# Relaxed DNA substrate specificity of transposases involved in programmed genome rearrangement

Matt W.G. Walker[1,2], Takahiko Akematsu[2,3], Erhan Aslan[2], Danylo J. Villano[2], Harrison S. Fried[2], Hui Lan[4], Samuel H. Sternberg[2,5*], Laura F. Landweber[1,2*]

[1] Department of Biological Sciences, Columbia University, New York, NY, 10027, USA

[2] Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, 10032, USA

[3] Department of Biology, Faculty of Science, Kanagawa University, Yokohama, Kanagawa, 221-8686, Japan

[4] Department of Biology, Barnard College, Columbia University, New York, NY, 10027, USA

[5] Howard Hughes Medical institute, Columbia University, New York, NY, 10032, USA

\* To whom correspondence should be addressed: laura.landweber@columbia.edu, shsternberg@gmail.com

1 **ABSTRACT**

2 During post-zygotic development, the ciliate *Oxytricha trifallax* undergoes massive
3 programmed genome rearrangement that involves over 225,000 DNA cleavage and joining
4 events. An *Oxytricha* family of Tc*1/mariner* transposons, known as Telomere-Bearing Elements
5 (TBEs), encodes a transposase that has been implicated in rearrangement, but its high copy
6 number (>34,000 paralogs) has precluded genetic strategies to investigate its DNA recognition
7 properties directly in *Oxytricha*. Here, we developed a heterologous strategy to assay TBE
8 transposase expression and activity in *E. coli*, revealing highly promiscuous DNA cleavage
9 properties. Systematic ChIP-seq experiments allowed us to define the DNA binding specificities
10 of multiple distinct transposase subfamilies, which exhibited a binding and cleavage preference
11 for short, degenerate sequence motifs that resemble features present within the TBE transposon
12 ends. The relaxed sequence preference is striking for autonomous transposases, which typically
13 recognize their end sequences with strict specificity to avoid compromising host fitness. Finally,
14 we developed a custom antibody to investigate TBE transposases in their native environment and
15 found that they precisely localize to the developing nucleus exclusively during the rearrangement
16 process. Collectively, this work establishes a robust heterologous workflow for the biochemical
17 investigation of enzymes that have been repurposed for large-scale genome rearrangements.

18 **INTRODUCTION**

19 The genomes of many organisms undergo programmed rearrangements that require
20 precise DNA elimination and assembly (Smith *et al*., 2020; Wang & Davis, 2014). Arguably the
21 most complex programmed genome rearrangements occur in the ciliated protozoan, *Oxytricha*
22 *trifallax*. Like most ciliates, *Oxytricha* is binucleated, with a separate germline micronucleus (MIC)
23 and somatic macronucleus (MAC) (Prescott, 1994). During the vegetative (asexual) life cycle, the
24 MIC genome replicates through mitosis, while the MAC genome divides through amitosis, with
25 chromosomes imperfectly segregated to the daughter nuclei. During sexual reproduction,
26 conjugating cells exchange haploid micronuclei that undergo syngamy, the parental MAC
27 degrades and one copy of the zygotic germline MIC develops into a new MAC, initiating an
28 intricate cascade of recombination events requiring over 225,000 DNA cleavage and joining
29 events (X. Chen *et al*., 2014; Prescott, 1994; Swart *et al*., 2013). This rearrangement process also
30 results in the full elimination of >90% of germline DNA.

31 The MIC is present in multiple copies per cell and is largely transcriptionally silent, in part
32 because a large portion of its precursor gene segments are in the wrong order relative to the
33 transcribed version in the MAC; individual gene segments are also interrupted by internally

34    eliminated sequences (IESs) that must be removed in order for retained macronuclear-destined

35    sequences (MDSs) to arrange to produce functional MAC genes (**Figure 1A**). In the MIC, MDSs

36    are flanked by short direct repeats, called *pointers*, only one of which remains after

37    rearrangement. MDS joining produces >16,000 new 'nanochromosomes' that each acquire two

38    telomeres and are amplified to high copy number (average ploidy ~1900), exhibit an average

39    length of 3.2 kb, and usually encode only one gene (X. Chen & Landweber, 2016; Lindblad *et al*.,

40    2019; Prescott, 1994; Swart *et al*., 2013). Previous studies identified key molecular components

41    that are involved in the rearrangement process: 27-nt Piwi-interacting small RNAs (piRNAs) mark

42    MDSs for retention (Fang *et al*., 2012), long noncoding RNAs transcribed from parental MAC

43    chromosomes program the MDS assembly process (Nowacki *et al*., 2008), and a large family of

44    germline-encoded transposases are essential for DNA processing (Nowacki *et al*., 2009).

45        TBE, or Telomere-Bearing Element, transposons are abundant in the MIC, present at

46    >34,000 copies that comprise almost 15% of the germline genome, but they are completely

47    eliminated during development. TBEs encode three distinct ORFs, one of which is a 42 kDa

48    transposase (**Figure 1B**). Phylogenetic analyses indicated that the TBE transposase belongs to

49    the Tc*1*/*mariner* family (Doak *et al*., 1994), which uses a DDE catalytic triad to perform the DNA

50    cleavage and strand joining chemistry required for transposition (Hickman & Dyda, 2016).

51    Phylogenetic analyses also revealed that there are three superfamilies for the TBEs – TBE1,

52    TBE2, and TBE3 – with TBE2 further subdivided into TBE2.1 and TBE2.2 (X. Chen & Landweber,

53    2016) . The TBE moniker originates from the terminal inverted repeat (TIR) sequences flanking

54    the elements, which encode telomere-like repeats $(G_4T_4)_4$. Tc*1*/*mariner* family elements typically

55    contain inverted repeats at their end sequences, which delineate the boundaries of the mobile

56    element and recruit the transposition machinery with strict specificity during mobilization. In

57    *Oxytricha*, TBEs are detected as extrachromosomal circular DNA (eccDNA) during genome

58    rearrangement, indicating that they are excised as circular products (Williams *et al*., 1993; Yerlici

59    *et al*., 2019), although direct mobilization activity has not been experimentally observed. While

60    TBE transposons likely spread and proliferate via transposition, extant TBE transposases may

61    instead play a predominant role in DNA excision, akin to the role of RAG1 recombinase in V(D)J

62    recombination, more often removing rather than mobilizing genomic elements. Unlike Tc*1*/*mariner*

63    elements, which recognize a TA target site and are flanked by two copies of that motif after

64    integration, TBEs are flanked by an ANT motif, which is strikingly consistent with the most

65    common motif of 3-bp pointer sequences (X. Chen *et al*., 2014). This overlap suggests that the

66    motif is either a requirement for transposase-mediated DNA elimination, or a remnant of

67    transposon integration events that led to the emergence of IESs (Klobutcher & Herrick, 1997).

68      Previous work on TBEs indicated their involvement in programmed DNA elimination.

69      Specifically, RNAi knockdown of transposase gene families led to stalled rearrangement of MAC

70      genomic loci, arrested development, and the accumulation of high molecular weight DNA along

71      with incompletely or improperly rearranged genes (Nowacki *et al*., 2009). However, failed

72      rearrangement only occurred when RNAi was directed against all 3 superfamilies of TBEs, limiting

73      the utility of this approach to investigate the activity of specific TBE-encoded transposases.

74      Furthermore, while both reverse and forward genetic tools in *Oxytricha* can target the MAC (Clay

75      *et al*., 2019; Fang *et al*., 2012), techniques for genetic manipulation of the germline MIC are

76      severely compromised by its low copy number and small size compared to the MAC, and this

77      severely limits our ability to determine how and where TBE transposases cleave DNA to effect

78      precise DNA elimination during development.

79      To circumvent these limitations, we developed a heterologous expression approach in

80      *Escherichia coli*, which allowed us to gain insights into DNA binding and cleavage activity of the

81      TBE transposase. We uncovered DNA binding preferences within the *E. coli* genome and its own

82      TBE transposon ends, and find that this binding activity is consistent with DNA cleavage, which

83      we detected using deep-sequencing to profile fragmented DNA. Our work presents the first direct

84      evidence of TBE transposase binding and cleavage activity, and establishes a robust workflow

85      for studying a previously intractable enzyme implicated in genome rearrangement.


86      **MATERIALS AND METHODS**

87      **Plasmid and *E. coli* strain construction.**

88      Plasmids and oligonucleotides used in this study are listed in **Supplementary Tables 1**

89      and **2**, respectively. *E. coli* codon-optimized open reading frames for TBE transposon-encoded

90      proteins were synthesized and cloned into expression vectors under the control of an IPTG-

91      inducible T7 promoter on medium-copy pET vectors with carbenicillin resistance. Epitope-tagged

92      protein expression vectors were constructed by appending a C-terminal FZZ tag (3×FLAG-tag,

93      TEV protease cleavage size, and ZZ domain of protein A, described in (Kataoka *et al*., 2010)).

94      Catalytically-inactivated derivatives of the TBE transposase proteins were constructed by

95      'around-the-horn' PCR, to mutate each residue of the catalytic triad (DDE) to alanine (<u>A</u>DE, D<u>A</u>E

96      and DD<u>A</u>). All plasmids were cloned into NEB Turbo cells (NEB), purified (QIAprep Spin Miniprep

97      Kit), and verified by Sanger sequencing (GENEWIZ).

98      Four custom *E. coli* BL21(DE3) strains were generated encoding a genomically-integrated

99      TBE transposon belonging to each of the TBE families (TBE1 from contig ctg7180000067530,

100     TBE2.1 from contig 7180000067905, TBE2.2 from contig ctg7180000069065, and TBE3 from

101    ctg7180000089059). These strains were generated by Himar1C9-mediated transposition. We
102    cloned representative TBE transposons from the four TBE families onto a temperature-sensitive
103    pSC101 plasmid, downstream of a kanamycin resistance cassette, which was flanked by *mariner*-
104    family transposon end sequences. The plasmid also encoded a transposase expression cassette,
105    which produced the Himar1C9 protein required for transposition of the *mariner* element. The
106    temperature-sensitive plasmid was cured before selecting for antibiotic resistance, which could
107    only be conferred after genomic integration of the TBE element. Specifically, after transformation
108    by heat shock, cells were recovered at the restrictive temperature and plated on LB agar plates
109    supplemented with kanamycin (50 µg/ml), and colonies that grew were confirmed by PCR to have
110    lost the temperature-sensitive TBE plasmid. The genomic integration was later confirmed for each
111    of the four strains by Illumina deep sequencing (Smear-Seq and ChIP-seq), which detected a
112    single copy of the transposon in the genome and revealed the integration site, allowing us to
113    generate complete *E. coli* reference genomes for each strain that were used for read mapping
114    and analysis.

115    **DNA fragmentation assay and time course experiments.**
116    The DNA fragmentation analysis relied on inducing transposase expression in liquid
117    media, performing a miniprep to isolate DNA, and separating DNA by electrophoresis on a 1%
118    agarose gel. First, chemically competent *E. coli* BL21(DE3) cells were transformed and single
119    colonies were grown overnight at 37 °C in LB medium with 100 µg/ml carbenicillin. The next day,
120    cultures were diluted 1:100 in fresh LB medium with 100 µg/ml carbenicillin and grown at 37 °C
121    until mid-logarithmic phase ($OD_{600}$=0.4). Cultures were then induced with 0.1 mM IPTG and
122    transferred to 16 °C. For the time course experiments, multiple 2 ml cultures were set up in
123    replicate. After the defined induction time, cell density was measured ($OD_{600}$), DNA was isolated
124    (QIAGEN Plasmid Mini Kit), and DNA concentration was measured (DeNovix
125    Spectrophotometer). 500 ng of DNA was loaded onto each lane of a 1% agarose gel stained with
126    SYBR Safe (Thermo Fisher Scientific), subjected to electrophoresis (130 V for 20 min), and
127    imaged (Bio-Rad GelDoc XR Imaging System).

128    **Structural predictions and sequence alignments.**
129    Structural models of the TBE transposase dimer were predicted using AlphaFold version
130    2.3.1 (commit 18e12d6) (Jumper *et al*., 2021), using the multimer model preset (--
131    model_preset=multimer) (Evans *et al*., 2022) , and displayed in ChimeraX (v1.6). Structural
132    models of the TBE transposase dimer interacting with 40-bp of transposon end DNA were

133  predicted using RoseTTAFoldNA (Baek *et al*., 2023). MSAs were generated by Clustal Omega

134  run on EMBL's Job Dispatcher (Madeira *et al*., 2022), and visualized in Jalview version 2.11.3.2

135  (Waterhouse *et al*., 2009) .


136  **NGS approach for sequencing fragmented DNA ("Smear-seq").**

137  Our approach to sequencing fragmented DNA relied on DNA isolation, dA-tailing, second-

138  strand synthesis, sonication, library prep, and next-generation sequencing (**Figure 2A**). First, the

139  wildtype or catalytically-inactivated transposase was expressed as in the DNA fragmentation

140  assay, wherein we diluted overnight cultures of *E. coli* BL21(DE3) cells expressing the

141  transposase 1:100 in LB with 100 µg/ml carbenicillin, induced mid-logarithmic phase cultures with

142  0.1 mM IPTG, and grew cultures for 18 h at 16 °C. DNA was isolated (QIAGEN Plasmid Mini Kit),

143  and A-tailing was performed using an approximate molar ratio of 1:1,000 of DNA:dATPs (NEB

144  Terminal Transferase). After incubating at 37 °C for 30 min and heat-inactivating at 70 °C, DNA

145  was cleaned and eluted in 40 µl Milli-Q H$_2$O (QIAquick PCR Purification Kit). Concentration was

146  measured by fluorometry (DeNovix dsDNA High Sensitivity Assay) before second-strand

147  synthesis using oligo(dT)18 primers and PrimeSTAR Max DNA Polymerase (Takara). DNA was

148  cleaned and eluted in 40 µl MilliQ H$_2$O and the concentration was measured by fluorometry before

149  normalized samples were loaded into a microTUBE AFA Fiber Crimp-Cap and TE buffer was

150  added to a total volume of 130 µl. Samples were then sheared by sonication on an M220 Focused-

151  ultrasonicator (Covaris) using the following SonoLab 7.2 settings: peak power, 50.0; duty factor,

152  20; cycles/burst, 200; treatment time, 150 s. After sonication, DNA was prepared for NGS using

153  the NEBNext Ultra II DNA Library Prep for Illumina (NEB). After adaptor ligation, DNA was purified

154  without size selection using AMPure XP beads (Beckman Coulter) and appended sequencing

155  indexes with 10 cycles of PCR and NEBNext Q5 Master Mix. Two-sided size selection was used

156  to clean and select ~450 bp fragments. To first remove small fragments, 0.55× AMPure XP bead

157  solution was added and samples were allowed to separate on a magnetic rack before the

158  supernatant containing small DNA fragments was removed. Then, to remove large fragments,

159  0.35× AMPure XP bead solution was added to each sample, separated on a magnetic rack, and

160  the supernatant discarded. Beads were then washed twice in 80% (v/v) EtOH and eluted in 12 µl

161  of TE buffer. Concentration was determined by fluorometry (DeNovix dsDNA High Sensitivity Kit)

162  and samples were sequenced using a NextSeq High Output Kit with 150 cycles (Illumina).

163  After sequencing, paired-end reads were filtered and trimmed using fastp (S. Chen *et al*.,

164  2018) and mapped to the *E. coli* BL21(DE3) reference genome and to the TBE expression plasmid

165  using bowtie2 (Langmead & Salzberg, 2012). Then, we used Samtools to sort, index, and filter

166    multi-mapping reads (Danecek *et al*., 2021). Finally, coverage was normalized to counts per

167    million (CPM) using the deepTools2 command bamCoverage (Ramírez *et al*., 2016).

168    *Oxytricha* **culture conditions and mating**

169    *Oxytricha* strains were cultured and mated as in Fang *et al*., 2012. Two wild type *Oxytricha*

170    strains with different mating types, JRB310 and JRB510 were cultured in Pringsheim media (0.11

171    mM $Na_2HPO_4$, 0.08 mM $MgSO_4$, 0.85 mM, $Ca(NO_3)_2$, 0.35 mM KCl, pH 7.0) supplemented with

172    *Chlamydomonas reinhardtii* and *Klebsiella* as the food source. To induce mating, cells were

173    allowed to deplete *Chlamydomonas* overnight, filtered through cheesecloth to remove any

174    remaining debris, and equal amounts of cells were mixed at a density of 5,000 cells/ml with a final

175    volume of 300 ml. Pairs typically form 3–4 hours post-mixing, and maximum pairing efficiency

176    (~70–80%) is generally achieved 12 hours post-mixing.

177    **Production of anti-TBE transposase antibodies**

178    The peptide KRQHLNSKPKRPLK was synthesized and used for immunization in rabbits

179    to yield anti-TBE transposase antibodies (GenScript). Rabbits were immunized three times at two-

180    week intervals. 4.28 mg antibody was purified and suspended in 7.5 ml PBS with 0.02% ProClin

181    300 preservative. The antibody was validated by Western hybridization in *E. coli* cells expressing

182    the TBE transposase, as described below.

183    **Western blotting of heterologous TBE protein expression in** *E. coli*.

184    Protein expression in *E. coli* was confirmed by Western analysis, which was performed

185    essentially as previously described (Walker *et al*., 2023) with minor modifications. *E. coli* overnight

186    cultures expressing FZZ- or MBP-tagged proteins were diluted 1:100 in LB supplemented with

187    100 µg/ml carbenicillin and grown at 37 °C to $OD_{600}$=0.4. Protein expression was induced with 0.1

188    mM IPTG, and cultures were transferred to 16 °C and grown for 18 h. The equivalent of 1 ml of

189    $OD_{600}$=1.5 culture was pelleted by centrifugation (12,000 x g for 5 min at 4 °C), and the pellet was

190    suspended in 150 µl lysis buffer (20 mM Tris-Cl [pH 7.5], 150 mM NaCl, 0.5 mg/ml lysozyme).

191    Samples were incubated at 25 °C for 10 mins before adding 10 mM DTT, 1% N-lauroyl sarcosine

192    and 2X SDS loading dye (100 mM Tris-Cl [pH 6.8], 4% (w/v) SDS, 30 % (v/v) glycerol) and 0.07%

193    (w/v) bromophenol blue). Samples were incubated at 95 °C for 10 min, and 5 µl of each sample

194    underwent separation on a 4-12% gradient SDS-PAGE gel (Bio-Rad, Mini-PROTEIN TGX) at 100

195    V for 90 min. Proteins were transferred to a PVDF membrane (Invitrogen iBlot 2 Transfer Stack).

196    Subsequent washing, blocking, and antibody incubation were performed by gentle nutation at

197    room temperature. Membranes were washed with 1× PBS + 0.1% Tween 20, then incubated for

198    1 h in blocking buffer (1× PBS, 0.1% Tween 20, 5% BSA), washed in 1× PBS + 0.1% Tween 20,

199    then incubated in primary antibody diluted in blocking buffer for 1 h at room temperature. For the

200    FZZ-tagged samples, 1:40,000 monoclonal ANTI-FLAG M2 antibody produced in mouse (Sigma-

201    Aldrich, F1804) was used. For the MBP-tagged sample, 1:10,000 of anti-TBE transposase

202    antibody (GenScript) was used. For the loading control, 1:5,000 anti-GAPDH monoclonal antibody

203    produced in mouse (Invitrogen, MA5-15738) was used. Membranes were washed three times in

204    1× PBS + 0.1% Tween 20, and primary antibodies were detected with either goat anti-mouse

205    IgG1 heavy chain HRP (Abcam, ab97240) or goat anti-rabbit IgG (H+L)-HRP (Bio-Rad, 1706515).

206    Finally, we performed enhanced chemiluminescence (ThermoFisher SuperSignal West Dura

207    Extended Duration Substrate) and imaged blots on an Amersham Imager 600 (GE).

208    **Western blotting of native TBE protein expression in *Oxytricha***

209    ~300,000 mating cells from different time points were collected by centrifugation at 130 *g*

210    for 2 min using a swinging bucket centrifuge (Sorvall RC6, Thermo Scientific). The supernatant

211    was discarded, and the cell pellet was suspended in the residual medium. Next, 2x SDS loading

212    buffer was directly added to the samples, which were then heated at 95 °C for 10 min. The

213    samples were run on TGX Stain-Free Precast Gels (Bio-Rad) and transferred to a 0.2 µm Immun-

214    Blot PVDF Membrane (Bio-Rad, 1620177) under semi-dry conditions using Trans-Blot SD blotter.

215    The membrane was blocked with 5% nonfat dry milk (biokemix, M0841) in 1x TBS-T for 1 hour at

216    RT and then incubated overnight at 4 °C on a rotator with a custom-made rabbit anti-TBE

217    transposase antibody (1:1000 dilution in blocker). The following day, the membrane was washed

218    five times with 1x TBS-T (5 min each) and incubated with 1:10,000 goat anti-rabbit IgG (H + L)-

219    HRP (Bio-Rad, 1706515) for 1 hour at RT. After final washes (five times with 1x TBS-T, 5 min

220    each), the chemiluminescence signal was detected using Clarity Western ECL reagent (Bio-Rad),

221    and the membrane was imaged with a ChemiDoc imager (Bio-Rad).

222    **Chromatin-immunoprecipitation followed by next-generation sequencing (ChIP-seq) in *E.***

223    ***coli***

224    ChIP-seq in *E. coli* was performed essentially as described (Bonocora & Wade, 2015;

225    Hoffmann *et al.*, 2022; Park, 2009), with some modifications. Cultures were prepared and protein

226    expression was induced as described in the Western blot method, above. After growing induced

227    cultures at 16 °C for 18 h, formaldehyde was added to a final concentration of 1% to each 40 ml

228    of culture and nutated for 20 min at room temperature. Formaldehyde was quenched by adding

229    4.6 ml of 2.5 M glycine, nutating for 10 min. Samples were then centrifuged at 4000 x g at 4 °C

230    for 8 min. The cell pellet was washed twice in cold 1× TBS, and then the equivalent of 40 ml of

231    $OD_{600}$=0.6 was transferred to a tube to normalize cell density across samples and pelleted again

232    by centrifugation. The pellet was transferred to a 1.5 ml Eppendorf tube and centrifuged at 10,000

233    x g at 4 °C for 5 min. The supernatant was discarded and cell pellets were flash-frozen in liquid

234    nitrogen.

235    Cell pellets were resuspended in lysis buffer (50 mM HEPES-KOH [pH 7.5], 0.1% (w/v)

236    sodium deoxycholate, 0.1% (w/v) SDS, 1 mM EDTA, 1% (v/v) Triton X-100, 150mM NaCl, 1×

237    cOmplete Roche protease inhibitor), and transferred to a 1 ml milliTUBE AFA Fiber (Covaris).

238    Samples were sonicated on an M220 Focused-ultrasonicator (Covaris) with the following

239    SonoLab 7.2 settings: minimum temperature, 4 °C; set point, 6 °C; maximum temperature, 8 °C;

240    peak power, 75.0; duty factor, 10; cycles/bursts, 200; sonication time, 17.5 min. After sonication,

241    10 µl of sheared cleared lysate was transferred to a separate tube as the pre-IP control ("input").

242    The remainder was used for immunoprecipitation.

243    For immunoprecipitation, 25 µl Dynabeads Protein G (Thermo Fisher Scientific) for each

244    sample was washed four times with 1× PBS + 0.5% BSA and mixed with 4 µl monoclonal anti-

245    FLAG M2 antibody produced in mouse (Sigma-Aldrich F1804), which was conjugated at 4 °C for

246    6 h. After conjugation, beads were washed four times with 1× PBS + 0.5% BSA, resuspended in

247    30 µl lysis buffer, then mixed with sonicated samples and rotated overnight at 4 °C.

248    The following day, beads were washed three times with lysis buffer lacking protease

249    inhibitor, then washed once with lysis buffer lacking protease inhibitor and supplemented with 500

250    mM NaCl, once with ChIP wash buffer (10 mM Tris–HCl [pH 8.0], 250 mM LiCl, 0.5% (w/v) sodium

251    deoxycholate, 0.5% (v/v) Nonidet-P40, 1 mM EDTA) and finally twice with TE buffer (10 mM Tris-

252    HCl [pH 8.0], 1 mM EDTA). After the final wash, beads were suspended in 200 µl elution buffer

253    (1% (w/v) SDS, 0.1 M $NaHCO_3$) and incubated at 65 °C for 1 h 15 min to release protein-DNA

254    complexes from the beads, vortexing every 15 min to resuspend the beads. While incubating the

255    immunoprecipitated samples, 10 µl of the frozen pre-IP samples were withdrawn and 190 µl

256    elution buffer was added. After the 65 °C incubation was complete, 10 µl of 5 M NaCl was added

257    to each pre-IP and IP sample, which were incubated at 65 °C overnight to reverse crosslinks.

258    After overnight incubation, RNA was digested by adding 1 µl RNase A (Thermo Fisher

259    Scientific) and incubating at 37 °C for 1 h. Protein was digested by adding 2.8 µl of 20 mg/ml

260    proteinase K (Thermo Fisher Scientific) and incubating at 55 °C for 1 h. DNA was purified

261    (QIAquick PCR Purification Kit) and eluted in 40 µl TE buffer before the concentration was

262    measured by fluorometry (DeNovix dsDNA Ultra High Sensitivity Kit). Samples were normalized

263     to the lowest concentration, and DNA was prepared for next-generation sequencing using the

264     NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB). ChIP-seq samples were sequenced on

265     a paired-end run using NextSeq Mid or High Output Kits with 150 cycles (Illumina).

266         Paired-end reads were filtered, trimmed, and mapped as described in the Smear-seq

267     workflow above. After read mapping, peaks were called using MACS2 (Zhang *et al.*, 2008), filtered

268     to exclude peaks with a score less than 20, and *de novo* motif prediction was performed using

269     Homer using the command 'findMotifsGenome.pl' with default settings and a window size of 100

270     (Heinz *et al.*, 2010).

271     **ChIP-seq in *Oxytricha***

272         *Oxytricha* cell lysates were prepared for ChIP-seq at 0 h, 36 h, and 48 h after mixing of

273     cells of compatible mating type. Two million cells were fixed in 1% formaldehyde (Sigma) for 2

274     min, and crosslinking was quenched by adding glycine to a final concentration of 125 mM,

275     followed by incubation for 5 min at RT on a rotator. Fixed cells were then lysed in 1 ml of ChIP

276     lysis buffer (50 mM HEPES/KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 1% SDS,

277     0.1% sodium deoxycholate, 1x Protease Inhibitors Cocktail, Sigma) for 15 min on ice. Chromatin

278     was sheared using Q800R3 Sonicator with 25 cycles of 30 seconds on/off with 40% amplitude,

279     spun down, and the cleared lysate was flash frozen at -80 °C.

280         Each 1 ml sample was diluted 10x in dilution buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl,

281     1mM EDTA pH 8.0, 1% Triton, 1x cOmplete Roche Protease Inhibitor), and pre-cleared in

282     Dynabeads Protein G. To bind the anti-TBE transposase antibody to the beads, 50 µl Dynabeads

283     Protein G for each sample was washed two times with 1× PBST and mixed with 17.5 µl of custom

284     anti-TBE transposase antibody produced in rabbits, which was conjugated at 4 °C for 6 h. After

285     conjugation, beads were washed once with 1× PBST, resuspended in 50 µl dilution buffer, then

286     mixed with the sonicated pre-cleared samples and rotated overnight at 4 °C. The following day,

287     the remaining ChIP-seq steps (protein/RNA digestion, DNA cleanup, and NGS library preparation)

288     were performed as described above, in the methods section for *E. coli* ChIP-seq.

289     **Indirect immunofluorescence**

290         Immunofluorescent staining experiments were performed as previously described (Fang

291     *et al.*, 2012) with minor modifications. Briefly, cells were fixed with 4% paraformaldehyde

292     (Electron Microscopy Sciences) for 10 min at RT on an end-to-end rotator, followed by two washes

293     with 1x PBS. Fixed cells were placed on a Poly-L-Lysine-coated (0.1 mg/ml, Gibco) hydrophobic

294     printed 12-well slide (Epredia, ER202W) and incubated overnight at 4 °C. The cells were then

295    permeabilized with permeabilization solution (0.5% Triton X-100 in 1x PBS) for 30 min at RT,

296    followed by a 5 min incubation in 0.1 M HCl. After a brief wash with 1x PBS, Image-iT™ FX Signal

297    Enhancer (Invitrogen) was applied, and the cells were incubated for 30 min at RT. Cells were

298    subsequently blocked with blocking buffer (0.2% Triton X-100, 0.5% IgG, and protease-free BSA

299    in 1x PBS) for 1 hour at RT. Rabbit anti-TBE transposase antibody, diluted 1:100 in blocking

300    buffer, was applied to the cells, and the slide was incubated overnight at 4 °C. Next day, the cells

301    were washed three times with washing solution (0.2% Triton X-100 in 1x PBS) for 10 min each at

302    RT. Alexa Fluor™ 488-conjugated goat anti-rabbit secondary antibody (1:1000 in blocking

303    solution; Invitrogen, A11034) was then applied, and the slide was incubated at 37 °C for 1 hour

304    in the dark. After a final series of three washes with washing solution (10 min each at RT), DAPI-

305    containing VECTASHIELD® Antifade Mounting Medium (Vector laboratories, H-1000-10) was

306    added to the samples, and the slide was covered with a long cover glass (Fisherbrand, 12545J).

307    Fluorescent images were captured using a Zeiss Axiovert 200 inverted microscope equipped with

308    an Axiocam 506 monochrome camera. Images were pseudo-colored using ImageJ64 when

309    merging two channels.


310    **RESULTS**

311    **Expression of the TBE transposase ORF induces widespread DNA fragmentation in *E. coli***

312            To study the function of a high copy number transposase implicated in genome

313    rearrangements in *Oxytricha*, we set out to heterologously express candidate TBE transposases

314    in *E. coli*, focusing our initial efforts on a representative transposon encoded in the *Oxytricha* MIC

315    genome and referred to as TBE2.1(7905), a member of the TBE2.1 family encoded on contig

316    7180000067905 (Chen *et al.*, 2014). When we expressed the protein in *E. coli* BL21(DE3) and

317    isolated and analyzed the cellular DNA fraction (**Figure 1B**), we observed a striking DNA

318    fragmentation phenotype, characterized by a significant and reproducible DNA smear upon

319    electrophoretic separation (**Figure 1C**). Apparent DNA degradation was visible as early as 3

320    hours after induction, and by 18 h, the DNA products ranged broadly in length. To determine

321    whether this phenotype was merely a cytotoxic byproduct of heterologous protein over-

322    expression, we cloned inactivated variants of the transposase, mutated in the predicted active

323    site from DDE to DAE. This and other single amino acid alanine substitutions in the DDE motif

324    abolished the DNA smearing phenotype (**Figure 1D**), providing the first direct evidence of TBE

325    transposase-mediated DNA cleavage activity. Of note, these experiments in *E. coli* lacked any

326    presence of bona fide TBE transposon DNA substrates.

327    We next set out to determine the composition of DNA products that appeared after

328    transposase induction, via a high-throughput sequencing approach here dubbed 'Smear-seq'

329    (**Figure 2A** and Materials and Methods). To prepare samples, we induced the WT or catalytically-

330    inactivated transposase for either 1 h or 18 h, extracted DNA by miniprep, and then performed 3'

331    polyadenylation and second strand synthesis, followed by ligation to sequencing adaptors, and

332    sequencing on an Illumina platform. After mapping sequencing reads to the TBE expression

333    vector and the *E. coli* BL21(DE3) genome, we quantified the relative composition of plasmid or

334    genomic DNA in each sample and found that after 18 h of induction, the WT transposase samples

335    primarily comprised genomic DNA (>95% of reads), whereas the 3 h timepoint and the

336    catalytically-inactivated transposase samples primarily comprised plasmid DNA (**Figure 2B**).

337    These data suggest that presence of the TBE transposase active site permits extensive genomic

338    DNA cleavage by the transposase in *E. coli*.

339    To test whether genomic DNA fragmentation was accompanied by cell death, we

340    measured cell density by spectrophotometry and observed a dramatic loss of cell density after

341    induction of the WT transposase, but not after induction of catalytically-inactivate transposase

342    (**Figure 2C**), consistent with the genomic DNA cleavage activity of the wildtype TBE enzyme.

343    Our deep-sequencing approach also allowed us to search for preferred DNA cleavage

344    sites, by examining the distribution of reads across the *E. coli* reference genome. We observed

345    reproducible features in the WT 18 h sample on megabase-scale genomic regions within the *E.*

346    *coli* genome, with consistent regions of enrichment and depletion and coverage profiles that were

347    nearly indistinguishable across two independent biological replicates (**Figure 2D**). These data

348    suggests that the TBE transposase may be preferentially cleaving consistent regions of the

349    genome.


350    **ChIP-seq reveals degenerate DNA sequence motifs bound by TBE transposases**

351    To more directly and globally profile specific DNA sites in the *E. coli* genome recognized

352    by the TBE transposase, we applied ChIP-seq (chromatin immunoprecipitation followed by DNA

353    sequencing) using the WT or catalytically-inactivated (DAE) transposase fused to a C-terminal

354    epitope tag (3xFLAG) (**Figure 3A**), reasoning that the DAE variant would prevent DNA

355    degradation and thus reveal the underlying substrate specificity. Examining the ChIP-seq read

356    coverage across the *E. coli* genome revealed narrow regions of enrichment and depletion that

357    were highly consistent across biological replicates (**Figure 3B**), in stark contrast to the more

358    diffuse Smear-seq signal (**Figure 2D**). We induced each transposase variant for 3 h and 18 h and

359    observed strong agreement between the ChIP-seq profiles for all samples excluding WT at 18 h,

360    which had more diminished peaks, we conjecture due to accumulated DNA cleavage activity

361    reducing the signal.

362         To test whether conserved DNA sequence features were present within the genomic

363    regions of enrichment, we performed peak-calling and *de novo* motif prediction. This analysis

364    revealed a short sequence motif (5'-ANTTTGA) as the top motif detected in all samples except

365    for WT 18 h, where no statistically significant motifs were found (**Figure 3C**). This motif was not

366    only called as the top motif in each sample, but also the only motif that met the significance

367    threshold in each of the three samples (**Figure S1A**). The absence of statistically significant motifs

368    in the WT 18 h sample is likely also due to transposase cleavage activity leading to cleavage and

369    product release, thus decreasing enrichment at preferred binding sites. Intriguingly, the short

370    sequence motif is also found within the TBE transposon terminal inverted repeats (**Figure 3C,**

371    **TIR**), which transposases typically recognize to mobilize an element.

372         We hypothesized that this motif would resemble sequence features found within the TBE

373    transposon terminal inverted repeat (TIR) ends, and that structural modeling might identity likely

374    points of interaction between the transposase and TIRs. To test this, we used AlphaFold Multimer

375    (Jumper *et al*., 2021) to predict the structure of the TBE transposase dimer, which suggested that

376    it was similar to the solved structure of another Tc*1/mariner* family transposase, Mos1, with an

377    RMSD of 7.991Å. (**Figure S1B**). Mos1, first isolated from *Drosophila mauritiana* (G. Bryan *et al*.,

378    1990; G. J. Bryan *et al*., 1987), encodes an N-terminal DNA binding domain (residues 1-112) and

379    C-terminal DD[D/E]-family catalytic domain (residues 126-345). Each N-terminal DNA binding

380    domain encodes two α-helical motifs (HTH1 and HTH2) which contact the *mariner* transposon

381    ends at positions 21-26 and 8-13, respectively (Richardson *et al*., 2009). Intriguingly, when we

382    predicted the structure of the TBE transposase with 40 bp of TIR DNA (RoseTTAFoldNA; Baek

383    *et al*., 2023), its two HTH motifs were positioned near the TTT trinucleotide motifs at the

384    transposon ends, with the DNA binding domains contacting TIR DNA at positions 24-27 and 16-

385    18, respectively (**Figure S1C**). The TTT trinucleotide motif was also detected in the *E. coli* ChIP-

386    seq peaks (**Figure 3C**), lending additional support to the interpretation that the transposase

387    exhibits DNA sequence preferences for this motif.

388         Next, we set out to test whether we could detect transposase enrichment within the native

389    TBE transposon end sequences. We generated a genomically-modified *E. coli* strain that

390    harbored a complete TBE transposon element in its circular chromosome and performed ChIP-

391    seq in this new genetic background (**Figure 3D**). We observed strong enrichment within the

392    transposon terminal inverted repeats (TIRs), offering the first evidence that the TBE transposase

393 recognizes its transposon ends and suggesting that the TBE2.1 transposase may be actively

394 targeting TBE elements during *Oxytricha* genome rearrangement.

395 To test whether the transposase could also recognize *Oxytricha* genomic elements that

396 are eliminated during programmed DNA rearrangement, we cloned a ~4.5-kb region of

397 micronuclear DNA covering the TEBPβ gene locus, which has the most thoroughly studied

398 rearrangement products (Nowacki *et al*., 2008, 2009). To avoid reintegrating the *Oxytricha*

399 genomic element into the *E. coli* genome, we first tested whether we could recapitulate ChIP-seq

400 results with a plasmid-encoded TBE transposon element, and indeed observed consistent

401 enrichment profiles between the genomic- and plasmid-encoded elements (**Figure S2A**). When

402 we examined the enrichment profile across the native TEBPβ locus, introduced on a bacterial

403 plasmid, we observed a variable pattern of enrichment clustered near eliminated regions (**Figure**

404 **S3B**), with read coverage highest in a region with a high-density of rearrangement junctions and

405 IESs. This result may indicate that other protein or nucleic acid features are responsible for

406 targeting transposase activity to IESs during rearrangement.

**TBE accessory proteins do not alter TBE DNA binding profiles in *E. coli***

408 The TBE transposon encodes two proteins in addition to the putative transposase (**Figure**

409 **1B**), which are under purifying selection and are proposed to function as accessory factors during

410 transposition (X. Chen & Landweber, 2016). This feature is surprising for Tc*1/mariner* elements,

411 which typically encode only their transposase and generally lack accessory proteins (Dupeyron

412 *et al*., 2020). One of the two *Oxytricha* proteins is estimated to be 57-kDa and comprises predicted

413 zinc-finger and kinase domains, while the other is estimated to be 22-kDa protein with no

414 predicted domains (X. Chen & Landweber, 2016) (**Figure S3A**). To assay in *E. coli* whether the

415 expression of these accessory proteins could affect DNA cleavage or binding of the TBE

416 transposase, we prepared epitope-tagged expression vectors for each component and co-

417 expressed them with untagged dual-expression vectors for the other components (**Figure S3B**).

418 We first confirmed protein expression of each element by Western blot and observed bands at

419 the expected size for each protein, as well as a second, smaller band for the 57-kDa sample that

420 might represent a protein degradation product (**Figure S3C**). ChIP-seq against tagged versions

421 of either accessory protein or the TBE transposase, in the presence or absence of the two other

422 untagged proteins, did not display differences in the DNA-binding profiles for any combination

423 relative to the pre-IP input profile, suggesting that the accessory proteins might not bind DNA with

424 specificity (**Figure S3D**) or that the *E. coli* expressed versions might not capture the *Oxytricha*

425     environment or modifications necessary for function. It is also possible that the accessory proteins

426     participate in later steps of transposition or repair during genome rearrangement.

**427     Transposase homologs exhibit similar DNA sequence preferences**

428     Phylogenetic analysis of the 42-kDa transposase ORF encoded within diverse TBE

429     transposons revealed that they cluster into four families — TBE1, TBE2.1, TBE2.2, and TBE3

430     (**Figure 4A**) (X. Chen & Landweber, 2016) — and allowed the generation of consensus protein

431     sequences for each family that share >70% amino acid identity (X. Chen & Landweber, 2016) and

432     possess conserved catalytic active sites (**Figure S4**). We wondered whether differences among

433     transposase families might confer differences in transposase activity and therefore repeated the

434     DNA fragmentation and ChIP-seq assays using the consensus sequences for each of the four

435     TBE families (**Figure 4B-D**). Each transposase representative induced a similar DNA

436     fragmentation phenotype, although the phenotype was less dramatic for the TBE3 family

437     sequence (**Figure 4B**). We also prepared catalytically-inactive variants of each transposase

438     family sequence and confirmed that DNA fragmentation required the presence of an intact

439     catalytic triad. ChIP-seq experiments using the consensus sequences revealed consistent profiles

440     of DNA enrichment across the *E. coli* genome, although the profiles for the TBE1 and TBE3

441     representatives were less pronounced (**Figure 4C**). The most noticeable difference when using

442     consensus sequences arose when we performed ChIP-seq in *E. coli* strains encoding

443     genomically-integrated TBE transposons from each family. In these genomic backgrounds, we

444     did not detect strong enrichment within the transposon TIRs (**Figure 4D**). This result could be

445     attributed to differences in the DNA binding preference, or to differences between the transposon

446     TIRs among families, which differ up to 47% (X. Chen & Landweber, 2016).

**447     Transposase binding correlates with DNA cleavage**

448     We wondered whether the DNA signatures from our ChIP-seq and Smear-seq approaches

449     overlapped, which would indicate correspondence between DNA binding and cleavage activity.

450     Visual inspection of the read coverage profiles from both datasets revealed an apparent

451     consistency between the ChIP-seq and 18 h WT Smear-seq samples, despite the more diffuse

452     mapping in the Smear-seq profiles compared to the ChIP-seq profiles, which could be due to

453     downstream DNA degradation after transposase-mediated DNA cleavage (**Figure 5A**, **Figure**

454     **S5A**). To test whether this qualitative similarity was statistically significant, we averaged the

455     genome-wide read coverage within 2-kb bins and compared the average coverage between

456     samples (**Figure 5B-D**, **Figure S5B**). This analysis revealed a moderate positive correlation (*r* =

457 0.48) between the ChIP-seq signal and the Smear-seq signal after 18 h of inducing the

458 catalytically active transposase (**Figure 5C**). The moderate correlation disappeared when we

459 compared the ChIP-seq signal to the Smear-seq signal after inducing the catalytically-inactive

460 transposase (**Figure 5D**). We also observed a remarkable consistency between Smear-seq

461 replicates at 18 hours using the WT transposase (*r* = 0.99 **Figure S5B**), reflecting reproducible

462 and thorough DNA degradation in the presence of catalytically active *Oxytricha* transposase

463 expressed in *E. coli*, confirming efficient expression. The correlation patterns remained consistent

464 across all bin sizes ranging from 500 bp to 10 kb (**Figure S5C**). Overall, the consistency between

465 regions of DNA binding, as measured by ChIP-seq, and regions of DNA cleavage, as measured

466 by Smear-seq, indicates that the DNA recognition and cleavage activity of the TBE transposase

467 is sequence-specific, favoring degenerate motifs that are partially present within the TBE

468 transposon ends.

469 **TBE transposase is expressed in the developing nucleus and preferentially associates**

470 **with TBE elements in *Oxytricha***

471 Efforts to study the biochemical properties of TBE transposases in vitro have been

472 hampered by protein insolubility (Orsolya Barabas, personal communication), so we decided to

473 pursue alternative experiments to monitor its native localization and activity in *Oxytricha*. We

474 selected a 14-amino-acid peptide corresponding to the predicted surface-exposed region

475 between the two N-terminal DNA binding domains (**Figure S6A**), which is conserved in TBE2.1

476 and TBE2.2 (**Figure S6B**), and purified polyclonal antibodies after rabbit immunization. Western

477 hybridization upon heterologous TBE transposase expression in *E. coli* revealed a clear band at

478 the expected size (**Figure S6C**), supporting the specificity of the antibody.

479 Next, we performed a mating time course of *Oxytricha* cells and collected lysates at 12-

480 hour intervals for anti-TBE transposase western analysis to monitor transposase expression. A

481 strong 42-kDa band was present at 36 and 48 h, corresponding to the expected peak of genome

482 rearrangement (**Figure 6A**) (X. Chen *et al.*, 2014; Williams *et al.*, 1993). To investigate TBE

483 transposase localization, we performed immunofluorescence at the same 12-hour intervals post-

484 mixing and observed a strong signal in developing somatic nuclei (**Figure 6B**). These data reveal

485 that TBE transposases are not only expressed during development, but that they specifically

486 localize to *Oxytricha*'s developing somatic macronucleus.

487 Finally, we performed ChIP-seq experiments in *Oxytricha* cells at 0, 36, and 48 hours post-

488 mixing. When we mapped sequence reads back to the TBE-family transposons and compared

489 the proportion of reads mapping to TBEs versus to the remainder of the micronuclear genome,

490  we observed a ~4-fold enrichment compared to input, and a striking >8-fold enrichment of TBEs

491  at developmental timepoints compared to the 0 h timepoint (**Figure 6C**), revealing that TBE

492  transposases preferentially target TBE transposon elements natively in *Oxytricha*. We did not

493  observe peaks of enrichment within the transposon end sequences themselves (**Figure S6D**),

494  which may be due to read-mapping artifacts; because we are mapping ChIP-seq reads to a single

495  TBE2.1(7905) element, the sequence variability flanking the high copy-number TBEs

496  heterogeneously distributed throughout the genome may cause coverage to drop out near the

497  boundary of the element. Taken together, these data reveal that *Oxytricha* TBE transposases are

498  expressed in their developing somatic nucleus, and preferentially associate with TBE-specific

499  DNA transposons.


500  **DISCUSSION**

501  By leveraging a heterologous expression strategy in *E. coli*, our study provides the first

502  insights into TBE transposase activity, revealing the enzyme's DNA sequence preferences for

503  binding and cleavage. Our heterologous expression approach addresses a critical gap in

504  understanding the DNA substrate preferences of this enzyme that participates in genome

505  rearrangement. Functional studies of the TBE transposase in *Oxytricha* have been limited by tools

506  to manipulate a high copy element in the germline genome. Our present approach overcomes

507  these limitations and provides insight into the mechanism of DNA targeting and cleavage. The

508  approach may be leveraged to study the activity of other transposases, which can be notoriously

509  difficult to purify due to their basic properties and aggregation propensities (Jaillet *et al.*, 2014),

510  limiting the ability to study their biochemical properties and activities *in vitro*.

511  After observing an unexpected DNA fragmentation pattern when we expressed the TBE

512  transposase in *E. coli*, we developed a deep sequencing approach ('Smear-seq') to investigate

513  the enzyme's genomic DNA fragmentation activity. By leveraging an orthogonal approach to

514  directly profile DNA binding sites of the catalytically inactive transposase using ChIP-seq and *de*

515  *novo* motif predictions, we discovered that the transposase preferentially binds short sequence

516  variations of the motif 5'-ANTTTGA, which is partially present within the TBE transposon ends.

517  Moreover, ChIP-seq revealed enrichment of the transposase within the native TBE end

518  sequences (TIRs) of the TBE2.1(7905) variant. We further applied our scalable approach to profile

519  DNA binding specificities for all major families of TBE transposases, revealing their ability to

520  induce DNA fragmentation and to display consistent enrichment profiles across the genome.

521  Finally, we found a moderately positive correlation between the ChIP-seq and Smear-seq signals,

522    indicating that the TBE transposase not only binds DNA at specific sites but can also induce DNA

523    fragmentation at these loci.

524          One limitation of the heterologous workflow is that it does not reconstitute native protein

525    interactions between the TBE transposase and other *Oxytricha* elements that are absent from the

526    *E. coli* expression system. We also acknowledge that our experiments primarily examined a single

527    TBE2.1 transposase variant and consensus sequences from the four transposon families, which

528    cannot represent the activity of the >34,000 specific TBE elements encoded in the MIC genome.

529    Importantly, the majority of germline-encoded transposase genes do not contain an intact catalytic

530    triad (X. Chen & Landweber, 2016), indicating that they have lost DNA cleavage activity and could

531    be pseudogenes. However, non-catalytic paralogs could also serve other functions as binding

532    partners or interactors, as in another model ciliate, *Paramecium tetraurelia* (Bischerour *et al*.,

533    2018).

534          Transposases involved in genome rearrangement have been extensively profiled in other

535    model ciliates, which also undergo large-scale genome rearrangements, albeit with reduced

536    complexity and at a smaller scale than *Oxytricha*. *Paramecium* genome rearrangement eliminates

537    >45,000 IESs and requires a domesticated PiggyMac (PGM) transposase (Arnaiz *et al*., 2012;

538    Baudry *et al*., 2009). PGM is homologous to the *piggyBac* transposase, originally isolated from a

539    cabbage looper moth, and it inserts precisely and exclusively into TTAA target sites (Fraser *et al*.,

540    1983). In *Paramecium*, PGM is expressed in the developing macronucleus shortly after mating,

541    and knockdown of PGM reduces cell viability and inhibits IES excision (Baudry *et al*., 2009). Not

542    all copies of PiggyMac are catalytically intact; five groups of catalytically-inactivated Pgm-like

543    proteins (Pgmls) interact with PiggyMac, and are required for both nuclear localization and DNA

544    elimination activity (Bischerour *et al*., 2018). Intriguingly, the DNA-targeting activity of the

545    catalytically-intact PGM is influenced by chromatin state, and the nucleosome remodeling protein

546    ISWI1 participates in IES nucleosome depletion, perhaps by increasing chromatin accessibility

547    for DNA cleavage by PGM (Singh *et al*., 2022). Although our work in *Oxytricha* indicates that the

548    TBE transposase is sequence-specific, we have yet to determine the extent to which chromatin

549    state modulates transposase recruitment or cleavage activity.

550          The genome of another well-studied ciliate, *Tetrahymena*, also encodes domesticated

551    *piggyBac*-like transposases that are implicated in DNA rearrangement. Three of these

552    transposases, Tpb1, Tpb2, and Tpb6 are also expressed shortly after mating and their knockdown

553    impairs viability and inhibits DNA elimination (Cheng *et al*., 2010, 2016). The targeting mechanism

554    of Tpb2 is based on recruitment to heterochromatin structures, as opposed to specific sequences

555    (Cheng *et al*., 2010), while Tpb1/6 are thought to require sequence recognition at the boundaries

556    of eliminated sequences (Cheng *et al*., 2016). Future work is needed to determine whether DNA

557    recruitment of the TBE transposase is dependent on heterochromatin structures in its native

558    environment; however, the short length of many *Oxytricha* IESs presents a challenge to chromatin

559    recognition (Chen *et al*. 2014).

560          The relaxed sequence specificity of the TBE transposases differs markedly from the strict

561    sequence-specific activity of most transposases that exclusively target their transposon end

562    sequences to minimize off-target cleavage activity that could compromise host fitness. Although

563    our work demonstrates that the TBE transposases do recognize their end sequences, they also

564    exhibit remarkable off-target activity, with the ability to reproducibly bind and cleave distinct

565    regions across the *E. coli* genome. Future work in *Oxytricha* is needed to confirm whether this

566    relaxed specificity is indeed critical to facilitate the extensive cleavage events that occur during

567    development of the somatic macronucleus. Together, this work reveals DNA binding and

568    cleavage activity of the TBE transposases involved in *Oxytricha* genome rearrangement, and

569    describes a facile heterologous approach to profile the activity of enzymes implicated in DNA

570    rearrangements.

**DATA AVAILABILITY**

High-throughput sequencing data and custom scripts used for analyses of high-throughput sequencing data are available at Zenodo (DOI 10.5281/zenodo.15040686). Datasets generated and analyzed in the current study are available from the corresponding authors on request.

**SUPPLEMENTARY DATA**

Supplementary Data are available online.

**ACKNOWLEDGEMENT**

**CONTRIBUTIONS**

M.W.G.W., T.A., S.H.S., and L.F.L. conceived and designed the project. M.W.G.W. and T.A. designed and constructed TBE protein expression vectors and performed DNA fragmentation assays in E. coli. The DNA fragmentation phenotype was originally identified by T.A. M.W.G.W. performed and analyzed the ChIP-seq and Smear-seq experiments and validated the TBE transposase antibody in *E. coli*. D.J.V. and T.A. designed the Smear-seq approach. E.A. performed immunofluorescence experiments and western blotting in *Oxytricha*, and prepared *Oxytricha* lysates for ChIP-seq. H.S.F. and D.J.V. assisted in the analysis of ChIP-seq data. H.L. assisted in the construction of expression vectors. M.W.G.W wrote the manuscript with assistance from S.H.S. and L.F.L and input from all authors.

**FUNDING**

599 **CONFLICT OF INTEREST**

600 M.W.G.W. is a co-founder of Can9 Bioengineering. S.H.S. is a co-founder and scientific advisor

601 to Dahlia Biosciences, a scientific advisor to CrisprBits and Prime Medicine, and an equity holder

602 in Dahlia Biosciences and CrisprBits.

**Figure 1**
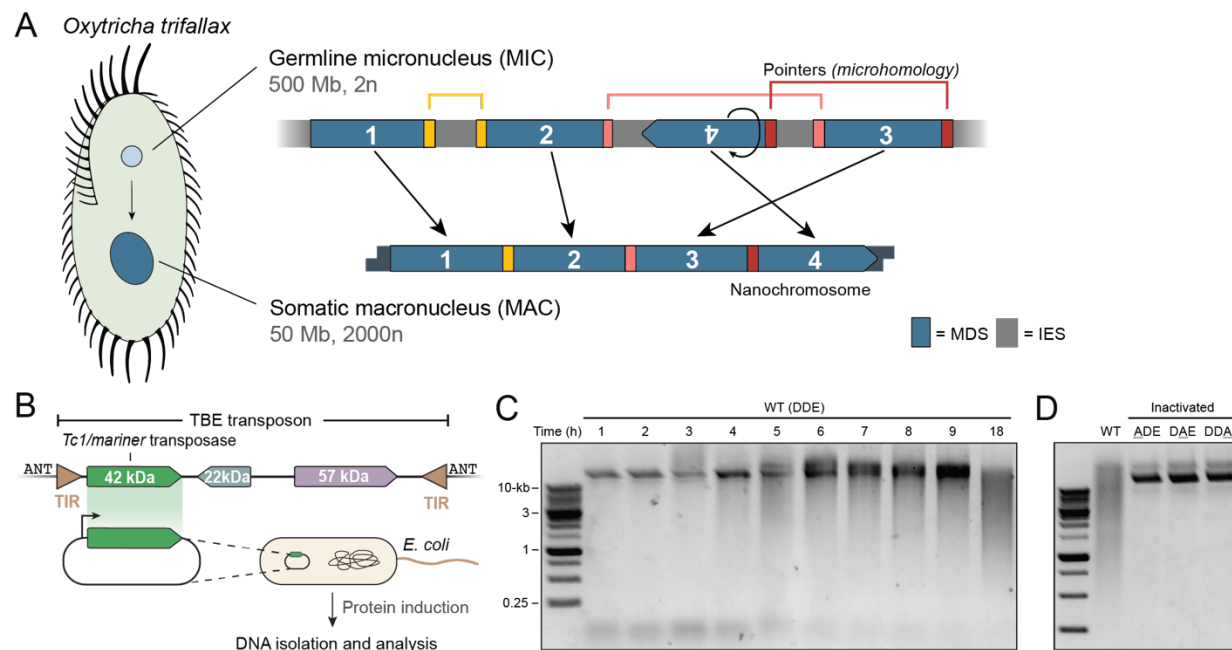


**Figure 1 | The TBE transposase induces DNA fragmentation in *E. coli.* (A)** Genome rearrangement in *Oxytricha trifallax*. *Oxytricha* cells contain a germline micronucleus (MIC), which is diploid (2n) and encodes ~500 Mb of genome content. The somatic macronucleus (MAC) is polyploid (2000n) and encodes ~50 Mb of genome content. Blue rectangles illustrate macronuclear destined sequences (MDSs), which are separated by internally eliminated sequences (IESs, gray) in the MIC. During rearrangement, IESs are eliminated and MDSs rearrange at short homologous sequences (pointers, colored bars) to form functional gene-length nanochromosomes. (**B**) Schematic of TBE transposon genetic architecture (top), and heterologous TBE transposase expression approach in *E. coli*. Terminal inverted repeats (TIRs) are indicated by brown inverted triangles, which are flanked by the ANT target-site duplication. (**C-D**) DNA fragmentation analysis upon TBE transposase induction in *E. coli*, using either the WT transposase with intact DDE motif (**C**), or catalytically inactivated transposase mutants bearing either an ADE mutation or DAE or DDA (**D**). 500 ng DNA was separated by electrophoresis on a 1% agarose gel and visualized using SYBR Safe.

619    **Figure 2**



620
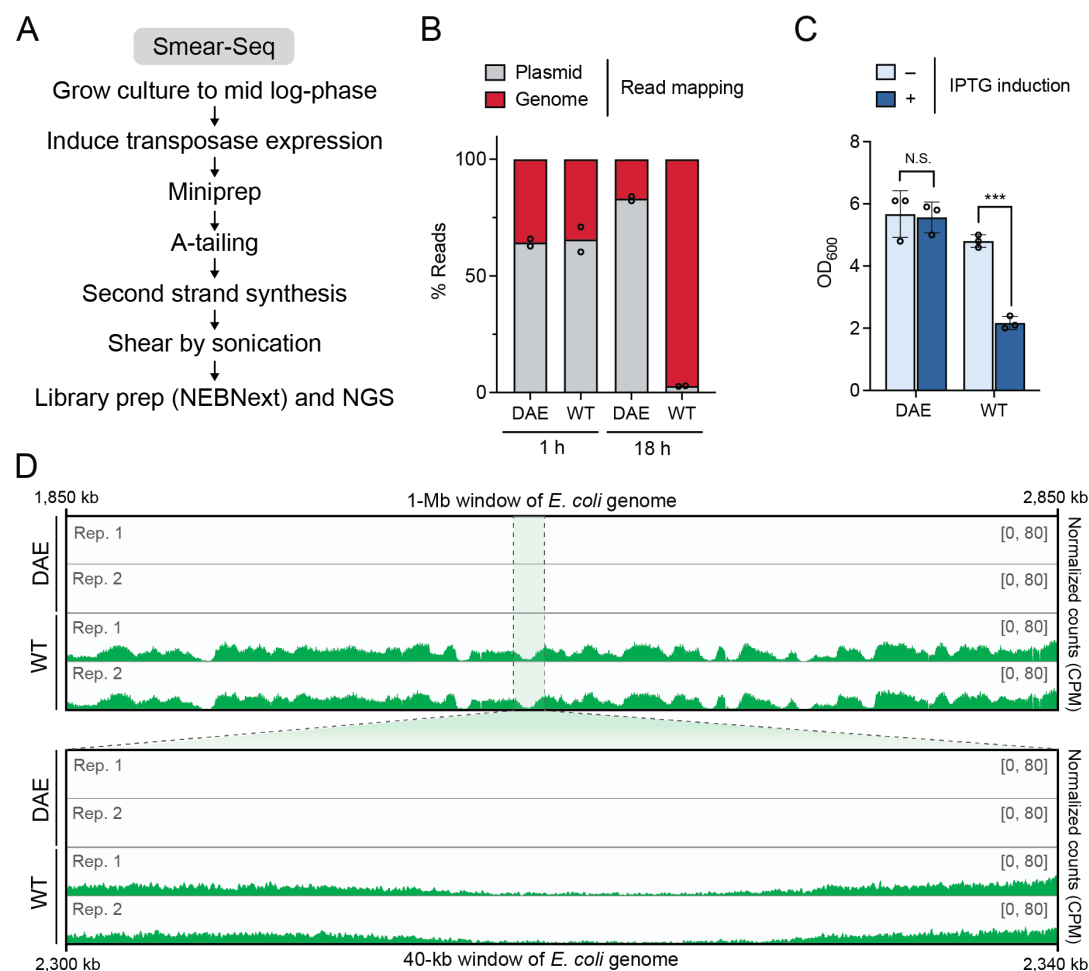
**Figure 2 | The TBE transposase cleaves genomic DNA and induces cell death.** (**A**) 'Smear-seq' workflow for sequencing fragmented DNA in *E. coli*. (**B**) Percentage of reads from each sample mapping to the transposase expression plasmid or to the *E. coli* genome. Each point represents one independent biological replicate. (**C**) Optical cell density ($OD_{600}$) of the WT or inactivated (DAE) transposase after 18 hours with and without IPTG induction. Data are shown as mean +/- SD for n=3 biological replicates; *** $p < 0.001$ (unpaired *t*-test); N.S., not significant. (**D**) Read coverage across the *E. coli* genome in a 1-Mb window (top) or 40-kb window (bottom). The y-axis represents read counts per million (CPM), with the axis limit set to 80 for each sample.

**Figure 3**
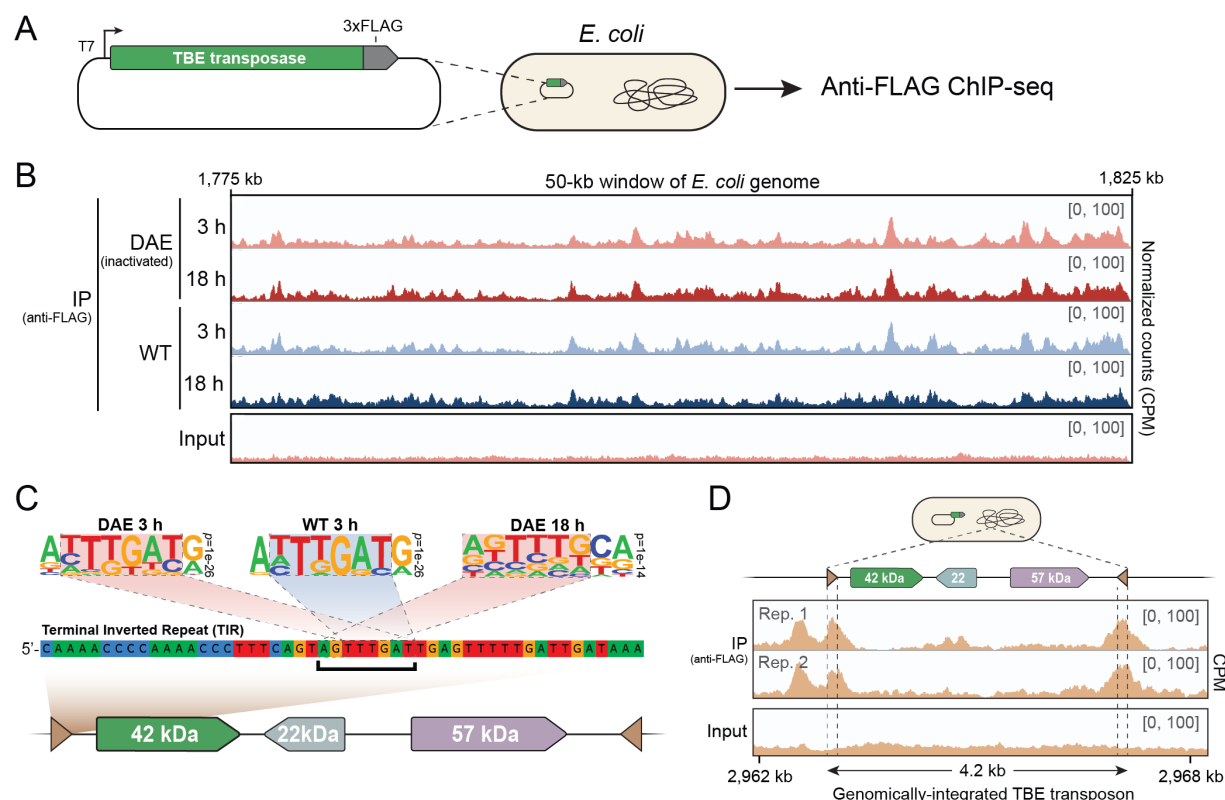


**Figure 3 | ChIP-seq reveals DNA sequence preferences for degenerate motifs partially encoded within the TBE transposon ends.** (**A**) ChIP-seq approach for determining DNA binding preferences of the TBE transposase in *E. coli.* (**B**) Read coverage across a representative 50-kb window of the *E. coli* genome, for the WT and catalytically-inactivated transposase (DAE). The y-axis represents counts per million (CPM) and the axis limit is set to 100. (**C**) Sequence logos represent the top statistically significant motif from *de novo* motif prediction (Homer; Heinz *et al.*, 2010). P-values indicate statistical significance; motifs are considered significant if their p-values are less than 1e-10. Motif prediction for the WT 18 h sample did not detect any statistically significant motifs (N.S.). Dashed boxes indicate regions that map to the TBE transposon ends, as illustrated by the first 50 nt of the TBE2.1 terminal inverted repeat (TIR). (**D**) Read coverage from ChIP-seq using a catalytically-inactive TBE2.1 transposase induced for 18 h in a strain harboring a genomically-integrated TBE2.1 transposon. Vertical dashed lines on the coverage plots indicate the TIR boundaries.

**Figure 4**
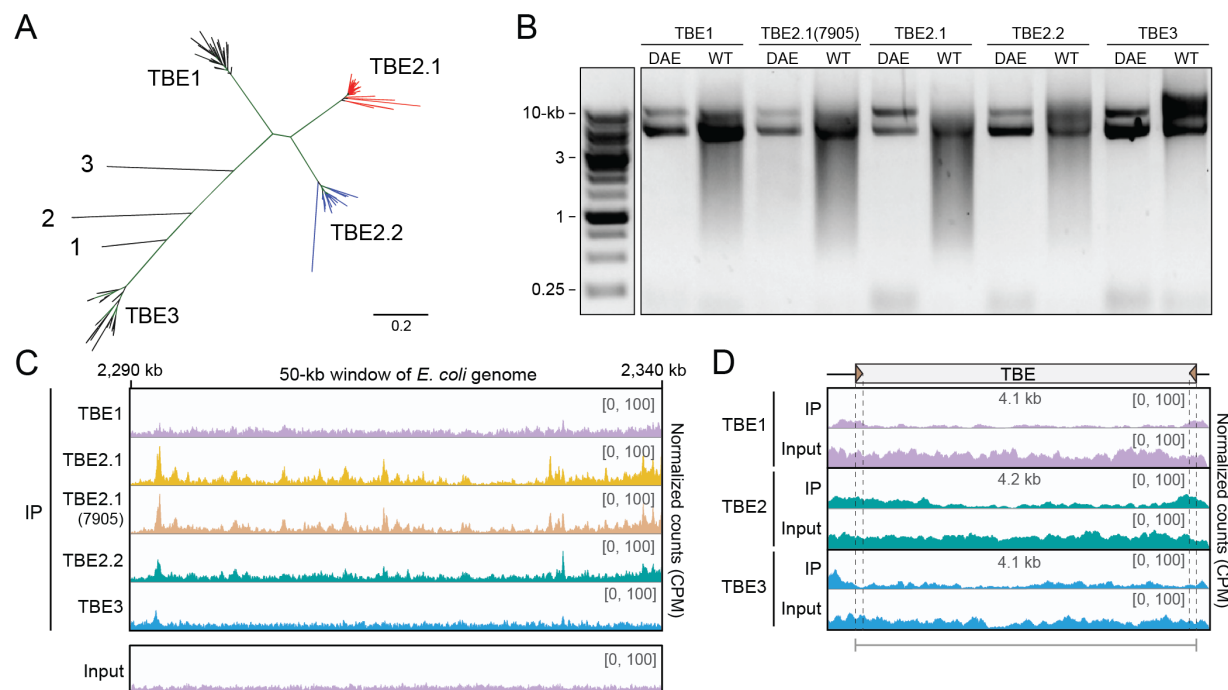


**Figure 4 | TBE transposase homologs are catalytically active and exhibit similar enrichment profiles across the *E. coli* genome.** (**A**) Unrooted Bayesian phylogenetic tree of four families of TBE elements, from (X. Chen & Landweber, 2016), built using the protein sequences of the 57 kDa protein. The scale indicates branch substitutions per site, and branches labeled 1-3 are the following TBE orthologs: 1: *Sterkiella histriomuscorum*; 2: *Tetramena sp.;* 3: *Laurentiella sp*. (**B**) DNA fragmentation assay for representative TBE transposase homologs from each of the four major families, upon heterologous expression in *E. coli*. Each sample was induced for 18 h before DNA was purified, and 500 ng DNA was separated by electrophoresis on a 1% agarose gel and visualized using SYBR Safe DNA Gel Stain. A catalytically inactivated mutant (DAE) was generated and cloned for a representative of each TBE family. For each transposase representative, only the WT transposase induces DNA fragmentation. (**C**) Read coverage across a representative 50-kb window of the *E. coli* genome from ChIP-seq using the catalytically-inactivated representative from each TBE group. The y-axis represents counts per million (CPM) and the axis limit is set to 100. (**D**) Read coverage in CPM from ChIP-seq in strains harboring a genomically-integrated TBE element corresponding to a representative from either TBE1 (from ctg7180000067530), TBE2 (from ctg7180000069065), or TBE3 (from ctg7180000089059). The y-axis represents CPM with the axis limit set to 100, and dashed lines indicate the TIR boundaries.
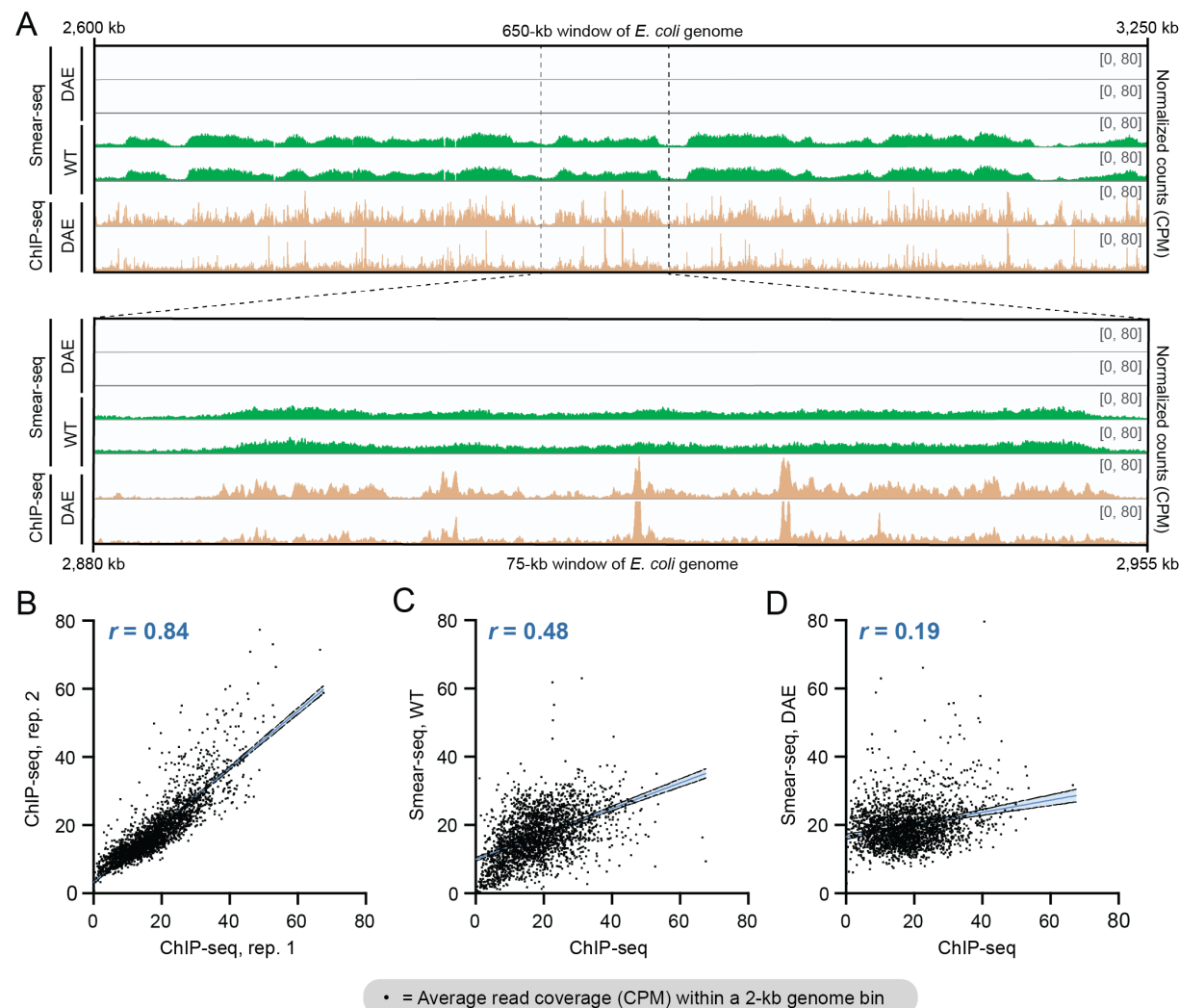
663 **Figure 5**



664

665 **Figure 5 | Transposase binding activity (ChIP-seq) correlates with cleavage activity**
666 **(Smear-seq).** (**A**) Read coverage profiles of the 18 h Smear-seq DAE and WT samples,
667 compared to the 18 h ChIP-seq DAE signal, in 650-kb (top) or 80-kb (bottom) bins across the *E.*
668 *coli* genome. The y-axis represents CPM with the axis limit set to 80. (**B-D**) Comparison between
669 average read coverage (CPM) across 2-kb bins of the *E. coli* genome. There is strong correlation
670 between ChIP-seq replicates (**B**), moderate correlation between ChIP-Seq DAE and 18 h WT
671 Smear-seq samples (**C**), and no correlation between the ChIP-seq DAE and the 18 h DAE Smear-
672 seq signal (**D**). *r* values represent the Pearson correlation, and linear regressions are shown with
673 95% confidence bands of the best-fit line.

674    **Figure 6**



<span style="text-align:center">= Developing somatic nucleus</span>

675
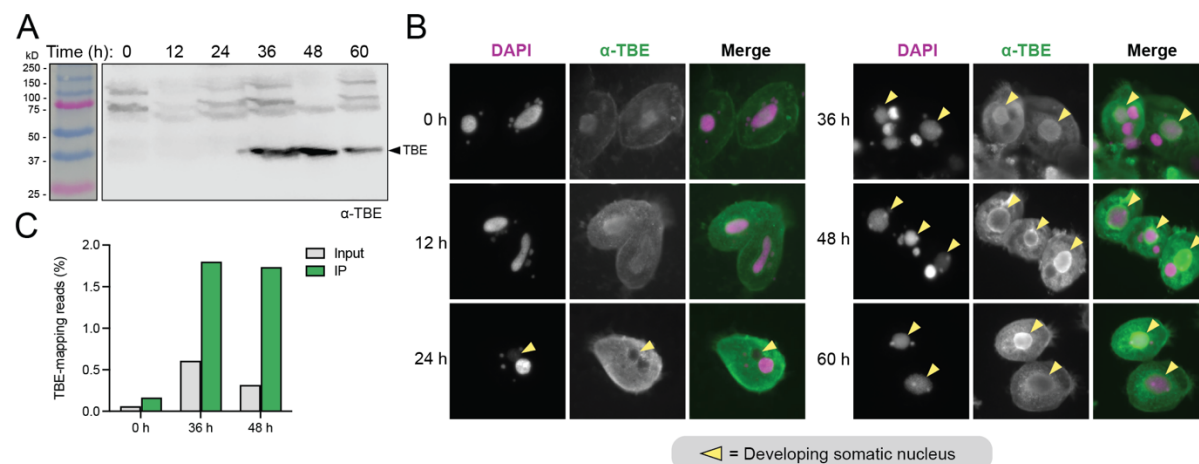
676    **Figure 6 | TBE transposase expression, localization, and transposon-DNA preference**
677    **during *Oxytricha* development.** (**A**) Western blot analysis using an anti-TBE transposase
678    antibody against *Oxytricha* cell lysates collected at 12-hour intervals post mixing of compatible
679    mating types. (**B**) DAPI staining of DNA and anti-TBE transposase immunostaining of *Oxytricha*
680    cells at 12-hour intervals post mixing. Arrowheads indicate developing somatic macronuclei. Note
681    that until 48 h, DAPI stains the parental macronuclei more strongly than the new macronuclei,
682    due to the copy number difference. (**C**) Percent of micronuclear genome-mapping ChIP-seq reads
683    that map to TBE2.1(7905) in the input and IP samples, at 0 h and two timepoints after mixing.

684    **SUPPLEMENTARY FIGURES**
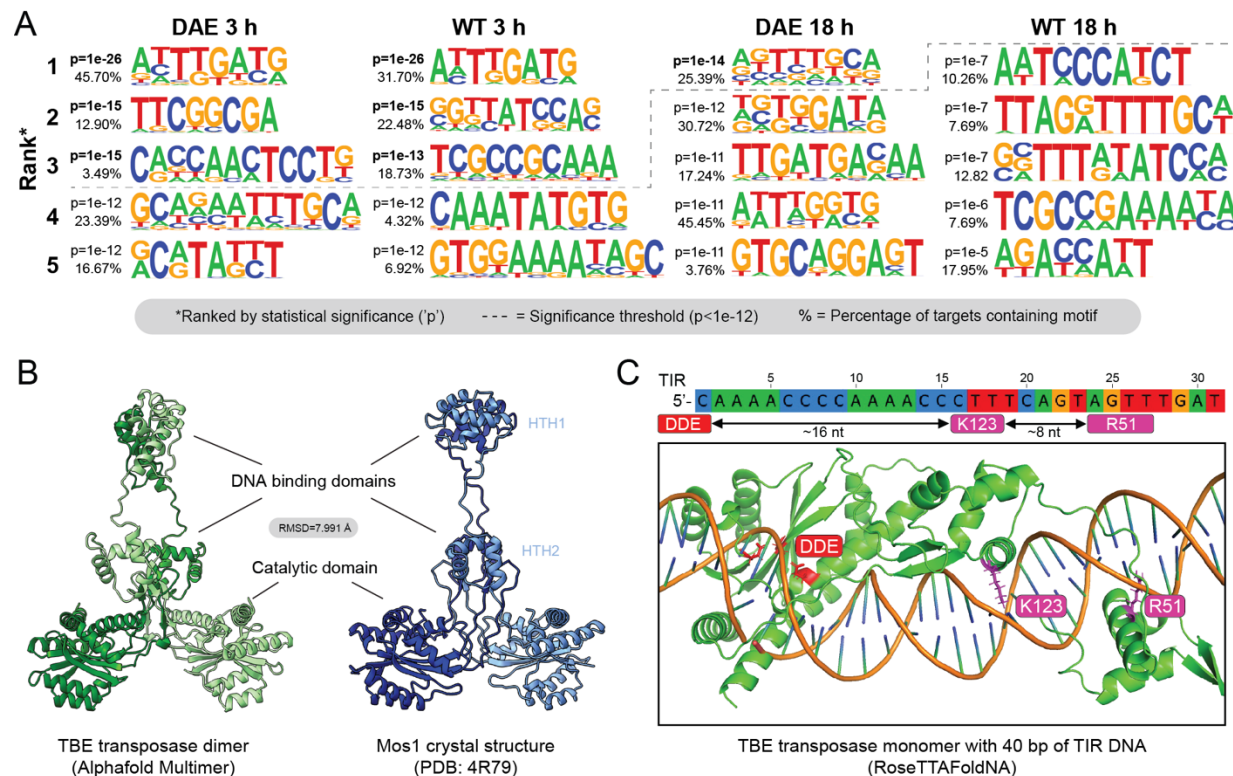


685

686    **Figure S1 |** *De novo* **motif analysis and structural prediction of the TBE transposase places**
687    **putative DNA binding domains near preferred motifs within transposon ends.** (**A**) Sequence
688    logos identified from *de novo* motif prediction (Homer; Heinz *et al.*, 2010) within peaks of ChIP-
689    seq read enrichment for the catalytically-inactivated (DAE) and WT transposase. Sequence logos
690    are ranked by statistical significance, and the top five most significant motifs from each sample
691    are shown. The dashed line indicates the cutoff for statistical significance (p-value<1e-12). The
692    percentage of ChIP-seq peaks containing each motif is listed to the left of each motif (%). (**B**)
693    AlphaFold-Multimer (Evans *et al.*, 2022) prediction of the TBE2.1(7905) transposase dimer (left).
694    Annotated DNA binding domains are based on structural homology to the Mos1 crystal structure
695    (right). DNA-binding helix-turn-helix (HTH) and catalytic domains are indicated.    (**C**)
696    RoseTTAFoldNA (Baek *et al.*, 2023) prediction of a TBE transposase monomer interacting with
697    TBE TIR DNA. The TIR DNA schematic (top) illustrates the predicted interaction sites of the DNA
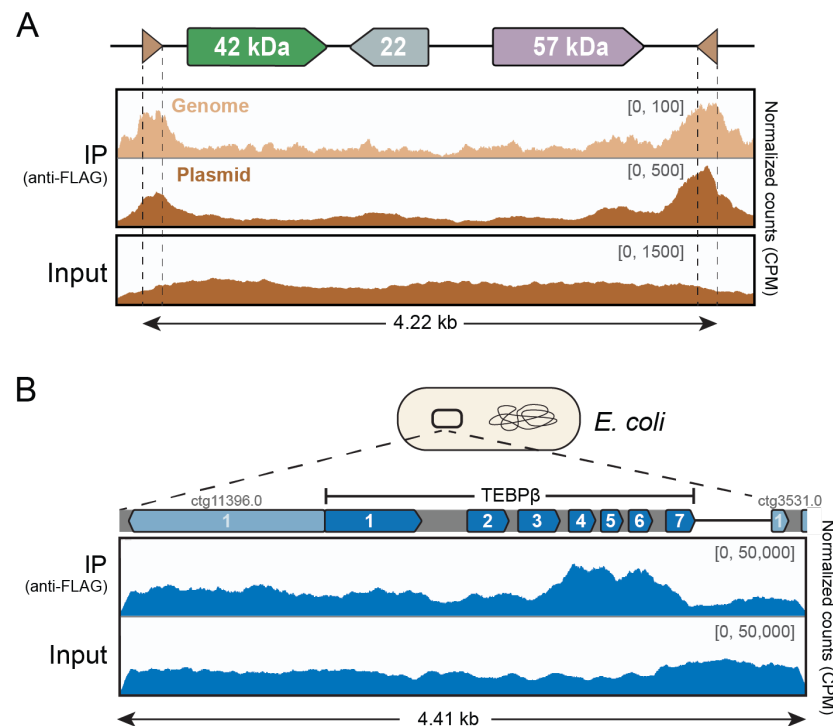698    binding domain residues (K123 and R51) based on the RoseTTAFoldNA model.

699 **Figure S2**



700

701 **Figure S2 | The TBE transposase does not exhibit a consistent pattern of enrichment**
702 **across *Oxytricha* MIC genomic DNA in *E. coli*.** (**A**) Read coverage from ChIP-seq of the TBE
703 transposase using either a genomically-integrated ("genome") or a plasmid-encoded complete
704 TBE transposon ("plasmid") in *E. coli*. The y-axis represents CPM and is set to the following
705 maximum for each sample: 100 CPM for the genome IP track; 500 CPM for the plasmid IP track;
706 and 1500 CPM for the plasmid input track. (**B**) ChIP-seq read coverage from the FZZ-tagged TBE
707 transposase expressed in *E. coli* cells containing a cloned *Oxytricha* MIC genomic region on a
708 plasmid containing the TEBPβ MIC locus and flanking regions. The y-axis represents CPM and
709 is set to 50,000 for both the IP and input samples. Dark blue boxes within the TEBPβ MIC locus
710 depict its 7 MDSs, gray boxes represent IESs, and pale blue boxes are MDSs for flanking loci;
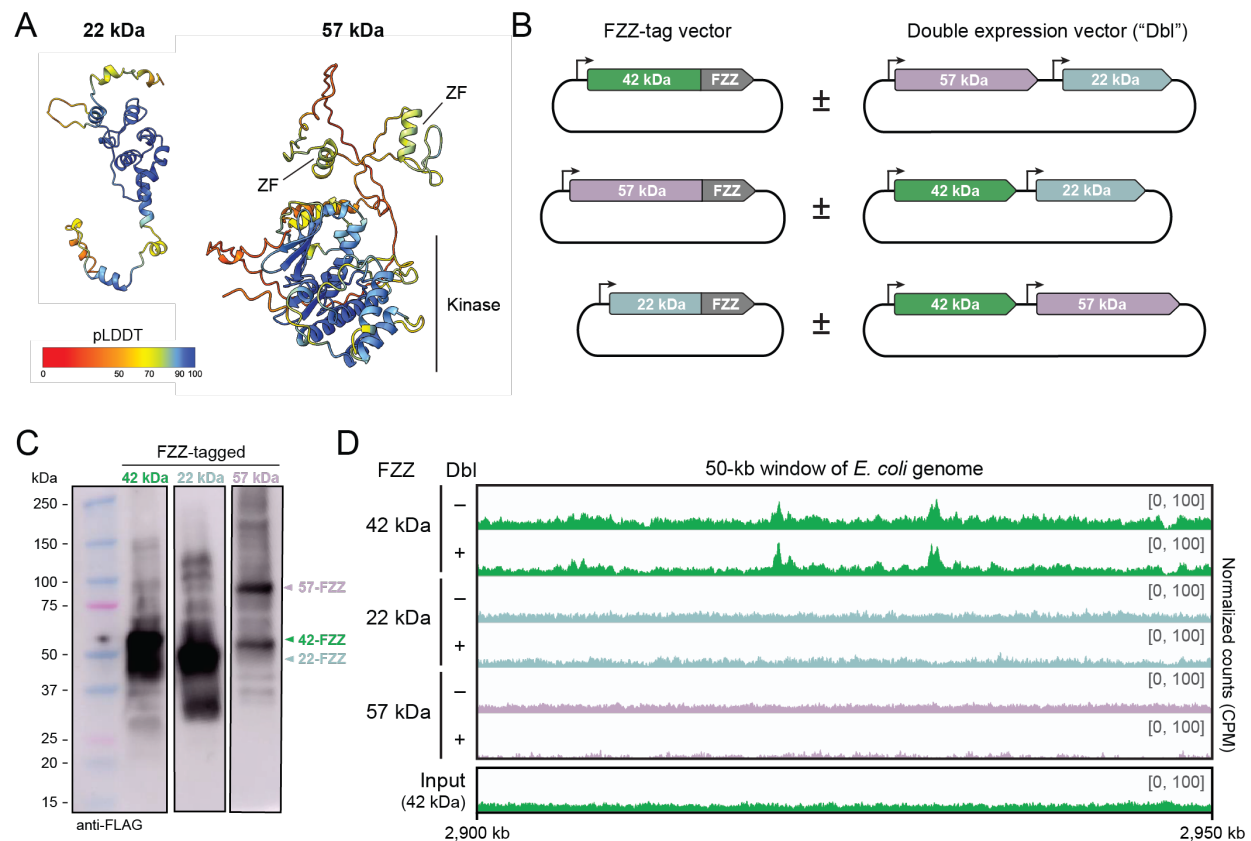711 figure approximately to scale.

712 **Figure S3**



713

714 **Figure S3 | Investigating the accessory proteins encoded within TBE transposons.**
715 (**A**) AlphaFold predictions of the 22-kDa and 57-kDa ORF, colored by pLDDT (predicted local
716 distance difference test) values. Predicted zinc finger ('ZF') and kinase domains from HHpred
717 (Söding *et al*., 2005) are annotated on the 57-kDa structure; the 22-kDa protein exhibits no high-
718 confidence domain predictions via HHpred. (**B**) Schematic of expression vector design for ChIP-
719 seq experiments in D. In each experiment (horizontal rows in D), strains contained pairwise
720 combinations of FZZ-tagged single-protein vectors in the presence and absence of untagged
721 double protein expression vectors ('Dbl'). (**C**) Western blot with an anti-FLAG antibody used
722 against each of the three FZZ-tagged expression vectors in *E. coli*. Expected sizes for the FZZ-
723 fusion proteins are 77 kDa (57-kDa ORF), 60 kDa (42-kDa ORF), and 42 kDa (22-kDa ORF). (**D**)
724 Read coverage across a representative 60-kb window of the *E. coli* genome for the FZZ-tagged
725 expression vectors, in the presence (+) and absence (-) of the corresponding untagged double-
726 expression vectors ('Dbl'). The y-axis represents counts per million (CPM), and the axis limit is
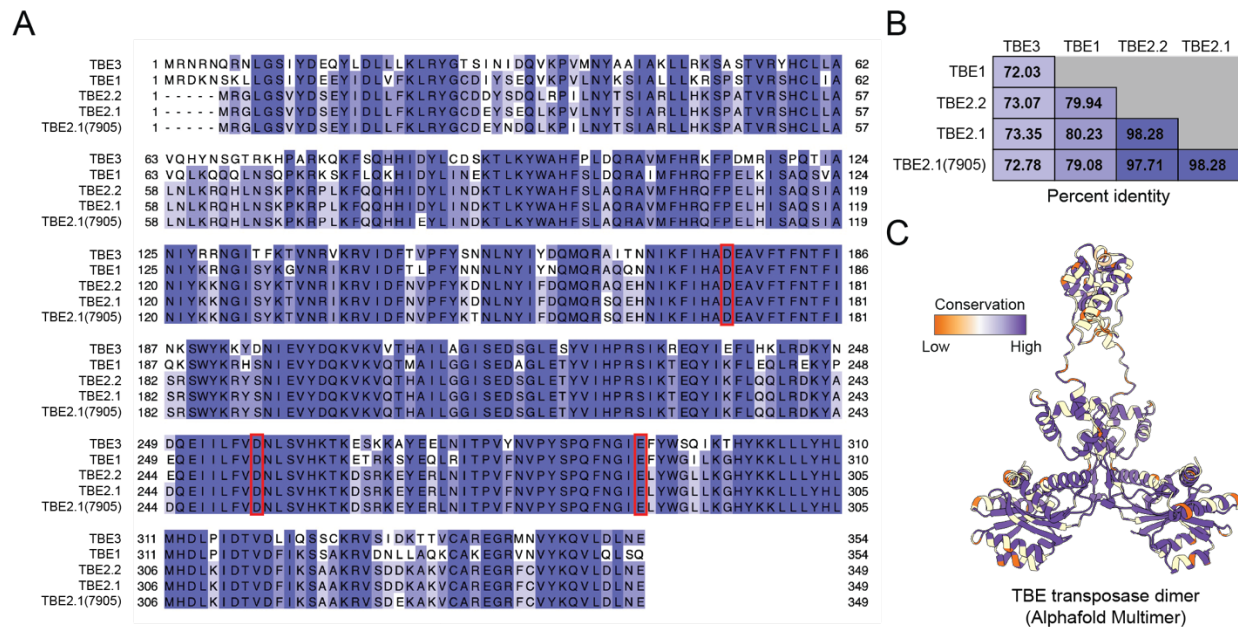727 set to 100.

728    **Figure S4**



729

730    **Figure S4 | Alignment and amino acid similarity across TBE transposase families.**

731    (**A**) Multiple sequence alignment of the four TBE family transposase consensus sequences (Chen

732    *et al.* 2016), alongside representative TBE2.1 family member (7905). Red boxes indicate residues

733    belonging to the DDE catalytic triad. (**B**) Sequence identity matrix of the percent amino acid

734    identity for the four TBE family consensus sequences and TBE2.1(7905). (**C**) TBE transposase

735    dimer structural prediction colored by sequence conservation across the TBE families.
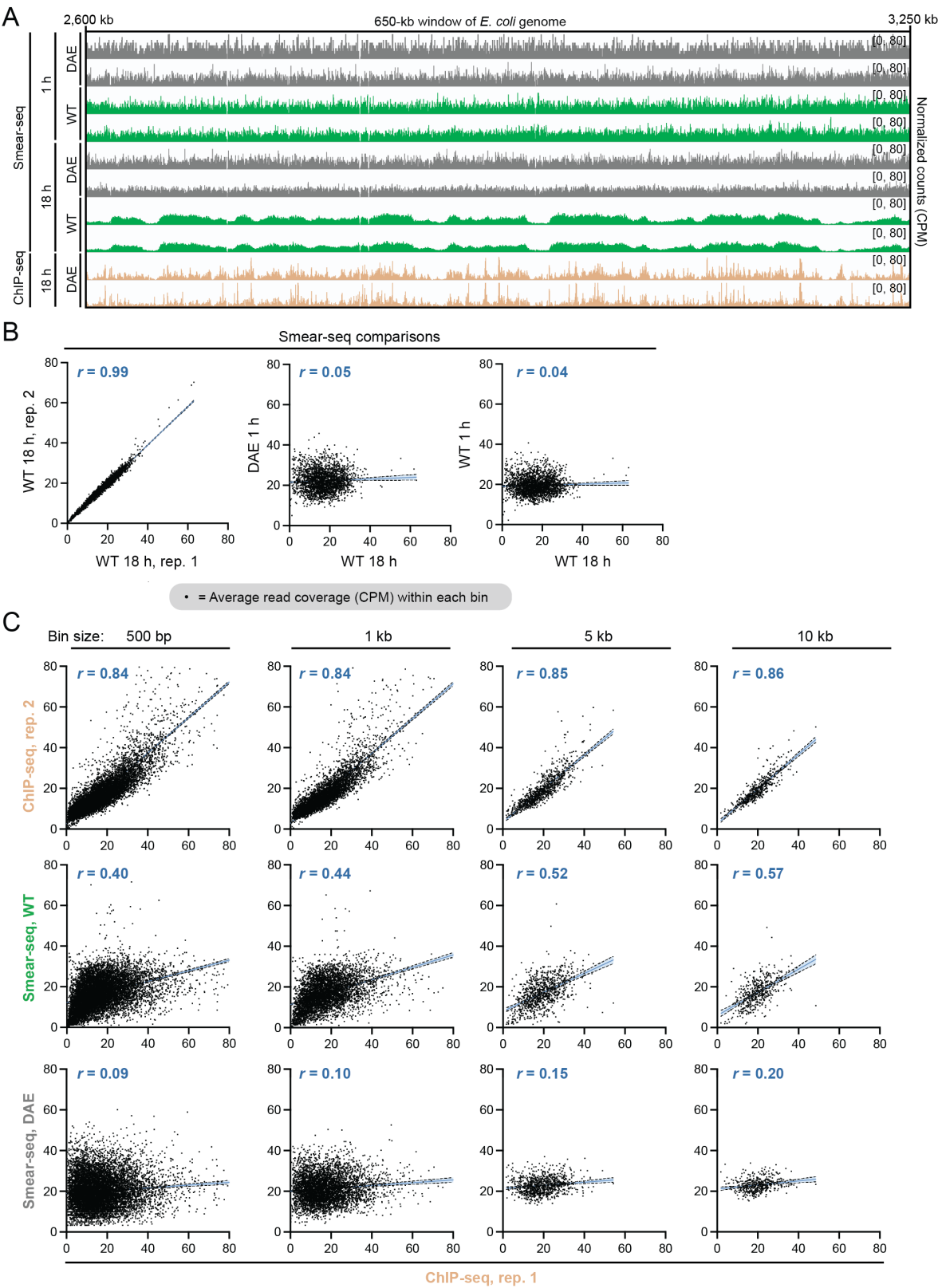
736     **Figure S5**



737

738 **Figure S5** (previous page) **| Comparison between ChIP-seq and Smear-seq datasets.**

739 (**A**) Read coverage profiles of all Smear-seq samples and the 18 h ChIP-seq DAE sample, in a

740 650-kb window of the *E. coli* genome. The y-axis represents CPM with the axis limit set to 80. (**B**)

741 Comparisons between average read coverage in CPM across 2-kb bins of the genome. The

742 Smear-seq signal at 18 h is strongly correlated across independent biological replicates. (**C**)

743 Coverage comparisons when bin sizes are varied between 500 bp and 10 kb. *r* values represent

744 the Pearson correlation, and linear regressions are shown with 95% confidence bands of the best-

745 fit line.

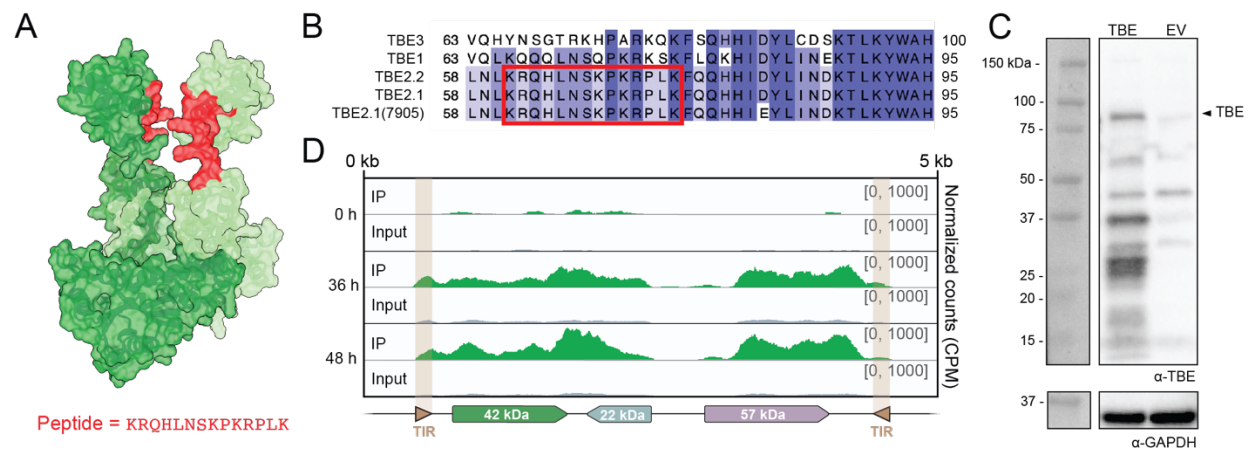**Figure S6**



Peptide = KRQHLNSKPKRPLK

**Figure S6 | Generating and validating an anti-TBE transposase antibody.** (**A**) Structural prediction of the TBE2.1 transposase dimer by AlphaFold-Multimer (Evans *et al.*, 2022); red coloring indicates the predicted surface-exposed region that was selected for peptide synthesis and antibody generation. (**B**) Partial multiple sequence alignment of the four TBE-family transposase consensus sequences (from **Figure S4A**). Red box indicates the 14 residues selected for peptide synthesis. (**C**) Western hybridization to the anti-TBE transposase antibody (top) and anti-GAPDH antibody loading control (bottom). The two samples are *E. coli* BL12(DE3) cells expressing either the MBP-tagged 42-kDa plasmid (TBE) or an empty vector control (EV). The expected size of the MBP-tagged 42-kDa ORF is 88.6 kDa. (**D**) Read coverage mapped across the representative TBE2.1(7905) element demonstrates transposon localization, concentrated at genic regions, at 36 and 48 h. The y-axis represents counts per million (CPM), and the axis limit is set to 1000.

## REFERENCES

Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.-M., Wilkes, C. D., Garnier, O., Labadie, K., Lauderdale, B. E., Mouël, A. L., Marmignon, A., Nowacki, M., Poulain, J., Prajer, M., Wincker, P., Meyer, E., Duharcourt, S., Duret, L., Bétermier, M. & Sperling, L. (2012). The *Paramecium* Germline Genome Provides a Niche for Intragenic Parasitic DNA: Evolutionary Dynamics of Internal Eliminated Sequences. *PLoS Genetics*, *8*(10), e1002984. https://doi.org/10.1371/journal.pgen.1002984

Baek, M., McHugh, R., Anishchenko, I., Jiang, H., Baker, D. & DiMaio, F. (2023). Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nature Methods*, 1–5. https://doi.org/10.1038/s41592-023-02086-5

Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E. & Bétermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes & Development*, *23*(21), 2478–2483. https://doi.org/10.1101/gad.547309

Bischerour, J., Bhullar, S., Wilkes, C. D., Régnier, V., Mathy, N., Dubois, E., Singh, A., Swart, E., Arnaiz, O., Sperling, L., Nowacki, M. & Bétermier, M. (2018). Six domesticated PiggyBac transposases together carry out programmed DNA elimination in *Paramecium*. *ELife*, *7*, e37927. https://doi.org/10.7554/elife.37927

Bonocora, R. P. & Wade, J. T. (2015). Bacterial Transcriptional Control, Methods and Protocols. *Methods in Molecular Biology*, *1276*, 327–340. https://doi.org/10.1007/978-1-4939-2392-2_20

Bryan, G., Garza, D. & Hartl, D. (1990). Insertion and excision of the transposable element *mariner* in Drosophila. *125*(1), 103–114. https://doi.org/10.1093/genetics/125.1.103

Bryan, G. J., Jacobson, J. W. & Hartl, D. L. (1987). Heritable somatic excision of a *Drosophila* transposon. *Science* (New York, N.Y.), *235*, 1636–1638. https://doi.org/10.1126/science.3029874

Chen, S., Zhou, Y., Chen, Y. & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen, X., Bracht, J. R., Goldman, A. D., Dolzhenko, E., Clay, D. M., Swart, E. C., Perlman, D. H., Doak, T. G., Stuart, A., Amemiya, C. T., Sebra, R. P. & Landweber, L. F. (2014). The Architecture of a Scrambled Genome Reveals Massive Levels of Genomic Rearrangement during Development. *Cell*, *158*(5), 1187–1198. https://doi.org/10.1016/j.cell.2014.07.034

Chen, X. & Landweber, L. F. (2016). Phylogenomic analysis reveals genome-wide purifying selection on TBE transposons in the ciliate *Oxytricha*. *Mobile DNA*, *7*(1), 2. https://doi.org/10.1186/s13100-016-0057-9

Clay, D. M., Kim, H. & Landweber, L. F. (2019). Transformation with Artificial Chromosomes in *Oxytricha trifallax* and Their Applications. *G3: Genes, Genomes, Genetics*, *9*(10), 3119–3127. https://doi.org/10.1534/g3.119.400298

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M. & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Doak, T. G., Doerder, F. P., Jahn, C. L. & Herrick, G. (1994). A proposed superfamily of transposase genes: Transposon-like elements in ciliated protozoa and a common "D35E" motif. *Proceedings of the National Academy of Sciences*, *91*(3), 942–946. https://doi.org/10.1073/pnas.91.3.942

Dupeyron, M., Baril, T., Bass, C. & Hayward, A. (2020). Phylogenetic analysis of the *Tc1/mariner* superfamily reveals the unexplored diversity of *pogo*-like elements. *Mobile DNA*, *11*(1), 21. https://doi.org/10.1186/s13100-020-00212-0

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., … Hassabis, D. (2022). Protein complex prediction with AlphaFold-Multimer. *BioRxiv*, 2021.10.04.463034. https://doi.org/10.1101/2021.10.04.463034

Fang, W., Wang, X., Bracht, J. R., Nowacki, M. & Landweber, L. F. (2012). Piwi-Interacting RNAs Protect DNA against Loss during *Oxytricha* Genome Rearrangement. *Cell*, *151*(6), 1243–1255. https://doi.org/10.1016/j.cell.2012.10.045

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. & Glass, C. K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime *cis*-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, *38*(4), 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Hickman, A. B. & Dyda, F. (2016). DNA Transposition at Work. *Chemical Reviews*, *116*(20), 12758–12784. https://doi.org/10.1021/acs.chemrev.6b00003

Hoffmann, F. T., Kim, M., Beh, L. Y., Wang, J., Vo, P. L. H., Gelsinger, D. R., George, J. T., Acree, C., Mohabir, J. T., Fernández, I. S. & Sternberg, S. H. (2022). Selective TnsC recruitment enhances the fidelity of RNA-guided transposition. *Nature*, *609*(7926), 384–393. https://doi.org/10.1038/s41586-022-05059-4

828 Jaillet, J., Dussaussois-Montagne, A., Renault, S. & Augé-Gouillou, C. (2014). Expression and
829     Purification of the Eukaryotic MBP-MOS1 Transposase from Sf21 Insect Cells. *Bio-Protocol*,
830     *4*(20), e1262. https://doi.org/10.21769/bioprotoc.1262

831 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool,
832     K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A.
833     J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021).
834     Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589.
835     https://doi.org/10.1038/s41586-021-03819-2

836 Kataoka, K., Schoeberl, U. E. & Mochizuki, K. (2010). Modules for C-terminal epitope tagging
837     of *Tetrahymena* genes. *Journal of Microbiological Methods*, *82*(3), 342–346.
838     https://doi.org/10.1016/j.mimet.2010.07.009

839 Klobutcher, L. A. & Herrick, G. (1997). Developmental genome reorganization in ciliated
840     protozoa: the transposon link. *Progress in Nucleic Acid Research and Molecular Biology*, *56*,
841     1–62. https://doi.org/10.1016/s0079-6603(08)61001-6

842 Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
843     *Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

844 Lindblad, K. A., Pathmanathan, J. S., Moreira, S., Bracht, J. R., Sebra, R. P., Hutton, E. R. &
845     Landweber, L. F. (2019). Capture of complete ciliate chromosomes in single sequencing
846     reads reveals widespread chromosome isoforms. *BMC Genomics*, *20*(1), 1037.
847     https://doi.org/10.1186/s12864-019-6189-9

848 Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N.,
849     Kolesnikov, A. & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-
850     EBI in 2022. *Nucleic Acids Research*, *50*(W1), W276–W279.
851     https://doi.org/10.1093/nar/gkac240

852 Nowacki, M., Higgins, B. P., Maquilan, G. M., Swart, E. C., Doak, T. G. & Landweber, L. F.
853     (2009). A Functional Role for Transposases in a Large Eukaryotic Genome. *Science*,
854     *324*(5929), 935–938. https://doi.org/10.1126/science.1170023

855 Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T. G. & Landweber, L. F. (2008).
856     RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature*,
857     *451*(7175), 153–158. https://doi.org/10.1038/nature06452

858 Park, P. J. (2009). ChIP–seq: advantages and challenges of a maturing technology. *Nature*
859     *Reviews Genetics*, *10*(10), 669–680. https://doi.org/10.1038/nrg2641

860 Prescott, D. M. (1994). The DNA of ciliated protozoa. *Microbiological Reviews*, *58*(2), 233–267.
861     https://doi.org/10.1128/mr.58.2.233-267.1994

862 Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S.,
863     Dündar, F. & Manke, T. (2016). deepTools2: a next generation web server for deep-
864     sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165.
865     https://doi.org/10.1093/nar/gkw257

866 Richardson, J. M., Colloms, S. D., Finnegan, D. J. & Walkinshaw, M. D. (2009). Molecular
867     Architecture of the Mos1 Paired-End Complex: The Structural Basis of DNA Transposition in
868     a Eukaryote. *Cell*, *138*(6), 1096–1108. https://doi.org/10.1016/j.cell.2009.07.012

869 Singh, A., Maurer-Alcalá, X. X., Solberg, T., Häußermann, L., Gisler, S., Ignarski, M., Swart, E.
870     C. & Nowacki, M. (2022). Chromatin remodeling is required for sRNA-guided DNA
871     elimination in *Paramecium*. *The EMBO Journal*, *41*(22), EMBJ2022111839.
872     https://doi.org/10.15252/embj.2022111839

873 Smith, J. J., Timoshevskiy, V. A. & Saraceno, C. (2020). Programmed DNA Elimination in
874     Vertebrates. *Annual Review of Animal Biosciences*, *9*(1), 1–29.
875     https://doi.org/10.1146/annurev-animal-061220-023220

876 Söding, J., Biegert, A. & Lupas, A. N. (2005). The HHpred interactive server for protein
877     homology detection and structure prediction. *Nucleic Acids Research*, *33*(suppl_2), W244–
878     W248. https://doi.org/10.1093/nar/gki408

879 Swart, E. C., Bracht, J. R., Magrini, V., Minx, P., Chen, X., Zhou, Y., Khurana, J. S., Goldman,
880     A. D., Nowacki, M., Schotanus, K., Jung, S., Fulton, R. S., Ly, A., McGrath, S., Haub, K.,
881     Wiggins, J. L., Storton, D., Matese, J. C., Parsons, L., … Landweber, L. F. (2013). The
882     *Oxytricha trifallax* Macronuclear Genome: A Complex Eukaryotic Genome with 16,000 Tiny
883     Chromosomes. *PLoS Biology*, *11*(1), e1001473. https://doi.org/10.1371/journal.pbio.1001473

884 Walker, M. W. G., Klompe, S. E., Zhang, D. J. & Sternberg, S. H. (2023). Novel molecular
885     requirements for CRISPR RNA-guided transposition. *Nucleic Acids Research*.
886     https://doi.org/10.1093/nar/gkad270

887 Wang, J. & Davis, R. E. (2014). Programmed DNA elimination in multicellular organisms.
888     *Current Opinion in Genetics & Development*, *27*, 26–34.
889     https://doi.org/10.1016/j.gde.2014.03.012

890 Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. (2009). Jalview
891     Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*,
892     *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

893 Williams, K., Doak, T. G. & Herrick, G. (1993). Developmental precise excision of *Oxytricha*
894     *trifallax* telomere-bearing elements and formation of circles closed by a copy of the flanking
895     target duplication. https://www.embopress.org/doi/epdf/10.1002/j.1460-2075.1993.tb06148.x

896 Yerlici, V. T., Lu, M. W., Hoge, C. R., Miller, R. V., Neme, R., Khurana, J. S., Bracht, J. R. &
897     Landweber, L. F. (2019). Programmed genome rearrangements in *Oxytricha* produce

898      transcriptionally active extrachromosomal circular DNA. *Nucleic Acids Research*, *47*(18),

899      9741–9760. https://doi.org/10.1093/nar/gkz725

900  Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C.,

901      Myers, R. M., Brown, M., Li, W. & Liu, X. S. (2008). Model-based Analysis of ChIP-Seq

902      (MACS). *Genome Biology*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137