**Practical 2: TE classification and nomenclature (aka Transposon quiz)**

1. **Is there any repeat in Repbase with similarity across most of the length of this sequence?**

* Paste all sequences into CENSOR to compare to Repbase. Choose default parameters.
* Click on repeat names for more info (note: Repbase restricts access after some traffic).
* <http://www.girinst.org/censor/index.php>

1. **Is there any similarity to nucleotide sequences in GenBank or gtRNAdb?**

* Paste all sequences for BLASTn against GenBank “*nucleotide collection (nr/nt)*”. Choose “*Somewhat similar sequences (blastn)*” and default parameters.
* <http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome>
* Paste the sequence for BLASTn against Genomic tRNA Database. Choose default parameters.
* <http://gtrnadb2009.ucsc.edu/blast.html>

1. **Is there any similarity to protein sequences or domains in GenBank or CDD?**

* Paste all sequences for BLASTx against GenBank “*Reference proteins (refseq\_protein)*”. Choose default parameters.
* <http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome>
* Paste the sequence into Conserved Domain search. Choose default parameters.
* <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

1. **Are there any peculiar motifs at the start/end of this sequence?**

* If the sequence starts with TG… and ends with CA…, it is likely an LTR retrotransposon.
* If the sequence starts with 10-30 bp that are reverse complement to the last 10-30 bp of the repeat, it is likely a DNA transposon.
* <http://reverse-complement.com/>

1. **Are there tandemly repeated motifs within this sequence?**

* Paste the sequence two times into MAFFT to align against itself. Choose default parameters and “Plot only” at the bottom of the page. Look at resulting dot plots.
* <http://mafft.cbrc.jp/alignment/server/>

1. **Is there anything peculiar about the secondary structure of this sequence?**

* If sequence <900 bp long, paste into mfold to estimate secondary structure. Choose default parameters. Look at the .jpg file of the highest ranked output.
* <http://www.unafold.org/mfold/applications/rna-folding-form.php>

|  | **Sequence** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **What is it?** | **01** | **02** | **03** | **04** | **05** | **06** | **07** | **08** | **09** | **10** |
| DNA transposon |  |  |  |  |  |  |  |  |  |  |
| Hepatitis B virus |  |  |  |  |  |  |  |  |  |  |
| HIV |  |  |  |  |  |  |  |  |  |  |
| LINE retrotransposon (CR1) |  |  |  |  |  |  |  |  |  |  |
| LINE retrotransposon (RTE) |  |  |  |  |  |  |  |  |  |  |
| LINE retrotransposon (CR1) nested within DNA transposon |  |  |  |  |  |  |  |  |  |  |
| LTR retrotransposon |  |  |  |  |  |  |  |  |  |  |
| Microsatellite |  |  |  |  |  |  |  |  |  |  |
| Palindrome with tRNA triplex |  |  |  |  |  |  |  |  |  |  |
| tRNA gene |  |  |  |  |  |  | y |  | y |  |
| SINE retrotransposon |  |  |  |  |  |  |  |  |  |  |
| Satellite |  |  |  |  |  |  |  |  |  |  |

**Hint: There is only one answer per sequence.**

If time allows: Suggest a name for each sequence according to nomenclature (taxon: *Bla bla*).

**Personal notes for transposon quiz**

| **Sequence** | **Notes** | **Suggestion for nomenclature** |
| --- | --- | --- |
| **01** | nBlast: no sim  Protein: n |  |
| **02** | MicTal-51094  class: DNA/hAT  nBlast: y  Protein: y |  |
| **03** | CR1-E1\_fAlb  Class: NonLTR/CR1  nBlast: y  Protein: y |  |
| **04** | hAT-N4\_AmRa  class: DNA/hAT  nBlast: y  Protein: y |  |
| **05** | ERV3-7\_PMaj-LTR  Class: ERV/ERV3  nBlast: y  Protein: y |  |
| **06** | AviRTE\_MUn  Class: onLTR/RTE  nBlast: y  Protein: y |  |
| **07** | TguSINE1  Class: NonLTR/SINE/SINE2  nBlast: y  tRNA: Xenopus\_tropicalis\_GL181548.trna6-AspGTC (8829-8900) Asp (GTC)  Protein: n |  |
| **08** | eBHBV1  eBHBV2  Class: IntegratedVirus/DNAV  nBlast: y  Protein: y |  |
| **09** | ERV2-8\_CAnn-LTR  Class: ERV/ERV2  nBlast: y  tRNA: Taeniopygia\_guttata\_chr3.trna2-IleTAT (15414414-15414507)  Protein: n |  |
| **10** | CR1-X1\_fAlb  Mariner-25\_SIn  Class: NonLTR/CR1 or DNA/Mariner  nBlast: y  Protein: y |  |