Python 3mers formula

Aadil Noon

2020-05-15

```
my_seq1 = "gctcgtgtctgcgaggcttggttccctgaaatgcaaagatgtactttcaatgccttcgacctattggtgcacgtggtgcc"
def count_kmers(read, k):
    """Count kmer occurrences in a given read.

    Parameters
    ----------
    read : string
        A single DNA sequence.
    k : int
        The value of k for which to count kmers.

    Returns
    -------
    counts : dictionary, {'string': int}
        A dictionary of counts keyed by their individual kmers (strings
        of length k).

    Examples
    --------
    >>> count_kmers("GATGAT", 3)
    {'ATG': 1, 'GAT': 2, 'TGA': 1}
    """
    counts = {}
    num_kmers = len(read) - k + 1
    for i in range(num_kmers):
        kmer = read[i:i+k]
        if kmer not in counts:
            counts[kmer] = 0
        counts[kmer] += 1
    return counts
count_kmers(my_seq1,3)
```

```
## {'gct': 2, 'ctc': 1, 'tcg': 2, 'cgt': 2, 'gtg': 4, 'tgt': 2, 'gtc': 1, 'tct': 1, 'ctg': 2, 'tgc': 5,
```

The code above creates a formula to see the unique possible codons in a DNA sequence