Landes Bioscience

www.landesbioscience.com

Supplemental Material to:

Nara Lee, Genaro Pimienta, Joan A Steitz, Walter N Moss

RNA families in Epstein–Barr virus

2013; 11(1)
http://dx.doi.org/10.4161/rna.27488

www.landesbioscience.com/journals/rnabiology/article/27488/

```
# STOCKHOLM 1.0
#=GF ID    EBER2
#=GF AC
#=GF DE    Epstein-Barr virus EBER2
#=GF AU    Moss W
#=GF SE    Published; PMID:17381320
#=GF SS    Published; PMID:17381320
#=GF TP    Gene; sRNA;
#=GF RN    [1]
#=GF RM    17381320
#=GF RT    The challenge of viral snRNPs.
#=GF RA    Conrad NK, Fok V, Cazalla D, Borah S, Steitz JA;
#=GF RL    Cold Spring Harb Symp Quant Biol. 2006;71:377-384.
#=GF RN    [2]
#=GF RM    17381320
#=GF RT    Structural analyses of EBER1 and EBER2 ribonucleoprotein particles present in
Epstein-
#=GF RT    Barr virus-infected cells.
#=GF RA    Glickman JN, Howe JG, Steitz JA;
#=GF RL    Journal of virology 1988;62(3):902-911.
#=GF WK    http://en.wikipedia.org/
#=GF SQ    4

EBV1|NC_007605          AGGACAGCCGUUGCCCUAGUGGUUUCG--GACACACCGCCAACGCUCAGU
GD2|HQ020558            AGGACAGCCGUUGCCCUAGUGGUUUCG--GACACACCGCCAACGC-----
MHV4|NC006146           AGGACAUUGGCUAGCCUAGAGGUUACGGACAGAGGGGGGUGGUGUGCGGG
CeHV12|AF200364         -GACCAUCGGCUGCCCUAAUGGUUACG--UAGGGAGGAGUUGUGUCCGGC
#=GC SS_cons            .<<<<<<<<<.<<<<<<..<<<..<<<..<...<<<<<<..........

EBV1|NC_007605          GCGGUGCUACCGACCCGAGGUCAAGU-CCCGGGGGAGGAGAAGAGAGGCU
GD2|HQ020558            ---------------------------CCCGGGGGAGGAGAAGAGCGGCU
MHV4|NC006146           UUCAUCCGUCCGUGCCGAGUACAAGUCCCGGGAAGGGGAGAAGAGCGGCU
CeHV12|AF200364         UUAGUCCGAGCGUACCGGGUACAAAUCCCGGGGGGUUGAGAAGAGCGGCU
#=GC SS_cons            >>>>>>...>>>>.>>>........<.<<<<<<<<<<<<<<...<<<<.

EBV1|NC_007605          UCCCGCCUAGAGCAUUUGCAAGUCAGGAUUCUCUAAUCCCUCUGGGAGAA
GD2|HQ020558            UCCCGCCUAGAGCAUUUGCAAGUCAGGAUUCUCUAAUCCCUCUGGGAGAA
MHV4|NC006146           CCCGCCUAUGUGCAAUUGA-AGUCAGGAUUCUCCACCCUCCCUGG-CAAC
CeHV12|AF200364         CCCG-CCAAGUGCAUUUGC-AGUCAGGAUUCUCUACCCUCCCCGGGUAAA
#=GC SS_cons            ....>>>>..................>>>>>>..>>>>.>>>>>>..>

EBV1|NC_007605          GGGUAUUCGGCUU---GUCCGCUAUUUUU
GD2|HQ020558            GGGUAUUCGGCUU---GUCCGCUGUUUUU
MHV4|NC006146           G-----CCAGCCAAU-GUCCGCU-CUUUU
CeHV12|AF200364         G-----CCAGCUCAUGGUCUGCU-CUUUU
#=GC SS_cons            >>>>>..>>>>.>...>>>>.........

//
```

```
# STOCKHOLM 1.0
#=GF ID    IRES_EBNA
#=GF AC    RF00448
#=GF DE    Epstein-Barr virus nuclear antigen (EBNA) IRES
#=GF AU    Moss WM
#=GF TP    Cis-reg; IRES;
#=GF DR    SO:0000243 SO:internal_ribosome_entry_site
#=GF DR    GO:0043022 GO:ribosome binding
#=GF RN    [1]
#=GF RM    12555070
#=GF RT    Epstein-Barr virus U leader exon contains an internal ribosome entry
#=GF RT    site.
#=GF RA    Isaksson A, Berggren M, Ricksten A;
#=GF RL    Oncogene 2003;22:572-581.
#=GF CC    This is a proposed revision of the structure presented in reference [1]. The
original
#=GF CC    model, based on predicted folding free energy, forbid pseudoknot folding. This
revision
#=GF CC    combines a program (DotKnot- http://dotknot.csse.uwa.edu.au) that allows for
pseudoknots,
#=GF CC    with comparative sequence analysis. A 46 nt region can form a conserved
pseudoknot.
#=GF WK    http://en.wikipedia.org/wiki/Epstein-Barr_virus_nuclear_antigen_(EBNA)_IRES
#=GF SQ    3

EBV1|NC_007605        GAAAAGGCGCGGGAUAGCGUGCGCUACCGGAU--GGCAGGUAACUUAGGA
EBV|M125531           -AGAGGAUGAAGACUA-----AGUCACAGGCUUAGCCAGGUAACUUAGGA
MHV4|NC006146         --UCAGCUGAUCGUUGUUGUCAUUCUUCAAUUUUUGCAGGUAACUCAGAA
#=GC SS_cons          ..<<<<<.<<<<....<<<.<<<.<<<<..........>>>>.AAAAAA.

EBV1|NC_007605        AGCGUUUCUUGAGCUUCCCUGGGAUGAGCGUUUGGGAGAGCUGAUUCUGC
EBV|M125531           AGCGUUUCUUGAGCUUCCC-------------UGGGAGAGCUGAUUCUGC
MHV4|NC006146         AGCGUUUCUUGAGCUUCCCGGGGAUGAGCGUCUGGGAGAGCUGAUUCUGC
#=GC SS_cons          ....BBBBBBBBBBBBaaaaaa..bbb.bbbbbbbbbb..<<<.<<<<<

EBV1|NC_007605        AGCCCAGAGAGUAGUCUCAGGGCAUCCUCUGGAGCCUGACCUGUGACCGU
EBV|M125531           AGCCCAGAGAGUAGUCUCAGGGCAUCCUCUGGAGCCUGACCUGUGAUCGU
MHV4|NC006146         AGACCGGCUAGGAGCCGCAGGGCGUUCUCCGGGGCGCGGCCCGAGACCGU
#=GC SS_cons          .<<<<<<<<.<<.<<<....>>>>>.>>>>>>.>>.<<..<<<<<<<...

EBV1|NC_007605        CGCAUCAUAGACCGCCAGUAGACCUGGGAGCAGAUUCACCGCCGCGGCCG
EBV|M125531           CGCAUCAUAGACCGCCAGUAGACCUGGGAGCAGAUUCACCGCCGCGGCCG
MHV4|NC006146         CGCGUCGUAGACAGCCAGGAGACCCGGGAGCAGAUUCACCGCAGCGGCCG
#=GC SS_cons          ...>>>>>>>..>>>>>.....>>>>..>>>>>.>>>.>>>>>>..>>>

EBV1|NC_007605        UCUCCUUUAAGUGUGAAUC
EBV|M125531           UCUCCUUUAAGACCCCACC
MHV4|NC006146         UCUCCCUCAAGGUGCUGUG
#=GC SS_cons          >..>>>>>..........


//
```

```
# STOCKHOLM 1.0
#=GF ID    EBVSIS1
#=GF AC
#=GF DE    Epstein-Barr virus stable intronic sequence RNA (ebv-sisRNA-1)
#=GF AU    Moss W
#=GF SE    Published; PMID:17381320
#=GF SS    Published; PMID:17381320
#=GF TP    Gene; sisRNA;
#=GF RN    [1]
#=GF RM    23937650
#=GF RT    Genome-wide analyses of Epstein-Barr virus reveal conserved RNA structures and a
novel
#=GF RT    stable intronic sequence RNA.
#=GF RA    Moss WM, Steitz JA;
#=GF RL    BMC Genomics 2013;14:543.
#=GF WK    http://en.wikipedia.org/
#=GF SQ    8

EBV1|NC_009334                GUAAGUGGAC-UUUAAUUUUUUCUGCUAAGCCCAACACUCCACCACACCC
CeHV12|AF200364               GUAAGUGGACUUUUCCAUUUUUCUGCUAAGCCCAACACUCCACCACAGCC
MHV4|NC_006146                GUAAGUGGAC-UUUCAUUUAUUCCGCUCAGCCCACCACUCCACCACACCC
PoHV3|AJ311194                GUAAGUGGAC-UUUAAGUUUUUCUGCUAAGCCCAACAUUCCACCACACCC
HVMF1|X77781|                 -UAAGUGGAC-UUUCAGUUAUUCUGCUCAGCCCACCACUCCACCACACCC
CeHV15|AJ311199               GUAAGUGGAC-UCUUAGUUACUCUGCUCAGCCCACCACUCCACCACACCC
CaHV3|NC_004367               GUAAGUGU---CUUUAAUUUUUAAGCCUCAUGCA-CAGCACACAAACACC
PoHV1|AJ311196                GUAAGUGGAC-UUUAAGUUUUUCUGCUAAGCCCUACACUCCACCACACCC
#=GC SS_cons                  ...<<<<<<.............>>>>>>.....................

EBV1|NC_009334                AGGCACACA---------CUACA----CACACCCACCCGUCU-CAG
CeHV12|AF200364               AGGCACACA---------CUACA----CACACCCGCCCAUCU-CAG
MHV4|NC_006146                AGGCACACA---------CUACA----CACAGCCACCCAUCU-CAG
PoHV3|AJ311194                AGGCACACA---------CUACA----CACACCCGCCCGACU-UAG
HVMF1|X77781|                 AGGCACACA---------CUACA----CACAGCCACCCGUCU-CAG
CeHV15|AJ311199               AGGCACACA---------CUACA----CACAGCCACCCGUCU-CAG
CaHV3|NC_004367               AGGCACACAAACAGCAGCCACACUUCCACACGCGCCCGUCUCCAG
PoHV1|AJ311196                AGGCACACA---------CUACA----CACACCCACCCGUCU-CAG
#=GC SS_cons                  ..........................................

//
```

```
# STOCKHOLM 1.0
#=GF ID    EBVSIS2-HP
#=GF AC
#=GF DE    Epstein-Barr virus stable intronic sequence RNA hairpin (ebv-sisRNA-2 HP)
#=GF AU    Moss W
#=GF SE    Published; PMID:17381320
#=GF SS    Published; PMID:17381320
#=GF TP    Gene; sisRNA;
#=GF RN    [1]
#=GF RM    23937650
#=GF RT    Genome-wide analyses of Epstein-Barr virus reveal conserved RNA structures and a
novel
#=GF RT    stable intronic sequence RNA.
#=GF RA    Moss WM, Steitz JA;
#=GF RL    BMC Genomics 2013;14:543.
#=GF WK    http://en.wikipedia.org/
#=GF SQ    4

EBV1|NC_009334          ----------GAGA----------------------CCAGUCCAGGG-G
CeHV12|AF200364         ACCG------GUGA----------------------CCAGCCAAGAGUC
MHV4|NC_006146          GCCAGCAAGCGAGA----------------------CCAGCCCAGAG-G
CaHV3|NC_004367         GGGGAGUGAAGCGAAAAUUGGGCCCAGGCCGCGCCUGCCAGCCCACAGAC
#=GC SS_cons            ....................................<<<<...<<<<.<

EBV1|NC_009334          CCCGA-----------------GGUUGGAGCCAGCGGGCACC-CGAGGU-
CeHV12|AF200364         CCCGA-----------------GGUUGGAGCCAGCAGGUGCC-CAGGGUC
MHV4|NC_006146          CCCGA-----------------GGUUGGAGCCGGCGGGCGCCACCAGGU-
CaHV3|NC_004367         CCCCAUUCGGACCUCCCCCAUUAUCCCGGCCUAACCGGGUCC-CAGGGGC
#=GC SS_cons                   <<.<<.................<<<.<<<.<<<<.<<<<.<<.<<<<<<.

EBV1|NC_009334          -CCCAGCACCCGGUC-CCUCCGGGGGGCAGAGACAGG--CAGGGCCCCCC
CeHV12|AF200364         CCCCAGAAGGCGGCCUCCUCUGGAGGCAGAGAGCCGGGCAGGGUCCCUCC
MHV4|NC_006146          -CCCAACACCCGGUCCCCCUCUAGUGGACAGAGACAGGGUUAGAGCCCUCC
CaHV3|NC_004367         AUCCGGGAGCCGGGCCCCCUCUG-----------------------CCGC
#=GC SS_cons                   .<<<.......<<<<.<<<<<<<<<<<<<<<<.<<<<.<<<<<<<<<<<

EBV1|NC_009334          GGCAGCUGGCCCCGAGGAGGCGCCCGG--AGUGGGGCCGGUCGGCUGGGC
CeHV12|AF200364         GGCGGCUGGACCCGAGGGGGCGCCCGGGUAGAGCAGCAGGCUGGCGGGGC
MHV4|NC_006146          GGCGGCCGGACCCGAGGAGGCGCCUGG---GCGAGGCUGGGUGGCUGGGC
CaHV3|NC_004367         UUUGGGCGACCCAACUGAGACCCCGAGGCUAGGGACUAGGCCGGCGGGCC
#=GC SS_cons                   <<<<<<<<<<<<<<<<<<<<<<<<<..<<<<<<<<<<<<<<<<<<<

EBV1|NC_009334          UGGCCGAGCCCGGGUCUGGGAGGUCUGGGGUGGCGAGCCUGCUGUCUCAG
CeHV12|AF200364         UGGCCCGGGUUUGGGUUUGGGGGUCGGGGGUGGCGAGCCUGGUGUCUCUG
MHV4|NC_006146          AGGCCGGGUCUCGGGUCUGGGGGUCUGUGGUGGUGAGCCUGGUGCCUCUG
CaHV3|NC_004367         GCCUCCGGGCCAGGGGCUGGACCCCCUGACCGGGAGGCGAGAGGAU----
#=GC SS_cons                   <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<.<

EBV1|NC_009334          GAGGGGCCUGGCUCCGCCGGGUGGCCCUGGGGUAAGUCUGGGAGGCAGAG
CeHV12|AF200364         GAGGGCCUGGGCACCGCCGAGUGGCCCUGGGGUAGCUCUGGAAGGCUGAG
MHV4|NC_006146          GAGG-CCCUGGCCCCGCCCGGUGGCCUUGGGGUAGCUCUGGGAGUCAGAG
CaHV3|NC_004367         AGGGACCCGGCCACCCCUUGGGGGUUCUGGGGACACGGUGGCUGGC----
#=GC SS_cons                   <<<<<<<<<<<<<<<<<.<<<<<<<<<<<<<..<<<<<<<.<<<.<<<

EBV1|NC_009334          GGACGGCCUAGGCCCGGGGAAGUGGAGGGGGAUCGCCCGGGUCUCUGUUG
CeHV12|AF200364         GGUCGGCUCAGGCACAGGGGAGUUGAGGGGGAUCGUCCGGGUCUCUGGUG
MHV4|NC_006146          AGUCGGCCUAGGCCCGGGG-AGUGGAGGGGGAUCGCCCGGGUCUCUGCUG
CaHV3|NC_004367         ---------------------------------UGCCCCGACAUCUAGU-
#=GC SS_cons                   <<.<<<<<..<<<<<<<<<..<<...<<<<<<<<<<<<<<.<<<<<...

EBV1|NC_009334          GCAGAGUCCGGGCGAUCCUCUGAGACCCUCCG-GGCCCGGACGGCCGCCC
CeHV12|AF200364         GCAGAGACCGGACGAUCCUCCAAGCCCCUCUGAGCCCCAAGCGGCCGCCC
MHV4|NC_006146          GCAGAGUCUGAGCGAUCCUCCGAGACUCCGGG-CCCCAAGGCGGCCGCCC
CaHV3|NC_004367         -UGGCCACUAGAGGGCUGUCUAAGUGUCUUGG---------CAGCCAGCA
#=GC SS_cons                   .>>>>>>>>>>>>>>>>>>>...>>.>>>>>.>>>>.....>>>>>>>>

EBV1|NC_009334          UCGGCCCCCCAGACAGACCCCAGGGUCUCCAGGCAGGGUCCGGCAUCUCC
CeHV12|AF200364         UCAGUCCCCCAGACAGACCCCAGGGUCGCCAGGCGGGGUCCUGGUCCUCC
```

```
MHV4|NC_006146          UCAGUCCCCCAGACAGACCCCAGGGUCGCCAGGCGGGGUCCGGCCUCUCC
CaHV3|NC_004367         ACCGUGUCCCCGAAACCCCCAAGGGGC----GGCCGGGUCCCUAUCCUCU
#=GC SS_cons                    >>.>>>.>>>>>>>..>>>>>>>>>>>>>>.>>>>>>>>>>>>>>>>>>>

EBV1|NC_009334          AGGGGCAGCAGGCUCACCACCACAGGCCCCCAGACCCGGGUCUCGGCCA
CeHV12|AF200364         AGGGACAGCAGGCUCGCCACCCCAGGCCUCCCAGACCCAAACCCGGGUCA
MHV4|NC_006146          UGGGGCAGCAGGCUCAACACCCCGGACCUCCCAGGCCCAAGACCCGGCCU
CaHV3|NC_004367         CG-----------CCUCCCGGUCAGGGGGUCCAGCCCCUGGCCCGGAGGC
#=GC SS_cons                    .>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>.>>>>>>>>

EBV1|NC_009334          GCCGAGCCGACC--GGC-CCCGCGCCGGGCGCCUCCUCGGAGCCAGCCCC
CeHV12|AF200364         GCCCCGCCAGCCU-GCCUCUCUCCCCGGGCGCCUCCUCGGGUCCAGCCGC
MHV4|NC_006146          GCCCGGCCUGCCC-GGC-CUCGCUCCGGGCGCCUCCUCGGGUUCAGCCGC
CaHV3|NC_004367         GGCCCGCCGGCCUAGUCCCUAGCCUCGGGGUCUCAGUUGGGUCCAACCCG
#=GC SS_cons                    >>>>>>>>>>>>..>>>.>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

EBV1|NC_009334          CGGGGU-UGGUUCUGCCCCUCUCUCUGUCCUUCAGAGGAACCA-------
CeHV12|AF200364         C-GGGU-GACCCCUGCCCCUCUCUCUGUCUUCCAGAGUGGGGCCGCCCUG
MHV4|NC_006146          U--GGC-UGGCUCUGCCCCUCUCUCUGUCCUCCAGGGGGAACC-------
CaHV3|NC_004367         AAGGGCAGAGUUUUGCCCGGUUCCCUGAAGUCUCAGGUGGGCC-------
#=GC SS_cons                    >>>>>...>>>>>>>>>.>>.>>>>>>>>>>>>>>>>>>>>>........

EBV1|NC_009334          -GGGACCUCGGGCAC------------------------------CCCAG
CeHV12|AF200364         UGGGGGAAC-GGGAC------------------------------CUCAG
MHV4|NC_006146          -AGGGGACC-GGCGC------------------------------CCCAG
CaHV3|NC_004367         -UAAAGCCCGCGCACUCACCACCACCCGUUAAACUAAUGGAAUGGCCUUG
#=GC SS_cons                    .>>>>>>>>>>>......................................

EBV1|NC_009334          AGCCCCUCGGGCCCGCCUCCAGGCGCCCUCCUGGU----------CUC
CeHV12|AF200364         AGCCCCUCGGGCCCGCCUCCAGGCGCCCUCCUGGU-------------
MHV4|NC_006146          AGCCCCUCGGGUCCGCCUCCAGGCGUCCUCCUGGU-------------
CaHV3|NC_004367         AGGUCCUAGGGUC-GCCUCCUGGUGGCCACUGCACCACCUAGCAGCCC
#=GC SS_cons                    .>>>>>>.>>.>>>>>>>>..>>>.>>>>.>>>>..............

//
```

Supplementary files

SI File 1. Stockholm alignment of conserved EBER2 sequences and consensus structure. Sequences are from three EBV strains: EBV1 (EBV type 1), GD1 (Guangdong 1), and GD2, as well as two from related lymphocryptoviruses: MHV4 and CeHV12. Base pairs indicated by matching “<” and “>” characters.

SI File 2. Stockholm alignment of conserved EBNA IRES sequences and consensus structure. Pseudoknot base pairs indicated by matching capital/lowercase letters “A/a” and “B/b” highlighted in green and blue, respectively.

SI File 3. Stockholm alignment of conserved ebv-sisRNA-1 sequences and consensus structure.

SI File 4. Stockholm alignment of conserved ebv-sisRNA-2 hairpin sequences and consensus structure.