

1 **Concise Title:**

2 **MIP-MAP: High Throughput Mapping of *Caenorhabditis elegans* Temperature**

3 **Sensitive Mutants via Molecular Inversion Probes**

4

5 **Mok CA^{*}, Au V[†], Thompson OA^{*}, Edgley ML[†], Gevirtzman L^{*}, Yochem J[‡], Lowry J[‡],**

6 **Memar N[§], Wallenfang M^{**}, Rasoloson D^{**}, Bowerman B[‡], Schnabel R^{††}, Seydoux**

7 **G^{**}, Moerman DG[†], Waterston RH^{*}**

8 *** Department of Genome Sciences, University of Washington, Seattle, Washington,**

9 **USA, 98195**

10 **† Department of Zoology and Michael Smith Laboratories, University of British**

11 **Columbia, Vancouver, British Columbia, Canada, V6T 1Z4**

12 **‡ Institute of Molecular Biology, University of Oregon, Eugene, Oregon, USA, 97403**

13 **§ Department Biology II, Biocenter, LMU Munich, Martinsried, Germany**

14 **** Dept. of Molecular Biology and Genetics, Johns Hopkins University School of**

15 **Medicine, Baltimore MD 21205, USA**

16 **†† Department of Developmental Genetics, Institute of Genetics, TU Braunschweig,**

17 **Spielmannstr. 7, 38106 Braunschweig, Germany**

18

19 **Running Title:**

20 **Mapping nematode TS alleles**

21

22 **Keywords:**

23 Genetic Mapping

24 *Caenorhabditis elegans*

25 Temperature Sensitive

26 Molecular Inversion Probes

27 Massively Multiplex Sequencing

28

29 **Corresponding Author:**

30 Calvin Mok

31 3720 15th Ave NE

32 Foege South, room S343E, Seattle, WA, USA

33 98195-5065

34 **Abstract**

35 Temperature sensitive (TS) alleles are important tools for the genetic and functional
36 analysis of essential genes in many model organisms. While isolating TS alleles is not
37 difficult, determining the TS-conferring mutation can be problematic. Even with whole-
38 genome sequencing (WGS) data there is a paucity of predictive methods for identifying
39 TS alleles from DNA sequence alone. We assembled 173 TS lethal mutants of
40 *Caenorhabditis elegans* and used WGS to identify several hundred mutations per strain.
41 We leveraged single molecule molecular inversion probes (MIPs) to sequence variant
42 sites at high depth in the cross-progeny of TS mutants and a mapping strain with
43 identified sequence variants but no apparent phenotypic differences from the reference
44 N2 strain. By sampling for variants at ~1Mb intervals across the genome we genetically
45 mapped mutant alleles at a resolution comparable to current standards in a process we
46 call MIP-MAP. The MIP-MAP protocol, however, permits high-throughput sequencing of
47 multiple TS mutation mapping libraries at less than 200K reads per library. Using MIP-
48 MAP on a subset of TS mutants, via a competitive selection assay and standard
49 recombinant mutant selection, we defined TS-associated intervals of 3Mb or less. Our
50 results suggest this collection of strains contains a diverse library of TS alleles for genes
51 involved in development and reproduction. MIP-MAP is a robust method to genetically
52 map mutations in both viable and essential genes. The MIPs protocol should allow high-
53 throughput tracking of genetic variants in any mixed population.

54 **Introduction**

55 The Million Mutation Project, a collection of genomic data for thousands of mutagenized
56 *C. elegans* strains (Thompson *et al.* 2013), has provided a dense library of mutant
57 alleles with which to study gene function and has been widely used since its release.
58 However, given that such strains generally contain only homozygous viable mutations,
59 they largely exclude strongly deleterious alleles of essential genes. Such genes
60 encompass a range of classes that include roles in cell division, development and
61 fertility – key components to all multicellular organisms. Temperature-sensitive (TS)
62 lethal alleles can facilitate the genetic analysis of essential genes through the
63 conditional modulation of their function without the complications that balancer
64 chromosomes can introduce when present in non-conditional lethal strains (Golden *et*
65 *al.* 2000; O'Rourke *et al.* 2011a; Lowry *et al.* 2015). There are several reported screens
66 for TS lethal alleles in *C. elegans* but to date, there are only a small portion of genes
67 with TS alleles identified (Zonies *et al.* 2010; Ehmke *et al.* 2014; Lowry *et al.* 2015).
68 Generating a comprehensive library of mutant strains with conditional lethal phenotypes
69 has the potential to expand our knowledge of essential genes, their required levels of
70 expression, the timing of their function(s), and the details of their protein and domain
71 structure. Furthermore, such a catalog could have broad implications for elucidating the
72 life cycle of *C. elegans* and organisms beyond this model. Such an endeavour,
73 however, requires the mapping and characterization of mutant alleles in a systematic
74 and high-throughput manner.

75 Chemical mutagenesis methods to generate TS alleles typically result in up to
76 several hundred mutations per strain (Sarin *et al.* 2008; Flibotte *et al.* 2010; Thompson

77 *et al.* 2013). Determining which of these molecular alterations is responsible for the TS
78 mutant phenotype is especially challenging when dealing with partially penetrant or
79 leaky alleles. Many mutations can initially be prioritized as candidates because they
80 alter the coding potential of a gene, but even after such filtering, more than 1/5 of the
81 collection's SNVs remain. In contrast to identifying constitutive loss-of-function alleles,
82 such as deletions, frame-shifts and nonsense mutations, there is, as yet, no universal
83 predictive method for recognizing which coding or perhaps noncoding molecular change
84 is most likely to give rise to a temperature sensitive phenotype (Perry *et al.* 1994;
85 Rogalski *et al.* 1995; Harfe *et al.* 1998; Poultney *et al.* 2011). When feasible, mutants
86 are initially outcrossed or backcrossed to remove extraneous mutations originating from
87 the mutagenesis process. Otherwise, these additional mutations may, to some degree,
88 affect development or other essential pathways, thus obfuscating the process of allele
89 characterization.

90 Genetic mapping is a general method for identifying the causative mutation in
91 such strains. Mapping with classic visible genetic markers in *C. elegans* can be a
92 laborious, iterative process. Single nucleotide variants between strains, such as
93 wildtype N2 and the Hawaiian strain CB4856, act as molecular genetic markers that
94 allow the parallel mapping of multiple sites across the genome in a single cross. The so-
95 called Snip-SNP assay is a popular method that exploits a subset of Hawaiian genome
96 SNVs that alter recognition sites of the restriction enzyme DraI (Davis *et al.* 2005).
97 However, the Hawaiian genome is proven to harbor various alleles (De Bono and
98 Bargmann 1998; Seidel *et al.* 2008, 2011; Andersen *et al.* 2014) that can negatively
99 alter the representation of segregant populations; CB4856 also has phenotypes of its

100 own that may interfere with the scoring of some behavioral phenotypes (Wicks *et al.*
101 2001). Despite these issues, the research community continues to leverage CB4856 as
102 a mapping strain to develop new methods of mapping complex mutations (Doitsidou *et*
103 *al.* 2010; O'Rourke *et al.* 2011b; Minevich *et al.* 2012; Smith *et al.* 2016).

104 More recently, approaches and tools have been developed that allow the
105 simultaneous measurement of SNV frequency in a cross-population. Included amongst
106 these are a combined-step WGS and SNP analysis complemented by analysis via the
107 CloudMap system as well as restriction site associated DNA polymorphism mapping
108 (Doitsidou *et al.* 2010; O'Rourke *et al.* 2011b; Minevich *et al.* 2012). At their core, these
109 mapping methods still rely on the basic principles of bulk segregant analysis except that
110 they now benefit from using massively parallel sequencing to examine molecular
111 markers across the genome in a single sequencing library. As currently implemented
112 these mapping methods use whole genome sequencing (WGS) to a depth of 18-40X or
113 more on populations generated from 20-50 F2 mutant phenotype animals (Doitsidou *et*
114 *al.* 2010; Wang *et al.* 2014; Jaramillo-Lambert *et al.* 2015; Lowry *et al.* 2015) to capture
115 data on both the recombinant genome landscape and the associated mutant allele. On
116 a large scale, mutant mapping strains is a labor-intensive exercise. The large amount of
117 sequence generation, the use of the Hawaiian strain - with its limitations - and the
118 issues with scaling to high-throughput suggest there is room for improvement. In
119 particular, molecular inversion probes (Turner *et al.* 2009) (MIPs), allow high-
120 throughput, target-based genome amplification and sequencing. A single MIP targets a
121 region via a pair of complementary annealing arms. Once annealed, the genomic
122 sequence between arms, also known as the gap-fill sequence, is copied by a DNA

123 polymerase and the entire single-stranded probe is circularized by ligase. Each
124 circularized probe represents a unique strand of DNA that can now be linearized and
125 sequenced. MIPs have proven very useful in targeted exome sequencing libraries
126 (Mamanova *et al.* 2010; Kiezun *et al.* 2013) and in screening samples for
127 subpopulations of variants. Recently, Hiatt *et al.*, introduced single molecule molecular
128 inversion probes (smMIPs) containing the addition of a randomized molecular barcode
129 to differentiate between low-frequency variants and those contributed by sequencing or
130 amplification artifacts in genomically diverse samples (Hiatt *et al.* 2013).

131 In complement to the Million Mutation Project collection we have sequenced a
132 group of 173 previously uncharacterized mutants obtained from multiple screens for TS
133 lethality, which reveals a rich catalog of genomic variants. To identify the specific
134 mutation(s) underlying these TS phenotypes, we developed an alternative strategy to
135 whole genome sequencing genetic mapping that exploits MIPs to specifically assay only
136 regions of the genome containing SNVs of interest. To avoid the pitfalls associated with
137 the Hawaiian strain we identified a polymorphic mapping strain from the Million Mutation
138 Project that resembles N2 in movement and growth but has fewer than 300 SNVS
139 across the genome. In pilot experiments, we demonstrate the utility of our MIP-based
140 method (MIP-MAP) in generating high-resolution genomic interval maps for *C. elegans*
141 mutant alleles. We further developed a competitive fitness approach to map the
142 associated genomic intervals for a subset of our TS lethal mutants. In crosses with our
143 mapping strain we use non-permissive temperatures to select against TS lethal
144 homozygotes over several generations to identify candidate intervals and query our
145 sequencing data to produce a list of candidate mutations. Our studies identify TS alleles

146 for a variety of novel genes not previously classified as essential to development or
147 reproduction, suggesting our collection can be a useful resource for studying essential
148 genes.

149 **Materials and Methods**

150 ***TS strain isolation and sequencing***

151 Set 1 strains originated from the Bowerman lab and were isolated from an ENU
152 mutagenesis protocol (Kemphues *et al.* 1988b) in EU1700, a *lin-2(e1309)* background
153 with an integrated *pie-1*-driven series of GFP fusions to beta-tubulin, histone 2B and a
154 PH domain. Mutagenized animals were screened for conditional TS embryonic lethality
155 by shifting F2 animals as L4s from 15°C to 26°C for 24-36 hours and identifying those
156 with an accumulation of dead embryos. These F2 animals were then shifted back to
157 15°C with the intention of isolating any strains that managed to recover and produce
158 some progeny. Recovered strains were then grown at 15°C for a few generations before
159 rescreening at 26°C to confirm for TS lethality.

160 Set 2 strains originated from the Schnabel lab and were isolated from EMS
161 mutagenesis (Brenner 1974; Ehmke *et al.* 2014) in an wildtype (N2) background. After 8
162 days of growth at 15°C, animals were singled to 96-well plates and visually screened for
163 TS lethality at 25°C. Positive wells were regrouped to new 96-well plates twice before
164 testing a second time at 25°C. Second-round positive strains were plated on individual
165 NGM plates before confirming TS lethality a third time at the non-permissive
166 temperature followed by detailed phenotype analysis.

167 Set 3 strains originated from the Seydoux lab and were isolated from EMS
168 mutagenesis (Kemphues *et al.* 1988b; Golden *et al.* 2000) in a JH50 background (*him-*

169 *3(e1147)IV; lin-2(e1309) axIS36 X*). F2 animals were screened by bleach
170 synchronization of F1 gravid adults followed by upshift at L4 from 15°C to 25°C for 20
171 hours and then down to 15°C for an additional 20 hours. F2 animals accumulating dead
172 embryos were singled to NGM plates at 15°C and examined 3 days later for F3
173 progeny, indicating the presence of a maternal effect embryonic TS lethal mutation.

174 SNV mutations for all 173 strains were sequenced and analyzed using the same
175 methods and custom pipelines described previously for the Million Mutation Project
176 (Thompson *et al.* 2013). SNV calls for each strain can be found in Supplemental Data
177 SD2.

178 ***MIP-MAP Design***

179 The smMIP design for these experiments was altered from those presented in Hyatt *et*
180 *al.* 2013, by placing the 12bp molecular barcode directly on the 3' end of the first read
181 primer sequence. This alteration allows MIP sequencing to proceed in a single-end
182 50bp read that includes the molecular barcode, ligation arm sequence, and 18bp of
183 gap-fill sequence. MIP annealing arms were designed using a custom R script that
184 generated multiple combinations of lengths and locations for the ligation and extension
185 arms. Ligation arm locations were limited to a maximum distance of 18bp either up- or
186 down-stream from the SNV of interest. From these possible variations, the optimal MIP
187 sequences were then chosen based on criteria as described elsewhere (Turner *et al.*
188 2009; O'Roak *et al.* 2012). MIP sequence information for VC20019 probes can be found
189 in Supplemental Data SD3.

190 ***MIP capture protocol.***

191 MIP pools were prepared as described (Turner *et al.* 2009). Briefly, MIPs were pooled in
192 equimolar amounts and treated with 50 units of polynucleotide kinase (PNK, from NEB)
193 for 45 minutes at 37° and then 20 minutes at 80° in a 100ul reaction. The 5'-
194 phosphorylated probes were diluted to 330nM for use in later steps. MIP libraries were
195 based on Hiatt *et al.* 2013. Annealing reactions containing 500ng of target genomic
196 DNA, 330 fmoles of MIP pool, and 1X Ampligase buffer (Epicentre) in 10ul were treated
197 for 3 minutes at 98°, 30 minutes at 85°, 60 minutes at 60° and 120 minutes at 56°. To
198 gap fill the product, 300 pmoles dNTPs, 7.5 umoles Betaine (Sigma), 20 nmoles NAD+,
199 1X Ampligase buffer (Epicentre), 5 units Ampligase, and 2 units Phusion DNA
200 polymerase (NEB) were added to the 10ul anneal reaction and incubated for 120
201 minutes at 56°, and 20 minutes at 72°. To degrade genomic template and any remaining
202 linear MIPs, 20 units Exonuclease I (NEB) and 50 units Exonuclease III (NEB) were
203 added and incubated for 45 minutes at 37°, and 20 minutes at 80°. 10ul of this capture
204 reaction was then amplified by 18 rounds of PCR (15 seconds at 98°, 15 seconds at
205 65°, 45 seconds at 72°) with 1 unit Kapa Hifi Hotstart TAQ, 10 nmoles dNTPs, and 25
206 pmoles each of forward and reverse primers in a 50ul reaction. Libraries were then size
207 selected between 250-450bp and purified with Agencourt AMPure XP beads before
208 sequencing with Illumina sequencing technology.

209 ***MIP sequencing analysis***

210 Sequencing data for each library was analysed using custom scripts written in R. Briefly,
211 each sequenced library was analysed for exact matches to the expected MIP ligation
212 arm sequences at their supposed positions within the read. This subset of matched
213 ligation arm reads was then analysed for specific matches to the expected reference or

214 mutant gap-fill sequence. This set was then trimmed by filtering the quality score at the
215 specific SNV positions to include only those with Phred score \geq Q30. In addition, any
216 duplicated molecular barcodes from this set were analysed to produce a single
217 consensus SNV read. If a duplicated molecular barcode could not find a majority
218 agreement across gap-fill reads, it was removed from further analysis. This uniquely
219 identified final set was used to count reference and SNV representation to produce a
220 percentage calculated using $\text{SNV}/(\text{reference} + \text{SNV})$ counts. Scripts used in this
221 analysis are available upon request.

222 ***Normalization of MIP data***

223 Each MIP was analysed at multiple dilutions to generate a series of curves to account
224 for the difference between any single raw SNV-MIP call and the average of the
225 combined SNV-MIP calls at each site. Normal distributions for each MIP were
226 generated to fit these expected offset curves (height (h), sigma (σ), mean (μ), and offset
227 (o)). These parameters (**Supplemental Data SD4**) were then used to re-adjust raw
228 SNV percentages to within the expected values using the formula

$$229 \quad \text{final.value} = o + h * e^{-0.5((\text{raw.value} - \mu) / \sigma)^2}$$

230 ***Mapping strain determination***

231 Based on Million Mutation Project sequencing data, a group of candidate strains was
232 chosen to identify an appropriate mapping strain. We chose candidates with a minimum
233 SNV density of 500 ± 500 kbp while also minimizing for the occurrence of deleterious
234 alleles (nonsense alleles, high Grantham scores, frameshift mutations etc.). This
235 analysis yielded 28 candidate strains, which were then phenotyped against our
236 reference N2 strain (VC2010) for growth, fecundity and developmental defects. The

237 Million Mutation Project strain VC20019 exhibited growth and development similar to N2
238 and was chosen as a mapping strain.

239 ***sma-9 bulk segregant mapping***

240 *sma-9(tm572)* mutants were backcrossed twice to N2 to remove extraneous mutations
241 before crossing with VC20019 males. Wild type phenotype F1 hermaphrodites were
242 picked to new plates and allowed to self-cross. F2 progeny with the small body (*sma*)
243 phenotype were then identified and pooled in groups of varying sizes to grow on 100mm
244 NGM plates until starved. Pools were isolated for genomic DNA and subsequently
245 analysed using MIPs.

246 ***Temperature sensitive competitive fitness mapping***

247 *hlh-1* and TS mutant mapping was conducted using a competitive fitness assay followed
248 by bulk segregant analysis. Briefly, mutant animals were crossed with either VC20019
249 or DM7448 (VC20019;Ex[*pmyo-3::YFP*]) males. These crosses were then grown at 15°
250 for 24-48 hours before shifting to 23° or 26°. Plates were allowed to starve and an agar
251 chunk was transferred to fresh 100mm OP50-seeded plates. Agar chunk size was
252 varied with 4cm² pieces having the best mapping results.

253 Alternatively, non-starved surviving F1s were individually chosen and plated to
254 fresh 100mm plates. When a population starved, progeny were sub-sampled by chunk-
255 transfer to fresh media. This process continued for at least 6 rounds at the non-
256 permissive temperature, with each transfer generally producing a new generation of
257 animals. The remaining animals on each plate after a transfer were harvested for
258 genomic samples used in subsequent MIP analyses.

259 ***Temperature sensitive mutant selective bulk segregant mapping***

260 TS mutants were crossed with DM7448 males and allowed to grow at 15°. F1 cross-
261 progeny were chosen based on positive YFP expression. F2 animals were transferred
262 singly to 96-well plates (liquid culture) at 15° and grown to starvation. Each well was
263 subsampled and tested for growth at the non-permissive temperature in liquid culture.
264 Populations that failed to thrive at the non-permissive condition were identified and
265 pooled together from the original 15° populations. This pool was then harvested for
266 genomic DNA used in subsequent MIPs analyses.

267 ***Data availability***

268 Strains are available upon request. File SD1 contains strain information including lab
269 origin, allele name, and any available phenotype information. File SD2 contains all
270 variant call information from the set of 173 strains described in this manuscript. Raw
271 sequencing files are available from the NCBI Sequence Read Archive
272 (<http://www.ncbi.nlm.nih.gov/sra>) under accession number SRP018046. File SD3
273 contains all of 89 MIP sequences and genomic targets used in the mapping analysis
274 presented in the manuscript as well as all primer sequences required for library
275 generation and usage on an Illumina-based sequencer. Sample MIP pools for capture
276 can be made available upon request as well. File SD4 contains all of the required
277 information for MIP normalization during the processing and analysis of sequence data.
278 Custom scripts used to analyse sequencing data are also available upon request.

279 **Results**

280 ***Sequencing a collection of embryonic lethal TS mutant strains***

281 We collected a total of 173 embryonic lethal TS mutant strains from three
282 different sources that were generated either in an N2 or *lin-2(e1309)* background using

283 different mutagenesis protocols (EMS or ENU) (Kemphues *et al.* 1988a; Ehmke *et al.*
284 2014) (see also Methods). After mutagenesis, candidate strains were identified through
285 screens for embryonic lethality phenotypes at temperatures of 25°C or 26°C (**Table 1**,
286 **Supplemental Data SD1**). More specifically, the 72 strains in the first set are comprised
287 predominantly of mutant strains with early embryo defects not primarily affecting cell
288 division or with lethality occurring after the 4-cell stage, a small number of strains with
289 some form of sterility, and a small group of strains with low penetrance lethality or
290 unconfirmed TS phenotypes. 67 strains make up the second set and encompass a wide
291 range of TS phenotypes across seven broad classes from general lethality to tissue-
292 specific developmental defects (**Table 1**). The third set consists of 34 strains identified
293 from a mutagenesis screen for maternal effect lethality. Combined, these sets cover a
294 diverse range of phenotypes that could reveal information on essential genes in a
295 variety of pathways.

296 Prior to mapping, we sequenced genomic samples for each strain to a target
297 minimum of 12-fold coverage (with a mean of approximately 14.5-fold coverage). Using
298 the bioinformatic pipeline developed for the Million Mutation Project, we find that on
299 average, each strain carries 328 SNVs including 70 missense changes, nearly 3
300 nonsense mutations, and approximately 1 splice site mutation per strain (**Table 2**), or <
301 1 change in protein coding potential per megabase of sequence within the *C. elegans*
302 genome. Among these changes, missense mutations are the most likely to be
303 associated with temperature-sensitive phenotypes (O'Rourke *et al.* 2011a; Lowry *et al.*
304 2015) although in rare instances other alterations may be responsible (Harfe *et al.*

305 1998). We have included a copy of the final SNV call data as part of our supplemental
306 data (**Supplemental Data SD2**).

307 Given the challenges in *de novo* approaches to identifying the causative alleles
308 from this collection, we turned to genetic mapping to narrow the list of candidates. We
309 sought to develop a method that 1) produced a linkage interval at a resolution that was
310 sufficient to limit the list to a handful of candidates and was comparable in resolution to
311 current methods; 2) replaced the Hawaiian isolate as a mapping strain with one more
312 genomically similar to N2; 3) allowed for a cost-efficient high-throughput workflow. We
313 postulated that mapping with smMIPs in a targeted sequence capture strategy (Hiatt et
314 al., 2013) would provide similar mapping resolution to current WGS methods, allow the
315 use of a strain minimally divergent from N2 and decrease the overall sequencing burden
316 per sample.

317 ***Targeted sequencing of select polymorphisms by MIP-MAP***

318 We re-engineered the original probe structure to adapt the single molecule targeting
319 capabilities of smMIPs for analysing SNV abundance as a method of mapping (hereby
320 referred to as MIP-MAP) (**Figure 1**). Each MIP is composed of an 80bp oligonucleotide
321 with a pair of annealing arms at the 3' (extension arm) and 5' (ligation arm) ends
322 totalling 40 basepairs. Between these two arms is a molecular barcode sequence –
323 randomized for each individual oligonucleotide – as well as a common MIP backbone
324 used in library preparation and amplification steps (**Figure 1A**). Each MIP is designed to
325 target a locus where a known variant is located within the first 18 basepairs of the 100-
326 150 basepairs of gap-fill sequence (**Figure 1AB**). The gap-fill sequence also serves to
327 unambiguously map the recovered sequence. The library is then amplified, at which

328 time an experimental index sequence is incorporated (**Figure 1C**). This design places
329 all the vital information within a single-end 50 basepair Illumina sequencing read
330 (**Figure 1D**). After sequencing, each sample library samples is demultiplexed based on
331 an experimental index after which molecular barcodes and gap-fill sequence data are
332 used to eliminate PCR duplicates (**Figure 1E, Methods**). The remaining unique
333 sequencing information is used to determine a MIP target's SNV representation versus
334 the total reads matching that locus as calculated by

$$335 \frac{SNV \text{ captures}}{Reference \text{ captures} + SNV \text{ captures}}$$

336 With a targeted sequencing strategy in hand, we next examined the parameters
337 that would influence our selection of probes across the *C. elegans* genome: mapping
338 resolution, mapping strain choice, and MIP accuracy. Mapping resolution is generally
339 limited by the number of F2 animals pooled rather than SNV density. With just 100 F2s
340 picked, average resolution is no more than 1 cM or about 1-3 Mb depending on the
341 region of the genome (Barnes *et al.* 1995; Rockman and Kruglyak 2009; Doitsidou *et al.*
342 2010). Accordingly, sampling ~100 SNV markers across the genome would match the
343 expected genetic resolution. At this density, two hundred thousand unique sequencing
344 reads would provide deep coverage of each SNV, allowing the accurate estimation of
345 even relatively rare linked SNVs – a vast improvement in sequencing burden versus
346 current paradigms.

347 Decreasing the density requirement of our mapping SNVs also permitted us to
348 survey the Million Mutation Project strains for potential alternatives to the Hawaiian
349 strain. The Million Mutation Project strains average ~400 mutations each, yet many are
350 roughly wild type in appearance. We screened for a collection of strains that carried

351 minimal numbers of predicted deleterious or missense coding variants and a relatively
352 even distribution of SNVs across the genome at a frequency of at least 1 per 500
353 kilobases. From 28 candidates, we identified a strain, VC20019, with developmental
354 timing and fecundity similar to that of our N2 wild type strain (VC2010). VC20019, also
355 referred to as the mapping strain, has a total of 269 mutations, 3 nonsense, and 51
356 missense alleles, thus having enough genomic diversity compared to N2 to meet our
357 molecular marker density requirements while having a reduced likelihood of potentially
358 negative interactions.

359 We designed MIPs targeting 96 SNV sites across the VC20019 genome,
360 generating intervals spaced an average of 0.98 ± 0.36 Mbp apart (**see Supplemental**
361 **Data SD3**). To identify the efficiency of annealing and overall fidelity of these probes,
362 we generated a series of VC20019 and N2 genomic DNA mixtures for capture reaction
363 and sequencing. From these libraries, we identified sequence-specific biases for each
364 MIP and generated normalization curves for subsequent capture/sequencing reactions
365 (**Supplemental Figure S1, Supplemental Data SD4, Methods**). We removed seven
366 VC20019 MIPs that consistently produced low sequencing read counts, high ratios of
367 non-specificity, or unpredictable allelic biases, bringing our total probe count to 89
368 probes (1.06 ± 0.43 Mbp intervals).

369 We also considered that the frequency for linked and therefore rare SNVs could
370 be adversely impacted by false-positive MIP-MAP reads. To gauge the impact of false-
371 positives on the accuracy of our calls, we designed an additional 176 MIP-MAP targets
372 based on SNVs in 44 strains from the Million Mutation Project collection. We analysed
373 the final SNV representations of each MIP capture library from a genomic DNA mixture

374 that excluded the selected Million Mutation Project target strains. We sequenced this set
375 of MIPs to an average depth of ~35K reads. We removed data from poorly amplifying
376 MIPs, leaving 174 MIP target loci that were expected to have 0% mutant SNV
377 representation. MIP reads with positive mutant calls must originate from gap-fill, PCR,
378 or sequencing errors and the overall percent mutant SNV representation was calculated
379 as the false-positive rate. As a further quality analysis step, we filtered out low-quality
380 sequencing reads at specific SNV sites and calculated the average false positive rate
381 was 0.0122% \pm 0.0129%. (**Supplemental Figure S2A, Methods**).

382 Lastly, we determined the sequencing efficiency of the MIP-MAP method by
383 comparing the population of unique molecular barcodes for each MIP versus the total
384 reads identified for that MIP. For each MIP in our false positive set of 174 MIP targets,
385 the mean number of sequencing reads per unique molecular barcode was 1.05 with a
386 mean depth of 35,000 reads per MIP. We analysed an additional 22 libraries from
387 another sequencing run, consisting of the 89 mapping strain MIPs in each library. The
388 mean reads per barcode across 1,958 data points was 1.01 with a mean depth of 7000
389 reads per MIP (**Supplemental Figure S2B**). These results suggest that we are
390 capturing a nearly 1:1 representation of the genomic sample, with few redundant reads
391 from non-unique capture sequences.

392 In summary, we re-engineered the smMIP format to produce MIP-MAP which targets
393 the genome with high fidelity and high sequencing efficiency at a mapping resolution
394 comparable to current methods. We identified a polymorphic strain similar to N2 with no
395 observed adverse phenotypes and a reduced likelihood of negative genetic interactions.
396 We next turned to examining the MIP-MAP protocol in mapping known mutant alleles

397 and then uncharacterized strains of varying difficulty, such as those from the TS
398 collection.

399 **Mapping *sma-9* and *hlh-1* mutations using MIP-MAP**

400 To investigate the mapping capabilities of the MIP-MAP method, we first focused on
401 mapping with two distinct phenotypes; a clearly identifiable *sma-9* small body mutant
402 and the well-characterized *hlh-1* TS lethal mutant. We proceeded by using the LW478,
403 *sma-9(tm572)X*, strain in a standard bulk segregant strategy similar to WGS (**Figure**
404 **2A**). Briefly, we crossed *sma-9(tm752)X* hermaphrodites with mapping strain males and
405 from F1 cross progeny chose groups of F2 *sma* phenotype animals (10, 25, 50, 75, 100,
406 200) to grow until starvation on OP50-seeded 100mm NGM plates before isolating
407 genomic DNA for MIP-MAP capture reaction and sequencing (**Figure 2, Supplemental**
408 **Figure S3**). Our expectation with such a strategy was to observe a region where
409 VC20019 SNVs were nearly depleted due to linkage disequilibrium in selecting for the
410 *sma-9* allele from the mutant strain. We generated MIP-MAP data by growing
411 populations seeded with 10 *Sma* phenotype F2s. These replicate samples had more
412 background signal across the genome than subsequent experiments and our interval of
413 interest was 1.5Mbp and 4.2Mbp in these samples (**Figure 2B**). Seeding populations
414 with 75 *Sma* phenotype F2s yielded similar mapping profiles that narrowed the likely
415 interval to 2.6Mbp (**Supplemental Figure S3C**). Likewise, two mapping sets seeded
416 with 200 F2 animals also consistently identified a 2.6Mbp interval containing *sma-9*
417 (**Figure 2C**). Upon closer inspection of our *sma-9* mapping intervals, we observed two
418 sites ~590kb apart that typically share nearly identical SNV representation, suggesting
419 linkage disequilibrium. It is not surprising, therefore, to find that *sma-9* is located

420 equidistantly between these points, making it less likely to generate a sharp single-point
421 peak even with a high number of F2s as input (**Figure 2D**). On average, the MIP targets
422 in our 75-F2 sets had 3778 unique reads, suggesting that each F2 haplotype was
423 sampled an average of 25 times. Therefore, the MIP-MAP technique can sufficiently
424 capture an accurate representation of the recombinant landscape across the population.

425 We next explored if MIP-MAP was applicable to more challenging mutant
426 phenotypes. For instance, isolating homozygotes of non-conditional lethal/sterility
427 alleles or identifying the correct homozygous F2 populations of low penetrance alleles
428 may not always be feasible. We approached this scenario with a competitive fitness
429 mapping assay (**Figure 2E**). We hypothesized that we could exploit the fitness defect of
430 an allele during the mapping process and gradually replace it with a more “fit” version
431 from the mapping strain. Using the *hlh-1(cc561)*II temperature sensitive embryonic
432 lethal allele, we attempted to passively map its locus by using non-permissive
433 temperatures to select against this allele over several generations. In contrast to the
434 *sma-9* mapping, we expected to identify a region of interest by the fixation of VC20019
435 SNVs towards 100% representation within the mapping profile. We refined the
436 parameters of the competitive fitness MIP-MAP method by varying the number of
437 progeny passaged at each generation. Briefly, we crossed *hlh-1* animals to VC20019
438 males at 15°C on 50mm NGM plates for approximately 24 hours before shifting these
439 plates to between 23°C and 26°C for the remainder of the experiment. We expected
440 only cross progeny F1s would thrive at this temperature. F1s were picked and self-
441 fertilized before randomly picking L1- or L2-staged F2 larvae. With each subsequent
442 generation, a subpopulation (50, 100, 200 or 400 animals) was passaged to a new plate

443 for expansion (**Figure 2F, Supplemental Figure S4A-C**). We continued in this manner
444 for 4-7 generations, collecting samples for MIP-MAP library preparation and sequencing
445 at each passaging.

446 Based on our observations from mapping *sma-9*, we hypothesized that at least
447 75 F2 animals (upwards of 150 recombinant chromosomes) homozygous for our region
448 of interest, were required to consistently generate the smallest mapping interval
449 compatible with the density of our markers. The mapping profile from a competitive
450 fitness assay might therefore be influenced by the number of unique recombinant
451 chromosomes that carry our region of interest in each generation. We reasoned that
452 subsampling in small numbers could limit recombinant representation and increase the
453 incidence of bottlenecking within the population. A chance reduction in recombinant
454 chromosome diversity could lead to quick fixation of a region of interest but with
455 reduced resolution. Conversely, subsampling in larger numbers could give a greater
456 range of recombinant genomes, reduce bottlenecking and generate better resolution
457 while likely requiring more time to reach fixation of the more “fit” wild type region.

458 With this in mind we varied our passaging size and observed that an
459 overrepresented region of VC20019-associated SNVs on LGII with varying degrees of
460 resolution (**Figure 2FG**). This candidate interval included the location of *hlh-1* and
461 suggested that we had correctly mapped the *cc561* allele. The smallest possible
462 mapping intervals encompassing the location of *cc561* are 1.7 and 1.9Mbp. When
463 passaging smaller populations (i.e., 50 animals per generation), fixation occurred
464 quickly but in some cases resulted in a larger mapping interval across LGII (**Figure 2F,**
465 **Supplemental Figure S4A**). The resolution of the MIP-MAP approach, however,

466 improved in the populations seeded with ≥ 200 randomly chosen F2s (**Supplemental**
467 **Figure S4BC**). Thus, sampling more animals generated the optimal minimal mapping
468 interval of 1.71 Mbp with only a one to two generation increase required for fixation.
469 To exclude the possibility of a sampling bias for healthier animals when hand-picking
470 our *hlh-1* mapping recombinants, we replicated these experiments with an alternate
471 procedure by transferring completely randomized populations via agar chunks of
472 starved animals for 5-9 generations (**Figure 2G, Supplemental Figure S4D-F,**
473 **Methods**). We again observed that larger transfer sizes associated with better
474 resolution but this approach also generated a more pronounced delay of 2-8 additional
475 generations before fixation.

476 We further validated the competitive fitness assay by investigating the general
477 fitness of the mapping strain in a cross to the reference strain VC2010 and growing
478 cross-progeny at a range of temperatures. We observed only partial fixation of LGV at
479 high temperature over 8 generations but other loci preferences within the genome
480 appeared negligible (**Supplemental Figure S5A-C**), thereby confirming that our *hlh-1*
481 results were not an artefact of the mapping strain used. We could therefore reliably
482 identify a strong temperature-sensitive lethal allele through a simplified competitive
483 fitness selection process. In addition, we noted the weak beginnings of fixation on LGII
484 when growing an *hlh-1* mapping population at 15°C (**Supplemental Figure S6**),
485 suggesting a small population growth defect was conferred by the *cc561* allele at the
486 so-called permissive growth temperature. We next turned to see if the MIP-MAP
487 competitive fitness assay could be applied to genetically mapping uncharacterized TS
488 lethal alleles.

489 **Using MIP-MAP to identify TS mutations that confer lethal phenotypes**

490 Our success with *hlh-1* led us to hypothesize that we could, with good resolution, map
491 mutant alleles from our large collection of sequenced TS lethal strains. Unlike our
492 previous trials, these mutants carry hundreds of variants – any one of which may
493 influence the outcomes of a competitive fitness assay. Furthermore, TS alleles of low or
494 weak penetrance could also prove troublesome given a reduced selection coefficient
495 during the propagation process. Regardless of these factors, from the collection of TS
496 mutants we chose 15 strains (**Table 3**) of high sequencing depth but varying
497 phenotypes to attempt genetically mapping with MIP-MAP.

498 Since these highly mutated strains could harbor mutations that reduce brood size or
499 mating efficiency, we modified our process to ensure the picking of cross-progeny by
500 crossing our TS mutants with DM7448 (a VC20019 strain carrying an
501 extrachromosomal *pmyo-3::YFP* marker) males and picking YFP-positive F1
502 hermaphrodites. These F1s were subjected to the previously-described competitive
503 fitness assay via chunk- or wash-transfer and in some instances we supplemented our
504 analysis with a multi-well liquid format of bulk segregant mapping for clarification
505 (**Figure 3A-D, Methods**). Our mapping results (**summarized in Figure 3E, Table 4**)
506 can be categorized into increasing levels of analysis difficulty: clear single TS loci
507 (**Figure 4AB, Supplemental Figure S7**); single TS loci with additional partial fitness
508 defective loci (**Figure 4CD, Supplemental Figure S8**); and multi-locus profiles resolved
509 by additional investigations (**Figure 4E-G, Figure 5A-C, Supplemental Figure S9**).

510 Three strains yielded clean, single-locus profiles with only one or two TS
511 candidate alleles in each interval. Among these strains, VC50174 mapped to an

512 adenine to thymine transversion resulting in an E196D mutation in the fifth exon of
513 *C09D4.3* (**Figure 4AB**), a gene with no prior phenotype data to suggest an essential
514 role in development. Available RNAseq data reports *C09D4.3* transcript abundance
515 peaking at the L4 stage, most especially in males (Hutter *et al.* 2009; Boeck *et al.* 2016)
516 and is suggested to have preferential expression in the germline (Grün *et al.* 2014).
517 Furthermore, we observed that VC50174 animals exhibited a failure of fertilization and
518 vacuolated spermatozoa. VC50022 mapped to an interval with a coding mutation in *tbb-*
519 *2*; one of two nearly functionally redundant *C. elegans* β -tubulin genes. Although not
520 entirely essential, multiple labs have identified semi-dominant TS-lethal alleles for this
521 gene (Wright AJ 2003; Ellis *et al.* 2004). In particular, the *tbb-2* mutation in VC50174 is
522 located within the same exon as a semi-dominant embryonic lethal allele *t1623* (Gönczy
523 *et al.* 1999). The other possible candidate, *F56D2.5*, is at the edge of this interval,
524 shows low expression throughout embryogenesis (Boeck *et al.* 2016) and has no useful
525 RNAi phenotype data. VC50260 maps to an interval encompassing a single coding
526 mutation for *npp-8* which has also been reported as an essential gene with a
527 documented TS allele, *ax146* (Asencio *et al.* 2012).

528 In three instances, our TS mapping profiles suggested the presence of additional
529 weak low-fitness alleles at secondary loci (**Figure 4CD, Supplemental Figure S8**). In
530 these cases, we combined mapping data from multiple time points with the expectation
531 that the TS-associated locus was the least competitive at non-permissive temperatures
532 and would therefore fixate at a faster rate. When we were unable to ambiguously
533 resolve the correct chromosome, we further mapped the strain with a liquid bulk-
534 segregant assay (**Figure 3A**). By growing singled F2 populations in 96-well liquid format

535 we could subsample each and phenotype for TS lethality, going back to the original
536 populations later to pool for MIP analysis. This method would generally produce a low
537 number of TS F2s (**Supplemental Table S1**) but could yield a single mapping signal
538 across the *C. elegans* genome. Of particular interest, the strain VC50383 yielded a
539 number of a candidates within the mapping interval. The most likely candidate,
540 however, was a *let-2* exon 16 guanine to adenine transition resulting in a G1385E
541 change which was confirmed through complementation testing. Coincidentally, another
542 set 3 strain, VC50380, also mapped to an interval that included another *let-2* guanine to
543 adenine transition (G1110E) in exon 14. *In toto*, there are 9 separate *let-2* mutations
544 from our collection that could further elucidate the complex nature of this gene (Meneely
545 and Herman 1981).

546 Our final group of nine TS strains represent multi-locus mapping profiles that fixed to
547 similar proportions within a short period of time (**Figure 4EFG, Supplemental Figure**
548 **S9**). These strains may carry additional mutations that confer disadvantages during the
549 competitive mapping process. Again, a combination of comparing fixation rates at each
550 locus and the liquid segregant mapping assay were used to resolve the TS-associated
551 locus. In three of these strains, our mappings included a clearly marked interval on
552 chromosome X (**Figure 4EFG**) which encompasses the *lin-2* gene. Strains from both
553 set 1 and set 3 harbor the *lin-2(e1309)* mutation which causes egg retention and
554 reduced brood size – negative influences in a competitive fitness assay. From these
555 strains in particular, VC50182 yielded two candidates within the interval on LGIII of
556 which *cts-1* (a citrate synthase) is the likely prime candidate. VC50182 animals

557 exhibited late arrest at the 2-fold stage, while reported *cts-1* RNAi phenotypes include
558 embryonic lethality and sterility (Kamath *et al.* 2003; Melo and Ruvkun 2012).

559 Also of note, in mapping the strain VC50141, we observed the concomitant but
560 incomplete fixation of two loci – one each on LGI and LGII – when using our competitive
561 fitness assay (**Figure 5A**). Similarly, the liquid-format bulk segregant assay identified
562 the same two linkage groups at approximately 30% mapping allele representation
563 versus an expected 0% generated from 75 TS-positive wells in 387 randomly chosen
564 F2s. This phenomenon suggested the presence of two TS alleles. The LGI interval
565 contained only a single coding change which corresponded to a guanine to adenine
566 transition (E994K) in the RNA polymerase homolog *ego-1* (Qiao *et al.* 1995; Smardon *et*
567 *al.* 2000) (**Figure 5B**). There are no previously reported conditional-lethal alleles for
568 *ego-1*. The LGII locus contained 4 possible candidates including a cytosine to thymine
569 (P632S) *zyg-9* mutation (**Figure 5C**). There are a large number of reported TS alleles
570 located across *zyg-9* (Bellanger *et al.* 2007), suggesting it is a good candidate for
571 VC50141. To confirm whether these two loci were independent TS alleles or if they
572 perhaps interacted to produce a synthetic TS lethal effect, we proceeded to isolate
573 these two alleles into separate strains (**Methods**) followed by scoring for lethal
574 phenotypes at 26°C. Based on our observations we concluded the presence of two TS-
575 lethal phenotypes, sterility and embryonic lethality, linked to the *ego-1* and *zyg-9* alleles
576 respectively. Thus our mapping methods were able to identify both *ego-1(ax458)I* and a
577 *zyg-9(ax3101)II* as the primary candidates for two independent TS-lethal mutations in
578 this strain.

579 Overall, for our set of TS strains, this method yielded relatively small mapping
580 intervals that contained greatly reduced numbers of candidate alleles (**Table 4**) -
581 simplifying the process of finding the causative allele. Of 15 candidate strains, 6 were
582 mapped with only the competitive fitness approach while 9 had additional bulk
583 segregant mapping performed, with the special case of VC50141 requiring additional
584 intervention (**Figure 3E**). In three strains, our mapping intervals contained only a single
585 relevant coding sequence candidate, while in twelve intervals we narrowed candidates
586 to a primary variant based on a combination of gene description, expression pattern,
587 RNAseq, and RNAi phenotypes data from Wormbase and strain phenotype
588 observations (**Table 5**). Of the sixteen mapping intervals identified only one (VC50352;
589 LGIV) could not be narrowed to a primary candidate based solely on a comparison of
590 the strain phenotype, mapping data and available literature. Of the fifteen TS intervals
591 with primary candidate genes, six have no reported TS alleles listed in Wormbase and
592 one (*C09D4.3*) has no previously reported embryonic lethal phenotypes. Of the other
593 candidates with no available TS alleles, *ego-1*, *let-805*, *mog-4*, and *par-5* alleles with
594 known phenotypes are only available from the Caenorhabditis Genetics Center as
595 viable strains with genetic balancers. The gene *cts-1* only has a single deletion strain
596 available which is also listed as sterile/lethal.

597 **Discussion**

598 Our large collection of sequenced *C.elegans* conditional mutant strains offers a rare
599 opportunity to study essential genes on a larger scale. A high-throughput mapping of
600 the causative alleles in this collection, however, is not a trivial task. We began our
601 analysis of these strains with the goal to generate a robust mapping strategy that was

602 affordable, high-throughput, comparable in resolution to current mapping paradigms and
603 flexible enough for the community's needs. Using single-molecule molecular inversion
604 probes we developed the MIP-MAP strategy which relies on targeted loci sequencing of
605 the *C.elegans* genome at great depth. MIP-MAP is differentiated from other mapping
606 methods by 1) its adaptability to different mapping and segregant selection schema, 2)
607 its independence from the usage of the Hawaiian strain CB4856, 3) by the reduction in
608 sequencing and data generation while maintaining 1Mb resolution, and 4) its unique
609 advantage in mapping the more recent collections of sequenced strains.

610 MIP-MAP successfully identified a phenotype-associated locus in the direct-
611 selection of *sma-9* F2 segregants – a common mapping paradigm. However, with the
612 *hlh-1(cc561)* temperature sensitive lethal allele, we successfully used a competitive
613 selection scheme to identify the *hlh-1* locus versus an alternative approach such as
614 picking dead larvae for WGS (Smith *et al.* 2016). This indirect approach to identifying
615 lethal loci may also useful in gauging the rate of fixation across multiple conditions to
616 elucidate mutation penetrance or gene function. In fact, our *hlh-1* data suggested that
617 even at the permissive temperature of 16°, *hlh-1(cc561)* animals have some level of
618 decreased fitness, despite a lack of reported lethality in this condition (Harfe *et al.*
619 1998).

620 Our results demonstrated the flexibility of this method to work with non-standard
621 mapping strains. We leveraged a Million Mutation Project strain with a mere 269 SNVs
622 and targeted strain-specific markers at 89 loci, using them to successfully map *sma-9*
623 and *hlh-1* mutant alleles to within ~2Mb intervals. Furthermore, we have shown with
624 VC20019 that the Million Mutation Project is also a library that contains polymorphic

625 mapping strains with fewer potential issues than more divergent genomes like the
626 Hawaiian strain. Based on our results it is even plausible that one could identify Million
627 Mutation Project strains compatible for mapping synthetic phenotypes. This strategy
628 would reduce the additional steps required in homozygosing background mutations
629 used in forward genetic screens – simplifying the cloning process.

630 In terms of cost, the MIPs themselves are a “one-time” purchase where a single
631 order at 25nM scale provides more than 2 million capture reactions per probe. MIP-MAP
632 also offers comparable resolution to WGS with far less extraneous sequencing data. For
633 example, each MIP-MAP library can be sequenced to an average 200,000 reads to
634 obtain mapping data. In contrast, whole genome sequencing at recent standards of
635 ~20X coverage (50bp, paired end) requires approximately 20 million reads - 2 additional
636 orders of magnitude with most of the sequence being reference genome data. It is
637 preferable to heavily multiplex samples during the sequencing process and thus lower
638 overall cost per sample. Choosing the appropriate mapping strategy (bulk segregant
639 versus competitive fitness) one could easily MIP-MAP fifty to one hundred times the
640 number of strains on a single HiSeq lane compared to conventional whole genome
641 sequencing mapping. The overall cost savings per strain and capabilities of MIP-MAP
642 are therefore suitable to a high-throughput analysis of collections such as our TS strains
643 or the Million Mutation Project. More recent large-scale genome wide association
644 studies may also benefit from this method when verifying hits from amongst multiple
645 candidate associations (Cook *et al.* 2016).

646 Although the MIP-MAP method was developed with sequenced mutant
647 collections in mind, it should be noted that it is also an advantageous companion

648 approach to standard genome sequencing and mapping of *de novo* strains. Million
649 Mutation Project findings suggest genome sequencing to a 12-fold depth (Thompson *et*
650 *al.* 2013) is sufficient to confidently identify SNVs and other mutations in a heavily
651 mutagenized background. In combination with MIP-MAP, this still represents a potential
652 sequencing savings of more than 7 million reads. Alternatively, once a mutation is
653 mapped to a particular locus, the genomic interval pull-down sequencing method
654 (O'Rourke *et al.* 2011b) can target chromosomal regions using *C. elegans* genomic
655 Fosmids. These isolated intervals can be sequenced to identify mutations for that
656 specific genomic interval. Thus, the throughput scalability of MIP-MAP still provides a
657 significant advantage in mapping novel strains.

658 Intriguingly, a number of our mappings identified the presence of additional non-
659 TS loci that conferred lowered fitness rates on the TS strains, suggesting that our TS
660 collection holds a diverse library of additional mutations worth cloning. In a similar
661 fashion, MIP-MAP could be useful in high-throughput mining of the Million Mutation
662 Project for specific phenotypes in a form of reverse-genetics screen. For instance,
663 rather than testing for temperature-sensitive lethality, MMP strains could be queried for
664 phenotypes related to food source, plate conditions, or RNAi interactions. More broadly,
665 MIP-MAP demonstrates, on a focused level, a method of barcoding strains within the
666 context of a mixed population. This low-cost high-throughput form of analysis could be
667 applied, for instance, to the analysis of mixed pools of multiple strains to track variant
668 fitness across a diverse set of conditions. Targeting strains in this manner might reveal
669 the functions of many novel genes. Given that the Million Mutation Project strains have

670 already been sequenced, using MIP-MAP to identify the cause of these phenotypes
671 would generate far less sequencing data than WGS methods.

672 Reviewing our TS-mapping workflow, five of nine strains were bulk segregant
673 mapped with more than 60 F2s chosen as TS-positive but none of these generated a
674 concise interval – rather they complemented the sharp intervals obtained by the
675 competitive fitness approach. Curiously, in three of these five sets the chosen F2
676 population was far less than the expected 25% ratio for a homozygous recessive
677 mutation (**Supplemental Table S1**). Indeed, in the remaining four strains mapped with
678 F2 counts below 60, these animals were identified at low frequency from very large
679 populations of potential F2s – further suggesting issues in either identifying the TS
680 phenotype or reduced phenotype penetrance in these populations. In *toto*, using bulk
681 segregant analysis we observed only three strains (VC50028, VC50374, and VC50380)
682 that generated mapping intervals comparable in resolution to data from the competitive
683 fitness approach. In that respect, a MIP-MAP competitive fitness assay over many
684 generations offers the ability to identify mutant loci that are of incomplete penetrance
685 and may prove problematic in standard bulk segregant mapping. From this we conclude
686 that for the remainder of our TS strains – especially those of apparently lower
687 penetrance – it is more ideal to complete both mapping approaches in parallel with the
688 benefit of reducing repetition of mapping crosses in downstream workflows.

689 The collection of *C. elegans* TS strains we present here is, to our knowledge, the
690 largest, fully-sequenced collection to date. Based on our mapping results we suspect
691 that many new and interesting alleles can be mined from these strains. For example,
692 our mapping of the strain VC50174 identified a sperm-specific essential gene *C09D4.3*.

693 Recent studies suggested that sperm-targeted RNAi screens may be inefficient due to
694 competition between endogenous and exogenous RNAi pathways. This hypothesis
695 supports the literature from multiple RNAi screens that report no embryonic lethal
696 phenotypes for *C09D4.3* (Fraser 2000; Maeda *et al.* 2001; Sönnichsen *et al.* 2005).
697 Finding this TS allele within our small set suggests that our overall collection of strains
698 has the potential to not only yield the first TS alleles for a number of essential genes but
699 also the potential to identify new essential genes not discovered by standard RNAi
700 screens. Further analysis of our collection as a whole may identify that genes with a
701 high frequency of mutant alleles in TS loci could help to prioritize candidate intervals.
702 Our data also suggests that ~6% of the strains in this collection may carry TS alleles for
703 genes not previously characterized as essential.

704 In summary, we present a large collection of sequenced temperature sensitive
705 mutants housing a variety of embryonic-lethal and sterility phenotypes. In conjunction
706 with this, we demonstrate the MIP-MAP methodology as a robust, cost-efficient, and
707 high-throughput genetic mapping format. It allows for a wide range of flexible sample
708 generation schema that can be implemented with relative ease. We believe MIP-MAP
709 can even identify alleles with weak penetrance phenotypes given the proper selection
710 conditions. Taken together, these two tools present a unique opportunity for the
711 nematode research community to investigate a diverse library of mutant strains and to
712 map these and other mutant alleles in a unique and accessible manner. We believe
713 both of these resources will be a beneficial addition to the nematode and scientific
714 community at large.

715

716 **Acknowledgments**

717 We thank Emily Turner, Joseph Hiatt and Choli Li for advice on MIP library preparation
718 and sequencing. C.A.M is supported by the Canadian Institutes for Health Research
719 MFE-135408. N.M is supported by the Internationales Graduierten Kolleg
720 Niedersachsen. Work in the B.B. laboratory was supported by an R01 grant GM114053
721 from the NIH. Work in the G.S. laboratory was supported by an R01 grant HD37047
722 from the NIH. G.S. is an Investigator of the Howard Hughes Medical Institute. Work in
723 the D.G.M laboratory was supported by the Canadian Institutes for Health Research
724 and the NHGRI (through R.H.W). D.G.M is a Senior Fellow of the Canadian Institute for
725 Advanced Research. Work in the R.H.W laboratory was supported by an ARRA GO
726 grant HG005921 from the NHGRI, an R21 grant HG007201-02 from the NIH, and by the
727 William H. Gates Chair of Biomedical Sciences.

728

729 **References**

- 730 Andersen E. C., Bloom J. S., Gerke J. P., Kruglyak L., 2014 A Variant in the
731 Neuropeptide Receptor npr-1 is a Major Determinant of *Caenorhabditis elegans*
732 Growth and Physiology. *PLoS Genet.* 10.
- 733 Asencio C., Davidson I. F., Santarella-Mellwig R., Ly-Hartig T. B. N., Mall M., *et al.*,
734 2012 Coordination of kinase and phosphatase activities by Lem4 enables nuclear
735 envelope reassembly during mitosis. *Cell* 150: 122–135.
- 736 Barnes T. M., Kohara Y., Coulson A., Hekimi S., 1995 Meiotic recombination, noncoding
737 DNA and genomic organization in *Caenorhabditis elegans*. *Genetics* 141: 159–79.
- 738 Bellanger J.-M., Carter J. C., Phillips J. B., Canard C., Bowerman B., *et al.*, 2007 ZYG-
739 9, TAC-1 and ZYG-8 together ensure correct microtubule function throughout the
740 cell cycle of *C. elegans* embryos. *J. Cell Sci.* 120: 2963–2973.
- 741 Boeck M. E., Huynh C., Gevirtzman L., Thompson O. A., Wang G., *et al.*, 2016 The
742 time-resolved transcriptome of *C. Elegans*. *Genome Res.* 26: 1441–1450.
- 743 Bono M. De, Bargmann C. I., 1998 Natural variation in a neuropeptide Y receptor
744 homolog modifies social behavior and food response in *C. elegans*. *Cell* 94: 679–
745 689.
- 746 Brenner S., 1974 *Caenorhabditis elegans*. *Methods* 77: 71–94.
- 747 Claycomb J. M., Batista P. J., Pang K. M., Gu W., Vasale J. J., *et al.*, 2009 The
748 Argonaute CSR-1 and Its 22G-RNA Cofactors Are Required for Holocentric
749 Chromosome Segregation. *Cell* 139: 123–134.
- 750 Cook D. E., Zdraljevic S., Roberts J. P., Andersen E. C., 2016 CeNDR , the
751 *Caenorhabditis elegans* natural diversity resource. : 1–8.

- 752 Cox G. N., Laufer J. S., Kusch M., Edgar R. S., 1980 Genetic and Phenotypic
753 Characterization of Roller Mutants of *Caenorhabditis elegans*. *Genetics* 95: 317–
754 39.
- 755 Davis M. W., Hammarlund M., Harrach T., Hullett P., Olsen S., *et al.*, 2005 Rapid single
756 nucleotide polymorphism mapping in *C. elegans*. *BMC Genomics* 6: 118.
- 757 Doitsidou M., Poole R. J., Sarin S., Bigelow H., Hobert O., 2010 *C. elegans* mutant
758 identification with a one-step whole-genome-sequencing and SNP mapping
759 strategy. *PLoS One* 5.
- 760 Ehmke M., Luthe K., Schnabel R., Döring F., 2014 S-adenosyl methionine synthetase
761 1 limits fat storage in *Caenorhabditis elegans*. *Genes Nutr.* 9.
- 762 Ellis G. C., Phillips J. B., O'Rourke S., Lyczak R., Bowerman B., 2004 Maternally
763 expressed and partially redundant beta-tubulins in *Caenorhabditis elegans* are
764 autoregulated. *J. Cell Sci.* 117: 457–464.
- 765 Encalada S. E., Martin P. R., Phillips J. B., Lyczak R., Hamill D. R., *et al.*, 2000 DNA
766 replication defects delay cell division and disrupt cell polarity in early
767 *Caenorhabditis elegans* embryos. *Dev. Biol.* 228: 225–238.
- 768 Flibotte S., Edgley M. L., Chaudhry I., Taylor J., Neil S. E., *et al.*, 2010 Whole-genome
769 profiling of mutagenesis in *Caenorhabditis elegans*. *Genetics* 185: 431–441.
- 770 Fraser A. G., 2000 Functional genomic analysis of *C. elegans* chromosome I by
771 systematic RNA interference. *Nature* 408: 325–330.
- 772 Golden A., Sadler P. L., Wallenfang M. R., Schumacher J. M., Hamill D. R., *et al.*, 2000
773 Metaphase to anaphase (*mat*) transition-defective mutants in *Caenorhabditis*
774 *elegans*. *J. Cell Biol.* 151: 1469–1482.

- 775 Gönczy P., Schnabel H., Kaletta T., Amores A. D., Hyman T., *et al.*, 1999 Dissection of
776 cell division processes in the one cell stage *Caenorhabditis elegans* embryo by
777 mutational analysis. *J. Cell Biol.* 144: 927–946.
- 778 Graham P. L., Schedl T., Kimble J., 1993 More *mog* genes that influence the switch
779 from spermatogenesis to oogenesis in the hermaphrodite germ line of
780 *Caenorhabditis elegans*. *Dev. Genet.* 14: 471–484.
- 781 Grün D., Kirchner M., Thierfelder N., Stoeckius M., Selbach M., *et al.*, 2014
782 Conservation of mRNA and protein expression during development of *C. elegans*.
783 *Cell Rep.* 6: 565–77.
- 784 Harfe B. D., Branda C. S., Krause M., Stern M. J., Fire A., 1998 MyoD and the
785 specification of muscle and non-muscle fates during postembryonic development of
786 the *C. elegans* mesoderm. *Development* 125: 2479–2488.
- 787 Hiatt J. B., Pritchard C. C., Salipante S. J., O’Roak B. J., Shendure J., 2013 Single
788 molecule molecular inversion probes for targeted, high-accuracy detection of low-
789 frequency variation. *Genome Res.* 23: 843–854.
- 790 Hutter H., Ng M.-P., Chen N., 2009 GExplore: a web server for integrated queries of
791 protein domains, gene expression and mutant phenotypes. *BMC Genomics* 10:
792 529.
- 793 Jaramillo-Lambert A., Fuchsman A. S., Fabritius A. S., Smith H. E., Golden A., 2015
794 Rapid and Efficient Identification of *Caenorhabditis elegans* Legacy Mutations
795 Using Hawaiian SNP-Based Mapping and Whole Genome Sequencing. *G3* 5:
796 1007–1019.
- 797 Kamath R. S., Fraser A. G., Dong Y., Poulin G., Durbin R., *et al.*, 2003 Systematic

798 functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421:
799 231–7.

800 Kemphues K. J., Kusch M., Wolf N., 1988a Maternal-effect lethal mutations on linkage
801 group II of *Caenorhabditis elegans*. *Genetics* 120: 977–986.

802 Kemphues K. J., Priess J. R., Morton D. G., Cheng N. S., 1988b Identification of genes
803 required for cytoplasmic localization in early *C. elegans* embryos. *Cell* 52: 311–320.

804 Kiezun A., Garimella K., Do R., Stitzel N. O., Benjamin M., *et al.*, 2013 Exome
805 sequencing and the genetic basis of complex traits. *NIH Public Access* 44: 623–
806 630.

807 Lowry J., Yochem J., Chuang C.-H., Sugioka K., Connolly A. A., *et al.*, 2015 High-
808 Throughput Cloning of Temperature-Sensitive *Caenorhabditis elegans* Mutants with
809 Adult Syncytial Germline Membrane Architecture Defects. *G3:
810 Genes|Genomes|Genetics*.

811 Ma X., Zhu Y., Li C., Xue P., Zhao Y., *et al.*, 2014 Characterisation of *Caenorhabditis*
812 *elegans* sperm transcriptome and proteome. *BMC Genomics* 15: 168.

813 Maeda I., Kohara Y., Yamamoto M., Sugimoto A., 2001 Large-scale analysis of gene
814 function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr. Biol.* 11: 171–
815 176.

816 Mamanova L., Coffey A. J., Scott C. E., Kozarewa I., Turner E. H., *et al.*, 2010 Target-
817 enrichment strategies for next-generation sequencing. *Nat. Methods* 7: 111–118.

818 Melo J. A., Ruvkun G., 2012 Inactivation of conserved *C. elegans* genes engages
819 pathogen- and xenobiotic-associated defenses. *Cell* 149: 452–466.

820 Meneely P. M., Herman R. K., 1979 Lethals, steriles and deficiencies in a region of the

- 821 X chromosome of *Caenorhabditis elegans*. *Genetics* 92: 99–115.
- 822 Meneely P. M., Herman R. K., 1981 Suppression and function of X-linked lethal and
823 sterile mutations in *Caenorhabditis elegans*. *Genetics* 97: 65–84.
- 824 Minevich G., Park D. S., Blankenberg D., Poole R. J., Hobert O., 2012 CloudMap: A
825 cloud-based pipeline for analysis of mutant genome sequences. *Genetics* 192:
826 1249–1269.
- 827 Morton D. G., Shakes D. C., Nugent S., Dichoso D., Wang W., *et al.*, 2002 The
828 *Caenorhabditis elegans* *par-5* gene encodes a 14-3-3 protein required for cellular
829 asymmetry in the early embryo. *Dev. Biol.* 241: 47–58.
- 830 O’Roak B. J., Vives L., Fu W., Egertson J. D., Stanaway I. B., *et al.*, 2012 Multiplex
831 Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum
832 Disorders. *Science* (80-.). 338: 1619–1622.
- 833 O’Rourke S. M., Carter C., Carter L., Christensen S. N., Jones M. P., *et al.*, 2011a A
834 survey of new temperature-sensitive, embryonic-lethal mutations in *C. elegans*: 24
835 alleles of thirteen genes. *PLoS One* 6: e16644.
- 836 O’Rourke S. M., Yochem J., Connolly A. A., Price M. H., Carter L., *et al.*, 2011b Rapid
837 mapping and identification of mutations in *Caenorhabditis elegans* by restriction
838 site-associated DNA mapping and genomic interval pull-down sequencing.
839 *Genetics* 189: 767–778.
- 840 Perry M. D., Trent C., Robertson B., Chamblin C., Wood W. B., 1994 Sequenced alleles
841 of the *Caenorhabditis elegans* sex-determining gene *her-1* include a novel class of
842 conditional promoter mutations. *Genetics* 138: 317–327.
- 843 Poultney C. S., Butterfoss G. L., Gutwein M. R., Drew K., Gresham D., *et al.*, 2011

- 844 Rational design of temperature-sensitive alleles using computational structure
845 prediction. PLoS One 6.
- 846 Qiao L., Lissemore J. L., Shu P., Smardon A., Gelber M. B., *et al.*, 1995 Enhancers of
847 *glp-1*, a gene required for cell-signaling in *Caenorhabditis elegans*, define a set of
848 genes required for germline development. Genetics 141: 551–569.
- 849 Rockman M. V., Kruglyak L., 2009 Recombinational Landscape and Population
850 Genomics of *Caenorhabditis elegans*. PLoS Genet. 5: e1000419.
- 851 Rogalski T. M., Gilchrist E. J., Mullen G. P., Moerman D. G., 1995 Mutations in the unc-
852 52 gene responsible for body wall muscle defects in adult *Caenorhabditis elegans*
853 are located in alternatively spliced exons. Genetics 139: 159–169.
- 854 Rual J. F., Ceron J., Koreth J., Hao T., Nicot A. S., *et al.*, 2004 Toward improving
855 *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library.
856 Genome Res. 14: 2162–2168.
- 857 Sarin S., Prabhu S., O’Meara M. M., Pe’er I., Hobert O., 2008 *Caenorhabditis elegans*
858 mutant allele identification by whole-genome sequencing. Nat. Methods 5: 865–
859 867.
- 860 Seidel H. S., Rockman M. V, Kruglyak L., 2008 Widespread genetic incompatibility in *C.*
861 *elegans* maintained by balancing selection. Science 319: 589–594.
- 862 Seidel H. S., Ailion M., Li J., Oudenaarden A. van, Rockman M. V., *et al.*, 2011 A novel
863 sperm-delivered toxin causes late-stage embryo lethality and transmission ratio
864 distortion in *C. elegans*. PLoS Biol. 9.
- 865 Simmer F., Moorman C., Linden A. M. Van Der, Kuijk E., Berghe P. V. E. Van Den, *et*
866 *al.*, 2003 Genome-wide RNAi of *C. elegans* using the hypersensitive *rrf-3* strain

- 867 reveals novel gene functions. PLoS Biol. 1.
- 868 Smardon A., Spoerke J. M., Stacey S. C., Klein M. E., MacKin N., *et al.*, 2000 EGO-1 is
869 related to RNA-directed RNA polymerase and functions in germ-line development
870 and RNA interference in *C. elegans*. Curr. Biol. 10: 169–178.
- 871 Smith H. E., Fabritius A. S., Jaramillo-Lambert A., Golden A., 2016 Mapping
872 Challenging Mutations by Whole-Genome Sequencing. G3
873 Genes|Genomes|Genetics 6: g3.116.028316.
- 874 Sönnichsen B., Koski L. B., Walsh a, Marschall P., Neumann B., *et al.*, 2005 Full-
875 genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. Nature
876 434: 462–9.
- 877 Thompson O., Edgley M., Strasbourger P., Flibotte S., Ewing B., *et al.*, 2013 The million
878 mutation project: A new approach to genetics in *Caenorhabditis elegans*. Genome
879 Res. 23: 1749–1762.
- 880 Turner E. H., Lee C., Ng S. B., Nickerson D. a, Shendure J., 2009 Massively parallel
881 exon capture and library-free resequencing across 16 genomes. Nat. Methods 6:
882 315–316.
- 883 Wang Y., Wang J. T., Rasoloson D., Stitzel M. L., O' Connell K. F., *et al.*, 2014
884 Identification of suppressors of *mbk-2/DYRK* by whole-genome sequencing. G3
885 (Bethesda). 4: 231–41.
- 886 Wicks S. R., Yeh R. T., Gish W. R., Waterston R. H., Plasterk R. H., 2001 Rapid gene
887 mapping in *Caenorhabditis elegans* using a high density polymorphism map. Nat.
888 Genet. 28: 160–164.
- 889 Wright AJ H. C., 2003 Mutations in a β -Tubulin Disrupt Spindle Orientation and

890 Microtubule Dynamics in the Early *Caenorhabditis elegans* Embryo. *Mol. Biol. Cell*

891 14: 4512–4525.

892 Zahreddine H., Zhang H., Diogon M., Nagamatsu Y., Labouesse M., 2010 CRT-

893 1/Calreticulin and the E3 Ligase EEL-1/HUWE1 Control Hemidesmosome

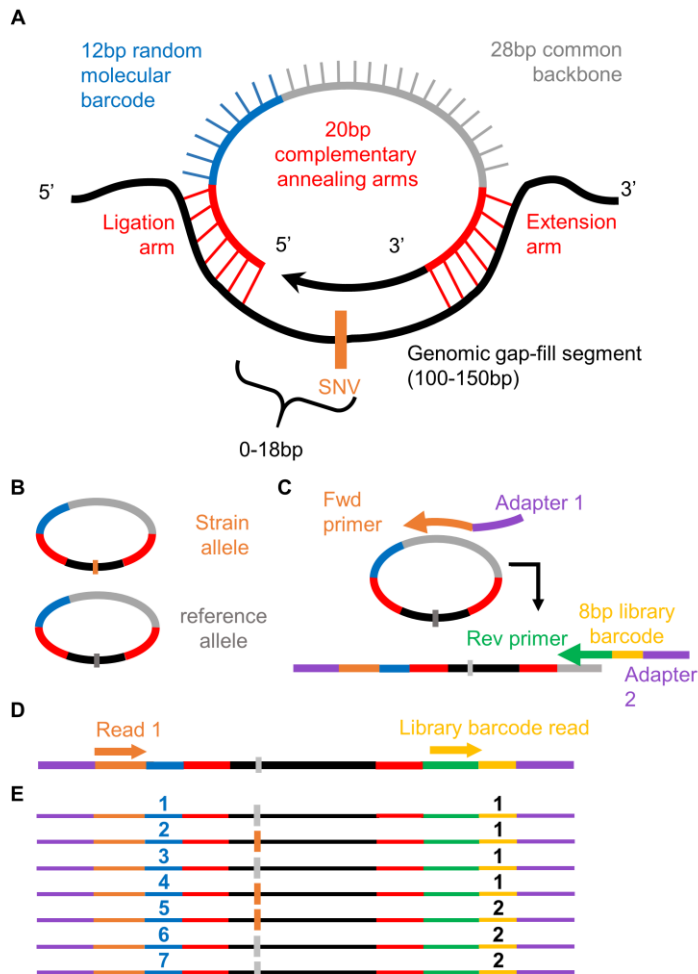
894 Maturation in *C. elegans* Development. *Curr. Biol.* 20: 322–327.

895 Zonies S., Motegi F., Hao Y., Seydoux G., 2010 Symmetry breaking and polarization of

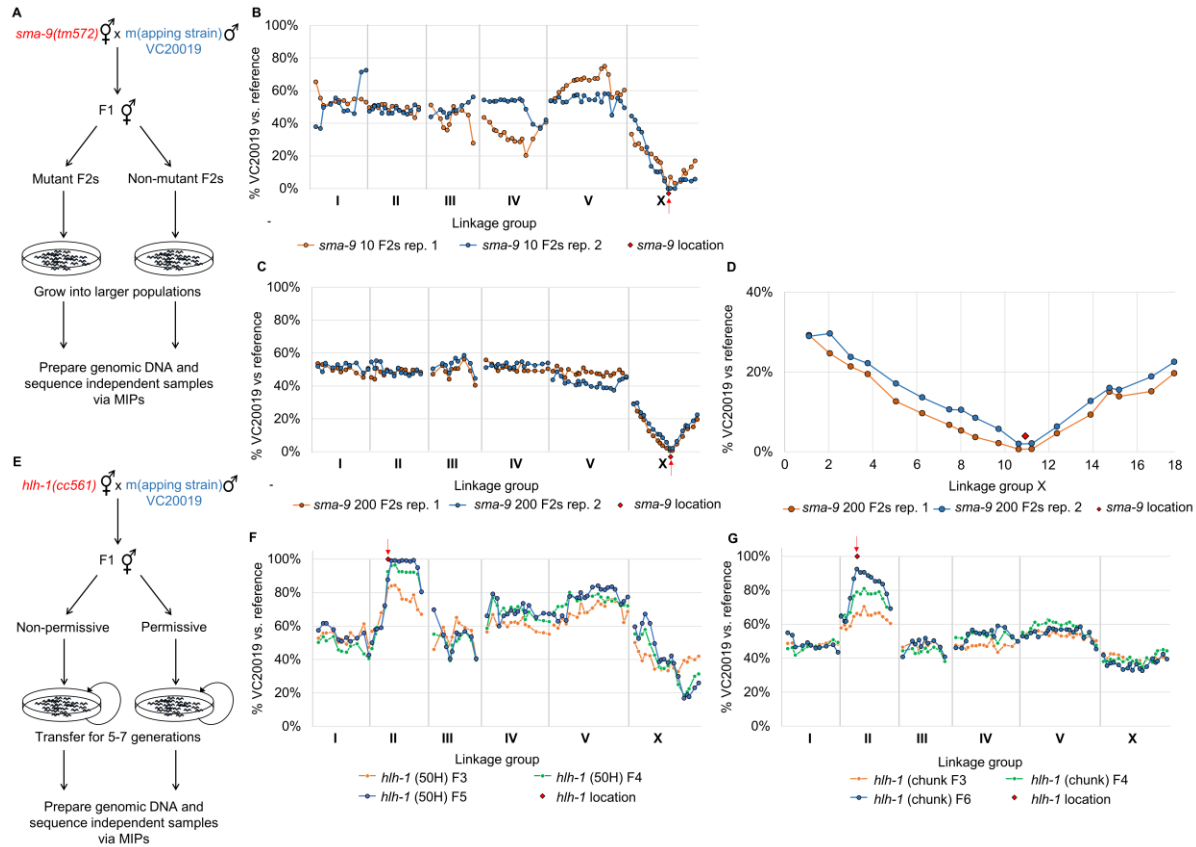
896 the *C. elegans* zygote by the polarity protein PAR-2. *Development* 137: 1669–1677.

897

898



899
 900 **Figure 1. MIPs design and workflow.** The design of the single-molecule Molecular
 901 Inversion Probes (smMIPs, Hiatt et al., 2013) was modified to relocate the molecular
 902 barcode. Overall the MIP design incorporates two 20bp annealing arms, a 12bp
 903 molecular tag and 28bp common backbone. After annealing to a target segment ranging
 904 in size from 100-150bp (A), the MIP is gap-filled with high-fidelity polymerase capturing
 905 the sequences of interest and circularized via ligation (B). The SNV of interest is located
 906 within an 18bp gap-fill window upstream of the ligation arm. Uncircularized DNA is then
 907 degraded via exonucleases. The remaining MIPs are linearized by combining with
 908 sequencing adapters and library barcodes during PCR amplification (C). A single end
 909 read captures both the molecular tag, 5' annealing sequence, and enough genomic
 910 sequence to confirm correct target capture. This information is then demultiplexed (E)
 911 and used as a means to compare sequence variant ratios.



913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

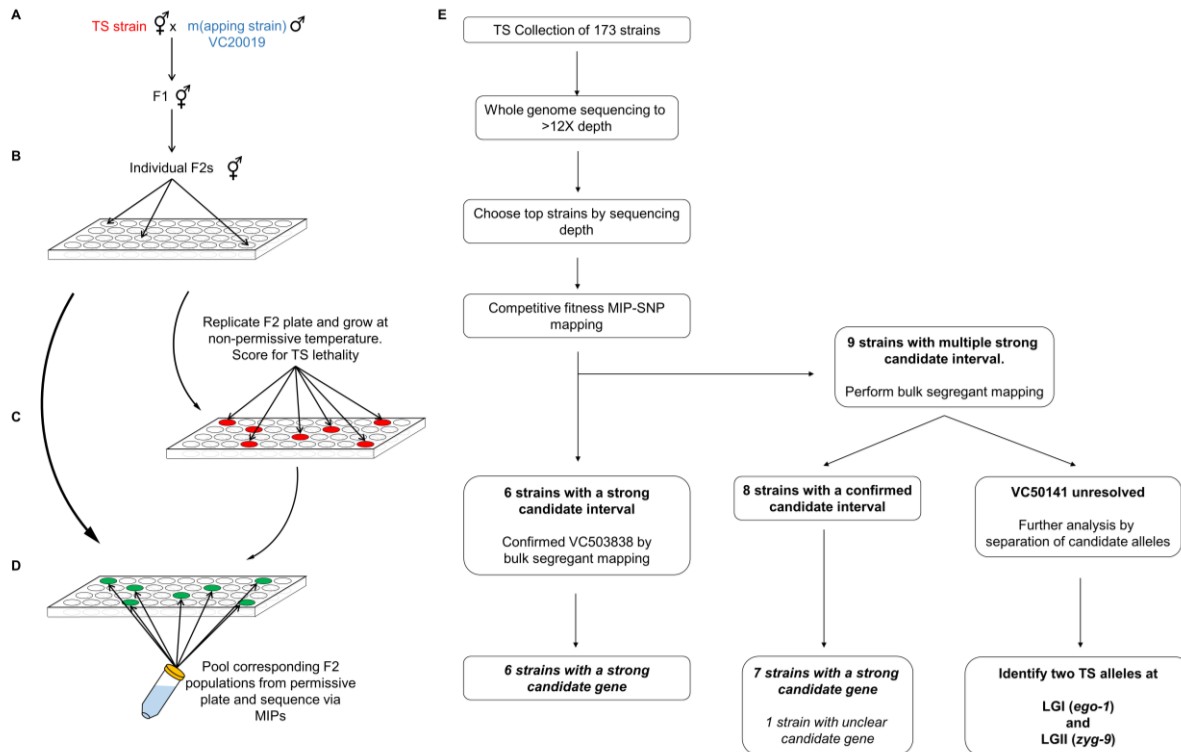
928

929

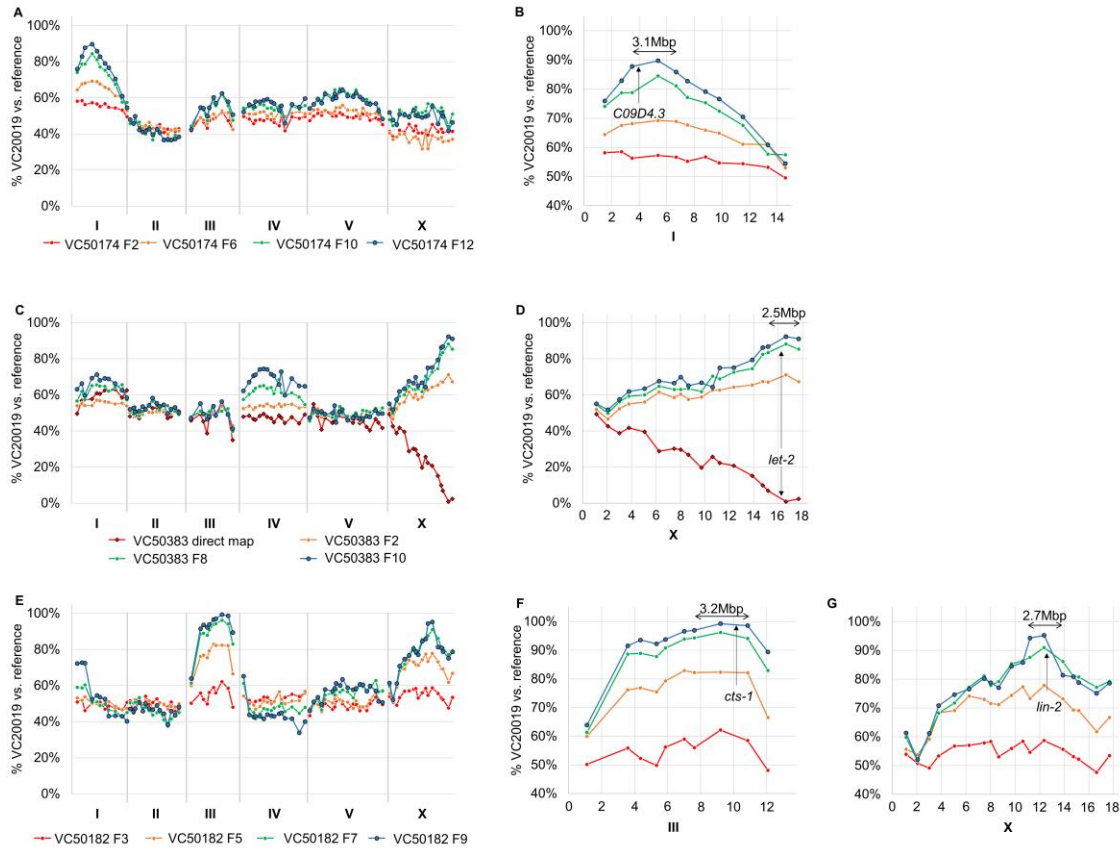
930

931

Figure 2. Mapping of *sma-9* and *hlh-1* via VC20019 and MIP-MAP sequencing. Bulk segregant mapping scheme for *sma-9* (A) using MIPs that target a number of VC20019-specific SNVs across the genome. Two replicate MIP-MAP samples of populations grown from 10 *sma* phenotype F2 animals sufficiently identified a *sma-9*-associated interval on X within a 1.5Mbp to 4.2Mbp window (B) and replicates using 200 *sma* F2 recombinants produced a 2.6Mbp window (C). Closer examination of the MIP target sites on LGX show *sma-9* is located equidistantly between two probes only 590kb apart (D). The *hlh-1(cc561)* temperature sensitive embryonic lethal allele was mapped using a competitive fitness mapping method (E). PD4605 hermaphrodites were mated with VC20019 males and non-permissive temperatures were used to select against the subsequent *hlh-1(cc561)* homozygous progeny. Starting with F1 cross progeny and transferring mixed-stage sub-populations of animals (F,G) for up to 7 generations, a region corresponding to the *hlh-1* locus on LGII was successfully identified by the fixation of VC20019-specific SNVs. A mapping interval of 2.8-7.8Mbp was identified by the F5 generation when transferring populations of 50 L1s (F). A 2.7Mb interval interval was observed by the F6 generation when transferring 4cm² chunks of mixed-stage animals (G). Each line present in (F) and (G) represents mapping data from a different generation for the same biological replicate.



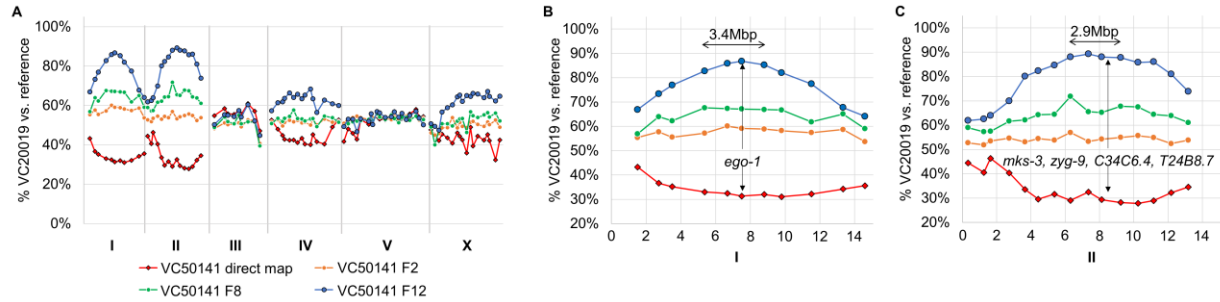
932 **Figure 3. Workflows for liquid bulk segregant mapping and TS mutant analysis.**
 933 (A) VC20019 males were mated to a TS mutant strain with cross progeny F1 animals
 934 used to generate F2 recombinants. Single F2s were used to seed wells (B), growing
 935 these populations at permissive temperatures before replicating them to grow in non-
 936 permissive conditions (C) and identify populations that failed to thrive. Dead populations
 937 (red) were then chosen from the original wells (green) and pooled together (D) to
 938 prepare a single genomic sample for MIP-MAP sequencing. Samples from multiple
 939 plates were combined in the mapping of a single mutant depending on allele penetrance
 940 or the number of positive F2 wells identified. The workflow (E) of mapping TS mutants
 941 shows 6 strains could be mapped by competitive fitness MIP-MAP alone; 8 strains
 942 benefitted from additional mapping via the liquid bulk segregant protocol; and 1 strain
 943 (VC50141) required additional analyses.



944

945 **Figure 4. Successful mapping of temperature-sensitive lethal mutant strains via**
946 **competitive fitness and liquid-format MIP-MAP techniques.** 16 TS lethal mutants
947 with genomic sequencing data were mapped by the competitive fitness approach used
948 with *hlh-1*. Mapping results yielded a diverse group of mapping profiles including clean
949 single-locus mappings as seen in VC50174 (AB); profiles such as VC50383 with small
950 additional loci or peaks appearing at later generations but resolved by additional liquid-
951 format bulk segregant mapping of phenotypes (CD); and multiple-locus profiles resolved
952 by identifying shared background mutations in related strains as in VC50182 (EF). Each
953 line present in a graph represents mapping data from a different generation for the
954 same experimental replicate or the results of a liquid-format bulk segregant mapping
955 experiment (indicated as direct map with diamond points).

956



957

958 **Figure 5. The TS strain VC50141 has two independent TS alleles of *ego-1* and *zyg-***

959 **9.** Both competitive fitness mapping and liquid-format bulk segregant mapping

960 confirmed that the TS lethality in VC50141 was related to two loci (A). The first interval

961 located on LGI contains only a single relevant coding mutation in *ego-1* (B). The second

962 locus, appearing to be of equal penetrance on LGII contains 4 candidates, of which *zyg-*

963 *9* is the most likely causative allele (C). Each line present in the graph represents

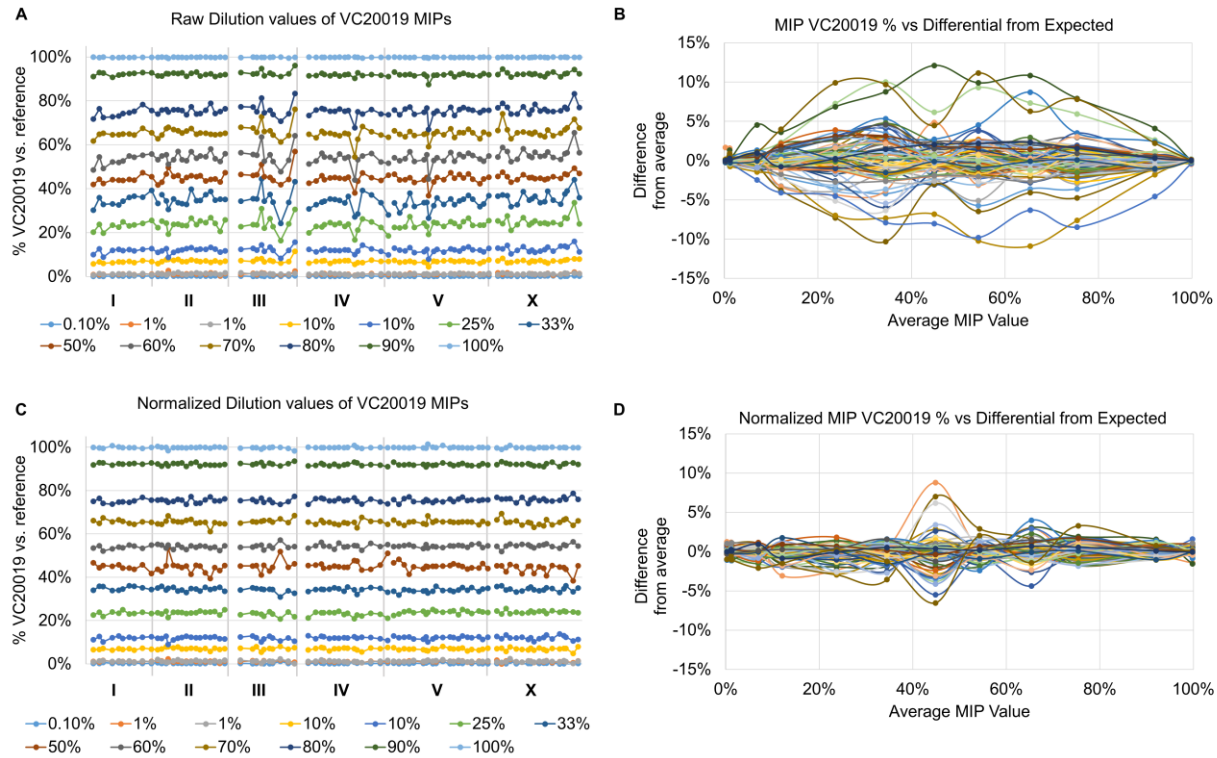
964 mapping data from a different generation for the same experimental replicate or the

965 results of a liquid-format bulk segregant mapping experiment (indicated as direct map

966 with diamond points).

967

968



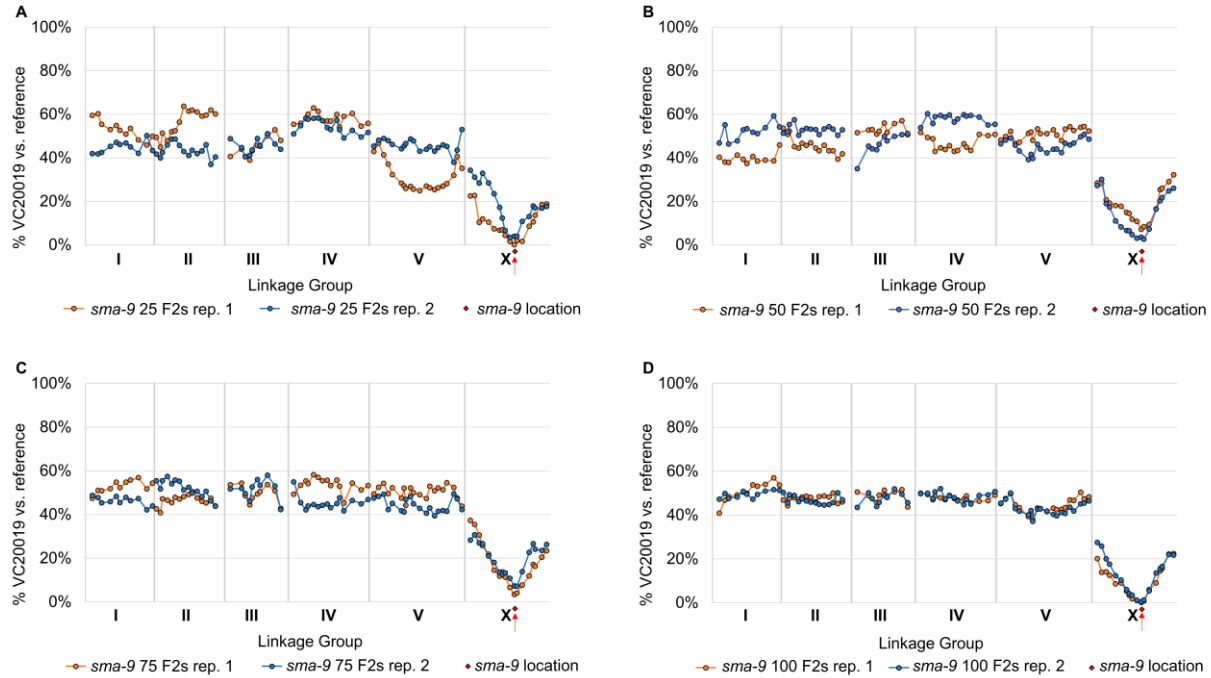
969

970 **Supplemental Figure S1. MIPs generate consistent results but may require**
971 **normalization.** The concordance of SNV ratios across multiple loci and sample
972 dilutions (A), identified a small subset of MIPs that generated incorrect read calls across
973 multiple dilutions. The difference between the called dilution value and the average
974 across all 90 MIPs was used to generate a behavioural curve (B). Behavioural curves
975 were used to normalize the MIP SNV ratios, generating more consistent calls across all
976 MIP targets (C, D).

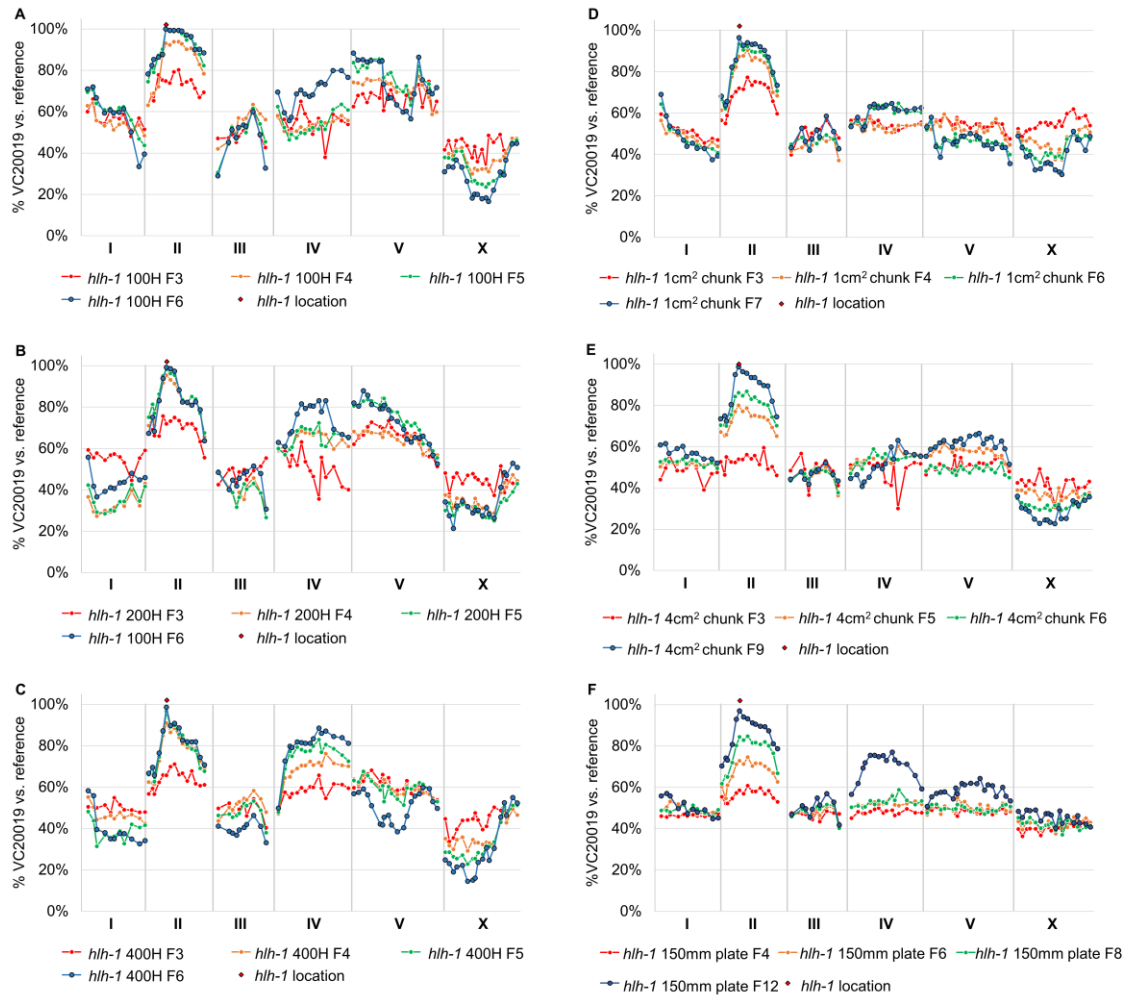


977

978 **Supplemental Figure S2. MIP accuracy and efficiency were tested on a range of**
979 **samples.** (A) The false positive rates of MIPs were investigated using a series of
980 probes targeting 172 genomic loci. A mean raw false positive read rate of 0.0651% ±
981 0.0696% was observed with a maximum of 0.543%. By filtering reads at the SNV site
982 for quality scores of Phred ≥ 30, the false positive rate was decreased to a mean of
983 0.0122% ± 0.0129% with a maximum value of 0.0785%. (B) MIP capture and library
984 preparation produces a high concordance of sequence reads vs. molecular tags. Using
985 24 sets of MIP capture reactions with the same 89 MIPs, the observed mean rate of
986 molecular tag duplication via PCR is 1.048 ± 0.006. The ratio of reads per tag was not
987 correlated to the total reads per MIP when analysing the same 24 sets by total read
988 count on a per MIP per capture reaction basis.
989

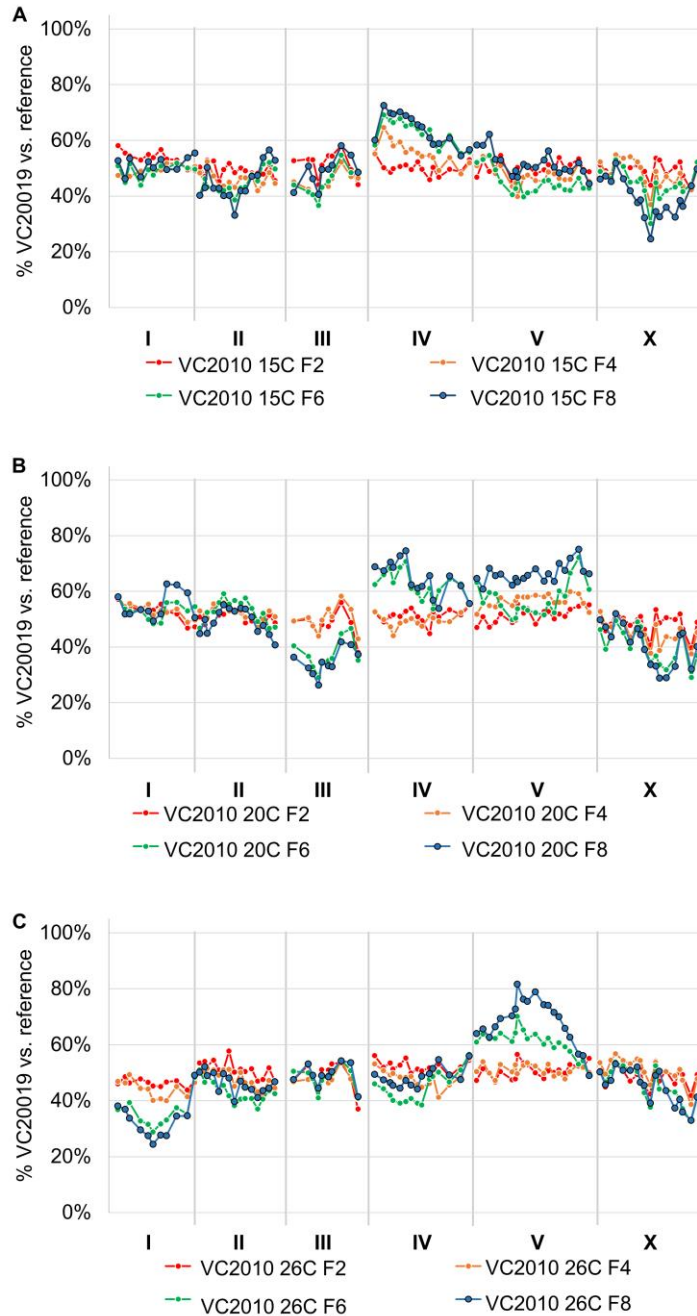


991 **Supplemental Figure S3. MIP-MAP profiles for *sma-9* mutants improves when**
992 **using more F2 recombinants.** A sufficient MIP-MAP profile to identify the *sma-9* locus
993 was generated with as few as 25 F2 *sma-9* phenotype animals (A) but mapping
994 improved with increasing numbers of F2s populations at 50 (B), 75 (C) and 100 (D) F2
995 animals.



996

997 **Supplemental Figure S4. MIP-MAP profiles for *hlh-1* mutants can be generated by**
998 **indirectly mapping in multiple schemes.** MIP-MAP mapping profiles for *hlh-1* were
999 reproduced using varying numbers of animals transferred each generation until fixation
1000 against the TS *hlh-1(cc561)* allele occurred. The size of population transferred each
1001 generation was tested at 100 (A), 200 (B), and 400 (C) animals. Random selection of
1002 animals was verified populations were transferred by chunking 1cm² (D) or 4cm² (E) of
1003 animals or washing the entire population and portioning ~10,000 animals onto new
1004 150mm plates (F). Each line present in a graph represents mapping data from a
1005 different generation for the same experimental replicate.



1006

1007

1008 **Supplemental Figure S5. VC20019 appears to have little to no temperature-**

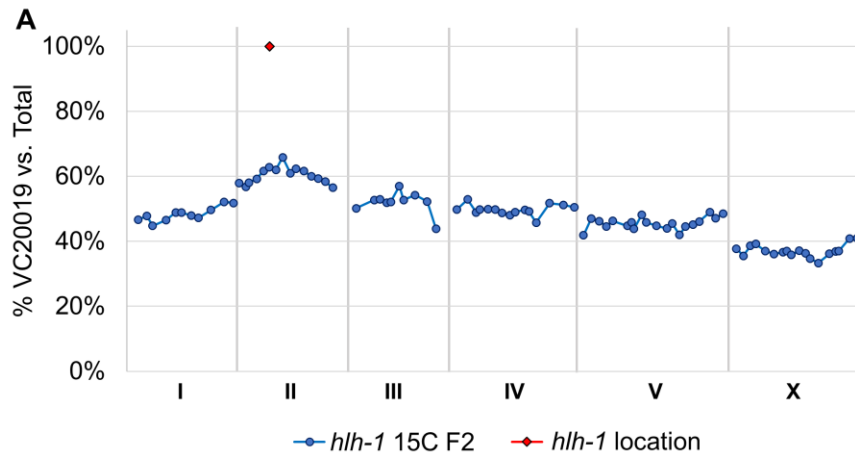
1009 **sensitivity.** The mapping profile of a cross between VC20019 and VC2010 was

1010 completed with very little fixation at any particular locus across the genome after

1011 mapping at 15°C (A), 20°C (B), and 25°C (C) for at least 8 generations. This data

1012 indicates that the mapping mutant VC20019 has comparable fitness to our N2 reference

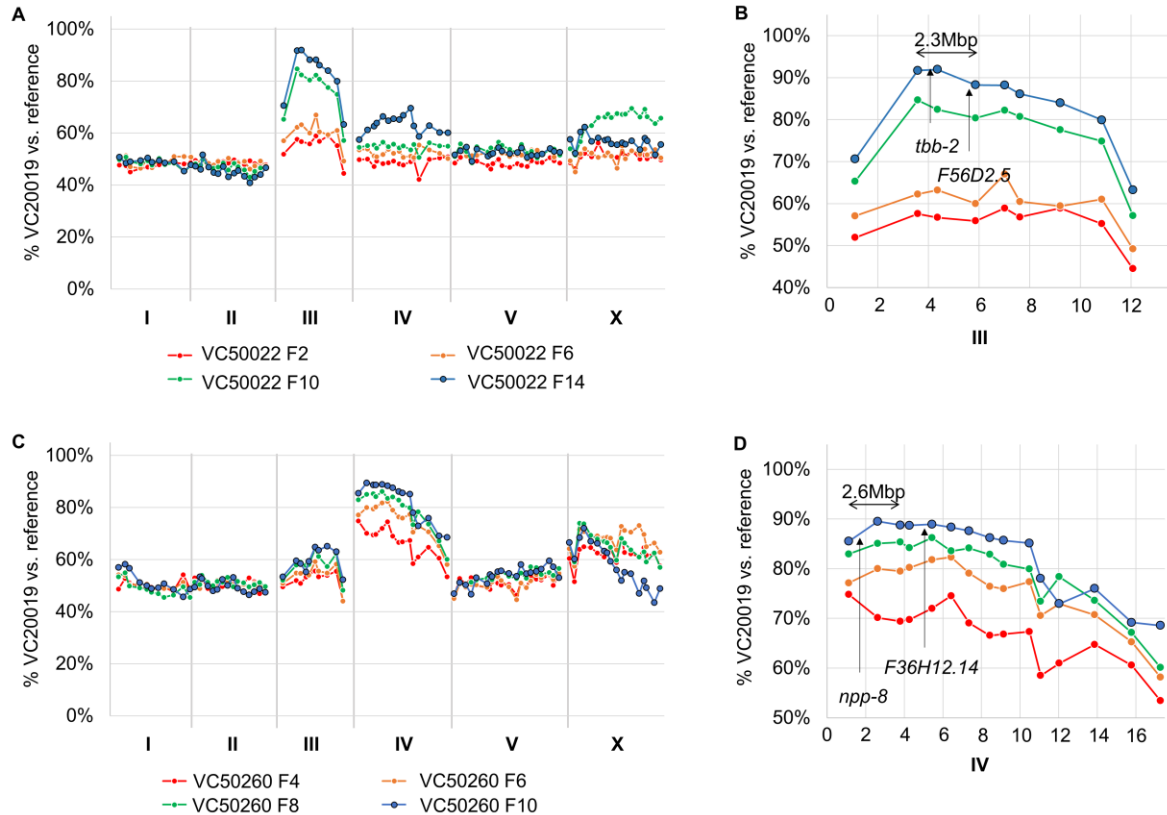
1013 *C. elegans* strain VC2010.



1014

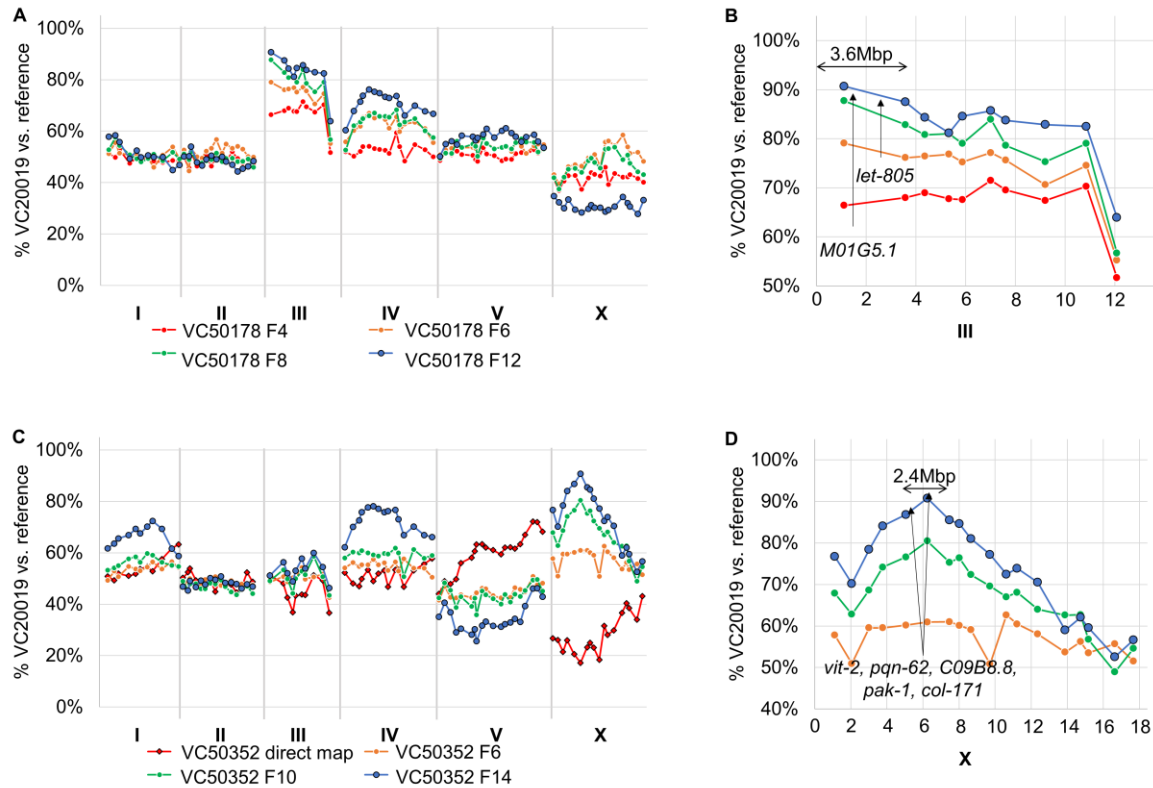
1015 **Supplemental Figure S6. The *hlh-1(cc561)* allele exhibits mild fitness defects at a**
1016 **permissive temperature.** While mapping the *hlh-1(cc561)* allele at 15°C the beginnings
1017 of fixation against the *hlh-1* interval was observed, even after a single generation at the
1018 permissive temperature.

1019



1020

1021 **Supplemental Figure S7. TS strains that resolved with a single locus.** VC50022 is
1022 an example of a clean TS mapping that identified a single locus without additional loci
1023 (A). In this strain there are only 2 relevant coding mutations within the identified interval
1024 on LGIII (B). The most likely candidate *tbb-2* has a number of embryonic lethal
1025 phenotypes. VC50260 (C) resolved without much indication of any other candidate loci.
1026 Within this interval on LGIV there was only 1 relevant coding variant which was located
1027 in *npp-8* (D). The next closest relevant coding variant is in *F36H12.14* nearly 1.5Mbp
1028 from the predicted interval.



1029

1030 **Supplemental Figure S8. TS strains with mapping profiles that identify additional**

1031 **weak alleles of low fitness.** The TS locus within VC50178 (A) was identified on the left

1032 arm of LGIII by following competitive fitness mapping over multiple generations. The

1033 predicted 3.6Mbp interval (B) has only two relevant coding changes with the likely

1034 candidate as *let-805* based on phenotyping of the VC50178 strain. The TS locus on

1035 LGX in VC50352 (C) was resolved with the liquid bulk segregant mapping assay. Within

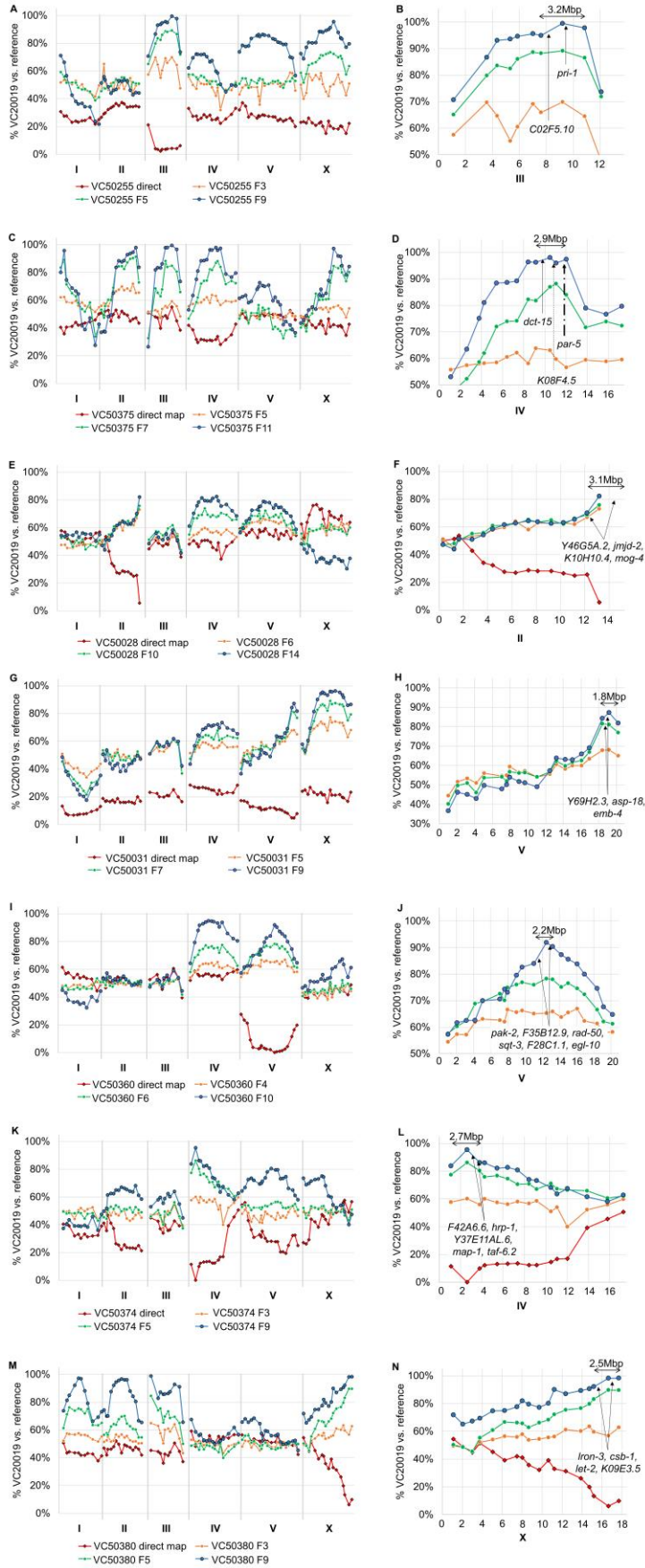
1036 this interval (D) are 7 relevant candidates with the prime candidate being in *pak-1* based

1037 on the VC50178 observed TS phenotype. It is possible the additional loci harbour weak

1038 or low-penetrance alleles involved in population fitness and are not apparent until the

1039 TS-allele has been removed from the population.

1040



1042 **Supplemental Figure S9. TS strains can carry multiple strong fitness-defective**
1043 **loci.** In three cases, the competitive fitness mapping protocol identified multiple loci
1044 within these strains that conferred an overall reduced fitness to the population at a
1045 fixation rate that made it difficult to identify the correct TS-specific locus. Additional
1046 mapping with the liquid-format bulk segregant assay elucidated the correct linkage
1047 group for the TS phenotype - albeit with lower resolution. The TS-associated locus in
1048 VC50255 was identified at LGIII (A) with the prime candidate as *pri-1* based on
1049 phenotype and Wormbase reports (B). A 2.9Mbp TS interval was identified in VC50375
1050 on LGIV (C) containing 3 candidates (D) with the prime candidate as *par-5* based on a
1051 shared cell-cycle delay phenotype in the strain. In VC50028 a 3.1Mbp interval on LGII
1052 (E) was identified with *mog-4* as the primary candidate based on shared phenotype data
1053 between the strain and Wormbase sources (F). The strain VC50031 had multiple loci
1054 with a TS-associated hit on a 1.8Mbp interval on LGV and a likely hit for *lin-2* on LGX
1055 (G). The primary candidate in this interval is *emb-4* (H) based on based on VC50031's
1056 sterility phenotype and a reported high early embryonic expression of this gene. The
1057 strain VC50360 produced a mapping profile (I) with two very strong candidate loci that
1058 were resolved to a 2.2Mbp interval on LGV (J) with a primary candidate of *sqt-3* as the
1059 strain's phenotype overlapped with reported RNAi and allele phenotypes. VC50374
1060 exhibited a number of peaks that were mirrored in liquid-format segregant mapping with
1061 the TS-locus resolved at LGIV (K). The 2.7Mbp interval (L) contained many candidates
1062 with the primary candidate as *taf-6.2*, a gene reported to have embryonic lethal
1063 phenotypes and TS lethal alleles. One *suc* allele is located in the same exon as the
1064 VC50374 SNV. The TS strain VC50380 (M) identified 4 strong low-fitness loci with
1065 liquid-format bulk segregant mapping resolving the TS locus to LGX. Within this interval
1066 (N) the primary candidate is *let-2* based on reported phenotypes compared to this
1067 strain.
1068

1069 Table 1. Temperature-sensitive collection phenotype summary

Set	General Phenotype	Total Strains
1	Failure of Fertilization	8
1	P1 Delay	11
1	Unclassified sterility	5
1	Early divisions normal, cytoplasmic clearing	3
1	Early division normal, late arrest	8
1	Early division normal, terminal phenotype unknown	22
1	Low penetrance embryonic lethality or unlikely TS strain	15
2	General larval lethality and sterility	39
2	Cytoplasm morphology abnormal	9
2	Cell division abnormalities	17
2	Cell death aberrations	10
2	Early gastrulation defects	1
2	Tissue-specific developmental defects	15
2	Morphology abnormalities	38
2	Undocumented Phenotype	3
3	Maternal effect lethality	34

1070 Note: Set 2 encompasses overlapping phenotypes from the 7 categories, with undocumented phenotype data for 3 strains

1071

1072 Table 2. Mutation summary from 173 strains

Set (Unique Genes)	Total SNVs	Coding exons			Non-coding exons (UTRs)	Introns		ncRNAs	Intergenic
		Missense	Nonsense	Synonymous		Splicing	Other		
Set 1 (4,579)	9,285	1,971	72	678	316	42	2,770	131	3,302
Set 2 (11,943)	38,965	8,291	356	3,396	1,087	142	11,078	595	14,020
Set 3 (4,269)	8,612	1,896	78	723	251	35	2,334	124	3,171
All strains	56,803	12,144	505	4,793	938	217	16,164	850	20,478
Unique Genes	14,327	7,972	493	3,937	1,497	215	7,623	779	

1073

1074

1075 Table 3. Table of temperature-sensitive mutants mapped

Strain	Set	WGS Coverage	Phenotype Penetrance	Timing of lethality and phenotype observations at 26°C
VC50022	3	16.6	100%	Arrest in early embryogenesis
VC50028	3	40.7	100%	Sterility of P ₀ when plated at L4. Any F1s produced arrest at early larval stages. Some sterility also noted at permissive temperatures.
VC50031	3	16.4	100%	Sterility of P ₀ when plated at L4. Any F1s produced arrest at early embryogenesis.
VC50141	3	16.9	100%	Sterility of P ₀ when plated at L4. Any F1s produced arrest at early embryogenesis.
VC50174	1	21.9	<80%	Sterility of P ₀ when plated at L4, but also have bag phenotype so F1 are still produced. Failure of fertilization with vacuolated spermatozoa observed.
VC50178	1	20.7	100%	P ₀ when plated at L4 still produce progeny but F1 population is observed to be sterile. Severe morphogenesis defect, possible 2-fold arrest.
VC50182	1	19.8	100%	Early divisions normal, late arrest. Mostly 2-fold arrest with movement.
VC50255	1	20.0	100%	Sterility of P ₀ when plated at L4 with bag phenotype of dead eggs likely due to embryonic arrest. Any F1 produced are sterile. P1 delay is observed in embryos.
VC50260	1	22.9	99%	Sterility of P ₀ when plated at L4 with bag phenotype of dead eggs likely due to embryonic arrest. Any F1 produced are sterile. P1 delay is observed in embryos.
VC50352	2	23.4	<100%	Slightly leaky as P ₀ can give rise to F1, and then F2 but these are usually sickly and sterile. In some cases they can make F3. Pretzel inviable, asynchronous cell cycle, later stage uncoordinated.

VC50360	2	22.2	100%	P ₀ have a mild roller phenotype, sluggish in appearance. F1 progeny appear to arrest in early embryogenesis. Pretzel inviable, pretzel deformed, cell adhesion/migration defect.
VC50374	2	20.9	99%	P ₀ when plated at L4 produce unhatched eggs of various forms and appears embryonic lethal. Any F1s produced are sterile. Arrest of development at premorphogenetic stage
VC50375	1	19.0	100%	F1 sterility with P1 delay observed.
VC50380	2	19.6	100%	P ₀ when plated at L4 appear active and produce eggs. F1 progeny arrest in late embryogenesis and L1. Pretzel inviable, abnormal cytoplasm morphology, spindle defects, aberrant pharyngeal cluster and pretzel deformed.
VC50383	2	21.7	100%	P ₀ plated at L4 appear active but produce eggs that fail to hatch. Pretzel inviable, larval lethal, and muscle defects

 1076

1077

1078 Table 4. Summary of temperature-sensitive mutant candidate mapping intervals

Strain	Set	Mapping methods	Mapping interval	Candidate gene(s)
VC50022	3	CF	LGIII:3579637-5328496	<i>tbb-2, F56D2.5</i>
VC50174	1	CF	LGI:3494925-6676830	<i>C09D4.3</i>
VC50178	1	CF	LGIII:0-3579637	<i>M01G5.1, let-805</i>
VC50182	1	CF	LGIII:7617321-10853887	<i>cts-1, enu-3</i>
VC50260	1	CF	LGIV:1125542-3797513	<i>npp-8</i>
VC50380	2	CF	LGX:15150772-17676467	<i>Iron-3, csb-1, let-2, K09E3.5</i>
VC50028	3	CF and 96-well	LGII:121799849-15279345	<i>cox-10, jmjd-2, K10H10.4, mog-4</i>
VC50031	3	CF and 96-well	LGV:18400066-20183925	<i>Y69H2.3, asp-18, emb-4</i>
VC50255	1	CF and 96-well	LGIII:7617321-10853887	<i>C02F5.10, pri-1</i>
VC50352	2	CF and 96-well	LGIV:5057999-7478926	<i>vit-2, pqn-62, Y34B4A.6, ifta-1, C09B8.8, pak-1, col-171</i>
VC50360	2	CF and 96-well	LGV:11035658-14223805	<i>pak-2, F35B12.9, rad-50, sqt-3, F28C1.1, egl-10</i>
VC50374	2	CF and 96-well	LGIV:1125542-3797513	<i>F42A6.6, hrp-1, Y37E11AL.6, map-1, taf-6.2</i>
VC50375	1	CF and 96-well	LGIV:9147769-12028068	<i>dct-15, K08F4.5, par-5</i>
VC50383	2	CF and 96-well	LGX:15180772-17676467	<i>rab-14, let-2, nas-39, acs-9, F52G3.1, gcy-11, crb-1</i>
VC50141	3	CF, 96-well, segregation testing	LGI:6676830-8797510 LGII:6305987-9232191	<i>ego-1</i> <i>mks-3, zyg-9, C34C6.4, T24B8.7</i>

1079 CF = competitive fitness mapping; 96-well = liquid bulk segregant mapping

1080

1081 Table 5. TS Candidate mutation information

Strain	Primary candidate	Allele	Candidate gene phenotype information	TS allele(s) available
VC50022	<i>tbb-2</i>	ax123	Early embryonic lethal ^a	<i>or362, or600</i>
VC50028	<i>mog-4</i>	ax87	Sterility ^b	<i>q233</i>
VC50031	<i>emb-4</i>	ax968	Larval lethal ^c	<i>hc60</i>
VC50141	<i>ego-1</i>	ax458	Embryonic arrest, and small oocytes with unusual chromosome morphology by RNAi ^d .	None. In same exon as <i>om54</i> and <i>om71</i>
	<i>zyg-9</i>	ax3101	Embryonic lethal by RNAi ^{c,e,f} . Failure to hatch in TS allele <i>or623</i> ^g .	<i>or623, or628, or593, or635, or634</i>
VC50174	<i>C09D4.3</i>	<i>or976</i>	No embryonic phenotypes reported. Peak expression in L4 and male samples. Expression present in sperm transcriptome ^h	None
VC50178	<i>let-805</i>	<i>or896</i>	Embryonic lethal ^{e,f} , maternal sterile ^{c,e} , larval lethal ^{c,e,i} .	None
VC50182	<i>cts-1</i>	<i>or1005</i>	Embryonic lethal ^{c,e,i} .	None
VC50255	<i>pri-1</i>	<i>or1187</i>	Delayed early cell division with prominent 3-cell stage ^j , and embryonic lethality ^{e,i} .	None
VC50260	<i>npp-8</i>	<i>or1044</i>	Embryonic lethal ^{c,e} .	<i>ax146</i>
VC50352	<i>pak-1</i>	<i>t3254</i>	Low range embryonic lethal, larval lethal phenotypes and larval lethal with muscle detachment ^k .	None
VC50360	<i>sqt-3</i>	<i>t3352</i>	Embryonic lethal ^e , dumpy ^{e,f} , roller phenotypes ^l by RNAi and various alleles.	<i>e24, sc8, sc43, sc63, sc80</i>
VC50374	<i>taf-6.2</i>	<i>t3399</i>	Embryonic lethal ^f	<i>ax701, ax514</i>
VC50375	<i>par-5</i>	<i>or996</i>	Embryonic lethal ^{c,e} , and cell cycle delayed ^m .	None
VC50380	<i>let-2</i>	<i>t3320</i>	Early larval lethal and embryonic lethal ⁿ .	<i>mn114, e1470, b246, mn101, mn109,</i>
VC50383		<i>t3381</i>		<i>mn111, g30</i>

1082 ^aO'Rourke et al. 2011; ^bGraham, Schedl, and Kimble 1993; ^cKamath et al. 2003; ^dClaycomb et al. 2009; ^eSimmer et al. 2003; ^fSönnichsen et al. 2005; ^gBellanger et
1083 al. 2007; ^hMa et al. 2014; ⁱRual et al. 2004; ^jEncalada et al. 2000; ^kZahreddine et al. 2010; ^lCox et al. 1980; ^mMorton et al. 2002; ⁿMeneely and Herman 1979
1084
1085

1086 Supplemental Table S1. 96-well liquid mapping number of F2 segregants used

Strain	Set	Direct Mapping number of F2s (chosen/total)	% F2s Picked	Mapping observations
VC50352	2	37/362	10.22	Clear peak by F10 but carried to F14. Direct mapping locus only fixates to 20%.
VC50255	1	66/440	15.00	Clearest interval by F9. Direct mapping has very broad interval across most of LGIII.
VC50375	1	109/453	24.06	Clear peak at F7 but also matched in amplitude by 4 other peaks. Direct map resolved chromosome but with poor resolution.
VC50028	3	No Data	No Data	Quick fixation observed on LGII to 80% by F6 but does not improve even at F15. Direct mapping mirrored exact interval.
VC50031	3	72/737	9.76	Clear interval on LGV by F7 but best at F9. Direct mapping has high signal noise but confirms LGV.
VC50141	3	75/387	19.37	Mapping doesn't clarify until between F8 and F12. Direct mapping fixates to ~30% on each locus.
VC50360	2	98/439	22.32	Clearest interval on LGV by F10. Direct mapping confirms but with a very broad interval across most of LGV.
VC50374	2	15/384	3.90	Clear interval on LGIV by F5 but best fixation by F9. Direct mapping managed to identify the interval with 15 F2s.
VC50380	2	42/384	10.93	Somewhat clear peak on LGX by F5 but reached fixation at F9 with 3 additional strong intervals. Direct mapping confirmed LGX with good resolution.

- 1087
- CF = competitive fitness mapping
- 1088
- 96-well = liquid bulk segregant mapping
- 1089