

Computing the accuracy of an alignment with respect to a reference

Thompson et. al. (Nucleic Acids Res. 1999 Jul 1;27(13):2682-90) describe how to compute the sum-of-pairs accuracy of the computed alignment with respect to the reference. The accuracy is defined as the number of correctly aligned nucleotides in the computed alignment divided by the number of aligned nucleotides in the reference or true alignment.

Let us do an example. Assume the true alignment of ACCTAG and ACTT is given by

```
ACCTAG
A-CTT-
```

and the computed alignment is

```
ACCTAG
AC-TT-
```

There are four aligned nucleotides in the reference. In the computed alignment the first pair of A's are also aligned in the reference. Therefore that is correct. The second pair of C's, however, are not aligned in the reference and therefore that is wrong. The remaining aligned nucleotides in the computed alignment are also correct. Thus the accuracy of the computed alignment is $\frac{3}{4}$ or as a percentage it is 75%. Note that we ignore gaps when computing accuracy.