Improving Transgenesis Efficiency and CRISPR-Associated Tools Through Codon Optimization and Native Intron Addition in *Pristionchus* Nematodes

Ziduan Han,1 Wen-Sui Lo,1 James W. Lightfoot, Hanh Witte, Shuai Sun, and Ralf J. Sommer2

Max Planck Institute for Developmental Biology, Tuebingen 72076, Germany

ORCID IDs: 0000-0001-8341-0678 (Z.H.); 0000-0002-2438-0015 (W.-S.L.); 0000-0002-5835-2135 (J.W.L.); 0000-0003-1503-7749 (R.J.S.)

**ABSTRACT** A lack of appropriate molecular tools is one obstacle that prevents in-depth mechanistic studies in many organisms. Transgenesis, clustered regularly interspaced short palindromic repeats (CRISPR)-associated engineering, and related tools are fundamental in the modern life sciences, but their applications are still limited to a few model organisms. In the phylum Nematoda, transgenesis can only be performed in a handful of species other than *Caenorhabditis elegans*, and additionally, other species suffer from significantly lower transgenesis efficiencies. We hypothesized that this may in part be due to incompatibilities of transgenes in the recipient organisms. Therefore, we investigated the genomic features of 10 nematode species from three of the major clades representing all different lifestyles. We found that these species show drastically different codon usage bias and intron composition. With these findings, we used the species *Pristionchus pacificus* as a proof of concept for codon optimization and native intron addition. Indeed, we were able to significantly improve transgenesis efficiency, a principle that may be usable in other nematode species. In addition, with the improved transgenes, we developed a fluorescent co-injection marker in *P. pacificus* for the detection of CRISPR-edited individuals, which helps considerably to reduce associated time and costs.

**KEYWORDS** nematodes; *C. elegans*; *P. pacificus*; transgenesis; CRISPR editing; codon usage bias; intron-mediated enhancement; parasitic nematodes

---

The utilization of transgenes has proven fundamental to many aspects of molecular biology and for functional genomic studies (Rubin and Spradling 1982; Mello et al. 1991; Chalfie et al. 1994; Clough and Bent 1998; Hutter 2012). For instance, easily applied and efficient transgenic methods have been instrumental in furthering our understanding of biological pathways and dissecting associated phenotypes. Additionally, it has facilitated the visualization of gene expression patterns and protein localization through the usage of fluorescent proteins such as GFP in a swathe of organisms (Chalfie et al. 1994). However, limiting factors for the successful establishment of transgenesis in an organism are the differing regulatory strategies and mechanisms found between species. In accordance with this and despite their ubiquitous usage, efficient transgenesis tools are frequently restricted to canonical model organisms.

Gene expression, including transgene expression, is regulated by a multitude of factors, including at the transcriptional and translational levels. One such regulatory mechanism is through codon usage bias (CUB). Here, the degenerate nature of the nucleotide triplet code ensures that each amino acid can be encoded by several synonymous codons, with the exception of the amino acids methionine and tryptophan (Sharp and Li 1987). Correspondingly, organism genomes show their own distinct usage of the code. This codon bias is more pronounced in genes with elevated expression levels. Specifically, highly expressed genes strongly favor a specific set of codons with the favored codons contributing to a more efficient translation process through faster ribosome elongation (Duret and Mouchiroud 1999; Plotkin and Kudla 2011). Further, artificial manipulation of the CUB can also alter gene expression dramatically (Redemann et al. 2011). In addition to CUB, regulatory regions of a gene are also thought to be
crucial for transcriptional control as evidence suggests a relationship between the exon–intron structure of a gene and its expression through a process termed “intron-mediated enhancement” (IME). Here, intron density positively correlates with both the level and extent of a gene’s expression (Castillo-Davis et al. 2002). As such, these phenomena have been exploited for the enhancement of molecular tools including improving transgenesis in a number of well-studied model organisms (Brinster et al. 1988; Bischof et al. 2007).

One such organism is the nematode Caenorhabditis elegans, where an abundance of molecular tools, including transgenesis, are available and its CUB and exon–intron structures are well characterized (Ragle et al. 2015). In particular, IME in C. elegans is strongly influenced by the position, number, and sequence of introns, and introns positioned near the 5’ end of a gene shows the greatest contribution to this effect (Okkema et al. 1993; Crane et al. 2019). Further, replacing native codons with favored codons increases the translation level of a protein (Redemann et al. 2011). However, C. elegans is far from the only nematode of significance in the phylum, with an array of parasitic nematodes of both animals and plants, as well as other free-living nematodes, also now frequently used for research. Despite this, transgenesis has only been successfully applied to a few nematode species outside of the genus Caenorhabditis (Higazi et al. 2002; Li et al. 2006; Schlager et al. 2009; Lok 2012), with problems arising due to efficient delivery of DNA materials to the gonad (Evans 2006) and compatibility of the DNA to the endogenous genetic machinery of the recipient. Therefore, despite recent advancements (Adams et al. 2019), low efficiency is still the bottleneck for most transgenic experiments in other nematode species.

In addition to C. elegans, another distantly related free-living nematode frequently used for research is Pristionchus pacificus (Sommer et al. 1996). This nematode has been established as a model system to study evolutionary developmental biology and, more specifically, the evolution of novelty. This is due to the nature of its mouth structure, which is phenotypically plastic and demonstrates two distinct variants. One morph exhibits two teeth while the other contains only a single tooth. The genetic network behind this developmental decision has been extensively studied and is heavily influenced by the nematode’s environment (Ragsdale et al. 2013; Kieninger et al. 2016; Bui et al. 2018; Sieriebriennikov et al. 2020). The presence of teeth in P. pacificus facilitates an additional behavior as they are capable of predating the larvae of other nematodes. Here, it has been observed that the mouth-form dimorphism strongly correlates with the predation behavior; as only the morphs possessing two teeth are active predators, whereas the single-toothed morphs are strict bacterial feeders (Wilecki et al. 2015; Moreno et al. 2019; Akduman et al. 2020). Furthermore, the predatory behavior coincides with the existence of a self-recognition system (Lightfoot et al. 2019) and environmental responses distinct from C. elegans (Hong and Sommer 2006; Moreno et al. 2016, 2017).

Outside of C. elegans, P. pacificus is arguably the most advanced nematode system in terms of the availability of molecular tools (Schlager et al. 2009; Witte et al. 2015; Okumura et al. 2017; Loer et al. 2019). However, previous methodologies resulted in low efficiencies of P. pacificus transgenics, with on average one to three F1 Roller(s) per 40 injected P0s (Schlager et al. 2009). Thus, P. pacificus suffers from a much less efficient transgenesis system compared with C. elegans for several potential reasons. First, it relies on the formation of complex arrays, which incorporate transgene DNA, genomic DNA fragments that must come from P. pacificus itself, and a co-injection marker, to be carried as heritable chromosome fragments. Second, the current versions of fluorescent proteins utilized in P. pacificus (Schlager et al. 2009) have not been adapted to its specific CUB and no attempts have yet been made to improve these fluorescent proteins further by investigating any potential IME. Together, these factors likely contribute to the varying degrees of generational transmission observed in P. pacificus transgenesis experiments and will likely hinder the successful development of other transgenic techniques including additional fluorescent proteins, calcium imaging, and optogenetics.

In this study, using publicly available data sets, we first computed the CUB and global intron structure in 10 nematode species to investigate the conservation of these factors across the phylum, making use of the most recent genomic and transcriptomic data sets. For this, we selected species living in different ecosystems including parasites of animals and plants. As each nematode species shows a distinct CUB and potential IME, we focused on P. pacificus and utilized these factors together with its spliced leaders (SLs), a specific but conserved transcriptional regulatory element in nematodes (Denker et al. 2002), to improve the efficiency of transgenesis in this species. Finally, with the improved transgenesis in P. pacificus, we established a new method using a fluorescent co-injection marker to identify potential clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-edited candidates, reducing the workload and cost for CRISPR/Cas9 screening.

**Materials and Methods**

**Obtaining genome annotations and transcription profiles**

We collected published annotations and transcriptomes of 10 nematode species representing three of the five major nematode clades (Blaxter et al. 1998): C. elegans (Lee et al. 2018; Liu et al. 2019) [WormBase web site (https://wormbase.org), release WS271 2019], C. briggsae (Grün et al. 2014), Haemonchus contortus (Laing et al. 2013), P. pacificus (Prabh et al. 2018; Rödelsperger et al. 2019), P. fissidentatus (Prabh et al. 2018; Rödelsperger et al. 2018), Strongyloides ratti (Hunt et al. 2016), Globodera pallida (Cotton et al. 2014), Bursaphelenchus xylophilus (Kikuchi et al. 2011; Tanaka et al. 2019), Brugia malayi (Choi et al. 2011; Foster et al. 2020), and Ascaris suum (Wang et al. 2011, 2017). To acquire the expression profiles of P. pacificus and P. fissidentatus, we retrieved RNA-sequencing (RNA-seq) data sets of P. pacificus and P. fissidentatus from the Sequence
Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra/; Supplemental Material, Table S1). We mapped raw reads to the reference genome of each species (Table S1) using Hisat2 (Kim et al. 2015) with default parameters, and quantified the numbers of reads mapping to each annotated gene using the package featureCounts (Liao et al. 2014). For other species, the gene expression results were directly downloaded from WormBase (Howe et al. 2016, 2017). Detailed information of the metadata is summarized in Table S1. For C. elegans and A. suum, whose annotations included isoform data, only the longest transcripts were used in downstream analyses.

**Codon usage computation**

To identify the CUBs of genes with different expression levels, the percentage codon usage for each gene was calculated using cusp from EMBOSS suite (Rice et al. 2000). We optimized the codon of proteins based on the most preferred codons of genes with high expression levels in *P. pacificus*.

**P. pacificus trans-spliced messenger RNA identification**

To identify the *P. pacificus* transcripts that contain SLs, we first performed RNA-seq using a ribosomal RNA (rRNA) depletion library. Briefly, total RNA of *P. pacificus* was extracted via Direct-zol RNA Miniprep (Zymo Research) and a RibZero rRNA Removal Kit (Human/Mouse/Rat; Illumina), and RNA libraries were constructed using the ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina). Sequencing was carried out on an Illumina HiSeq 3000 sequencer with one-sixth of a lane. We used Trinity (Grabherr et al. 2011) for *de novo* transcriptome assembly, and identified the transcripts with SLs by the consensus SL sequences at 5’ ends (SL1: TACC CAAGTTTGAG; and SL2: CAGTATCTCAAG) (Blaxter 2006). We used MEME SUITE (Bailey et al. 2009) to identify the motifs of 3’ sequences of the trans-splice sites.

**Statistics**

We performed the chi-square test to test whether the frequencies of synonymous codons in the most highly expressed genes (11th bin) were deviated from the frequencies of genome-wide synonymous codons. We performed the one-tailed Kolmogorov–Smirnov test to compare the intron length distributions between *C. elegans* and the other species. We calculated the Pearson’s correlation coefficient to measure the linear relationship between intron length and gene expression level. We performed the Wilcoxon signed-rank test to test whether genes that contained SLs were different in expression level from genes without SLs.

**Plasmid construction and microinjection**

The optimized *egl-20p::GFP* and *egl-20p::TurboRFP* (red fluorescent protein) were modified based on a pUC19 backbone from a previous study (Schlager et al. 2009). Full sequences of these plasmids in text files can be found in the supplemental materials. Modified GFP and TurboRFP sequences were synthesized from Integrated DNA Technologies (IDT; Coralville, IA) and cloned into the pUC19 backbone using Gibson Assembly Master Mix (New England Biolabs, Beverly, MA) following the manufacturers’ protocols. Plasmids were extracted using the QiAprep Spin Miniprep kit (QIAGEN, Valencia, CA). Three introns from the rRNA gene *Ppa-rps-1* (gene ID: PPA18896; El paco annotation v2) (Rödelberger et al. 2019) were added into the sequence of GFP or turboRFP from 5’ to 3’ and were roughly evenly spaced (“Fire Lab Vector Kit 1995”): intron 1, gtgacgctctctgttgtaatgggtgtgtgaaactttaataat cattggattcttaacctatattttttcg; intron 2, gtgaagctatacattagcggtctttatcttgtatcctgggggttggttggagaggagatatattatatttttaaatattttcag; and intron 3, gtaggtggctgctatatttagctagcagatgaacttttttcag. For the two-intron codon-optimized *egl-20p::GFP* and *egl-20p::TurboRFP*, intron 1 (the most 5’ intron) was removed using a Q5 Site-Directed Mutagenesis kit (New England Biolabs), and both intron 1 and 2 were removed for the one-intron *egl-20p::GFP* and *egl-20p::TurboRFP*.

*P. pacificus* microinjections were performed following the standard protocol (Schlager et al. 2009; Witte et al. 2015). Plasmids were diluted to 50 ng/μl for microinjection using TE buffer. Well-fed *P. pacificus* (strain PS312) young hermaphrodites (preferably not carrying any eggs) were used for injections. The injection mix for the co-injection marker-assisted CRISPR/Cas9 editing was modified from those of Witte et al. (2015) and Dokshin et al. (2018), and the mix contained 0.5 μg/μl Cas9 nuclease (catalog# 1081058; IDT), 0.1 μg/μl trans-activating CRISPR RNA (catalog# 1072534; IDT), 0.056 μg/μl guide RNA (CRISPR/Cas9 RNA; IDT), and 0.05 μg/μl co-injecting plasmid. Potential CRISPR-edited alleles were amplified by PCR and sequenced using Sanger sequencing. Alternatively, the PCR amplicons were run on a 4% TBE agarose gel to detect heteroduplex formation (Bhattacharya and Van Meir 2019).

**Data availability**

The raw sequence data of *P. pacificus* rRNA-depleted RNA-seq have been deposited at the SRA under BioProject identified PRJNA658248. Supplemental material available at figshare: https://doi.org/10.25386/geneic.13090322.

**Results**

**Codon usage is divergent among nematode species**

To enhance transgene expression in diverse nematode species, we first obtained a comprehensive view of CUB in nematodes. For that, we calculated CUB in 10 species of eight nematode genera, representing three of the five major clades of the phylum Nematoda (Figure 1A and Figure S1; Blaxter et al. 1998). For a given amino acid, we found a favored codon in highly expressed genes in every species. The frequencies of different codons in the most highly expressed genes (11th bin) deviate significantly from their genome-wide frequencies (for all comparisons, 0 < P < 4.19×10−11, chi-square test). There was no clear pattern between CUBs and phylogenetic relationships or lifestyles (free-living or parasitic). However, genome-wide GC content may be
Figure 1. Codon preferences in nematodes species as a function of expression levels. (A) The codon usage bias of C. elegans, H. contortus, P. pacificus, S. ratti, G. pallida, B. xylophilus, B. malayi, and A. suum. The protein-coding genes are binned based on the transcripts per kilobase million value from expression level low to high with a log2 scale into 11 bins (x-axis). The dots represent the average codon usage frequency of a given bin. (B) Gene grouping and codon usage bias for P. pacificus, P. fissionatus, C. elegans, and C. briggsae. Figures in the first row show the number of genes (y-axis) grouped from low to high expression with a log2 scale into 11 bins (same as A).
one major factor correlating with the CUB (Mitreva et al. 2006). For example, *S. ratti* and *B. malayi* have low-GC-content genomes, and subsequently the codon usage is also biased toward AT-rich codons. In these species, GCA and GCT are more preferred than GCC and GCG for alanine. Intriguingly, species with a similar GC content can still exhibit drastically different patterns in CUB. For example, *P. pacificus* and *H. contortus*, which both have ~43% GC content, show differing codon preferences for coding proline and alanine (Figure S1). Thus, our new analysis of CUB confirms previous studies that codon usage is divergent among nematode species. Note that the species considered here belong to very different nematode taxa and are phylogenetically only distantly related.

**Codon usage adaptation is conserved within genera**

To study the evolution of codon usage between more closely related nematodes, we focused on two well-studied nematode genera *Caenorhabditis* and *Pristionchus*. *C. elegans* and *P. pacificus* share a common ancestor around 100 million years ago, and they have a distinct CUB. The most dramatic examples of this can be seen in the amino acids glutamine, glutamic acid, and lysine, where *P. pacificus* and *C. elegans* favor the opposing codons. However, within the genus *Pristionchus* the CUB appears conserved, as in *P. dissidentatus*, also a species within the *Pristionchus* genus (Rödelsperger et al. 2018), and we found it shares a highly similar CUB with that observed in *P. pacificus* (Figure 1B and Figure S2). Similarly, in *Caenorhabditis*, the CUB is conserved between *C. elegans* and *C. briggsae* (Figure S3). This finding strongly suggests that codon usage adaptation evolved more ancestrally than the speciation events within the genera *Pristionchus* and *Caenorhabditis*, and that CUB is conserved between closely related species.

**Global intron structure and SL1 frequency shows distinct patterns**

As the presence and distributions of introns also contribute to gene regulation (Castillo-Davis et al. 2002), we next

---

**Figure 2** Global intron length distribution in diverse nematodes. (A) The intron length distribution of eight nematode species. Vertical lines indicate the median lengths of introns of each species, while the numbers in gray indicate the modes of intron lengths (bp). (B) Median intron length as a function of the gene expression. The protein-coding genes are binned by expression level from low to high with a log scale into 11 bins. (C) Elevated expression level of SL1-operated genes. Consensus sequence of the SL1 trans-splice sites in *P. pacificus* (top). SL1- and SL2-spliced genes have a higher expression level than those that are not trans-spliced (left). The proportion of trans-spliced genes is positively associated with expression level (right). SL1, spliced leader 1; TPM, transcripts per kilobase million.
investigated global intron composition to potentially understand the IME of genes across nematodes. Unexpectedly, introns across eight annotated genomes showed distinct features in terms of the general pattern observed (for all comparisons $P = 0.0$, Kolmogorov–Smirnov test) and the median intron size. When intron length was plotted by frequency, a unimodal pattern was detected in *C. elegans*, *H. contortus*, *P. pacificus*, and *S. ratti*, whereas a bimodal pattern was observed in *B. xylophilus* (Figure 2A). Further, in the clade three nematodes (*B. malayi* and *A. suum*) intron length appeared to be much longer and with a wider distribution.

Further analysis of the average intron length revealed that the introns of *C. elegans* have a mode (the most abundant number) of 47 nt and a median of 63 nt in length (Figure 2A), while the distantly related *S. ratti* has even shorter introns with a median of 52 nt and a more homogeneous distribution. In *H. contortus*, *P. pacificus*, *B. xylophilus*, and *G. pallida*, the distribution of intron length shows a greater range compared with *C. elegans*, although an accumulation of introns with a size between 40 and 60 nt is also detectable. When comparing intron size with gene expression level, we found that in *C. elegans* ($r = -0.55$, $P = 0.038$, Pearson’s correlation), *H. contortus* ($r = -0.68$, $P = 0.010$, Pearson’s correlation), *S. ratti* ($r = -0.70$, $P = 0.008$, Pearson’s correlation), *G. pallida* ($r = -0.91$, $P < 0.001$, Pearson’s correlation), *B. malayi* ($r = -0.91$, $P = 0.001$, Pearson’s correlation), and *A. suum* ($r = -0.95$, $P < 0.001$, Pearson’s correlation), the intron size was negatively correlated with gene expression level. However, this correlation was not observed in the other species (Figure 2B).

Finally, we investigated another gene regulatory element, SL1. Nematodes have a specific trans-splicing mechanism at the 5’ end of many premature messenger RNAs (mRNAs), which is trimmed and replaced by an SL sequence (Denker et al. 2002). This mechanism is thought to increase translation (Yang et al. 2017). Using a *P. pacificus* RNA-depleted RNA-seq library instead of deeply sequenced mRNA-enriched RNA-seq data sets, which are traditionally used in *C. elegans* (Allen et al. 2011), we identified a total of 5982 genes in *P. pacificus* that were SL1-operated, and 922 genes that were SL2-operated. These genes have an SL1 3’ splice site with a consensus sequence “TTTCAG” (Figure 2C), which is also conserved in *C. elegans* (Yang et al. 2017). Globally, higher expression levels were observed in *P. pacificus* genes associated with SL1 compared with genes without splicing leaders ($P = 0.0$, Wilcoxon signed-rank test). Therefore, this suggests that SL1 increases translation in *P. pacificus*, a similar phenomenon to that observed in *C. elegans* (Yang et al. 2017). While the published nematode data sets are not sufficient for us to survey the SL1 trans-spliced genes of other nematode species, given the fact that the sequences of SLs are conserved among nematodes (Guiliano and Blaxter 2006), the SL trans-splicing could be a highly conserved mechanism in the Nematoda phylum.

**Optimization of GFP and TurboRFP sequences and increased transgenesis efficiency**

With our observations of the large variations in CUB and potential IME regulating gene expression across nematodes, we decided to focus on a single species and attempt to improve its transgenesis efficiency. Therefore, we focused on establishing two fluorescent proteins for use in the free-living nematode *P. pacificus*. These were based on the previously utilized TurbORFP (Schlager et al. 2009) and on GFP (Fire Lab Vector Kit 1995), which are commonly utilized across the *C. elegans* community. In *P. pacificus*, TurbORFP has been used to successfully produces transgenic lines; however, this was only at a low transmission efficiency. The GFP previously used in *P. pacificus* was optimized according to *C. elegans’* CUB and hardly generated detectable fluorescence. Therefore, we replaced the codons in these two fluorescent proteins with two sets of codon usages: the CUB found associated with the top 10% most highly expressed genes and with the top 3% most highly expressed genes of *P. pacificus* (Table S2). Alongside this, we also attempted to optimize both fluorescent proteins further through the addition of native introns to increase its transcription. We selected the native introns of the gene *Ppa-rps-1* as it is highly expressed through all life stages and has four relatively short introns. The three shorter introns of *Ppa-rps-1* were added into the reading frame of the codon-optimized GFP and TurbORFP. Finally, we added an SL1 3’ splice site sequence immediately upstream of the start codon of both fluorescent proteins (illustrated in Figure 3).
In a first set of experiments, we performed all three optimization steps (CUB, native intron addition, and SL1 3′ splice site sequence) simultaneously and used the previously established egl-20 promoter to drive fluorescent protein expression. We were able to obtain GFP transcriptional reporter lines with robust and intense signals (Figure 3). More importantly, we considerably improved the efficiency of transgenesis of both GFP (PZH008) and TurboRFP (PZH009) constructs (P = 0.02 and P = 0.003, respectively; Table 1).

Note that we still experienced variability in the efficiency; possibly due to factors such as injector and age of the specimen, the efficiency increased to 20% of injected animals. While we did not systematically test all variables individually due to the enormous costs that would have been associated with such studies, we confirmed the increase in efficiency by the subsequent removal of introns. Indeed, intron removal coincided with a decrease in transgenic efficiency (Table 1). Together, we found that the codon-optimized three-intron GFP and TurboRFP had greater efficiency compared with the previous nonoptimized TurboRFP (Schlager et al. 2009). However, for unknown reasons, utilizing the CUB of the top 3% highly expressed genes did not further increase the efficiency (Table S3).

**Fluorescent co-injection marker-assisted CRISPR genome editing**

With the establishment of reliable and robust transgenic markers in *P. pacificus*, we next attempted to implement these tools to reduce the workload and the cost of screening potential CRISPR/Cas9 alleles. Therefore, we tried to establish a method that employed the optimized fluorescent markers to identify potential mutants induced with CRISPR/Cas9 (Figure 4A). A fluorescent marker can indicate well-injected specimens, which carry an increased likelihood of successfully induced CRISPR/Cas9 mutations. Therefore, using the egl-20p::TurboRFP (PZH009) as a co-injection marker, our experienced injectors obtained between 1 and 5 P0s (on average 2.5) producing RFP-positive F1 progeny from 30 well-injected nematodes (Figure 4B). Furthermore, progeny cooccurring on RFP injection marker-positive plates also frequently carried CRISPR/Cas9-induced mutations at high efficiency (77% of the identified plates).

### Table 1 Improved transgenesis efficiency using transcriptional reporter constructs with codon optimization and intron addition in *P. pacificus*

| Construct          | Number of introns | Injected P0s | Number of P0s with fluorescent F1s | Efficiency (%) |
|--------------------|-------------------|--------------|-----------------------------------|---------------|
| egl-20p::GFP       | 3                 | 49           | 11                                | 22 (P = 0.02) |
| egl-20p::TurboRFP  | 3                 | 55           | 16                                | 29 (P = 0.003)|
| egl-20p::GFP       | 2                 | 40           | 4                                 | 10 (P = 0.39) |
| egl-20p::GFP       | 1                 | 12           | 0                                 | 0 (P = 0.48)  |
| egl-20p::TurboRFP  | 1                 | 18           | 0                                 | 0 (P = 0.34)  |
| Ppa-prl-1*         | NA                | NA           | NA                                | 5             |

The GFP and TurboRFP sequences were optimized using *P. pacificus* favored codons (from top 10% highly expressed) with addition of native introns.

* Data from Schlager et al. 2009, summarized from over 3000 P0 injections. Chi-square tests were performed between Ppa-prl-1 and optimized constructs.

*Figure 4* Newly established fluorescent co-injection marker-assisted CRISPR genome editing in *P. pacificus*. (A) An illustration of the workflow for CRISPR genome editing in *P. pacificus*. (B) Using an egl-20p::TurboRFP construct (PZH009) as a co-injection marker, an average of 2.5 P0s had RFP + F1s from 30 well-injected *P. pacificus* P0s (left). (B) Next, 8–16 F1s were selected from each P0 with RFP + F1s to detect CRISPR alleles. There was a 77% chance that the P0-contained RFP + F1s also contained CRISPR-edited F1s (right). These data were accumulated from 18 independent experiments. Note, there were two additional experiments with no RFP + F1 detected, but CRISPR editing still occurred. CRISPR, clustered regularly interspaced short palindromic repeats; RFP, red fluorescent protein.
allowing the number of progeny necessary to be screened to isolate a CRISPR/Cas9 mutant to be greatly reduced (Figure 4B). Thus, the improved fluorescent-based co-injection marker strongly assisted the detection of CRISPR-generated edits in *P. pacificus*. We would like to note here that this fluorescent marker-assisted CRISPR method is compatible with knockouts and shorter repair templates (<120 nt), but does not seem to work with longer repair templates.

Discussion

The usage of transgenic tools is fundamental to successful studies in molecular biology; however, their efficiency is not uniform between organisms. This is, in part, likely due to differences in gene regulatory mechanisms between different species. In canonical model organisms, the development of efficient transgenic tools is aided by the existence of large scientific communities capable of refining and optimizing their application; however, this is not usually possible in other systems. Although the delivery of DNA to the germline can be an obstacle in nematode species (Kranse et al. 2020), delivery via microinjection is not a hindrance in *Pristionchus*, since we have generally achieved a higher efficiency for CRISPR knockouts compared with transgenesis. Here, we have revealed large differences in CUB and IME across nematodes, which likely contribute to gene regulatory differences between species. As a proof of principle, we investigated a single nematode species, *P. pacificus*, whereby we have successfully exploited its favored CUB and IME to develop *P. pacificus*-adapted fluorescent transgenic proteins. Additionally, we have shown that these adapted proteins containing *P. pacificus* gene regulatory requirements demonstrate a dramatically increased expression efficiency. It has recently been shown in *C. elegans* that the 5′ intron contributes the most to the elevated level of gene expression (Crane et al. 2019). Our results in *P. pacificus* agree with this finding because transgenesis efficiency decreased when the 5′ intron was removed. Transgenes with constructs that were modified using the top 3% CUB did not further improve efficiency. We can only speculate that this might be due to the most highly favored codons causing ribosomal traffic jams (Plotkin and Kudla 2011). Nevertheless, these improvements allow transgenes to be utilized as co-injection markers to reduce the screening time and costs of CRISPR/Cas9 genome editing. Thus, our method provides an alternative to the existing *Pristionchus* co-CRISPR method, in which the identification of CRISPR candidates relies on a Dpy phenotype (Nakayama et al. 2020).

By means of an initial bioinformatic analysis of the species-specific CUB and IME, our experiments demonstrate the potential to develop optimized transgenic tools and explore distinctive attributes that were not previously possible. While we did not systematically test the specific contributions of CUB, IME, and SL1, they likely all play important roles in transcription and translation for the increased transgenesis efficiency in *P. pacificus* (Redemann et al. 2011; Yang et al. 2017; Crane et al. 2019), but it is important to note that this principle could be further utilized to optimize genetically encoded calcium indicators and optogenetic tools to explore *Pristionchus*-specific behaviors, and genetic ablation methods to investigate aspects of anatomy and physiology. We hypothesize that, using knowledge of species-specific genomic features, it is possible to establish transgenic tool kits in other free-living nematodes, and additionally in parasitic nematode systems that have a significant impact on world health (Brindley et al. 2009) and crop production (Nicol et al. 2011).

Acknowledgments

The authors would like to thank the members of the Sommer laboratory for collecting CRISPR editing data. Adrian Streit provided insightful thoughts about this manuscript. This work was supported by the Max Planck Society (funding awarded to R.J.S.); an Alexander von Humboldt Foundation Postdoctoral fellowship (awarded to Z.H.), and a Chinese Scholarship Council Ph.D. fellowship (awarded to S.S.). Construct maps in Snapgene files are available upon request.

Literature Cited

Adams, S., P. Pathak, H. Shao, J. B. Lok, and A. Pires-daSilva, 2019 Liposome-based transfection enhances RNAi and CRISPR-mediated mutagenesis in non-model nematode systems. Sci. Rep. 9: 483. https://doi.org/10.1038/s41598-018-37036-1

Akduman, N., J. W. Lightfoot, W. Rösseler, H. Witte, W.-S. Lo et al., 2020 Bacterial vitamin B12 production enhances nematode predatory behavior. ISME J. 14: 1494–1507 (erratum: ISME J. 14: 1911). https://doi.org/10.1038/s41396-020-0626-2

Allen, M. A., L. W. Hillier, R. H. Waterston, and T. Blumenthal, 2011 A global analysis of C. elegans trans-splicing. Genome Res. 21: 255–264. https://doi.org/10.1101/gr.113811.110

Bailey, T. L., M. Boden, F. A. Buske, M. Frith, C. E. Grant et al., 2009 MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37: W202–W208. https://doi.org/10.1093/nar/gkp335

Bhattacharya, D., and E. G. Van Meir, 2019 A simple genotyping method to detect small CRISPR-Cas9 induced indels by agarose gel electrophoresis. Sci. Rep. 9: 4437. https://doi.org/10.1038/s41598-019-39950-4

Bischof, J., R. K. Maeda, M. Hediger, F. Karch, and K. Basler, 2007 An optimized transgenesis system for Drosophila using germ-line-specific C31 integrases. Proc. Natl. Acad. Sci. USA 104: 3312–3317. https://doi.org/10.1073/pnas.0611511104

Blaxter, M. L., P. De Ley, J. R. Garey, L. X. Liu, P. Scheldeman et al., 1998 A molecular evolutionary framework for the phylum Nematoda. Nature 392: 71–75. https://doi.org/10.1038/32160

Brindley, P. J., M. Mitreva, E. Ghedin, and S. Lustigman, 2009 Helminth genomics: the implications for human health. PLoS Negl. Trop. Dis. 3: e538. https://doi.org/10.1371/journal.pntd.0000538

Brinster, R. L., J. M. Allen, R. R. Behringer, R. E. Gelinas, and R. D. Palmiter, 1988 Introns increase transcriptional efficiency in transgenic mice. Proc. Natl. Acad. Sci. USA 85: 836–840. https://doi.org/10.1073/pnas.85.3.836
Bui, L. T., N. A. Ivers, and E. J. Ragsdale, 2018 A sulfotransferase dosage-dependently regulates mouthpart polyphenolism in the nematode Pristionchus pacificus. Nat. Commun. 9: 4119. https://doi.org/10.1038/s41467-018-05612-8

Castillo-Davis, C. I., S. L. Mekhedov, D. L. Hartl, E. V. Koonin, and F. A. Kondrashov, 2002 Selection for short introns in highly expressed genes. Nat. Genet. 31: 415–418. https://doi.org/10.1038/ng940

Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher, 1994 Green fluorescent protein as a marker for gene expression. Science 263: 802–805. https://doi.org/10.1126/science.8303295

Choi, Y.-J., E. Ghedin, M. Berriman, N. Holroyd et al., 2011 A deep sequencing approach to comparatively analyze the transcriptome of lifestyle stages of the filarial worm, Brugia malayi. PLoS Negl. Trop. Dis. 5: e1409. https://doi.org/10.1371/journal.pntd.0001409

Clough, S. J., and A. F. Bent, 1999 Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 16: 735–743. https://doi.org/10.1046/j.1365-313x.199800343.x

Cotton, J. A., C. J. Lilley, L. M. Jones, T. Kikuchi, A. J. Reid et al., 2014 The genome and life-stage specific transcriptomes of Globodera pallida elucidate key aspects of plant parasitism by a cyst nematode. Genome Biol. 15: R43. https://doi.org/10.1186/gb-2014-15-3-r43

Crane, M. M., B. Sands, C. Battaglia, B. Johnson, S. Yun et al., 2019 In vivo measurements reveal a single 5’-intron is sufficient to increase protein expression level in Caenorhabditis elegans. Sci. Rep. 9: 9192. https://doi.org/10.1038/s41598-019-45517-0

Denker, J. A., D. M. Zuckerman, P. A. Maroney, and T. W. Nilsen, 2002 New components of the spliced leader RNA required for nematode trans-splicing. Nature 417: 667–670. https://doi.org/10.1038/nature00783

Dokshin, G. A., K. S. Ghanta, K. M. Piscopo, and C. C. Mello, 2018 Robust genome editing with short single-stranded and particle bombardment. Exp. Parasiol. 100: 95–102. https://doi.org/10.1016/S0014-4894(02)00004-8

Hong, R. L., and R. J. Sommer, 2006 Pristionchus pacificus: a well-rouneded nematode. Bioessays 28: 651–659. https://doi.org/10.1002/bies.20404

Howe, K. L., B. J. Bolt, S. Cain, J. Chan, W. J. Chen et al., 2016 WormBase 2016: expanding to enable helmhmin geno- methodic research. Nucleic Acids Res. 44: D774–D780. https://doi.org/10.1093/nar/gkv1217

Howe, K. L., B. J. Bolt, M. Shafie, P. Kersey, and M. Berriman, 2017 WormBase ParaSite - a comprehensive resource for helmhmin genomes. Mol. Biochem. Parasitol. 215: 2–10. https://doi.org/10.1016/j.molbiopara.2016.11.005

Hunt, V. L., I. J. Tsai, A. Coghlan, A. J. Reid, N. Holroyd et al., 2016 The genomic basis of parasitism in the Strongyloides clade of nematodes. Nat. Genet. 48: 299–307. https://doi.org/10.1038/ng3495

Hutter, H., 2012 Fluorescent protein methods: strategies and applications. Methods Cell Biol. 107: 67–92. https://doi.org/10.1016/B978-0-12-394620-1.00003-5

Kieninger, M. R., N. A. Ivers, C. Rödelsperger, G. V. Markov, R. J. Sommer et al., 2016 The nuclear hormone receptor NHR-40 acts downstream of the sulfatase EUD-1 as part of a developmental plasticity switch in Pristionchus. Curr. Biol. 26: 2174–2179. https://doi.org/10.1016/j.cub.2016.06.018

Kikuchi, T., J. A. Cotton, J. J. Dalzell, K. Hasegawa, N. Kanaki et al., 2011 Genomic insights into the origin of parasitism in the emerging plant pathogen Bursaphelenchus xylophilus. PLoS Pathog. 7: e1002219. https://doi.org/10.1371/journal.ppat.1002219

Kim, D., B. Langmead, and S. L. Salzberg, 2015 HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12: 357–360. https://doi.org/10.1038/nmeth.3317

Krane O., H. Beasley, S. Adams, A. P. da Silva, C. Bell, et al., 2020 Towards genetic modification of plant-parasitic nematodes: delivery of macromolecules to adults and expression of exogenous mRNA in second stage juveniles. bioRxiv. doi.org/10.1101/2020.07.15.193052 (Preprint posted July 15, 2020).

Laing, R., T. Kikuchi, A. Martinelli, I. J. Tsai, R. N. Beech et al., 2013 The genome and transcriptome of Haemonchus contor-

Laing, R., T. Kikuchi, A. Martinelli, I. J. Tsai, R. N. Beech et al., 2013 The genome and transcriptome of Haemonchus contor-

Lee, R. Y., K. L. Howe, T. W. Harris, V. Arnaboldi, S. Cain et al., 2018 WormBase 2017: molting into a new stage. Nucleic Acids Res. 46: D869–D874. https://doi.org/10.1093/nar/gkx998

Liao, Y., G. K. Smyth, and W. Shi, 2014 featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30: 923–930. https://doi.org/10.1093/bioinformatics/btu656

Lightfoot, J. W., M. Wilecki, C. Rödelsperger, E. Moreno, V. Susoy et al., 2019 Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes. Science 364: 86–89. https://doi.org/10.1126/science.aav9856

Li, X., H. C. Massey, Jr., T. J. Nolan, G. A. Schad, K. Kraus et al., 2006 Successful transgenesis of the parasitic nematode Strongyloides stercoralis requires endogenous non-coding control elements. Int. J. Parasitol. 36: 671–679. https://doi.org/10.1016/j.ijpara.2005.12.007

Liu, Y., K. G. Kaval, A. van Hoof, and D. A. Garsin, 2019 Heme peroxidase HPX-2 protects Caenorhabditis elegans from pathogens. PLoS Genet. 15: e1007944. https://doi.org/10.1371/journal.pgen.1007944

Loer, C., H. Witte, R. Sommer, and O. Hobert, 2019 An antibody staining protocol variation for nematodes that adds heat-induced antigen retrieval (HIER). MicroPubl Biol 2019. Available at:
Mitreva, M., M. C. Wendell, J. Martin, T. Wylie, Y. Yin et al., 2006 Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. Genome Biol. 7: R75. https://doi.org/10.1186/gb-2006-7-8-r75

Nicol, J. B., 2012 Nucleotide acid transfection and transgenesis in parasitic nematodes. Parasitology 139: 574–588. https://doi.org/10.1017/S0031182011001387

Moreno, E., A. McGaughran, C. Rödelsperger, M. Zimmer, and R. J. Sommer, 2015 Molecular cloning of a dominant roller mutant and establishment of DNA-mediated transformation in the nematode Pristionchus pacificus. Genetics 135: 385–375. https://doi.org/10.1101/gr.186783.114

Redemann, S., S. Schloissnig, S. Ernst, A. Pozniakowsky, S. Ayloo et al., 2011 Codon adaptation-based control of protein expression in C. elegans. Nat. Methods 8: 250–252. https://doi.org/10.1038/nmeth.1565

Rubin, G. M., and A. C. Spradling, 1982 Genetic transformation of Drosophila with transposable element vectors. Science 218: 348–353. https://doi.org/10.1126/science.6289436

Wang, J., B. Czech, A. Crunk, A. Wallace, M. Mitreva et al., 2011 Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. Genome Res. 21: 1462–1477. https://doi.org/10.1101/gr.121426.111

Wang, J., S. Gao, Y. Mostovoy, Y. Kang, M. Zagoskin et al., 2017 Comparative genome analysis of programmed DNA elimination in nematodes. Genome Res. 27: 2001–2014. https://doi.org/10.1101/gr.225730.117

Communicating editor: O. Hobert

956 Z. Han et al.