Universidade Federal do Rio Grande do Sul
Instituto de Informática Aplicada
Biologia Computacional - Lista 10
Prof. Dr. Márcio Dorn

Nome: Vicente Merlo
Matrícula: 244950

O algoritmo simplesmente itera sobre todas as sequências de tamanho N, verificando qual delas tem uma distância de no máximo M com relação a outras sequências também de tamanho N que não sejam a original.
A função distância é a verificadora de mutações. Ela retorna quantos caracteres diferentes existem entra uma sequência e outra.
Os resultados foram, sendo representados como sequência => número de motifs encontrados:

Sequência 1:

N=8, M=2

{'gaatctat': 1, 'tggctatc': 1, 'cgctatct': 1, 'tctatgcg': 1, 'ctgtgcga': 1, 'atctatgc': 1, 'tctgtgcg': 1, 'tgcgtttc': 1, 'gacgctat': 1, 'ctatgcgt': 1, 'ggctatcc': 1, 'gctatctg': 1, 'tgcgaatc': 1, 'gctatcca': 1, 'acgctatc': 1, 'ctctgtgc': 1}

N=5 M=3

{'ctggc': 23, 'gatag': 18, 'cgtac': 25, 'ctgtg': 19, 'tatct': 26, 'gtttc': 27, 'tccac': 20, 'tatcc': 25, 'ctctg': 22, 'tgtgc': 23, 'ggcta': 24, 'ctatg': 25, 'ctgat': 26, 'ctatc': 50, 'tacat': 25, 'acgta': 18, 'tgata': 22, 'tctgt': 24, 'gtaca': 23, 'aatct': 24, 'cctga': 22, 'atctg': 22, 'cgaat': 21, 'ttcca': 23, 'acgct': 22, 'taggt': 19, 'atcta': 24, 'gtcct': 26, 'tatgc': 26, 'cgttt': 28, 'caacc': 23, 'aacca': 23, 'gcgtt': 21, 'tgcgt': 22, 'atagg': 17, 'ataga': 18, 'gacgc': 20, 'ggtcc': 23, 'catag': 24, 'agacg': 19, 'gaatc': 19, 'ccacg': 19, 'tggct': 26, 'tagac': 23, 'tctgg': 21, 'atcca': 21, 'gtgcg': 22, 'tgcga': 22, 'tctat': 30, 'acata': 25, 'cctct': 24, 'tccaa': 23, 'cgcta': 24, 'tcctc': 30, 'atgcg': 23, 'tttcc': 26, 'ccaac': 22, 'cacgt': 21, 'gctat': 48, 'aggtc': 24, 'gcgaa': 17}

N=3 M=1

{'tat': 39, 'tgt': 14, 'tct': 42, 'ttt': 9, 'tgc': 22, 'tag': 12, 'tgg': 7, 'tac': 14, 'ttc': 12, 'tcc': 30, 'gta': 9, 'gtc': 10, 'gcg': 12, 'gtg': 9, 'gtt': 8, 'gct': 22, 'ggt': 8, 'cga': 12, 'cgc': 8, 'gat': 11, 'ggc': 6, 'gac': 7, 'cgt': 16, 'gaa': 4, 'aat': 6, 'acc': 10, 'atg': 12, 'aca': 9, 'acg': 14, 'atc': 30, 'aac': 8, 'ata': 22, 'tga': 6, 'cct': 24, 'aga': 6, 'cat': 11, 'agg': 5, 'ctg': 24, 'cta': 42, 'ctc': 13, 'cac': 7, 'caa': 10, 'cca': 30}

Sequência 2:

N=8 M=2

{'accagaag': 1, 'acgtacac': 1, 'aaccagaa': 1, 'ctgaaaca': 1, 'acgtacgt': 1, 'aacctgaa': 1, 'ccggcaac': 1, 'cagaacca': 1, 'tacgtaca': 1, 'accggcaa': 1, 'gaaccaga': 2, 'caacctga': 1, 'gatacgta': 1, 'tacgtacg': 1, 'atacgtac': 1, 'gtacgtac': 1, 'cctgaaac': 1, 'cgtacgta': 1, 'gaaacaaa': 1, 'acctgaaa': 2}

N=5 M=3

{'aacca': 30, 'ccggc': 21, 'ttgat': 16, 'caacc': 29, 'gatac': 21, 'actgg': 23, 'ctggt': 18, 'aacgc': 27, 'atttg': 19, 'aagtg': 24, 'tacgt': 34, 'tgaaa': 27, 'tacat': 22, 'acgta': 52, 'tgata': 20, 'gtaca': 48, 'gtgta': 19, 'gtacg': 18, 'cctga': 19, 'tacac': 28, 'acgct': 21, 'ctgaa': 22, 'caaac': 30, 'tactg': 20, 'agtac': 21,

'gtact': 20, 'acctg': 23, 'tggtg': 14, 'aaacg': 29, 'gaaac': 28, 'gaagt': 23, 'accgg': 27, 'aacaa': 30, 'atacg': 24, 'agaag': 27, 'ggtgt': 16, 'agaac': 25, 'aacct': 24, 'cgtac': 46, 'accag': 27, 'cattt': 21, 'acaaa': 32, 'aaaca': 33, 'acacc': 31, 'gctca': 23, 'ccaga': 26, 'cgctc': 19, 'tgtac': 19, 'tttga': 18, 'tcaga': 21, 'cagaa': 58, 'cggca': 25, 'acatt': 24, 'gcaac': 27, 'caccg': 26, 'gaacc': 21, 'ggcaa': 28, 'ctcag': 22}

N=3 M=1

{'tgt': 9, 'ttt': 3, 'tga': 12, 'tgg': 5, 'tac': 45, 'ttg': 6, 'caa': 22, 'tca': 7, 'gca': 13, 'gta': 40, 'gtg': 16, 'gct': 5, 'ctc': 4, 'ggt': 8, 'cgc': 6, 'gat': 6, 'aag': 11, 'cgg': 9, 'ggc': 2, 'aaa': 34, 'cgt': 18, 'gaa': 36, 'acc': 39, 'aca': 51, 'acg': 33, 'aac': 60, 'ata': 12, 'cct': 7, 'aga': 22, 'cat': 9, 'att': 5, 'ctg': 18, 'act': 14, 'cac': 16, 'ccg': 10, 'agt': 18, 'cag': 20, 'cca': 9}

Sequência 3:

N=8 M=2

{}

N=5 M=3

{'gggct': 20, 'tcttc': 24, 'gtggc': 20, 'tccgt': 22, 'agtcc': 23, 'gtgca': 20, 'gatgt': 23, 'caccc': 22, 'atgta': 20, 'ttcgt': 20, 'aattt': 21, 'ctctg': 21, 'ccctc': 21, 'tataa': 19, 'ttctt': 23, 'ataag': 20, 'gccaa': 21, 'tgcac': 21, 'cgaaa': 20, 'cctct': 26, 'gtccg': 23, 'ccaac': 22, 'tgatg': 23, 'cgagg': 19, 'aaatt': 20, 'acgag': 17, 'ctctt': 22, 'gcacc': 21, 'caacg': 21, 'tcttt': 22, 'gacga': 19, 'ccgtg': 23, 'gaggg': 24, 'ggctg': 19, 'tggcc': 21, 'gaaaa': 22, 'aaagt': 20, 'ggctc': 23, 'taaga': 19, 'tcgtg': 17, 'ctgat': 23, 'agacg': 22, 'aaaat': 21, 'cgtgc': 21, 'tggct': 20, 'agggc': 20, 'cgtgg': 20, 'tctgg': 21, 'acgaa': 20, 'ctggc': 21, 'aaaag': 20, 'aagac': 23, 'aagtc': 23, 'tttct': 23, 'aacga': 23, 'gctga': 22, 'gtata': 21, 'accct': 25, 'ggcca': 21, 'ctttc': 23, 'gctct': 20, 'tgtat': 23, 'cttcg': 20}

N=3 M=1

{'tat': 9, 'tgt': 13, 'tct': 33, 'ttt': 22, 'tgc': 10, 'tga': 8, 'tgg': 14, 'ttc': 16, 'tcg': 9, 'tcc': 10, 'gca': 6, 'gta': 6, 'gcc': 11, 'gtc': 12, 'gtg': 16, 'gac': 10, 'gct': 16, 'ctt': 22, 'cga': 14, 'gat': 7, 'ggg': 9, 'ggc': 21, 'gag': 8, 'acg': 16, 'caa': 10, 'cgt': 16, 'gaa': 11, 'acc': 6, 'atg': 11, 'aag': 24, 'aaa': 48, 'aac': 10, 'ata': 8, 'agg': 10, 'cct': 12, 'aga': 10, 'aat': 11, 'att': 8, 'ctg': 18, 'ctc': 20, 'cac': 6, 'ccg': 8, 'agt': 7, 'cca': 7, 'ccc': 9, 'taa': 8}

Sequência 4:

N=8 M=2

{'attacatc': 1, 'acgtacgt': 2, 'tacgtata': 1, 'cttacgta': 1, 'cctattac': 2, 'gtaactat': 1, 'cgatgtaa': 1, 'gctgtaac': 1, 'tgtaacta': 1, 'aactatta': 1, 'catcttac': 1, 'actattac': 1, 'attacctg': 1, 'ttacctgc': 1, 'tacgtacg': 1, 'tgtaagtc': 1, 'tattacat': 2, 'tattacct': 2, 'gatgtaag': 1, 'tcttacgt': 3, 'ctattacc': 1, 'ctattaca': 1, 'ccctatta': 1, 'agctgtaa': 1, 'atgtaagt': 2, 'ttacgtac': 2, 'cgtacgta': 1, 'ctgtaact': 1, 'gtacgtat': 2}

N=5 M=3

{'tcata': 26, 'agtca': 21, 'actat': 29, 'ctgta': 20, 'cctat': 25, 'gatgt': 18, 'cccct': 30, 'cccta': 30, 'caccc': 26, 'atgta': 23, 'tccga': 24, 'tatac': 29, 'ctccg': 26, 'gccac': 24, 'gtacg': 15, 'gcctc': 23, 'cttac': 27, 'tacat': 27, 'acgta': 42, 'tctta': 25, 'tacct': 24, 'cctgc': 26, 'ccgat': 23, 'attac': 50, 'cctcc': 26, 'gtcat': 25, 'gtaac': 20, 'ttacc': 26, 'gtaag': 17, 'agcct': 20, 'acctg': 20, 'cgtat': 24, 'tacgt': 40, 'catct': 30, 'atagc': 21, 'tagct': 21, 'agctg': 16, 'catag': 22, 'taact': 23, 'acatc': 27, 'cgtac': 21, 'ttacg': 19, 'tgcca': 25, 'ttaca': 27, 'cgatg': 17, 'taagt': 20, 'atctt': 27, 'aagtc': 20, 'gtata': 24, 'ctatt': 56, 'tgtaa': 38, 'tatta': 56, 'aacta': 26, 'acccc': 28, 'ctgcc': 25, 'ccacc': 27, 'gctgt': 17}

N=3 M=1

{'tat': 48, 'tgt': 18, 'tct': 12, 'tgc': 10, 'tag': 11, 'taa': 28, 'tac': 70, 'tta': 39, 'tcc': 14, 'tca': 9, 'gta': 44, 'gcc': 16, 'gtc': 8, 'gct': 8, 'ctt': 14, 'cga': 5, 'gat': 6, 'aca': 9, 'act': 13, 'cgt': 22, 'acc': 28, 'atg': 10, 'aag': 5, 'acg': 16, 'atc': 12, 'aac': 12, 'ata': 30, 'cct': 42, 'agc': 14, 'cat': 24, 'att': 16, 'ctg': 14, 'cta': 32, 'ctc': 10, 'cac': 11, 'ccg': 10, 'agt': 9, 'cca': 11, 'ccc': 26}

Sequência 5:

N=8 M=2

{'gcgttttg': 1, 'tcgatcgt': 1, 'tggtcgtc': 1, 'cgtcatgg': 1, 'cgttttgg': 1, 'ttggtcgt': 2, 'cgctcgat': 1, 'tcgtcgta': 1, 'gtcgtcgt': 1, 'cgatcgtt': 1, 'gcgtcatg': 1}

N=5 M=3

{'tcatg': 24, 'cgctc': 20, 'cgcgt': 25, 'taacg': 23, 'ggggt': 28, 'ttttg': 30, 'gatcg': 24, 'tgtta': 27, 'cgatc': 21, 'tggtc': 29, 'ctgtt': 31, 'tatac': 21, 'ttggt': 32, 'acgcg': 23, 'gtagg': 29, 'cgggg': 28, 'gtcgt': 50, 'acgta': 20, 'gtacg': 25, 'cgtta': 23, 'gtcat': 22, 'tacaa': 16, 'taggt': 28, 'gcgtc': 21, 'gcggg': 29, 'cgttt': 23, 'tcgat': 24, 'tggcg': 29, 'tgcgt': 25, 'gcgtt': 25, 'ggtat': 24, 'acgct': 21, 'tcgtc': 27, 'ataca': 17, 'tcgta': 21, 'catgg': 25, 'ggtcg': 29, 'atcgt': 22, 'cgtca': 21, 'cgtcg': 24, 'cgtag': 23, 'gtatg': 27, 'cgtac': 20, 'ttaac': 24, 'ttata': 24, 'tacgc': 22, 'caacg': 18, 'tcgtt': 23, 'acaac': 14, 'gctcg': 26, 'gtttt': 26, 'atggc': 26, 'gttaa': 26, 'aacgt': 18, 'ggcgg': 27, 'tttgg': 34, 'tatgc': 19, 'atgcg': 22, 'gggta': 25, 'gttat': 25, 'aacgc': 17, 'ctcga': 24}

N=3 M=1

{'tat': 20, 'tgt': 16, 'ttt': 20, 'tgc': 8, 'tag': 12, 'tgg': 26, 'tac': 16, 'tcg': 56, 'tta': 20, 'ttg': 14, 'tca': 8, 'gta': 33, 'gtc': 33, 'gcg': 36, 'gtt': 45, 'gct': 10, 'ggt': 51, 'cga': 10, 'cgc': 24, 'gat': 10, 'cgg': 15, 'ggg': 24, 'ggc': 11, 'cgt': 84, 'atg': 18, 'aca': 5, 'acg': 39, 'atc': 9, 'aac': 8, 'ata': 9, 'agg': 10, 'cat': 10, 'ctg': 5, 'ctc': 7, 'caa': 3, 'taa': 9}