

The complete nucleotide sequence of the Tst1 retrotransposon of potato

Anne Camirand¹ and Normand Brisson*

Département de biochimie, Université de Montréal, PO Box 6128, Station A, Montréal, Québec H3C 3J7 and ¹McGill Cancer Center, 3655 Drummond, Montréal, Québec H3G 1Y6, Canada

Submitted July 10, 1990

EMBL accession no. X52287

Tst1, a 5060 bp-long *copia* -like element was found in the fifth intron of the single-copy gene encoding starch phosphorylase in potato (*Solanum tuberosum* cv. Kennebec) (1). The element is inserted in a reverse orientation to that of the phosphorylase gene and appears to be present in only one copy per haploid potato genome. Tst1 is flanked by a putative 5 bp-long host target sequence repeat, and possesses long terminal repeats of 285 and 283 bp at its 5' and 3' ends, respectively. Its internal domain defines four overlapping ORFs. ORF1 codes for 109 amino acids. ORF2 codes for 334 amino acids and contains regions homologous to the RNA binding site and protease of other retroviral elements (2). ORF3 and 4 code for 675 and 390 amino acids respectively. Suppression of the stop codon at the end of ORF3 would place ORF4 in phase with ORF3 and define a single ORF of 1069 amino acids encoding an integrase and a reverse transcriptase. The overall order of the four domains is identical to that reported in the *copia* element of *Drosophila* (3).

ACKNOWLEDGEMENTS

This work was supported by a fellowship from the Fonds de la Recherche en Santé du Québec to A. Camirand, a fellowship from the Natural Sciences and Engineering Research Council of Canada to N. Brisson, and by a grant from the Medical Research Council of Canada.

REFERENCES

1. Camirand, A., St-Pierre, B., Marineau, C. and Brisson, N. (1990) *Mol. Gen. Genet.* in press.
2. Doolittle, R.F., Feng, D.-F., Johnson, M.S. and McClure, M.A. (1989) *Quart. Rev. Biol.* **64**, 1-30.
3. Mount, S.M. and Rubin, G.M. (1985) *Mol. Cell. Biol.* **5**, 1630-1638.
4. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

1	TCTGTAATGA	GTAGGCAGAT	TTACTCATAA	CTGCTCCACC	ATTGAGCTGC	CAGATTTGAT
61	TATACCTGCT	CCACAATAA	ACGTGGCAGA	TTGATTAGA	ATCTGCTCCT	GTAATATGCG
121	GATTAAGAATA	TATATTCTTA	GGAAATATG	TAGAGGAAA	TCCTTAAT	TAAGGATTTCC
181	CTAGAAATA	GGATTTGAT	ACTTTATTCG	GTCTTACTT	GTCTCTTAT	AAATCTGTA
241	CAAACCACTC	GAATAAAT	ACATTTTCCG	GTATCAAA	TTTGGTCTTA	TCAGAGCCGA
301	CCATTTCCCC	TAATACATA	TGGATAAACC	AGCTTTTCC	AGAGCGATC	TCTTGAACC
361	ATCAACAGGA	ACGTCTCGCA	AAAATCTGA	TGACACAGG	AGGCAATCT	TCTCACATCC
421	CTTCCAGTT	AACCTCTCAT	CGATTAATG	GGAGAACTA	TCTGGATGG	GCACATCCG
481	TAAGCTTGC	AATCGATGG	CGTGAATG	TGGGACATC	GACCGAGAG	ACAAAGAAC
541	CGGAGTGG	AGAGATCAGA	GAATCACTC	GTGATTCCTT	GGTAAATAA	TTCTGCGAA
601	CTAGTATAG	GTAAATCTTA	TCTCTTTTG	CCACTGCCA	AGGATGCATG	GGAAAGCTT
661	AGGAGTCTAT	ATTCGATT	AGAAAATGCC	TCTCAAAAT	TGGAGTAAA	AATGAACTA
721	TGGCAGACTA	AGCAGGTTGA	GAATGAATG	ACTACTTAT	ATAATGAAT	GTGTCCCTTG
781	TGGCAGGAGT	TAGACCAGT	TTACAATGAT	GAATGGAGT	GTCTATGGA	TAGTGTGAA
841	GCCAAAAAAA	AGGAAGAAA	TGAGAGAGTT	TACCTTTTT	TAGCAAGAT	AAACCGAGAG
901	TTTGACAAAG	TCAGGTCGG	AATCTTAGA	AAAACTCTC	TGCCCTTCC	TGAATTTTT
961	CTGAAATTA	GGACAGAGA	AATCTGGAG	AAAGTTAGT	TAAAACCTGA	TTTAAATGT
1021	GAATCAAAAC	CTGTATAGA	CTTTACAGT	CTTGTACAG	TGAAGAATGA	AGAGGACAAA
1081	AAGAAAAGC	CATTGTGTGA	TCATTTCAA	AAATATGGC	ACACCCTGA	AACCTGTTGG
1141	AAGATTCATG	GGAAACCCT	AACTGGGAA	AACAAGGAG	TTGATGATCA	TGCCTTTCAA
1201	TCTCAAAATG	GTACAGCCCT	CCAAACAAT	TATTTGTATC	AGGGGACAA	ACCTTCTCCA
1261	GAACGCTCT	TATTTGATG	GGAACTATC	GAATTCCTG	ACAACTCTC	TCAATCTGA
1321	CAATTCCTG	CAATGCACC	AAATCTTCT	AATCTCTCT	GTCTTTTGC	TGAACCTGA
1381	TCATCTGAT	CGTGTCTTT	TCTAGTGTG	AATTCACAAC	AAATAGATCT	TTGGACTGTA
1441	GATTCAGAG	CTAGTACCA	CATGACTGG	AGTTCCGTT	TTTTCTCAT	TTACACTCT
1501	TGTGCAGGA	ACAGTAAAT	CAAGTAGCT	GTCTTTCC	TCTCTGCAAT	TGCTGAAAA
1561	GAACATATA	AACCTATAC	CATCTCTGT	ACTTCAAT	ACTCTCTAT	TTCCAAGCT
1621	CTCTGTGAT	CTAGTCTTT	TCCGTAAT	AACCCGCTT	CTTAAATGT	GTGTTATTT
1681	TTATTCGAT	TTGTGTAAT	TTGAGAAAA	GGTCTGGGG	AAGATGATT	GCAGTGCTAG
1741	AGAACTCAG	GGCTTTTAT	TTCTTGACAA	CGGGAACAAC	TCACTACAG	TGAATCTAT
1801	TTTTTAAAC	TCTACTTTT	TTTTGAATA	AGTCAGTAT	TGGCACTAT	GGCTTTGACA
1861	TCTTGAATTT	TATTTATTA	GACATTTTA	ACTCAGTTA	TTTTGAATA	AAAATCCATC
1921	CTTATCCAA	TGTGAATTT	GTGAAATGG	TAAGCATCA	TGATATACT	TTTTCTCTC
1981	TCAAAGATC	CAGGCTCAA	AACTTTTAC	AATGATTCAT	AGTATGTTT	GGTTCACCGC
2041	TAGAATTTCA	ACAATGTTG	GAAGAGGTT	GTGTAACCT	TTTATGAGC	ATCATACTAG
2101	TTTGAAGGCA	GTGTTTAT	TGAGGGCAA	ATCTGAGTT	AAAAATGTT	TTGAGACTG
2161	CCATGCTATG	GTGAAACAC	AGTCAATGA	GAATAACTG	ATTTTTCGA	GTGCAATGG
2221	CGGTGAATTT	TTAATGAAC	AATTAGAAG	TTTTTTCAG	AAAACCTGG	TAGTACACCA
2281	AAGCTTTCT	CCAGATCAC	CTCAACAAA	TGGATCTGC	GAAGAAAAG	ATAGCATCT
2341	TTTTAGAGGCA	ACGAGAGCTT	TAATGTTTC	TAGTAAAGT	CCACAACCC	TTTGGGGAGA
2401	GGCTCTCTG	ACTGCCACT	ACTTATTTA	TAGAATGCA	TCTGCCCTC	TGAAATTTA
2461	GACACCTTT	AAAGTTTCA	AGCAAGTTT	TCTAGTTCT	AGGTAACCTA	CAGATCTACC

2521	TCTGAGAGTT	TTTGGTTGTA	CAACATTTG	TCATGTTCA	AATCGTAGTA	AACTGAAAC
2581	GCGGGCTAAG	AACTGTTAT	TTGAGGGTA	TGCCCAAGT	CAAAAGGAT	ATAAGTTGA
2641	TGATCCACAT	GCTAGGAAGA	TAATTGTCAC	AATGGACCTA	ACTTTCTTG	AGTCGCACT
2701	CTATTTTACT	ACTCATCTC	AGGGGAGTA	TCATTTAGT	GAAGATTCAT	TTTTTGTGAT
2761	ATTTGAGAA	CTAGATAA	AGCAAAATG	GAGCTAATA	TTACAAATA	ATACTGATG
2821	GAGAGATGA	GGGGAAGTA	TAAATAAGT	TGATCCAAGA	GATGATAGG	ACCAAGTGA
2881	CTTAATGATA	AAAACCTAGA	AAATCAAACT	TGAACCTGA	GGCCCTCAA	ATGCAAAAA
2941	TAAAAATGG	AATAGAGAAC	AAAAACAGA	AATGAGGGT	TATTCGAGAA	GGACCAGAAC
3001	TCAGAAAAA	AGGACCGAAG	ATTCTCAAC	CTGCCAAAA	TCAGCCCA	AAGACCTTAC
3061	TGTAATCAA	GTACTCTTC	CACTGATTC	TATTTCTAA	TCTTAGATC	TAGATCTACC
3121	TATTCGAAA	CGCAAAGGT	TAGAAAATC	TCTAATAAT	CCATTTCTA	TGTTTGTGTC
3181	CTACAAAAA	TTGTCTCTA	GTATTCAGC	CTTACTTCA	CAACTCTTA	GTGTTGAGT
3241	TCCAACAAT	GTGAGGATG	CTCTACAAG	TCCGATGCA	AAGGAGGAT	TTTTGGAGGA
3301	ATTCGAGCT	CTTGAAAAA	ATGAGACATG	GGAGTTAGT	GATTTACCCG	AGGAAAGAA
3361	ACCATGGGT	TGTAAATGG	TGTTCAACC	CAAACTCAA	TCAGACGGAT	CCTTGGAAAG
3421	GTATAAGCC	CACCTAGTAG	TTAAGATCA	CACAGACATA	TCGATGCATG	ACTATCTCA
3481	GACGTTTGT	CCAGTAGTA	AGCTAAACT	AAATAGATG	CTGTTGACAA	TCGCAGTAA
3541	TCTGAGTGG	TCTTTTAGC	AAATGAACT	GAAGATGTT	TTCTGAACG	GGCACCCTG
3601	GGAGAAGTC	TATATGATC	CCCAACAGG	TTTTGAAGG	AAGTACAAGT	CAAAAATATG
3661	CAGGCTTAGA	AGATCTCTT	ATGGTCTAA	ACAATCTCA	AGGCTTGTG	TCGAAAGGTT
3721	TACTCAATTT	GTGAAAAGC	AAGGATATG	GCAAGGACA	GCAGATACA	CAATGTTTAC
3781	TCAGACTTCA	CTAGAAAGG	AAACAACCT	TCTTATAGT	TATGTTGACG	ATATCTCTC
3841	CACAGAGAT	GATGTTGTT	AGATAAAAA	TCTAAGGAA	CGTCTGCTC	CAGAATTTGA
3901	GATAAGGAC	TTAGGCCCG	TAAAGTACT	TCTTGGCAT	GAGGTTGCAC	GATCGAAGAA
3961	AGGAATATA	GTGTCACAA	GGAAGTATG	TCTTGATCT	TAAAAGAAA	CAGGAATGAG
4021	TGTTGTTAG	CAACTGAAA	CTCAAATTA	TCCAAATCT	AGTTTGTAA	AGGAAAGAAA
4081	ATGATTTGAT	AAGGGTCAAT	ATCAGAGAT	GGTGGCCAG	TTGATTTACT	GTTCACATC
4141	TAGACTTGT	ATTTCTTTG	CTGTGAGCT	AGTCAATCAG	TTCATGAT	ATCCACGAG
4201	AGAACACCAA	GAAGCTGTG	ATCGAACTA	AAGTATCTA	AAGATTTAC	CTGGGAAAG
4261	GTGTTCTCT	AAGAAGATG	AGCAAAGAG	TCTTGGGCT	TATACGGATG	CAGATTGGCG
4321	TGTTCTACT	ATTGATAGA	GTCTACTAC	TGGATATGG	ACATTTGTT	GGGAAATTT
4381	GGTACATGG	AGAATGAAG	AGCAAAATG	GTGGCTCGA	AGTGTAGT	AGGCTGAGTA
4441	TCTGTCATG	CTCTCCGGAA	TTTTGAAAT	ATTGTCGTC	AGAGATTTT	TGGAAGAACT
4501	AAAGAAACT	GTGATTTTC	CAATGAATG	GTATTTGAC	AAATGGGCT	CCATAGCAT
4561	TGCAATPACT	CCAGTTCAAC	ATGATAGAC	AAAGATCTG	GAAGTACAG	ACACATCTT
4621	AAAAGAAAG	TGAAGATGG	AGTGTGTCG	ATTTCTTTG	TGCCAACAC	AGCAAGGTT
4681	CAAAATATA	TCACAAAAG	TCTTTTCAA	ACTACTTTG	ATCCCTTTG	TAGCAATTTA
4741	GGCATTTTG	CTATTACAT	GCCGACTGA	GGGGGATTT	TGATTTGAT	GCCAGATTTA
4801	GTCAATTTCT	CTCTCAACT	TGAGTTGCA	GATTTGATTA	TAATTTGCT	CACAAATAAA
4861	GTGTCAGAT	TGAGAATCT	GCCTGTA	TATGGATTA	GAATATATA	TCTAGGAAA
4921	ACTTTTAGA	GGAAATCTCT	TAATCAAGG	ATTTCTTAAG	AATTAGGAT	TGATTTGTT
4981	ATTTGGTCT	TACTTGTCT	CTTAAATAA	CTGTACAAC	ACTAGCAATA	AAATATACT
5041	TTGAGTATC	TAAACATTA				

Nucleotide sequence of the Tst1 element of potato as obtained by dideoxy sequencing (4) of both strands. LTR: long terminal repeat. First strand primer site: nucleotides 286 to 307. Second strand primer site: nucleotides 4769 to 4777. Small boxes: ATGs of the four ORFs. Underlined: stop codons of the four ORFs.

* To whom correspondence should be addressed