Supplemental File S1

phiC31 integrase for recombination mediated single copy insertion and genome manipulation in *C. elegans*

Fang-Jung Yang[1,*], Chiao-Nung Chen[2,*], Tiffany Chang[1,*], Ting-Wei Cheng[3], Ni-Chen Chang[1,a], Chia-Yi Kao[1,b], Chih-Chi Lee[1,c], Yu-Ching Huang[1,d], Jung-Chen Hsu[1], Jengyi Li[1], Meiyeh J. Lu[1], Shih-Peng Chan[2,3,#], John Wang[1,#]

[1]Biodiversity Research Center, Academia Sinica, Taipei, Taiwan
[2]Genome and Systems Biology Degree Program, College of Life Science, National Taiwan University, Taipei 10617, Taiwan
[3]Graduate Institute of Microbiology, College of Medicine, National Taiwan University, Taipei 10051, Taiwan

[a]Present Address: Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA
[b]Present Address: Institute of Biomedical Science, Taipei, Taiwan
[c]Present Address: Department of Evolutionary and Environmental Biology, University of Haifa, Haifa, Israel
[d]Present Address: CH Biotech R&D Co., Ltd., Nantou City, Taiwan

* equal contribution
# corresponding authors

Shih-Peng Chan, Ph.D.
Associate Professor
Graduate Institute of Microbiology
College of Medicine
National Taiwan University
Taipei, Taiwan

John Wang, Ph.D.
Associate Research Fellow
Biodiversity Research Center
Academia Sinica
Taipei, Taiwan

List of Supplementary Files, Figures, and Tables:

Supplementary File S1. This file. Supplemental methods and results.
Supplementary File S2. Sequence and annotation for the plasmid used to generate the phiC31 landing pad via MosSCI. SnapGene compatible file.
Supplementary File S3. PhiC31 recombination mediated cassette exchange (RMCE) protocol.
Supplementary File S4. Fasta file containing 7 sequences: attB_retrofitted_Sinv47D10_short (original and partially corrected) and attB_retrofitted_Sinv47D10 (original and partially corrected) BACs as well as the LG II sequences, which carry the *antIs31*, *antIs32*, and *antIs33* sequences, for BRC0566, BRC0906, and BRC0882, respectively.

Supplementary Figure S1. Construction of pCG150_double_phiC31_attB plasmid.
Supplementary Figure S2. Design of dual color operon reporters.
Supplementary Figure S3. Ladder PCR of single-copy insertions of transgenes.
Supplementary Figure S4. Confirmation that phiC31 expressed from the *antIs30* transgene is spliced.
Supplementary Figure S5. phiC31 integrase mediated RMCE using *antIs30* in *C. elegans*.
Supplementary Figure S6. Validation of 33.4 kb BAC integration.
Supplementary Figure S7. Large BAC integration attempt.

Supplementary Table S1. Strains used and generated in this study.
Supplementary Table S2. Information for RMCE and other plasmids used in this study.
Supplementary Table S3. Primers used in this study and associated important sequences.
Supplementary Table S4. PCR assay results for candidate insertion lines derived from the longer BAC, attB_retrofitted_Si47D10

Supplementary Methods and Results

S1A. BAC retrofitting and BAC shortening

For the BAC insertion experiments, the fire ant BAC 47D10 was retrofitted with a cassette containing *Cbr-unc-119(+)*, the kanamycin resistance gene, and two "outward" facing *attB* sites flanking *sur-5p::GFP*, to yield attB_retrofitted_Sinv47D10 using recombineering (Quick & Easy BAC Modification Kit, Gene Bridges). Retrofitting targeted the chloramphenicol resistance gene (*CmR*) on the BAC vector. The retrofitted cassette was made as follows. First, a plasmid (158attB-3-8) was built with two "outward" facing *attB* sites (40 bp) flanking *sur-5p::GFP* using standard cloning. Next, *Cbr-unc-119(+)* was cloned into the *Nhe*I site. After, an Invitrogen Gateway compatible fragment consisting of *attR1*, *CmR*, *ccdB*, and *attR2* was cloned into the *Dra*III site. Then, using Gateway cloning, the kanamycin resistance gene (*kanR*) was swapped into the *attR1 attR2* site. Finally, this complete cassette was amplified with primers with homology overlap to the BAC vector *CmR* gene (Cm_recombi_F and Cm_recombi_R; primer sequences in Sup. Table S3) and the PCR product was gel purified (Viogen gel isolation kit) prior to recombineering. Candidate recombineered BACs could be identified by kanamycin resistance and chloramphenicol sensitivity. The recombineered BAC was checked for the presence of each of the cassette elements by PCR and for the proper insertion junctions.

To shorten the attB_retrofitted_Sinv47D10 BAC, ~104 kb of the fire ant sequence was replaced by *ampR* via recombineering. The primers used to amplify *ampR* from plasmid 158attB-3-8 were Del_mark03_20kB_amp_F and Del_mark03_20kB_amp_R (primer sequences in Sup. Table S3) which have homology to the relevant fire ant sequences. The PCR product was gel purified prior to recombineering. Candidate recombineered BACs could be identified by ampicillin and kanamycin resistance. The shorted BAC, attB_retrofitted_Sinv47D10_short, was validated by flanking PCR around the introduced *ampR* gene.

S1B. Oxford Nanopore Technologies (ONT) sequencing and analysis, methods and results

Sequencing
We used the ONT platform to sequence the two BACs (attB_retrofitted_Sinv47D10_short [hereafter "Short BAC"] and attB_retrofitted_Sinv47D10 ["Long BAC"]) and three strains (BRC0566 (*antIs31*), the landing pad strain; BRC0906 (*antIs32*), carrying a shorter BAC insertion; and BRC0882 (*antIs33*), carrying a longer BAC insertion).

For the two BACs, *E. coli* cultures were grown in 3 ml of LB plus 25 μg/ml kanamycin overnight at 37°C. The next morning this starter culture was added to 27 ml of room temperature LB plus 25 μg/ml kanamycin and incubated at 37°C. After 5 h, BAC DNA was purified using the Qiagen Genomic-tip 20/g kit (Qiagen).

For, Short BAC, two methods of fragmentation were tried. BAC DNA was partially digested with *Eco*RI (37°C for 5 min) or fragmented by tagmentation (Nexteria XT DNA Library Kit). These two

were then barcoded (ONT EXP-NBD104) and sequenced together on an ONT GridION flowcell. Based on the analysis of these two datasets (below), we only used tagmentation for Long BAC. After, Long BAC was barcoded and sequenced as a pool with BRC0566 and BRC0906 (below).

For the three strains, genomic DNA (gDNA) was isolated using the phenol:chloroform method. Since the bulk of the gDNA was already <20 kb, samples were directly prepared for sequencing (i.e., no additional fragmentation). gDNA samples from BRC0566 and BRC0906 were barcoded (ONT EXP-NBD104), pooled along with Long BAC, and then sequenced on an ONT GridION flowcell. The gDNA samples for BRC0882 and another aliquot of BRC0906 were not barcoded and were instead prepared using the Ligation Sequencing Kit (ONT SQK-LSK109). These were run serially on a flowcell, with a wash in between.

Bioinformatics analysis
Short BAC assembly
*Eco*RI digested and tagmented Short BAC sequence reads were assembled separately with the first 5,000 reads using canu (v1.4, with parameter "`genomeSize=0.1m`") (Koren et al. 2017). An assembly was also done using the pooled first 5,000 reads from each of the samples (i.e., 10,000 reads).

The *Eco*RI digested sequencing reads were dominated by read lengths corresponding to fully digested *Eco*RI digested fragments, despite partial digestion, so it was unclear how well this sample would assemble. Both the *Eco*RI only and the pooled assemblies failed to produce a complete BAC assembly. Thus, these data were not analyzed further (and also not included in the NCBI short read archive). The assembly using tagmented DNA sequence reads produced a complete assembly. Because the BAC ends were known from Sanger sequencing, we could verify that the assembly was correct by comparing to the fire ant reference genome. Next, for polishing, the first 50,000 reads from the tagmentation data set were mapped against the assembly using minimap2 (v2.15, with parameter "`-ax map-ont`") (Li 2018) and then polished 1x using Racon (v1.3.2) (Vaser et al. 2017).

Long BAC assembly
The first 5,000 reads were assembled as above with canu except using parameter "`genomeSize=0.15m`". Polishing was done, as above, except with the entire sequence data set and 5x.

Analysis of sequencing results for the landing pad strain BRC0566
We manually added the predicted sequence of the landing pad into LG II of the *C. elegans* genome (version WS280) and then used minimap2 to map the ONT reads to the modified genome. In an initial analysis, samtools (Danecek et al. 2021) was used to convert the mapping output file from sam to bam format, sort, and then index the reads. Then using IGV (Robinson et al. 2011) to visualize the mapping, we noticed that the regions of the landing pad corresponding to *GFP* and *unc-119(-)* had extra read depth compared to the flanking region (Fig. A).
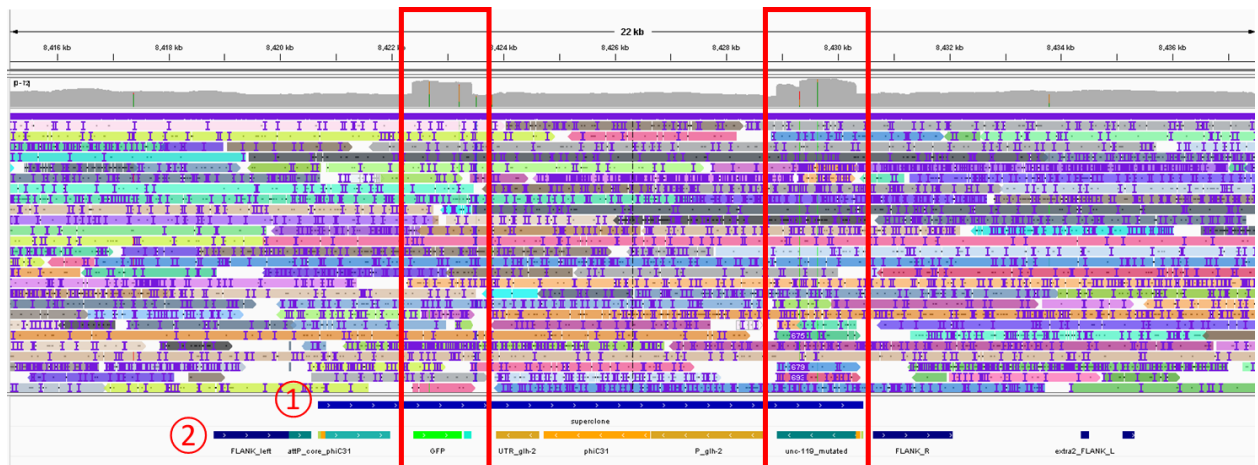
Fig. A. Unfiltered ONT reads mapping to the landing pad strain BRC0566. IGV view of ONT reads. Most of the landing pad has depth similar to the flanking LG II genomic sequence, however *GFP* and *unc-119(-)* have inflated depth. This extra depth is contamination from Short BAC (attB_retrofitted_Sinv47D10_short). ① Blue bar indicates extent of the landing pad. ②Other colored segments denote genes within the landing pad or flanking regions.
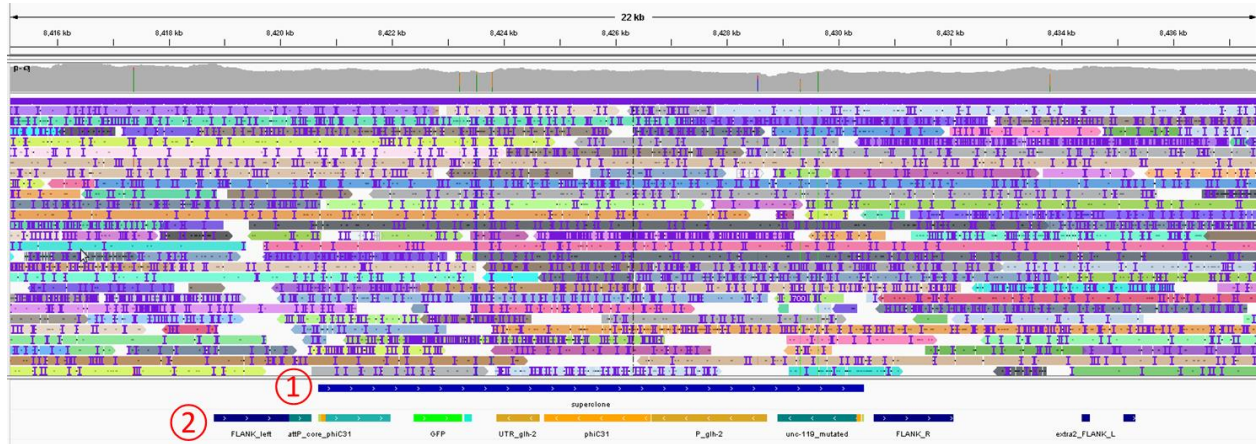

Fig. B. Filtered ONT reads mapping to the landing pad strain BRC0566. IGV view of ONT reads. Reads were filtered if there was ≥500 of soft clipped bases on either side. Coverage is now uniform across the entire landing pad, as expected. ① Blue bar indicates extent of the landing pad. ②Other colored segments denote genes within the landing pad or flanking regions.

Inspection of the reads revealed that many reads mapping to *GFP* or *unc-119(-)*had long unmapped (soft clipped) flanking segments, and these segments would map best to Short BAC (which was in the same pool). Therefore, we added a filtering step to remove such reads based on having long soft clipped segments (samclip (Seemann) with parameter "`--max 500`"). Inspection of the post-filtered reads with IGV showed that the landing pad had coverage depth that was similar to the flanking region (Fig. B). Furthermore, three very long reads crossed the

entire landing pad cleanly. Together these results confirm that the *antIs31* landing pad is complete and in one copy, as expected.

Analysis of sequencing results for the Short BAC insertion strain BRC0906 (*antIs32*)
We manually added the predicted assembled sequence between the *attB* sites of Short BAC into the landing pad of LG II of the BRC0566 genome (above). In the first round of sequencing we used barcoded ONT reads that were pooled with Short BAC and BRC0566. Initial analysis, as above, of mapped reads to BRC0906 also indicated contamination from Short BAC. However, in this case, since the *antIs32* insertion overlaps extensively with Short BAC, simple filtering could not remove the contaminating reads. These data were not analyzed further (and also not included in the NCBI short read archive).

Thus, we sequenced a second aliquot of the same gDNA sample for BRC0906 by itself without barcoding. In this reanalysis we added two bacterial genomes to the reference genome which were contributing appreciable amounts of reads to this sequencing run. The first is the food *E. coli* OP50, and the other is *Stenotrophomonas maltophilia* strain ISMMS3 (possibly contaminating the NGM plates). We also only mapped reads ≥5,000 bp, to remove short reads and to eliminate the 3.6 kb lambda phage ONT internal control, which has overlap with an ~190 bp segment of the inserted BAC vector. We did not remove reads with long soft clipped segments, and in fact there were very few such reads, as there was no (or very little) contamination from other samples. This mapping result is in main Fig. 3C and shows that the *antIs32* is a very large, precise, single copy integration of ~33.4 kb.

Analysis of sequencing results for the Long BAC insertion strain BRC0882 (*antIs33*)
We manually added the predicted assembled sequence between the *attB* sites of Long BAC into the landing pad of LG II of the BRC0566 genome (above). We included the genomes from *E. coli* OP50 and *S. maltophilia* as part of the reference. We then filtered and mapped the reads to the BRC0882 genome as for the second BRC0906 sample. This mapping result is in Sup Fig. 7C and shows that *antIs33* is a complex, multicopy (2-4x) insertion.

Data availability for ONT analysis
Because Short BAC and Long BAC were assembled and polished only with ONT sequencing reads, there are still errors in the sequence. We used these sequences for building the BRC0906 and BRC0882 genomes. Since those analyses were primarily for assessing depth of coverage, error-free sequence was less important. These "original" assembled and polished sequences are called "attB_retrofitted_Sinv47D10_short_ORIGINAL" and " attB_retrofitted_Sinv47D10_ORIGINAL" in the fasta file Sup. File S4.

The BAC sequences were manually corrected for errors in the landing pad gene coding regions before submitting to NCBI GenBank. There are likely other errors. For attB_retrofitted_Sinv47D10_short, the *GFP* sequence was corrected by examining the raw reads in IGV; the correct sequence was clearly in the majority. For attB_retrofitted_Sinv47D10, both the *GFP* and *kanR* genes were corrected similarly. These sequences are called

"attB_retrofitted_Sinv47D10_short_PARTIALLY_CORRECTED_GBK_OK668367" and "attB_retrofitted_Sinv47D10_PARTIALLY_CORRECTED_GBK_OK668368" in Sup. File S4.

LG II for BRC0566, BRC0906, and BRC0882 are also in Sup. File S4. To reduce the file size, the other chromosomes are not included in Sup. File S4. The *C. elegans* genome used was version WS280. The *E. coli* OP50 was genome is from accession CP059137 (version CP059137.1). The *S. maltophilia* strain ISMMS3 genome is from accession CP011010 (version CP011010.1)

Raw Oxford Nanopore sequence data are available at NCBI SRA under accession numbers: SRR16511097, SRR16511098, SRR16511099, SRR16519834, and SRR16519835 (Project ID: PRJNA773226). The GenBank accession number for BAC attB_retrofitted_Sinv47D10_short is OK668367 and for BAC attB_retrofitted_Sinv47D10 is OK668368.

S1C. Generation of chromosomal inversion on LG IV

To generate a large chromosomal inversion, we first inserted an *attP* site into the intron of *dpy-13* and an *attB* site into an intron of *unc-30*, as follows. The intronic *dpy-13* or *unc-30* target sgRNA, the co-CRISPR *dpy-10* sgRNA, and appropriate repair templates were injected into N2 worms, and then Roller (Rol) F1 individuals were cloned out. After >8 eggs had been laid, the F1 individuals were genotyped by single worm PCR using flanking primers for putative insertion of the *attB* or *attP* sequence (primers and *attB/P* sequences are in Sup. Table S3). For positive F1 individuals, at least 8 F2 individuals were singled out to obtain homozygous insertions, which were again determined by single worm PCR. From homozygous F2 lines, non-Rol, non-Dpy-10 F3 individuals were picked to establish strains for the insertion alleles. Insertion alleles were then validated by sequencing of the PCR products obtained using primers flanking the insertion site. The insertion alleles obtained were *antSi50* (*attP*, BRC0658) in *dpy-13* and *antSi51* (*attB*, BRC0601) in *unc-30*; these strains are wild-type. Both strains were backcrossed to N2 3x to generate BRC0609 and BRC0669.

Recombination was used to place the *attB* and *attP* insertions onto the same chromosome to generate BRC0671 (*antSi50 antSi51*) as follows. We first built *antSi50 unc-30* and *dpy-13 antSi51* double mutants. To do this, N2 males were crossed to *unc-30*(*ok613*) hermaphrodites to obtain *unc-30/+* F1 males that were then crossed to *antSi50* [*attP* at *dpy-13*] hermaphrodites. Half of the F2 should be *antSi50 +/+ unc-30* and can be identified by their segregation of Unc F3. Multiple Unc F3 were singled out and allowed to lay at least several eggs and then the F3 parents were genotyped by PCR for *antSi50*. Such rare, positive, recombinant F3 individuals would be *antSi50 unc-30/+ unc-30*. We then screened the F4 by PCR for homozygous *antSi50 unc-30* double mutants. *dpy-13(e184) antSi51* was obtained similarly. Next, we crossed N2 males to *dpy-13 antSi51* hermaphrodites to obtain *dpy-13 antSi51/+ +* F1 males that were then crossed to *antSi50 unc-30* hermaphrodites. Half of the nonDpy nonUnc F2 should be *antSi50 unc-30/dpy-13 antSi51* and can be identified by their segregation of Unc F3. From these plates, several nonDpy nonUnc F3 were singled out. Recombinant F3 individuals would be *antSi50 unc-30/antSi50 antSi51* or *antSi50 antSi51/dpy-13 antSi51*; such individual would segregate only 1/4 Unc or 1/4 Dpy (i.e.,

no Dpy Unc) individuals in the F4. We then picked multiple non-Unc or non-Dpy F4 siblings to obtain homozygous *antSi50 antSi51* double mutants, which was verified by PCR. One of the lines was designated BRC0671.

The chromosomal inversion was generated as follows. First BRC0566 (*antIs31; unc-119(ed9)*) was outcrossed to N2 to remove *unc-119*(*ed9*) to yield BRC0664. Next, N2 males were crossed to BRC0664 and then heterozygous F1 males were crossed to BRC0671 to produce *antIs31/+; antSi50 antSi51/+ +* or *+/+; antSi50 antSi51/+ +* F2 cross progeny. *myo-2p::GFP* is difficult to detect in adults, so it was scored in the F3 larval progeny. F3 and F4 progeny were screened for Dpy Unc individuals, indicative of an inversion mediated by phiC31 integrase. Inversions were confirmed in the independently derived Dpy Unc lines by PCR and sequencing of the junctions.

Recombination rates between *dpy-13* and *unc-30* in the inversion allele, *antIn1*, and in the wild-type orientation (*dpy-13(e184) unc-30(ok613)*) were determined from the recombinant offspring of the *dpy-13 unc-30/+ +* heterozygotes. First, N2 males were crossed to the respective Dpy Unc strains and 5 or 6 heterozygous *dpy-13 unc-30/+ +* F1 offspring were singled out. Next, from each F1, 20-30 non-Dpy non-Unc F2 progeny were singled out. Then, F2 genotypes were scored by examining the F3 segregation pattern; F2 recombinant genotypes would reflect recombination events in the F1 mother. We chose this approach rather than scoring phenotypically visible recombination events (Dpy non-Unc or Unc non-Dpy) directly in the F2 generation because we were concerned of possible mis-scoring among Dpy non-Unc, Unc non-Dpy, and Dpy Unc individuals. Additionally, these three classes of individuals sometimes die after having only a few progeny, making verification of the mother F2's phenotype problematic.

Parent genotype: *dpy unc/+ +*

Progeny genotypes

		(1-r)/2	(1-r)/2	r/2	r/2
		d u	*+ +*	*d +*	*+ u*
(1-r)/2	*d u*	x	pick	x	x
(1-r)/2	*+ +*	pick	pick	pick	pick
r/2	*d +*	x	pick	x	pick
r/2	*+ u*	x	pick	pick	x

Fig. C. Punnett square of progeny genotypes, their expected frequencies, and progeny examined. For the recombination analysis, only nonDpy nonUnc progeny picked, as indicated. White, orange, and green indicate different probabilities of obtaining that genotype combination. Orange and green also indicate recombinant genotypes. Recombinant genotypes / total picked is equal to the frequency of observed plates with recombinants and is $(4r-2r^2)/(3-2r+r^2)$. *d, dpy*; *u, unc*, r, recombination rate; x, not picked.

Then the recombination rate was calculated based on the fraction of non-Dpy non-Unc F2 that segregated non-Dpy Unc or Dpy non-Unc and the Punnett Square expectations for single recombinants per chromosome (i.e., no double or more recombination events). In this case,

observed recombinants (14/143 or 0.0979) = $(4r-2r^2)/(3-2r+r^2)$, where, r is recombination rate. Rearranging and solving for the quadratic formula, then r is 7.27% or 7.27cM.

S1D. Additional details on longer BAC integration attempt which would insert ~137 kb

The ability to generate very large, precise, single copy transgenes (10s to 100+ kb) at specific locations would facilitate more complex experiments. Using phiC31 mediated RCME, we succeeded for a shorter ~33.4 kb BAC transgene (main text and main Fig. 3), which is fosmid-sized (up to ~40 kb).

BACs can be much bigger (~100 to ~350 kb). Thus, to test if we could precisely integrate a >100 kb transgene, we attempted to integrate the longer BAC, attB_retrofitted_Sinv47D10. This BAC was obtained by retrofitting the fire ant BAC clone 47D10 with a cassette that contains *Cbr-unc-119(+)*, *sur-5p::GFP*, and two *attB* sites. The *attB* sites are oriented such that the entire BAC (including vector sequence), except for *sur-5p::GFP*, would be inserted upon proper RMCE (Sup. Fig. 7A), resulting in a ~137 kb insertion.

We injected this retrofitted BAC either with a *myo-3p::mCherry* co-injection marker (n=319) or with the addition of spermine (30 µM) and spermidine (70 µM) but without a co-injection marker (n=110) into the phiC31 landing pad strain and obtained 138 F1 non-Unc individuals. We obtained 22 integrated lines (18 with co-injection marker and 4 with spermine and spermidine) based on segregating exclusively non-Unc progeny. Of these, 18 lines were from independent F1's. The last four lines were not completely independent, with two lines derived from each of two F1's. In both pairs, although they had the same F1 ancestor, the lines had different integration outcomes (Sup. Table S4). We conducted PCR assays for proper insertion for all the lines, and these results as well as their fluorescence phenotypes and additional information are in Sup. Table S4.

Of the 22 integrated lines, three did not express GFP or mCherry (main Table 1), which would be expected of potential RMCE integrants with this BAC. However, none were proper, clean single copy insertions after further validation. For one line, BRC0780 (*antIs48*), PCR assays showed that only one junction was correct and only *Cbr-unc-119* was inserted, indicating that the insertion was incomplete and possibly complex (since the other junction did not yield a PCR result). For a second line, BRC0766 (*antIs40*), although the junction assays were consistent with RMCE, there were deletions within the BAC. The third line BRC0882 (*antIs33*) appeared correct based on the PCR assays. To investigate this strain further, we conducted Southern blot assays with three probes (Sup. Fig. S7B). Two of the probes hybridized to only a single fragment, which would be consistent with a single copy or a low copy number insertion with no rearrangement in this region. However, a third probe hybridized to three fragments, indicating that multiple copies of at least the probe-targeting region was inserted. Finally, we sequenced this strain using the Oxford Nanopore Technologies (ONT) platform. Based on sequencing depth, the *antIs33* insertion carries approximately 2-to-4 copies depending on transgene region (Sup. Fig 7C).

We observed GFP fluorescence in the remaining 19 lines. Three lines also expressed mCherry. To try to understand these integration events better, we conducted PCR analysis (Sup. Table S4). Seven of these lines had proper junctions, were positive for all BAC PCR assays, and had lost phiC31, consistent with phiC31 mediated RMCE insertion of the complete BAC. However, as GFP could still be observed and given the results for the *antIs33* insertion, these integrants likely carry multiple copies. The remaining 12 lines yielded positive PCR assays for only one proper junction, two junctions but of opposite orientation, or three junctions. The last is difficult to explain but one possibility is heterozygosity for two insertions, one on each homologous chromosome. We did not pursue this possibility. Of these 12 lines, six carried internal deletions of the BAC. These results indicate that the insertions were complex and are compatible with the model suggested by Nonet (Nonet 2020) that the formation of extrachromosomal arrays, composed of rearrangements within the BAC and/or recombination with the co-injection marker, often occurs prior to integration that probably occurs in the F1.

Overall, these injection results indicate that inserting single copy >100 kb BAC-sized constructs through phiC31 mediated recombination in *C. elegans* is not readily possible with our current injection conditions. One reason may be that BACs are difficult to purify without accidental shearing. Likewise, shearing can occur during the injection step. Spermine and spermidine are often added to the injection buffer to protect BACs during injections in mice (Van Keuren et al. 2009). We added spermine and spermidine for injections of 110 (of 429 total) P0's, but all four insertion lines were incorrect (Sup. Table S4). The concentration of DNA injected may also be important (Van Keuren et al. 2009). Optimization in these and other areas may ultimately permit successful phiC31 mediated integration of >100 kb BACs.

References

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM et al. 2021. Twelve years of samtools and bcftools. Gigascience. 10(2).

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27(5):722-736.

Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 34(18):3094-3100.

Nonet ML. 2020. Efficient transgenesis in caenorhabditis elegans using flp recombinase-mediated cassette exchange. Genetics. 215(4):903-921.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat Biotechnol. 29(1):24-26.

Samclip. c2020. [accessed 2021 August]. https://github.com/tseemann/samclip.

Van Keuren ML, Gavrilina GB, Filipiak WE, Zeidler MG, Saunders TL. 2009. Generating transgenic mice from bacterial artificial chromosomes: Transgenesis efficiency, integration and expression outcomes. Transgenic Res. 18(5):769-785.

Vaser R, Sovic I, Nagarajan N, Sikic M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27(5):737-746.