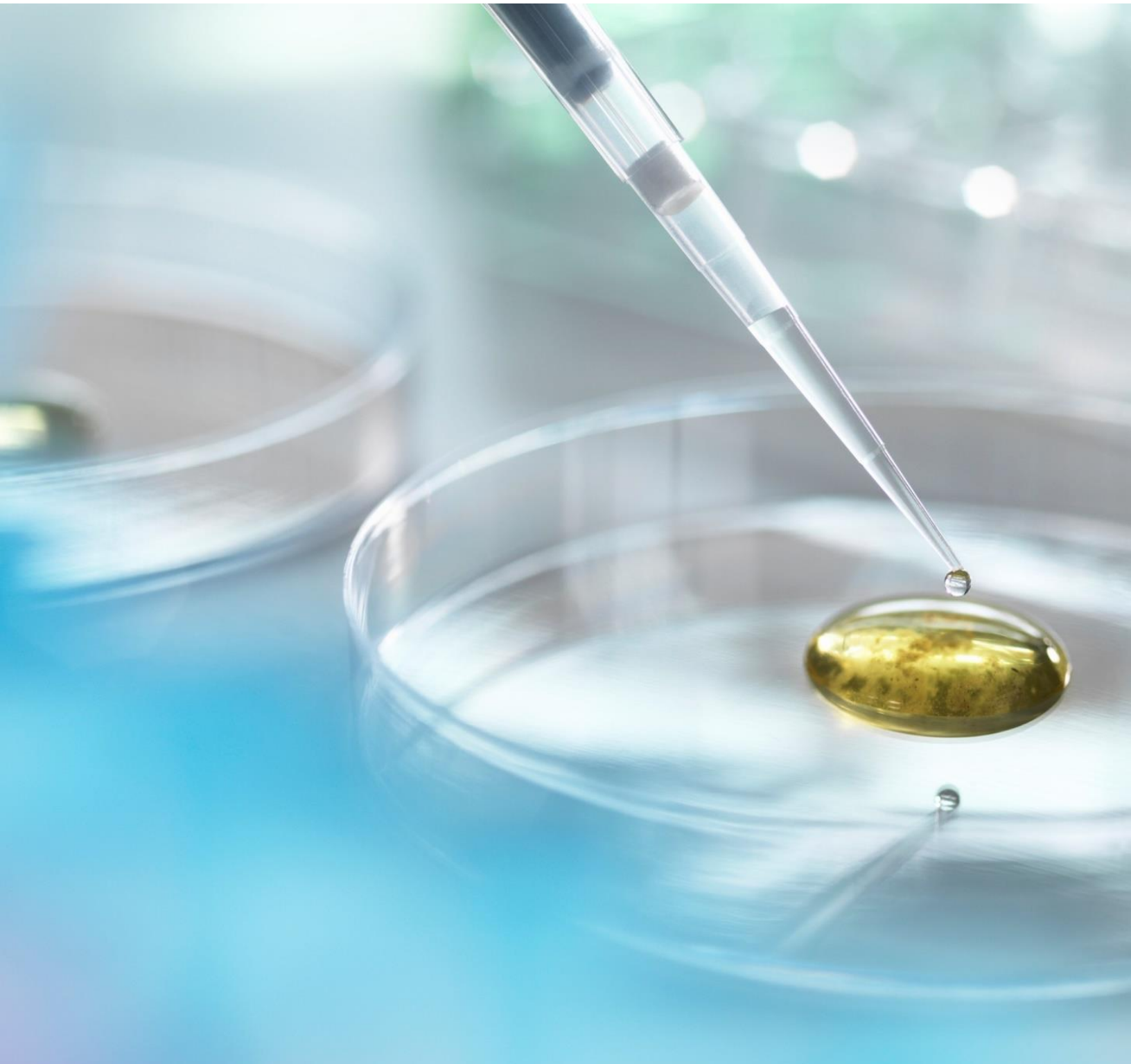
An overview of search and match algorithms complexity and performance

sarah Mohsen fahmy / DR:sara el-metwally /genomics



RELATED WORK

Pattern P can be matched in String T by adding four empty spaces before the pattern and two after.

```
String  T:  A C C T C G A G T
                    | | |
Pattern P:  _ _ _ _ C G A _ _
```

In (Yeh and Cheng 2008), They used Levenshtein distance applied to images and videos to determine feature vectors. For instance:

Input A: 

Input B: 

To find maximum matches we remove last triangle in input A.

In (Amir et al. 2004),They proposed new string distance like Levenshtein distance implemented with Message Passing Interface (MPI).

In (Knuth et al. 1977), This traditional algorithm is now known as KMP string matching algorithm which used for pattern matching in strings.

In (Hussain et al. 2013), named Bidirectional Exact Pattern Matching (BDEPM) which uses pointers in string matching.

```
String  T:  A C C T C G A G T
                  ↑| | |↑
Pattern P:  _ _ _ _ C G A _ _
```

In (Alsmadi and Nuser 2012), They evaluated two algorithms for DNA string comparison. The Longest Common Substring (LCS) algorithm, and Longest Common Sub-Sequence (LCSS) algorithms. In the following example, the highlighted letters, CTCT, in the sequences is LCSS of the specified sequences.

```
String  T:  A C G T C G A G T
              |   |   |     |
Pattern P:  _ C _ T C _ _ _ T
```

Different types of string matching algorithms are explored in (Singla and Garg 2012), concluding that for string matching, Boyer Moore algorithm is the best. In Aho-Corasick performs better than the CommentZ-Walter algorithm.