1.

a) Same protein may vary in different organisms

b) Yak Genome Sequences can be found here:

<http://www.ncbi.nlm.nih.gov/nuccore/AGSK00000000>

c) Follow this link for statistics: <http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html>

2. Descriptive

3. Descriptive

4. Create a graph where the vertices are letters of the alphabet, with directed edges (x,y) if there is a two character substring xy.   The shortest common superstring is an Eulerian (not Hamiltonian) path problem on this graph, which can be efficiently solved.   One must add edges between vertices which have non-equal in/out degrees until it becomes Eulerian.

5. Build suffix tree for A, search for B as long as there is a match. When you stop, if it is a leaf node, then it is the desired longest suffix of A that exactly matches the prefix of B.

6.

Explain similar to the original palindrome finding problem with cases to consider, but use reverse compliment of the original string, instead of the reverse one.