

PiggyBac sequences documentation

(c) Hyung-song Nam, PhD, June 4, 2022

Pertains to Addgene plasmids

#133885 pBluescript-PB

#133886 p15A-PB

#159093 PB-CA retrofitted

#177151 pBluescript-minimal PB

I. PiggyBac transposon sequence

NCBI accession number J04364

<https://www.ncbi.nlm.nih.gov/nuccore/J04364>

See Solodushko et al., 2013 <https://doi.org/10.1038/gt.2013.52> for nomenclature, i.e., Terminal Repeats, Internal Domain, Terminal Domain

♦

Hyperactive piggyBac transposase, hyPB

Yusa et al., 2011

<https://doi.org/10.1073/pnas.1008322108>

Mutations to the piggyBac transposase sequence:

I30V

S103P

G165S

M282V

S509G

N538K

N571S - 571, not 570 as in the paper above

★

Excision only piggyBac transposase, exPB

Li et al., 2013

<https://doi.org/10.1073/pnas.1305987110>

Mutations in addition to the hyPB mutations:

R372A

K375A

D450N

go to next page

II. piggyBac Terminal Domains alignments

First line piggyBac transposon sequence

Second line pBS-PB Terminal Domains sequences

Third line PB-CA Terminal Domains sequences

piggyBac transposase ORF in **RED**

Ends of the Terminal Domain sequences in **BLACK**

Fri Jun 03, 2022 23:24 MDT

piggyBac transposon.gb from 1 to 2472

Alignment to

pBS-PB transposon sequence.gb-- Matches:682; Mismatches:8; Gaps:1782;

Unattempted:0

PB-CA transposon sequence.gb-- Matches:793; Mismatches:9; Gaps:1670;

Unattempted:0

```

      *           *           *           *           *
1>ccctagaaagatagtctgcgtaaaattgacgcatgcattcttgaaatatt>50
1>ccctagaaagatagtctgcgtaaaattgacgcatgcattcttgaaatatt>50
1>ccctagaaagatagtctgcgtaaaattgacgcatgcattcttgaaatatt>50

      *           *           *           *           *
51>gctctctcttttctaaatagcgcggaatccgctcgctgtgcatttaggacatc>100
51>gctctctcttttctaaatagcgcggaatccgctcgctgtgcatttaggacatc>100
51>gctctctcttttctaaatagcgcggaatccgctcgctgtgcatttaggacatc>100

      *           *           *           *           *
101>tcagtcgccgcttggagctcccgtgaggcgtgcttgtcaatgcggtaagt>150
101>tcagtcgccgcttggagctcccgtgaggcgtgcttgtcaatgcggtaagt>150
101>tcagtcgccgcttggagctcccgtgaggcgtgcttgtcaatgcggtaagt>150

      *           *           *           *           *
151>gtcactgattttgaactataacgaccgctgagtcaaaatgacgcatgat>200
151>gtcactgattttgaactataacgaccgctgagtcaaaatgacgcatgat>200
151>gtcactgattttgaactataacgaccgctgagtcaaaatgacgcatgat>200

      *           *           *           *           *
201>tatcttttacgtgacttttaagatttaactcatacgataattatattggt>250
201>tatcttttacgtgacttttaagatttaactcatacgataattatattggt>250
201>tatcttttacgtgacttttaagatttaactcatacgataattatattggt>250

      *           *           *           *           *
251>atttcatgttctacttacgtgataaacttattatatatatatattttcttggt>300
251>atttcatgttctacttacgtgataaacttattatatatatatattttcttggt>300
251>atttcatgttctacttacgtgataaacttattatatatatatattttcttggt>300

      *           *           *           *           *

```

301>atagatatcgtgactaatatataataaaa**atgggtagttcttttagacgatg**>350
301>atagatatcgt**gactaatatataataaaa**-----n-n-n>332
301>atagatatcgtgactaatatataataaaaatgggtagttcttttagacgatg>350

* * * * *
351>**agcatatcctctctgctcttctgcaaagcgatgacgagcttggttggtgag**>400
333>-n-n-n-n----->336
351>agcatatcctctctgctcttctgcaa**agcgatgacgagcttggttggtg**---->396

* * * * *
401>**gattctgacagtgaatatcagatcacgtaagtgaagatgacgtccagag**>450
336>----->336
397>-----n-n-n-n-n-n-n----->404

* * * * *
451>**cgatacagaagaagcgtttatagatgaggtacatgaagtgcagccaacgt**>500
336>----->336
404>----->404

* * * * *
501>**caagcggtagtgaaatattagacgaacaaaatgttattgaacaaccaggt**>550
336>----->336
404>----->404

* * * * *
551>**tcttcattggcttctaacagaatcttgaccttgccacagaggactattag**>600
336>----->336
404>----->404

* * * * *
601>**aggtgaagaataaacattggttggtcaacttcaaagtccacgagggcgtagcc**>650
336>----->336
404>----->404

* * * * *
651>**gagtcctctgcactgaacattgttcagatctcaaagaggtccgacgcgtatg**>700
336>----->336
404>----->404

* * * * *
701>**tgccgcgaatatatatgaccacttttatgcttcaaactatTTTTTactga**>750
336>----->336
404>----->404

* * * * *

751>tgagataatttcggaaattgtaaaatggacaaatgctgagatatcattga>800
336>----->336
404>----->404

* * * * *
801>aacgtcgggaatctatgacaggtgctacatttcgtgacacgaatgaagat>850
336>----->336
404>----->404

* * * * *
851>gaaatctatgccttcttttggtattctggtaatgacagcagtgagaaaaga>900
336>----->336
404>----->404

* * * * *
901>taaccacatgtccacagatgacctctttgatcgatctttgtcaatggtgt>950
336>----->336
404>----->404

* * * * *
951>acgtctctgtaatgagtcgtgatcgttttgatttttgatacgatgtctt>1000
336>----->336
404>----->404

* * * * *
1001>agaatggatgacaaaagtatacggccacacttcgagaaaacgatgtatt>1050
336>----->336
404>----->404

* * * * *
1051>tactcctgttagaaaaatatgggatctctttatccatcagtcatacaaa>1100
336>----->336
404>----->404

* * * * *
1101>attacactccaggggctcatttgaccatagatgaacagttacttggtttt>1150
336>----->336
404>----->404

* * * * *
1151>agaggacggtgtccgtttaggatgtatatcccaaacaagccaagtaagta>1200
336>----->336
404>----->404

* * * * *

1201>tggaataaaaaatcctcatgatgtgtgacagtgggtacgaagtatatgataa>1250
336>----->336
404>----->404

* * * * *
1251>atggaatgccttattttggaagaggaacacagaccaacggagtaccactc>1300
336>----->336
404>----->404

* * * * *
1301>ggtgaatactacgtgaaggagttatcaaagcctgtgcacggtagttgtcg>1350
336>----->336
404>----->404

* * * * *
1351>taatattacgtgtgacaattgggttcacctcaatccctttggcaaaaaact>1400
336>----->336
404>----->404

* * * * *
1401>tactacaagaaccgtataagttaaccattgtgggaaccgtgcgatcaaac>1450
336>----->336
404>----->404

* * * * *
1451>aaacgcgagataccggaagtactgaaaaacagtcgctccaggccagtggg>1500
336>----->336
404>----->404

* * * * *
1501>aacatcgatgttttgttttgacggacccttactctcgtctcatataaac>1550
336>----->336
404>----->404

* * * * *
1551>cgaagccagctaagatggtatacttattatcatcttgtgatgaggatgct>1600
336>----->336
404>----->404

* * * * *
1601>tctatcaacgaaagtaccggtaaacgcgcaaattggttatgtattataatca>1650
336>----->336
404>----->404

* * * * *

1651>aactaaaggcggagtggacacgctagaccaaatgtgttctgtgatgacct>1700
336>----->336
404>----->404

* * * * *
1701>gcagtaggaagacgaataggtggcctatggcattattgtacggaatgata>1750
336>----->336
404>----->404

* * * * *
1751>aacattgcctgcataaattcttttattatatacagccataatgtcagtag>1800
336>----->336
404>----->404

* * * * *
1801>caagggagaaaagggttcaaagtcgcaaaaaatttatgagaaacctttaca>1850
336>----->336
404>----->404

* * * * *
1851>tgagcctgacgtcatcggtttatgcgtaagcgtttagaagctcctactttg>1900
336>----->336
404>----->404

* * * * *
1901>aagagatatttgcgcgataatatctctaataattttgccaaatgaagtgcc>1950
336>----->336
404>----->404

* * * * *
1951>tggtacatcagatgacagtactgaagagccagtaatgaaaaaacgtactt>2000
336>----->336
404>----->404

* * * * *
2001>actgtacttactgcccctctaaaataaggcgaaaggcaaatgcatcgtgc>2050
336>----->336
404>----->404

* * * * *
2051>aaaaaatgcaaaaaagttatttgcgagagcataatattgatattgtgcca>2100
336>----->336
405>-----cgagagcataatattgatattgtgcca>430

* * * * *

```

2101>aagttgtttctgactgactaataagtataatttgtttctattatgtataa>2150
337>-----taataagtataatttgtttctattatgtatag>368
431>aagttgtttctgactgactaataagtataatttgtttctattatgtatag>480

```

```

      *           *           *           *           *
2151>gttaagctaattacttatttttataataacaacatgactgtttttaaaagtac>2200
369>gttaagctaattacttatttttataataacaacatgactgtttttaaaagtac>418
481>gttaagctaattacttatttttataataacaacatgactgtttttaaaagtac>530

```

```

      *           *           *           *           *
2201>aaaataagtttatttttgtaaaagagagaatgtttaaaagttttgttact>2250
419>aaaataagtttatttttgtaaaagagagaatgtttaaaagttttgttact>468
531>aaaataagtttatttttgtaaaagagagaatgtttaaaagttttgttact>580

```

```

      *           *           *           *           *
2251>ttatagaagaaatttttgagtttttgtttttttttaataaataaataaaca>2300
469>ttatagaagaaatttttgagtttttgtttttttttaataaataaataaaca>518
581>ttatagaagaaatttttgagtttttgtttttttttaataaataaataaaca>630

```

```

      *           *           *           *           *
2301>taaataaaattgtttggtgaattttattattagtagtaagtgtaaatataa>2350
519>taaataaaattgtttggtgaattttattattagtagtaagtgtaaatataa>568
631>taaataaaattgtttggtgaattttattattagtagtaagtgtaaatataa>680

```

```

      *           *           *           *           *
2351>taaaacttaatatctattcaaattaataaataaacctcgatatacagacc>2400
569>taaaacttaatatctattcaaattaataaataaacctcgatatacagacc>618
681>taaaacttaatatctattcaaattaataaataaacctcgatatacagacc>730

```

```

      *           *           *           *           *
2401>gataaaacacatgcgtcaatttttacgcatgattatctttaacgtacgtca>2450
619>gataaaacacatgcgtcaatttttacgcatgattatctttaacgtacgtca>668
731>gataaaacacatgcgtcaatttttacgcatgattatctttaacgtacgtca>780

```

```

      *           *
2451>caatatgattatctttctaggg>2472
669>caatatgattatctttctaggg>690
781>caatatgattatctttctaggg>802

```

As is evident above, the Terminal Domains in the plasmid PB-CA extends additional 67 bp into the piggyBac transposase ORF sequence at the 5' end and additional 44 bp at the 3' end, compared to the Terminal Domains in the plasmid pBluescript-PB. Basically, in the plasmid PB-CA, the Terminal Domains overlap somewhat with the piggyBac transposase ORF, whereas in the plasmid pBluescript-PB, the Terminal Domains end near the 5' and 3' ends of the piggyBac transposase ORF.

By the way, the PB TD sequences in the plasmid pBluescript-PB arose from alignments of plasmid sequences from Dr. Sen Wu, Capecchi laboratory and NCBI sequences (one shown above and additional plasmids).

go to next page

III. The flanking baculovirus sequences

The baculovirus sequences near the site into which the transposon had integrated

NCBI accession number MH077961

<https://www.ncbi.nlm.nih.gov/nuccore/MH077961>

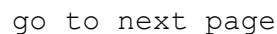
or

NCBI accession number KY792989

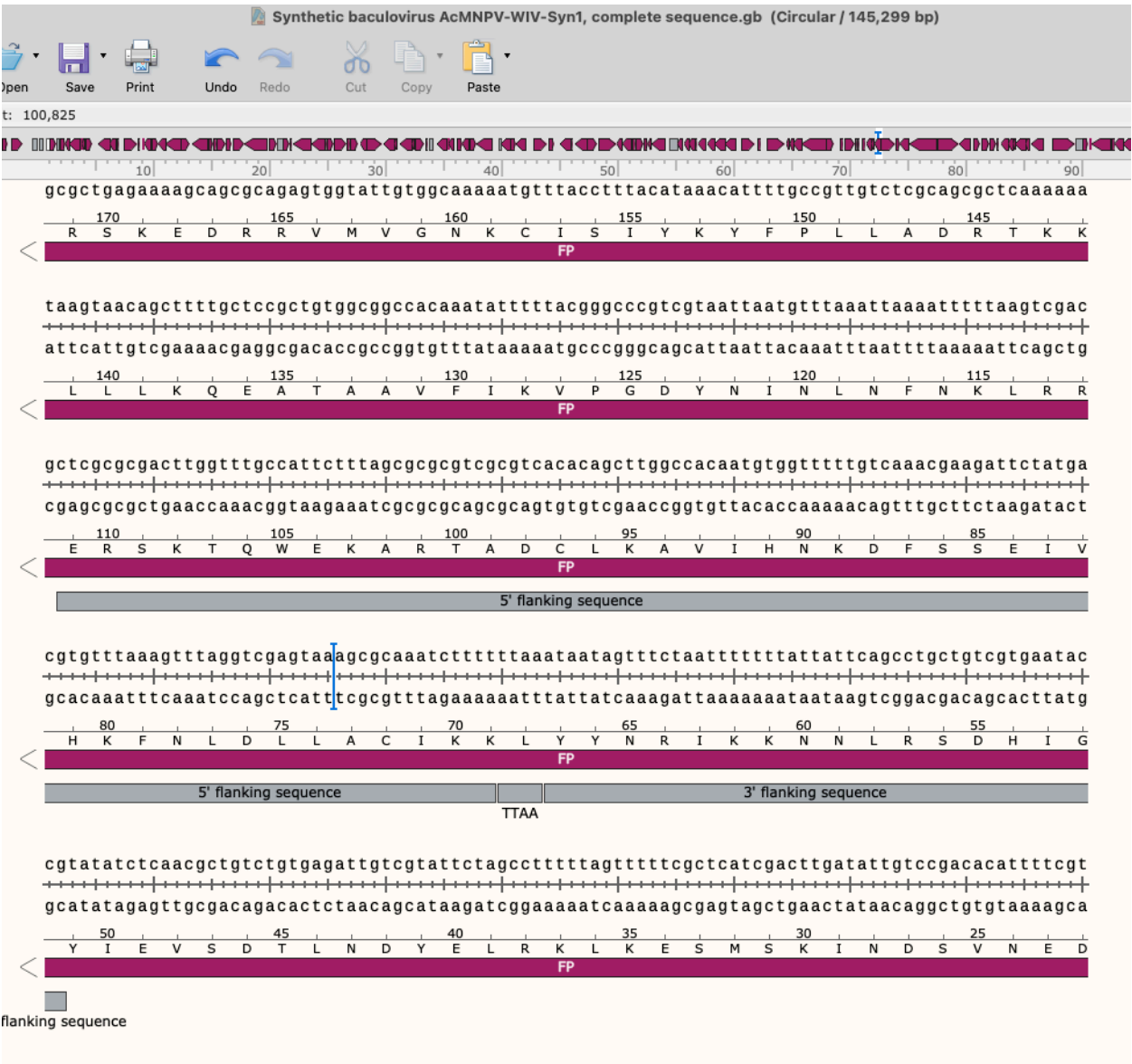
<https://www.ncbi.nlm.nih.gov/nuccore/KY792989>

go to next page

Homologies of flanking sequences in PB-CA to MH077961



Homologies of flanking sequences in PB-CA to KY792989



go to next page

V. Baculovirus flanking sequences alignments

First line Baculovirus sequence

Second line PB-CA flanking sequence

Third line pBS-PB flanking sequence

TTAA in **RED**

Ends of the PB plasmids' flanking sequences in **BLACK**

Fri Jun 03, 2022 23:17 MDT

Baculovirus.gb from 1 to 315

Alignment to

pBS-PB flanking sequences.gb-- Matches:165; Mismatches:8; Gaps:150;

Unattempted:0

PB-CA flanking sequences.gb-- Matches:210; Mismatches:6; Gaps:103;

Unattempted:0

```

      *           *           *           *           *
1>tggcggccacaaatatttttacgggcccgtcgttaattaatgtttaaatta>50
1>~~~~~t----->1
1>~~~~~a-----a-----g---c-t-g-----g-----t-->8

      *           *           *           *           *
51>aaatttttaagtcgacgctcgcgcgacttggtttgccattcttttagcgcg>100
2>--a-----t--a-g---g-g---c--g-----a--c-----c-c->13
9>--a-----c----ctcgcgcgacttggtttgccattcttttagcgcg>43

      *           *           *           *           *
101>cgtcgcgtcacacagcttgccacaatgtgggttttgtcaaacgaagatt>150
14>-----a-a-----cacaatgtggtttttgtcaaacgaagatt>44
44>cgtcgcgtcacacagcttgccacaatgtgggttttgtcaaacgaagatt>93

      *           *           *           *           *
151>ctatgacgtgtttaaagtttaggtcgagtaaagcgcaaattctttt---tt>197
45>ctatgacgtgtttaaagtttaggtcgagtaaagcgcaaattcttttnnnnn>94
94>ctatgacgtgtttaaagtttaggtcgagtaaagcgcaaattcttttnnnnn>143

      *           *           *           *           *
198>aa-ataatagttttctaatttttttattattcagcctgctgtcgtgaatac>246
95>nnnataatagttttctaatttttttattattcagcctgctgtcgtgaatac>144
144>nnnataatagttttctaatttttttattattcagcctgctgtcgtgaatac>193

      *           *           *           *           *
247>cg-t--atatctcaacgctgtct-gtgagattgtcgtattctagcctttt>292
145>cggggggaggt-tc--c-ct-t-tagtgag---g--g--tt--a-a-t-t->175
194>cg----a-g-ct---c-c-----a-a-t-tcg----c---cc-t-a>212

      *           *

```

```
293>tagtttttcgctcatcgacttga>315
176>--g-----cg--c---g-c~~~~>181
213>tag-t----g---a--g--t~~~~>220
```

As is evident above, the 5' PB-CA flanking sequence has additional 54 bp of the flanking sequences from the baculovirus compared to the pBluescript-PB. The 3' flanking sequences are the same in the two plasmids.

go to next page

The PB Terminal Repeats and Terminal Domains sequences in p15A-PB are from the plasmid pBluescript-PB, and they are the same.

✧

The plasmid pBS-minimal PB doesn't have the baculovirus flanking sequences nor much of the Terminal Domains sequences and the Internal Domains sequences. The DNA sequences flanking either sides of the minimal PB Terminal Repeats in that plasmid are vector backbone sequences from plasmid pFNF (Addgene #22687). To confirm the minimal PB Terminal Repeats sequence, see Troyanovsky et al., 2016 <https://doi.org/10.1038/mtna.2016.76>.

end of the document