

# Supplementary material to: i-ADHoRe 2.0: An improved tool to detect degenerated genomic homology using genomic profiles

Cedric Simillion      Koen Janssens      Lieven Sterck  
Yves Van de Peer

August 17, 2007

In order to compare the performance of both alignment methods, i-ADHoRe was run twice on the same dataset, once using the original Needleman-Wunsch (NW) based alignment algorithm and once using the new Greedy-Graph (GG) based alignment method. The output alignments of both runs were compared to assess the difference of both alignment methods. Only pairs of multiplicons that are identical between both runs were considered. Because the detection of a higher-level multiplicon is directly dependant on the alignment of the profile used to detect it, most output multiplicons between both runs are slightly different from each other. Only those pairs of multiplicons were considered for comparison where the first and last genes of each segment were identical. Note that during the construction of profiles, the i-ADHore algorithm sometimes inverts part of a segment. Depending on the alignment method used, different breakpoints can be chosen for these inversions, which can lead to a different number of genes in alignments otherwise considered identical.

The quality of an alignment was assessed by counting the number of misaligned genes. A gene was considered to be misaligned if in a given profile a homolog was present on a different segment and this homolog was positioned in a different column than the reference gene. Conversely, a gene was considered aligned if a homolog was positioned in the same column in the alignment. The number of alignable genes is then the sum of both the number of misaligned and aligned genes. The lower the number of unaligned genes, the better the quality of an alignment is.

Table 1 shows the comparison of both alignment methods for the 21 pairs of identical alignments between both runs. In 11 cases the alignment created with the GG-method had a smaller fraction of unaligned genes whereas only in once case the NW-method gave a better result. For 9 cases no difference was observed, but for 6 of these this was because the number of unaligned genes was zero with both methods. This data shows clearly that the GG alignment method outperforms the original NW method.

	Needleman-Wunsch (NW) method			Greedy Graph-based (GG) method			best performing
	#alignable	#misaligned	fraction mis-aligned	#alignable	#misaligned	fraction mis-aligned	
1	47	4	8.51%	47	1	2.13%	GG
2	28	7	25.00%	30	9	30.00%	NW
3	26	0	0.00%	26	0	0.00%	-
4	20	3	15.00%	22	1	4.55%	GG
5	53	9	16.98%	53	9	16.98%	-
6	26	0	0.00%	26	0	0.00%	-
7	27	2	7.41%	25	0	0.00%	GG
8	59	12	20.34%	61	12	19.67%	GG
9	49	8	16.33%	49	7	14.29%	GG
10	24	0	0.00%	24	0	0.00%	-
11	20	2	10.00%	20	0	0.00%	GG
12	38	5	13.16%	38	4	10.53%	GG
13	27	9	33.33%	27	1	3.70%	GG
14	19	0	0.00%	19	0	0.00%	-
15	32	4	12.50%	32	4	12.50%	-
16	18	0	0.00%	18	0	0.00%	-
17	38	2	5.26%	38	0	0.00%	GG
18	23	1	4.35%	29	1	3.45%	GG
19	27	0	0.00%	27	0	0.00%	-
20	56	9	16.07%	56	9	16.07%	-
21	42	6	14.29%	44	5	11.36%	GG
Total	699	83	11.87%	711	63	8.86%	GG

Table 1: Comparison of the Needleman-Wunsch and Greedy Graph based alignment methods.