

Molecular basis of essential and morphological variations across 12 balancer strains in *C. elegans*

Filip Cotra^{1*}, Tatiana Maroilley^{1*}, Maja Tarailo-Graovac^{1§}

¹Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta, Canada

[§]To whom correspondence should be addressed: maja.tarailograovac@ucalgary.ca

*These authors contributed equally.

Abstract

Balancers are primarily chromosomal rearrangements that allow lethal or sterile mutations to be stably maintained as heterozygotes. Strains with balanced lethal/sterile mutations are available at the Caenorhabditis Genetics Center. Such strains harbor morphological markers, with associated molecular changes, that are in *trans* to the balancer. In many cases, only the genetic position (in cM) has been described for balanced mutations or morphological markers. We used short-read whole genome sequencing to uncover the genomic position of those variants (balanced mutations and linked markers) and predicted their effect. We investigated 12 different strains, and characterized at a molecular level 12 variants.

A

Variant	Type	Gene	Strain	Mutagen	Position	AA* change	Effect	Flanking sequences**
Balanced mutations								
<i>bn126</i> WBVar00000502	431-bp Deletion	<i>klp-19</i>	SS746	UV+TMP	III:13,306,061-13,306,492	p.K216/- (Y43F4B.6)	Exon 2 truncation	CATCGTTGAAGCGGACTCAGCCGGCGGATCCTTCG_13306061_TTC_13306492_TCCACGAACACTGTGAAAACCGCATGAGACTGCTGGACAT
<i>e1466</i> WBVar00144022	Substitution	<i>mcm-4</i>	CB3475	EMS	I:2,321,422	p.Q88* (Y39G10AR.14a.1)	Stop gained	GTCTCCATTCCCGAATCT-G/A-GACGGTAGCATGATCGAGC
<i>e1500</i> WBVar00144047	Substitution	<i>unc-93</i>	MT690	Unknown	III:3,646,850	p.G388R (C46F11.1a.1)	Missense	TTTTCATGATTGTACTGCG-G/A-GACAAGTCGCGGAAATAT
<i>md2415</i> WBVar00088863	2,730-bp Deletion	<i>unc-13</i>	RM2431	TMP	I:7,450,592-7,453,322	- (ZK524.2d.1)	Exons 24-30 truncation	TATATAATGCGAGTAATTGAAGAAGATCAAAAAAATCT_7450592_7453322_GACGAGACTGGGCTCATTTTGTCTCAGGATCTTTCACAGCGTCAAA
<i>t1568</i> WBVar00249146	Substitution	<i>cyk-1</i>	GE2722	EMS	III:7,024,608	p.Q1212* (F11H8.4a.1)	Stop gained	CTTCTTTCTGTGAGATT-G/A-TTGACGTCTCCGGTTC
Morphological markers								
<i>h1040</i> WBVar00146729	572-bp Deletion	<i>unc-54</i>	KR2839	UV	I:14,857,166-14,857,738	- (F11C3.3.3)	Exon 6 truncation	CTTCTCGGATTGAAGAAGAG_14857166_14857738_GGCGTCTTGGCCTTGAA
<i>mn153</i> WBVar00089070	Substitution	<i>let-2</i>	SP423	EMS	X:16,387,977	- (F01G12.5a.1)	Intron 4 Splice acceptor	TCCAACCTGGTCCAGGGGGT-C/T-TGCAACAACGATGACTCAG
<i>s2170</i> WBVar00242254	Substitution	<i>dpy-1</i>	DP246	EMS	III:2,036,920	p.G1239E (M01E10.2a.1)	Missense	GAGCCAAATCTGATGAGCAT-C/T-CACGTGGATCGACGAACGA
<i>tm9708</i> WBVar02148984	Substitution	<i>unc-5</i>	FX30257	CRISPR-Cas9	IV:5,509,656	- (B0273.4a)	Intron 2	GGCAAGTAGAAGAAAAAGT-G/A-TCAAGAAAAATGAGAAATTA
<i>tm9709</i> WBVar02148981	179-bp Deletion	<i>dpy-5</i>	FX30235	CRISPR-Cas9	I:5,432,754-5,432,933	- (F27C1.8)	Exon 1 truncation	AGTGCCTCTGGCTCTCTGGGGCTCTGGGTTTC_5432754_5432933_CCATCACATCATACATGTCATTGATGTCATTTGTGATGACAAGTGACAG
<i>tm9710</i> WBVar02148976	1-bp Deletion	<i>dpy-2</i>	FX30144	CRISPR-Cas9	II:6,714,967	p.G129Efs (T14B4.6.1)	Frameshift	CGGGAGAACGTGGACCATCA-G/-GAGACTCTGACTTCCAGC
<i>tm9723</i> WBVar02148982	90-bp Deletion	<i>unc-9</i>	FX30237	CRISPR-Cas9	X:13,215,724-13,215,814	- (R12H7.1)	Exon 8 truncation	ACCGAAATTAACCCATTTCAGGATGGTGTTCCTACTTC_13215724_13215814_GCACGAGTGTGACACACCCCAACTCTGATGCTTTTCAATT

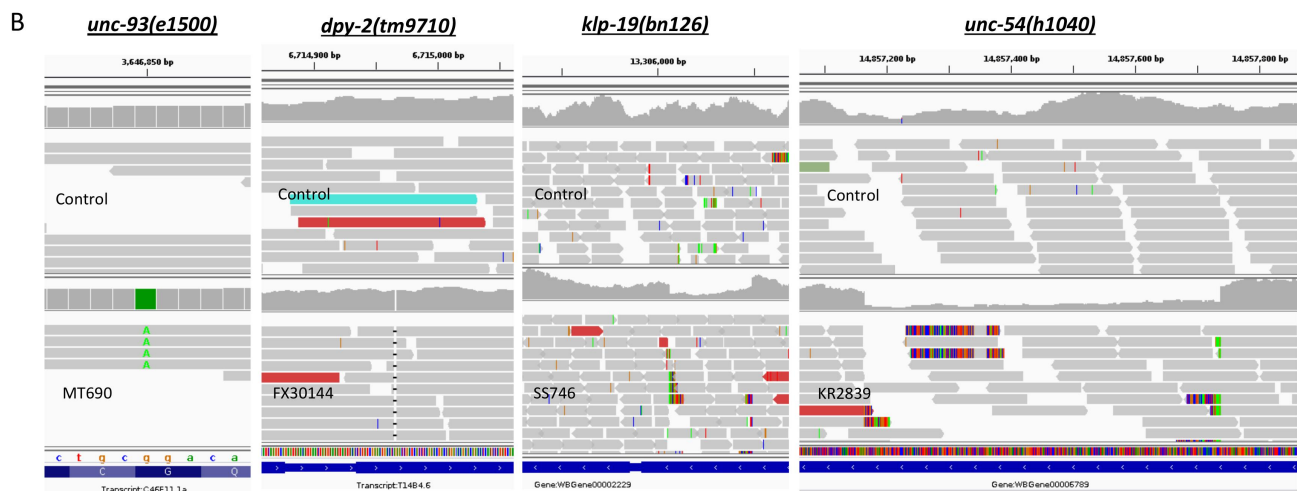


Figure 1. Molecular basis of balanced mutations and morphological markers across 12 balancer strains:

A) Table summarizing balanced mutations (TOP) and morphological markers (BOTTOM) resolved at a molecular level (genomic position and variant effect) in this study. *Amino Acid change. Between parenthesis is the transcript used as reference. For frameshift mutations, we indicated the amino acid after which the frameshift occurs. ** The flanking sequences were retrieved from Sanger sequencing for *bn126*, *md2415*, *tm9709*, and *tm9723*. B) Visualization of the genomic position of four morphological markers or balanced mutations in the Illumina short-read whole genome sequencing data using the Integrative Genomics Viewer v2.16 (IGV) in the control genome (N2) and the strain of interest. For each sample, the information is displayed on two different tracks. The histogram of the top track represents the coverage per base. A sudden drop in the coverage can signal a deletion. The arrows in the bottom track represent the reads aligned along the reference genome. A grey read is a read aligned to the reference genome with limited differences. At a deletion breakpoint, reads colored in red have their mate aligned on the reference genome at a larger distance than the inner distance initially used during the library preparation. The reference sequence and corresponding annotation is displayed at the very bottom of each

screenshot. From left to right: [unc-93\(e1500\)](#) in [MT690](#), [dpy-2\(tm9710\)](#) in [FX30144](#), [klp-19\(bn126\)](#) in [SS746](#), and [unc-54\(h1040\)](#) in [KR2839](#).

Description

Balancer chromosomes have been used for decades to maintain variations in *Caenorhabditis elegans* (*C. elegans*) that impede the organism's ability to propagate, causing lethality or sterility (Edgley et al., 2006). Balancers rely on complex genomic rearrangements created through mutagenesis, originally implicating random processes like X-rays or UVs, and more recently via CRISPR-Cas9 (Dejima et al., 2018). We have recently used short-read whole genome sequencing (srWGS) technology to explore the molecular structure of balancers at base-pair resolution from various different strains available at the *Caenorhabditis* Genetics Center (CGC). Balancer strains carry balancer chromosomes, balanced detrimental variants and additional distinctive small variants that provide visible markers for either a balancer chromosome or the balanced lethal mutations, called morphological markers (Edgley et al., 2006). These markers are used to track worms as homozygotes will display a unique and distinctive phenotype (e.g. Uncoordinated-Unc or Dumpy-Dpy) while heterozygotes will present a wild-type phenotype. In addition, balancer strains available at the CGC carry mutations preserved in a heterozygous state on balancer chromosomes, such as the [bn126](#) variant in the essential gene, [klp-19](#) (Powers et al., 2004), that elicit phenotypes like temperature sensitive sterility, developmental defects, or homozygous lethality or sterility. Using our published srWGS datasets (Maroille et al., 2021, 2022), we have mapped to base-pair resolution morphological markers and balanced mutations for which the genomic position was so far unknown. Such information is crucial to predict the effect of such variants and better understand the molecular mechanism, as well as the genotype-phenotype relationship.

We previously sequenced balancer strains with paired-end Illumina srWGS to map and explore the structure of genetic balancers (Maroille et al., 2021, 2022). Some alleles (balanced mutation and/or markers) reported in the genotype of the strains included in both studies were yet to be mapped at the genomic level. Indeed, while one mutation ([bn126](#)) was known to be a deletion, its genetic position had been interpolated based on phenotypic observations and genetic mapping. We have also included balancer strains created with CRISPR-Cas9 (Dejima et al., 2018) to retrieve the genomic positions for some morphological markers included in those genomes.

Here, we are reporting the genomic location of 12 variants across 12 strains (listed in Figure 1A). Reads associated with the genes affected by each variation were manually and visually examined to identify their potential location. As a proof of concept, four of these variations (deletions [bn126](#), [md2415](#), [tm9709](#), and [tm9723](#)) were confirmed through PCR and Sanger sequencing.

As morphological markers and balanced mutations have been described affecting specific genes to create their unique morphological phenotype or lethality/sterility, we visually analysed the sequence of those genes within their respective strains, in comparison to the genome of [N2](#) which we sequenced and used as a control (Figure 1B). Often, only one variant could be detected in the gene of interest, absent in the [N2](#) genome. This made identification unambiguous, as each variation was associated with only one possible originating mutation. Of these variants, six were base substitutions while six were deletions (including one 1-bp deletion). [bn126](#) was identified as a deletion of 431 bp through position III:13,306,061-13,306,492 (Figure 1B). It causes the truncation of exon 2 of [klp-19](#). [e1466](#) is a G>A transition at position I:2,321,422 (Figure 1B), resulting in a nonsense mutation affecting a glutamine (Q/*). [e1500](#) was identified as a missense mutation causing a G>A transition at position III:3,646,850, thus changing glycine to arginine (G/R). [h1040](#) was found to be a deletion of 572 bp, through position I:14,857,166-14,857,738 (Figure 1B), truncating exon 6 of [unc-54](#). [md2415](#) was also identified as a deletion, removing 2,730 bp through position I:7,450,592-7,453,322, truncating partially exon 24 and exon 30, and entirely exons 25 to 29 of [unc-13](#). [mn153](#) is a C>T transition at position X:16,387,977, falling within an intronic region, and thus causing no amino acid change, but predicted to affect splicing (splice acceptor variant). [s2170](#) was identified as a C>T transition at position III:2,036,920, acting as a missense mutation by changing glycine to valine (G/E). [t1568](#) was found to be a G>A transition at position III:7,024,608, resulting in the creation of a stop codon. [tm9708](#) was revealed to be a G>A transition at position IV:5,509,656 and is an intronic variant. The prediction of its effect would require further functional analyses (RNA-Seq). [tm9709](#) was found to be a deletion of 179 bp through position I:5,432,754-5,432,933, truncating partially exon 1 of [dpy-5](#). [tm9710](#) was identified as a single base deletion at position II:6,714,967, removing a guanine and causing a frameshift (Figure 1B). Finally, [tm9723](#) was found to be a deletion of 90 base pairs through position X:13,215,724-13,215,814, truncating partially exon 8 of [unc-9](#).

Using PCR and Sanger sequencing, we experimentally validated the genomic positions retrieved from the srWGS analysis for four deletions: [klp-19\(bn126\)](#) in [SS746](#), [md2415\(unc-13\)](#) in [RM2431](#), [dpy-5\(tm9709\)](#) in [FX30235](#) and [unc-9\(tm9723\)](#) in [FX30237](#). Primers used are described in Methods and Sanger sequences can be found in Figure 1A. Deletions analysed in this study were found to have been submitted to mutagenesis processes prone to DNA breakage such as UV/TMP and Psoralen (that intercalates in DNA and under UVA exposure, can create covalent interstrand cross-links). The substitutions, however,

were predominantly found in strains exposed to EMS. These observations agree with previous studies (Gengyo-Ando & Mitani, 2000).

Mapping morphological markers and lethal/sterile mutations preserved on balancer chromosomes to the base-pair resolution may give greater insight into how they impact the genes they affect and the phenotype or survival of their associated strains. It also offers the opportunity to genotype strains at molecular level, to complement phenotyping. It has the potential to uncover important gene regions for biochemical pathways and support the documentation of variation, expanding opportunities for future research while solidifying prior efforts.

Methods

Obtaining Strains and maintenance

All strains used in this analysis were provided by the CGC, which is funded by the NIH Office of Research Infrastructure Programs (P40 OD010440). All strains were maintained at 16°C and kept on standard NGM plates streaked with [OP50](#).

Whole Genome Sequencing

Genomic DNA was collected and extracted as described in Maroilley et al. (2022). In brief, the DNA was extracted using Qiagen blood and tissue kit (13323) from ~100 mg of worm tissue and diluted with 10 nM Tris-HCl (pH 8.0). We sequenced the genomes with PCR-free library preparation protocol and NovaSeq 600 Illumina sequencing technology.

Read Alignment and Analysis

FastQC v0.11.9 (Andrews S., 2010) was used to analyze the sequence quality of the fastq files obtained from sequencing, Trimmomatic v0.39 (Bolger et al., 2014) was used to trim adaptors from the end of Illumina reads, and Burrows-Wheeler Aligner (BWA v0.7.17) (Li, 2013) was used to align the trimmed sequences using the Maximal Exact Match (MEM) algorithm.

Identification of Variations

Sequences were analyzed using Integrative Genomics Viewer v2.16 (IGV) to identify potential sources for each variation (Robinson et al., 2011), making use of the WS245 *C. elegans* reference to identify regions associated with the genes of interest. Along with the strains thought to contain the variations, [N2](#) was observed in a similar manner to filter out any mutations contained within, as these surely could not be the cause of the observed phenotypic impact associated with the variations. The effect of variants was predicted using Variant Effect Predictor Ensembl Release 109 (McLaren et al., 2016).

Experimental Validation

We confirmed four of the deletions by PCR and sanger sequencing. The primers used are as follows: for [bn126](#), 5'-TTGTCCACCAAGTGCCGTTTC-3' and 5'-CTCGTCAGTGCGTTGTTCCAAAG-3'; for [md2415](#) 5'-TCCTTAAGACACCCTCCTACA-3' and 5'-AGTCCTTTGCCACCTGTAAA-3'; for [tm9709](#) 5'-CGAGACCTGCCATCGTATTT-3' and 5'-CGAGACCTGCCATCGTATTT-3'; for [tm9723](#) 5'-GTAGTTAGAACTTTCCCGTTCA-3' and 5'-AAACCACAGAAGGAAGGAGAAG-3'.

Reagents

STRAIN	GENOTYPE	AVAILABLE FROM
SS746	klp-19(bn126)/mT1 [dpy-10(e128)] III.	CGC
CB3475	mcm-4(e1466)/szT1 [lon-2(e678)] I; +/szT1 X.	CGC
MT690	nDf6/unc-93(e1500) dpy-17(e164) III.	CGC
KR2839	hDf15 unc-75(e950)/hIn1 [unc-54(h1040)] I.	CGC
RM2431	unc-13(md2415)/hT1 I; +/hT1 V.	CGC
SP423	mnDp1 (X;V)/+ V; unc-3(e151) let-2(mn153) X.	CGC

GE2722	cyk-1(t1568) unc-32(e189)/qC1 [dpy-19(e1259) glp-1(q339)] III; him-3(e1147) IV.	CGC
DP246	unc-45(st601)/sC1 [dpy-1(s2170)] III.	CGC
FX30257	tmC25 [unc-5(tm9708)] IV.	CGC
FX30235	tmC20 [dpy-5(tm9709)] I.	CGC
FX30144	tmC6 [dpy-2(tm9710)] II.	CGC
FX30237	tmC24 [unc-9(tm9723)] X.	CGC

Acknowledgements: This work was supported by funding from Canadian Institute of Health Research, CIHR-Project grant number PJT-156068. The research was enabled by utilizing the Compute Canada (www.computeCanada.ca) computing resources. All strains were provided by the CGC, which is funded by the NIH Office of research Infrastructure Programs (P40 OD01440).

References

- Andrews S. (2010). *FastQC: a quality control tool for high throughput sequence data*. <https://github.com/s-andrews/FastQC>
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120. PubMed ID: [24695404](#)
- Dejima K, Hori S, Iwata S, Suehiro Y, Yoshina S, Motohashi T, Mitani S. 2018. An Aneuploidy-Free and Structurally Defined Balancer Chromosome Toolkit for *Caenorhabditis elegans*. *Cell Reports* 22: 232-241. PubMed ID: [29298424](#)
- Edgley ML, Baillie DL, Riddle DL, Rose AM. 2006. Genetic balancers. *WormBook* : 1-32. PubMed ID: [18050450](#)
- Gengyo-Ando K, Mitani S. 2000. Characterization of mutations induced by ethyl methanesulfonate, UV, and trimethylpsoralen in the nematode *Caenorhabditis elegans*. *Biochem Biophys Res Commun* 269: 64-9. PubMed ID: [10694478](#)
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. <http://arxiv.org/abs/1303.3997>
- Maroille T, Flibotte S, Jean F, Rodrigues Alves Barbosa V, Galbraith A, Chida AR, et al., Tarailo-Graovac M. 2023. Genome sequencing of *C. elegans* balancer strains reveals previously unappreciated complex genomic rearrangements. *Genome Res* 33: 154-167. PubMed ID: [36617680](#)
- Maroille T, Li X, Oldach M, Jean F, Stasiuk SJ, Tarailo-Graovac M. 2021. Deciphering complex genome rearrangements in *C. elegans* using short-read whole genome sequencing. *Sci Rep* 11: 18258. PubMed ID: [34521941](#)
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol* 17: 122. PubMed ID: [27268795](#)
- Powers J, Rose DJ, Saunders A, Dunkelbarger S, Strome S, Saxton WM. 2004. Loss of KLP-19 polar ejection force causes misorientation and missegregation of holocentric chromosomes. *J Cell Biol* 166: 991-1001. PubMed ID: [15452142](#)
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* 29: 24-6. PubMed ID: [21221095](#)

Funding: Canadian Institute of Health Research, CIHR-Project grant number PJT-156068

Author Contributions: Filip Cotra: formal analysis, validation, writing - original draft. Tatiana Maroille: conceptualization, formal analysis, funding acquisition, writing - original draft. Maja Tarailo-Graovac: conceptualization, funding acquisition, supervision, writing - review editing.

Reviewed By: Anonymous

Curated By: Stavros Diamantakis

History: Received April 28, 2023 **Revision Received** May 5, 2023 **Accepted** May 23, 2023 **Published Online** May 26, 2023

Copyright: © 2023 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Cotra, F; Maroilley, T; Tarailo-Graovac, M (2023). Molecular basis of essential and morphological variations across 12 balancer strains in *C. elegans*. microPublication Biology. [10.17912/micropub.biology.000846](https://doi.org/10.17912/micropub.biology.000846)