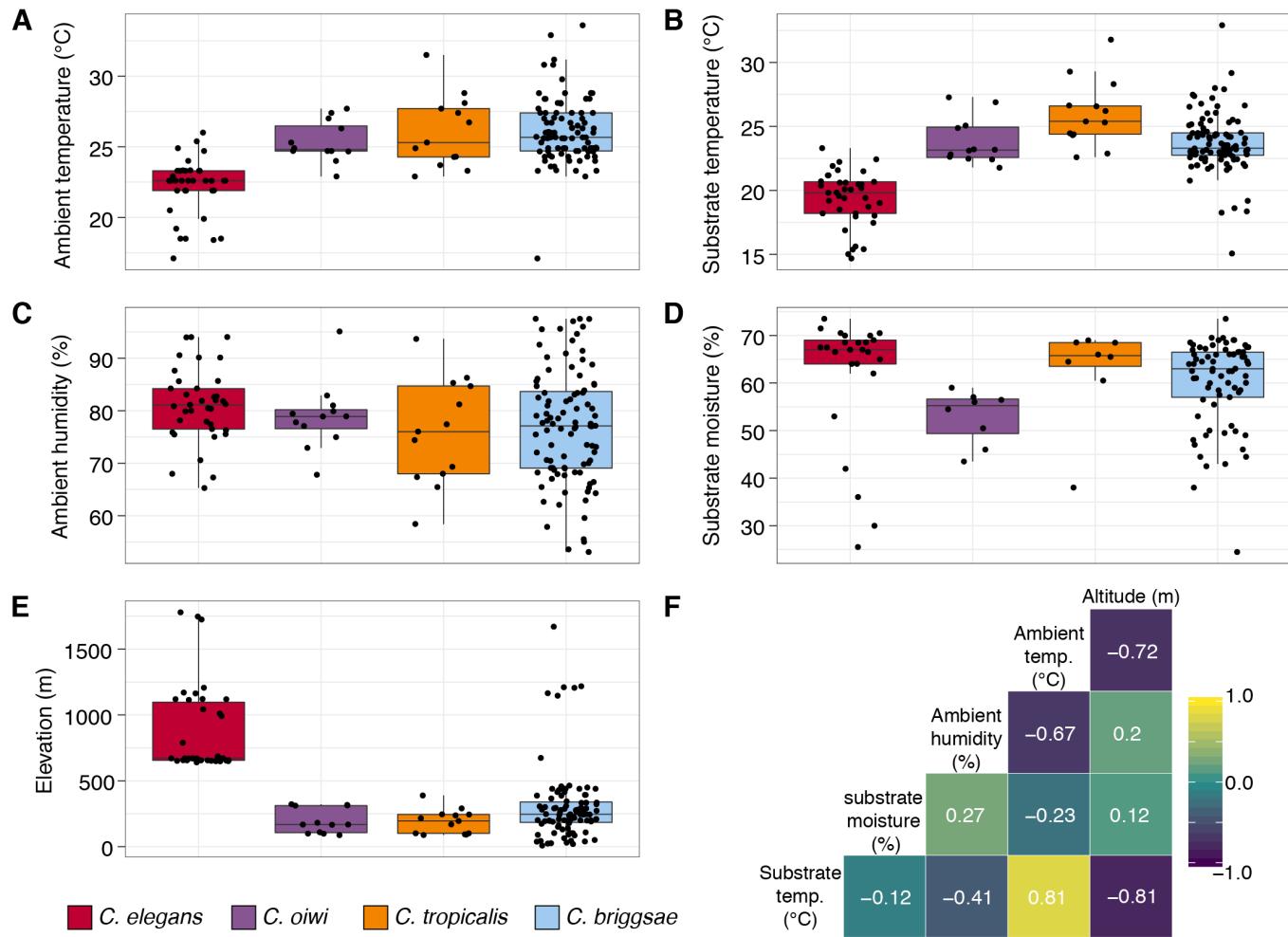
156 **Figure 2 - Collection categories by substrate type.** (A) The percentage of each collection category is shown by
157 substrate type. The collection categories are colored according to the legend at the right, and the total number of
158 samples for each substrate are shown to the right of bars. (B) The percentage of collections is shown by substrate type
159 for each *Caenorhabditis* species (excluding *C. kamaaina*, $n = 2$). (C) Examples of substrate photographs for
160 *Caenorhabditis* species are shown. The *C. oiwi* leaf litter cell is blank because *C. oiwi* was only isolated from flowers
161 and fruit.
162

163 The enrichment of *C. briggsae*, *C. tropicalis*, and *C. oiwi* on flowers might indicate that this substrate class
164 has a higher nutrient quality for these species. If this hypothesis is correct, we might expect to see a greater
165 incidence of proliferating populations on flower substrates than other substrates. However, we saw no
166 observable association between large population size (approximate number of nematodes on collection
167 plate) and substrate class for *C. briggsae* (Spearman's $\rho = -0.0197$, $p = 0.57$ flower vs. leaf litter),
168 *C. tropicalis* (Spearman's $\rho = -0.26$, $p = 0.73$ flower vs. leaf litter), nor *C. oiwi* (Spearman's $\rho = 0.258$, p
169 = 0.21 flower vs. fruit/nut/vegetable), which suggests that other factors might drive the observed flower

170 enrichment or that we are limited by the small sample size. Taken together, these data suggest that the
171 *Caenorhabditis* species we isolated do not exhibit substrate specificity, despite flower-substrate preferences
172 of *C. briggsae*, *C. tropicalis*, and *C. oiwai*, which is different from some other species in the genus that
173 demonstrate substrate specificity (e.g., *C. astrocarya* and *C. inopinata*) (Ferrari et al., 2017; Kanzaki et al.,
174 2018).

175

176 In addition to recording substrate classes, we measured elevation, ambient temperature and humidity, and
177 substrate temperature and moisture to determine if these niche parameters were important for individual
178 *Caenorhabditis* species (**Figure 3**; see Methods).



179
180 **Figure 3 - Environmental parameter values for sites where *Caenorhabditis* species were isolated.** (A-E) Tukey
181 box plots are plotted by species (colors) for different environmental parameters. Each dot corresponds to a unique
182 sampling site where that species was identified. In cases where two *Caenorhabditis* species were identified from the
183 same sample ($n = 3$), the same parameter values are plotted for both species. All p-values were calculated using Kuskal-
184 Wallis test and Dunn test for multiple comparisons with p values adjusted using the Bonferroni method; comparisons
185 not mentioned were not significant ($\alpha = 0.05$). (A) Ambient temperature (°C) was typically cooler at the sites where
186 *C. elegans* were isolated compared to sites for all other *Caenorhabditis* species (Dunn test, $p < 0.005$). (B) Substrate
187 temperature (°C) was also generally cooler for *C. elegans* than all other *Caenorhabditis* species (Dunn test, $p <$
188 0.00001). (C) Ambient humidity (%) did not differ significantly among the *Caenorhabditis*-positive sites. (D) Substrate
189 moisture (%) was generally greater for *C. elegans* than *C. oiwai* (Dunn test, $p = 0.002$). (E) Elevation (meters) was
190 typically greater at sites where *C. elegans* were isolated compared to sites for all other *Caenorhabditis* species (Dunn
191 test, $p < 0.00001$). (F) A correlation matrix for the environmental parameters was made using sample data from the
192 *Caenorhabditis* species shown. The parameter labels for the matrix are printed on the diagonal, and the pearson
193 correlation coefficients are printed in the cells. The color scale also indicates the strength and sign of the correlations
194 shown in the matrix.

195 Consistent with previous *C. elegans* collections in tropical regions (Andersen et al., 2012; Dolgin et al., 2008),
196 all *C. elegans* isolates were collected from elevations greater than 500 meters and were generally found at
197 higher elevations than other *Caenorhabditis* species (**Figure 3E**; mean = 867 m; elevation: Dunn test, $p <$
198 0.00001). We also found that *C. elegans*-positive collections tended to be at cooler ambient and substrate
199 temperatures than other *Caenorhabditis* species (ambient temperature: Dunn test, $p < 0.005$; substrate
200 temperature: Dunn test, $p < 0.00001$), although these two environmental parameters were correlated with
201 elevation (**Figure 3F**). Notably, the average substrate temperatures for *C. elegans* (19.4 °C), *C. tropicalis*
202 (26.0 °C), and *C. briggsae* (23.7°C) positive collections are close to the optimal growth temperatures for
203 these species in the laboratory setting (**Figure 3B**) (Poulet et al., 2015). Our collections also indicate that
204 *C. oiwi* tends to be found on drier substrates than *C. elegans* (**Figure 3D**; Dunn test, $p = 0.0021$), but we
205 observed no differences among species for ambient humidity (**Figure 3C**). Given the similar substrate and
206 environmental parameter preferences of *C. tropicalis*, *C. briggsae*, and *C. oiwi*, we next asked if these
207 species colocalized at either the local (< 30 m²) or substrate (< 10 cm²) scales. To sample at the local scale,
208 we collected samples from 20 gridsects (see Methods; **Supplemental Figure 1**) and observed no
209 colocalization of these three species, although only 16% of the total collections were a part of a gridsect. At
210 the substrate scale, we found *C. tropicalis* and *C. briggsae* cohabitating on two of 108 substrates with either
211 species present and *C. oiwi* and *C. briggsae* cohabitating on one of 107 substrates with either species
212 present (**Supplemental Figure 2**). Among 95 substrates with *C. briggsae*, we observed nine instances of *C.*
213 *briggsae* cohabitating with other PCR-positive species. We did not collect any samples that harbored
214 *C. elegans* and any other *Caenorhabditis* species. Taken together, these cohabitation results highlight the
215 ubiquitous nature of *C. briggsae* on the Hawaiian Islands and further suggests that the niche of *C. elegans*
216 might be distinct from *C. tropicalis*, *C. briggsae*, and *C. oiwi* on the Hawaiian Islands.
217

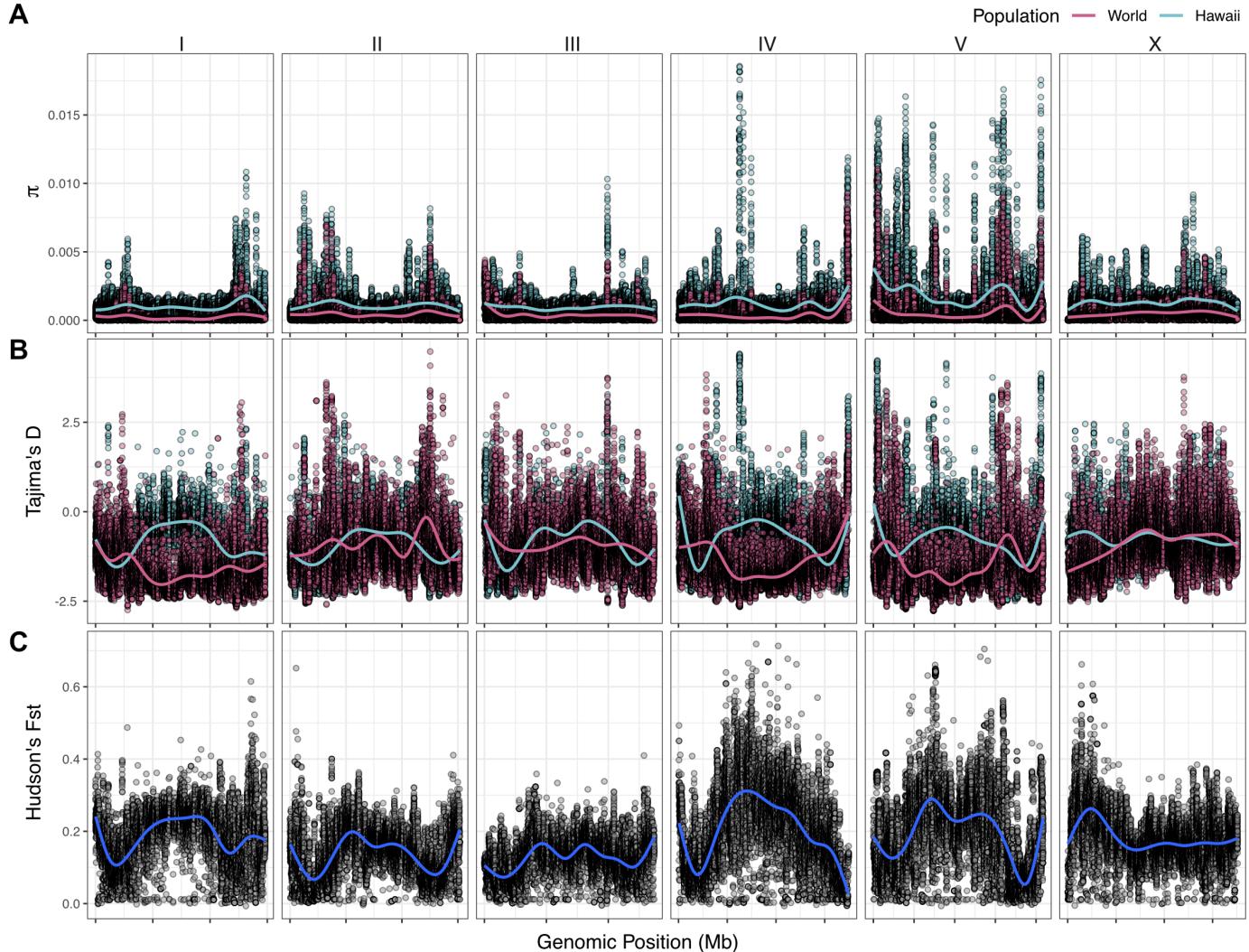
218 Hawaiian *C. elegans* are divergent from the global population

219 We previously showed that two *C. elegans* isolates from Hawaii are highly divergent relative to wild isolates
220 from other regions of the world and represent a large portion of the genetic diversity found within the species
221 (Andersen et al., 2012). Since this analysis, an additional 15 isolates have been collected from the islands
222 and show similarly high levels of genetic diversity (Cook et al., 2016; Hahnel et al., 2018). To better
223 characterize the genetic diversity in Hawaii, we acquired whole-genome sequence data from 95 *C. elegans*
224 isolates that we collected in this study. By analyzing the variant composition of these 95 isolates, we identified
225 26 distinct genome-wide haplotypes that we refer to as isotypes (see Methods). Within these 26 isotypes,
226 we identified approximately 1.54 million single nucleotide variants (SNVs) that passed our filtering strategy
227 (see Methods; hard-filter VCF; **Supplemental Table 4**), which is 27.6% greater than the total number of
228 SNVs identified in all of the 233 non-Hawaiian isotypes included in this study. We found that distinct isotypes
229 are frequently isolated within close proximity to one another in Hawaii. We identified up to seven unique
230 isotypes colocalized within a single gridsect (less than 30 m²) (**Supplemental Figure 3A**). We also found that
231 colocalization occurred at the substrate level; among the 38 substrates from which we isolated *C. elegans*,
232 12 contained two or more isotypes (**Supplemental Figure 3B**). The variant data from all 43 Hawaiian
233 isotypes (26 new with 17 previously described Hawaiian isotypes) allowed us to perform detailed analyses
234 of Hawaiian genetic diversity.
235

236 Consistent with what is known about the *C. elegans* global population (Andersen et al., 2012), we observed
237 a high degree of genome-wide relatedness among a majority of non-Hawaiian isotypes (**Supplemental**
238 **Figure 4**). By contrast, the Hawaiian isotypes are all diverged from the non-Hawaiian population with the
239 exception of five non-Hawaiian isotypes. Among these exceptions, ECA36 and QX1211 were collected from
240 urban gardens in New Zealand and San Francisco, CA respectively, and grouped with some of the most
241 divergent isotypes from Hawaii. More surprisingly, three non-Pacific Rim isotypes also grouped with the
242 Hawaiian isotypes. These include JU2879, MY16, and MY23. JU2879 was isolated from a rotting apple in

243 Mexico City, Mexico and both MY isotypes were isolated from garden composts in Nordrhein-Westfalen,
244 Germany, separated by approximately 5 km. Within the Hawaiian population, genome-wide relatedness
245 revealed a high degree of divergence (**Supplemental Figure 4**). This trend is further supported by elevated
246 levels of genome-wide average nucleotide diversity (π) in the Hawaiian population relative to the non-
247 Hawaiian population, which we found to be three-fold higher (Hawaii $\pi = 0.00124$; non-Hawaiian $\pi =$
248 0.000408, **Figure 4A; Supplemental Data 2**).

A



249

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

256

257 The genomic distribution of diversity followed a similar pattern across chromosomes for both populations,
258 wherein chromosome centers and tips exhibited lower diversity on average than chromosome arms (**Figure**
259 **4A; Supplemental Data 2**). This pattern is likely explained by lower recombination rates, higher gene
260 densities, and elevated levels of background selection on chromosome centers (Consortium, 1998; Cutter
261 and Payseur, 2003; Rockman et al., 2010). Interestingly, we observed discrete peaks of diversity in specific
262 genomic regions (e.g., chr IV center), which suggests that balancing selection might maintain diversity at
263 these loci in both populations (**Figure 4A; Supplemental Data 2**). This hypothesis is supported by
264 corresponding spikes in Tajima's D (**Figure 4B; Supplemental Data 3**) (Tajima, 1989). Alternatively, higher

265 values of Tajima's D might indicate a population contraction, but the discrete nature of these peaks makes
266 this possibility less likely. A third possible explanation is that uncharacterized structural variation (e.g.,
267 duplication and divergence) exists in these regions. Nevertheless, the variant sites within these discrete
268 peaks in π and Tajima's D are unlikely the result of sequencing errors because they are identified across
269 multiple samples (see Methods).

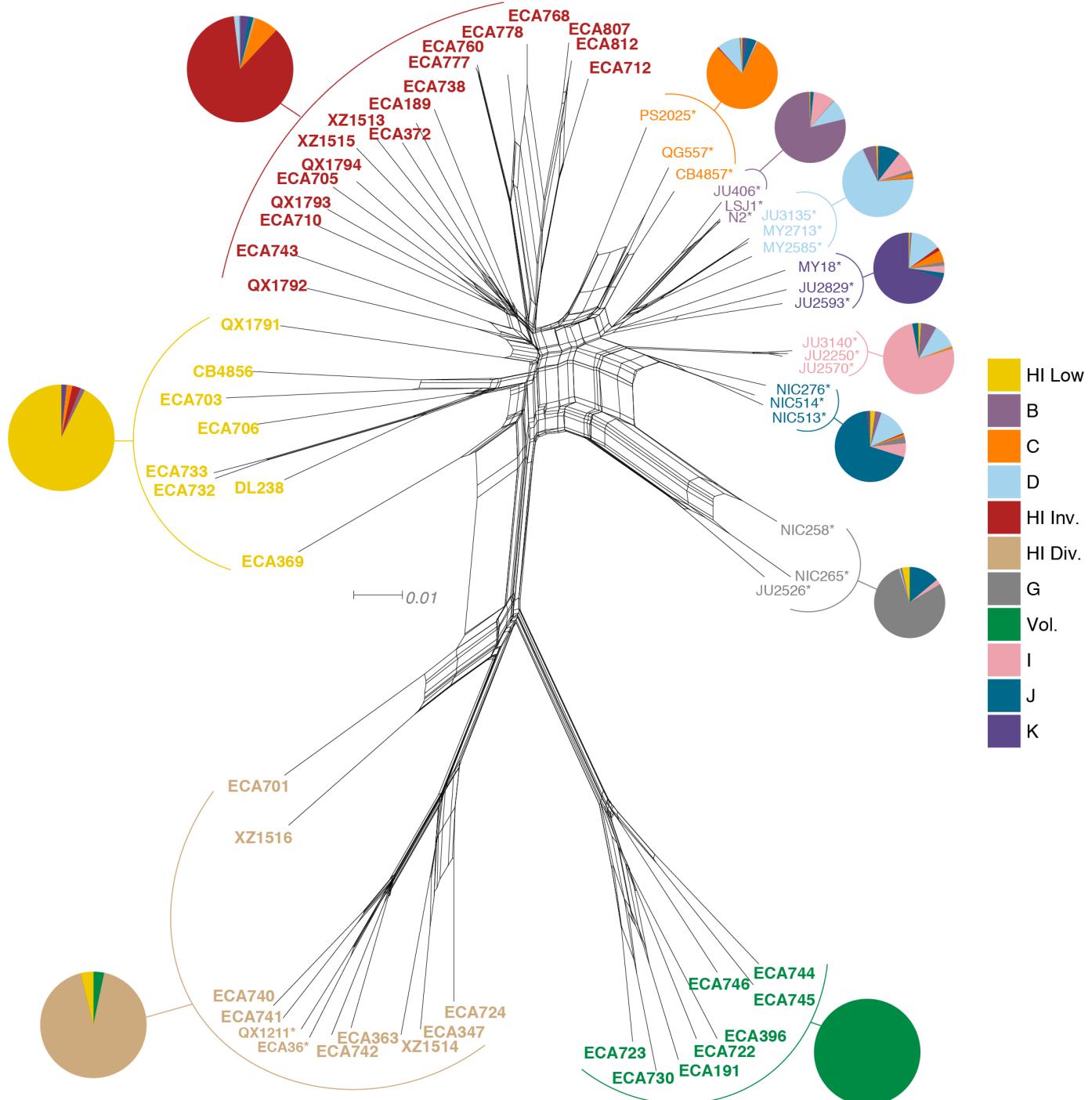
270

271 Our previous analysis showed that 70–90% of isotypes contain reduced levels of diversity across several
272 megabases (Mb) on chromosomes I, IV, V, and X (Andersen et al., 2012). This reduced diversity was
273 hypothesized to be caused by selective sweeps that occurred within the last few hundred years, potentially
274 through drastic alterations of global environments by humans. The two Hawaiian isotypes, CB4856 and
275 DL238, did not share this pattern of reduced diversity, suggesting that they avoided the selective pressure.
276 Consistent with this previous analysis, we did not observe signatures of selection in the Hawaiian population
277 on chromosomes I, IV, V, and X, as measured by Tajima's D (**Figure 4B; Supplemental Data 3**), which
278 suggests that the Hawaiian and non-Hawaiian populations have distinct evolutionary histories. This
279 distinction is also captured in genome-wide Hudson's F_{ST} , where the divergence between the two populations
280 is highest in regions of the genome impacted by the selective sweeps (**Figure 4C; Supplemental Data 2**)
281 (Bhatia et al., 2013; Hudson et al., 1992). Taken together, these data suggest that the Hawaiian population
282 has largely been isolated from the selective pressures thought to be associated with human activity in many
283 regions of the world.

284

285 **C. elegans** population structure on Hawaii

286 To assess population structure among all 276 isotypes, we performed admixture analysis (see Methods).
287 This analysis suggested that the *C. elegans* species is composed of at least 11 ancestral populations (K), as
288 indicated by the minimization of cross-validation (CV) error between Ks 11-15 (**Supplemental Figure 5**).
289 The population assignments for K=11 closely aligned to the relatedness clusters we observed in a neighbor-
290 joining network of all Hawaiian strains and the species-wide tree (**Figure 5, Supplemental Figure 4**). For
291 Ks 11-15, the majority of Hawaiian isotypes consistently exhibit no admixture with non-Hawaiian ancestral
292 populations. However, a minority of Hawaiian isotypes are consistently either admixed with non-Hawaiian
293 populations (e.g. K=11, 14, and 15) or assigned to ancestral populations that contain non-Hawaiian isotypes
294 (e.g. K=12 and 13) (**Supplemental Figure 5**). These data support that a subset of Hawaiian isotypes are
295 consistently shown to exhibit a greater degree of genetic relatedness with non-Hawaiian isotypes across
296 different population subdivisions. Together, we found at least four distinct subpopulations on the Hawaiian
297 Islands and at least seven additional non-Hawaiian subpopulations comprise the remainder of
298 subpopulations from around the globe (**Supplemental Figure 6**).



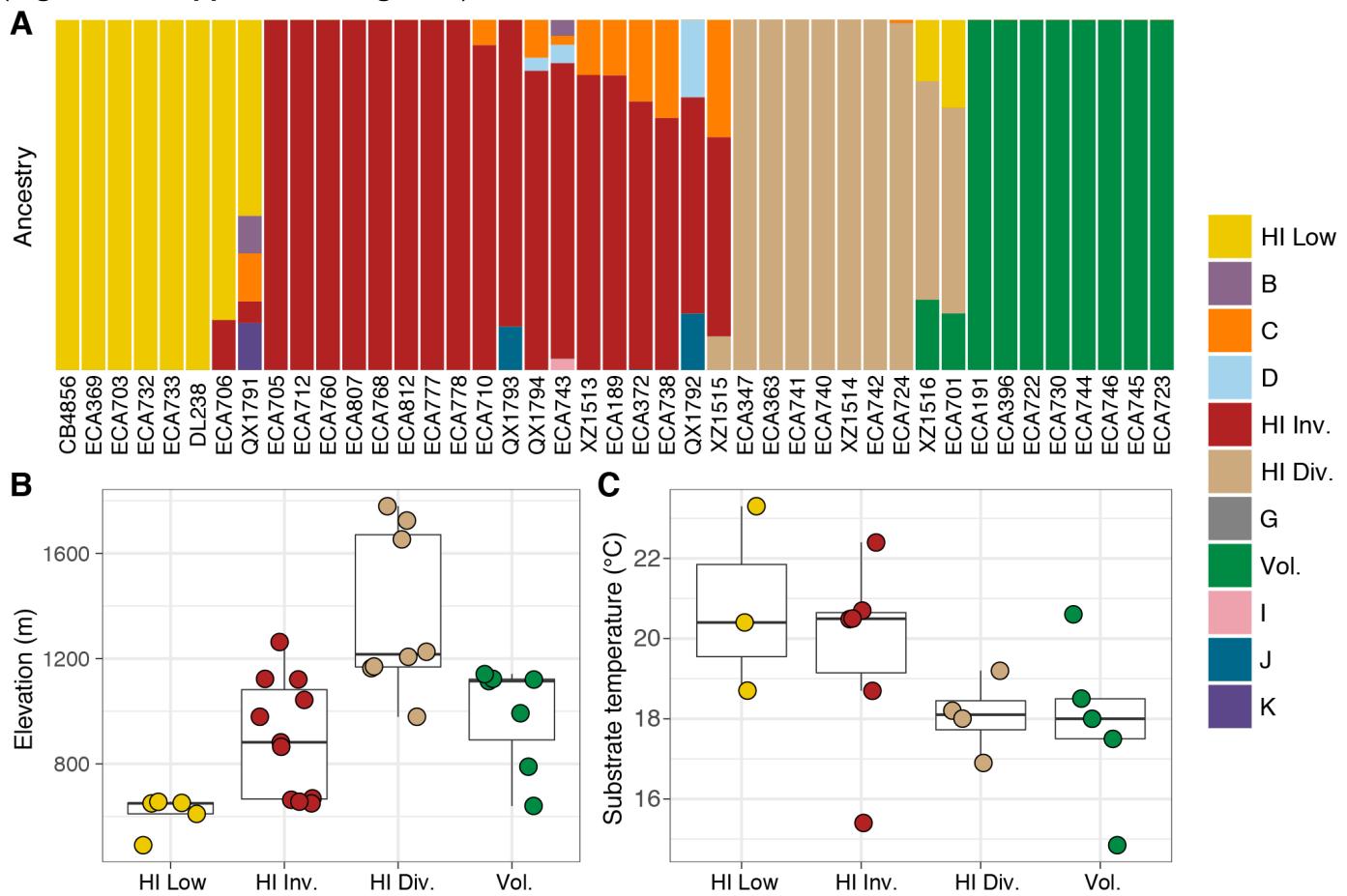
299

300 **Figure 5 - Relatedness of the Hawaiian *C. elegans* isolates.** Neighbor-joining net showing the genetic relatedness
301 of the Hawaiian *C. elegans* population relative to a representative set of non-admixed, non-Hawaiian individuals from
302 each population defined by ADMIXTURE (K=11). Colors of labels indicate the ancestral population assignment from
303 ADMIXTURE (K=11), including the seven global populations (B-K) and the four Hawaiian populations: Hawaiian
304 Invaded, Hawaiian Low, Hawaiian Divergent, and Volcano. Isotypes labeled with an asterisk are representative of non-
305 admixed, non-Hawaiian isolates from each population defined by ADMIXTURE (K=11). Pie charts represent ancestral
306 population proportions for all isolates within the full admixture population.

307

308 The majority of isolates assigned to the seven non-Hawaiian ancestral populations exhibit a high degree of
309 admixture with one another (at K=11), indicating that these populations are not well differentiated. By
310 contrast, isolates assigned to three of the four Hawaiian ancestral populations showed almost no admixture.
311 We refer to the four Hawaiian populations as Volcano, Hawaiian Divergent, Hawaiian Invaded, and Hawaiian

312 Low for the following reasons. All eight isotypes in the Volcano population were isolated on the Big Island of
313 Hawaii at high elevation in wet rainforests primarily composed of ferns, ‘Ōhi‘a lehua, and koa trees. We chose
314 to name this population ‘Volcano’ because the majority of isotypes were isolated from the town of Volcano.
315 The Hawaiian Divergent population is named for the two highly divergent isotypes, XZ1516 and ECA701,
316 which were isolated from Kauai, the oldest Hawaiian island sampled. However, we emphasize that the
317 population assignment of these two highly divergent isotypes might not be correct given that they each
318 contain many unique variants that were filtered from the admixture analysis. The Hawaiian Invaded
319 population is named because many of the isotypes assigned to this population exhibited admixture with non-
320 Hawaiian ancestral populations, which is suggestive of an invasion of non-Hawaiian alleles into Hawaii
321 (**Figure 6A, Supplemental Figure 7**).

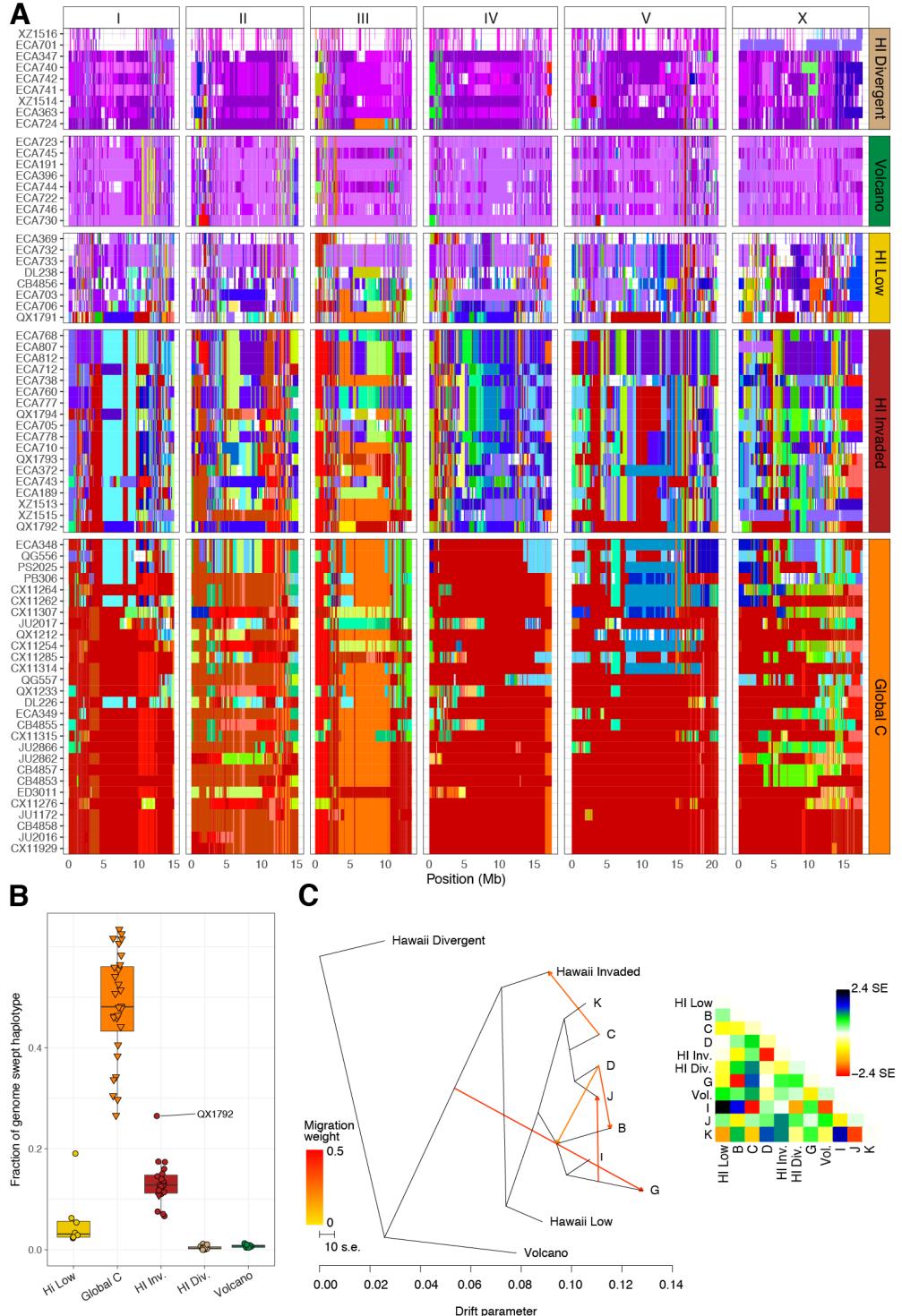


322
323 **Figure 6 - Environmental parameters of Hawaiian *C. elegans* isotypes.** (A) The inferred ancestral population
324 fractions for each Hawaiian isotype as estimated by ADMIXTURE (K=11; run on the entire *C. elegans* population) are
325 shown. The bar colors represent the ADMIXTURE population assignment for the isotypes named on the x-axis. (B-C)
326 Tukey box plots are shown by ADMIXTURE population assignments (colors) for different environmental parameters.
327 We used the average values of environmental parameters from geographically clustered collections to avoid biasing
328 our results by local oversampling (See Methods - Environmental parameter analysis). All p-values were calculated using
329 Kuskal-Wallis test and Dunn test for multiple comparisons with p values adjusted using the Bonferroni method;
330 comparisons not mentioned were not significant ($\alpha = 0.05$). (B) The collection site elevations for Hawaiian isotypes
331 colored by the ADMIXTURE population assignments are shown. The Hawaiian Low and the Hawaiian Invaded
332 populations were typically found at lower elevations than the Hawaiian Divergent population (Dunn test, p -values =
333 0.000168, and 0.037 respectively). (C) The substrate temperatures for Hawaiian isotypes colored by the ADMIXTURE
334 population assignments are shown.
335
336 The Hawaiian Low population is named because isotypes assigned to this population tended to be isolated
337 at lower elevations than those assigned to the other Hawaiian populations (See Methods, **Figure 6B**). The

338 population structure of the Hawaiian isotypes suggests that geographic associations within the Hawaiian *C.*
339 *elegans* population exist either by elevation or by island.
340

341 Within the Hawaiian Invaded population, one of the 19 isotypes was isolated from outside of Hawaii (MY23),
342 and 11 of 18 Hawaiian isotypes showed admixture with various non-Hawaiian populations, particularly the
343 non-Hawaiian population C (**Figure 6, Supplemental Figure 6**). By contrast, just one individual assigned to
344 the global C population was admixed with the Hawaiian Invaded population (**Supplemental Figure 6**). This
345 result suggested that these populations either share ancestry or recent gene flow occurred between them.
346 To distinguish between these possibilities, we explicitly tested for the presence of gene flow among all
347 subpopulations using TreeMix (Pickrell and Pritchard, 2012), which estimates the historical relationships
348 among populations accounting for both population splits and migration events. We found evidence of gene
349 flow between the Hawaiian Invaded population and the non-Hawaiian population C (**Figure 7C; Supplemental Figure 8**). The topological position of the fourth highest-weight migration event identified by
350 TreeMix (*i.e.*, C→Hawaiian Invaded) suggested that the evidence of gene flow is not caused by incomplete
351 assortment of ancestral alleles (*i.e.*, the migration arrows connect the ‘C’ and Hawaiian Invaded lineages at
352 the branch tips) (**Figure 7C; Supplemental Figure 8**). Importantly, TreeMix cannot distinguish the direction
353 of migration between these subpopulations.
354

355
356 To further assess evidence of gene flow between the Hawaiian and globally distributed subpopulations, we
357 analyzed the haplotype structure across the genomes of all 276 *C. elegans* isotypes (Browning and
358 Browning, 2016). Within the Hawaiian Divergent, Volcano, and Hawaiian Low populations, we observed
359 haplotypes that were largely absent from the non-Hawaiian isotypes. By contrast, the Hawaiian Invaded
360 population shared haplotypes that were commonly found in non-Hawaiian isotypes assigned to non-Hawaiian
361 populations. For example, the isotypes in the Hawaiian Invaded population exhibiting admixture with the
362 global C population share haplotype arrangements on the left and center of Chr III (red and orange Chr III,
363 **Figure 7A**). We also found evidence of the globally swept haplotype in all of the isotypes from the Hawaiian
364 Invaded population, particularly on chromosomes I, V, and X, but less so on chromosome IV (**Figure 7B, Supplemental Figure 9**). By contrast, greater than 50% of chromosome IV contained the swept haplotype
365 in all of the isotypes from the global C population (**Supplemental Figure 9**). Taken together, our data showed
366 that the Hawaiian isotypes from the Volcano, Hawaiian Divergent, and Hawaiian Low populations have
367 avoided the selective sweeps that are pervasive across most regions of the globe, and individuals within the
368 Hawaiian Invaded subpopulation have likely been outcrossed with these swept haplotypes.
369



370

371 **Figure 7 - Evidence of migration between the Hawaiian and world populations.** (A) The inferred blocks of identity
372 by descent (IBD) across the genome are shown. The genomic position is plotted on the x-axis for each isotype plotted
373 on the y-axis. The block colors correspond to a uniquely defined IBD group. The dark red blocks correspond to the most
374 common global haplotype (i.e., the swept haplotype on chr I, IV, V, and left of X). Genomic regions with no color
375 represent regions for which no IBD groups could be determined. The four Hawaiian populations are shown in the top
376 four facets excluding non-Hawaiian isotypes. The bottom facet shows the global C population. (B) The total fraction of
377 the genome with the swept haplotype is shown by ancestral population. The data points correspond to isotypes and are
378 colored by their assigned ancestral populations. The Hawaiian isotypes are plotted as circles and non-Hawaiian isotypes
379 are plotted as triangles. Hawaiian isotypes with greater than 25% of their genome swept are labelled. (C) The inferred
380 relationship among the ancestral populations allowing for five migration events (ADMIXTURE, K=11). The heat map to
381 the right represents the residual fit to the migration model.

382 Discussion

383 We sought to deeply sample the natural genetic variation within the *C. elegans* species to better understand
384 the evolutionary history and driving forces of genome evolution in this powerful model system. Because the
385 Hawaiian Islands have been shown to harbor highly divergent strains relative to most regions of the world,
386 we choose to sample extensively on these islands. We developed a streamlined collection procedure that
387 facilitated our collection of over 2,000 samples across five Hawaiian islands. From these collections, we
388 isolated over 2,500 nematodes and used molecular data to partition 427 of these isolates into 13 distinct
389 taxa, mostly from the *Rhabditidae* family. In total, we identified and cryogenically preserved 95 new
390 *C. elegans* isolates that represent 26 genetically distinct isotypes. These isotypes represent the largest single
391 *C. elegans* collection effort on any island system and contain 27% more SNVs than all 233 non-Hawaiian
392 isotypes combined. Our findings confirm high diversity in Hawaii matching previous studies (Andersen et al.,
393 2012; Wicks et al., 2001). Furthermore, we document the first evidence of outcrossing between Hawaiian
394 and global populations.
395

396 The origins of *C. elegans*

397 The higher genetic diversity in the Hawaii population might indicate that it represents an ancient population,
398 similar to African populations in humans (Nielsen et al., 2017; Ramachandran et al., 2005). The possibility
399 that the *C. elegans* species might have originated from the Hawaiian Islands, or migrated there from adjacent
400 landmasses shortly after speciation, requires that the Hawaiian Islands predate the split between *C. elegans*
401 and its closest known relative *C. inopinata*, which is estimated to be 10.5 million years (Kanzaki et al., 2018).
402 The extant Hawaii Islands we sampled range in age from the still-forming Big Island to the 5.1 million year
403 old Kauai, but the now submerged Emperor Seamounts represent approximately 70 million years of stable
404 land masses over the Pacific Hotspot (Neall and Trewick, 2008). Therefore, older land masses might have
405 donated colonists to younger islands maintaining the Hawaiian *C. elegans* populations over millions of years
406 and allowing the accumulation of genetic diversity. The higher genetic diversity in the Hawaiian Islands may
407 also be driven by population demography on the Islands. It is possible that Hawaii harbors larger, more
408 temporally stable effective population sizes than other regions of the world that have been sampled. Under
409 a neutral model, populations with a larger effective population size are expected to have a greater number
410 of neutral polymorphisms (Kimura, 1991). These larger, more stable effective population sizes are plausible
411 in Hawaii given the abundant supply of available habitat, e.g. rotting fruits and vegetable matter, and stable
412 temperatures throughout the year. The Hawaiian climate is particularly less variable than many temperate
413 regions where *C. elegans* populations must overwinter and are known to exhibit seasonal population
414 expansions and contractions (Frézal and Félix, 2015), and tend to be dominated by highly related genotypes
415 from year to year (Richaud et al., 2018). Ultimately, the pattern of genetic variation in Hawaiian populations
416 is likely influenced by a combination of demographic history (e.g., changes in population size, short- and
417 long-range migration events, and admixture) as well as evolutionary processes such as natural selection,
418 recombination, and mutation. To further untangle the evolutionary history of this species, additional samples
419 from natural areas around the globe and in particular the Pacific Rim will be required.
420

421 Out of Hawaii or invasion of Hawaii?

422 Our data support outcrossing between the Hawaiian Invaded population and the less-diverse global C
423 population. Moreover, most strains from the Hawaiian Invaded and non-Hawaiian populations share portions
424 of the globally swept haplotype. Within the Hawaiian Invaded population, isotypes share smaller portions of
425 the swept haplotype relative to isotypes from the non-Hawaiian populations. It remains unclear whether
426 sharing of globally swept haplotypes can be explained by emigration of nematodes from Hawaii (out of
427 Hawaii) or immigration of nematodes to Hawaii (invasion of Hawaii). In either case, the Hawaiian Islands are
428 geographically isolated, which should theoretically restrict gene flow to and from the Islands. However,
429 Hawaii's position as a global trade-hub makes gene flow with the rest of the world more likely (Frankham,

430 1997). Although we do not have direct evidence to discriminate between these possibilities, the ‘Out of
431 Hawaii’ hypothesis might have occurred through long-range dispersal of genotypes similar to those found in
432 the Hawaiian Invaded population, which then underwent selection over multiple generations to resemble the
433 more swept genotypes found across the globe. Migration out of Hawaii could have been aided by the
434 transition of the Hawaiian economy towards large-scale production and export of sugarcane and tropical
435 fruits, which began in the late nineteenth century (Bartholomew et al., 2012). If correct, then this situation is
436 similar to what is thought to have occurred within *Drosophila melanogaster* where the fruit trade might have
437 facilitated recent migrations from native regions to oceanic islands (David and Capy, 1988; Hales et al.,
438 2015). Alternatively, the pattern of haplotype sharing could be explained by an ‘invasion of Hawaii’ scenario,
439 wherein swept haplotypes have invaded Hawaii. This scenario could threaten the genetic diversity of the
440 Hawaiian populations if the invading alleles confer strong fitness advantages as is expected for swept
441 haplotypes (Andersen et al., 2012). However, if an invasion of Hawaii is currently underway, we have little
442 evidence to support the selection of the globally swept haplotypes in Hawaii. First, the Hawaiian Invaded
443 population only contains small fractions of the swept haplotypes on chromosomes I, V, X, and even smaller
444 fractions on chromosome IV. Second, it would take a considerable number of generations to create the small
445 fractions of the swept haplotypes that we observe in the Hawaiian Invaded population because of the low
446 outcrossing rates and high incidence of outbreeding depression in *C. elegans* (Dolgin et al., 2007).

447

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

449 We used a publicly available weather data from the National Oceanic and Atmospheric Administration and
450 the National Climatic Data Center to measure the variation in seasonal temperatures for locations close to
451 the sites where isotypes were collected (Evans et al., 2017). We found that the Hawaiian populations
452 experienced less seasonal variability in temperature than any of the non-Hawaiian populations
453 (**Supplemental Figure 10**). These findings raise the possibility that the ancestral niche of *C. elegans* might
454 be similar to the thermally stable Hawaiian habitats where genetic diversity is highest. However, factors other
455 than seasonal temperature variation might also characterize the ancestral niche of *C. elegans*. The Hawaiian
456 Divergent population was enriched at higher elevation, which has been less impacted by human activities in
457 Hawaii since the time of Polynesian colonization (Alison Kay, 1994). By contrast, the Hawaiian Invaded
458 population is found at lower elevations. Although it remains unclear what factors restrict gene flow between
459 the non-admixed and Hawaiian Invaded populations, it is possible that selective pressures associated with
460 human impact contribute to their isolation. This possibility would be consistent with the hypothesis that the
461 global sweeps, present in the Hawaiian Invaded population, originated through positive selection acting on
462 loci that confer fitness advantages in human-associated habitats (Andersen et al., 2012). Taken together, we
463 suspect that the ancestral niche of *C. elegans* is likely to be similar to the thermally stable, high elevation
464 Hawaiian habitats where human impacts are less prevalent.

465

466 **Unravelling the evolutionary history of *C. elegans***

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

478

479 Methods

480

481 Strains

482 Nematodes were reared at 20°C using OP50 bacteria grown on modified nematode growth medium (NGMA),
483 containing 1% agar and 0.7% agarose to prevent animals from burrowing (Andersen et al., 2014). In total,
484 169 *C. briggsae*, 100 *C. elegans*, 21 *C. tropicalis*, 15 *C. oiwi*, and four *C. kamaaina* wild isolates were
485 collected. Of these strains, 95 *C. elegans*, 19 *C. tropicalis*, and 12 *C. oiwi* wild isolates were cryopreserved
486 and are available upon request along with the other *C. elegans* strains included in our analysis
487 (**Supplemental File 2**). The type specimen for *C. oiwi* (ECA1100) is also deposited at the *Caenorhabditis*
488 Genetics Center (**Supplemental File 1**).

489

490 Sampling strategy

491 We sampled nematodes at 2,263 sites across five Hawaiian Islands during August 2017. Before travelling to
492 Hawaii, general sampling locations were selected based on accessibility via hiking trails and by proximity to
493 where *C. elegans* had been collected previously (Andersen et al., 2012; Cook et al., 2016; Hahnel et al.,
494 2018; Hodgkin and Doniach, 1997). Sampling hikes with large elevation changes were prioritized to ensure
495 that we sampled across a broad range of environmental parameters. On these hikes, we opportunistically
496 sampled substrates known to harbor *C. elegans*, including fruits, nuts, flowers, stems, leaf litter, compost,
497 soil, wood, and live arthropods and molluscs (Ferrari et al., 2017; Frézal and Félix, 2015; Schulenburg and
498 Félix, 2017). In 20 locations, we performed extensive local sampling in an approximately 30 square meter
499 area that we refer to as a 'gridsect'. The gridsect comprised a center sampling point with additional sampling
500 sites at one, two, and three meters away from the center in six directions with each direction 60° apart from
501 each other (**Supplemental Figure 1**).

502

503 Field sampling and environmental data collection

504 To characterize the *Caenorhabditis* abiotic niche, we collected and organized data for several environmental
505 parameters at each sampling site using a customizable geographic data-collection application called
506 Fulcrum®. We named our customized Fulcrum® application 'Nematode field sampling' and used the
507 following workflow to enter the environmental data into the application while in the field. First, we used a
508 mobile device camera to scan a unique collection barcode from a pre-labelled plastic collection bag. This
509 barcode is referred to as a collection label or 'C-label' in the application and is used to associate a particular
510 sample with its environmental and nematode isolation data. Next, we entered the substrate type, landscape,
511 and sky view data into the application using drop down menus and photographed the sample in place using
512 a mobile device camera. The GPS coordinates for the sample are automatically recorded in the photo
513 metadata. We then measured the surface temperature of the sample using an infrared thermometer
514 Lasergrip 1080 (Etekcity, Anaheim, CA), its moisture content using a handheld pin-type wood moisture meter
515 MD912 (Dr. Meter, Los Angeles, CA), and the ambient temperature and humidity near the sample using a
516 combined thermometer and hygrometer device GM1362 (GoerTek, Weifang, China). These measurements
517 were entered into the appropriate fields in the application (**Supplemental Table 3**). Finally, we transferred
518 the sample into a collection bag and stored it in a cool location before we attempted to isolate nematodes.
519 Seventy samples in our raw data had missing GPS coordinates or GPS coordinates that were distant from
520 actual sampling locations after visual inspection using satellite imagery. The positions for these samples
521 were corrected using the average position of the two samples collected before and after the errant data point
522 or by manually assigning estimated positions.

523

524 Nematode isolation

525 Following each collection, the substrate sample was transferred from the barcoded collection bag to an
526 identically barcoded 10 cm NGMA plate seeded with OP50 bacteria. For 1,989 of the 2,263 samples
527 collected, we isolated nematodes that crawled off the substrates onto the collection plates approximately 47
528 hours after the samples were collected from the field (mean = 46.9 h, std. dev. = 19.5 h). The remaining 274
529 samples were shipped overnight from Hawaii to Northwestern University in collection bags, and the
530 nematodes were isolated approximately 172 hours after sample collection (mean = 172.5 h, std. dev. = 17.9
531 h). For each collection plate, up to seven gravid nematodes were isolated by transferring them individually
532 to pre-labeled 3.5 cm NGMA isolation plates seeded with OP50 bacteria. We refer to these isolation plates
533 as 'S-plates' in the Fulcrum® application we called 'Nematode isolation' (**Supplemental Table 4**). At the time
534 of isolation, we recorded the approximate number of nematodes on the collection plate and whether males
535 or dauers were present. Importantly, male and dauer observations from samples shipped from Hawaii were
536 not recorded to avoid bias caused by the long handling time of these samples. We merged the collection,
537 isolation, and environmental data together into a single data file with the 'process_fulcrum_data.R' script that
538 can be found in the scripts folder of the GitHub repo (https://github.com/AndersenLab/Hawaii_Manuscript)
539 (**Supplemental Data 4**).

540

541 Nematode identification

542 The isolated nematodes were stored at 20°C for approximately 14 days (mean = 14.3 d, std. Dev. = 4.9 d)
543 but were not passaged during this time to avoid multiple generations of proliferation. For initial genotyping,
544 five to ten nematodes were lysed in 8 µl of lysis solution (100 mM KCl, 20 mM Tris pH 8.2, 5 mM MgCl₂,
545 0.9% IGEPAL, 0.9% Tween 20, 0.02% gelatin with proteinase K added to a final concentration of 0.4 mg/ml)
546 then frozen at -80°C for up to 12 hours. The lysed material was thawed on ice, and 1 µl was loaded directly
547 into 40 µl reactions with primers spanning a portion of the ITS2 region (Internal Transcribed Spacer) between
548 the 5.8S and 28S rDNA genes with forward primer oECA305 (GCTGCGTTATTACCAACGAATTGCARAC)
549 and reverse primer oECA202 (GCGGTATTGCTACTACCAYYAMGATCTGC) (Kiontke et al., 2011). The
550 PCR used the following conditions: three minutes denaturation step at 95°C; then 34 cycles of 95°C for 15
551 seconds, 60°C for 15 seconds, and 72°C for two minutes; followed by a five-minute elongation step at 72°C.
552 The presence of ITS2 PCR products was visualized on a 2% agarose gel in 1X TAE buffer. Isolates that did
553 not yield an ITS2 PCR product were labelled as 'PCR-negative', and those reactions that yielded the
554 expected 2 kb ITS2 PCR product were labelled as 'PCR-positive'. We then used Sanger sequencing to
555 sequence the ITS2 PCR products with forward primer oECA305. We classified *Caenorhabditis* species by
556 comparing the ITS2 sequences to the National Center for Biotechnology Information (NCBI) database using
557 the BLAST algorithm. Isolates with sequences that aligned best to genera other than *Caenorhabditis* were
558 only classified to the genus level. For every isolate where the BLAST results either aligned to *C. elegans*,
559 had an unexpectedly high number of mismatches in the center of the read, or did not match any known
560 sequences because of poor sequence quality, we performed another independent lysis and PCR using high-
561 quality Taq polymerase (cat# RR001C, TaKaRa) to confirm our original results. For this confirmation, we
562 used the forward primer oECA305 and the reverse primer oECA306 (CACTTCAAGCAACCCGAC) to
563 sequence the confirmation ITS2 amplicon in both directions. The sequence chromatograms were then quality
564 trimmed by eye with Unipro UGENE software (version 1.27.0) and compared to known nematode species in
565 the NCBI sequence database using the BLAST algorithm. We used the consensus alignment of the forward
566 and reverse reads to confirm our original results. For *C. elegans*, five of the 100 strains perished before we
567 could confirm their identity. We also confirmed that several strains that best aligned to *C. kamaaina* shared
568 a large number of mismatches in the center of the ITS2 amplicon, suggesting they belonged to a new species.
569 For these strains, we performed reciprocal mating tests with *C. kamaaina* to infer the new species by the
570 biological species concept (Félix et al., 2014). None of these crosses produced viable progeny, suggesting
571 that these isolates represent a new *Caenorhabditis* species (**Supplemental File 1**).

572

573 Illumina library construction and whole-genome sequencing

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

586

587 Variant calling

588 To ensure reproducible data analysis, all genomic analyses were performed using pipelines generated in the
589 Nextflow workflow management system framework (Di Tommaso et al., 2017). Each Nextflow pipeline used
590 in this study is briefly described below (**Supplemental Table 5**). All pipelines follow the “*pipeline name-nf*”
591 naming convention and full descriptions can be found on the Andersen lab dry-guide website:
592 (<http://andersenlab.org/dry-guide/pipeline-overview/>).

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

603 After isotypes are assigned, we used *alignment-nf* with BWA (v0.7.17-r1188) (Li, 2013; Li and Durbin,
604 2009) to align trimmed sequence data for distinct isotypes to the N2 reference genome (WS245) (Lee et al.,
605 2018). Next, we called SNVs using *wi-nf*, which uses the BCFtools (v.1.9) (Danecek et al., 2014). The *wi-nf*
606 pipeline generates two population-wide VCFs that we refer to as the soft-filtered and hard-filtered VCFs
607 (**Supplemental Table 2**). After variant calling, a soft-filtered VCF was generated for each sample by
608 appending the following soft-filters to variant sites: Depth (FORMAT/DP) > 10; Mapping Quality (INFO/MQ)
609 > 40; Variant quality (QUAL) > 10; (Allelic Depth (FORMAT/AD) / Number of high quality bases
610 (FORMAT/DP)) ratio > 0.5. These soft-filters were appended to the FT field of the VCF using *VCF-kit* (Cook
611 and Andersen, 2017). Next, sample VCFs were merged using the merge utility of BCFtools. Once the
612 population VCF was generated, variant sites with greater than 90% missing genotypes (high_missing)
613 or greater than 10% heterozygosity (high_heterozygosity) were flagged. We refer to this VCF as the soft-filtered
614 VCF. To construct the hard-filtered VCF, we removed all variants that did not pass the filters described above.
615 Both the soft- and hard-filtered isotype-level VCFs are available to download on the CeNDR website (version
616 20180527) (Cook et al., 2017).

617 We further pruned the hard-filtered VCF to contain sites with no missing genotype calls and removed
618 sites in high linkage disequilibrium (LD) using PLINK (v1.9) (Chang et al., 2015; Purcell et al., 2007) with the
619 *--indep-pairwise 50 1 0.95* command. The predicted variant effects were appended to the VCF using SnpEff
620 (v 4.3) (Cingolani et al., 2012). We further annotated this VCF with exons, G-quartets, transcription factor

621 binding sites, histone binding sites, miRNA binding sites, splice sites, ancestral alleles (XZ1516 set as
622 ancestor), the genetic map position, and repetitive elements using vcfanno (v 0.2.8) (Pedersen et al., 2016).
623 All annotations were obtained from WS266. We removed regions that were annotated as repetitive. We
624 named this VCF the ‘PopGen VCF’ (**Supplemental Data 4; Supplemental Table 2**).
625

626 Phylogenetic analyses

627 We characterized the relatedness of the *C. elegans* population using RAxML-ng with the GTR DNA
628 substitution model and maximum likelihood estimation to find the parameter values that maximize the
629 phylogenetic likelihood function, and thus provide the best explanation for the observed data (Kozlov et al.,
630 2019). We used the vcf2phylip.py script (Ortiz, n.d.) to convert the ‘PopGen VCF’ (**Supplemental Data 4**) to
631 the PHYLIP format (Felsenstein, 1993) required to run RAxML-ng. To construct the tree that included 276
632 strains, we used the GTR evolutionary model available in RAxML-ng (Lanave et al., 1984; Tavaré, 1986).
633 Trees were visualized using the ggtree (v1.10.5) R package (Yu et al., 2017). To construct the neighbor-net
634 phylogeny, we used SplitsTree4 (Huson and Bryant, 2006).
635

636 Population genetic statistics

637 Genome-wide pi, Hudson’s F_{ST} , and Tajima’s D were calculated using the PopGenome package in R (Pfeifer
638 et al., 2014). All statistics were calculated along sliding windows with a 10 kb window size and a 1 kb step
639 size.
640

641 Admixture analysis

642 We performed admixture analysis using ADMIXTURE (v1.3.0) (Alexander et al., 2009). Prior to running
643 ADMIXTURE, we LD-pruned the ‘PopGen VCF’ (**Supplemental Data 4**) using PLINK (v1.9) (Chang et al.,
644 2015; Purcell et al., 2007) with the command `--indep-pairwise 50 10 0.8`. We also removed variants only
645 present in one isotype. We ran ADMIXTURE ten independent times for K sizes ranging from 2 to 20 for all
646 276 isotypes. Visualization of admixture results was performed using the pophelper (v2.2.5) R package
647 (Francis, 2017). We chose K=11 for future analyses because the cross-validation (CV) error approached
648 minimization at this K (**Supplemental Figure 5**). Furthermore, K=11 subset the Hawaiian isotypes into four
649 distinct populations, which exactly matched the subsets obtained from running ADMIXTURE on just the 43
650 Hawaiian isotypes at K=4 (K=4 minimized CV for ADMIXTURE with Hawaiian isotypes only, (**Supplemental**
651 **Figure 11**). We performed TreeMix analysis on K=11 for zero to five migration events (Pickrell and Pritchard,
652 2012).
653

654 Haplotype analysis

655 We determined identity-by-descent (IBD) of strains using IBDSeq (Browning and Browning, 2013) run on the
656 ‘PopGen VCF’ (**Supplemental Data 4**) with the following parameters: `minalleles=0.01, ibdtrim=0, r2max=0.8`.
657 IBD segments were then used to infer haplotype structure among isotypes as described previously (Andersen
658 et al., 2012). After haplotypes were identified, we defined the most common haplotype found on
659 chromosomes I, IV, V, and X as the swept haplotype. We then retained the swept haplotypes within isotypes
660 that passed the following per chromosome filters: total length > 1 Mb; total length / maximum population-
661 wide swept haplotype length > 0.03. We classified chromosomes within isotypes as swept if the sum of the
662 retained swept haplotypes for a chromosome was > 3% of the maximum population wide swept haplotype
663 length for that chromosome.
664

665 Environmental parameter analysis

666 We calculated the pairwise distances among all *C. elegans*-positive collections on Hawaii and detected five
667 distinct geographic clusters, each of which contain collections that are within 20 meters of one another. The
668 largest of these clusters comprised 18 collections in the Kalopa State Recreation Area on the Big Island of

669 Hawaii. This cluster contained 11 collections from gridsect-3 and seven additional collections within 20
670 meters from the edge of the gridsect. The other four geographic clusters contain four or fewer collections
671 each. We used the average values of environmental parameters from geographically clustered collections to
672 avoid biasing our results by local oversampling. We applied this strategy to the comparison of environmental
673 parameters between the Hawaiian admixture populations and used the Kuskal-Wallis test to detect
674 differences ($\alpha = 0.05$).
675

676 Acknowledgements

677 We thank the members of the Andersen lab for editing the manuscript for flow and content and for making
678 reagents used in the experiments presented. We are grateful to landowners who gave us permission to
679 collect nematodes on their property. We also thank individuals who have helped us collect additional strains.
680 We would also like to thank the Hawaii Department of Land and Natural Resources as well as the Natural
681 Area Reserves System for permitting, support for these studies, and general advice about the Hawaiian
682 Islands. Additionally, Dr. Sam Gon from The Nature Conservancy Hawai'i Program helped with the naming
683 of *Caenorhabditis oiwi*. This research was supported by start-up funds from Weinberg College of Arts and
684 Sciences and the Molecular Biosciences department. KK is supported by NSF DEB 0922012 to D. H. A.
685 Fitch.
686

687 Competing interests

688 The authors declare no conflicts of interest.
689

690 References

- 691 Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated
692 individuals. *Genome Res* **19**:1655–1664.
693 Alison Kay E. 1994. A Natural History of the Hawaiian Islands: Selected Readings II. University of Hawaii
694 Press.
695 Andersen EC, Bloom JS, Gerke JP, Kruglyak L. 2014. A variant in the neuropeptide receptor npr-1 is a
696 major determinant of *Caenorhabditis elegans* growth and physiology. *PLoS Genet* **10**:e1004156.
697 Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, Félix M-A, Kruglyak L. 2012.
698 Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat Genet*
699 **44**:285–290.
700 Barrière A, Félix M-A. 2014. Isolation of *C. elegans* and related nematodes. *WormBook* 1–19.
701 Bartholomew DP, Hawkins RA, Lopez JA. 2012. Hawaii Pineapple: The Rise and Fall of an Industry.
702 *HortScience* **47**:1390–1398.
703 Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting FST: the impact of
704 rare variants. *Genome Res* **23**:1514–1521.
705 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data.
706 *Bioinformatics* **30**:2114–2120.
707 Brenner S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* **77**:71–94.
708 Browning BL, Browning SR. 2016. Genotype Imputation with Millions of Reference Samples. *Am J Hum
709 Genet* **98**:116–126.
710 Browning BL, Browning SR. 2013. Detecting Identity by Descent and Estimating Genotype Error Rates in
711 Sequence Data. *Am J Hum Genet* **93**:840–851.
712 Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. 1994. Green fluorescent protein as a marker for
713 gene expression. *Science* **263**:802–805.
714 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to
715 the challenge of larger and richer datasets. *Gigascience* **4**:7.
716 Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program
717 for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the
718 genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**:80–92.
719 Consortium TCES. 1998. Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating

- 720 Biology. *Science* 1–8.
- 721 Cook DE, Andersen EC. 2017. VCF-kit: assorted utilities for the variant call format. *Bioinformatics*
722 **33**:1581–1582.
- 723 Cook DE, Zdraljevic S, Roberts JP, Andersen EC. 2017. CeNDR, the *Caenorhabditis elegans* natural
724 diversity resource. *Nucleic Acids Res* **45**:D650–D657.
- 725 Cook DE, Zdraljevic S, Tanny RE, Seo B, Riccardi DD, Noble LM, Rockman MV, Alkema MJ, Braendle C,
726 Kammenga JE, Wang J, Kruglyak L, Félix M-A, Lee J, Andersen EC. 2016. The Genetic Basis of
727 Natural Variation in *Caenorhabditis elegans* Telomere Length. *Genetics genetics*. 116.191148.
- 728 Cutter AD, Payseur BA. 2003. Selection at linked sites in the partial selfer *Caenorhabditis elegans*. *Mol Biol
729 Evol* **20**:665–673.
- 730 Danecek P, Schiffels S, Durbin R. 2014. Multiallelic calling model in bcftools (-m) 10–11.
- 731 David JR, Capy P. 1988. Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet
732 4*:106–111.
- 733 Dey A, Chan CKW, Thomas CG, Cutter AD. 2013. Molecular hyperdiversity defines populations of the
734 nematode *Caenorhabditis brenneri*. *Proc Natl Acad Sci U S A* **110**:11056–11060.
- 735 Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. 2017. Nextflow enables
736 reproducible computational workflows. *Nat Biotechnol* **35**:316–319.
- 737 Dolgin ES, Charlesworth B, Baird SE, Cutter AD. 2007. Inbreeding and outbreeding depression in
738 *Caenorhabditis* nematodes. *Evolution* **61**:1339–1352.
- 739 Dolgin ES, Félix M-A, Cutter AD. 2008. Hakuna Nematoda: genetic and phenotypic diversity in African
740 isolates of *Caenorhabditis elegans* and *C. briggsae*. *Heredity* **100**:304–315.
- 741 Evans KS, Zhao Y, Brady SC, Long L, McGrath PT, Andersen EC. 2017. Correlations of Genotype with
742 Climate Parameters Suggest *Caenorhabditis elegans* Niche Adaptations. *G3* **7**:289–298.
- 743 Félix M-A, Braendle C, Cutter AD. 2014. A streamlined system for species diagnosis in *Caenorhabditis*
744 (Nematoda: Rhabditidae) with name designations for 15 distinct biological species. *PLoS One
745 9*:e94723.
- 746 Félix M-A, Duveau F. 2012. Population dynamics and habitat sharing of natural populations of
747 *Caenorhabditis elegans* and *C. briggsae*. *BMC Biol* **10**:59.
- 748 Félix M-A, Jovelin R, Ferrari C, Han S, Cho YR, Andersen EC, Cutter AD, Braendle C. 2013. Species
749 richness, distribution and genetic diversity of *Caenorhabditis* nematodes in a remote tropical rainforest.
750 *BMC Evol Biol* **13**:10.
- 751 Felsenstein J. 1993. Phylogeny inference package. *Department of Genetics, University of Washington*.
- 752 Ferrari C, Salle R, Callemeyn-Torre N, Jovelin R, Cutter AD, Braendle C. 2017. Ephemeral-habitat
753 colonization and neotropical species richness of *Caenorhabditis* nematodes. *BMC Ecol* **17**:43.
- 754 Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic
755 interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**:806–811.
- 756 Francis RM. 2017. pophelper: an R package and web app to analyse and visualize population structure.
757 *Mol Ecol Resour* **17**:27–32.
- 758 Frankham R. 1997. Do island populations have less genetic variation than mainland populations? *Heredity
759 78 (Pt 3)*:311–327.
- 760 Frézal L, Félix M-A. 2015. *C. elegans* outside the Petri dish. *eLife* **4**. doi:10.7554/eLife.05849
- 761 Grishok A, Tabara H, Mello CC. 2000. Genetic requirements for inheritance of RNAi in *C. elegans*. *Science
762 287*:2494–2497.
- 763 Hahnel SR, Zdraljevic S, Rodriguez BC, Zhao Y, McGrath PT, Andersen EC. 2018. Extreme allelic
764 heterogeneity at a *Caenorhabditis elegans* beta-tubulin locus explains natural resistance to
765 benzimidazoles. *bioRxiv*. doi:10.1101/372623
- 766 Hales KG, Korey CA, Larracuente AM, Roberts DM. 2015. Genetics on the Fly: A Primer on the *Drosophila*
767 Model System. *Genetics* **201**:815–842.
- 768 Hodgkin JA, Brenner S. 1977. Mutations causing transformation of sexual phenotype in the nematode
769 *Caenorhabditis elegans*. *Genetics* **86**:275–287.
- 770 Hodgkin J, Doniach T. 1997. Natural variation and copulatory plug formation in *Caenorhabditis elegans*.
771 *Genetics* **146**:149–164.
- 772 Huang R-E, Ren X, Qiu Y, Zhao Z. 2014. Description of *Caenorhabditis sinica* sp. n. (Nematoda:
773 Rhabditidae), a nematode species used in comparative biology for *C. elegans*. *PLoS One* **9**:e110957.
- 774 Hudson RR, Boos DD, Kaplan NL. 1992. A statistical test for detecting geographic subdivision. *Mol Biol*

- 775 *Evol* **9**:138–151.
- 776 Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*
777 **23**:254–267.
- 778 Kanzaki N, Tsai IJ, Tanaka R, Hunt VL, Liu D, Tsuyama K, Maeda Y, Namai S, Kumagai R, Tracey A,
779 Holroyd N, Doyle SR, Woodruff GC, Murase K, Kitazume H, Chai C, Akagi A, Panda O, Ke H-M,
780 Schroeder FC, Wang J, Berriman M, Sternberg PW, Sugimoto A, Kikuchi T. 2018. Biology and
781 genome of a newly discovered sibling species of *Caenorhabditis elegans*. *Nat Commun* **9**:3216.
- 782 Kimura M. 1991. The neutral theory of molecular evolution: a review of recent evidence. *Jpn J Genet*
783 **66**:367–386.
- 784 Kiontke KC, Félix M-A, Ailion M, Rockman MV, Braendle C, Pénigault J-B, Fitch DHA. 2011. A phylogeny
785 and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. *BMC Evol
786 Biol* **11**:339.
- 787 Koch R, van Luenen HG, van der Horst M, Thijssen KL, Plasterk RH. 2000. Single nucleotide
788 polymorphisms in wild isolates of *Caenorhabditis elegans*. *Genome Res* **10**:1690–1696.
- 789 Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: A fast, scalable, and user-
790 friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*.
791 doi:10.1093/bioinformatics/btz305
- 792 Lanave C, Preparata G, Saccone C, Serio G. 1984. A new method for calculating evolutionary substitution
793 rates. *J Mol Evol* **20**:86–93.
- 794 Lee D, Zdraljevic S, Cook DE, Frézal L, Hsu J-C, Sterken MG, Riksen JAG, Wang J, Kammenga JE,
795 Braendle C, Félix M-A, Schroeder FC, Andersen EC. 2019. Selection and gene flow shape niche-
796 associated copy-number variation of pheromone receptor genes. *bioRxiv*. doi:10.1101/580803
- 797 Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs
798 with antisense complementarity to lin-14. *Cell* **75**:843–854.
- 799 Lee RYN, Howe KL, Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Davis P, Gao S, Grove C, Kishore
800 R, Muller H-M, Nakamura C, Nuin P, Paulini M, Raciti D, Rodgers F, Russell M, Schindelman G, Tuli
801 MA, Van Auken K, Wang Q, Williams G, Wright A, Yook K, Berriman M, Kersey P, Schedl T, Stein L,
802 Sternberg PW. 2018. WormBase 2017: molting into a new stage. *Nucleic Acids Res* **46**:D869–D874.
- 803 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-
804 bioGN]*. doi:arXiv:1303.3997 [q-bio.GN]
- 805 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
806 *Bioinformatics* **25**:1754–1760.
- 807 McGrath PT, Rockman MV, Zimmer M, Jang H, Macosko EZ, Kruglyak L, Bargmann CI. 2009. Quantitative
808 mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors.
809 *Neuron* **61**:692–699.
- 810 Neall VE, Trewick SA. 2008. The age and origin of the Pacific islands: a geological overview. *Philos Trans
811 R Soc Lond B Biol Sci* **363**:3293–3308.
- 812 Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. 2017. Tracing the peopling of the
813 world through genomics. *Nature* **541**:302–310.
- 814 Ortiz EM. n.d. vcf2phylo. Github.
- 815 Pedersen BS, Layer RM, Quinlan AR. 2016. Vcfanno: fast, flexible annotation of genetic variants. *Genome
816 Biol* **17**:118.
- 817 Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: An Efficient Swiss Army
818 Knife for Population Genomic Analyses in R. *Mol Biol Evol* **31**:1929–1936.
- 819 Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele
820 frequency data. *PLoS Genet* **8**:e1002967.
- 821 Poulet N, Vielle A, Gimond C, Ferrari C, Braendle C. 2015. Evolutionarily divergent thermal sensitivity of
822 germline development and fertility in hermaphroditic *Caenorhabditis* nematodes. *Evol Dev* **17**:380–
823 397.
- 824 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW,
825 Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based
826 linkage analyses. *Am J Hum Genet* **81**:559–575.
- 827 Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. 2005.
828 Support from the relationship of genetic and geographic distance in human populations for a serial
829 founder effect originating in Africa. *Proc Natl Acad Sci U S A* **102**:15942–15947.

- 830 Richaud A, Zhang G, Lee D, Lee J, Félix M-A. 2018. The Local Coexistence Pattern of Selfing Genotypes
831 in *Caenorhabditis elegans* Natural Metapopulations. *Genetics* **208**:807–821.
- 832 Rockman MV, Kruglyak L. 2009. Recombinational Landscape and Population Genomics of *Caenorhabditis*
833 *elegans*. *PLoS Genet* **5**:e1000419.
- 834 Rockman MV, Skrovanek SS, Kruglyak L. 2010. Selection at linked sites shapes heritable phenotypic
835 variation in *C. elegans*. *Science* **330**:372–376.
- 836 Schulenburg H, Félix M-A. 2017. The Natural Biotic Environment of *Caenorhabditis elegans*. *Genetics*
837 **206**:55–86.
- 838 Stevens L, Félix M-A, Beltran T, Braendle C, Caerel C, Fausett S, Fitch D, Frézal L, Gosse C, Kaur T,
839 Kiontke K, Newton MD, Noble LM, Richaud A, Rockman MV, Sudhaus W, Blaxter M. 2019.
840 Comparative genomics of 10 new *Caenorhabditis* species. *Evolution Letters* **3**:217–236.
- 841 Stevens L, Félix M-A, Beltran T, Braendle C, Caerel C, Fausett S, Fitch DHA, Frézal L, Kaur T, Kiontke
842 KC, Newton MD, Noble LM, Richaud A, Rockman MV, Sudhaus W, Blaxter M. 2018. Comparative
843 genomics of ten new *Caenorhabditis* species. *bioRxiv*. doi:10.1101/398446
- 844 Sudhaus W, Kiontke K. 2007. Comparison of the cryptic nematode species *Caenorhabditis brenneri* sp. n.
845 and *C. remanei* (Nematoda: Rhabditidae) with the stem species pattern of the *Caenorhabditis Elegans*
846 group. *Zootaxa* **1456**:45–62.
- 847 Sulston JE, Schierenberg E, White JG, Thomson JN. 1983. The embryonic cell lineage of the nematode
848 *Caenorhabditis elegans*. *Dev Biol* **100**:64–119.
- 849 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.
850 *Genetics* **123**:585–595.
- 851 Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on*
852 *mathematics in the life sciences* **17**:57–86.
- 853 Thomas CG, Wang W, Jovelin R, Ghosh R, Lomasko T, Trinh Q, Kruglyak L, Stein LD, Cutter AD. 2015.
854 Full-genome evolutionary histories of selfing, splitting, and selection in *Caenorhabditis*. *Genome Res*
855 **25**:667–678.
- 856 Wicks SR, Yeh RT, Gish WR, Waterston RH, Plasterk RHA. 2001. Rapid gene mapping in *Caenorhabditis*
857 *elegans* using a high density polymorphism map. *Nat Genet* **28**:160.
- 858 Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree : an r package for visualization and annotation of
859 phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**:28–36.