Genome structural variation in natural *Caenorhabditis elegans* populations

Kyle J. Lesack, James D. Wasmuth*

Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

* Corresponding author: jwasmuth@ucalgary.ca

Abstract

Structural variation, involving large alterations in chromosome structure, drives genetic diversification and the emergence of new phenotypes. These changes are widespread in natural populations and play an important role in adaptation and speciation. For many species, research has been limited to laboratory adapted strains and experimental evolution, which may not reflect the diversity of structural variants in the wild. Furthermore, technological limitations have proved to be a major barrier to accurate and comprehensive variant calling. In this study, we used PacBio sequencing data from 14 wild *Caenorhabditis elegans* strains to characterize structural variants in a natural population. With long-reads, we overcame limitations associated with short-read approaches and leveraged population-level data to further refine the accuracy of variant calls. We found that large, rapidly evolving gene families, such as GPCR, F-Box, and C-Type lectin, were prominent among the variants predicted to have phenotypic consequences. These results shed light on the significant role of structural variants in the evolution of *Caenorhabditis elegans* in its wild habitats and the limitations of treating N2 as the reference wild-type.

Keywords: structural variation; *Caenorhabditis elegans*; environmental adaptation

Introduction

Structural variants (SVs) are large-scale alterations in a genome—such as deletions, duplications, and inversions—that affect gene content, order, or dosage. These variants can have profound phenotypic effects by disrupting genes, altering gene dosage, or rearranging regulatory elements. While often deleterious, SVs can also drive adaptive evolution through mechanisms such as gene duplication followed by neofunctionalization or chromosomal inversions that suppress recombination and maintain beneficial allele combinations (Faria et al., 2019; Katju and Bergthorsson, 2013; Kirkpatrick, 2010).

Despite their importance, SVs remain poorly characterized in most species, especially at the population level. This is partly due to the technical limitations of short-read DNA sequencing, which struggles to resolve large or complex variants, particularly in repetitive regions (Koboldt, 2020; Lesack et al., 2022). Long-read sequencing has substantially improved the resolution of SVs and providing more reliable alignments (Smolka et al., 2024). In *Caenorhabditis elegans*, much of our knowledge about SVs comes from laboratory strains and mutation accumulation experiments. These studies have revealed a high rate of spontaneous duplications and the formation of complex alleles (Farslow et al., 2015; Konrad et al., 2018). For example, a recombinant inbred line showed increased fitness due to a partial duplication and inversion of five different regions of *rcan-1*, a gene which regulates thermotactic behavior (Zhao et al., 2020). However, laboratory strains, such as N2, carry numerous derived alleles and pseudogenised genes, limiting their relevance for understanding variation in natural populations (Andersen and Rockman, 2022; Sterken et al., 2015). Comparative studies between N2 and wild isolates, particularly the Hawaiian strain CB4856, have revealed extensive structural divergence

44 (Lee et al., 2021; Vergara et al., 2014). In 12 natural strains, more than 5% of the N2 genome is
45 affected by SVs, with CB4856 showing over 79,000 insertion/deletion events, more than 800
46 unaligned segments, and 61 hyper-divergent regions (Lee et al., 2021; Thompson et al., 2015).
47 The *Caenorhabditis elegans* Natural Diversity Resource (CeNDR) has expanded access
48 to genomic data from wild strains (Cook et al., 2017). This collection has facilitated the
49 discovery of genetic loci underlying xenobiotic metabolism and anthelmintic drug resistance
50 (Hahnel et al., 2018; Zdraljevic et al., 2019). In this study, we used PacBio long-read sequencing
51 data from 14 wild *C. elegans* strains to generate a high-confidence, population-scale map of SVs.
52 We identified thousands of deletions, duplications, and inversions—many affecting genes
53 involved in environmental response, stress, and immune function.
54

Materials and Methods

Strains

We analysed PacBio sequencing data from 14 *C. elegans* strains (DL238, ECA36, ECA396, EG4725, JU1400, JU2526, JU2600, JU310, MY2147, MY2693, NIC2, NIC526, QX1794, XZ1516) sourced from CeNDR and collected from Hawaii, New Zealand, and Western Europe (Figure S1).

Sequence alignment and structural variant calling

Reads were aligned to the N2 reference genome using NGMLR (v0.2.7), and alignments were sorted using Picard (v2.27.5) (Broad Institute, 2025; Sedlazeck et al., 2018). SVs were called using Sniffles2 (v2.0.7), filtered to include only deletions, duplications, and inversions ≥ 100bp (Smolka et al., 2024). SV calls were refined using Iris (v1.0.04) for improved breakpoint resolution and merged across strains using Jasmine (v1.1.5) and Iris (v1.0.4) (Kirsche et al., 2023).

SV annotation and functional profiling

We used Ensmbl's Variant Effect Predictor (v.109.3) to predict the functional consequences of each SV (McLaren et al., 2016). SVs classified as "high impact" were selected for functional enrichment using easyGSEA (v1.3.1) and annotated using WormCat (v2) (Cheng et al., 2021; Holdorf et al., 2020). Genes were assigned to statistically significant terms across WormCat's hierarchical levels (FDR ≤ 0.05), though no single level was exclusively used due to trade-offs in resolution and coverage.

75 ## Data presentation

76 SV densities along chromosomes were plotted using karyoploteR (v1.26.0) (Gel and
77 Serra, 2017). SV allele frequencies were visualized UpSetPlot (v0.8.0) (Lex et al., 2014). Only
78 the largest intersections—25 duplications, 20 for deletions and inversions—were displayed. The
79 remaining plots were created using ggplot2 (v.3.4.2) (Wickham, 2016).

80 ## Data availability

81 The code, and Snakemake pipeline, used in this project is freely available:
82 https://github.com/kyleLesack/sv_analysis_cendr_pacbio

83 # Results and Discussion

84 We identified 14,473 deletions, 3,149 duplications, and 1,748 inversions in 14 wild *C.*
85 *elegans* strains (Table S1). SVs were unevenly distributed across the genome (Figure S2).
86 Deletions and duplications were enriched at the autosomal arms, whereas inversions were more
87 evenly distributed. These patterns are consistent with previous findings; SVs preferentially occur
88 in low gene density, high recombination regions (Cutter et al., 2009). Chromosome arms may
89 promote SV formation through recombination-based mechanisms while protecting essential
90 genes located in denser chromosome centres (Carvalho and Lupski, 2016). Among all SVs, 64%
91 were singletons—variants unique to a single strain—suggesting high inter-strain diversity (Table
92 S2). Hawaiian strains had the highest number and proportion of singleton variants while Western
93 European strains had the fewest (Table 1). The high ancestral diversity in Hawaiian strains is
94 known, with reduced diversity in non-Hawaiian lineages due to selective sweeps (Crombie et al.,
95 2019). Conversely, 25 variants were found in all 14 strains: 15 deletions, two duplications, and
96 eight inversions (Table S2; Figures S3, S4, S5).

97 To prioritise variants with likely functional consequences, we focused on high-impact
98 SVs. This subset included 1,485 deletions, 214 duplications, and three inversions, collectively
99 affecting 1,258 genes (Figure 1; Tables S3, S4, S5). Deletions were enriched for splicing
100 disruptions, transcript ablations, and frameshift variants (Table 2). Duplications primarily
101 resulted in transcript amplifications, and all three high impact inversions were classified as
102 stop/loss mutations. Functional enrichment revealed strong overrepresentation in categories
103 related to environmental stress, including GPCRs, F-box proteins, C-type lectins (CLECs), and
104 detoxification enzymes (Figure 2; Table S6). Many of these genes were in hyper-variable

105 regions, which have been shown to maintain ancestral variation through balancing selection (Lee

106 et al., 2021).

107

Strain	Region	Deletions			Duplications			Inversions		
		Total	Unique	Unique Proportion	Total	Unique	Unique Proportion	Total	Unique	Unique Proportion
DL238	Hawaii	2343	804	0.34	485	209	0.43	333	95	0.29
ECA396	Hawaii	3412	1643	0.48	519	300	0.58	399	140	0.35
QX1794	Hawaii	2363	756	0.32	459	186	0.41	311	74	0.24
XZ1516	Hawaii	4007	2440	0.61	443	266	0.60	519	301	0.58
ECA36	New Zealand	3522	1671	0.47	586	308	0.53	398	147	0.37
EG4725	Western Europe	1913	290	0.15	452	94	0.21	280	42	0.15
JU1400	Western Europe	1692	225	0.13	417	120	0.29	240	34	0.14
JU2526	Western Europe	2257	452	0.20	442	96	0.22	309	55	0.18
JU2600	Western Europe	1519	290	0.19	385	87	0.23	238	39	0.16
JU310	Western Europe	1366	170	0.12	385	72	0.19	217	16	0.07
MY2147	Western Europe	1279	217	0.17	364	85	0.23	199	29	0.15
MY2693	Western Europe	1251	97	0.08	385	84	0.22	258	35	0.14
NIC2	Western Europe	457	80	0.18	248	77	0.31	150	12	0.08
NIC526	Western Europe	1544	179	0.12	402	78	0.19	257	22	0.09

108 **Table 1:** Singleton variants among individual *C. elegans* strains.

109

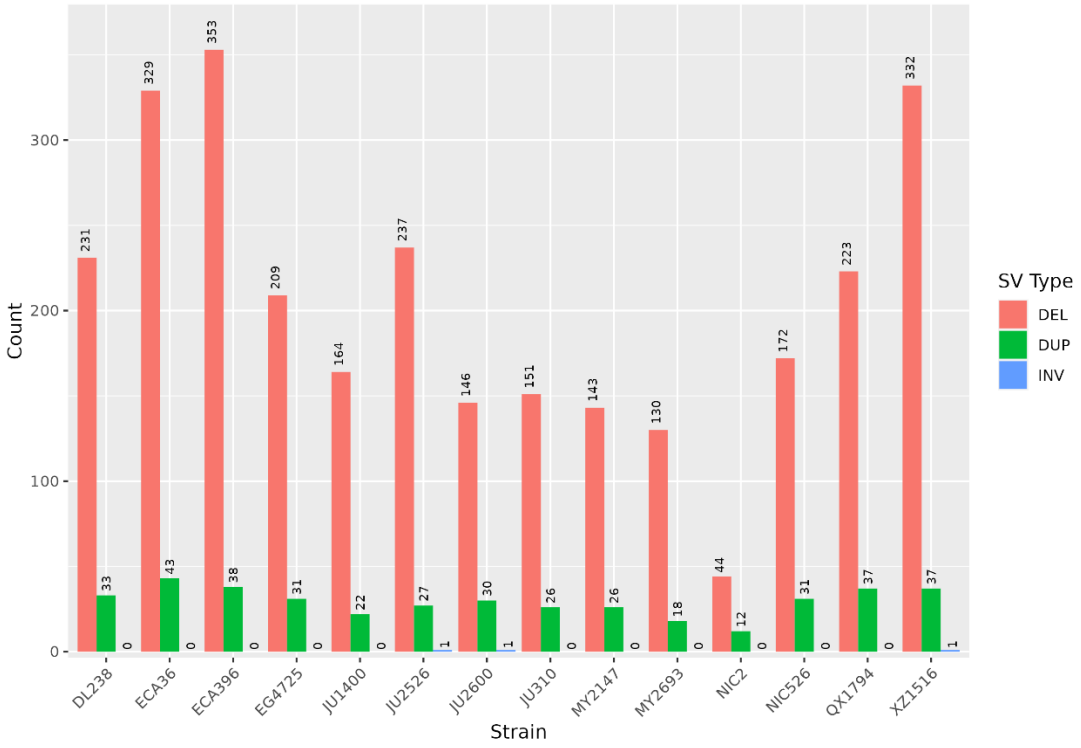

Figure 1: Number of structural variants determined to have high functional impact.

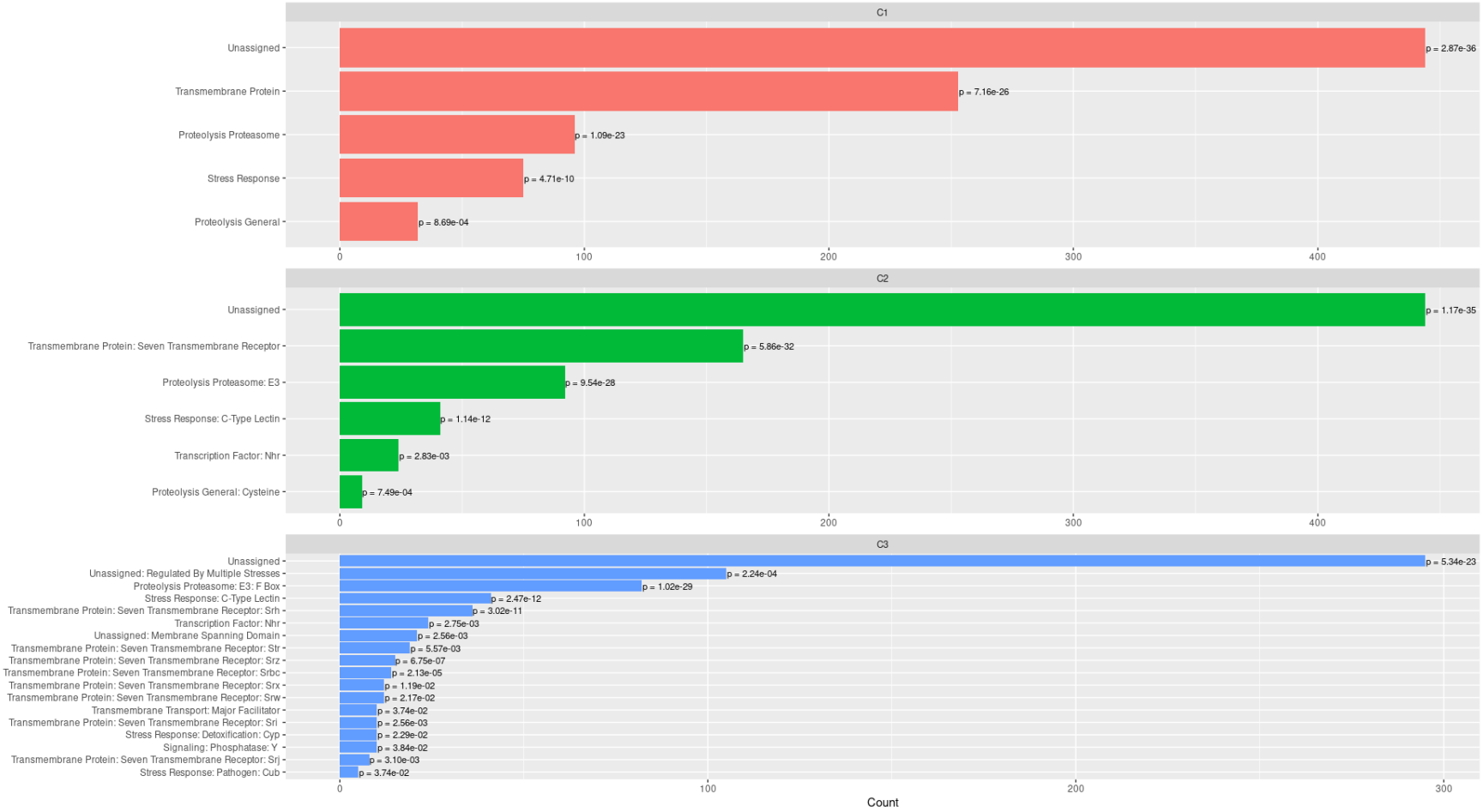

Figure 2: Statistically significant overrepresented WormCat terms for putative high impact SVs. The red, green, and blue colours represent the categories for the C1, C2, and C3 levels, respectively.

117

SV Type	Term	Description‡	Count
Deletion (n = 1,485)	splice_donor_variant	A splice variant that changes the 2-base region at the 5' end of an intron	733
	splice_acceptor_variant	A splice variant that changes the 2-base region at the 3' end of an intron	724
	transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature	411
	frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	223
	stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	120
	start_lost	A codon variant that changes at least one base of the canonical start codon	20
	stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	2
Duplication (n = 241)	transcript_amplification	A feature amplification of a region containing a transcript	240
	start_lost	A codon variant that changes at least one base of the canonical start codon	1
Inversion (n=3)	start_lost	A codon variant that changes at least one base of the canonical start codon	3

118 **Table 2:** Predicted Consequences of High Severity Structural Variants.

119 ‡ Consequence descriptions downloaded on May 22, 2023 from:
120 https://useast.ensembl.org/info/genome/variation/prediction/predicted_data.html

121

122 **GPCRs and sensory behaviour**

123 Seven-transmembrane G-protein-coupled receptors (GPCRs), involved in
124 chemosensation and behaviour, were among the most affected gene families. Of the 253 genes in
125 the "Transmembrane protein" category, 152 were GPCRs. The SRH family was enriched, with
126 36 genes, which is consistent with previous reports of rapid expansion and diversification in *C.*
127 *elegans* (Thomas and Robertson, 2008; Vidal et al., 2018). A duplication of *srh-234* was present

128 in two strains. The expression of *srh-234* is regulated by the chemosensation of food (Gruner et
129 al., 2014). Two deletions are GPCRs were noteworthy: *sri-14* is involved in response to high and
130 toxic concentrations of diacetyl (Taniguchi et al., 2014); and *str-208* is implicated in erratic
131 movement (Daigle et al., 2022). While the loss of these genes may reduce responsiveness to
132 adverse stimuli or affect locomotion, it is not yet known whether the associated structural
133 variants confer an adaptive benefit, impose a cost, or are selectively neutral.

134 **F-box genes and immune adaptation**

135 SVs affected 128 proteolysis-related genes, including 74 F-box genes. The F-box gene
136 family is considerably expanded in *C. elegans* relative to other *Caenorhabditis* species with
137 several members under positive selection in the non-Hawaiian strains (Ma et al., 2021). One of
138 these is *fbxa-158*, which promoted a thermotolerance-associated intracellular pathogen response,
139 but we found it was lost in three non-Hawaiian strains, EG4725, JU1400, and MY2693, and one
140 Hawaiian strain, ECA396 (Panek et al., 2020). Additional immune-related SVs may reflect
141 adaptations to local microbial environments. For example, *hecw-1*, deleted in the Hawaiian
142 strains ECA396 and ZX1516, is an E3 ubiquitin ligase that modulate pathogen avoidance (Chang
143 et al., 2011).

144 **C-type lectins and pathogen response**

145 Thirty-one C-type lectin (CLEC) genes were affected by SVs. CLECs are a rapidly
146 evolving gene family involved in innate immunity and pathogen recognition (Pees et al., 2016).
147 Seven CLECs were affected by deletions—*clec-2, clec-10, clec-13, clec-67, clec-122, clec-143*,
148 and *clec-174*—all known to be modulated by bacterial pathogens (Pan et al., 2021; Sahu et al.,
149 2012; White and Herman, 2018). A further three CLECs were duplicated; *clec-70* is induced by

150 *Microbacterium nematophilum* and *Staphylococcus aureus*, while *clec-189* and *clec-258* await
151 functional annotation but may represent novel immune effectors (Pan et al., 2021).

152 Detoxification and xenobiotic response

153 Several families of detoxification genes were enriched in SVs, including cytochrome P450s
154 (CYPs), glutathione S-transferases (GSTs), UDP-glucuronosyltransferases (UGTs), and nuclear
155 hormone receptors (NHRs). Two caffeine inducible CYPs, *cyp-33e2* and *cyp-35a4* harbored
156 deletions; the latter gene also regulates fat storage (Aarnio et al., 2011; Min et al., 2015). Other
157 deletion-affected CYP genes included: *cyp-33c5* and *cyp-13a5*, which are involved in cadmium
158 detoxification; and *cyp-33d3*, which is inducible by ethidium bromide and low pH (Cong et al.,
159 2020; Cui et al., 2007). In other families, notable deletions included: *ugt-14* is upregulated in
160 response to benzimidazole-class anthelmintics (Stasiuk et al., 2019); and *nhr-196* is regulated by
161 exposure to the mycotoxin aflatoxin B1 (Karengera et al., 2022).

162 Evolutionary and Functional Implications

163 Our results highlight the utility of long-read sequencing for characterizing SVs at a
164 population scale. The high diversity of SVs in wild *C. elegans* populations—particularly those in
165 gene families related to sensory perception, immunity, and detoxification—emphasizes the
166 limitations of relying solely on N2 as a wild-type reference (Andersen and Rockman, 2022;
167 Sterken et al., 2015). Many of the identified SVs may confer adaptive advantages, particularly in
168 geographically and microbially diverse environments.

169 Importantly, many SVs overlap with hyper-divergent regions—genomic hotspots for
170 adaptive diversity—shown in *C. elegans* and the parasitic *Heligmosomoides polygyrus* to contain
171 genes involved in sensory perception and stress response (Lee et al., 2021; Stevens et al., 2023).

172 Going forward, hybrid approaches that integrate long- and short-read sequencing and
173 transcriptomic data from wild isolates will help resolve the regulatory and phenotypic
174 consequences of SVs (Zhang et al., 2022). In particular, eQTL mapping that includes SVs could
175 offer insights into complex traits not explained by single-nucleotide variation alone (Boyling et
176 al., 2022). Additionally, functional studies targeting SVs genes currently categorized as
177 "Unassigned" may reveal previously uncharacterized molecular mechanisms.
178

179

Acknowledgments

We thank Dr. Erik Andersen and members of the Andersen lab (Johns Hopkins University) for discussions on using CeNDR datasets. We also acknowledge the high-performance computing resources made available by Research Computing and the Faculty of Veterinary Medicine at the University of Calgary.

Funding

This work was funded through grants to JDW from the Natural Sciences and Engineering Research Council of Canada (NSERC; #04589-2020).

References

Aarnio, V., Lehtonen, M., Storvik, M., Callaway, J.C., Lakso, M., Wong, G., 2011. Caenorhabditis Elegans Mutants Predict Regulation of Fatty Acids and Endocannabinoids by the CYP-35A Gene Family. Front. Pharmacol. 2. https://doi.org/10.3389/fphar.2011.00012

Andersen, E.C., Rockman, M.V., 2022. Natural genetic variation as a tool for discovery in *Caenorhabditis* nematodes. Genetics 220, iyab156. https://doi.org/10.1093/genetics/iyab156

Boyling, A., Perez-Siles, G., Kennerson, M.L., 2022. Structural Variation at a Disease Mutation Hotspot: Strategies to Investigate Gene Regulation and the 3D Genome. Front. Genet. 13, 842860. https://doi.org/10.3389/fgene.2022.842860

Broad Institute, 2025. broadinstitute/picard.

Carvalho, C.M.B., Lupski, J.R., 2016. Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet 17, 224–238. https://doi.org/10.1038/nrg.2015.25

Chang, H.C., Paek, J., Kim, D.H., 2011. Natural polymorphisms in C. elegans HECW-1 E3 ligase affect pathogen avoidance behaviour. Nature 480, 525–529. https://doi.org/10.1038/nature10643

Cheng, X., Yan, J., Liu, Y., Wang, J., Taubert, S., 2021. eVITTA: a web-based visualization and inference toolbox for transcriptome analysis. Nucleic Acids Research 49, W207–W215. https://doi.org/10.1093/nar/gkab366

211 Cong, Y., Yang, H., Zhang, P., Xie, Y., Cao, X., Zhang, L., 2020. Transcriptome Analysis of the
212 Nematode Caenorhabditis elegans in Acidic Stress Environments. Front. Physiol. 11,
213 1107. https://doi.org/10.3389/fphys.2020.01107
214 Cook, D.E., Zdraljevic, S., Roberts, J.P., Andersen, E.C., 2017. CeNDR, the Caenorhabditis
215 elegans natural diversity resource. Nucleic Acids Res 45, D650–D657.
216 https://doi.org/10.1093/nar/gkw893
217 Crombie, T.A., Zdraljevic, S., Cook, D.E., Tanny, R.E., Brady, S.C., Wang, Y., Evans, K.S.,
218 Hahnel, S., Lee, D., Rodriguez, B.C., Zhang, G., Van Der Zwagg, J., Kiontke, K.,
219 Andersen, E.C., 2019. Deep sampling of Hawaiian Caenorhabditis elegans reveals high
220 genetic diversity and admixture with global populations. eLife 8, e50465.
221 https://doi.org/10.7554/eLife.50465
222 Cui, Y., McBride, S.J., Boyd, W.A., Alper, S., Freedman, J.H., 2007. Toxicogenomic analysis of
223 Caenorhabditis elegans reveals novel genes and pathways involved in the resistance to
224 cadmium toxicity. Genome Biol 8, R122. https://doi.org/10.1186/gb-2007-8-6-r122
225 Cutter, A.D., Dey, A., Murray, R.L., 2009. Evolution of the Caenorhabditis elegans Genome.
226 Molecular Biology and Evolution 26, 1199–1234.
227 https://doi.org/10.1093/molbev/msp048
228 Daigle, A.T., Deiss, T.C., Melde, R.H., Bergthorsson, U., Katju, V., 2022. Bergerac strains of
229 *Caenorhabditis elegans* revisited: expansion of Tc *1* elements imposes a significant
230 genomic and fitness cost. G3 Genes|Genomes|Genetics 12, jkac214.
231 https://doi.org/10.1093/g3journal/jkac214
232 Faria, R., Johannesson, K., Butlin, R.K., Westram, A.M., 2019. Evolving Inversions. Trends in
233 Ecology & Evolution 34, 239–248. https://doi.org/10.1016/j.tree.2018.12.005

234 Farslow, J.C., Lipinski, K.J., Packard, L.B., Edgley, M.L., Taylor, J., Flibotte, S., Moerman,
235 D.G., Katju, V., Bergthorsson, U., 2015. Rapid Increase in frequency of gene copy-
236 number variants during experimental evolution in Caenorhabditis elegans. BMC
237 Genomics 16, 1044. https://doi.org/10.1186/s12864-015-2253-2
238 Gel, B., Serra, E., 2017. karyoploteR: an R/Bioconductor package to plot customizable genomes
239 displaying arbitrary data. Bioinformatics 33, 3088–3090.
240 https://doi.org/10.1093/bioinformatics/btx346
241 Gruner, M., Nelson, D., Winbush, A., Hintz, R., Ryu, L., Chung, S.H., Kim, K., Gabel, C.V., Van
242 Der Linden, A.M., 2014. Feeding State, Insulin and NPR-1 Modulate Chemoreceptor
243 Gene Expression via Integration of Sensory and Circuit Inputs. PLoS Genet 10,
244 e1004707. https://doi.org/10.1371/journal.pgen.1004707
245 Hahnel, S.R., Zdraljevic, S., Rodriguez, B.C., Zhao, Y., McGrath, P.T., Andersen, E.C., 2018.
246 Extreme allelic heterogeneity at a Caenorhabditis elegans beta-tubulin locus explains
247 natural resistance to benzimidazoles. PLoS Pathog 14, e1007226.
248 https://doi.org/10.1371/journal.ppat.1007226
249 Holdorf, A.D., Higgins, D.P., Hart, A.C., Boag, P.R., Pazour, G.J., Walhout, A.J.M., Walker,
250 A.K., 2020. WormCat: An Online Tool for Annotation and Visualization of
251 *Caenorhabditis elegans* Genome-Scale Data. Genetics 214, 279–294.
252 https://doi.org/10.1534/genetics.119.302919
253 Karengera, A., Sterken, M.G., Kammenga, J.E., Riksen, J.A.G., Dinkla, I.J.T., Murk, A.J., 2022.
254 Differential expression of genes in C. elegans reveals transcriptional responses to
255 indirect-acting xenobiotic compounds and insensitivity to 2,3,7,8-

256 tetrachlorodibenzodioxin. Ecotoxicology and Environmental Safety 233, 113344.
257 https://doi.org/10.1016/j.ecoenv.2022.113344
258 Katju, V., Bergthorsson, U., 2013. Copy-number changes in evolution: rates, fitness effects and
259 adaptive significance. Front. Genet. 4. https://doi.org/10.3389/fgene.2013.00273
260 Kirkpatrick, M., 2010. How and Why Chromosome Inversions Evolve. PLoS Biol 8, e1000501.
261 https://doi.org/10.1371/journal.pbio.1000501
262 Kirsche, M., Prabhu, G., Sherman, R., Ni, B., Battle, A., Aganezov, S., Schatz, M.C., 2023.
263 Jasmine and Iris: population-scale structural variant comparison and analysis. Nat
264 Methods 20, 408–417. https://doi.org/10.1038/s41592-022-01753-3
265 Koboldt, D.C., 2020. Best practices for variant calling in clinical sequencing. Genome Med 12,
266 91. https://doi.org/10.1186/s13073-020-00791-w
267 Konrad, A., Flibotte, S., Taylor, J., Waterston, R.H., Moerman, D.G., Bergthorsson, U., Katju, V.,
268 2018. Mutational and transcriptional landscape of spontaneous gene duplications and
269 deletions in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U.S.A. 115, 7386–7391.
270 https://doi.org/10.1073/pnas.1801930115
271 Lee, D., Zdraljevic, S., Stevens, L., Wang, Y., Tanny, R.E., Crombie, T.A., Cook, D.E., Webster,
272 A.K., Chirakar, R., Baugh, L.R., Sterken, M.G., Braendle, C., Félix, M.-A., Rockman,
273 M.V., Andersen, E.C., 2021. Balancing selection maintains hyper-divergent haplotypes in
274 Caenorhabditis elegans. Nat Ecol Evol 5, 794–807. https://doi.org/10.1038/s41559-021-
275 01435-x
276 Lesack, K., Mariene, G.M., Andersen, E.C., Wasmuth, J.D., 2022. Different structural variant
277 prediction tools yield considerably different results in Caenorhabditis elegans. PLOS
278 ONE 17, e0278424. https://doi.org/10.1371/journal.pone.0278424

279 Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., Pfister, H., 2014. UpSet: Visualization of
280 Intersecting Sets. IEEE Trans Vis Comput Graph 20, 1983–1992.
281 https://doi.org/10.1109/TVCG.2014.2346248
282 Ma, F., Lau, C.Y., Zheng, C., 2021. Large Genetic Diversity and Strong Positive Selection in F-
283 Box and GPCR Genes among the Wild Isolates of Caenorhabditis elegans. Genome Biol
284 Evol 13, evab048. https://doi.org/10.1093/gbe/evab048
285 McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P.,
286 Cunningham, F., 2016. The Ensembl Variant Effect Predictor. Genome Biol 17, 122.
287 https://doi.org/10.1186/s13059-016-0974-4
288 Min, H., Kawasaki, I., Gong, J., Shim, Y.-H., 2015. Caffeine Induces High Expression of cyp-
289 35A Family Genes and Inhibits the Early Larval Development in Caenorhabditis elegans.
290 Molecules and Cells 38, 236–242. https://doi.org/10.14348/molcells.2015.2282
291 Pan, W., Huang, X., Guo, Z., Nagarajan, R., Mylonakis, E., 2021. Identification and Functional
292 Analysis of Cytokine-Like Protein CLEC-47 in Caenorhabditis elegans. mBio 12,
293 e02579-21. https://doi.org/10.1128/mBio.02579-21
294 Panek, J., Gang, S.S., Reddy, K.C., Luallen, R.J., Fulzele, A., Bennett, E.J., Troemel, E.R., 2020.
295 A cullin-RING ubiquitin ligase promotes thermotolerance as part of the intracellular
296 pathogen response in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. U.S.A. 117, 7950–
297 7960. https://doi.org/10.1073/pnas.1918417117
298 Pees, B., Yang, W., Zárate-Potes, A., Schulenburg, H., Dierking, K., 2016. High Innate Immune
299 Specificity through Diversified C-Type Lectin-Like Domain Proteins in Invertebrates. J
300 Innate Immun 8, 129–142. https://doi.org/10.1159/000441475

301 Sahu, S.N., Lewis, J., Patel, I., Bozdag, S., Lee, J.H., LeClerc, J.E., Cinar, H.N., 2012. Genomic
302 Analysis of Immune Response against Vibrio cholerae Hemolysin in Caenorhabditis
303 elegans. PLoS ONE 7, e38200. https://doi.org/10.1371/journal.pone.0038200
304 Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., Schatz,
305 M.C., 2018. Accurate detection of complex structural variations using single-molecule
306 sequencing. Nat Methods 15, 461–468. https://doi.org/10.1038/s41592-018-0001-7
307 Smolka, M., Paulin, L.F., Grochowski, C.M., Horner, D.W., Mahmoud, M., Behera, S., Kalef-
308 Ezra, E., Gandhi, M., Hong, K., Pehlivan, D., Scholz, S.W., Carvalho, C.M.B., Proukakis,
309 C., Sedlazeck, F.J., 2024. Detection of mosaic and population-level structural variants
310 with Sniffles2. Nat Biotechnol 1–10. https://doi.org/10.1038/s41587-023-02024-y
311 Stasiuk, S.J., MacNevin, G., Workentine, M.L., Gray, D., Redman, E., Bartley, D., Morrison, A.,
312 Sharma, N., Colwell, D., Ro, D.K., Gilleard, J.S., 2019. Similarities and differences in the
313 biotransformation and transcriptomic responses of Caenorhabditis elegans and
314 Haemonchus contortus to five different benzimidazole drugs. Int J Parasitol Drugs Drug
315 Resist 11, 13–29. https://doi.org/10.1016/j.ijpddr.2019.09.001
316 Sterken, M.G., Snoek, L.B., Kammenga, J.E., Andersen, E.C., 2015. The laboratory
317 domestication of Caenorhabditis elegans. Trends in Genetics 31, 224–231.
318 https://doi.org/10.1016/j.tig.2015.02.009
319 Stevens, L., Martínez-Ugalde, I., King, E., Wagah, M., Absolon, D., Bancroft, R., Gonzalez de la
320 Rosa, P., Hall, J.L., Kieninger, M., Kloch, A., Pelan, S., Robertson, E., Pedersen, A.B.,
321 Abreu-Goodger, C., Buck, A.H., Blaxter, M., 2023. Ancient diversity in host-parasite
322 interaction genes in a model parasitic nematode. Nat Commun 14, 7776.
323 https://doi.org/10.1038/s41467-023-43556-w

324 Taniguchi, G., Uozumi, T., Kiriyama, K., Kamizaki, T., Hirotsu, T., 2014. Screening of Odor-
325 Receptor Pairs in *Caenorhabditis elegans* Reveals Different Receptors for High and Low
326 Odor Concentrations. Sci. Signal. 7. https://doi.org/10.1126/scisignal.2005136
327 Thomas, J.H., Robertson, H.M., 2008. The Caenorhabditis chemoreceptor gene families. BMC
328 Biol 6, 42. https://doi.org/10.1186/1741-7007-6-42
329 Thompson, O.A., Snoek, L.B., Nijveen, H., Sterken, M.G., Volkers, R.J.M., Brenchley, R., Van'T
330 Hof, A., Bevers, R.P.J., Cossins, A.R., Yanai, I., Hajnal, A., Schmid, T., Perkins, J.D.,
331 Spencer, D., Kruglyak, L., Andersen, E.C., Moerman, D.G., Hillier, L.W., Kammenga,
332 J.E., Waterston, R.H., 2015. Remarkably Divergent Regions Punctuate the Genome
333 Assembly of the *Caenorhabditis elegans* Hawaiian Strain CB4856. Genetics 200, 975–
334 989. https://doi.org/10.1534/genetics.115.175950
335 Vergara, I.A., Tarailo-Graovac, M., Frech, C., Wang, J., Qin, Z., Zhang, T., She, R., Chu, J.S.,
336 Wang, K., Chen, N., 2014. Genome-wide variations in a natural isolate of the nematode
337 Caenorhabditis elegans. BMC Genomics 15, 255. https://doi.org/10.1186/1471-2164-15-
338 255
339 Vidal, B., Aghayeva, U., Sun, H., Wang, C., Glenwinkel, L., Bayer, E.A., Hobert, O., 2018. An
340 atlas of Caenorhabditis elegans chemoreceptor expression. PLoS Biol 16, e2004218.
341 https://doi.org/10.1371/journal.pbio.2004218
342 White, C.V., Herman, M.A., 2018. Transcriptomic, Functional, and Network Analyses Reveal
343 Novel Genes Involved in the Interaction Between Caenorhabditis elegans and
344 Stenotrophomonas maltophilia. Front. Cell. Infect. Microbiol. 8, 266.
345 https://doi.org/10.3389/fcimb.2018.00266

346 Wickham, H., 2016. ggplot2, Use R! Springer International Publishing, Cham.
347 https://doi.org/10.1007/978-3-319-24277-4
348 Zdraljevic, S., Fox, B.W., Strand, C., Panda, O., Tenjo, F.J., Brady, S.C., Crombie, T.A., Doench,
349 J.G., Schroeder, F.C., Andersen, E.C., 2019. Natural variation in C. elegans arsenic
350 toxicity is explained by differences in branched chain amino acid metabolism. eLife 8,
351 e40260. https://doi.org/10.7554/eLife.40260
352 Zhang, G., Roberto, N.M., Lee, D., Hahnel, S.R., Andersen, E.C., 2022. The impact of species-
353 wide gene expression variation on Caenorhabditis elegans complex traits. Nat Commun
354 13, 3462. https://doi.org/10.1038/s41467-022-31208-4
355 Zhao, Y., Long, L., Wan, J., Biliya, S., Brady, S.C., Lee, D., Ojemakinde, A., Andersen, E.C.,
356 Vannberg, F.O., Lu, H., McGrath, P.T., 2020. A spontaneous complex structural variant in
357 rcan-1 increases exploratory behavior and laboratory fitness of Caenorhabditis elegans.
358 PLoS Genet 16, e1008606. https://doi.org/10.1371/journal.pgen.1008606
359
360
361
362