

TCTCATGTCTCACATTTCAAAAATAGGACCATTCCCTTTCCAACAGTCCCCCAGGGTCTTAGCTCATTTCCAACATTA  
GGCCAAAAGTCCAAATCCAGTCTAATCTGTGACGAGGCAAGTCCCTTTACCTATAAACCTGTAAAATCCAAAGCA  
AATTAATTAATCTTCTTAGATACAGTGGAGGTACAGGCATTAATAAAAACCCATTCCAAATGGGAGAAATGGCCAAAA  
CAAAGGGGCTAAAAGGCCCACTGAGGCAATAATTAATCTAAGGCTCCAGAATAATCTTTGATTCATGTCTCATAT  
CTAGGTAAGCTGGTACAAGTAGAGGACTCCCACAGCCTTGGCAGTTCCTGCCCCGTGGCTTTGCAGGGTACAGCC  
CCCCATTATGGCTGGTTCATGGGCTGGTGTGTCTGGCTTTCCAGGCTCACAGTGCAAGATATCACAGTGTATTC  
TACCTTGTGTGGTATTGGAAGAATAAGGCCCTCTTTTAAAAGTTTCCATAAGCCTATACCCCCTGAAGATCGTT  
TGGGGGCTTACAACCATAAGATTTTTAATAAAAAAGGGCGCCAAACCATTTGGCCATTTAAGAGAGGGTTCGCGCC  
CTAAAACGCAAATTTTGCCTTGGAATAAAGGATTTCCCTTATATCCTCTGAAATCTAAGCAGAGGTTTCCAAACC  
TGAATTCGGTTTTCTGCATACCTGCAGGACCAACACCATGTGGAAGCTGCCAAGGAATAGGGCTTGCATCCTCTG  
AAGCTATGGCCTGAGCTGTACCTTGGCCCATTTTAGCCATGGTTGGAGCTGAAGCAGCTGGAATGNAGGGGACCAT  
GTCCCAAGCTTGTACAGAGCAGTGGGNCCTGACCAGCCCCAACTTTTGCATCTCTTAGTCACTGTTTGGAGGG  
GCTGCTGTGAAGTCTCTGACATGCCCAGGAGACATTTTCCCCATTGTCTTGGTGATTAACATTTGGCTTCTTGTTA  
CTTATATAAAATTTCTGCAGCAGGCATGAATTTCTCCCCAGAAAATGGGTTTTTCTTTTCTGTACCTCTTCAGGCT  
ACAATTTTTCCAAATTTCTTATTTCTCTGCTTCCCTTGAATGCTTTGCTGTTTAGAAATTTCTTGCGCCAGATACCC  
TAAATCATCTCTCAAGTTCAAAGTTCACAGATCTTAGGGCAGGGGCAGAATGCTGCCAGTCTCTTTGCATAG  
CAAAGTCACCTCCACTCCAGTCCAACAAGTTCCTCATCTCCATCTAAGACCCCCACAGACTGGATGTTATTGTC  
CATATCTCTATCAGCATTTTGGTCAAAGCCATCAATAAGTCTCTAGGAAGTTCAGATTTTCCCACATTTTCCCTA  
TTTACTTCTGAGCCCTCCAACTGTTCCAGCCTTACCTATTACCCAGTTCCAAAGTCACCTTCTACATTTTTGGGT  
ATCCTTATAGCAGCACCCCACTCTTTGCAGTGCAAGTTTATTGTATTAGTCCATTTCTCATGCTGCTAATAAAGACA  
TACCCAAGACTGGGTAGTTTATCAAGGAAAAGGTTTAAATTAACCTCACAGTTAAGCATGGCTGGGGAGGCCCTCAAG  
AAACTTAAAAAACTAGCAAAATGGGAAGCAAAATATGTGGCAGCTGGAGAAGTGCCGAGCAAAAGGGGGAAGAGATC  
CTTATAAAGCCATCAGATCTCATGAGAACTCCCTATCACCAGAACAGGATGAGGGCAGCCACCCCATGATTCAGT  
ACCTCCCACTAGGTCCCTCCCTTGACATGTGGGCAGTATGGAACCTACAATTCAGATGAGATTTTGGGTAGGGAT  
ATAGCCAAACCATATCATCTTTTATGTAAAATCTTAATATGCCCTGAAATTAATCTTTTGGATCTCAAATTCACT  
TAGACCATAAAGCCATGATGAACTAGTTTACTCTTTGTATTGTACTTTTTGCCTTTGTACCTCAAGAATTGAA  
ACCCTTCTGAGGACATTTGGTTAATTTCTGCCCTCTCTAGGCTGTGCCTTTGTACCTCCCTCCGATTTCTTACT  
TTGAGCACATAGAACATTAAGAGGACTTAGGAAGCTTCCACTCAGCTCTTGTGAGCCTCTGGTCTGTATTAGAA  
TATGTTTTTGTTTAACATGTCTTGTACAAGCCTCTTAGAAGGGTTTGACCATTACTACTTTCAGGGTACGCTTTTAG  
AAACAAGCATCACATATTTACAGCCTAGGAATGAAGATGCATCATGGGCTCATGTCTGTTGTCTTAGCTCACATAG  
TCTGGCTTTCAGTAGTCCAGGACTTCTAGTACTGCAGTTGGAGTTCAGGTGCCTGGTATGTTGAATAAGTGGTGT  
AAGCAAAGTACCTCACAGTAAAGGCTGTTTTCTGTAGATTATAACCAATCTTCAATGCTTTATTAATGCTTTCTGA  
TTCTCCTGTGATCATCTCCTTAGAAGTATCAAGACCTGGGCTTCTTATTCAGTATTTTGGTCTTTTGGATTCTC  
ACTGTTTATGCAAGCTTCAGTGTATATTTTTTAAACATCTCACAGTTTTCCATGGTTCCTATAACTAGAGACTATT  
TGTTTTCAGTGAACAAATTTCCAAAAGATATTCAGGAACTCGGGAAACACATATACTTCCAAAACACTGGGCAGAA  
ATTTTCAGCCCCCTCATCTGCCGCTAAGATAGACAAATTTAGACATTTCTGAATTACTACCTCATAAATCAATAAAA  
TTGATACAAAAATTAACAGAATTGATAGCAAGTCAATGGCATGAGGGTTGTGAGAAAATAATAAAACCGAAAATTA  
GTACAGTTTTCTGGAGGTTAGTTTATCAGTACACAATGAAAGTCATAAATTTGATTTAACCTAGGAACACAACCTTC  
TGAAATTTATCCAAGGAAAAGAACTTCAAGAAGAAAAGAAAATTAATGTGTCATATAGCTAAAAATTTGAAAACAGTCC  
AAAACCCAGTTTTTTCAGAAATGTCGACTATTACAATGGAATGGAATATTTACCACCAGAATCATTTGTCTCTAAATA  
TAATGACATTTGACCTTTGACATAATGAACGTGGGCATTTTGGGAACATATGGGAATTAATGAGGGAATGATCTAAAT  
CTGATTTCTCACTGTAAAACCTGAGGCTCATGGTTTTTCACTGTTGGGAAGCACAGTATGTCTTTGGTATTAAGAC  
AAATTTGGAAAATACAGTACTGTTTTTGTATTTAAAGAGAGGGAGATCCCCCTGTTTTAAAATAAGAACTGTCTTCCCT  
GCATTTTGCAGTCCAGTATTTCTAGTATTTCCAAATAGTACTTTAATCAATGGGACCTCTCTCTTTTTTAAAA  
AAATAAATTTGTTGTTTGGTAATGGTGAACATTTGTTTCACTGAGAAAGACCTGCCCTCAGAGGGGAGTATTGACCAG  
GCATCTTTCTTTCTGTCTGTCTGTCTGTCTGTCTTTCTAGCAGCTCCACGATATGAATGTGCTCCGGTGAACCTT  
TTTTCTGCAGGATCCGACTTTGCAGCCGATCAGGTCAGACCAATATGAATCCGATTTATTGCATATGAGTCTGAAAT  
CACTCACTCGTGCAGATACAAAAATAAACCTGGTTTTTACCAACAAGTTTTTATTTGTATTTTCGCTTAGAAGAAATAAA  
TTCATGATTTGGTATTTTGGATGAGAAAGAAATATTTTCATGAGAACTAGATAATAATGCAATAGGAAAAAGGTCC  
AGAAAAATAAATCTTAACTCTAGGAGCATATAAGAGTCACTAGATGTGAAGGACAGTTTAAACATCACCATTGTTTT  
GGGTTTTTTGTGTTTGTGTTTTGTTTTTCACTCACATTTAAAATTAATTTTTTGAACCTTTTAGTAGCTTTAAAGAA  
AGCAAGAGATTTATATTTGTGAGAAAGATTTAAAAGATCTAGAGACAGTGAACCAGAAGGAAATAGATGAAAATGT  
TAACAGTAGTTAATTTGCTGGTGGATTTACTACTCAATTTTTTATCTTCTAAATTTGCTGCAATATATTTATTTGGT  
TTCTTACGTTCTATACATTTATATTTATATATATATTTTTGTTATGTACATGACTTTTCATAATTTTTAAGATGAT  
GTTGTGATTTGCCAGTTTTAATATTTGAAGTTACTAAATGCTCTAGATGTTTLAGAGAGAAGAAAAAACCCATGAAA  
TAGTCTGATATTTCCAAGATTTCCATGGAAATTTGGGGACTTGGCTCATAAGATCTTGTATGCTTTGGGAGGGACAGCCT  
GAGGAAGGGCGCGGCAGTGGGAGGGGCTGTGTTTCAATTTGCAAGTAAGTTGTATGAAGATCTTAGAACAGCGCCTGC  
CTCTAGTAAGTACTATATGTGTTTTCTGATGTTAATAAATATTTTTGGGGTGTCTTCTTTGAGCCAGACACTTGGCC  
ACTGCCAGGATCCAAGCAGAGTTCCTTGCCTTATATACCTTAAATTCAGGAAGGAGGAGACCAATAATAAACACATA  
TTCAGATGCCTATGTGTTGGCTGGTGATAAGTTATGTGTAGACAGTAAAGCAAAAAGCAGGAGTTCGGGAAGTGG  
GAGATTGCTGTTGTATTCAGGGTTATCCAGGAGCTCACGAAGCCTAATTTGAAACAGCAGCCAGAAGGAGGGCAGG

GGGTGAGCCACACCAGTATCTGAGGGTGGCAAATGTCTGAGCAAGTGTGTTCCAGGCAGAGGGAGCAGCAGGTCCA  
GAGGCTCTGGGGTGGGGCATGCCCTGGCCGTGGGAGGAATGGCACGGAGCTGCTATGGCTGGAGCGGAGGAAGGAG  
CGAGGATGGAAGAAGATGGGGAAAGCAGGGACCATGCCACAGGGAGCTGTGACTTCTGCATGAAATGTTTTATTTAA  
GTTTTACTACTTATTTTTAAAACTTTTTACTTTTGAGAATTTTCAAACATAACACGAAAGAAGATAGTCTAATGAACC  
AATATGGTCTCATCACCAGCTTTTTAATAACTTGAATCCGTGGCCACTCTGGCCCTCATCTCTGCCAGCAGCAGTT  
CCTTTACTCCCTCTTCTCAAGTTACCGGGTAGCGCAGCTCAGATCTCATCTGATTTCCCTCCATCAGGATTTTCAGAA  
TGTCAATCTAAGAGAGAAAAGTTCTCATTAATGTTTGTGAGGCTCAAAAATGGTCCCTTCTTTGGCCAGCAGGAATCC  
CTTTGGGTGGCCCTGTACCTCTGTGAGAGTGGCCCTGTGGTATTTGACAGCTCCATTTGTTTTAAGCAAGAGACA  
CCCCAGACCCGTTTTAGACATTTCTTTTTTTTTTTTTTTTTTTTTTTTGAGACGGAGTTTCGCCTTTGTTACCCAGGCTG  
GAGTGCAATGGCGCAATCTCGGCTCACCGCAACCTCCGCCCCCTGGGTTCAGGCAATTTCTCTGCCTCAGCCTCCT  
GAGTAGCTGGGATTACAGGCACGCACCACCGTGCCAGCTAATTTTTTTGTAATTTTTTAGTAGAGACGGGGTTTTAC  
CATGTTGACCAGGATGGTCTCGATCTGTTGACCTCGTGATCCACCCGCCCTCGGCCCTCCCAAAGTGCTGGGATTACA  
GGCTTGAGCCACCGCACCCGGCCCTTTAGACATTTCTTAACCCAGGCTTGGAGTCAGTCATTTCTCTAAGGATCCC  
TGGTTCCATGTGGGAAATGGCATTATATTTGGAGGCCAAATCTATGTAATAGGGATGGTTATTTGCTGCAGGTTTG  
TCATTGCTTCTAGGAATTTTCAGTGAACAGAGGTAGGAAACATTTCAATTTTCAGTTTCCCCCACCTTTTTTTTTTAAG  
ACAGAGAAGGAAGGTCAAATAATGTGTGAATATAAAACACTACATTAATATTTTTGTTTTGGCCAGGCACAGTGGCT  
CACTCCTGTAATCCAGCACTTTGGGAGGCCAAGGTAGGCAGATCACCTGAGGTCAGGAGTTTGAGACCAGCGTGG  
CTGACATGGTGAAACCCCAAGTGTGGTGGCACATACCTTTAATCCAGCTACTTGGGAGGCTGAGGCAGAAGAATC  
CGCTTGAACCCGGGAGGCAGAGGTTGCAGAGCCAAGATTTGTGCTGTTGCACCTCCAGCCTGGGCAACAGAGTGAGAC  
TATTTTTTAAAAAAGAAATGAGGTGGAGTGAGAATGAACTGGAGAATCTCTGACAGTAAAGACTCCCTTAGGAT  
TGGTGTACACCTTTATGTGCTGCGAAAAGAAATGTGAGGTGGATAAAAGGTGATGTAAGTGGTGATTTATTCGTC  
TGGGTTTTCTGATTTTAGGTAAGTGGCACCTCCCAGCAGAAGCATCTGCTGGGCAGCTGCAGTGATAACACAGGG  
TGACGATCTTTGAGTCATCAATCAGAGCCCCCTACCAAGGTTCTTGGTGAGGAGTGGAAAGGGCATGCCCTCTGGG  
CCAGAGCAGGGCTTGTGAGCTCGGTTTTATCCTCAGCTTTTTCTTTCTTCTCAGCCAGATGCCTGTATGGACGG  
AGCGCCTGACCACATGTGAGGGTCAGTGATGAGGGCATCACTGAAGCTGGAGACGGAATGGTTTTTAGAGAGTAT  
AGAAAGGGATAATAAGAAATGGCTCCTGGGCTTGATGCAAGGGACCCTGTTTTCTCTTGTGACAGCTGATGTAG  
ATGTTGACATGAGACAGATGAGACAGCTATGCACTGCTCTCACAGCAGAAGGGGAAAAGTCAAACCTGGGAAT  
TCAGTGCAATGGGGTCAGGAGAAAAGTCGTGGAAACACTAGTTCTGTTTTTACACTCAGTAACCTAAACCAGCTCT  
TTACACAACCTCTTCTTGTAAATAATATCAAAAACCTTAAGCTAACATATTTCTTTCTTTTTGAAAACAAACATGTT  
TAAACATTTTAAATGTGGGTATGTCTCTGTTGGAAATTAACAGTAAAAAATTTCTCCAGCCTTGTTCACAGTGG  
GCCTGGAGGGTGCATGCGGGACCAGCACTAATATCTTTAATCTCCTCAGTGGAGGTGAAGACAGAGAGCAAGAGAA  
GCGTCCCTCCCCCTCCGGATCATCCCTTATCTGATTCATGTCCATCACCCTGCCAGCCAGAATCGGAATCAGAG  
CCTTGCTGTCTCACTGTGGTTGAAAAGATTCACTCTGGTTATGAAGGTAAGTCAGTAGATAAGATGCAGAAATGTC  
AGCAATCAGATAGGAACATGGGAAAGTCTGTTTTATGTCAGTCAGAGAACAAAAAAGAGCACTTGGGTCTCCAGTT  
TGACCCTTAAACTCTTCAGGGGTTTGAAGCTGTTTTTCACTGATGTCTGTTTACTGACCTTGGCTGCTTGTGTAGT  
CATTGCCTTCGTGGTCTGCATGGCTGAACGAATGGTGCAGCAGATGCAATCTTCTAGTCTCTCTCTTCTGGTGA  
CCCATTCAGTGTGGCCTTCTGCTTGTCTTGTGTGAAAGAAGAGTGTGAAATTTGTTTCCAGAGAGGGCCACAAGAG  
TGATTTTATGTTTTTGAAGTGTGAAACTCCCTGAGTAAAATGTATATCATATCTTTTTACGAGCAGAGCTATATTTT  
TAGTGCAGGTATTAATCCACAGGCAGCTTCCATGTATTTCTTAAAGTCAGTTTGGATATAATATGACATACTGTCT  
CCTACAAATCACCATCCTGTGCAACTGTCAACAGTAGAAACCACAGGGCTTATGGGAAAAGGGGTTTAGGAGCA  
CAAACTCAAAAACCTTAGTTGCACATTAATAATAGACAGGAATCAAAAATGATAACATAGTTTACACAGTGTGAA  
ATGATTAGGAAATATACAAACACTACAGTAAATATGGCTCATACCACAGGATGGACTGAGGTAACAGGTAGATGTT  
TGAGGTATATATGATGTTTTTGTCTTATTTCTTTTTTGTATTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGTTTT  
GTTCACTCAGGTGAGCAGATGTGTGATCGTGGCTCATTGAAGCCTCAGACTCCAAGGCTTAAGTATCTCTCTGG  
TCTCAGCCTCCAAGTAAATAGGACCGCAGGCCTCACTGTCAATGCCCCACTCATTTTATTTACTTATTTATTTTT  
TGTAGAGACAGGGTCACTATGTTGCCAGGCTGGTCTCAAATTCCTGGGCTCTAGTGATCTTCCCTGCCTTGGCC  
TCCTAAAGTGCTGGGATTATAGGTATGAGCCATCACACTTGTTTTTTGTGTATTTCGTAAGTACTTCCCTTCAACTGG  
ATGCTGTTTTCTGCATTCACTTAAAGCTTTCTTGA AAAATGAAATAGTGCATATGCAAATGAGATAGTCGCATTTGCT  
CAAACCTGTTCCCTAATATATCAGTCACTTCATTATTTTATTTATTTTTTTTTTTTTTTGAGTCTCGCTCTGTCCCCAG  
GCTGGAGTGCAGTGGCACAATTCACTGCAACCTCCACCTCCAGGTTCAAGCGATTCTCATGCCTCAGCCTCCCGA  
ATACCTGGGATTACAGGTGCCTGCCCAATGCCCTGCTAATTTTTGTATTTTTTAGTAGAGACAGGTTTTTACCTTC  
TTGGGCAGGCTGATCTCAAATTCCTGACCTCAAGTATCAGCCTCCCTTGGCCCTCCCGAAGTGCTGGGATTACAGG  
CGTGAGCCACCAAGCACTTCATTTTTAAAGCAAGCATTAGAACAGAACCTACCATAACCAATTTTTTTTTTTGAGATG  
AAGTCTTGCTCTGTCTCCAGGCTGGAGTGCAGTGGCATGATCTTGGCTCACTCCAACCTCCGCTCCAGGTTCA  
AGTGATTCTCCTACCTCAGCATCCAAGTAGCTGGGATTATAGATGTGCACCAACACTCTGGCTAATTTTTTGTGTT  
TTTACTAGAGACGGTTTACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACATCGAATATCCATATGGTACTAAA  
TAATATTTGTAATATTTTTGTTTGTGTTAGTTATATGCAGCTACTAAATCTGATTAACCTTTAAGTTATTCAGTT  
TTAACTATTAATTTTTTATGATAAGCTAGGCATGGTGCCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGA  
ATCTCTTGAACCTGGGAGGCGGAGATTCAGTGTAGCCGAGATTTGTGCCACTGCACCTCCAGCCTGGGGGATAGAGTG  
AGATTATGTCTCAAAAAAAGAGAGATAACTATTTAATGTGCCATTTCTTTCTGTCTCTGTTTTAAATTTGAC  
TTTTGTAGGCTTCATATTTGCCCTAAATAAAATGCAGTTAGTAATGAAATTAACCTTAACTGTAGCTTTGGCCGC

TTGAAGGCCAGGCCGCTAAATGAGATTTGTACCATTTTGTTCAGGATTTTCAGAGAGACCATTGAGGTTACAGAAG  
GAGTTTATTTTAGGTGTACACTGGCCAGCTGGGTCCCTGACTGGGAACAAAATGGCTGAGCCCTTATCAAAGGCA  
GGGCTTTACTTTTTATACGTGAAATACACAGTGGCACTAAAACAAGTTTGCAATTTATCAAGCAATGATAGGCACC  
TAACTTAGGCTTACATATGACTGTTGCTACGAGACCCAGATGTTTCGCTATCTGTCAAGCTCAAGGTTTCTAGCCCA  
AGCCTTATCTGTTGCTTCTCCTACAGTGGCCAGAAGGCTGGAGTCTGCTGCTTGGGCATTTCTTATGACCTTC  
GCTGTGCTACAGAACTTACAGTTATTGGCTATAGAGAACAAGAATATAAGAATTTAAAAACTTACACAACCTTGCA  
AAGCAGGACACAATTCATGGGATAGAGAAGGACTCAGGGAAGTTTGCCTTATAAAAAAAAAAATTTAATTTCTGCGT  
CTCTTCTGTATTGATTTTATTTTATTAATTTTCTCTTCAATTTTCATAAAGATTTTGTCTCTTTTAAATAT  
TTTAAGTAGTTAAGTGAATAATCAGATATTCACCTTTAAATAAGTCTCTTTTTTTTTTTCATAGCTCCTCGAATCCCA  
GTGAATAAAAAGAGTCTTGACCACAACACACACCCAGGGTGTGTTTTTCTTGAAGTAGATGAAAAGTAAGTACTTC  
TTTAAGCCTAAAAGAAATTTGTTTCTGGAAATAAATATAAATGTGAAGAAGTTACATTATGTTTGCATGTTTATAC  
ATTATATAACTGAAATCTATCCATTTACCTGTTTGTATTTTTGTTTACATGGAAACATGTTATCTAAATTTTTAATAT  
AAATATCCTGCATATATAAAGTGATTACAAGTTACATTTGATTTTCAACACAGTATACTTTAGGTAGGCCAAGAAT  
TCTTTCATTTTACAGAGGAAGACAACCCAGGGGCCGTTTGCCTGAGAACAGATGTGTAGTATCTGATTGTGCTCAA  
ACTGGGAGCAAGACTCTGGTCTTCTAGCACCTATTCTGGTGTATTTTCCACTTGACAGGCGTATCTCAAATCAGCT  
TAGTATGCAATTTCTGTTAGTCTGAAGTGGGGCAAAGGCAGAGAGAACAGATGATTTTTTTGTAAAGAAATGGTCT  
TTATTTTGAACCTTATGTCTTCTACTTTTGGGGGATGTTGAAATAGTCTTTGTTATTTAAAAGATAGTAATTGAT  
AGCCCACTTTGCGTAATGGCGCTTTGTAAAAATCTTTATTTTGGAGTAAATGTTAAGTATTCCTTACATTTCAGT  
ATTCCCAAGTGTTTTTTCCCACAAGTTAATTTTAAAAGTTTTACTTAACTCAATGAAAATTTAAAAATCTGGGAAC  
CCACACACCACACCAAAACATAACTGTAAAGCGAAAAAGTCAAAAACATTTTTTTTGTATCGTTGTGTGTTTTTTGTTT  
TTTGTTTTTTTTTTTTTTTTGGAGACAGTCTCACTGTGTCCACCAGGCTGGAGTGTAGTGGCGTAATCTCCACACACTG  
CAACCTCCGTCTCCCTGGTTC AAGCGATTCTCCTGCTCAGTCTCCCAAGTAGCTGGGATACAGGCAGCGTGCCA  
CCATGCCCGGCTAATTTTTGTATTTTTAGATGAGATCAGGTTTCCACCATATTGGCCTCAATCTCCTGACCTCTTGA  
TCCGCCCCTTGGCTTCCAGAGGGCTAGTATTACAGGCATGAGCCACCATGCCAGCCTTGATAGTTTTTTTTAA  
TTAAAAATTTTAAACATCTGAAAGTGAGTTTATTTGATATTTACTCTGAAACCAAATAGTAATGGGTGAGATACTT  
AATCTGTATAAAGTAAGAAAATTTAGTTAAATATAAATCTGAAAAATAGCTTAAATCTGTATTGATATTAGATCCCA  
GTTACTCAGGAGGCTAAGGCAGGTAATCTTGAACCTGGGAGATGGAGGTTGCAGTGAAGCAAGATCATGCCACTG  
CACTCCAGCCTGGCAACAGAGTAAGACTCTGTCTCAAAAACAACAAATAAAAAGAAAATGTTTGTGGGTTTTCTACAC  
ATGGAAAATACTTACAGAATCCAGTTTTTTTTCTGAGGCATCCTTAGTCTTACCATTCTAAGTTTCCAATGAGCCTT  
TATCCTGAAACTATATCTGTATACTAGCTGTTTGTCTATCTGATGTATCTTTTTGGATGACATTGTTTTTATGTAA  
TTTTAATTATGCTGTATGAATTAGCAACCTGCTTTATTCATTTAACATCATCAGCATCTTCTATGTTACGATATG  
CTTTTTGTAGACTTTTTTAAATGGCAACATAAACATAATGGAGGGCATCGGAGTCTCTAAGGAACCTGACATGGTT  
TTCTTTTCAAGATCTGATTTAGTAAGCTTACAGTGAAGCTTATGATTTGACCTCTCTGCAGAGCACGTCTAGAACTT  
AATGGAAGACATGGGCTCCTGGCTGCCCTAATCATGAGACTGATTTGTGCTAGAACACAAAGTTGATAATTAAGA  
AATAACAGCTGCCAAATAATATCATTTATATGGCTCAGAAAGGACTTCATTCCTTCTTCCACCTCCTACCTCCC  
AAGTACGGAAACCTATTGGGAAATGAACATGAAATCCAATGATTTGCTTGGTATGAACATAGGAAACCGCAGCCAGT  
AAGATGCCCTCATGAGGCTTCCATCTGGGGCTCAGTGTGGTCTGTGCTCCCCACCCAGTCAAGGTTAGAGATAT  
ATCCTAGATTTTTGAGGCTCTATTTTCAGTCTTTGATTTAAATATATGATCTGTATTTTTAGCAAGTTCCCTAGATGGT  
TCTGATGCAACGTGTGAGATAACCAACGTGTTAGATTACACAACCATTTCCCCCAGGATCGACATTTGGGTTCTTT  
TTTTGTTTTTATAAATGCTAAGATAAACAGCTTTGCATAAAATGTTTTCAAATTTACAATGAAATTAAGAAGTATTT  
TTAAAAGTATATAAATATAGGAATCAACTGATGAGTGTAAACATGCTACCGGTGTTGTGTATCTGTATCCCAGAGC  
AGTGCCTTTGCTGGGTTACCTACCTCAGGACCTGATTGGAACATCAATCTTAACTTACCTGCACCCTGAAGATCGT  
TCTCTGATGGTTGCCATACACCAAAAAGGTAAGGACCTACTTCTTTATTTGGAAGAAATGTTTTTCTATAATTGATT  
TGTTCTGATTTTTCTTTTCACTCATTAGAGCGCAAGTCTTTAACAGAGACAGTAGATTGAAACTGCTAAAGC  
AATTGTTTTCTATTTAAACATACTAAAACAGGCGCAGGCATGTTGGGTACACCTGTAATCCCAACACTTTGGGAGG  
CCAAGGTGGGCGAATCGCCTGAGGTGAGGAGTGGAGACAGCCTGGCCAACATGGTGAACCCATCTATACTAAA  
AATACAAAAATTAGCCGGGCATGGTGGCACACGCTGTAATCCCAGCTACTCAGGAGGCTGCGGCACGTAGAATTGC  
TTGACCCCTGGGAGGCGGAGTGCAGTGAACCAAGATCACATCCGGCCACTCCACCCCTGGGCAACAGAGCAAGGT  
TGTGTTTTCAAAAAAAAAACAACACTAAAACATTTACAAATAAGTGGCTAAAAAAGGACATTTGTAGTTAGTATCTGC  
ATTAATAAATCCCATTTTCACTTTTTTATTTTATATAGCTTGAAGTATGCAGGGCATCCTCCCTTTGAACATTTCT  
CCCATTCGATTTCTGACTCAAAATGGAGACTACATCATTCTGGATTCCAGTTGGTCCAGCTTCGTGAATCCCTGGA  
CCCGGAAGGTTTCTTTTCACTATTGGTTCGGCATAAAGTTTGAACGTAAGCTAGTCAAGTTTTTCATTTTTTCTAAAACA  
TACCTTGTATCAAAATAACTCCATAGCTTATTGACTGCCTTCTGACTCAGTTAGTTGACACATGAACTAAATAAA  
GCACAGATGTTTGGTGCATTTGGAAGATAGTTTACAGTATGTACCATGAGCCTTAAAGTAAGACATAATGTCATC  
CTTAAAGTAGCTTAAATGTCCATTTTTTAAAGTGAACCTCTGGAATATTTGCTTGTATGATTTATATATTGGTCAATGA  
AAATTTTTTGCAGAGATCACATAACATGGAAAAATATTTGGCACATTTAAATGTAAAAATTAGCCTACAAAATTTATAC  
TTAGAATTATAATTATGTTTAAAAACATAAACACACTATGGAATAAAGAAAAGAGAATGAAAGGAAATAAAAATAGG  
AATTTTGGTTAAGTGTAGGTACTAGGACTTATAGGGTTTTTTCTTTTTTTTTCTCTCTCTATGTTGAAGATCTCT  
TTAGTGAGCATGTA AAAACTTTTTATATCAGCCTTCTGCCATCAAATTCACCTTATTTATGTCTATTTTTGTTTAT  
TTGGTTTTTTTTTGGAAAAGGTTCTCCCTCCTGTTGCTCAAGGGTGAATGTAGCTAAGTATTCAGCTCAAGCT  
ATGACCTTGACTTCTTGCTCAGTATCCTCCCACTTCAGCCTCTGATAGCTGGCCACAGGTACCTCCAACCAT

TCAGTAATTCTGTATTTTGTAGAGCTGGGTTTTGCCAGTTGCCAGGTAGGTCTCAAACCTCCTGGGCTCAAGCATC  
 CTCCTGCCTCAGCCTCCCAGAGTGCTGGGATTTTGAAGCTGTAAGCCACCACCCAGCCTGTGTCCATTTTTTTAA  
 TATTTCTTTTTCTGCCAGCATAAGTTGACACCAGCTTGTGGCTTGCTGTTTTCTGTGAGTGTACATAACAGCATC  
 ACGTGGAAATGGAAC TAGACCAGCCCGCCCTCTACTTCTTGTGCCTACGGTACTCCCAGGCCCTTGAGTTTTTGGT  
 CTTTTTTTTGAGACAGGGTCTCACTCTGTCAACCAGGCTGGAGTACATGGCACAATTGCGGCTTCCTCAGCCTCTC  
 CAGTAGCTGGGATTACAGGTGCACACCACCACCCTGGCTCATTTTTTTTTTGTATTTTTTAATAGAGACAGGGTTTCA  
 TCACATTGGCCAGGCTAGTCTGAAACTCCTGACCTCAAATGATCTGCCACCTTGGCCTCCCAAACCTGCTGGGCTT  
 ACAGGCATGAACCATCATGCCAACCCCTTTGGTCCATTTTTTAATTGAATTTCTAGGCCCTTTGACCTGGCCCTAGGT  
 TTCCCTTCGATGCTTCTTTTTTAAGCATGGAAATTTTACTTGACTCCAGAATCATGCTGAGATGGAGAAGTTTGGT  
 TTTTTGGTATCTGTCTATACAGCATTGTCTACGTGAGGAAGAACTGAGAACAGTAGATCTACCAGTGAGACATAT  
 TTGATCACCTGTTAATATTACATTAGAGATGATGCAGAGCCATGGTGAGAAGAGGTCTCAAGGCTGCTCTTACCTC  
 TTTCTGCTCCAGGGTTTTGGCCTCTCCGCAGAGCACATCTAGAACTTAATGGAAGACATGGGCTCCTGGCCACCCTA  
 ACTCATGAGACTGATTTGTACTAGAACACGAAATTTTATAAATTGAGAAAATAACAGTTGCCAAGTAATTACCATTTGT  
 ATTGCTCAGAAAAGGACTTCTCCCTCTCCCCACCCCATACCTCCAAGTACCATAACCTATTGGGAAATGAACAT  
 GAAATCTGATGGTTGCTTGGCCTGAATGTAGGAAACCACAGCTAGTAAGATGCCCTCATGAGGCTTCCATCTGGGG  
 CTCGGTGATGTTCTGTGTACCCAGTCAGGGTTAGAGACATATCCTAGATTTCTGAAGCTGTCTTTTCAGTGGTTT  
 GATTAATAATATATGTCAATTACCCAACATATAATGAATATTCAGAATGTTTTCTATTATTTTTAAAAATCACACTGG  
 TGTTTACATCTTAACAACAAAGATTATGTGAGAAAAGGATTTAAGGCTAGTAGCTATTAATCTTGCTATAAAAATA  
 ATAATCACATATTTATAAAGTATAAGGCTCCGTGTAAGACAGGGTGC TAGGCCCTGGGAATGTGATGGTAATCAGA  
 ATGTTCTCTTGACCATCAGGGAATCCACACTGTCATGGGAGAGACAGTAAATAAAAAATGGCACTGTTGTATATAAC  
 TAGAGTTTTGTTAAGTGTGACAACCTGAAATGGGAAGGGGGCTGCGAGACTGTTGAGCAAGGGACCCTCCCCATGTCT  
 CTCTCAGTCTTGGCAGCAGTGCTGCAGGCAGGCTGGTTAGGTGTCTCACCTCTCATTTTACAGGTA AAAATCCAT  
 ACTCAGTGAAATAGAGTAACCTTCCCAGGGATCACCTGGGGAAGAGTTGGAAGTGGAGACTGCAGCCAGGTTTTCT  
 TTCTTACTCTGTTCCTTACTTTACATACACAGCAATGTATATTTTATGTCATGGCATTGTTGTTGCTATTTGATGAAGGCTA  
 TCTCCTTATTGCCTGCAGATTAATCATTTACATTAGATAAATTACTGATTTAAAATACACTTTACTGCCTCATGCT  
 TATAATCCCAGCCTCTATAAAAAAATTTAAAAACAAAATAGCTGGGCATGGGGGTATGTAACATAGTCCCAGCTAC  
 CTAGGAGGCTGAGGCAAGAGGGTTACTTGAGTCCAGGAGTTCAAGGACGTAGTGAGCCACAATTGCCCTACTGCAC  
 TCCAGGCTGGGCAAAAAGAACAGACCTTGTCTCTTTAAAAAAAAAAAAAAAAAAAAATTTGCGCGCGGACGACGACGC  
 GCGCACACACACAGACAGAGTTTTGCAGAAATTTTACAACTAGGAATATCTAAATCTCATAGTATAAAGCAATCT  
 GGTCTATTTCCATTTTACTCAGAAAAAAAATTTCCAAATACCTCTGAAAGGTTTTCAAGCTCACAATAAATGTACAG  
 GATGTGTTTATATCAGGAACCAGCTGTACTTTTTATTCCCAGATCAGTCAATCATGTATTTGTTTCTTTAACAATA  
 TGTAGTGGGTGCCTGCTATGGACCAGACACCATTCCAGGCACGAAGATCCAGGAGTGGAAAAAGTAACAAAGTA  
 CTTACCTTGTGGAGCTTACGCCGAGTGGGATATAGGGAGAGATGGTTACATAAAACGTATAGTGTGTGTCAGGTTG  
 TGAAATCAAGTTGTATATTCAGGTTACTGCTTGGCACAGACTTATGGTATACCTGACACCATGCAGATAAATATTGA  
 TCTAGATATAGTCTAGATGCGAAGGGAAAGAAAGGTGATTGGTGCAGCTGTTCCACTGTGATGGGGGCATGTGCAG  
 GGTAAAGTGAATATTCCTATACATGGAGATATTTATCCTGCTTTTATGACACTGTGAGCATCCTATGATTTTCCCTG  
 ATAATAATTTATGCCTTTGGGAAACTCAGTTTTTTAGGTTACTGAGCTGCATATATCAGTTACTTGTGTAACCGGAG  
 GCTGGCATAGTGATAAATGACAGCTCAGTTTTTGTGGTCCAGACTAGGCGGACTCAAATCCTAACTATCTGCTGCTT  
 CTCAGTTATGTGAACTTGGGCCAGTTCTCGGGATGAGACAGTCTGACTCATCCCCGACTTTGCATATCAGGGAT  
 AATTCCTAATTTCTGGGCGTTGTTGTGAGGACTCTGTGAGCTGATTCAGCTGATTCAGTATGCCTCATGGCATACT  
 GAATGTTGGTATGTACATGGCACATGTACATGGCAGTACAGTGTGGTGGCACACAGTGTGGACTCTAGCTTAGACT  
 CTGGTCCCTTCGTGAACACCTGCCTCCCCCTTACCTGTTAGGCCACTGGAAAAGGCTCTTAATCAGGTAAGGAG  
 AAAATACTGTAAAGCACTTGAATAATGAGGGTTTTCAAATGTTAGGTATTTGTATGTTATAAAAAACTAGTA  
 TAGTCCGGTGCAGTGGTGCACCTCCTGTAGGCCAGCAGTTTGGGAGGCCAAGGCGAGTGGATCAGTTGAGGTCAG  
 GAGTTGAGACCAACCTGACCAATGTGATGAAACCTGTCTCTATTAAAAATACAAACCTTCGCTGGGTGTGGTGG  
 TATGCGCCAATAAGCCAGCTACTCAGGAGGCTGAGGCAGGAAGATCAGTTGAATCTGGGAGGCAGAGGTTGCAGT  
 GAGCCAAGATTATGCCACTGCACCTCCAGCCTGGGTAACAAAATGAGACTCCATCTCAAAAACAAACAGAAAAGGGC  
 CAGGCGTGGTGGCTCACACCTGTAATCAAAGAACCTTGGAGAGGCCAAGGCAGGTAGATCACAAGGTGAGGATTTG  
 AGACCAGCCTGGCCAACATGGTGAATTTCTGTCTCTACTAAAAACACAAAAAATTAGCCAGGCATGGTGGCGTGTCT  
 CCTGTAATCCAGCTACAGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCAGAGGTTGTGGTGGCCAAAG  
 ATTGACCAATTGCATTCAGCTATTACACTCCAGCCTGGGCAATGGAGTGGGACTCCATCTCAAAAAAAAAAAAAAC  
 TAGTACAGTGCCCTGTACATAGAAGGTTCCAAGTATACGTGAGTTTTCTTCCCTTCTTTGTGATCCTCTCTGTTGAC  
 ATGTGTGCCATTCACTGTATTAGGTCACATATTTGAAGATGCTCTAGAATTTGAGTAGAAAATAAGACTGGAAATTTG  
 TAACAGCTGCATTTTACCACCTGGGGTGGGTTGCAGAACCCTTAAATTAGATAGCAAAAAGTGTAGGTTGCATTA  
 ACGGTAAAATGCCATTTGGCAAACTGGGGTCTTTAAAAATCTTTACAGTAAATTAACCTCAGTTACTGTCAATTTTTGA  
 AACCAGTTATTTGCTGAAGATATTTTAGTAGCAAAGGAGGTTACTTGTCACTAAGTTTAGATTTGTACACATATAG  
 CACTGAAACCATTATAGAAAAGATGACTGCTTTAGAGAACTTCACTGGCCTGTTAATGTCTCATTTGTTTCTTAGT  
 ATTTTCTGAACCTCCATGAGTAGAATTTAGCCTCCTATTTTTCTCTTTGGAGTCAAAGAATTTGTTTTGTAAGCA  
 TAGGCTATTTCTGGATAGCCTATTAATAAATACTTTCAAAAAATGTTGCTCATCTCTCACTGCATTTTCTAGGAG  
 CCCACTAAATGAGGATGTTTTTGTACCCAAAATAAAAAAGATGAACAATAATGACAAAGACATAACAGAATTACAA

GAACAAATTTACAACTTCTCTTACAGGTAAGTGTGAAATTTTTAAGAATGCAAAATTCCTCTGAATTGTGTTTTGT  
TTTTAAAGTGCTCAGTAAAATCACTTCACAAAAATAAATTATAAAGGGAGACAAATCTTTGCCCTCTGCATTCCAGA  
AATTTTTGTCTTTCTCTTTTTATTTTTAAATTTACATTTTGTCTTTTTTCCTTATCCTGCATTACACAGAATATTTT  
TGTAAGAACCCATTTTTAAGACAGCAAACCAAGTGACAGAGAATTTCTTACATTTCCAGAGGCACACTAGTGTGCTCAGCC  
TTTGGGTTTTGGCTTCAGTATTTTTTAAATAATTTCAATTTTTAAAGATGGAAGGAGAGGGATATAGGTCTCAGTTT  
CTCTTAACTACTGTAAGGTAAAAAGTGGAAAAATTTATGTGTTGAAAAGAAATACCAAAGGTATTTTTCTACAAGAGA  
TTCAATGAGTAAGTGGTAAATAAACTAATGGCTTCACACAGAGTAACTAACAGCCCTCAGTGGTTTTTCTCAGAATGT  
TGCCCATAAAATCTCTAGGGTGCATGAAACACCTGCCATGTTCTTGGACAAAGTATAATATATAGTTTGTCTCTCTC  
TGTCTAAAACCTGGGCACAAAGAACTAGACTTTTATTCTTCATGTTTTTTCAACACATTAATTTACCTTATGGGGTTTT  
TTTGACCAATAGCTGTGCTTACACAGTACTTTTAGAAAACAAACATTTACCACCTCATTGTTCTGGTCCCAAGGT  
GGCCCTGCCTACCTAGACTAGAACATAAGACCGTTCCCTTGAGAATTACTGAAATGACGGAGGCATCCCTTTGGTG  
TGTCAGGACCAACCTGCACACCCTAATTTGAGATTTTAGAAAATTGTCATCTTAATTTTTAAGATGATTCTAGATGAG  
CTCTGAGGTGGCTGCAATTAGGCACCCAGCAATTTCTGAAACTGTTCCCTCTGCTTCTAGCCAGTTCACGTGAGT  
GTGTCCAGCGGCTATGGGAGCCTGGGAAGCAGCGGGTTCACAGGAGCAGCTTGTCCAGCATGGCCCTCCTCCAGT  
CCAGTGGGCCTGCGCGGAGGAGACAAAGGCGGAGCAGGTGCGTGGGCTCATGTCACCTTCTTGGCCAGGCATCGC  
ATTCTCTGTGTTACTGCGGTTCTGCACGTGGTGACATCTTAGTATATGTTCCCTGACTTAAAGACCTCATCTGATAA  
CAGATATTTTAACTCACTTTGGTTTAGAAAAAATGCCTGTAGGCCAGGCACAGTGGCTCATGCCTGTAATCCAG  
CACTTTGGGAGGCTGAGGTGGGCAATCAGGAGGTGAGGAGTTAAAAACCAACCTGACCAACATGGAGAAGCCCT  
CTCTACTAAAAATACAAAAAATTAGCCGGGCATGGTGGCATGTGCCATAATCCCAGCTATTCAGGAGACTGAGGC  
AGGAGAGTTGCTTTAACCTGGGAGGCAAGAGGTTTGCAAATGAGCCCAAAGAATCGCGCCACTGCAACTTCAAGCC  
TGGCACGAAAAGAGCGAGACTCCCATTCTCCAAAAAGAAAACAACAAACAGTCTGTAACAAAGAAAATACCCACCT  
TAACAGAAAAGTGATATTTGCTTTAATTTGTTATGCTCAATTTGTATGCATTTCTAAAAATTAGCATATTTGTGAACAGAT  
TTATTTAGAAATTTGTTGGTACCAATATACTGTGGTGTGGTGCAGGAACTGAATTAAGCAGTTAATGTGCTTTAG  
TCATTATGTCTTCAGAGAAAACAGCTAAGTGTATTTTGTCTATGTATCTTTTAGTGGACATTTTTATAATTTTTTA  
AACTTACTTTTACTCTCTTTCCCTTAGATGACCTTGCAGCGGCTATGCCAGTGTCAACAAAATTAAGAATCGA  
GGTCAGCAGCTTACATTTGAGTCAAGGACCAATCATCAGTCAAGCCAGTGCAGGGGATGTGCGCAGAACTGAAT  
GTGGTGTGAGTCAAGCAGGCAAGGCTTCCCTCCATGGGCTGCCATCCCTGTGTAAACATCACACATA  
TGATTCATGAGCATAAAGCCATGACCCCGTGTGGTGTCTTTTCATATTTCTTGATTTGCCTAGGTAGCAATAGTAA  
TAGTAGTGATAATAAAGGCAGCAGCAATACTATTTCTGAAAGTAAATGGTACTAGAAATGCATTTGTATTAATAGTT  
GCAGTTTGGCTTTTACTGAAAAATTAACCTTTTCTGAAAAATACCAACAAGTTACAGATTTTCCCTCCCTACATAGGCAT  
CTCTTTAAACAGAGTGGATTTTAGAATGCATTTTAAATAATGAAGTATATTAATAATAATCCCTTTCTGTCAAT  
TTAAACAGAAAATCTTTGTGACCTAAATTCAGTGAAAACTTTGGCTTTTTTCTTGACATATATTTACTTTTCTCA  
AATACGAGGTAAGGAGCACTCCGAAAGTACTGGATCAGAGTAAATTTGGCATCATTTTGAAGGTTAACCTAGTTG  
CACTGTATGGCATTGGGAAGTTCAATAGTATAGAAATCAAGTTAGATTGACATTTAAATGTAACATACTTTCCCTT  
CATTTTTCTTCTGAAATGTTAATATAATTCGGAAGTTGTGATGGATCTTGGTCACTAAATGTTATTAATCTCTATC  
ATTGTCAAATGTAGACATTTTACACAGTAGGGGTGAAATTTCTAAGATTAGTAATGATAACAGCTGTATTTTTTCAG  
AACTTGACATAAATTGACAGCACAAAAATCAAACCTGCAGCTCCACAGGTGGCATCACTGGGAACCTTCTTAGTTT  
GTACTGCTGTAACAAAAATACCAAGACTAGGTAATGGATAAAGAACAGAAACCTATTTCTCACAGATTTGGAGGCTG  
GGAAGTCCAAGATCAGAGCACACCAGTTTCAATTTGAGACCAGATGAGGTCTTTCATGGTCTGAACATTTATCCTTGG  
GAGGGGACAAACGCTGTGTCTTACATGGCACAAGGGACTGAAAGGCAGAAAGATAGTTCCCTCCAGCCCTATTGA  
AAGGTTACAAATCCCATTCATAAGAACAGAGCCCTCATGACTTTGTCACTTTGTAAGGCCATGCCTGTTATCACTG  
TCACATTTGGGTTTTAAGTTCCAACACAGGGATGCTGGAGGGATACACACATTTCAAACATATAGTAGAGCCCACTGGA  
CTCCGACATAGAGCAAAATCCTGTCTTTTCAATTTTTTTATTTTTCTCTTTTGATAGGATGTTAAGCATTCTATTGGTT  
TCCTCTCTGTGTACATTTAAGGGCACAAGATAATAAATACAGCTGCTTTTTGTATTCTTTGCTCCCTCCAGACCCCA  
CATTTCCCTCTGCAAAAATAATTTATCTTAAATACAATTTTTACTATAGATCATTTATATGGAATAATGGGTCCTAG  
CAGAGAGAACTAGGACATGGGTCTAGCAGAGAACTCTTTTTAGGAGGTTTTTACAGCTAAAATGTGTAATGTTT  
TTCTCATTTGTCTGGTCTGCCAGTCTACTCACAGCTGCCCTGGGCCCTCAGATCACGTGGGGGCTCCTCGATGCTCAT  
GGTGGCATCACAGTTTTTACATACTGACCTTGGTCTCATCTTAGAATGGCTTCCCTTCTCACTTCTGCATATTTCA  
GTTTCTCATCATCTACCATATATTTTTGTAATGCCATTGGTGTCTATTTTTTTTTTTGAAGGCTGTATCTTTAAGTAGAT  
GGCAATATTCATCACTGACTTTTTATAAAAAATTAAGCTGCAGTGCCTCTGGCTACATATAGCCACGATATCGTGA  
AGCACACGGATGTGACTCTGGGGAAGAAGGAGATATCTTGCCATTTGGTATAGCTTTGGAGTACCAATGTTTCTG  
GGAGATCTTTATTTTATTAACATTTTGGAGCTTACAAATTTAAAGAGCTATGGAACCTTTAATACTCTTTACTTG  
AGAGACATGAGGATGCAATTTTAAAATGTCATAATTTAAATGTTAGAAATTTACTAGAGATATATACCAATGAAAA  
GGAAAATGCTTAGTTAAGCATGATTATGTTTACATGCTTTACAGGAACTTAAGGTTTTAAAATTTATATATTTACC  
TTTTCTTACAATATAAGAGCAAATAATGTTTGAAGTAGAATAAGAAGGGAGATCTTAAGTTAATATGGGAGGGCA  
AGGAAAAACAGGGATATGTAAGCCCTTTTACAGCTACGTGTACAGAGGCCAACTACAAGAGACATCATTTGAGGAAAT  
CCAGGAAAAAGTGAAGTGTCCATGTCTAGGGTTCATGGTATACCTGAGTTAGAGAAGATGACCCATAAGTTGAATAGAT  
TCAGAGCAAGATTACTTCCAGTTTCTGCTCCAGGATGTAGAGAGCCAGAAATAGCATCACCCCAACCTACAACCTGA  
AAAATCTGGATGAAAGGCAAGTTCAAGACCTTTCTTGACCTCATAATGCTGAGGTTACAGAGCAACCAACAGGCCT  
AAAATCTAAGGAGAGAAACATATTTGCAGAGAGGAGCAAGATGCAGCACTGGATTACTGTGCACCTGAGACTGGGT  
TGTGGGGCACCGGGAAGGCTCAGCTTCCAGATAGGATGTGTTCCCTCAGGCCCTTGTCCCTCACTCACCCAAGAAT

GGGAGAGAGGACCCACCAGACATGGTGAGCTCGTTTAAAGTAAATATTGGACCCCAAAGAGTTTTGCAGAAAGTG  
ATTTATTTAGGCAGAGAGAAAAGGAGAAAAGCAGCTCCCACGATGTTGTGGTAGGAGATCCAGAGATGGAGTC  
CTCCCAGATGAAAGAGAAGGGGGCTTTTAACTGGGAGTTATTCCTGCCCTATTCAAGTTTTCTGTCTATGAAAG  
GAGTTCTTTAACTTCTGTCTGGATTCTAACCATTTGATTTGTTGTTTGGCCAGAGTCTCCCCTAGCTTTTCATG  
GGTTCTAGACAAAAGGAAGCTTACTTGGACAGTCCATCTCCCACGGGAAAGTTAAACAAAGACCTCTAAGCTCCCA  
GATGGAGACATGGGTGGAGGTATGGTGCTGGGAAGTTCTTGCCTTTGAATCCACGCCCTTTTGGACCTGGGAAG  
AGAATACATTCCTCACACTGATGGCCTGAGTCAAGTAATTAGCAGAAGACAGACCTCTCCAAGGCACAATGCCA  
AGGAAAGACCAGAAGCTGAAAGTGGAGCAGATAATGAGAAAAACAAAAAGTTATCCACCAAACCAGCCACCAAC  
TGAGGCAAAAAGTGAACCTAGAGGAATTTGAAAGCTGTTGTACATGCTATAAAAACCTTATAAATACTCAAACCTGG  
CCTAAAAATTAGATTAATTCAAACACAGGTCTAAAAGGAAAGGCATCACCAGTCTCTGTTGTCTTACACAAGTTAT  
GCAGTTCTCAACAAGAAATTTATGTGGCATAATGGATAGGTAAGAAGGGGAAAAAATCCCAAGAGATAAAACCATCA  
TCAGAACGAGACTTAGATAGGACACAGAAGTTTGAAGTATCTGACAGGGAACCTGAAAATATTATGAATGATGCATG  
CTTAAAGACTATTGACAACTGGGCACATACAGGACCAGGTGGGCAATCACAGCATAACAGATGGGCATGAAAGAAATA  
CTCAAATGGAAATGTAACATCTGTAACATCAGTGCTAGAGATAAGGAACACATTTGATGGCACTCAGCACAGTCTA  
GAGAAGAATCATTATTTAAGCTGAAATAGAGAAAAGAAGTGAAGCAAACAGAGTATTCTAAGAATTTGGAGACAG  
CCTCAAAGTATCTAACATAGGTATAATTTGGAATCTCAGAAGAAATATTTGAATAAATAACAATAATTTTCTACATA  
GCCAAAGAATATCAAGTAAGTTAAATACAAAACAAAAATAAACAGAACACTTGGGCATATCATATACAAACTGAT  
GAAAAAGAAAACTTTGAAAGCAAAACAGAGAAAACAAAGATAGGAATTTATGACAGCCATTTCTTCAGAAACATGCA  
CACTAAAAAGGTGATAAGAGTAACATCTTCAACCAGAAAACTGTCAGCACAGAATTTCTGTACCTAGTGAAATGA  
TCTTTAAGAAAAATAAGGTGATATAAAGATTTTCTCAAACAAAAATCATGGGAATTTATTTCCAGCACACTTATTC  
CATACGAAATGCTAAAAGGAATTTCTTAAAGACAGGAGATACGATACCAGACCAAAAACCTTGGATCTACACAAAGAAG  
TAAAAATGCACTGGAAATTTGAATAAATGAAGATAAAAAGAAATTTCTATTTTTTAAATTTTTTAAAAATACTCATAAAGGT  
AACAGACTAAAGCAAAACCAGCAAACTAGATTTATAGCATAACATAAGAGAAATCTTTGTGACAGTACTGCTACAGA  
GGTTGGGGGATAGGCATTTGGGAATATACCATCATTAGGTCCTAATACTGCATTTGAAACAGTATTTTATTTATTTGG  
AGGTGGACACTAAGTATTTAAGAATACATATTTGGAGCTTTTAGGAAAACAATTTGAAAATATTTTTAAGTATGATA  
ATAAGTTAATAGAGGAGGTAAAAATGGAGTTTAAAGCACTAACCCAAAAATGTGGCCAAAAACCAATGGAACTGAAG  
CTGGGAAGTCTAGTACTCAGAAGACTGAGTGAGAGGATCCTTGAGCCAGGAGTTCTAGTTTCAGCTGGGCA  
ACATAGCCATACCCCATCTCTAAAAATACAGTTAACAATAAACAGCAAATGGAACAAATAGAAAACAACCTGGTGGC  
TGATATATTGTAATCCAACAATATCAATAATCACTAATGTATTGAATATGAATGTGATTTCTTTTAAATCTCACAGTA  
ATCTTTTCTATATAGATTATCCCCAGGACACATATTTACAGAATGGAATTTGGTAGAAATGGATGAGAAAAGAAATCC  
ACCTACTTCTATTTTTTACAAAAACCCATGGTAAAAATAAACAGGTAAAAGCAAAAGAGTAGAAAAGATATACCAT  
GCTAACACAAATCAAAAAGAAAGTGGGAAGCTGTATTATTTATTTATTTGTACAAAGTAGACTTCAGAAAAAGGAAAGTT  
ACCAGAGTTAGAGAGTGAGATTGTATAACAATTAAGGGGTTGATTTCCCTAAGAACATAGAACAATCTTACATGTT  
TATGAATCCAATACCAGAGCATCAAAATACATGAAATACAAGCTATAAACTGTAAGGAGAAATAGACAAAGTCAC  
AATGATAATTGGAGGACTCANCACATTTTCTTTTTCTCTCTCTTTTTTTTTTTTTTTTTTTGAGATAAAGTCTCACTCTG  
TAGCCAGGCTGGAGTGTAGTGGCGCAATCTTGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGTGATTTCTCCTGC  
CTCAGCCTCTCCAGTAGCTGGGACTACAGGCACCGGCTAGTTTTTGTATTTTTTAGTAGAGAAGGGGTTTCACCATG  
TTAACAGAAATGGTCTCAATCTCCTGACCTGTGATCCACCCGCCTCAGCCTCCCAAAGTGCTGGGATTACATATGC  
GAGCTACCACGCCTGGCTAGACTTCAGCACCTTTCTGTAACCTGATAAGAAGTAGTAGTCAGAAAATTAGTAATGCT  
ATAGAAGACCTGAACAACACCATCAACAAGGTCTGCCTATTTGACATTTATAATAAAGAACAATTTGACCCGAACAAC  
AGAATACATGTTTTTTCATGCAACAGTCACTCAAAAACAGATCATAGTATGGCATAATTACAAACCTTAAAAATTTTTTA  
AAGAAAAGAGATCATACTCTGTTTATACTCTGACCATAATGGATGAAACTAGAAATCAGTAACAGAGGTATCTAGA  
AAATCTCCATATGCTTTGAAATTAAGTAACTTCACTATCTAAAATAATACACGGGTCAAAGAGGAAGTTTCAAGGG  
AAATCAAAAACAATTTGAACCTGAACAAAATGAAAAGACCACATAAAAATATGTGGGATACACCTAAAACAGTCTTAA  
AGTGAATTTTACAGCATGAGTGTATATATAGAAAAAAGAAGATGGCCTCAACTCACTTATCTAAGCTTCTCCTCTT  
AAAAAGTAAAAAGAAAGAAACAATTTATATCTCAAGCAGTAAAGAAGGAAATAATAAAGACCAGATATCAATGTAAC  
TGAAAAAGAGAAAATAAGTTAAATCAAAGGGTAATAAAAATTAGTAAGTCTTTTGGCCAAACTGAAAGGAAGGGCAGGTT  
GGGCTTTATCAAAAATTTAAAACTTCTTTTTTGAATACTGATAACTATGGCCATATCGCCCTGAACGTGCATAAT  
CTCATCTGAAAGACTGATAAAAAGACAAAAGACAAAATATTTGCTGATTATATATACAATAATGAAGTTATATCC  
AGTGATACATAAAGAAATGCCAAAACCTCAAGAAGGCAAAACAACCCAATTTAGAACATTTTTCTGAGGAAGATACATAA  
ACAGCAAATAGCATACAAAAAATGCTGAACCTAAGTCAATAGACAAAAAAAATTTAAACCATATGAAACACCACC  
ACACATCGGTGAGAATAGCTAAATATTTTTTTTTTAAATTTGACAGTACCAAGTATAGGTAAGGATGCAGAGGGACTG  
GGGCACTCACCTATTTCTGGAGGTAATGCCAACTGTATAGCCACTTTGGGAAAGTTTGGACATTTCCCAATAAAGT  
CAAACATACTCTACCAGCTATTATGATTTACCCAACCTGAAATAGAACAAGTCACACAAAAACCTGTATGTGAATAT  
TATACTGTTACTCTCCAAACACAAAACAATCCAAATGGCATTTCAGCAGGAGAATGGATAAACTGTATATATCCACAC  
AATGAAATGCTACTCAGCAGTGAAGAGAAAAGAACTTGATTCACAGAGCACGGTAGACAAAAATCTTAAATACATTT  
TGTTAAAAAGAAAGGCAGAACCAAAAATGTTACATATTTGTACAATTTCCATCTGGAAAAGCATTTCTTTCTGGAATGTTCT  
AGATCTGGAAAACAGATCAGTATTTACTATGGCTGGGTGGAGGAGGAGTTGACTACAGAGGGCCAGCTGTAGGGAC  
ATTTTTAAGTCATGGAAATGTTCTTTTGGTTACTGTGGTGATGGATGCACAACCTTGCTTTGCATTTGCCAAAACAA  
GATTTTGAATTCGTGCACTCCATCTCACCCCAAAATGCAGAGAGCACCTAGGATGAGATGCAGGCTTTTGCACAAG  
TGAATCCGAATGCATTACAGGTGATCAGCCTAAACATACGGAAGGGATGGGGAAGAACAGAGCTCTGTAGCTTTGG

AAAACAGTGTCTGCAACTGGATATCCTAGAGCTAATACAAATTATGACTGTGCACAAACATTTGTACTCAAGTTGGTG  
TATTCGTTTCCCACCAGAGTACAGGTTAGCAATTTCTAAAGTGACTTTTATACTCGGGTTGAACAAATCATGGAGGCT  
GCTGTAGGTAACGATAACCAGCCTCCTCACTATCAGAGGAGAGCCACACATAACCGAATGGAAGGCACCTCCGTGGT  
GCTGGATTGGCGCTCACGGTTTCAGTGTGGACTCATGCCCTTTTGTTTACATACAGAGGGATGGACAGACACAGG  
TTAGTATACTATGCCAATAATGAGTGAAGACTAAAACAGATGGTTTCAGGCAAAGATGATCAAGGAATGAAACTATA  
GGTGAAAGGTTAATGGGGAACTTAACTTTTCATTAGAAGGTGGAAGTTTAAAAGAAACCGAATATTTGTGTGATTT  
CAAAAACCTTATTATCACCTTGTAAAATAGTGACTTTACACAGGACAAACAAACCTGAGAGACACCACCTTAACCAA  
TGATCAAGGTAACAAAACATATTAACATCATGTACCCTGGGATGAGGCCCTAGGAAGGACGCTGTACCCCAAATGT  
GTCACCTCGACCTAATCATGATAAGGTGTCAGAGAACTCCAATCAAAGGACATCCCGCTAAAGTTCTCACCTGTG  
TTCTCCAGCAGACCTGAAGGCAAAAGCAACATTTAGTAAATAAACAGTGAAATCCAAAGAAGGCCCTGGATTCCAG  
TTAGTAGAATGGTTCCACTGTTCAATTTCCTAGTTCCGGTCATTTTTATTATGATTGTATACCATGGTAACAGTGGAA  
GCTGGGTGGTGGGCATATAGGAACTCTGTTGTACTCTTTTTGCAACTCTTCCAGTCTAAAATTTATTTCAAAATTA  
AAGTTAAAAATAATTAATGAGGCCAGGCATGGTGGCTCACACCTGTAATCCCAGCACTCTGGGAGGCCGAGGCGGG  
CAGATCACAAAGTTCAGAAAGTTCGAGACCAGCCTGACCAACATAGTGAAACCCAGTCTCTACTAAAAATACAAAAAC  
TAGTTGGGCATGGTGACGGGTGCCATAATCCTAGCTACTTGGGAAGCTGAGACAGGAAAATCTCTTGAACCCAGG  
AGGTGGAGGTTGCAGTGAGTCAAGATCGCGCCGCTGCATGCACCTCCATTTCTGGGCAACAGAGCAAGGCTCTGTCTC  
AAGGAAAAAAAAAAAAAGGTTAATGAGTGAAGGTAATATAGTCATTTTTCCATAATACCCTCTGAAAATAGAGAA  
GACTTTTCCAAGCCAGCTAAAATGACCAGGCTTAAACCTTACAAAACCTCTCAATAGTACTTGTACATAATTAGT  
TGCTCCTTAAGAAAGTTTTTCATTTCTTGGTCTGCTGGGTGGCTGAACCTTGCAGGGTTTTCTGTGATCTTTCTAGCTG  
CCCTTTTTTCAGTCTCCTTTATGGGATCATTTGTCCTTTACAGGAGGTTTTTAGTGCTTAAGCCTGGAGGGTTTTTTT  
TGTTTTTTTTTTAAAGACAGAGTCTCGCTCTGTCCACCAGGCACCAGGCTGGAGTGTAGTGACGCAATCTCAGCTCA  
CTGCAACCAAGCCTGGAGGTTTTTTCTTCTCTGTACCTGTGCTATCTACAATGATGGCCACTTGCCACATTTGG  
CAATTTAAATTCAAATTCATACGATTAATAATTTAGTTTTCTCGGTTGTTCTAGCCACATTCAGGTGCTGAATAGG  
TACGTTAGTGGCTACCATACTAATCAGCACCATAATAGTTACGTCGATCATCATGGAAGGTTCTGTCCAGGCATAG  
GTTCCATTTGGGAGATGCTGTGTACTTACACAGCAACATGTAAGACACACAGGCTGCCCAAATTCACCCTTAG  
TAATTCAGCAGGTAGTACGTTGTACAGGTAGTGTACAGTCAAGTTCGAAAGTTCCCTCATAGCCTTTAACCTAT  
GGTTTCATTTCTGATAAATCTTTGCTGGAACCTTTGTTTCACTCCTTATTTTTTCTAAATACAGTTGTAGCCATTGT  
ATGTATAATACGCTCTGCACCTTATTTTTGTTTACACCTGGATATCAGTGCATTTTTAGATCTGCCCTCATTAATAGCT  
GCAGAATATTCTATTGATACATATATGATAAAGTCCCATGATTTCTGTTTTGCTAAGAAAGTATATGTTCTTAATTA  
CATGCTGAGTACTTTTTAATTGCACATCGCTTTATTTCTGTTTTCAGGTGAGTATAAGGCCTGTACTTCTTCCACCAA  
ATATTGAAAAATAATAGCATGTACGCTGAGCCTCACGAGGATCTGAGGAACGACGAGCACAGCCCATCTATCAAA  
AGATTAACGTATCGAGAGCGTCAATCAGGTGTGAGACCATGTTAGTCTTCTCCAAAAGCAGATGCTGTCACTGCAG  
ACACTTACAGAATTTTAAATAGCTGCTCTGACTTGGGTTTTTCACTATAAATCAATAGATATTTATGAAATGCCTA  
ACACATATAGTTTATAATACACTGAACATTACGGAAGGCACAAAAGCTAGAAGATGCATCCCTTCTTTGAAGGAA  
CTGTGCTATAGTTGAGAAGACAAGACCTAATTTTTATCAGAATTAATGGTGTAAAGGCAAGCTTGTCTAACCTGTGG  
CCTATAGGCCACATGTGGCACAGGACGGCTTTGAATACAATCTGGCGCAAATTCGTTAGTTTTTCTTACCACAGAA  
AATTTTTGTATTTGTTTTTTGGGTTTTTTAGCTGATTTTTTTTTTCTTTGGGTTTTTTTATATTTTAAGCTAGCACT  
AAGACAAGTATTCTTTCCAGTGTGGCCCAGGGAAGCCAAAAGATTGGACACCCATGGTGTAAAGAGATGAAGGTTAA  
TTCAAGACTGTAGTAAAAGAAATACTACAGTCCAGTGAATGAGGCTGGGATGGTCCAGGGAAGACTTTATCAAGAAG  
AAAGTACAAACTACCCTGAACAGAACTGGCATGAGGAAAGCGAAGAGATGGGAAATGGCAGTTAGGGAACAGGTAA  
GAGTATGTGGCCTAATTTATGTTGTTACAGATACCTGAAGAGCTACCACATTTCCAGCTTTGAAAAGAAAGTGTATCT  
CCTGTACAAAACAACCTTCTTCATCTCAGAAGAAGACAAAACAGAACCACAAGGCAGATGATGTCCAGGCCTTACA  
GGGTAACAAGAATGCCCTCAGAGTTAAATTAAGAAGACTGTAATACATCCTTCCCTTCTTCTTAAATGTCTAA  
TCTTGATTTTGTGTAAAGCATAAAATTTCTGATTTTCATGTGACTTATGGCAAATAAGGTTTAAATTTCTCAGTT  
ACAAATTTCTGTAACCTGGGCTTAGGTAAATTTCTTGTGAGGAAATTTTTTTTTTCACTGTCAAGTTTCTCTTACAT  
CACCAGCTGGTTTGCAAAATCCCAGTCAATCCTAAGCCAGAAATGCCAGCAAATGGACGGTCCACAGACACAGGAGG  
AGGAGCTCCACAGATCCTGTCCACGACAGTGTGAGCTTGGGGTCCGGCATGAGCCAGTGTGGTTACAGCAGCACC  
ATTGTCCACATCCCACCCGACAGAGACAGGTATCACACTCGCCTCTGACTTTGAAAACATACTCATCTTTAACTACA  
TTGTGATGAGAAAAAAAAGTCATGAATACCATTGCTCAGTTACAACCTTCTAAATTACATGTTTTTCTTGCTTTA  
ATTCACTTAAAAATATTTCTGGAGACGTTATCATGTTTATGTCAGTTACATATTTTGGACCTTATTTCTCAGTTATG  
TTATTTGTCAATGGAATCAGTTGACACATATGAAGCACTTAATACAGTACCTAGTAGGTAGGTAGTACTCCATTGG  
TGTTATTCATTATTAATGCAATTATTAATGGGACTGTGTTAGGTTACCTCGTTGCTGACTTTTAAAGAATGGGCCT  
GCGTTTTGCCCATTTCTCATGGGCTGCGTGATAGTGGGCAAGTTAAGTTACTTTACCTATTTGTGCCTTAATTGTTT  
CACACATAAAAGTGAAGATATTAACCTGCAAGGGAATANATTTGTGATCGCTGGTCCCTCAGACGTAACCTCATTAGA  
GTCAGAATGTAAACCGAGGGAATGCATATGAGCTTCTCAGAAAGCAAAAGACATAGAGATATGTTTTTGGAAACAT  
CTGGGCTCTATTTCTGCTTTTTTCTCCCTGGGATAAAGATGTGGCCCTGTACATGTTGAGGCTGATCTAAGATCAGG  
AGAAAACCAATACCAGTAAAAATTTCTTAAAGCACATGGACCTTAGTATAAATTCATTAAAAACAGTAATTTCTGCT  
TTAATTCACTTAAAAATATTTCTGGAGACGTTATCATGTTTATGTCAGTTACATATTTTGGACCTTATTTCTCAGTT  
ATGTTATTTGTCAATGGAATCAGTTGACACATATGAAGCACTTAATACAGTACCTAGTAGGTAGGTAGTACTCCATT  
TGGTGTATTCATTATTAATGCAATTAATTAATGGGACTGTGTTAGGTACCTCGTTGCTGACTTTTAAAGGGGCTGGC  
TTTTGCCCATTTCTCATGGGCTGCGTGATAGTGGGCAAGTTAAGTTACTTTTACCTATTTGTGCCTTAATTTGTTCCAC

ACATAAAGTGAAGATATTAACCTGCAAGGGAATACATTTGTGATCGTGGTCCCTCAGACATAACTCATTAGAGTCAG  
AATGTAAACCGAGGGAATGCATATGAGCTTCTCAGAAAAGCAAAAGACATAGAGATACGTTTTTTTTGGAACATCTGGG  
CTTTATTTCTGCTTTTTCTCCCTGGGATAAAGATGTGGCCCTGTACATGTTGAGGCTGATCTAAGATCAGGAGAAA  
CCAATACCAGTAAAAATTTCCCTAAAGCACATGGACCTTAGTATAAAATTCATTA AAAACAGTAATATAAAAAGCAGAG  
TACAGAAAAGCAGAGTGGCAGAGTCAAAAAAGAGCTTTTATTTTATTATTTTTTTTTTTGTATTTTTTAGTAGAGGCGGG  
GTTTAAATCCTGTATAGTTGTGATGACCTTAGTCACACATATATTCAGATTTTTGTATTCGTGTGATAGCAATGAA  
TAGAGTACTTGTAAAACTCTTCATTTACACCATAAATTACTTGGTAAATTTATTCCTCGGCTTCTTTTTATTCTCT  
TTTATTGCCGCTGTCTTTGACTTTTTTCATTGCTCGTAATAAATTTTGGTAGATGCAGGATGGTGACAAGTATGT  
TAAATTGTGTTTTACATGATCTCTGCTTTATCTTTGTAGTCCCTAGTCCATCAGCCAGTATGCATTTAGCC  
ATTTCCATGTGCCACTGTGTGCTTTAGTTCTTGATTAGGGTGATGTCATTA AATCTTTCCAACAAGTTGTGAGAT  
GGAGAGCATGGTCCCTGACAGGAGAACTATGGATTGTTGGGAAAAGACCAGAGAGCATTACGTTTTCTAATCAGTTAA  
AAGTAGAGCTTCACTGGACCTGAGCTAGCCCCCAGACAGGAGCATATGCACATTTTTAATATTCTGCATTCCTG  
TGTGGCTGTGTAGCTGTTACAGTACAGCCTATACAGAAAAGGGAGGGTCTCACCCAATAGGGAAAAGCTTTTTTTGG  
GCCGGGCGCGGTGGCTTACGCCCTTAAATCCCAGCACTTCGGGAGACTGAGGCAGGTGGATAACAAGGTCAGGGGAT  
CAAGACCATCTTGGCCAACATGGTGGAAACCCCGCCTCTACTAAAAATACAAAAAAAAAAATAAAAATAAAAGCTC  
TTTTTTGACTCTGCCACTCTGCTTTCTGTACTCTGCTTTTATATTACTGTTTTTAAATGAATTTATACTAAGGTCCA  
TGTGCTTTAGGAAATTTTTACTGGTATTGGTTTTCTCTGATCTTAGATCAGCCTCAACATGTACAGGGCCACATCT  
TTATCCCAGGGAGAAAAAGCAGAAATAGAGCCAGATGTTCCAAAAACGTATCTCTATGTCTTTTGCTTTCTGAG  
AAGCTCATATGCATTCCTCAGTTTACATTTCTGACTCTAATGAGTTATGTCTGAGGACCACGATCACAAATGTATT  
CCCTTGCAAGTTAATATCTTCACTTTATGTGTGGAACAATTAAGGCACAAAATAGGTAAAGTAACTTAACTTGCCCA  
CTATCACGCAGCCTATGAGTAGTAGAATAGGGAATCTATTATTGCCTATTAGTCTCCAGGAGTCATTGCCGTTATA  
GATATTTTGAACCTAAAGCAGGATCTAATTAAGATGGAATGTAGGCCAGGCACAGTGGCTCATGCATGTAATCCCA  
CCACTTTGGGAGGNMNNMNNMNAACGCAAAATTCATAAACTTTCTTAAAACATTATGAGATTTGGTTTTTGTTTTT  
TGGGTTTTTAGCCCATCAGCTATCATTTAGTATTAGTATATTTTTATGTGTGGCCCAAGACAACATTTTTCCAGTG  
TGGCCAGGGAAGCCAAAAGATTGGACACCCATGGTGTAAAGAGATGAAGGTTAATTC AAGACTGTAGTAAAAGAAA  
TACTACAGTCCAGTGATTGAGGCTGGGATGGTACAGGAACTTTGTTGGAACAGGTAAGAGTATGGGCTTAATTAATGTT  
GAACTGGCATGAGGAAAGCGAAGAGATGGGAAATGGCAGTGTGGGAAACAGGTAAGAGTATGGGCTTAATTAATGTT  
GTTACAGATACCTGAAGAGCTACAACATTCAGCTTTGAAAAGAAAAGTGTATCTCCTGTACAAAATACAACCTCTTC  
ATCCTCAGAAGAAGACAAAACAGAACCACAAGGCAGACGATGTCCAGGCCTTACAGGGTAAACAAGAATGCCCTCAG  
AGTTAAATTA AAAAGAACTGTAAATACATCCTTCCATTTCTCTTTAATGTCTAATCTTTATTTTGTGTAAGCATA  
AAATTCTGATTTTATGTGACTGCTTTATGGCAAATAAGGTTTAAATTTCTCAGTTACAACTTCTGTAACTGGGTC  
TTAGGTAAATTTCTTGTGGGAAATTTTTTTTTTCACTGTGAGTTTCTCTTACATCACCAGCTGGTTTTGCAAATCCCA  
GTCATACCTAAGCCAGAAATGCCAGCAATGGACGGTCCACAGACACAGGAGGAGGAGCTCCACAGATCCTGTCCA  
CGACAGTGTGAGCTTGGGGTGGGCATGAGCCAGTGTGGTTACAGCAGCACCATTGTCCACATCCCACCCGCGAGA  
GACAGGTATCACACTCGCTCTGACTTTGAAAACATAATCATCTTTAACTACATTTGTGATGAGAAAACAAAAGTCT  
TATAGACCATTTGCTCAGTTACAACCTCTAAATTACATGTTTTTCTTGTCTTTAANNMNNMNNMNTCCCAAAATGC  
TGGGATTACAGGCGTGATCCATTGCTCCAGCCCTAAAATAATTTTTTAAAAGATGGTATATGAGATACAATGGCT  
AAAAAGCCAATCAATAAAAATGAAAATTTTGAATAACTCAAAAAAGAAAGGCGAGGAAAGGAGGATAAAACAAAAAG  
AAGCAGAGAAAAACAAACAAATAACAAGAGAAATAGCTCTAAATCCAGTTCTGTGAGTAATGACATAAAATGACTT  
AGCACTTCAATTA AAAAGCAAAAATTTGGACCAAGCATAGTGGTTACACCTGTAATCTCAGGGACTCAGGAGGCCA  
AGGAGGTAATGCGTCCCCTGAGGGCAGGAGTTTGGACAGTCTCGGGGCGAATATAGTGAAACCCCAATCTCTAG  
CAAAGAACCATTTGTTAAGGTTGAGCCTAGGTGTGGTGATCCACACGCGTGTAGGTTCCAGGCTACGTATGGGAGG  
CAGTGGCAGAAGAATGCAGTTGAGCTTAGTTTGGCCATCTTGAGTTGAGTCCAGTCCAGTGCAGTCTTATCACATG  
TCTGTACTCCAGCTGAGTGACAAAGCTAAGACCTGTCTCAAAAAAAGTAAANACATAAGAAGGCAAGAGAAAT  
TCTCAAAAACAGATTTAAAAAAAATCAAGACCTAAGCTATATGCTGTCCACAAAGTACAATTA AAAATTCACCAATGGGT  
TGAAAATAAATGAATGGAAAAAGATGCACAAGTCAAACAATAAGCATAAGAAGGCTGCAGTGCCTTATTAATATGT  
GCAAAGTAGATTTT CAGGACAGTCATTTTATAATGACACAGTGGTCAAGTTCAGGAAGACATACCAGTCATAAAATATG  
TATTCATAATAACAGCTTTGAAAAGCTCTGAAGCAAAAATGTGGAAATAGACTTTCCATAATTTCTATTCAGAAAAT  
TGTACTACTCCCTCTCTGCAGTTGATAAAAAATATATGGAGAAAAAATTAAGTAAACAAACATGATCTCTGCAACAC  
TATTC AACCACCTTGAATCTAATTTGAGACTTAGAGAACATGAGACCTAATAACTGCATAATTACACCAGTCTTTCA  
GGTACACACAGTATGTTCCCTAAAGAAAATAAAGGCTGTTGGGGCTGTCTGTGTCTCCAATTTAAAAGAACTTGAT  
GGTGTGATGTGCTACAACCTGTGTTCTCTGGCTCGTGGAAAGTTTAAAGTTAAAACCTCGAGCTATATTTAGACCCGT  
AATCACTCTTAGATAGTCCAGACGAACACACTTTGCTAATACCTTACTGAGTTCCGAATAAGAGATAAGTGCACAG  
CGAGACAACAATTACGANAAATATTTCCAAATCGACGTGATATCTGAAAGTATACGCATACCTAAATATGATCATCG  
GGATGACAGCTAAAACGGGGGCTCTGTAAGACTCAGTTAGCTCTTTGAAAATGCTAGTTT TAGCTAGAAGAGTACGT  
CTCTAATCAGTGATTTGCGCGGTGTTGGCGTCAACGCCATGTGATAGTCCCAGGACTACTTTGGGGGAGAGCCCGA  
GACGGGCGTTGGATCACTGATGGTCAAAGAGATCAAGACCATCCTGGTCAAACATAGGTGGAACCCGCCGTCTCT  
ACTATAAATACAAAAAATTAGCTGGGCATAGTGGCATGTGCCATAATCCCAGCTACTCAGGAGGCTGAATCAGTG  
ATGAATTGCCGTAACCCGAGAGGCTGGAGTGTGCTGGTAATGCCGAGATCTGCTGCCATCGCACTCCATGCCGTG  
GGTAAACAAGAGCGAAACTCCGTCTCAAAAATAAAAATAAAGTTAGCCAAGGAAGAGTGAAGTAATTTCCAACGATGGA  
ATGAATGGA AAAAGATAAAAAGCAAAAGTCAACAAAATAGAAAACAGACCATAGAGAAATTTAAAGCTAAAAGTTGGT



TCTTTGAACAAGATGAACAAAATTTTTATAAACCCCTTATCCAGTACAGAGTGGTCTTAGTTCATGTGGCCTGTAC  
 AGTAAGTCTTGAAATCAGCTAGTATAACATCCCTCAAACCTTACTCTTTTTCAAGGTTGCTTTGAATATGCTAGGT  
 CTTTTGTCTGCATAAATTTGTGACTCAGTTTCTCAGTATGTTTATTCCTTATTATGGGGTTATATTTTTCTGTGCT  
 TTGTATGCCTTGGTTGATTTTTTTTTTAATTTGGATGCCAGATCCTGGGGAATTTGGGGTACCAATATTTTTGTTTTC  
 CTGTAATATTCTTGATCTTTATTTCTGGGCGGTTCCCTTNCCTTGAAACGATTTGCCAATCTTTTCAGGTTTGCTTT  
 TAATTTGTGGCAAGACAAAATACTGTGTAGTCTAGGCCGTTTTTTTCTCTGCTACCAAGGCAAGACCCCTCTGAGT  
 GCTCCATGTGGTGCCTCAGAAGTGAGGAGATTCTCTAGTCTGGCGATGGAGCCGCACTGTCCGCACTTGNATTGCT  
 ACAGACGTGGGGTCACTCTCGATGTCTGAAACCTATTGCCCTATATATTTTTGGCCATTTTTTTCTTTAAATCTC  
 AGACAGGAAGGTAAGTTTGGTCCCTGTTACTCCAACCTGGCTAGAAGTGAAGCCCTGTGGTGTAAACAGGCCACT  
 GCCAGGTATGGATGGGGCCAAACCGCCTGGATACTGCACCACAGCAGGAGGCCAGCATGGATGTCCACCTCTAC  
 AGTGTCCACCACAAAATCTTTTGTCTTAGAAGAATAACCAACCCCAAAAATATATACAACATGCTATATCAATA  
 AAAGAAAATTTAATGTACAACATTTCTGTATCAATAATAAATAGTTAAATAACCAGATATAAAAAAGAGGCAACCCA  
 AAATCCATTACATAGATGCTCACTTTAGTAAACTGGTATGCATATAAAAATGGAATTCACAGCAGTGTAGGAGTATG  
 AGACAACCTGTCCAGGCATTGATAAAAGTTCTCCAAGGGATACTGTACAGTAAAAGGTGCATTGGTGGCTAATGTAC  
 ATATGTGCTTTTGGGTAAAGATGGGGCCAGGGAGGAGGGTAATCTATATTCATGCTTGTTCATATGTAATAAAGAA  
 ACTGGAAGGGCTGGTGCAGTGGCTTATGCTAATAATCCCCGGCACTTTGGGAGGCCAAGGTGTATGGATCACCTG  
 AGGTCAAGAGTTCGAGACCAGCCTGGCCAACATCGTAAAACCCCATCTTTACTATAAACTACAAAATTAGCCAGG  
 CATGGTGGCATGTGCTGTAATCTCAGTACTTGGGAAGCTGAGGCAGGAGAATCACCTGAACTCAGGAAGCAGAG  
 GTTGCACTGAGCCGAGTTAGCGCCATTGCACTCCAGCCTGGGCAATAAGAGCGACACTCCATCTCAACAAAATAAG  
 TGTAAATAATTGAGGATCCAGATGGCGTGGTGAAGAACCCAGGATTTGAAGCTCTTGGTGAACGCACGGAGAGTGA  
 GTCAGGCCGCATTTCCAGACGGATCTTTGTGGCCACAGAACGGGGAAAATTTCCAGGTATAAAAAGAGACGCGGGAT  
 GCCAGGCAGAGGTTTGGCTGGTGCAGCCAGCAGCCGGTGCAGCTGGCGGCCCTGGCTGCTATAGCGCACTGGTCCC  
 GGTGCCCTGTTGAACCCGGCAATATGAGACTTGAGAGGGCAGATTGGCATATCTATCTGACTGAATGGGACTTGGAC  
 AGTGAGCCAGGCAAGGAGACTCCGGGAAACGGCGTTTGGGGTAGCGCAGTGGGACAAACAAAATGGTGATTTCCAAA  
 CGCTCCAGGTGGAGAGTTTCACTGCAGGCACAGCTGAACCCAGGACGGTGCAGCTCGGTGGGGGGGGGGGGGGTCT  
 GCCATTACCAGAGACTCGTGCCTTACTGAGGTACACGCCCATTTGCTGACGCAGCCTGGACGTTGCCGAGGCCAAC  
 CCGCCACAACAGAGAGACTCCACCCGAGGCAGCCGCGCAAGGCGGAGACCACCACAGCAGGACAGAGC  
 CTGCAGCAACAGGGTGAACCTCACCTTAGCAGGGCGGAGCCTCGGCAGGCAAAATAGTGACTAGACTGCCTCCTAGC  
 TGGGCAGGACAGCGCAACGGCACTCACAAAAGAAAGCCCCAACACCCCGAGACAGAGCATCTGAAAAAAGGGGT  
 TTATGAGTTCTGCTGCAGCAGAAATTTAAATGTAGCAGCCTAACAGCCCTGAATGAACAACAGAGCTCACAGCTCAGC  
 ACTTGAGCTCCTATAAAAGTATAGACCGTCTCCTCAAGCAGCTCCCTGACCCCTCTATATCCAAAAGACTGACATTT  
 GGCAGGCATCATTCAGGGACAAAAGATAGCAGAAAAAGAAACCAGTAGCATCCCTCACTGTTCTGCAGCTGCTATAG  
 TGCTACCCCAAGACAAGCAGGGCTTGAATGGACCTCAGCAGTTGTACAGTGGAGGGACTAGACTGGTAGAAGGAA  
 AACCAAGTAACAGAAAATACTTCATCATCAGCAATCTGGGCGTACACTCAGAGACCCAATCAAAAAGTCAGCAACTA  
 CTCAGGCGACAGGTGGATAAATCCACAAAGATGGGAAGAAACCAGCGCAAAAAGGAGGAAAACACCCGAAACCAGA  
 AACTTCACCTCCTAGAAAAGGACCAAGACTCCTCACAGCAAGGGAACAAAGCTGGACGGAGAATGAATGTGACGA  
 AATGACGGAATTAGACTTCAGAAGGTGGATAATGAGAACTTTTTGTGAGCTAAGAGAATGTTTTTAAATCAATGC  
 AAAGAAAATAAGAACCTTGAAAAAAGATTTTTAAAAAAGAATCGAGGAAANGATAACAAGAATGGATACCCTTAGAG  
 AGGAATATGGAAAAGAGTAAAGGAGCTGGGAAAAACAGAAATGAACGAGAACCCTCCCGAAGCANGCAACAAGGTTTC  
 ACAGCCGAATTGACCCAAGCAGAAAGAAATATCAGAAGTCGAAGATCAACTCACATGAAATAAAACGGAGAAAC  
 CAGATCAGAGAAAAAGCGCACAAAAGGNATGAACAAAGTCTCCAAGAAATGTGGGACTATGTGAAGAGACCTAACC  
 CTGACGTTGATAGGTGTACCAGAATGTGACGAAGAGAATGAATCCAAGCTGGAAAATACTCTTCAGGACATCATCC  
 AGGAAAATTTCCCCACCTAGCAAGACAGGCCAACACTCAAATGCAGGAAATACAGAGAACCACAAAAGATATTC  
 CGCAAGAAGAGCAACCCCAAGGCACATAATCGTCAGATTCAACAAGGTTGAAATAAAGGAGAAAATACTAAGGCCA  
 GCCAGAGAAAAGGTCGGGTCAACCCACAAAGGGAAGCCCTCAGACTCACAGCAGACTCTCTCGGACGAAAACACTAC  
 AAGCCAGAAGAGAGTGGGGGCCAATATTAACATCTTAAAGAAAAGAACTTTCAACCCAGAATTTTCATATCCAGC  
 CAAACTGAGCTTCAGAAGTGAAGGAAAAATAAAATCCTTTGCGAACAAGCAAGTACTCAGAGATTTTGTACCACC  
 AGGCCTGCTTTACAAGAGCTCCTGAAAAGAGGCCTACACATAGAAAAGGAACAACCAGTACCAGCCATTTCCAAAATC  
 AACTAAAATGCTAAAAGAGCATCAACATAATGAAGAATCTACAACAACTAACGGGCAAAAACAGCCACNTTAGCATCA  
 AAATGGCAGTATCAAAATCACACATAACAATATTAACCTAAATGTAAATGGACTAAATGCACCAATCAAAAAGACA  
 CAGACTGGCAAAATGGATAAAAAATCCAAAACCCATCAGTGTGCTGTATCCAGGAAACCCATCTCACATGCAAGGAT  
 ACACAAAGGCTCAAAAATAAAGGATGGAGGAAGATTTACCAAGCAAAATGGAGAGCAAAAAA  
 AGCAGGAGTTGCAATTTCTCATCTCTGATAAAAATAGACTTTTAAAGCAACAAAGATCAAAGAGACAAAGAAGGCCATT  
 TACATAATGGTACACAGGATCGATACACCAAGAAAAAGCTAAACGATCCTAAATATTTTTGGCACCAATGCAGGAA  
 GCCCCAGAATACTTACGGCAAGTTCTTAAATGACTTTACAAGAAGCACTTAGTACTCCCCAACACATTTATTGGTGG  
 GGGCAGACTTTTACACACTNCCACTGGTCAATTTTGAACGAATTCACCAGGACAGAAAAATCAACAAGGATATCCA  
 GGGCTTGAACCTCAGACCTGGAAACAAGCAAACTGATAGACATTTACAGAACTCTCCACCCCAAAATCCACAGAATAT  
 ACATTTCTCTCAGCACCACATCACACCTACTCTAAAATTTGACCACATAATTTGGAAGTAAAGCACTCCTCAGCAAA  
 GCAAAAACAAATGAAATCATAACAGTCTCTCAGACCATAGTGAATCAAGTTAGAATCAGAATTCAGAAAACCAACC  
 CAGAACCGCACAGCTTCATGGAACTGAACAACCTGGCTCTTGAATGTTGACTGGATAAACAATGAAATAAAGGCAG  
 AAATAAAGAAGTTCTTCAAAAACCAATGAGAACGAAGACACAACATGCCAGAATCTCTGGGACACATTTAAAGCAGT

CTCTAGAGGAAAAATATATAGCAGTAAGTGTCCATATGAGGAGAATGGAGAGATCCAAAATTTGACACCATATCGTCA  
AAATTGAAAGAGCTAGAGGAGCAAGATCAAAAAAATCAAAACCCTGCAGAAGACAAGAAATAACTAAGATCAGAG  
CTGAACTGAAAGGAGATAGAGACACGAAAAACCTTCAAAAAATCAATAAATCCGAGAGCTGGTTTTTTGAAAAGAT  
CAACAAAATAGACCACTAGCCAGATTGATTAAGAAAGAAAAGACAGAACAACCAATAGATGCAATAAAAAATGATA  
AAGGGGAAATCACCACAGATTCCACAGAAAATCAAAACCATCATCAGAGAATATTACAAACAACCTCTATGCACATAA  
ACTAGTAAACCTGGAAGAAAATGATTAATTTCTGGACTCCTGTGTCTCCCAAGCCTAAACCAGGAGGAAGCTGAA  
ACTATGAATAGACCAATAACAAGGTCAGAAGTCGAGGCAGCAATTAAGAGCCTACCACACAAAAAAGCCCAGGTC  
CAGATGGGTTACAGCCGAATTTCTACCAGACACACAAAAGAGGAGCTGGTACCATTCTTCTGAAACTATTTGAAAC  
AATCCAAAAAGAGGGAATCCTTCCCAAATCATTTTATGAGACCAATATCATCCTGTATACCAAAACCTAGCAGAGAC  
CCAACAAGAAAAAGAAAATTCAGGCCAATATCCATGATGAACATAGATGCAAAAAATCTTCAATAAAAAATTTGGCAA  
GCCAATTGCAACAGCAAAATCAAAAAACGTATCCATCATGATCAAGTAGGATTTCATCCTGGGGATGCAAGGCTGGTT  
CAACATATGCAAGTCTATAAACGTAATTTACCACATAATCAGAACCAAAAACAAAAACCACATGATTATCTCAATT  
GACGCAGAGAAGGCATTGACAAAAATCAACAGCCCTTTATGCTAAAAACCTCAATAAACTCGGTATCGATGGAA  
AGTATCTCAAAGTAATAAAAAGCTATTTATGACAAAACCAACAGCCAATATCATACTGAATGGGCAAAAACTGGAAGC  
ATTCCTTTGAAATCCAGCACTAAACAAGGATGCCCTCTTTCACCCTCCTATTTCAATATAGTACTGGAAGTTCTA  
GCCAGAGCAATCAGGCAAGAAAAAGAAAATAAAGGATTTAGATAGGAAAGGTGGAAGCCAAATTTGTCTCTATTTG  
CAGACGACATGATAGTATACCTAGAAGACCCCATCGCCTCAGCCCAAAAACTCCTGAAACTGATAAGCAACTTCAG  
CAAAGTCTCAGGATATTTAATCAAAGTGCAAAAAATCACAAGCATTCCTCTACACCAATAACAGACTTAAAGAAAGC  
CAAATCAAGAACGAACCTGCCATTACAAATTTGCTACAAAAAGAAATAAAATACCTAGGAATACAACCTCACAGGAAACG  
TAAGGGACCTCTTCAAGGAAAACTACAAACCACTGCTCAACGAAATAAGAGAGGACACAAACAGATGGAGAAACAT  
TCCATGTTTCATGGTTTTGGAAGAAATTAATATCGTGAAAAATGGCTATACTGACCAAAATAATTTACAGAATCAACGCT  
ATCCCCATCAAGCTACCATTGACTTTCTTACAGAACTGGAAAAAAGCACCATGAACTTCATATGGAACCAAAAGA  
GAGCCCGCATAGCCAAGTCAATTTCAAGCAAAAAAGAACACAGCGGGGGGCATCACACTACCGGATTTCAAACCTATA  
CTATAAGGCTACAGTAATCAAAACAGCATGGTACTGGTACCAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NN  
NN  
AACCACATTTGAGATAGCATCTCACACCAGTCAGAATGGCGATCATTAATAAACTCTGGAGACAACAGATGCTGGAGA  
GGATGTGGAGAAAAAGGAATACTTTTACACTGTTGGTGGGAGTGTAAATTTAGTTCAACCATTGTGGAAGACAGTGT  
GGCAATTCCTCAAGGCCTTAGAAAAGAAAATTCGGTTTTGACCCAGCAATCCATTACTGGGTATATATCCAAAGGA  
CTATAAAATCGTTCTACTATAAGGACACATGCACACGAATGTTTCATTGCAGCACTGTTTACAATAGCAAAGACCTGG  
AATCAACCCAAATGCCCATCAATGATAGACTGGATTGGGAAAACGTGGCACATATACACCGTGGAAATATTATGCAG  
CAATCAGAAAATGATGAGTTTGTGTCTGTTTTGTAGGGACATGGATGAATCTGGAGAACATCATTTCTCAGCAAACCTGAC  
ACAAGAACAGAAAATGAAATACCACATATTTCTCACTCATAGGCGGGTGTGAAAAATGAGAACACATGGACACAGG  
GAGGGGAGTACTAAACACTGGGGTCTATTGGGGGGAAAAGGGGAGGGCCAGCGGGGGGGGGGAGCTGGGGAGGGA  
TAGCCTGGGGAGAAAATGCCAAATGTGGGTGAAGGGGAGAAAGGAAGCAAAACACACTGCCATGTGTGTACCTACGC  
AAAATTGCAGGTTGCAACAGTACCCCCNAACTAAANGCAATAGAAAAAGGTGAAAAAAAATAAATAGATAAATA  
CGAAACTTAGCCAGGTTTTGGGTGGTACACGCCTGTTAAGTCCCAGCCTACTGCGGGAGGCTCGAGGCAGAATCGC  
TTGAACCCAGGAAGATGGTGTACAGTGAGCCGAGATCATGCCACTGCCTCCAGCCTGACAACAGAGCAAGACTTC  
ACTCCAGCGGAATGGTCTTCTGGAAGGGAAAACAAATGAGAACCTAGACAGCATGGACCAGGGGTGAGGATAATTTCA  
CTACATGTCATCTATTTAATACTTGTGATCTTTGAACCACTTGAACCTATATTTACCTATTTCAAAAAGCTAAATTT  
AAAGATACACACAAGAAAATTTGTGAGAAATTCAGGGAGACAATGTTTACAATAAAAAGCAAGACTTAAATAAATATG  
TGCATATTGANTATTACAACCTGGGACATGTGTGATGCATTTCTATGCAAAGAATATAATGGAAGACAGAAAATCAAG  
ATTTCTGTGCTAGCTAATGAAGCTTTTGGTATTCCTGTAGTTGTACTAAATGTTTTGTAACCTTAACCTGNNNNNN  
NNNNAAGTGCAAAATTCACAAGCATCCTCTACACCATAACAGACTAAAAGAAAGCCAAATCAAGAACGACTGCCATT  
CACAATGCTACAAAAATTCGAGCTCGGTACCCGGGACCCACAAAATACCTAGGAATAACAACCTCACAGGAAACGTAA  
GGGACCTCTTCAAGGAAAACTACAAACCACTGCTCAACGAAATAAGAGAGGACACAAACAGATGGAGAAACATTCC  
ATGTTTCATGGTTTTGGAAGAAATTAATATCGTGAAAAATGGCTATACTGACCAAAATAATTTACAGAATCAACGCTATC  
CCCATCAAGCTACCATTGACTTTCTTTCACAGAACTGGAAAAAAGCACCATGAACTTCATATGGAACCAAAAGAGAG  
CCCGCATAGCCAAGTCAATTTCTAAGCAAAAAAGAACACAGCGGGGGGCATCACACTACCGGATTTCAAACCTATACTA  
TAAGGCTACAGTAATCAAAAACAGCATGGTACTGGTACCAAAACAGAGATATAGACCAATGGAACAAAACAGAGGCA  
TCGGAGGCAACACAACATATCTACGACCATACAATCTTTGATAAAACCTAACAAAAACAAGCAATGGGGAAAGGATT  
CCCTGTTTAATAAATGGTGTGGGAAAACCTGGCTAGCCATGTGCAGAAAGCAGAAACTGGACCCCTTCTGACAAC  
TTACTACTAAAATTAACCTCAGATGGATTAAAGACTTAAACATAAGACCTGGCACCATAAAAACCTTAGAAGTAAAT  
GTAGGCAAAACCATCCAGGACATAGGAGTAGGCAAGGACTTCATGAACAAAACACCAAGAGCATTGGCAACAAAAG  
CCAAAATAGACAAAATGGGACCTAATGAAACTCCACAGCTTCTGCACGGCAAAAAGAAAACAATCACTAGAGTGAATCG  
GCAACCAACAGAAATGGGAAAAAATGTTTTGCAGTTTACCCATCTGACAAAAGGGCTGATATCCAGAATTTACAAAGAA  
CTCAAAACAGATTTATAGGAAAAAACAAGCCCATTCAAAAATGGGCAAGGATATGAACAGACACTTTACAAAAGA  
AGACATATATGAGGCCAACCAATCATATGAAAAAATGGTCACTCATCACTGGTCACTCAGAGAGATGCAAAATCAAAAC  
ACATTGAGATAGCATCTCACACCAGTCAGAATTTGGCGATCATTAATAANNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNCTTAGAAAAGAGAAAATTCGGTTTTGACCCACCAATCCATTACTGGGTATATATCCAAAGGACTATAAATCGTT  
CTACTATAAGGACACATGCACACGAATGTTTCATTGCAGCACTGTTTACAATAGCAAAGACCTGGAATCAACCCAAA

TGCCCATCGATGATAGACTGGATTGGGAAAACGTGGCACATATACACCATGGAATATTATGCAGCAATCAGAAATG  
ATGAGTTTGTGTCTGTTTGTAGGGACATGGATGAATCTGGAGAACATCATTCTCAGCAAACCTGACACAAGAACAGAA  
AATGAAATACCACATATTTCTCACTCATAGGCGGGTGTGAAAAATGAGAACACATGGACACAGGGAGGGGAGTACT  
AAACACTGGGGTCTATTTGGGGGGAAAAAGGGGAGGGCCAGCGGGGGGGCGGGGGAGCTGGGAAGGGATAGCCTGGGG  
AGAAATGCCAAATGTGGATGAAGGGGAGAAAGGAAGCAAAACACACTGCCTATGTGTGTACCTACGCAAATGTCTTG  
CATGTTCTGCACATGTACCCCAAACTTAAATGCAATAAAAAAATTTTAAAAAATTAATAGATAAAATACAAAA  
CTTAGCCAGGTTTGGTGGTACACGCCCTTTAGTCCCAGCCTACTCGGGAGGCTGAGGCAGAATCGCTTGAACCCAGG  
AAGATGGTGTACAGTGAGCCGAGATCATGCCACTGCCTCCAGCCTGACAACAGAGCAAGACTTCACTCCAGCGGA  
ATGGTCTTCTGGAAGGAAACAAATGAGAAGTACAGACTGAGCAGCATGGACAGGGGTGAGGATAAATTTACTACATGTCTCT  
ATTTAATACTTGTCTGATCTTTGAACCACTTGAACCTATATTACCTATTTCAAAAGTAAATTTAAAGATACACACAAG  
AAATTTTGCAGAATTCAGGGAGACAAAAATGTTTACAATAAAAGCAAGATTTAAATAAATATGTGCATATTGATAT  
TACAACCTGGGAAATGTGTGTGCATTCTATGCAAAGAATATAATACGGAAGACAGAAAATCAAGATATTTCTGTGCT  
AGGCTAATGAAGCTTTTTTGGTATTCTGTAGTTGTACTAATGATCTATGTCAAAATGTAACTGCAAGGCAGTTAA  
CAAAGTACACAACCTAAATAAATAAATAAATGGATAGAAATTAATAATGAAATACATCTTGTTCACCTCAGCGCA  
CGCGGATGCCACCCTCACTGTGTGAGCCAGGACCTGAACATGCGGCCAGCCCCCTTGACCTCGGAAGAATTTAAA  
CACGTGGACTCACTGCGCTGTTCTGTGCGCCAGGGATGCCACCCCTTGTGTGAGCCCAGGACCCTGAACATGC  
GGCCAGCCCCTTTGACCTCGGAGGAATTTAAACACGGTGGGCTCACTGCGGCTGTTCTGTCCGGGCATACCCCT  
GGAGAAGGAAGAGCAGAATTAATGTTGATAAATTCGAGAAAAGATCCTGTGCATCACCCCTATAGTTCCCTATCTT  
CAGCAAGAAAGCAGGGGCAAAGCTAAATTTATTCATATTTTTCAAGGTAGGTGAGTTTTTAAAAATAAATGCCAT  
TGGCCAGGCATGGTGGCTCATGCCGTGTAATCCCAGCACTTTGGGAGGCTGAGGCAGGCAGATTACAGGGTCAAGGA  
ATCAAGACCATCCTGGCCAACATGGTGAACCCCGTCTCTACTAAAAATAAAAAATTAACAGGATGTGGTGGCGGGC  
ACCTGTAGTCCCAGCTACTTGGGAGCCTAAGGCAGGAGAATTGCTTGAACCCAAGAGGTGGAGGTTGTGGTGGAGCT  
GAGATCCCGCCACTGCTCTTCAGCCTGGTGACAGGGAGACTCCGTCTCAAAAAATAAATAAATGCCATCAATCTATG  
TAGATGTTATAAATTTATATCTAAGGACTCAAAGATAAGGAGTGAACAATAGGAGTTTTTACTTGTAAAGAACTGATG  
GGGAGATGCTGAAAACATCTGTTTATATCAGCAATTTCTCTAGGCCCTTTCTTAGATCCTGTATCTGAGAACTGT  
CTTTTCATGAGAACTGATCAAGAGATGCTGAAATAAGTTTACTCGATAGCAATGATTTTTTACCCGGAATTTTGT  
TGCTTTTTTAGAGCCACTTTTTGCCATTCTGGTAGTATTGGTAAATGGGCATTCCTTTGTGAGCTCTTTATAAAG  
TACCTGTAAATGGCATCCTTTCTTCTTTTTGGGTTTAAATTTCCATAAATTTGCTTCTGCCTTTATCCTTCCAGGA  
GATTCACCTTTCCAACAGACGAGGTTCGGCCGGCTGCAGGAAAGGGAAGCACAAGCGCAAGAAGCTGCTGGAGCCAC  
CAGGCACGACGACTCGAACCGCCGGCTCCGGTACCAGGGGAGCGCATCTGAACGCAAGCCCTGTGTCGCCCT  
CCGTGGCCCTGCTTCCACACTCCTCAGGCCAACCTTCCCAGTTGCTGCCATGGTACCAGCCAGGCCCTTATCT  
CGTCCCAGGCTTTCCCTCCAGCTGTGACCTGTCCCGGAAGAGAATATGCCGTTCCCGGAAGTGCACCCGGAAGTC  
CTGCCTGGGCCACCCTCATCCAAGGACCAGCAGACTTGCCTAGGTTTTCCCTTCTCCTTGTGATACTTTTATGA  
CCGTTTTCTGCCTGACCCCTCTGTCTGCCCTCTGTTGTCTCCATCGTTCTTTCCATGTCCATTCCTGGGGGAGAC  
AGCCTCTTCTGCAATATTACCCCTCAATGTGACCAATGACTCCAAACCTGGACCCACCCCTTCAGTCACCAACCAA  
AGGAGAGAGGAGGAAAAAGTGGGAAGCACAAGAAAGGGGCACCCATTCATTTCCCTCAAGGAGCAGTTCCGCCCTTG  
AGTTAAACTTACTTCAGGAAGAGATGCCAGACCCTCTGAATCTCCAGATGAGATGTGCCACAAGCTGAGTATGT  
AAGTGTGCTCATTTTCAACACTCAAATGACAAAGTAAATAACTTTACTAAAGTTAAGGTGCTTGCCTTGGGACTC  
AGTGCTGACATCCTCAGTGGGTAATCCGCATGGCTACTGTGAGTGCGCCATTAGCCTAATTTATGCATAAACACGA  
AACACAAATGCCTGGCTGTTTGGAAAGTAGACATGAAAGAATCGTCTATATTCTGGATTCTTTCTTTGTGAGGTTT  
TTCCCAAAAGTGCTATCTTTCAAGTTTTCTGTGTGAACCTGGATTCCCTCCACTTTTTTAAAAATTCGATCCAGGTATTA  
CCCTTAAGTGTTCATTTGCCTATAATATATGTTTGGCTTACCTACAACCTGGGGTTATGGTCTGGTAAGATGGGAATAA  
GCCATATTGACAGTTTATCAATTTTACTTCTTCATGTACAGTACAGGACAACAGGTGTGTATGAAGCAGTAGCAG  
CTACTTTATTTGTATTATCAGTAGAGAAGGCAGTTTATAGTTGCACACTTAAACATGTATCTTTTCAAAATCAA  
AGGTTGTTTGGTCTGCTTCTTTGAACTTTTCACTCCCTTTACCTAGAAAGCCATTCCTCAGGTTTATTTAGCAGGTT  
AACTTTGGAATCTGACATTCCTATATAGAATCCAAGACAAGTGTGAATATTTACAAAATTTTCTTTGACCTTTA  
AGGTACGATATCTTCTTTAAAAAGATATTCAAAATCTGTGATCCAAATCAAGAAAATATGCTCACTGTTGAAAGCC  
TGATATTAGCATATCCTAATAAATCTAAGGCAAGGAAATGTTTTATGGTGAATAAAAATTTGGGCTTTGGAGTAAG  
ACATGGATGCAAACCCAGCTCTGTCACTAGTTAGGTAGATTCACTTAGACAAGTTATTTACCTTTCTTTGCCTC  
TGTTTCTTATTAGGAAAAATGGAAAGGTATATCCCTCATAGCTTACTCTGGGGGCTAAATGATGCAGTATTTATAT  
AGTACTTCTCCCATGCAGTAGGCACTCCTAGAGAGCTGGAAGTTTGTTTTGTTCATTATATATATTCCTAGTGT  
GACACAGTGTTTAGAACATGTAAAAATGTTGAGTATATATTTGCTGAATAACAAAAAATCACCTGTGGGTGCCTGG  
CATGTAGTAGGTAAGTACTGAGAAAGTAGTAGTGCCTTCATATCACTTACAAAGGAACCTTATCTTAGAGCTAGGCTAGA  
AAGCTTTTACTTCTTAGTTATGTATATAAATTTTATTTCCCTTAAATATGTGAAGTTGAAGATCTGCCAGTTACAGT  
AGTCTCCCTTACCAGGGGATATGTTTCAGACCCTGTGGATGTCGATGTCGGGACCACAGATAGTTCTGAACGCGT  
ATATACTGTGTGTTTTTCTTTACATGACTATCCATACCTATGGTAAAGTTTCATTTATAAATTAAGAAACAGTA  
AGAGTTTAAACCAGCAGTGACTAATAACATAGGACAGTTATAGCAATATATCTCATCACAGTTTCATGGATAGAAGAC  
TCATTTCTTCCCAAGATTTTAGCACCCCTCAGCATATGGGTTTTTTTTCTTTCTTTCTTTGAGAACCTTTACCTTTTAC  
TTAAATGAAGTACTTTATGACTTATCTTTAGCATATCCCAATTTGGTGGCATCATTACTCTTGCACCTTTGGGGCCGT  
TATTAAGTAAAGTAAAGTGTACTTTGAACACAAGCACTACGATACCATGGCAGTACGATCTGATAACCAAGACGACTA  
CTGATTGACTAACAGGTGGGTGGCACATACAGTGTGGCAGGAAGCAATGAGGCAGCTGAGTATTTCTGTGACACTAT

TCAGAATGGCTCGCAATTTTAAACTTAGGAACTGTTTCTGGAATTCCTCATTACTGTTTTTGGACCTCGGTTGAC  
CACAGGTTACCGAACTACAGAATGCAAAACCATGAATAAAGGAGAACTACTGTATTTTGGGATAAGAAGATTAAA  
GTGTCTTTTATGTGTCTTACTTTCTAGCAGTGTGTTATAGACAACAGTGGCAGTGAGGGCAGTCCCTGCTGCTACC  
CCTGCACTGTTTACAGGATCATCTCCAGGGAGACTCCATCGCATCCTACTGCCGGTGTGCTGTCCCCAGGGTCAC  
CTCCAGGGAGACTCCATCGCATCCTACTGCCAGCGTGTGTCCACAGGGTCACCTCCCAGGGAGGCACCATCCCA  
TCCTTCTGCCAGCGAGCTGTCCCCAGGGTCACCTCCCAGGGAGACTCCATCACGTCTACTGCCAGCGTGTGTCC  
ATGGGATCACTTCCAGGGAGACTACATCCATCCTACTGCCAGCATGCTGTCCCCGGGGTCAGTCCCAGGGAGA  
CTCCATCGCATCCTACTGCCAGCATACTGTCCCCGGAGTCACCTCCCAGGGAGACTCCATCCGGACTGGTCTGTC  
AGCATCAGGTAATGGATCAGGACAACCAATGTTTCAAATCCATTACCTGTTTGTGATGAGCTCTCACTGTGTACCCC  
AGGCACTCATGTGCAGTACAGGGTTTTTTTTGTTTTTCCTTTTGTCAAAAAAAAAAAAAATGGGGGCATATTTACAC  
ACAATAAAATTCACCGATTTTAGGAGTACAAACAATTCTGACAGATGTGCACCTTAGTGTAAGCACCACCACAATC  
ATCATATAGAACATTTCTGTCACCCCAAAAAGTTCTGTGACCCTCACCAGTCTGTCTCTGCCTTGACCTTTGGTCT  
CCTGGCAGGTACTGATTTTCTTTGAAACACTTTAGTTTTCTTTTTCTATAAATTTATGTATTGACTCAGACAGTATG  
CCATCTTTTATATAGGATTTCTTTAGTTAGCATGATGCTGTTGAGACGAAACACACCACCCGTGGCCTTTAAAGC  
TTAGGTCTGTGATCCATTTTGTGTTAAAGGATGAATCTTTCCCCATTAATTTGCCATGACACCTTTGTTGA  
ATATCTAATTTATCCTGCCCTGTTTGTGTTAAAGGATGAATCTTTCCCCATTAATTTGCCATGACACCTTTGTTGA  
AAATAACTTGATATCATATGTGTGGGTCTATTTCTGGACTCTCTTTTCTGTTCCACTGATCTGTATGGCCAGCTTT  
ACGGCCAATACCACACTTTGGAGTAATGTAACCTTCAAGTAAGTCTTGAAGTCTGGTGAATGAGTCTGGTGAGTAG  
TCCCTCAACTTAATTTCTTTTCAAATGATTTTGAATATTTCTAGATCCTATGTCTGTATAAATTTATAGACTCCA  
TTTTCTCAGTTTTCTAAAAAAAAAATCTTCTGGTGTGTTGATTGAAATTTGTGTTGAATCTATAGATGATCAATTTTCG  
AAGAATTTATCTTCTAATGATACTGAGTATTTTGATTTCGCAAGTACGATTTCTCTATTTAGGTCTTTAACTCTTCT  
CAACAATGTTTTATAGTTTTTTCAGCCTGCAACTCTTAGACATATTTTGTAAATTTATCTCTTAAGGATTTTATAAT  
TGTTGATGCTATTGTAATGGCTTTTAAATTTTTGATTTTCAATTTGTTTGTGGTAGCATATAGAAATTTAGTCA  
AATTGTATATTTTAACTTACATTTCTGCAGTTTTGCTAAACTCATTTATTTAGTTTTAGTTGCTGTTTTGTAGATTC  
TGCAGAATTTTCTACACCGTCTTTATTTATCTGCAAAATAGAGTTTTACTTCTTTTTTTCCATCTGTTTATCTTTGAT  
TTGCTTTGTCTCAATGAATGACTAGGACCCCCAGCAGCAAGATAAGCAAAAGAAATAGTTCTTACTTATTTTTGA  
GGCCAGAATTAACAATCAAAAACAGACAAAGACATTTAAGTAAAGAAAGGAAATGGCCAATATTTGCAGAG  
ATGCAATAATGTGTAACAAATATAGAAAGGGTTTTAATAAAAAATAGGAACATAGACACTTTTTGCCTGGTTCAGAGT  
CTTAAGGGAAAGCAGTTAGTGTTCATCATAGTATGATGTAATGTAGTTTTCTGATGCCCTATCAGATGAGAGTC  
CCTTCTACTGCTGCCTGCTGARATGCCTTTTCGCTGAGGATGCTCTGCTCTGTAGAGATACTCGTGGTTTTCTCCC  
GCCTTCTATTACTGTGGCAAATTAACCTCCATTTTCTGACAATAAATTAACCTTGCATTTGGTCTTGTCTTTTTATC  
CTTTTTTATATGTTGTTGGATTATCTTTCTAATATTTTGTACACATTTATGTCATCTCTGCAGAATATTTGGTCCAT  
TTCTTTTCTTACTTACAATGTCTTTGTCTGGTTTTGGTATTGGTAATTTCTGGCCTCAAAAAATGAGCTAGGAACT  
ATTTCTTGTATTTCTGTTTTCTGGAGGATTTGTGTTGAAGTATTACTTCTTTGTATGTTTGGTAGAATTTAC  
CAGTGAATCTCTCTGGCCTGGAGTTTTTCTCAGTGAGAAGGATTTTAAACAACAATTTAAATTTCTCTAATAGATA  
TAGGTCTAATCAGGTTATCAGTTTTCTAATGATCTTTGGTCAATTTGTATCTTCTAAAAAATGTATCTATTTTCATCT  
AATTTGTTGAGTGTATTGGCATAAAGTTTTAATAATATTCCTTTTTCTTATTAATGTCTGTGGGTTTTATAGTAAT  
GTACCTCTCTTACTGCTGGTATTCGCATTTCTTCTGATCAATCTGACTGTGTGTTATCAATTTAATAACTTTTTCA  
AAGAAGCAACTTTTTGATCTCAGTGGCGAATTTCTATACTTTAACGTGTGAGGGGAGGGAGTCTTTATTTGAATGATA  
CTCTAATTTATTTCTTGAGATTGAACCTCGAAGTGCCTGGGCTGGAGTGCAGTATCTGGCTCACTGCAGCC  
TTCACCTCCCGGGTTCAAGTGATTCTCTTGCCTCAGCCTCTGAGTAGCTGGGATTACAGGTGTCCACCACCATGC  
CCAGTTAATTTTTTTATCTATAAAAAGGAATGCTAAAAGTTGGGATTAAGGCATGAGCACCACATCTGGCCTATGC  
TCTAATTTTTTACTTCTTTTCTTCTTACTTTGGATTTTATTTGTTCTTTTTCTTATTTCTTAAAGTAGAAGTTT  
AGATCATTGAATTAAGTTTTTTTTTCCCCAATATAAACATTTAATGTTATTGAATGCTCTAACCATTTGATCCCT  
CAAAATTTGGTATGTTATTTTTATTTTTTCTTCACTTAAATAGTTCTTTTTTTTTTTTTTTAAGAGCCAAGGTCTC  
ACTTGACTGCCAGGCTGGAGTGCAGTTGTATGATCCCAGCTCACTGCAGCCTTAAACTCCTGGGCTCAACAGCCC  
TCTGACCTCAGCTTCTGAGTAGCTGGGACTACAGGCATATGCCACTGCGCCTGGCCAGTCTAAAATATTTCTAAT  
TTCACTTGTATTTTCTTCTTTGACCTTTGGGTTATTTAGAAAGGTTTTGATTTCTAATATCTGGAGACTTTTTAGA  
TATTTTTATTTTATTAATTTCTAATTTAAATGTAGTCAGAGAACATTTTTTTTATATTTGTAAATTTACTGCCACAT  
TGTGTAGTCATCTGGAAGTTGGTCCATCTCGATGAATGTTTCATGTGCATTTGAAAAGAGTGTGTGTTCTGCTATT  
GTTGAGTTACGTGTTCTATAAATGTAATTTAGATCAAGTTGATTGATGATGTTGATCCAACCTTCTCTATTTTAACT  
TGATTTTCTATCTGTTTATAAGTTATTATGACATGGCTCTTGTATGTCCAGCTATAAATTTGTGGATTTCTCTATTT  
TTTCTTATAGTGCAATGAGTACTTACTTCAATGTATATATTAGGTGTCTATACACTGAAGGTTGTTATATCCTCTT  
GATGAATTTGATCTTTTTTTTTTTTTATTTTTGAGACGGAGTCTCACTCTGTTGCCAGGCCTAAGTGCCATGGCACA  
TCACAGCTCACTGCAACCTCTGCCTCCCAGGTTAAAGCAATTTCTCCTATCTCAGCCTCCCAAGTAGCTGGGACTAC  
AGTTGCATGCCACCATGCCCAAGCTAATTTTTGTATTTTTGGTAGACACGGGGTTTTACCATGTTTGGCAGGATGGT  
CTCAATGTCTTAACTCGTGATCCACCTGCCTCGGCCTCCCAGAGTGTGGGATTACAGGCGTAAGCCACCCGACC  
CAACCAGATTGATCCCTTTGTCCTGAAATATGTCTTGCATGAAGTCTGCTTTGTCTGATATCAGTGTATCCAGTC  
CAGCTTTATTTGATTAGTGTTTACATGGTATTTTGTTTTTCATCTCTTTACTTTTAACTTATCTGTGTCTTTATA  
TTTTAAAGTGAATCTCTATAACCACCACAGAGTTGGGATTTGCTTTTTATCCATGTGATGATCTCGGCCTTTTACT  
GCAGTCTTGTCTAGACTATTTACGTTTAAATGTGACTATCGATATGGTTAGGTTAAGTTAAATCTTTTGTCTTGT

TGCTGTTTTCTGTTGGTCTCATCTGCCCTTCTTTTTTTCCTCCTACTCTTCTCCTTTTTGGGTTGTGTATTGGCTT  
TTCAAGTATAACCAATTTTTTTAAGTTTTTGGCTCTAGGATTTACTTTTTTTTTATTTTTATTTTTTCATGGAGACTG  
ACTCTCGCTCTGTCGCCCAGGCTGGAGTGCATGGCATGATCTCGGCTCACTGCAACCTTACCTCCCAGATTCAA  
GCTTTTTCTCCTGCCTTGCCCTCAGCTTCCCAACTAGCCGGTACTACAGACATGTGCCACCACACCAGCTGATTTTT  
TCTATTTTTAGTGGAGACAGCATTTACCGTGTGGCCAGGCTGGTCTCAAACCTTCTGACCTCAGGTGATCTGCC  
ACCTCAGCTTCCCAAAGTGTGGAAATTACAGGTGTGAGCCACCATGCCCAGCCTGCTCTAGGATTTACAGTGTACA  
TCTTTTACTTATCACAGTTTACCTTCAAATATTATACCTATTGTATCACTTTTATATATCAGAACCTTAAAATGGAG  
TACTTCCATTTCCCCCTTGTCTTTGTGCTATTTTTGTGCTTTTAAACATGTGTAAAAAGCCCACAATAACATT  
GTTGTTATTTTTGCTTTAAACAGTAATAACCTTTTTAAAGTCTTCTTGTGTGGGTGTTTCATTGTTGTCTTTGTGTT  
TTTTTGACACAGGCTTCACTCTGTTACCAGGTAGAGTGCATGGTGCATCATAGCTCACTGTAGCCTCAAACG  
CCTGAAACAAGTGATCCTTCAGCCTCGGCCCTCTGAATAAGTGGGACTACAGGCATGCACCACCACGCTAAGCTG  
ATTTTTTATTCTTATGGAGATGGGATCTCACTATGTTGCCCAGGCTGGTTTTCAGACTCCTGGCCTCAAGCAGTGGT  
CTCATCTTGGCCTCCCAAAGTGCCAGGATTACAGGTGTAAGCCACCATACCCAATCTAAAGATATTTTTTAGTGAG  
GAAAATTTTTTATTACGTTTACCCATTTAGTTTTTTTTTAAAAAAATTAGACAAATCTTTTTGTTTTGTATTTTTAA  
TTCCTTCAGAGTCAATGAGAAGGATCTAGACGAGTCAAATGCAAGCTAGGAAAGGAATGATCCTAAAGCTTCATAA  
GCAGCAATAAAAAGTATGGTCTTTGGGATGAAGTTTCTCTCTTGTGACAGGAGGAAGGTGAGGAGGGAGTACATATG  
TGATTAATTGTGTATATTTGGTATCATGAAGCTGAGCAATTTGCTTTCTGATGCCCTCCTTTTTTCTTTGTAAAATA  
GGTAAGAGTGAGGGAGGAAAATAGAGAATTTGAAAATAATCATGCTGGGATAGGTTAGTGCATGGACCAGAGGATG  
GTAGCAGGATTTCCAGGCATTTTGTATGCTCCATTTGAGGTTGGTGTATGATGAATTTATTGTAGAAACACTGTACCC  
CATATGTGACATTTTCCGTGCATTTGGCTACTTTGGTTTTAAGTAAGAACAAGCCAATAATTGGCTTTATTGGAGT  
TGGCATTTTGTATGGTTATGAAAAGTCAGAGGGAAAAAGAAGTCTAAAGTATTCACAAGGCGTGTACTGAATGACG  
AACTGAGGACTTAGGCTGAATAGTGAAGGAACTAGGAGAAGACTAAGATAAATTTGTGGTTGAAAATAATTGAACAA  
GCCAATGAGAAAAGTGAAGCTGAGATGTTAATGATTTTTGGAGATGAAACAGTTATGGGTGATGATAAAGGACAATGT  
GTGTAGCCTTGAGAGTGATAGCTCATTAGGAATGGAAAAAGAGATCTTGGGAAGACCAGAGAAGCAGGAGGCTGTT  
CACAGGGAGGAGGATGTTTCTGGAAAATAATGGTTGGGCTGGCTGATGAGGACTGTCTGAGGCAAGTCTTTGAAGA  
ATGAGAACCAGCTTGGCAAATCAGTCAGACAGGCTTTTTCTAATTTATTTGTTGTTGATTTTTGACAGGAAGCAGTG  
**ACTGCAGTATACTTTACTGGTGTCTTACTCTTCAAATCTCAAATGGGGCAGCAATCTCAGGATA**  
**TACAGAAAAAAGAAATATTTCCCATGTCCAGGCTCAGGCCACTGGAGAATGATACAACAGACCCTGAGTGATT**  
**CTCATGACGTATCAGGGACCTGAGAG**GTAAAGAAAGCACTTTAGAAAGCCACTTTTTATATTTTTCATGGTTTTCTTT  
TTTTCTTTTTTGGAGATGGAGTCTCAGTCACCCAGGCTGGAGTACAGTGGCACAATCTCTGCTCACTACAACCCTA  
CCCCCAGGTTCAAGTCATTTCTCCTGCCCTCAGCCTCCTGAATATCTGGGATTACAGGTGCACACCACCACACCTGG  
CTAATTTTTGCATTTTTTAGTAGGGACGAGGTTTTTGTATTTCTTTTTTTTTTCTTCGAGACAGAGTCTCACTCTGTC  
ATCCAGGCTGGAGTGCAGTGGTGCATCTAAGCTCACTGCTACCCTCCGCTCACCTGTTCAAGCGGTTCTCCTGCC  
TCAGCCTCCTGAGTAGTTTGGATTACAGGCATGCACCATCACATCTGGGTTAGTTTTTTGTATTTTTTAGTAGAGAT  
GGCGTTTACCATGTTGTTTCAAGTTGGTCTCAAACCTCCTGGTCTCATGTGATTCACCCACTTCAAGCCTCCCAAAG  
TGCTGGGACTACAGGCATGAGCCACTGCACCTGGCCGAGACAGGTTTTACCATGTTGGCCAGGCTGGTCTTGACCC  
CTGACCCGGGTGATCCACCACCTCATCTCCCAAGTGTGGATCCGGGTGTAGCACCCGGCCAGCTGTGTTGTTCAAT  
GAGGTGTTAAAGACAGACATAAATTTAATGTATTTTTTTATTTTCTCCTCCTTGGGAACATAAATATCATATTA  
TTATGTGCGAAGTCATCATGATGATGATTTCTTAAAAAAGGATGTGATGATAGTGCCACCTATATTGAGTTCCA  
GACTAAGCAAAGCTTACACATCGTGAGACAAGGAAACAGAAAGTCTGTGCTTCCCAAGATGGCACACACGTACAAG  
AACTTTTCATTTAGACTCTCGATTTTTTTTTCTTGTCTAAAAGAAGAGCTAAGGGAATCTCTTTCTTCCATAC  
ACCTCATTAACAACATCATACAGATATTTCCACCAAGAAAAAAGTCTACTTTACCTACTTGACAAGGAAATGCTCT  
AGTTATTTTTTAATTGTATGCCCTCATATGTTAGAAAAATGTAAGGTGTATTAACCAAAAAAGAAATGCTATTGCT  
AGATGACATGATAAAGAACCTGTGTCTGATTGAGGATCTTACGATTTCTGTTTGTGTTTCAAGGTTAAAGCAAGTTGT  
**CCTAAGAGAAGACTAGAAAGGCTAGAGAGTATGAAGCAGCAGCCCCAGTTTTTCTCATGGGCAGCAGGAGGAG**  
**CTGGCCAAGGTGTATAATTGGATTCAAAGCCAGGCTATCCCTCAGGAAATTTGACGTTCAA**GTAAAGCACAGTAATAG  
TGGCTGTCTACTCATGTGTTTTTAAACAGGTGGTGGTTTTTAAAAATAGTTCTTGTTTTTTCCAGGATCCCACAC  
CAACCTTTATTTCTACCATTACACTTCGTTGTAAGTGGTTAATAATTTCTTCTTAACCATTAGACCACCAATTA  
GTAAAACACTACGAGCTTAAAAATAATTAGAGAACCAAGTTAAGCCGGTTAATAGAGGTTTTGGACCCACGGTATCTTT  
TGGCATTATAAGTGGTTTTTTTTCTTCCACCAGCTAAAAGTTGGGAGACAACAAAGAGTCAAAACACCTGCATTGA  
TGTGTGATTGGCTGCATTTGACAAGTGAAGTGTGTTGTTGTTTCAAATATGTCTGCATGTTTGAAGAGTTAAAA  
AGATGTAAGATTACACCCGAGGTGGTATTATCATGTCAAGCAGAACTTTAATGGGACCTAAGTGTCTTTTTGTGTT  
ATGAGGTTCTCAATGAGATCATTGTCAATACCTGAACCAGTAAAGCTTTGATTTACCAATGGTCCCAACCAGT  
GTTGGTGGCAGCTGTGGTAATGGTTGAGATGGTTGTACTAGTAATGAGGCTTTTATTTCCACAAGTATCAAGTTAT  
GGGAGGTTGAAGAGTCAAAGTAAAAGCAATAGCAAATAGGCCCTAGCCAGTGGCTCACATCTGTAATCCCAGCACT  
TTGGGCGGCCACGGTAAGAGGATCAGGAGTTGGAGACCACACTGGGCAGCACAGCCAAACCTCATCTCTAATAAAAA  
ATCAAAAAAGTAGACGGTGTGGTAGCTTGCACTATAGTCCCAACTACTCAGGAGGCTGAGGCAGGAGGATCACT  
TGAGCCTGGGAAATGAGGCTACAATGAGCTATGATGTCAACACTGCACCTCCAGCCAGTCTCAAAAAAGAAAAAGT  
GT  
CTTATAAAGTAAGGCATACGATGAAATCATATACCCCCCTTCTTATACCTCAACCATCTTAAAAATTGACGTATAAT  
ACGTGTGGTATAATATATGTCTTCTTAACTATATCTATGTATATATTTCCGGTTTTTTAATTTTACGATTTCATA

TCTCCATGATCCACTTCCCGAAATCTCTTCACTGAGCAACCTTGAACGTCGACCCATGAAACATAACTTTCCCCAC  
TTTTACTACTGTTGGTGAGAGTATAATTTAGTTCAACCATCATGGAAGACTGTGTGGCGATTCCCTCAAGGATCTAGA  
AATGGAAATTCATGTGACCCAGCAATCCCATTACTGGCTATATACCCAAAGGATTAGAAATCATTCTGTCATAAA  
GACACATGCACATGTATGTTTCGATGTGATGCTGTTTACGATAGCAAAGACCTGGAACCAACCCAAATGCCCATCGA  
TGATAGACTGGCCAAAAAATGTGGCACAATATACCATATGCAGCCATTA AAAACGATGAATTCACGTCCTTTG  
TAGAGACATGGATGAACCTAGAAAACCATCATCTCAGCAAACGACACAAGAACAGAAAACAGACACCCGCATGTT  
TTCACTCATAGGTGTGTGATGAACAATGAGAGCTCTTGGGCACAAGGAAGGGAACAACACTTACAGGGCTAGGTGA  
GAGGGTGAAGGGGAGGAACAGTGGGGTAGGGTGGAGGGGATGGGACGAGGGAACGCCAGAAAAGAAATGCCTGATG  
TAAGCAAGCGAGCAAGGGGAGAATGAAGACAGCAAACCACCATGGCATAATATGTACCTATGCAACAATCATGCAGG  
ATGCAGGACATGCACATGTACCCAAAGCTCAAATACAATTA AAAAACATAAATTTCCCTTCCCTCTTCAAACCTCCA  
CTTCCACCCTGTTCTATCTTCTCTTTGTTCTTTTTTCTTCTGTATGCACAGCAATTTGTTTTTTTTATCCCAACAC  
TTGAAGTCCAATTGCCAATAATCTTGCTCTCAATCCACACTTCTCCCCCAAGGTTTACAACAAGTTTTCTTCCTT  
CACGCTTCCAGAGTAGCTGGGATTGCAGGTGCACACCACCATGCCAGCTAATTTGTTTTTACCATGTTGGCCAGGC  
TGGTCTCAAAGTCTGACCTCAAGTGATCCACCTGCCCTCAGCTCCCAAAGTGCTGGAATTACAGGCGAGAGCCAT  
ATGAACTGATACTCTACATACTATGAGTGAATCATAATTTATGGCTTGTGGACTTTTTATGGTTACGGTTGCAA  
GGTAAATCTTTGAACATCCTGTCACCTTTCGACCTACTTGAATTTAGATTTTAAAGTGAGTAAACAGCATATATCTGG  
TGCTTTTAAAAATCTGTTCACTCAATCTCTTTTGTATTGGAGAATTTAGCCCATTTACATTTAAAGTAATTAAGTGT  
AAAGAAGGACTTACATGTGTCACCTTGTCTATTTGTTTTCTATGTGCCTTAAAAGGTTTTTTCCCTCATTTCCCTGCA  
TTACTGTCACTTTTTGTGTTTGGTTGGTTTCTTGAAGTGAATGTTACGCCCTGCTCTTTTTCCCTGTTGTGTATATG  
GTGTAGCTATTTTCTTTGTGATAACCATGGGGATTACGTTTACGATTTCTAACATTTATATCACTCTAGTTTGAATTT  
ATACCAGCTTAACTTTGGTAACTTACAAAACTCTGCTTCTATAGAGTTCCATCTCCATCTTCTTTTCACTTACTGA  
AGGTATAAAAATAGGTATTTGTATATTTGTGCTTCCAAAAACATCATTTTTTATGTATTAATCTATTAATTTATATAG  
AAATCAAAAATATGGGATTACAAATCAAAGTTAAAATAAATACTGGCTTTTTATAGACTAGTTTTCTTTAAATTTATTA  
GTTTTCTAAAAACAAGAAGTGGAAATATACATAAATTTGTTGCAATAAATAATTTTACCTTCACTGGAATCTTTATTT  
TTTTCATACAGATTCAAGTTACTGTCTTTGATTTCAANCCCTGAAGGATTCCTGTAGCATTTCTTGGATGATAAGT  
CTAGTGGTGAGGCTGGCTTTGTTATCTGGGAATGTCTTCACTTCCCTTACTTTTTGAAGAACAGTATGGCCAGAC  
AAAGCATTTCTTAGTTGATGTTTTTTTTCTTTTAAAGCAATTTGAATACAATGGCCCACTGCCCTTTGGCCCTCCACA  
GTTTTCTGATGAGATACTGCAGATAACCTTATCAAGAATCTTCTGCATGTGATGAGTTGCTTTTTCCCTTGTGCTT  
TTACGAGTCTCTTTGTTTTTAACTTCACTATTTAATTACAATGTGTCTCAGTGTGGATCTCTGAGTTGTGAGTA  
CATCCTAGTTGGAGTTCATTTAGCCTCTTGGATGGTTATATTCATGTCTTTTCACTCAACTCTGGGGAGTTTTTCAGTG  
TTTTTTTTTTTTTTTTTAAATAATCTTTTTGCTCTTTCTCTCTATCTTCCCTGAGGACCTCCATAGTGCTTACGT  
TGCTCCACTTGATGGTATCCACATGTCTCCAGGCTCTGTTGACTTTTTCTTCACTGCTTTTTCTTTTTGTTTCTC  
AGACTTGACGATTTCAACTGTCTGTCTGCATGTACAGATTTCTTCTTCTTCTGCTCACATCTGCTTTTGAATCC  
CTCTAGTGAATTTTTTATTTCACTTTTGTACTTTGTATTTACAGAATTTCTTCTTCTGGCTTCTTTTTGGGTTTTCTC  
TCTCTTTATTGATATTTCCGTTTGTGTTGATTTTCTTGTATTTTCTCCATGCCTTTCTTTAATTTCTTTCA  
TCATATTTTCTTTTCTTTTTTTT TTTAAGACAAGGCTCCTCACTCTGTACCTAGGCTGGAATGCAATGGCATGATCCT  
GGCTCACTGCAGCTCAGCCTCCTGAGCTCAGGCAATCCTGCCACATCAGCTTCCCAAGTAGCTGGCACCCTACC  
ATGACCAGCTAATTTTTTTAGTTTTTTGTAGAGATGGGGTTTTTGCCACATTTGCCAGGCTGGTCTTGAACGCTTGAGC  
TCAAGCAATCCACCTGCTTTGGCTCCTGAAGTGCTGGGATTATAGCCGTGGGCCACTGCACCCAGCTTTTTTTTTT  
TTCTTTTTCTTTTTTTTTTTTTTTTAAAGGCACTCTCACTTTGTCAACCAGGCTGGAGTGCAGTGGCATGATCACAGC  
TCACTGCAGCCCTCATCTCCTAGTTTTCAGGTGATCCTCTCAAGCTGGGACTACGAGTGTGCACCACAATGTCTGGC  
TAATTTTATTTTATTTTTTTTAGAGATGAGGTCTCACTATGTTGCCAGCCTCAAGCCATCCACCTGCCTGGGCCTT  
CCAAAGTGCTGGGATTAAAGGCAAAAAGCAACAAGCCCCACCGCCCTAAGGACACTGACCAAGGTA AAAAANN  
NNNNNNNAAGTACATATGATTGCAGGCGAGCTGACATGCAATGGTTAATCTATGCACTTATAAGAGTGGGA  
GACTGGAGTTTCTCCATAGGAGTCAGGCTGGTCTAGAACCTCGTGAATGAGGTTATCCATCCAGAGTACAGAGAAGCA  
AAGTTCTGGCAGGACAAGCATGAGCCACTGCCCCAGCCAAGGCAATACATTCATGTGGAGAAAAGATATAAATAAGC  
TTTATAGGTGACAAGGCAGATTTACAGAAAATTTTTCAATTAGTTCTTTTTTGGTGACGGATTACCTTTACTTTAC  
TTTGACAGGACCTCCTTATTTCTACTGTGAAGAGGAGACAATCATCGTATCGCAGGCCCACGCTGGTCAAGACCAC  
TTGACCAGGTTGGTCAAGAAACCAGAGACGAGCGGTCCGGAGGAAGATGCGTGTAAATCCAGCACTTTGGGAGG  
CCGAGGTGGGTGGATCACGAGGTCAAGAGATCGAGACCATCCTGGTCAACATGGTGAACCCCGTCTCTACTAAAA  
ATACAAAAAATATCTGGGCATGGTGGCACGTGCCTGTAATCCAGTTACTCAGGAGGCTTAGGCGGGGAGAATTTCT  
TTGAACCCAGGAGGCGGAGGTTGCGGTGAGCCAAGATCGCGCCATTGCACCTCCAGCCTGGGTAACAAGAGCGAAAC  
TCCATCTCAAAAAA AAAAAAATTTGTCTTGGGCCGCACATAAAAATATACTAACAATAGCTGATGAGCGAAAAAAT  
AATAAGTCAACAAAAAACCTCATGTTTTAAGAAAGTTTACAAACTTGTGTTGGGCCACATATCAAACCCAACCAG  
CGCTCCAGGGGCCCTGACCACCAGGTGGACAACCTTGGGCTGTAATTTGCCATTTCTTGTAAAGCCCCAAGAATTT  
GTTTTTTTTTTTTGGTTGAAAAATAGGCATGTGAATCTAAGTCTGTAATTTCTGGAACCATATTTCTCTCTTTCCCAGG  
GTGTATTGATTATTTGTTTTTCAATTTTTGATTTGTTGGAAGACAGGGCCCTTATTTCCCACCCTGACTCCCACAAGCT  
GTATGTAAGCTGCTCCAGGAATATGTGCCTTGCCAGGGTCAAGGATGGGTAGTTGCTACTGTGCTTCAAGCTGAA  
ATTGATCACAGACAACCACAGCTTACCCTTAAGCATTTCCCCGAAAGTTGCAGGCCTTCAAATACACTCAAGTTC  
CAAATGTTACCTCGGACTGATTTTGTGAGTGCAGTTGTTGCTTAGGTGAGGAGATAGAGTCTTAGTCTCTCTACT  
CTGCCATCTTCTCAGAATCATCTCCCATCTCCATTTTCAAAGAAAAGAATGAGTGTCTGTGATGTAAGAATGACT

GGATAGGGTCTCACCAAAAAGCAGGAGCTGGGCTGGTTACTGTAAATGTCTTTGTTTTGTTTCATTGTCGTTCTTC  
TGGTTTTAAATAGACACGTATATGAAGGATGAGCCCCACAGCAATTTAAACAAAAGCGCAGTCCATGCACCACTTA  
GTGACGGGTCTGTTTTGTGCAAGCATCCTGGAGTGTCTCACACAAACCGAGCTGGGGTAGCTTGCTACACTCACT  
CCTAGGTTCTATGGGGAAGCTGTTTGATTCTAGCCTGAATACTGTAGTCAGTTGTAACACAGTGGCACTTGTGTAT  
CACACATAGAAAAGGTACCATTTAAAATATGGTAGAATAATCTTAAAAGAAAGACCACCATCATATATGTAGTCTGT  
TGTTGACTGAAATACCATGCGGCACATTACTCTATTTAAACAATTTGTCATGTGCCTTTTAATAATAAAGAAGAGC  
TCTAATGAATATTTTCAGACATGAATAAGTTGGGGAGTTAGAAACATGTTGCTTGACCAGCCTTTTACTGTTTTAA  
AACTCTTCAGGTGACATTGACATCAGCTAACGCACCTGCTTTTGTCTTTTTTTGGAGGCCTGTGTCAATTGCGACA  
GTGCAGATTCAGGTGATGATGCAGCCACATCCAGTGGTCAGGCAGAAGACAGCAGCTGCAATTGCATGGGATGCTA  
CACTCCCCTGCCAGGGGAGCCTGACCCAGAGGACAGCAGGCATGTCAATCCGATGTGTTG

**Anexo 4.** Sequência do gene *Per3* de sagüi (CJLAB). As regiões destacadas em azul representam os éxons, enquanto que a região promotora está destacada em rosa. Os íntrons estão marcados em marrom. Os nucleotídeos destacados em amarelo representam os nucleotídeos adquiridos pelas as sequências referências disponíveis na plataforma UCSC e na plataforma ENSEMBL.

## **Analysis of transposable elements in clock genes in the human genome**

Flávia Cal Sabino<sup>1\*</sup>, Alexandre Rossi Paschoal<sup>2</sup>, Rosa Maria R. P. S. Castro<sup>3</sup>,  
Mario Pedrazzoli<sup>1</sup>, Sergio Tufik<sup>1</sup>.

<sup>1</sup> Departamento de Psicobiologia, Universidade Federal de São Paulo, 04024-002, São Paulo, SP, Brazil.

<sup>2</sup> Programa interunidades de Bioinformática, Instituto de Matemática e Estatística,

Universidade de São Paulo, 05508-900, São Paulo, SP, Brazil.

<sup>3</sup>Núcleo de Psicopatologia, Políticas Públicas de Saúde Mental e Ações Comunicativas em Saúde Pública da Universidade de São Paulo - NUPSI/USP, Faculdade de Saúde Pública, Universidade de São Paulo Av. Dr. Arnaldo, 715 CEP: 01246-904 - São Paulo, SP, Brazil.

\*Corresponding author:

Flávia Cal Sabino  
Departamento de Psicobiologia  
Universidade Federal de São Paulo  
Rua Napoleão de Barros, 925  
04024-002, São Paulo, SP, Brasil  
[flaviacal@gmail.com](mailto:flaviacal@gmail.com)  
Phone: +55-11-21490155 R. 251



## Abstract

Transposable elements are a powerful biological mechanism for increasing variability in genomes. Transposable elements promote some evolutionary changes such as recombination or acquisition of specific functions. On the other hand, circadian rhythms are very important for physiological regulation of a myriad of functions, including system clock control, body temperature, feeding behavior, hormone secretion and others.

In this report, we analyze the distribution of transposable elements in major clock genes (*BMAL1*, *CLOCK*, *CRY1*, *CRY2*, *CSNK1D*, *CSNK1E*, *PER1*, *PER2* and *PER3*) in the human genome. We identified 885 transposable insertions in the 9 clock genes, of which almost 94% are located in intronic regions and less than 1% in coding regions. A large proportion of identified transposable elements (~83%) are related to SINE and LINE classes. We also analyzed the type of classes of transposable element, gene location and degree of proximity between TE and transcription factor binding sites (TFBS) and CpG islands.

This is the first reported study of the distribution of transposable elements in circadian rhythm-related genes. These results outline the distribution of TEs in upstream, exon, intron and downstream regions of these genes, and near TFBS and CpG islands and suggest they could influence the maintenance of circadian rhythm.

**Abbreviations:** TE: transposable elements;

**Keywords:** transposable elements, clock genes, circadian rhythms.

## INTRODUCTION

Transposed elements (TEs) are mobile genetic sequences that are able to move or replicate within the genome. They represent about 50%, 38% and 15-22% of the human, mouse and fruit fly genomes, respectively (Lorenc and Makalowski, 2003; Waterston et al., 2002; Consortium, ICGS, 2004; Biemont and Vieira, 2006). TEs are classified by their mode of replication. Class 1 transposed elements (also called retrotransposons) are propagated by reverse transcription of an RNA intermediate. Long interspersed repeat elements (LINE), short interspersed repeat elements (SINE) and retrovirus-like elements with long-terminal repeats (LTR) belong to this class. In contrast, elements of class 2 transposed elements (DNA transposons) move through a mechanism of “cut-and-paste” (Hedges and Batzer, 2005; Wicker et al., 2007).

Initially, TEs were believed to be genomic parasites that existed exclusively for self-replication in the host genome (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). TE sequences were referred to as “junk DNA” and were believed by some to be the “draft” for natural selection in evolutionary experiments (Makalowski, 2000). However, in recent years, it has become clear that TEs are extremely important in genome function (Kidwell and Lisch, 2001) and that they influence various aspects of gene structure, function and expression (Britten, 1996; Lorenc and Makalowski, 2003; Jordan et al., 2003; van de Lagemaat et al., 2003; Ganko, 2006; Almeida et al., 2007).

Some effects of TEs are negative and some of these are caused by unequal homologous recombination events. These events can be deleterious to the fitness of the organism (Deininger and Batzer, 1999; Callinan and Batzer, 2006) and have been related to human genetic disorders (Teugels et al., 2005; Callinan and Batzer, 2006; Franke et al., 2009). Some, however, increase genetic diversity and genomic expansion and contribute positively to genome organization (Deininger et al., 2003; Hedges and Batzer, 2005). TEs interact and influence the evolution of the entire host genome (Lorenc and Makalowski, 2003), as well as the expression of specific genes when inserted in promoter regions (Brosius, 1999). They introduce new regulatory sequences and can interfere with the epigenetic status (Duhl et al., 1994; Waterland and Jirtle, 2003; Feschotte, 2008) of a promoter region, creating new patterns of

expression of nearby genes (Britten, 1996; Britten, 1997; Medstrand et al., 2005). Brosius (1999) identified almost 80 cases of regulatory elements in vertebrate genes that are derived from TEs.

When inserted into introns, TEs can affect alternative splicing sites (Varagona et al., 1992; Davis et al., 1998). There are many cases of TE domestication in which the element is incorporated into an exonic region (Nekrutenko and Li, 2001; Lorenc and Makalowski, 2003; Ganko et al., 2003; Britten, 2004, Almeida et al., 2007). When such a TE is present inside an open reading frame, it is known as a TE-cassette. In this manner, TEs can contribute to the evolution of coding regions, resulting in the production of new protein isoforms and generation of genetic novelties (Nekrutenko and Li, 2001; Lorenc and Makalowski, 2003; Ganko et al., 2003; Britten, 2004). When inserted into the 3'UTR, these elements may serve to introduce alternative polyadenylation. TE insertion elongates the UTR and serves as a binding site for microRNA (Sugiura et al., 1992; Mager et al., 1999; Smalheiser and Torvik, 2006; Feschotte, 2008).

Based on the above, transposed elements possess high potential to contribute to protein variability. However, there are few studies investigating how TE-cassettes influence protein properties and functions (Lorenc and Makalowski, 2003).

Circadian rhythms have been widely studied in several organisms ranging from bacteria to human, and play an important role in control of physiological processes in living systems (Lowrey and Takahashi, 2004; Bell-Pedersen, et al., 2005, Hazen et al., 2009). Their main roles include control of the sleep-wake cycle, sleep disorders (Panossian and Avidan, 2009), system clock control, body temperature, feeding behavior, hormone secretion, drug and xenobiotic metabolism, glucose homeostasis, and cell-cycle progression (Hastings et al., 2008). The circadian system entrains endogenous processes to the solar day of 24 hours (Gibson et al., 2009), recognizing the local time given by natural light/dark cycle, measuring time passage and thereby contributing to the regulation of homeostasis and seasonal behavior (Scwartz & Meijer, 2004). A major feature of the circadian cycle is the rhythmic expression of several messenger RNAs and proteins. It is likely that genetic variation in clock genes

contributes significantly to physiological variation and also to differences in disease susceptibility (Takahashi et al., 2008).

In this study, we focus on TEs in clock genes, which are very important in the maintenance of biological systems that are regulated in a circadian manner. The clock genes are essential for maintaining circadian rhythms under constant or entraining conditions (Cermakian and Boivin, 2003; Reppert and Weaver, 2002). Although there are several studies with clock genes, there is a gap in knowledge of the relationship of TE to these genes. So far, very few reports have described the contributions of TE to the function of the clock genes and other genes that display strong circadian oscillations (Puech et al., 1997; Dupressoir et al., 1999; Gotter & Reppert, 2001). For this reason, the aim of the present study is to analyze the presence of TEs in the nine main human clock genes (*BMAL1*, *CLOCK*, *CRY1*, *CRY2*, *CSNK1D*, *CSNK1E*, *PER1*, *PER2* and *PER3*) as a basis for considering the possible influence of TEs in the expression of these genes.

## METHODS

### *Data collection and analysis*

All gene sequences (exons and introns) from *BMAL1* (ENST00000389707), *CLOCK* (ENSG00000134852), *CRY1* (ENST00000008527), *CRY2* (ENST00000263762), *CSNK1D* (ENST00000314028), *CSNK1E* (ENSG00000213923), *PER1* (ENSG00000179094), *PER2* (ENSG00000132326) and *PER3* (ENSG00000049246) were extracted from Ensembl and converted into a FASTA-formatted sequence in January 2009. The FASTA sequence also included 3 kb of the upstream and downstream regions of each gene, as used in Almeida et al. (2007) and based on studies in mammalian genome (Jordan et al., 2003; van de Lagemaat et al., 2003; Thornburg et al., 2006; Landry and Mager, 2003).

### *Bioinformatics analysis of TE in clock gene sequences*

RepeatMasker 3.1.8 (unpublished; <http://www.repeatmasker.org>) software was used to screened gene sequence files to identify TE insertions. This software uses the library Repbase (Jurka et al., 2005) of repeat sequences for alignment and identification of TEs.

For all genes described, the following parameters were used to search the TE: “wu-blast” as the search engine; “slow” to obtain a search 0-5% more sensitive than default; “noslow” to not mask low complexity DNA or simple repeats; “norna” to not mask small RNA (pseudo) genes; “species human” to specify the species or clade of the input sequence. All the alignments with at least 100 nucleotides (nt) were analyzed. All these parameters avoid spurious results.

A TE insertion was designated as a TE-cassette when a fragment of a TE was inserted into an mRNA coding sequence (Gotea and Makalowski, 2006). TE insertions were also mapped in exonic, intronic, upstream and downstream regions. TE was considered to be in exonic regions if they were completely inside the exonic region or if at least 5% of the TE overlapped the exonic region.

For all these tasks, an in-house PERL script was developed to filter and optimize the manually curated analysis of the RepeatMasker output.

### *TFBS and CpG Island analysis*

Transcription factor binding sites (TFBS) and CpG islands present in the 3 kb upstream and the first exon regions of each gene were analyzed against the RepeatMasker output. The information on transcription factor binding sites and CpG islands was extracted from the UCSC Genome Browser (Kuhn et al., 2009; Karolchik et al., 2007) (Human Genome). Only transcription factor binding sites consisting of at least 4 nt were considered.

### *Correlation between gene length and number of TE*

We also analyzed the relationship between the lengths of all nine clock genes and the number of transposable elements found using Pearson correlation.

## RESULTS AND DISCUSSION

### *General distribution of TE insertions in clock genes*

Table 1 shows the number of TEs and gene lengths; the latter includes both 3 kb upstream and downstream regions (see Methods). It is evident that the relationship of gene length and TE number in genes is not the same for all genes. This raises a question about the analysis of the correlation between number of TEs and clock genes length, as we discuss below.

Within the nine clock genes analyzed, a total of 885 TE insertions were present, with an average of 98.3 insertions/gene. The *CRY1* gene displayed the largest number of insertions (189 TE), while *PER1* showed only nine. SINE and LINE classes comprise ~83% of the total TE insertions identified. *CSNK1E* gene possessed the most SINE elements (112), while *CSNK1D* (14) and *PER1* (6) contained a smaller number of SINE (Table 2). Of the *PERIOD* homologs (*PER1-3*), the *PER3* gene showed 101 TE insertions, more than twice as many as the sum of *PER1* (9) and *PER2* (37). The same occurred for *CRYTOCROMES* and *CASEIN* homology genes, of which *CRY1* and *CSNK1E* display many more TE insertions. No LTR DNA classes were found in the *PER1* gene.

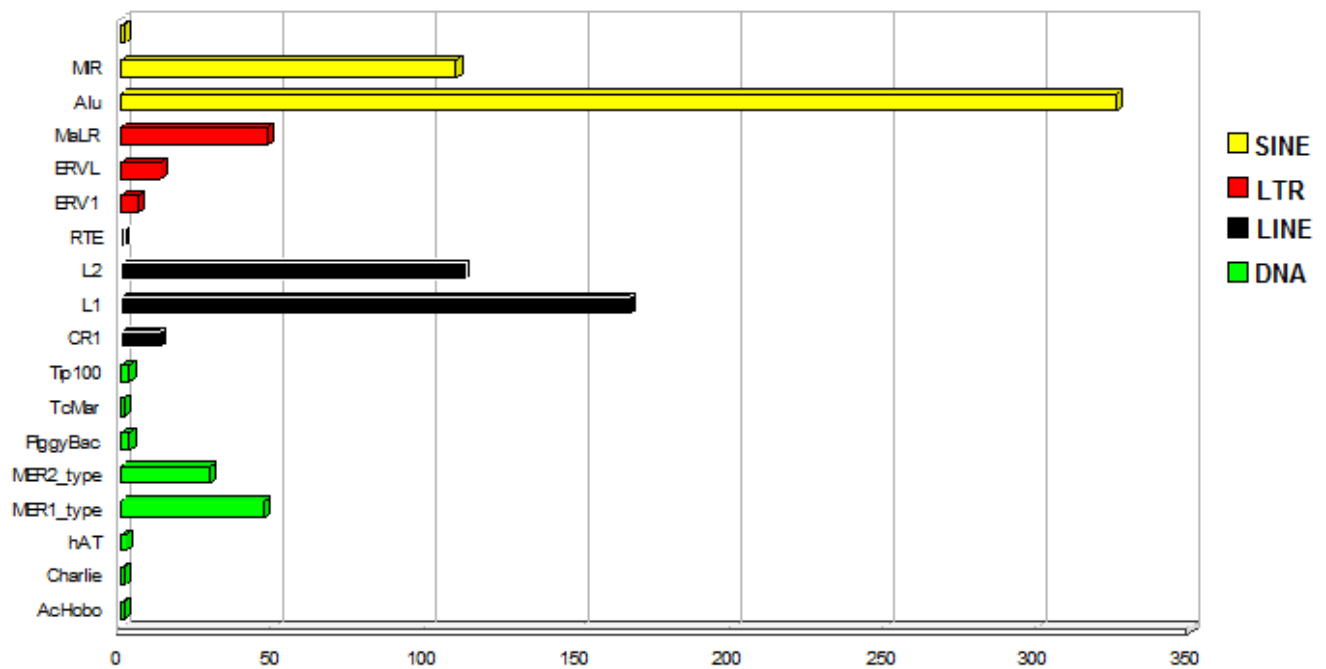
Figure 1 shows the large number of types of Alu and MIR (SINE-class) and types L1 and L2 (LINE- class) elements that appear in the clock genes analyzed. Alu elements are the most abundant family of repeats in the human genome [48] and these results agree with this evidence. Alu repeats are known to be related to several human diseases [48].

**Table 1.** List of each clock gene analysed with the number of TEs and gene length. This last one have include the 3kb in upstream/downstream region.

<b>Gene</b>	<b>Amount of TEs</b>	<b>Gene Length</b>
Bmal1	146	115480 bp
Clock	177	120338 bp
Cry1	189	108181 bp
Cry2	49	41768 bp
CSNK1D	22	35328 bp
CSNK1E	155	113830 bp
PER1	9	17894 bp
PER2	37	50529 bp
PER3	101	66858 bp
<b>Total</b>	<b>885</b>	<b>670206</b>
<b>Average</b>	<b>98,3</b>	<b>74467 bp</b>

**Table 2.** Distribuion of TE insertions in genes and regions

	<b>Upstream</b>				<b>Exon</b>			<b>Downstream</b>			<b>Intron</b>					<b>Total</b>
	LINE	SINE	LTR	<b>Total</b>	DNA	SINE	<b>Total</b>	SINE	LINE	<b>Total</b>	DNA	LINE	LTR	SINE	<b>Total</b>	
Bmal1	4	1		<b>5</b>	1	1	<b>2</b>	2	0	<b>2</b>	15	61	7	54	<b>137</b>	146
Clock		2	1	<b>3</b>	0	0	<b>0</b>	0	0	<b>0</b>	21	62	15	76	<b>174</b>	177
Cry1	3		3	<b>6</b>	0	0	<b>0</b>	5	0	<b>5</b>	20	67	23	68	<b>178</b>	189
Cry2	2	4		<b>6</b>	0	0	<b>0</b>	1	0	<b>1</b>	2	9	4	27	<b>42</b>	49
CSNK1D		1	2	<b>3</b>	0	0	<b>0</b>	0	0	<b>0</b>	1	5	0	13	<b>19</b>	22
CSNK1E				<b>0</b>	0	1	<b>1</b>	4	0	<b>4</b>	14	28	1	107	<b>150</b>	155
PER1	1	2		<b>3</b>	0	0	<b>0</b>	2	2	<b>4</b>		0	0	2	<b>2</b>	9
PER2	1			<b>1</b>	0	1	<b>1</b>	0	1	<b>1</b>	4	8	5	17	<b>34</b>	37
PER3	2	1		<b>3</b>	0	0	<b>0</b>	3	0	<b>3</b>	9	38	6	42	<b>95</b>	101
<b>Total</b>	13	11	6	<b>30</b>	1	3	<b>4</b>	17	3	<b>20</b>	86	278	61	406	<b>831</b>	<b>885</b>
<b>Average</b>	2	2	2		0	0		2	0		11	31	7	45		98



**Figure 1.** The distribution of repeats by class of TE in nine clock genes

### *Correlation between the length of clock genes and the number of TE*

Our results show that, for each of the nine genes, a high and positive correlation between gene length and number of TE was observed. We obtained a Pearson correlation with 0.967 and p-value of 0.000. However, a larger sample is needed to evaluate whether this correlation also holds for other genes that were not included in our analysis.

### *TE insertions mapped in upstream, exon, intron and downstream regions*

It is possible to group the distribution of TE insertions in the gene regions (upstream, downstream, exonic and intronic regions). As showed in Table 2, the TEs are distributed along the gene regions with 30 TEs (3.3%) in upstream regions, 20 TEs (2.2%) in downstream regions, only 4 TEs (<1%) in exonic regions and 831 TEs (~94%) in intronic regions.

The majority of TE insertions occur in introns, with the exception of *PER1*



which contains only two TE in this region. Four of the genes (*PER1*, *CSNK1D*, *PER2* and *CRY2*) together contain only ~12% (97) of the intronic TEs identified, while *PER3*, *BMAL1*, *CSNK1E*, *CLOCK*, *CRY1* genes contain the other 88% (734) of the intronic TEs. There is a divergent number of TE insertions in introns between paralog genes (*PERs* 1-2-3; *CRY1* vs. *CRY2*; and *CSNK1D* vs. *CSNK1E* – Table 2). On the other hand, the homologs *PER1* and *PER3* have similar numbers of TEs in upstream/downstream regions compared against *PER2*. This distribution is not similar in *CASEIN* genes, and in *CRYTOCROMOS* genes it is only similar in the upstream region.

*BMAL1*, *CSNK1E* and *PER2* display TEs in exons. *BMAL1* presents two TE insertions overlapped with exonic regions. One is a SINE/MIR type with the gene 24% (28 nt) overlapped in exon; the other is a DNA/MER1 type with ~43 (101 nt) in the exonic region. Both *CSNK1E* and *PER2* contain SINE/Alu elements of 253 and 306 nt, respectively, completely inside the exonic region.

In downstream regions, the *CRY2* and *PER2* genes each presented only one TE, while *CSNK1E* and *PER1* each contained four TE insertions. In upstream regions, the number is the same in *CLOCK*, *CSNK1D*, *PER1* and *PER3*, each with three TEs and in *CRY1* and *CRY2*, each of which has six upstream TE insertions.

*CSNK1D* and *CLOCK* have TEs only in introns and in upstream regions, while in *CSNK1E* there are no upstream TEs. Finally, LINE and LTR class elements are not found in exonic regions, while DNA class elements are also absent in both upstream and downstream regions. No TEs were found in the downstream regions of *CLOCK* and *CSNK1D* genes.

#### *TE insertion orientation in relation to clock genes*

A TE insertion can be in the sense or opposite orientation with respect to the host gene. In each of the nine clock genes, the frequencies of sense and antisense insertion orientations were analyzed. In general, the distribution of sense was 50.73% and antisense 49.27%. This results show a very similar distribution of TE insertions with respect to orientation. Orientation was also analyzed with respect to TE class. SINE class elements display somewhat more

TE in antisense (25.42%) than in sense (23.95%) orientation. On the other hand, DNA, LTR and LINE are somewhat preferentially inserted in the sense (5.08%, 4.63% and 17.06%, respectively), than in the antisense (4.75%, 2.94% and 16.16%, respectively) direction.

Analysis with respect to gene region insertion shows that upstream and intronic TEs display slightly more sense insertion (1.92% and 49.15%, respectively) than antisense insertion (1.47% and 44.75%, respectively). As previously mentioned, only four TE insertions were found in exonic regions, two in sense and two in antisense orientation. Downstream regions are the only regions that display somewhat more TEs in antisense (1.36%) than sense (0.9%) orientation.

#### *TFBS and CpG islands near TE insertions*

We looked for TE insertions near transcription factor binding sites (TFBSs) and CpG islands, based the search on UCSC Genome Browser data. *BMAL1*, *CSNK1D* and *CSNK1E* showed no TE elements within 500 nt distance of a CpG island or TFBS (see Methods). Considering the same distance as a limit, we observed the following results: *CRY1* gene showed one LINE/L2 element 453 nt upstream from a CpG island; *CRY2* contained one SINE/MIR located 181 nt upstream from a CpG island; *CLOCK* gene contained two TFBS located 226 nt and 261 nt downstream from one LTR/MaLR element; *PER1* contained 3 different elements (SINE/MIR, SINE/Alu and LINE/L2) near several TFBS. The first TE (SINE/MIR) is located 328 nucleotides upstream of a TFBS. The second TE (SINE/Alu) is surrounded by eight TFBS upstream and six downstream. The eight TFBS observed lie between the SINE/MIR and SINE/Alu elements, while the latter six TFBS are located between the second SINE/Alu and the third TE (LINE/L2). *PER2* contains a LINE/L2 element 276 nt upstream from one TFBS, and *PER3* has a LINE/L2 element 371 nt upstream from a CpG island and another SINE/Alu element overlapping the end of this same CpG island. 149 nt and 379 nt downstream of this SINE/Alu element, another CpG island and one TFBS, respectively, occur.

Although we mapped both TFBS/CpG islands and TE elements, we found that only *PER3* contains a CpG island overlapped with a TE element. No TE elements overlapped with TFBS.

These results raise a question about the influence of TE proximity to TFBS or CpG islands and whether such proximity might affect the final circadian phenotype. However, further experiments are necessary to confirm this possibility.

## **CONCLUSION**

It is known that transposable elements play an important role in genetic variability of organisms, and it has been shown that this variability sometimes results in positive consequences. In this report, we provide information on the presence of transposable elements within the sequences of nine human clock genes. The transposable elements report here are located for the most part in non-coding regions, with a very small number present in coding sequences. The results also show a difference in the numbers of TEs present in paralogous genes and show that clock gene length is correlated with number of TE insertions.

Our results showing the distribution of TEs between upstream, exonic, intronic and downstream regions of genes and near TFBS and CpG islands suggest that TEs can possibly influence the maintenance of circadian rhythm, and open future opportunities for studies on how TE insertions near TFBS/CpG islands may interfere with host clock protein functions.

### **Authors' contributions**

FCS performed the initial review of the literature for writing the manuscript that was analyzed by ARP. FCS and ARP contributed equally to writing, analysis and discussion of the report. ARP participated in application of RepeatMasker software and performed the statistical correlation analysis. RMRPSC and MP helped to draft the manuscript. ST helped to review the final version of the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We thank Artur Trancoso Lopo de Queiróz and Pablo Nehab Hess for their suggestions and review of the manuscript. Also we thank FAPESP, AFIP and CEPID for financial support.

## Reference

1. Lorenc A, Makalowski W: Transposable elements and vertebrate protein diversity. *Genetica* 2003, 118:183-91.
2. Waterston RH, Lindblad-Toh K, et al: Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002, 420:520-62.
3. Consortium ICGS: Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 2004, 432:695-716.
4. Biemont C, Vieira C: Genetics: junk DNA as an evolutionary force. *Nature* 2006, 443:521-524.
5. Wicker T, Sabot F, et al: A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 2007, 8:973-82.
6. Hedges DJ, Batzer MA: From the margins of the genome: mobile elements shape primate evolution. *Bioessays* 2005, 27:785-94.
7. Doolittle WF, Sapienza C: Selfish genes, the phenotype paradigm and genome evolution. *Nature* 1980, 284:601-3.
8. Orgel LE, Crick FH: Selfish DNA: the ultimate parasite. *Nature* 1980, 284:604-7.
9. Makalowski W: Genomic scrap yard: how genomes utilize all that junk. *Gene* 2000, 259:61-7.
10. Kidwell MG, Lisch DR: Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* 2001, 55:1-24.
11. Britten R: Cases of ancient mobile element DNA insertions that now affect gene regulation. *Mol Phylogenet Evol* 1996, 5:13-7.
12. Almeida LM, Silva IT, Silva WAJ, Castro JP, Riggs PK, Carareto CM, Amaral ME: The contribution of transposable elements to *Bos taurus* gene structure. *Gene* 2007, 390:180-189.
13. Jordan IK, Rogozin IB, Glazko GV, V KE: Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet* 2003, 19:68-72.
14. Ganko EW, Greene CS, Lewis JA, Bhattacharjee V, F MJ: LTR retrotransposon-gene associations in *Drosophila melanogaster*. *J Mol Evol* 2006, 62:111-20.
15. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P: Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet* 2003, 19:530-6.
16. Callinan PA, Batzer MA: Retrotransposable elements and human disease. *Genome Dyn* 2006, 1:104-15.
17. Deininger PL, Batzer MA: Alu repeats and human disease. *Mol Genet Metab* 1999, 67:183-93.

18. Franke G, Bausch B, et al: Alu-Alu recombination underlies the vast majority of large VHL germline deletions: Molecular characterization and genotype-phenotype correlations in VHL patients. *Hum Mutat* 2009, 30:776-86.
19. Teugels E, De Brakeleer S, et al: De novo Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes. *Hum Mutat* 2005, 26:284.
20. Deininger PL, Moran JV, Batzer MA, Kazazian HHJ: Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev* 2003, 13:651-8.
21. Brosius J: Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* 1999, 107:209-38.
22. Feschotte C: Transposable elements and the evolution of regulatory networks. *Nat Rev Genet. Review* 2008, 9:397-405.
23. Waterland RA, Jirtle RL: Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* 2003, 23:5293-300.
24. Duhl DM, Vrieling H, Miller KA, Wol\_ GL, Barsh GS: Neomorphic agouti mutations in obese yellow mice. *Nat. Genet.* 1994, 8:59-65.
25. Medstrand P, van de Lagemaat LN, et al: Impact of transposable elements on the evolution of mammalian gene regulation. *Cytogenet Genome Res* 2005, 110:342-52.
26. Britten R: Mobile elements inserted in the distant past have taken on important functions. *Gene* 1997, 205:177-82.
27. Varagona MJ, Purugganan M, R WS: Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. *Plant Cell* 1992, 4:811-20.
28. Davis MB, Dietz J, Standiford DM, Emerson CPJ: Transposable element insertions respecify alternative exon splicing in three *Drosophila* myosin heavy chain mutants. *Genetics* 1998, 150:1105-14.
29. Ganko EW, Bhattacharjee V, Schliekelman P, McDonald J: Evidence for the contribution of LTR retrotransposons to *C. elegans* gene evolution. *Mol Biol Evol* 2003, 20:1925-31.
30. Nekrutenko A, Li WH: Transposable elements are found in a large number of human protein-coding genes. *Trends Genet* 2001, 17:619-21.
31. Britten RJ: Coding sequences of functioning human genes derived entirely from mobile element sequences. *Proc Natl Acad Sci* 2004, 101:16825-30.
32. Smalheiser NR, Torvik VI: Alu elements within human mRNAs are probable microRNA targets. *Trends Genet* 2006, 22:532-6.
33. Mager DL, Hunter DG, Schertzer M, Freeman JD: Endogenous retroviruses provide the primary polyadenylation signal for two new human genes (HHLA2 and HHLA3). *Genomics* 1999, 59:255-63.
34. Sugiura N, Hagiwara H, Hirose S: Molecular cloning of porcine soluble angiotensin-binding protein. *J Biol Chem* 1992, 267:18067-72.
35. Cermakian N, Boivin DB: A molecular perspective of human circadian rhythm disorders. *Brain Res Brain Res Rev* 2003, 42:204-20.
36. Reppert SM, Weaver DR: Coordination of circadian timing in mammals. *Nature* 2002, 418:935-41.
37. Hazen SP, Naef F, et al: Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays. *Genome Biol* 2009, 10:R17.
38. Bell-Pedersen D, Cassone VM, et al: Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nat Rev Genet* 2005.
39. Lowrey PL, Takahashi JS: Mammalian circadian biology: elucidating genome-wide levels of temporal organization. *Annu Rev Genomics Hum Genet* 2004, 5:407-41.
40. Panossian LA, Avidan AY: Review of sleep disorders. *Med Clin North Am* 2009, 93:407-25, ix.

41. Hastings MH, Maywood ES, Reddy AB: Two decades of circadian time. *J Neuroendocrinol* 2008, 20:812-9.
42. Gibson EM, Williams, W P 3rd, et al: Aging in the circadian system: considerations for health, disease prevention and longevity. *Exp Gerontol* 2009, 44:51-6.
43. Schwartz WJ, Meijer JH: Real-time imaging reveals spatiotemporal dynamics of cellular circadian clocks. *Trends Neurosci* 2004, 27:513-6.
44. Takahashi JS, Hong HK, et al: The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nat Rev Genet* 2008, 9:764-75.
45. Gotter AL, Reppert SM: Analysis of human Per4. *Brain Res Mol Brain Res* 2001, 92:19-26.
46. Dupressoir A, Barbot W, et al: Characterization of a mammalian gene related to the yeast CCR4 general transcription factor and revealed by transposon insertion. *J Biol Chem* 1999, 274:31068-75.
47. Puech A, Dupressoir A, et al: Characterization of two age-induced intracisternal A-particle-related transcripts in the mouse liver. Transcriptional read-through into an open reading frame with similarities to the yeast ccr4 transcription factor. *J Biol Chem* 1997, 272:5995-6003.
48. Price AL, Eskin E, Pevzner PA: Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. *Genome Res.* 2004, 14:2245-52.
49. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: The human genome browser at UCSC. *Genome Res.* 2002, 12:996{1006.
50. Karolchik D, Kuhn RM, Baertsch R, Barber GP, Clawson H, Diekhans M, Giardine B, Harte RA, Hinrichs AS, Hsu F, Miller W, Pedersen JS, Pohl A, Raney BJ, Rhead B, Rosenbloom KR, Smith KE, Stanke M, Thakapallayil A, Trumbower H, Wang T, Zweig AS, Haussler D, Kent WJ: The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res.* 2008, 36:D773-9.
51. Thornburg BG, Gotea V, Makalowski W: Transposable elements as a significant source of transcription regulating signals. *Gene* 2006, 365:104-10.
52. Landry JR, Mager DL: Functional analysis of the endogenous retroviral promoter of the human endothelin B receptor gene. *J Virol.* 2003, 77:7459-66.
53. RepeatMasker [<http://www.repeatmasker.org>].
54. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* 2005, 110:462-467.
55. Gotea V, Makalowski W: Do transposable elements really contribute to proteomes? *Trends Genet* 2006, 22:260-7.

## Anexo 6: *In silico* annotation of non-coding RNA candidates

Alexandre Rossi Paschoal<sup>1</sup>, Francis Morais Franco Nunes<sup>2</sup>, Alexandre dos Santos Cristino<sup>3</sup>, Zilá Luz Paulino Simões<sup>2</sup>, Alan Mitchell Durham<sup>1</sup>

<sup>1</sup> Bioinformatic Program, Institute of Mathematic and Statistic, University of Sao Paulo - USP, Matão Street, 1010, SP, Brazil

<sup>2</sup> Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto e Faculdade de Medicina de Ribeirão Preto, São Paulo, SP, Brazil

<sup>3</sup> The Queensland Brain Institute, The University of Queensland, Brisbane, Australia

### Abstract

In a previous report - “*Prediction non-coding RNA in comparative analyze approach in Apis mellifera and Nasonia vitripennis. X-meeting, 2008*” - we used genomic comparison to obtain a dataset with more than 1500 putative ncRNA candidates from a prediction of *Drosophila melanogaster* against *Apis mellifera* and *Nasonia vitripennis*. Here, we present a pipeline designed to annotate these candidates. This pipeline provides a set of 5 annotation modules that can be used independently: (i) RNA positive similarity; (ii) RNA structural search; (iii) RNA type characterization; (iv) False positive candidate identification; and (v) Expression evidence. Different from other approaches, we run all steps in all candidates, presenting a single integrated report that, we believe provides more solid evidence about the evidence of ncRNA candidates. In the first module (i) we use BLAST to perform a similarity search against a large collection of more than 8 million ncRNA sequences downloaded from 25 different RNA-specific databases and from the ncRNA genes dataset of the UCSC Genome Browser. In the second module (ii), we use Rfam and RNAfold to gather structural evidence. Rfam contains probabilistic characterizations of more than 1300 ncRNA families (covariance models of each family) and is searched using the script *cmsearch* from the INFERNAL package. RNAfold is used to obtain a candidate fold of the sequence and to estimate the free energy value of this fold. The third module (iii) uses 10 different softwares (Aragorn, BRUCE, tRNAscan-SE, SnoReport, snoscan, HHMMiR, TripletSVM, miR-abela, miRFinder and Mipred) that are specific for different ncRNA types (snoRNA, miRNA, tRNA and tmRNA). The fourth module uses BLAST to perform a similarity search against the Swissprot database and the CPC (Coding Potential Calculator) tool to find potential false positives (candidates are not eliminated to allow for ncRNA candidates that can be expressed in the complementary DNA strand). Finally, the last module (v) uses similarity search against a database of expressed sequences (e.g. EST – Expressed Sequence Tag). The pipeline produces either CSV or HTML formatted output for manual curation. The pipeline was used to process ncRNA candidates obtained from comparative genomics from two organisms: *Apis mellifera* and *Nasonia vitripennis*. For *Apis mellifera* we identified 12% of putative false positive, 51% sequences without any evidence and 37% of confirmed ncRNAs (which 28% miRNA, 5% tRNA, 2% rRNA, ~1% snoRNA/sRNA and ~1% other ncRNA). For *Nasonia vitripennis* we identified 11% of putative false positives, 58%, sequences without any evidence and ~31% of confirmed ncRNAs (which 22% miRNA, 4% tRNA, ~3% rRNA, ~1% snoRNA/sRNA and ~1% other ncRNA). With the exception of the expressed sequences database, the same approach can be applied to other genomes. It is important to note that the pipeline reported here can be extended to include the comparative genomics step described in the previous report or used for candidates obtained from other sources.

**Társio-das-Filipinas (*Tarsius syrichta*)**

ENSP00000366755/1-1210 PDPPVCPLLSPSFLPCPFLGATA--SSAISPSMSSAMSPTL-DPPPSVTSQRREEEKWEA  
 ENSTSY00000007096/1-1127 XXX--XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX--XXXXXXXXXXXXXXXXXXXXXXXXXXXX

ENSP00000366755/1-1210 QSEGHFPFITSRSSSPLQLNLLQEEMPRPSESPDQMRNRTCPOTEYQCVTGNGG--SE-SSP  
 ENSTSY00000007096/1-1127 XX-----XX-XX--

ENSP00000366755/1-1210 ATTGALSTGSPP-REN--PSHPTASALSTGSPPMKNPS---HPTASALSTGSPPMKNPSH  
 ENSTSY00000007096/1-1127 -----XQC-VT-----GHRGNKK---SC-----  
 : : . : .

ENSP00000366755/1-1210 PTASTLSMGLPPSRTPSHPTATVLSTGSPSPSESPRSTGSAASGSSDSSIYLTSSVYSSKI  
 ENSTSY00000007096/1-1127 -----PTASELSMVLHKEPSGSGSAASGSSDSSIYLANSDYSSEI  
 \*\*\*: \*\* .\*\*\*\* \*\*\*\*\*:.\* \*\*:\*

ENSP00000366755/1-1210 SONGQOS-QDV--QKK-----ETFPNVAEPIWRMIRQTPERILMTYQVPERVKEVVLK  
 ENSTSY00000007096/1-1127 CQNGQOSR-HVQN--KETFPNL-----AESIWRMIQQTPECILMTYQVPERIKEVVLK  
 .\*\*\*\*\* . \* \* \*\*\*.\*\*\*\*\*:\*\*\*\* \*\*\*\*\*:\*\*\*\*\*

ENSP00000366755/1-1210 EDLEKLESMRQQQPQFESHGQKEELAKVYNWISQTVTQEIDIQACVTCENEDSADGAATS  
 ENSTSY00000007096/1-1127 EDLRKLESMRHQQPQFESHGQMEELANVYTWIQRQTIPQEIDTQGCITCKSKDSVDPATEF  
 \*\*\*.\*\*\*\*\*:\*\*\*\*\* \*\*\*:\*\*.\* \*\*\*:\*\*.\* \*\*\*:\*\*.\* \*\*\*:\*\*.\* \*\*\*:\*\*.\*

**Tatu-de-nove-bandas (*Dasyus novemcinctus*)**

ENSP00000366755/1-1210 ISPSMSSAMSPTLDPPPSVTSQRREEEKWEAQSEGHFPFITSRSSSPLQLNLLQEEMPRPS  
 ENSDNOP00000007727/1-1079 MPPSV-SSETPNLEPPSSV-SPQGVEEKWEAQSEEHPCTESSCSSLQLNLLQEEMPRSS  
 :.\*\*: \* :.\*:\*.\*\*.\* \* : \*\*\*\*\* \*\* \* .\*\*\*\*\*.\*

ENSP00000366755/1-1210 ESPDQMRNRTCPOTEYQCVTGNGG--SE-SSPATTGALSTGSPPRENPSHPTASALSTGS  
 ENSDNOP00000007727/1-1079 ESSDQLKRDICPETEYQYITGN--GGN-KG-----  
 \*\*.\*\*\*:.\*: \*\*:\* \*\*:\* . .

ENSP00000366755/1-1210 PPMKNPSHPTASALSTGSPPMKNPSHPTASTLSMGLPPSRTPSHPTATVLSTGSPSPSESP  
 ENSDNOP00000007727/1-1079 -----NH-----STASELS-LLLHKDSP  
 .\*\*:\* \*\* .:\*\*

ENSP00000366755/1-1210 SRTGSAASGSSDSSIYLTSSVYSSKISQNGQOSQDVQKKEETFPNVAEPIWRMIRQTPER  
 ENSDNOP00000007727/1-1079 FRTGSAASGSSDSSIYSTSSDSSSEISQNEQKSQDVQEKEPFPNLAESIWRMVEQTPER  
 \*\*\*\*\* \*\* \* \*\*:\* \*\*:\* \*\*:\* \*\*:\* \*\*:\* \*\*:\* \*\*:\* \*\*:\* \*\*:\* \*\*:\*

**Gálago (*Otolemur garnettii*)**

ENSP00000366755/1-1210 SFLPCPFLGATASSAISPSMSSAMSPTLDPPPSVTSQRREEEKWEAQSEGHFPFITSRSSS  
 ENSOGAP00000008770/1-1163 SFFPYPFLGATDSSEIPPPT-AAVAPYPEPSPSVTIQTVGVEEQWETGGEGHPFISRRSS  
 \*\*.\* \*\*\*\*\* \* \* . \* :\*:.\* :\*.\*\*\*\*\* \* \*\*:\*: .\*\*\*\*\*:\*\*\*\*\*

ENSP00000366755/1-1210 PLQLNLLQEEMPRPSESPDQMRNRTCPOTEYQCVTGNGG--SE-SSPATTGALSTGSP-  
 ENSOGAP00000008770/1-1163 PLQLNLLQEEMPKSSEFSDHMKRDICPKPEYXXXXX--XXX-XXXXXXXXXXXX--X  
 \*\*\*\*\*:.\* \* .\*:.\*: \*\*:\* \*\*:

ENSP00000366755/1-1210 REN--PSHPTASALSTGSPPMKNPSHPTASALSTGSPPMKNPSHPTASTLSMGLPP-----  
 ENSOGAP00000008770/1-1163 XX-XX-----XXXX-----XXXXXXXXXXXXXXXXXXXXXXXXXXXXX-----XXXXX





**Cão (*Canis familiaris*)**

ENSP00000366755/1-1210 PSFLPCPFLGATASSAISPSMSSAMSPTLDPPPSVTSQRREEEKWEAQSEGHFFITSRSS  
 ENSCAFP00000029103/1-1203 PSFCPYAFLGAAGSSGTPPFV-SAVAPHLEQLSSVLSQRQAEGRWEMPHGEHHCINSRSS  
 \*\*\* \* .\*\*\*\*:.\*. \* : \*\*:\* \* : .\*\* \*\*\*: \* :\*\* \* \* .\*\*\*\*

ENSP00000366755/1-1210 SPLQLNLLQEEMPRPSESPDQMRNTCPQTEYQCVT-G-NNG--SE-SSPATTGALSTGS  
 ENSCAFP00000029103/1-1203 SPLQLNLLQEDMLRSCSSDQVRGDIYQKLNMEGVLG-R---SGS-KKNPFTASELSMAL  
 \*\*\*\*\*:\* \*..\*.\*.\*:\* : : : \* \* ..\* \*.. \*\* .

ENSP00000366755/1-1210 PPREN-PSHPTASALSTGSPPMKNPSHPTASALSTGSPPMKNPSHPTASTL----SMGLP  
 ENSCAFP00000029103/1-1203 LPEE-SPSGAG-----STASGSGSKPFGFGTGV--EPPRAHMHHT--SQ  
 \*.\* \*\* . :.:.\*. .. .: :

ENSP00000366755/1-1210 PSRTPSHPTATV-----LSTGSPSES-PS---RTG--SAA-S-GSSDSSIYLTSSVYS  
 ENSCAFP00000029103/1-1203 ITQHLDDPFADDEPMCLIIYDFSPYCERG-KKKS---QV--LQ-R-SSDSSIYLASSDYS  
 :. :.\* \* : \*\* .\* . \*\*\*\*\*:\*\* \*\*

ENSP00000366755/1-1210 SKISQNGQQSQDVQKKEITFPNVAEPIWRMIRQTPERILMTYQVPERVKEVVLKEDLEKL  
 ENSCAFP00000029103/1-1203 SEITSNQQQFQGVQKKEITFPLAEESMWRMIKQTPECILMTYQVPERVTEAVLREDLEKL  
 \*.:.\*.\*\*\*\* \*.\* \*\*\*. :.\*. :.\*.\*.\*.\*.\*\*\*\*\*.\*. :.\*.\*\*\*\*\*

**Golfinho-comum (*Tursiops truncatus*)**

ENSP00000366755/1-1210 LPCPFLGATASSAISPSMSSAMSPTLDPPPSVTSQRREEEKW-EAQSEGHFFITSRSSSP  
 ENSTTRP00000009871/1-1106 SPYPFLGAAGSSEIPASV-SAVPPNPSPSSAISQRSGEGKWG-TQNEGHPLINKSSSP  
 \* \*\*\*\*\*:.\* \*..\*:\* \*\*:\*.\*. \*\*.\*. \*\*\* \* \*\* :\*.\*\*\*\*\*:\*. :\*\*\*\*

ENSP00000366755/1-1210 LQLNLLQEEMPRP-S--ESPDQMRNTCPQTEYQCVTGNG--SE-SSPATTGALSTGSP  
 ENSTTRP00000009871/1-1106 LQLDLLQENRPRSC-AS-----PE--AEY---HFTTSELSTVSLH  
 \*\*\*:\*\*\*\*: \*\*. : \*\*.\* \*\* \*

ENSP00000366755/1-1210 REN-PSHPTASALSTGSPPMKNPSHPTASALSTGSPPMKNPSHPTASTLSMGLPPSRTPS  
 ENSTTRP00000009871/1-1106 EA-SPPGAG-----SAAAG-----  
 . \* . . :.\*.

ENSP00000366755/1-1210 HPTATVLTSTGSPSES-PSRTGSAASGSSDSSIYLTSSVYSKISQNGQQSQDVQKKEITFP  
 ENSTTRP00000009871/1-1106 -----SSGSSINFATSDYSSEIFQNGQQSQDVQKKEITFP  
 \*\*.\*\*\* :.:\* \*\*\*: \* \*\*\*\*\*:\*\*\*\*\*

ENSP00000366755/1-1210 NVAEPIWRMIRQTPERILMTYQVPERVKEVVLKEDLEKLESMRQQQPQFQSHGQKEELAK  
 ENSTTRP00000009871/1-1106 NLAESIWRMIEQTPECILMTYQVPERVKEIVLKDLENLESMRRRQPQFQSHGQREELAN  
 \*:\*\*\*.\*\*\*\*\*.\* \*\*\*.\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*:

**Elefante africano (*Loxodonta africana*)**

ENSP00000366755/1-1210 AISPSMSSAMSPTLDPPPSVTSQRREEEKWEAQSEGHFFITSRSSSPLQLNLLQEEMPRP  
 ENSLAFP00000000486/1-926 EIPPSS-SPEAPGLEPPSSVTNQRRAERWETHGEEHPFISSRSSSPLQLDLLQEEMPGS  
 \*.\* \* .:\* \*.\*.\*.\*.\* \*.\*.\*.\*.\*.\*\*\*\*\*:\*\*\*\*\*

ENSP00000366755/1-1210 SESPDQMRNTCPQTEYQCVTGNG--SE-SSPATTGALSTGSPPREN-PSHPTASALST  
 ENSLAFP00000000486/1-926 SQSPDQMKRDVCSEAECQHVLGN--SGN-NNCCPASELSTVLLRQN-S-----  
 \*:\*\*\*\*\*:\*. :.\*.\* \*\* . . . . . \*\* :

ENSP00000366755/1-1210 GSPPMKNPSHPTASALSTGSPPMKNPSHPTASTLSMGLPPSRTPSHPTATVLTSTGSPSE  
 ENSLAFP00000000486/1-926 -----





**Morcego (*Myotis lucifugus*)**

ENSP00000366755/1-1210 REEEKWEAQSEGHFPFITSRSSSPLQLNLLQEEMPRPSESPDQMRNTCPQTEYQCVTGNN  
 ENSMLUP00000014533/1-922 GAGQSWE-QSPGCPGPSSRGSSPLQLALLQEDKRPSCESEPEQVR-----  
 :.\* \*\* \* \* \* :.\*.\*\*\*\*\* \*\*\*: .\*.\*\*\*:\*:\*

ENSP00000366755/1-1210 GSESSPATTTGALSTGSPPRENPSHPTASALSTGSPPMKNPSHPTASALSTGSPPMKNPSH  
 ENSMLUP00000014533/1-922 -----

ENSP00000366755/1-1210 PTASTLSMGLPPSRTPSHPTATVLTGSPSESPSRTGSAASGSSDSSIYLTSSVYSSKI  
 ENSMLUP00000014533/1-922 -----

ENSP00000366755/1-1210 SQNGQSQDVQKKETFPNVAEPIWRMIRQTPERILMTYQVPERVKEVVLKEDLEKLESM  
 ENSMLUP00000014533/1-922 -----

**Lemur (*Microcebus murinus*)**

ENSP00000366755/1-1210 CPFLGATASSAISPSMSSAMSPTLDPPPSVTSQRREEEKWEAQSEGHFPFITSRSSSPLQL  
 ENSMICP00000007691/1-1191 HPFLGAAGSFEI-----SAMAPNLGPPPAVTSPRGAEWETHSGEHLFINSRSSSPLQL  
 \*\*\*\*\*:\* \* \*\*\*:.\*.\*\*\*:\* \*\* \* \*\*:\*:\*:\* \* \*\*.\*\*\*\*\*\*

ENSP00000366755/1-1210 NLLQEEMPRPSESPDQMRNTCPQTEYQCVTGNN--SE-SSPATTTGALSTGSP-REN-  
 ENSMICP00000007691/1-1191 DLLQEEMPRSEFSDHMKRDVYLEAEHXXXXXX--XXX-XXXXXXXXXXXXXXXX-XXX-X  
 :\*\*\*\*\*.\* \*.\*:\*:\*.. :\*:\*

ENSP00000366755/1-1210 PSHPTASALSTGSPPMKNPSHPTASALSTGSPPMKNPSHPTASTLSMGLPP-----  
 ENSMICP00000007691/1-1191 X-----XXXX-----XXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXX-----XXXXXXXXXX

ENSP00000366755/1-1210 -----SRT-P---SHPTATVLTGSPSESPSRTGSAASGSSDSSIYLTSSVYSSKIS  
 ENSMICP00000007691/1-1191 XX-----XXXXXSSDSSSDYSSEIS  
 : \*\* \*\*\*:\*\*

ENSP00000366755/1-1210 QNGQSQDVQKKETFPNVAEPIWRMIRQTPERILMTYQVPERVKEVVLKEDLEKLESMR  
 ENSMICP00000007691/1-1191 QNGQSQDAQKKETFPNVAEPIWRMIQQTPEIILMTYQVPEXXXXXXXXXXXXXXXXXXXXXXXXXX  
 \*\*\*\*\*.\*\*\*\*\*\*.\*:\*.\*\*\*\*\*\*:\* \*\* \* \*\*\*\*\*

**Camundongo (*Mus musculus*)**

ENSP00000366755/1-1210 EEKWEAQSEGHFPFITSRSSSPLQLNLLQEEMPRPSESPDQMRNTCPQTEYQCVTGNN--  
 ENSMUSP00000099493/1-1113 EENWEAHSEELPFISSRSSSPLQLNLLQEEMPAPESEADAVRRGAGPDAKHHCVTGP--S  
 \*\*:\*:\*:\* \*\*\*.\*\*\*\*\* \*\*\*\*\*.\* :\*:\* :\*:::\*:\*\*\*

ENSP00000366755/1-1210 -SE-SSPATTTGALSTGSPPREN-PSHPTASALSTGSPPMKNPSHPTASALSTGSPPMKNP  
 ENSMUSP00000099493/1-1113 GS-RSRHCTSGELATATAQQE-S-----  
 \* \* .:\* \* \*:\*:\* :\*

ENSP00000366755/1-1210 SHPTASTLSMGLPPSRTPSHPTATVLTGSPSESPSRTGSAASGSSDSSIYLTSSVYSS  
 ENSMUSP00000099493/1-1113 -----AAAGSSASSIYFSSTDYAS  
 :\*\*\*\*\* \*\*\*::\*:\* \*

ENSP00000366755/1-1210 KISQNGQSQDVQKKETFPNVAEPIWRMIRQTPERILMTYQVPERVKEVVLKEDLEKLE  
 ENSMUSP00000099493/1-1113 EVSENQRQPQDRQRDEALPGAAEESIWRMIERTPECVLTMTYQVPERGREEVVKQDLEKLE  
 :\*: \* \*:\* \*:\*:\*:\*.\*.\*\*\*\*\*.\*:\* :\*\*\*\*\* :\* \*\*\*:\*\*\*\*\*:

**Cuíca (*Monodelphis domestica*)**

ENSP00000366755/1-1210 KWEAQS-EGHPFIT-SRSSSPLQLNLLQEEMP-R-P-S--ES--PDQMRNTCPQTEYQC  
 ENSMODP00000031107/1-1209 EWALQSEQEHPLFHSRSSSPLQLNLLQEEMPG-A-P-EP--LQ--YTGSEADPGAQHQS  
 :\* \*\* : \*\*::: \*\*\*\*\* \*\*::\* \*\*::\*.

ENSP00000366755/1-1210 VTG-NNG--SE-SSPATTGALSTGSPPREN-PSHPTASALSTGSPPMKNPSHPTASALST  
 ENSMODP00000031107/1-1209 KAGD---NKK-SH-NSVAKELFTLLLQKD-SPAGAGSVASGSGSVKS-----  
 :\* . :: \* \* :: \* : . : \* .:\*\*

ENSP00000366755/1-1210 GSPPMKNPSHPTASTLSMGLPP-SRT-PSHPTATVLSTGSPPSES PSRTGSAA-----  
 ENSMODP00000031107/1-1209 -----NS--S-----G-----S-DANGSFSGTSDHLKAAVP  
 \* \* :... \*\*:

ENSP00000366755/1-1210 -----SGSSDSSIYLTSSVYSSKISQNGQQSQDV  
 ENSMODP00000031107/1-1209 REGDLPAGHLGHLAKSVLEKCFQIFSSFWGFFIGSSDSSKYFASDDYSSETSQRQKS---  
 \*\*\*\*\* \*::\*\* \*\*\*: \*\*\* \*:

ENSP00000366755/1-1210 QKKETFPNVAEPIWRMIRQTPERILMTYQVPERVKEVVLKEDLEKLESMRQQQPQFQSHG  
 ENSMODP00000031107/1-1209 EEKETPHKSGRDAIWGMIKQTPECVLMYQVPERVKEDVLKEDLEKLESMRQHQPHTDG  
 :\*\*\* : ...\* \*.\*\*\*\* :\*\*\*\*\* \*\*\*\*\*:\*\*\*:\*\*\*:

**Rato (*Rattus norvegicus*)**

ENSP00000366755/1-1210 RREEEKWEAQS-EGHPFITSRSSSPLQLNLLQEEMPRPSESPDQMRNTCPQTEYQCVTG  
 ENSRNOP00000024932/1-1115 RRVEENWETHSEEEHPFISSRSSSPLQLNLLQEEMPAPSEYADAVRRGACPDALQCVTG  
 \*\* \*:\*:\*:\* \* \* \*\*\*:\*\*\*\*\* \*\* \* . \* : \*\* : \*\* : \*\* : \*\* : \*\*

ENSP00000366755/1-1210 NNG--SE-SSPATTGALSTGSPPREN-PSHPTASALSTGSPPMKNPSHPTASALSTGSPP  
 ENSRNOP00000024932/1-1115 N--SGS-RSPPCATGELATASVQQE-S-----  
 \* \* \* . \* : \* \* \* . \* : \*

ENSP00000366755/1-1210 MKNPSHPTASTLSMGLPPSRTPSHPTATVLSTGSPPSES PSRTGSAASGSSDSSIYLTSS  
 ENSRNOP00000024932/1-1115 -----SSAAASGSSASSVHGSGS  
 :.:\*\*\*\*\* \*\*::: .\*

ENSP00000366755/1-1210 VYSSKISQNGQQSQDVQKKETFPNVAEPIWRMIRQTPERILMTYQVPERVKEVVLKEDL  
 ENSRNOP00000024932/1-1115 DYTSEVSENGQRSQDRQDRAFSGAAEESIWRMIERTPECVLMYQVPERGRDVTLREDL  
 \*.\*:\*:\*.\*\*\*.\*\*\* \*:.:.\*...\*\*.\*\*\*.\*:\*\*\* :\*\*\*\*\* :.:\*\*.\*\*\*

**Bicho-preguiça (*Choloepus hoffmanni*)**

ENSP00000366755/1-1210 ASSAISPSMSSAMPTLDPPPSVTSQRREEEKWEAQS-EGHPFITSRSSSPLQLNLLQEEM  
 ENSCHOP00000006567/1-1111 DSSEMPPSV-SSVTPDLEPLPSGIHQQRVEEKWESQSEAHPFINSRSSSPLHLNLLQ-EI  
 \*\* :.\*:\*: \*:::\* \*.\* \*\* \*:\* \*\*\*\*\*:\*\*\*.\*\*\*.\*\*\*\*\*:\*\*\*\*\* \*:

ENSP00000366755/1-1210 PRPSESPDQMRNTCPQTEYQCVTGNNNG--SE-S-SPATTGALSTGSPP-REN-PSHPTA  
 ENSCHOP00000006567/1-1111 PRLSESV-QMKRXXXXXXXXXXXXXXXXXX--XXX-XXXXXXXXXXXX-XXXXXXXXXX-XX-XXXX  
 \*\* \*\* \*:\*:

ENSP00000366755/1-1210 SALSTGSPPMKNPSHPTASALSTGSPPMKNPSHPTASTLSMGLPP-----SRT-  
 ENSCHOP00000006567/1-1111 XXXXX-----XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XXXXXXXXXXXXXXXXXX

ENSP00000366755/1-1210 PSHPTATVLSTGSPPSES PSRTGSAASGSSDSSIYLTSSVYSSKISQNGQQSQDVQKKET  
 ENSCHOP00000006567/1-1111 -XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXSGDGSIHFASSD-YSEISQNLKSDVQEEET  
 \*.\*.\*:\*:\* \*\* \*:\*\*\*\*\*\* :\*\*\*\*\*:\*\*\*

ENSP00000366755/1-1210 FPNVAEPEIWRMIRQTPERILMTYQVPERVKEVVLKEDLEKLESMRQQQPQFSLGQKEEL  
 ENSCHOP00000006567/1-1111 FPNLAVESIWRMIEHTPEGILMTYQVPERVKEVVLKEDLEKLDMSRQQQPRFSLGQKEEL  
 \*\*\*:\* \*.\*\*\*\*\*.:\*\*\* \*\*\*\*\*:\*\*\*\*\*:\*\*\* \*\*\*\*\*

**Esquilos (*Spermophilus tridecemlineatus*)**

ENSP00000366755/1-1210 PPVCPLLSPSFLPCPFLGATASSAISPSMSSAMSPTLDPPPSVTSQRREEEKWEAQSEGH  
 ENSSTOP00000012561/1-1099 PRICPLWSPSPFPYFLGATDSSEILPSV-SAVAPNLEPPSVNS-GRVEENWEMQKEEH  
 \* :\*\*\* \*\*\*\* \* \*\*\*\*\* \*\* \* \*\*: \*\*:\*.\*.\*\*\*\*\*.\* \* \*\*:\*\*\* \*. \* \*

ENSP00000366755/1-1210 PFITSRSSSPLQLNLLQEEMPRPSES PDQMRNTCPQTEYQCVTGNNG--SE-SSPATTG  
 ENSSTOP00000012561/1-1099 LFISSRSSSPLQLDLLHEDMPASPEPPDAVRRDAGPEATDHCVSGG--GGSS-RSSS----  
 \*\*:\*\*\*\*\*:\*\*\*:\*. \* ..\*. \* :\*\*\*: \*:: \*:: :\*\*\*:\*. \* \*\*.

ENSP00000366755/1-1210 ALSTGSPPREN-PSHPTASALSTGSPMKNPSHPTASALSTGSPMKNPSHPTASTLSMG  
 ENSSTOP00000012561/1-1099 AAPLPSHPTR-S-----  
 \* . \* \* .

ENSP00000366755/1-1210 LPPSRTPSHPTATVTLSTGSPSESPSRTGSAAGSSDSSIYLTSSVYSSKISQNGQQSQD  
 ENSSTOP00000012561/1-1099 -----PSGAGSAAGSSSSCIFFTTSTDYSESSENGQQSQD  
 \*\* :\*\*\*\*\*.\*.\*:\*\*\*: \*\*\*: \* :\*\*\*\*\*

ENSP00000366755/1-1210 VQKKETFPNVAEPEIWRMIRQTPERILMTYQVPERVKEVVLKEDLEKLESMRQQQPQFSLGQKEEL  
 ENSSTOP00000012561/1-1099 TQKTSAFHSLAEDPIWKMIAQTPECVLMYQVPERVKEVVLKEDLEKLES LRQQQPQFSLGQKEEL  
 .\*. \* .:\*\*\*:\*\*\*:\*\*\* \*\*\*\*\* :\*\*\*\*\*:\*\*\*\*\*:\*\*\*\*\*.

**Tupaia (*Tupaia belangeri*)**

ENSP00000366755/1-1210 PSFLPCPFLGATASSAISPSMSSAMSPTLDPPPSVTSQRREEEKWEAQSEGH PFITSRSS  
 ENSTBEP00000003120/1-1131 PSFSPYSFLGATGSPDMPPSV-STVTPDLEPPFSGINQRRVQEKWETQSEDHPFINSRSS  
 \*\*\* \* .\*\*\*\*\*.\*. :.\*\*\*: \*::.\* \*.\* \* .\*\*\* :\*\*\*\*\*:\*\*\*.\*\*\*.\*\*\*.\*\*\*

ENSP00000366755/1-1210 SPLQLNLLQEEMPRPSES PDQMRNTCPQTEYQCVTGNNGSESSPATTGALSTGSPPREN  
 ENSTBEP00000003120/1-1131 SPLQLNLLQEEMPRSES SDWMKDTCL--EYQHVTGN-----  
 \*\*\*\*\*.\*\*\*.\* \*::.\*\* \*\*\* \*\*\*\*

ENSP00000366755/1-1210 PSHPTASALSTGSPMKNPSHPTASALSTGSPMKNPSHPTASTLSMGLPPSRTPSHPTA  
 ENSTBEP00000003120/1-1131 -----RGNKIHHPAAGKLSAV-----  
 . :.\*. .

ENSP00000366755/1-1210 TVLSTGSPSESPSRTGSAAGSSDSSIYLTSSVYSSKISQNGQQSQDVQKKETFPNVAE  
 ENSTBEP00000003120/1-1131 -----LHKEPSGTGSTASXX  
 .\*\*\* \*\*.\*

ENSP00000366755/1-1210 EPIWRMIRQTPERILMTYQVPERVKEVVLKEDLEKLESMRQQQPQFSLGQKEELAKVYNW  
 ENSTBEP00000003120/1-1131 XXXXXXXXXXXXXXXXXXXXXXXKVVLDLQKLERMRQQQPKFSLGQKEELAKVRSW  
 \*\*\*\*\*:\*\*\*:\*\*\* \*\*\*\*\*:\*\*\* \*\*\*\*\* .\*

**Canguru (*Macropus eugenii*)**

ENSP00000366755/1-1210 PPSVTSQRREEEKWEAQSEGH PFITSRSSSPLQLNLLQEEMPRPSES PDQMRNTCPQ  
 ENSMEUP00000009308/1-1062 PSPASSQRSAAEEWVLQSDQEHPLFHSRSSSPLQLNLLQEEMPGAPEPLPPMRQEAD--  
 \*...:\*\*\* \*\*:\* \* \* : \*::: \*\*\*\*\* \*\*\*\*\* ..\* \*\*:::

ENSP00000366755/1-1210 TEYQCVTGNNG-SE-SSPATTGALSTGSPPREN-PSHPTASALSTGSPMKNPSHPTASA  
 ENSMEUP00000009308/1-1062 -----TG-VQ-NSMAKGLFTLLHLD-SPTGADSAAG-----  
 . . : :.\* \* :: \* : . :.\* .

```

ENSP00000366755/1-1210  LSTGSPPMKNPSHPTASTLSMGLPPSRTPSHPTATVLSHGSPSESPSRGTGSAASGSSDS
ENSMEUP0000009308/1-1062 -----SSDS
                                                                    ****

ENSP00000366755/1-1210  SIYLTSSVYSSKISQNGQQSQDVQKKEETFPNVAEPIWRMIRQTPERILMTYQVPERVKE
ENSMEUP0000009308/1-1062 SRYFASSDSSSEISQNGQKS---EKKEAPHKSGRDAIWGMIEQTPKCILMTYQVPERVKE
* *::** **:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
                                                                    ****

ENSP00000366755/1-1210  VVLKEDLEKLESMRQQQPQFESHGQKEELAKVYNWIQSQTVTQEIDIQACVTCEN--ED--
ENSMEUP0000009308/1-1062 DVLKEDLEKLESMKQHQPQFTEGQKEELAEVHPWIQSHTVPREIDTHGCVTCDSKR--SL
*****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

```

**Anexo 7.** Parte da seqüência da região do VNTR da proteína PER3, das espécies cujos genomas estão disponíveis na plataforma ENSEMBL, alinhada com da seqüência referente da proteína de humano separadamente (par a par). Esses alinhamentos estão disponíveis na plataforma ENSEMBL. A seqüência do VNTR do humano aparece destacada em cinza.