Transposable elements (TEs) are mobile genetic elements, present in all domains of life. They commonly encode a single transposase enzyme, that performs the excision and reintegration reactions, and these enzymes have been used in mutagenesis and creation of next-generation sequencing libraries. All transposases have some bias in the DNA sequence they bind to when reintegrating the TE DNA. We sought to identify a transposase that showed minimal sequence bias and could be produced recombinantly, using information from the literature and a novel bioinformatic analysis, resulting in the selection of the hATx-6 transposase from *Hydra vulgaris* (aka *Hydra magnipapillata*) for further study. This transposase was tested and shown to be active both in vitro and in vivo, and we were able to demonstrate very low sequence bias in its integration preference. This transposase could be an excellent candidate for use in biotechnology, such as the creation of next-generation sequencing libraries.

Keywords: DNA cut and paste transposons, transposable element, target site duplication, hAT transposases, HERMES, assessing target site specificity

INTRODUCTION

Transposable elements (TEs) are defined as "DNA sequences that are able to move from one location to another in the genome (Munoz-Lopez and Garcia-Perez, 2010) and are commonly found in genomes of almost every known organism" (Aziz et al., 2010). TE's can be divided into two groups, retrotransposons and DNA transposons. Retrotransposons mobilize with an RNA intermediate step and generally have more complex mechanisms, while DNA TE's do not require RNA intermediates, and are often simpler systems. In many cases DNA transposition is catalyzed by a single protein, which performs the entire excision and reintegration steps; these enzymes are known as a transposases.

This capacity to perform all the reaction steps makes the transposases useful tools for genomic editing applications. One of the earliest applications was random mutagenesis, where TE's would be inserted into the genome at random sites to discover phenotypes associated with the insertion mutations (Goryshin et al., 2000). TE's can also be used to insert genes in a targeted manner, into the genomes of various organisms. The most common examples, particularly in vertebrates, are piggyBac and Sleeping Beauty (Gao and Liu, 2017; Qing et al., 2017). Recently, TE's in the Tn7 family that contain Cas subtype I-F and subtype V-K targeting modules, as well as through fusions of Cas-9 to the Hsmar1 transposase, have been used to mediate targeted gene insertions in *E. coli* (Bhatt and Chalmers, 2019; Klompe et al., 2019; Strecker et al., 2019).

The most common application for transposases currently is in the construction of libraries for next-generation sequencing (NGS) (Adey et al., 2010). To generate a library, the genomic DNA of the sample needs to be fragmented into small oligonucleotides followed by ligation of specific adapters to the cleaved ends in order to amplify and sequence the library. Classically, the cleavage of the genomic
DNA was achieved by physical fragmentation using sonication or enzymatically by endonuclease digestion, while the adapter ligation required a separate enzymatic reaction. Alternatively, one can use a transposase to perform the cleavage of the genomic DNA and the adapter ligation can be conducted in a single reaction (Figure 1). To construct an NGS library, first oligonucleotides (containing TE specific end sequences and an adapter) are mixed with the transposase (both Tn5 and MuA based systems are used commercially, as Nextera™ and MuSeek™, respectively), resulting in a transposase-DNA complex. These complexes are mixed with the purified genomic DNA, where the transposase integrates the transposon ends (with adapters attached) into the genomic DNA. As the two oligonucleotides are not connected to each other, a break is created in the genomic DNA.

The most widely used transposase for multiple applications (random mutagenesis, ATAC-seq, NGS library prep (Buenrostro et al., 2015; Gao and Liu, 2017)) is Tn5, first isolated from the genome of Escherichia coli (Berg et al., 1975). Tn5 transposase is currently the most studied system and was the first transposase to have its crystal structure solved (Davies et al., 2000). The version of the transposase used in applications has been modified from the wild-type form with several mutations (Wiegand and Reznikoff, 1992; Weinreich et al., 1994; Zhou and Reznikoff, 1997; Goryshin and Reznikoff, 1998). The E54K mutation improves the binding of the transposase to the TE ends. M56A does not directly affect activity but does prevent the expression of an N-terminally truncated form of the transposase (the inhibitor protein) which can bind and inhibit the functional transposase. The final mutation L372P increases activity by reducing the flexibility of the C-terminal portion of the transposase, which in the wild-type form interferes with the N-terminal portion as it binds to the TE end sequence. This hyperactive Tn5 shows a GC bias in insertion site preference, though a more recent study has used a series of four further mutations to partially reduce the bias (Green et al., 2012; Kia et al., 2017). The bias of Tn5 does not inhibit its use with genomes of GC content 30% or higher but is a significant limitation below that threshold (Kia et al., 2017).

In this study we sought to identify and characterize of a new transposase in a novel approach utilizing bioinformatic tools. We took the data set from Arensburger 2011 as a starting point, which collated 299 TE’s of the hAT (hobo/Ac/Tam3) superfamily (Arensburger et al., 2011), which was first discovered in plants (Kempken and Windhofer, 2001). hAT transposases are attractive targets as some members of this superfamily have been demonstrated to have low target site specificity (Vollbrecht et al., 2010; Lazarow et al., 2012; Michel et al., 2017). Our goal was to identify an unknown transposase with low target site preference from this dataset, that could be purified, characterized, and used in NGS. Analyzing the predicted insertion sites bioinformatically, several candidate transposases were identified. After selecting the 6th putative transposase (hATx-6), which had the lowest predicted site

**FIGURE 1 | Schematic representation of Tagmentation**

1. **Elements of a Transposon**
   - Representation of the elements of a transposon, with the transposase gene in red and the sequences for the Left Inverted Repeat (LIR) in blue and right inverted repeat (RIR) in green. The chromosome is represented as black lines.
   - Insertion of a transposon element resulting in Target Site Duplication (TSD).

2. **Transposon/DNA complex**
   - Schematic representation of Tagmentation. A transposase (red pentagon) is in complex with DNA containing LIR and RIR, each containing either the sequence for Illumina P5 (brown) or P7 (magenta) sequences. The resulting transposase reaction results in TSD which is filled with initial extension, amplified for 5 rounds with PCR and the products are sequenced bidirectionally with Illumina primers.
specificity, we purified the candidate transposase by over expressing it in *E. coli*. We characterized the transposase through *in vivo* papillation assays in *E. coli* and *in vitro* transposition assays. We then analyzed the insertion site preference using genomic DNA with various GC contents, to determine the insertion site preference, and its viability for NGS library preparation.

### MATERIALS AND METHODS

#### Materials

We grew cells in Rich Media (10 g/L Tryptone, 5 g/L Yeast Extract, 5 g/L NaCl, NaOH to pH 7.2) or in LB (10 g/L Tryptone, 5 g/L Yeast Extract, 10 g/L NaCl, 10 mM MgCl\(_2\) NaOH to pH 7.2). Restriction enzymes, HiFi master mix, and NEBNext Ultra II FS kit were all supplied by New England Biolabs (NEB). *E. coli* and *Rhodopseudomonas palustris* genomic DNA was prepared by Lofstrand Labs Limited and human NA19240 was purchased from the Coriell Institute (Llamas et al., 2019). *Haemophilus influenzae* genomic DNA was the kind gift of Keith Lunnen (NEB).

#### Bioinformatics

Biopython scripts used are shown in Supplementary Table S1. Previously published data on the target site duplication indicated that a number of TE’s in the genomes of cnidarians had low target site specificity (Arensburger et al., 2011). We constructed scripts that searched for hAT TE’s in the genome of *Hydra vulgaris* (BioProject PRJNA12876), using a TE end as the query for a BLAST search. We then examined the target site duplication (TSD) to see if at least seven of the eight bases were identical at each end of a putative insertion in the genome. If true, a weblogo of the sequence was generated. We selected TE’s that had weblogos indicating low site preference in the TSD for further study (Figure 2).

To identify the transposase gene, we used a modified script that extracted the sequences of the TE’s. We examined the sequences for open reading frames that 1) were sufficient size to encode a hAT transposase (~500 amino acids), 2) had homology to a hAT transposase by BLAST, and/or 3) contained the “hAT dimerization domain” (pfam 05699) when submitted to HMMR (https://www.ebi.ac.uk/Tools/hmmer). We lined up the candidate sequences in SeaView (Gouy et al., 2010) and derived a consensus sequence for the transposase (Supplementary Figure S1).

We analyzed sequencing results by first identifying and caching reads that had the 20 base-pair left end sequence of the hATx-6 TE (see DNA sequencing below), and removing the 20 bases, with the program cutadapt (Martin, 2011). We then used a Python script to create a position weight matrix of the first 28 bases at the end of each sequence, the first 8 bases being the TSD and the following 20 bases comprising the flanking region.

#### Bacteria and Plasmids

We used NEB 10-beta (NEB C3019) as a host for cloning and expression, and T7 Express for the papillation assay (NEB C2566).

We composed a DNA sequence that encoded the hATx-6 transposase suitable for expression in *E. coli* using the Codon Optimization tool from IDT (https://www.idtdna.com/CodonOpt). We then divided the gene sequence into three sections with 20 base overlaps and had the sections synthesized (IDT gBlocks) (Supplementary Figure S2). The gene was assembled with the vector using the NEBuilder HiFi
assembly Master Mix (NEB E2621). We then cloned the transposase into pMAL-c5X cut with XmnI and SbfI, assembling with a PCR fragment made using primers pMAL-c5X-hatx-6-F and pMAL-c5X-hatx-6-R. We named the plasmid pMAL-c5X-hATx-6 (Supplementary Figure S2).

Oligonucleotides are listed in Supplementary Table S2. We constructed the TE donor for the papillation assay by first assembling a PCR fragment made using pSYX20 (Morgan et al., 1996) as a template and primers pSYX20 F1330 and pSYX20 R4827 with a PCR fragment of lacZ made with primers clacZ_fwd2 and clacZ_rev2, using the NEBuilder HiFi assembly master mix. We then added 30 base-pairs of the right and left ends of the hATx-6 TE to the resulting plasmid by Q5 site-directed mutagenesis in two steps, first with the primers hATx-6-RE30-pSYXlac fwd and hATx-6-RE30-pSYXlac rev for the right end, then the primers hATx-6-LE30-pSYXlac fwd and hATx-6-LE30-pSYXlac rev for the left end, flanking the lacZ gene (Supplementary Table S2).

hATx-6 Transposase Expression

We first did small-scale expression trials by transforming NEB 10-beta with the expression plasmid and inoculating 10 ml LB containing 100 μg/ml ampicillin and incubating at 37°C overnight. We then inoculated 20 ml of the same medium with 0.2 ml of the overnight culture and grew to an OD600 of 0.5. We added IPTG to 0.3 mM, separated the culture into 4 5 ml aliquots and incubated them at 37°C for 2 h, 30°C for 4 h, 25°C for 8 h and 16°C for 16 h. In a separate experiment, we grew an overnight as described above and inoculated 5 ml with 0.05 ml of the overnight. We incubated the culture at 37°C until it reached an OD600 of 0.5, then shifted it to 18°C for 16 h. In both cases, we harvested the cells by centrifugation then lysed by detergent lysis as described (Walker et al., 2010). We analyzed samples by SDS-PAGE and looked for a band of the appropriate size.

hATx-6 Transposase Purification

Overnights grown at 37°C on a roller drum were used to inoculate 60 ml rich media in a 250 ml flask. This inoculum was used to inoculate 6 L of Rich + ampicillin of NEB 10-beta 

Transposase Activity Assay

We mixed 1–5 μl of MBP-hATx-6 at 9.2 μg/μl, 1.0 μM of TE end(s) and 50 ng of supercoiled pBR322 DNA in a 20 μl reaction containing 25 mM MOPS pH 7.0, 0.2 mM TCEP, 5 mM MgCl₂, 50 μg/ml BSA, 100 mM NaCl, 0.1 mM EDTA and incubated at 30°C for 30 min. We stopped the reaction with 5 μl of 6x Purple Gel Loading Dye (NEB) and loaded 15 μl on a 1% agarose gel containing 0.5 μg/ml ethidium bromide. We measured activity by conversion of the supercoiled plasmid to linear and nicked forms. For assays where the TE ends were labelled with biotin, we followed transfer of the end to the plasmid DNA by electrophoresing the DNA in the gel to a positively charged nylon membrane (Sigma # 11417240001) using a Bio-Rad Trans-Blot apparatus at 200 mA for 30 min. We washed and labelled the membrane using the Chemiluminescent Nucleic Acid Detection kit (ThermoFisher #89880) and visualized on an Amersham Typhoon Gel Imaging scanner (GE Life Sciences) with 488 nm laser and the Cy2 filter.

Papillation Assay

We made competent cells from T7 Express containing the pSYX-RE30-lacZ-LE30 TE donor plasmid, then transformed with 1 ng of pMAL-c5X-hATx-6 transposase plasmid and plated on LB plates containing 0.1% lactose, 20 μg/ml X-Gal, 100 μg/ml ampicillin, diluted appropriately to obtain ~100 colonies per plate. We incubated the plates at 37°C for 5 days, then stored them in the dark at 4°C overnight to enhance the X-Gal signal.

DNA Sequencing

We first fragmented and tagged genomic DNA from E. coli, Haemophilus influenzae, Rhodopseudomonas palustris and human leukocytes in a tagmentation reaction containing 37 μg MBP-hATx-6, 50 ng genomic DNA, 0.1 μM LE20 in a 30 μl reaction containing 25 mM MOPS pH 7.0, 0.2 mM TCEP, 5 mM MgCl₂, 50 μg/ml BSA, 100 mM NaCl, 0.1 mM EDTA. We purified the reaction using a Monarch DNA Clean-up kit and eluted the DNA in 30 μl of 10 mM Tris-Cl 0.1 mM EDTA pH 7.5. We then took 26 μl of this sample and took it through the NEBNext Ultra II FS workflow. The libraries were quantified using the NEBNext Library Quant kit for Illumina and sequenced on a MiSeq (Illumina).

RESULTS

Screen for Transposases

We undertook a search for a hAT transposase with low specificity in target site selection, using a bioinformatic approach to assess the bias at the target site duplication created when the TE inserts at a
new site (see Materials and Methods, Bioinformatics, Figure 3). This search led us to the hATx-6 TE in Hydra vulgaris (aka Hydra magnipapillata). A weblogo was constructed from the 54 insertions in the Hydra genome that had a recognizable target site duplication (TSD). The transposase ORF is shown as a box with transposase written in it. Data set from Arensburger with 299 curated TE’s of the hAT (hobo/Ac/Tam3) superfamily (Arensburger et al., 2011) was used to search for unique transposases. TE sequence is searched throughout the genomic DNA, to identify matching sequences. Identification of TSD sequences that would be directly adjacent to the TE sequence. Identification of the other TE and TSD sequences downstream of the original. Create a weblogo of TSD and utilize the info to select for the appropriate TE. Identification of the transposase ORF within the transposon sequence. A consensus sequence from homologs of the selected TE is created, and the engineered synthetic TE is subcloned into an expression plasmid, transformed into E. coli and purified. The purified TE is assayed for activity in vitro.

any of them could yield an active transposase (Supplementary Figure S1).

Purification of hATx-6 Transposase
We created a consensus sequence (Supplementary Figure S1) from the 10 hATx-6 ORFs and assembled a gene encoding the putative consensus sequence transposase from synthesized oligonucleotides. We cloned the gene into pMAL-c5X and discovered that expression at 18°C overnight without induction resulted in the best yield of full-length MBP-transposase fusion protein. The protein prepared under these conditions was typically a mixture of full length fusion protein and one or two smaller truncation products, presumably C-terminal truncations as one step in the purification was binding of the N-terminal MBP to an amylose column. We obtained approximately 10 mg of MBP-hATx-6 from 6 L of culture, approximately half of which was full-length.

In vivo and In vitro Activity of hATx-6
To test the activity of hATx-6 in vivo, we performed a papillation assay (Foster et al., 1981) with hATx-6 (Figure 5A). In this assay, a transposon (located on a donor plasmid) contains a lacZ gene lacking a promoter. When transposition occurs, there is a chance that the integration site will be downstream of an endogenous promoter in the host genome, leading to expression of lacZ gene. Expression of LacZ will cause blue colouration in those cells, due to the presence of X-gal in the agar medium. The number of individual papillae on each colony is indicative of the transposition frequency, it can be clearly seen that the hATx-6 transposase demonstrates multiple transposition events per colony. This was encouraging as it indicates that the hATx-6 transposase is expressed, folds into its native form and is active against the chromosome of E. coli.

We tested the purified hATx-6 transposase for activity in vitro by using it to cleave supercoiled plasmid in the presence of 20 or 30 base-pair left and right ends of the hATx-6 TE (Figures 5B–F). The activity was not significantly different if we added only the right end or only the left end (Figure 5D), and the shorter oligonucleotide gave reproducibly higher activity than the longer one (Figure 5E). In order to test whether the transposase was joining the TE ends to the plasmid DNA, we labelled the end with biotin and blotted the gel to a nylon membrane and probed with streptavidin-HRP to visualize the reaction products (Figure 5F). These results indicated that both single-end joining and double end joining reactions were occurring and could be detected in these simple in vitro reactions.

Sequence Bias of hATx-6 Transposase
To assess the bias of the hATx-6 transposase in an in vitro reaction, we used the transposase to fragment genomic DNA from E. coli (51% GC), Haemophilus influenzae (38% GC), Rhodopseudomonas palustris (65% GC) and human (41% GC) (Figure 4). The reaction stalled at an average fragment size of approximately 2 kb. We then took these fragments and used them to prepare an Illumina library using the Ultra II FS kit. Sequencing results were uploaded to the NCBI SRA database and can be accessed with PRJNA762372 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA762372). The
resulting samples were sequenced on a MiSeq using the paired-end protocol, and the reads were analyzed as follows. We first used the program Cutadapt to identify and isolate reads with the 20 base-pair TE end. About one third of the reads fell in this category. We then combined the two sets of reads produced from the paired-end protocol and created a position weight matrix from the first 28 bases of each read, the first 8 being the TSD and the remaining being the 20 flanking bases (Figure 4A). In the TSD, base 1 showed a moderate G bias, bases 3 and 8 have a moderate C bias, and bases 6 and 7 have a strong A bias. In the flanking region, base 2 has a moderate C bias and base 3 has a strong A bias. Comparing these to the same type of analysis of the insertions in the Hydra genome (Figure 4B) gave qualitatively the same result.

To assess whether there was a measurable difference between Tn5 and hATx-6 target site bias, we ran Blo.go sequence logos of a previously published Tn5 data set (Green et al., 2012) in Candida glabrata subtelomeric DNA, and the hATx-6 data set in Haemophilus influenzae DNA (Figure 6). These included the
target site duplication region and the flanking 10 bases, and background frequencies were adjusted for the genomic DNA used. The Tn5 insertion site shows GC bias across both the TSD region and the flanking 10 bases. hATx-6 shows no significant GC bias across both the TSD region or the flanking 10 bases. Some individual bases display strong (over 0.5 bits) bias, in particular sites 2 and 7 of the 8 base TSD. hATx-6 shows no significant GC bias across both the TSD region or the flanking 10 bases. Some individual bases display strong (over 0.5 bits) bias, in particular sites 2 and 7 of the 8 base TSD. Hermes has a strong bias for T at base 2 and A at base 7 of the 8 base TSD, while the other 6 bases all show strong bias for 2 bases (Gangadharan et al., 2010). All of these transposases

DISCUSSION

For most applications in molecular biology, the ideal transposase is one that has minimal bias at the target site. In the Tn5 system, the most commonly used for applications, there is a general GC bias in its target binding region and flanking DNA (Green et al., 2012). The ITm/mariner systems (such as ISY100 or Sleeping Beauty) there is an absolute bias for a TA dinucleotide TSD, but fairly limited sequence bias in the flanking regions (Feng and Colloms, 2007; Ivics and Izsvák, 2015). Many hAT superfamily members, such as Hermes, show strong bias at specific sites in the TSD region. Hermes has a strong bias for T at base 2 and A at base 7 of the 8 base TSD, while the other 6 bases all show strong bias for 2 bases (Gangadharan et al., 2010). All of these transposases
suffer from some target site biases, which inhibit the range of applications for which they would be applicable. hATx-6 does show some target site preference, more than expected based on the initial bioinformatics. However, when compared with Tn5 (Figure 6), the sequence bias overall does not indicate a significant bias towards GC or AT rich genomes. There is also much lower bias in the flanking DNA. Taken together, these results indicate that hATx-6 would be preferable for in vitro applications, and in particular for genomes with low GC content.

While what we have shown here is that the hATx-6 transposase is functional and has low target site preference, characteristics of the enzyme could be improved. Screening a mutant library using papillation assay could be used to further enhance its activity, as has been done in other systems (Zhou and Reznikoff, 1997; Mátés et al., 2009; Voigt et al., 2016). The X-ray crystal structure of the hAT transposases Hermes has shown it to exist as octamers, consisting of four dimeric transposase units (Hickman et al., 2014). It is highly likely that the hATx-6 transposase has a similar structure, though further studies to determine this would be of interest. Structural clues might lead to mechanistic details in explaining the low insertion sequence site specificities of hATx-6. Expression in E. coli has a tendency to produce some truncated forms of the protein, which might combine with full-length protein in the octamer and reduce or eliminate its activity, in a similar manner to the inhibitor protein of Tn5 (Braam et al., 1999). Modifications of the growth conditions, application of different host strains, or mutagenesis might

**FIGURE 6** Target site preferences of Tn5 and hATx-6 transposases. Bloocks sequence logos of (A) Tn5 into C. glabrata subtelomeric DNA and (B) hATx-6 into H. influenzae DNA. The Target Site Duplication (TSD) and the flanking 10 bases as ‘+x’ are indicated. Base frequencies are shown compared with the overall DNA content of (A) C. glabrata A: 0.32, C: 0.19, G: 0.18, T: 0.31 and (B) H. influenzae A: 0.31, C: 0.19, G: 0.19, T: 0.31.
improve the production of full-length, active protein. Mutagenesis could also be applied to optimize the solubility, as at present a portion of the protein produced is insoluble. Overcoming these difficulties would require additional work but given the favorable characteristics of the hATx-6 transposase, would be worthwhile.

Identification of new TEIs by the examination of target site preference is a methodology not used previously to screen new elements. This approach is a useful methodology for identifying new transposases aimed at particular applications. Instead of modifying a TE to fit an application to which it may not be suited, we might be able to identify the TE optimal for a specific purpose.

**DATA AVAILABILITY STATEMENT**

The original contributions presented in the study are publicly available in NCBI [https://www.ncbi.nlm.nih.gov/bioproject/762372 under accession number PRJNA762372.](https://www.ncbi.nlm.nih.gov/bioproject/762372)

**REFERENCES**

Adley, A., Morrison, H. G., Asan, A., Xun, X., Kitzman, J. O., Turner, E. H., et al. (2010). Rapid, Low-Input, Low-Bias Construction of Shotgun Fragment Libraries by High-Density In Vitro Transposition. *Genome Biol.* 11 (12), R119. doi:10.1186/gb-2010-11-12-r119

Arensburger, P., Hice, R. H., Zhou, L., Smith, R. C., Tom, A. C., Wright, J. A., et al. (2011). Phylogenetic and Functional Characterization of the hAT Transposon Superfamily. *Genetics* 188 (1), 45–57. doi:10.1534/genetics.111.126813

Aziz, R. K., Breitbart, M., and Edwards, R. A. (2010). Transposases Are The Most Abundant, Most Ubiquitous Genes in Nature. *Nucleic Acids Res.* 38 (13), 4207–4217. doi:10.1093/nar/gkq140

Berg, D. E., Davies, J., Allet, B., and Rochaix, J. D. (1975). Transposition of R Factor Genes to Bacteriophage Lambda. *Proc. Natl. Acad. Sci.* 72 (9), 3628–3632. doi:10.1073/pnas.72.9.3628

Bhatt, S., and Chalmers, R. (2019). Targeted DNA Transposition In Vitro Using a dCas9-Transposase Fusion Protein. *Nucleic Acids Res.* 47 (15), 8126–8135. doi:10.1093/nar/gkz552

Braam, L. A. M., Goryshin, I. Y., and Reznikoff, W. S. (1999). A Mechanism for Tn5 inhibition. *J. Biol. Chem.* 274 (1), 86–92. doi:10.1074/jbc.274.1.86

Buenrostro, J. D., Wu, B., Chang, H. Y., and Greenleaf, W. J. (2015). ATAC-seq: A Method for Assessing Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* 109, 21 29 1–21 29 9. doi:10.1002/0471142727.mb2129s109

Davies, D. R., Goryshin, I. Y., Reznikoff, W. S., and Rayment, I. (2000). Three-Dimensional Structure of the Tn 5 Symaptic Complex Transposition Intermediate. *Science* 289 (5476), 77–85. doi:10.1126/science.289.5476.77

Feng, X., and Colomys, S. D. (2007). In Vitro transposition of IS1001, a Bacterial Insertion Sequence Belonging to the Tc1/Mariner Family. *Mol. Microbiol.* 65 (6), 1432–1443. doi:10.1111/j.1365-2958.2007.05842.x

Foster, T. J., Lundblad, V., Hanley-Way, S., Halling, S. M., and Kleckner, N. (1981). Three Tn10-Associated Excision Events: Relationship to Transposition and Role of Direct and Inverted Reverses. *Cell* 23 (1), 215–227. doi:10.1016/0092-8674(81)90286-5

Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S. J., and Craig, N. L. (2010). DNA Transposon Hermes Inserts into DNA in Nucleosome-free Regions. *In Vitro* *Proc. Natl. Acad. Sci.* 107 (51), 21966–21972. doi:10.1073/pnas.1016382107

Gao, M., and Liu, D. (2017). CRISPR/Cas9-Based Pten Knock-Out and Sleeping Beauty Transposon-Mediated Nras Knock-In Induces Hepatocellular Carcinoma and Hepatic Lipid Accumulation in Mice. *Cancer Biol. Ther.* 18 (7), 505–512. doi:10.1080/15384047.2017.1323597

**AUTHOR CONTRIBUTIONS**

Goryshin, I. Y., Jendrisak, J., Hoffman, L. M., Meis, R., and Reznikoff, W. S. (2000). Insertional Transposon Mutagenesis by Electroporation of Released Tn5 Transposition Complexes. *Nat. Biotechnol.* 18 (1), 97–100. doi:10.1038/72017

Goryshin, I. Y., and Reznikoff, W. S. (1998). Tn5 In Vitro Transposition. *J. Biol. Chem.* 273 (13), 7367–7374. doi:10.1074/jbc.273.13.7367

Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol. Biol. Evol.* 27 (2), 221–224. doi:10.1093/molbev/msp259

Green, B., Bouchier, C., Fairhead, C., Craig, N. L., and Cormack, B. P. (2012). Insertion Site Preference of Mu, Tn5, and Tn7 Transposons. *Mobile DNA* 3 (1), 3–9. doi:10.1186/1759-8753-3-3

Hickman, A. B., Ewis, H. E., Li, X., Knapp, J. A., Laver, T., Doss, A.-L., et al. (2014). Structural Basis of hAT Transposon End Recognition by Hermes, an Octameric DNA Transposase from *Musca D. Cell.* 158 (2), 353–367. doi:10.1016/j.cell.2014.05.037

Ivics, Z., and Izsvák, Z. (2015). Sleeping Beauty Transposition. *Microb. Spectr.* 3 (4), MDNA300422014. doi:10.1128/microbiolspec.MDNA3-0042-2014

Kempken, F., and Windhofer, F. (2001). The hAT Family: A Versatile Transposon Group Common to Plants, Fungi, Animals, and Man. *Chromosoma* 110 (1), 1–9. doi:10.1007/s004120000118

Kia, A., Gloeckner, C., Osothpraporn, T., Gormley, N., Bomati, E., Stephenson, M., et al. (2017). Improved Genome Sequencing Using an Engineered Transposase. *BMC Biotechnol.* 17 (1), 6. doi:10.1186/s12896-016-0326-1

Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., and Sternberg, S. H. (2019). Transposon-Encoded CRISPR-Cas Systems Direct RNA-Guided DNA Integration. *Nature* 571 (7764), 219–225. doi:10.1038/s41586-019-1323-z

Lazarow, K., Du, M.-L., Weinier, R., and Kunze, R. (2012). A Hyperactive Transposase of the Maize Transposable Element Activator (Ac). *Genetics* 191 (3), 747–756. doi:10.1534/genetics.112.139642

Llamas, B., Narzisi, G., Schneider, V., Audano, A. P., Biederstedt, E., Blauvelt, L., et al. (2019). A Strategy for Building and Using a Human Reference Pangenome. *PLoS One* 14(8), 1. doi:10.1371/journal.pone.0221971

Martin, M. (2011). Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBNet* 17, 10–12. doi:10.14806/ej.17.1.200

Mátés, L., Chuah, M. K. L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., et al. (2009). Molecular Evolution of a Novel Hyperactive Sleeping Beauty Transposase Enables Robust Stable Gene Transfer in Vertebrates. *Nat. Genet.* 41 (6), 753–761. doi:10.1038/ng.343

Michel, A. H., Hatakeyama, R., Kimmig, P., Arter, M., Peter, M., Matos, J., et al. (2017). Functional Mapping of Yeast Genomes by Saturated Transposition. *Elife* 6, e23570. doi:10.7554/eLife.23570

**ACKNOWLEDGMENTS**

The authors would like to thank Dr. Brian Green and Professor Brendan Cormack for the Tn5 site specificity dataset and the late Don Comb for the creative, collaborative environment of New England Biolabs which permitted this study.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: [https://www.frontiersin.org/articles/10.3389/fmolb.2021.734154/full#supplementary-material](https://www.frontiersin.org/articles/10.3389/fmolb.2021.734154/full#supplementary-material)
Morgan, R. D., Camp, R. R., Wilson, G. G., and Xu, S. Y. (1996). Molecular Cloning and Expression of NlaIII Restriction-Modification System in E. C. Gene 183 (1-2), 215–218. doi:10.1016/s0378-1119(96)00561-6
Munoz-Lopez, M., and Garcia-Perez, J. (2010). DNA Transposons: Nature and Applications in Genomics. Curr. Genomics 11 (2), 115–128. doi:10.2174/138920210790886871
Qing, X., Walter, J., Jarazo, J., Arias-Fuenzalida, J., Hillje, A.-L., and Schwamborn, J. C. (2017). CRISPR/Cas9 and PiggyBac-Mediated Footprint-Free LRRK2-G2019s Knock-In Reveals Neuronal Complexity Phenotypes and α-Synuclein Modulation in Dopaminergic Neurons. Stem Cel Res. 24, 44–50. doi:10.1016/j.scr.2017.08.013
Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., et al. (2019). RNA-Guided DNA Insertion with CRISPR-Associated Transposases. Science 365 (6448), 48–53. doi:10.1126/science.aax9181
Voigt, F., Wiedemann, L., Zuliani, C., Querques, I., Sebe, A., Mátés, L., et al. (2016). Sleeping Beauty Transposase Structure Allows Rational Design of Hyperactive Variants for Genetic Engineering. Nat. Commun. 7, 11126. doi:10.1038/ncomms11126
Vollbrecht, E., Duvick, J., Scharer, J. P., Ahern, K. R., Dewatthanawong, P., Xu, L., et al. (2010). Genome-Wide Distribution of TransposedDissociationElements in Maize. Plant Cell 22 (6), 1667–1685. doi:10.1105/tpc.109.073452
Walker, I. H., Hsieh, P.-C., and Riggs, P. D. (2010). Mutations in Maltose-Binding Protein that Alter Affinity and Solubility Properties. Appl. Microbiol. Biotechnol. 88 (1), 187–197. doi:10.1007/s00253-010-2696-γ
Weinreich, M. D., Gasch, A., and Reznikoff, W. S. (1994). Evidence that the Cis Preference of the Tn5 Transposase Is Caused by Nonproductive Multimerization. Genes Develop. 8 (19), 2363–2374. doi:10.1101/gad.8.19.2363
Wiegand, T. W., and Reznikoff, W. S. (1992). Characterization of Two Hypertransposing Tn5 Mutants. J. Bacteriol. 174 (4), 1229–1239. doi:10.1128/jb.174.4.1229-1239.1992
Zhou, M., and Reznikoff, W. S. (1997). Tn 5 Transposase Mutants that Alter DNA Binding Specificity 1Edited by G. Smith. J. Mol. Biol. 271 (3), 362–373. doi:10.1006/jmbi.1997.1188

Conflict of Interest: Authors PR, GB-H, JH, GR, LE and MB was employed by company New England Biolabs, United States.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Riggs, Blundell-Hunter, Hagelberger, Ren, Ettwiller and Berkmen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.