Simultaneous Alignment and Structure Prediction of RNAs

Are Three Input Sequences Better than Two?[*]

Beeta Masoumi and Marcel Turcotte

School of Information Technology and Engineering
University of Ottawa
Ottawa, Ontario, Canada
{bmasoumi,turcotte}@site.uottawa.ca

A Web Appendix: Supplementary Material

---

[*] *2005 International Workshop on Bioinformatics Research and Applications*

Table 1. tRNA dataset.

Id	Length	Description
RD0260	77	Asp Phage T5 (Virus)
RD0500	76	Asp *Haloferax volcanii* (Archae)
RD4800	71	Asp *Aedes albopictus* (Mitochondria, Animal)
RE2140	76	Glu *Synechocystis sp.* (Eubacteria)
RE6781	76	Glu *Hordeum vulgare* (Chloroplast)
RF6320	76	Phe *Schizosaccharomyces pombe* (Cytoplasm, Fungi)
RL0503	88	Leu *Haloferax volcanii* (Archae)
RL1141	89	Leu *Mycoplasma capricolum* (Eubacteria)
RS0380	88	Ser *Halobacterium cutirubrum* (Archae)
RS1141	92	Ser *Mycoplasma capricolum* (Eubacteria)

Table 2. 5S rRNA dataset.

Id	Length	Description
AJ131594	117	*Delftia acidovorans*
AJ251080	117	*Geobacillus stearothermophilus*
K02682	120	*Micrococcus luteus*
M10816	119	*Geobacillus stearothermophilus*
M16532	121	*Thermus sp.*
M25591	117	*Geobacillus stearothermophilus*
V00336	120	*Escherichia coli*
X02024	119	*Sporosarcina pasteurii*
X02627	120	*Agrobacterium tumefaciens*
X04585	119	*Rhodobacter capsulatus*
X08000	122	*Arthrobacter oxydans*
X08002	122	*Arthrobacter globiformis*

Table 3. Sensitivity for the tRNA dataset.

Id	N$_{xd}$	N$_{d}$	Min$_{xd}$	Min$_{d}$	Max$_{xd}$	Max$_{d}$	Ave$_{xd}$	Ave$_{d}$
RD0260	4	5	95	57	100	100	98.8	90.5
RD0500	4	5	76	47	95	95	81.0	80.0
RD4800	5	5	95	57	100	100	99.0	91.4
RE2140	2	4	100	95	100	100	100.0	98.8
RE6781	2	4	100	81	100	100	100.0	95.2
RF6320	4	5	95	47	100	100	96.4	89.5
RL0503	1	2	95	95	95	95	95.8	95.8
RL1141	2	3	92	68	92	92	92.0	84.0
RS0380	1	2	92	80	92	80	92.0	80.0
RS1141	2	3	88	65	88	92	88.5	82.1

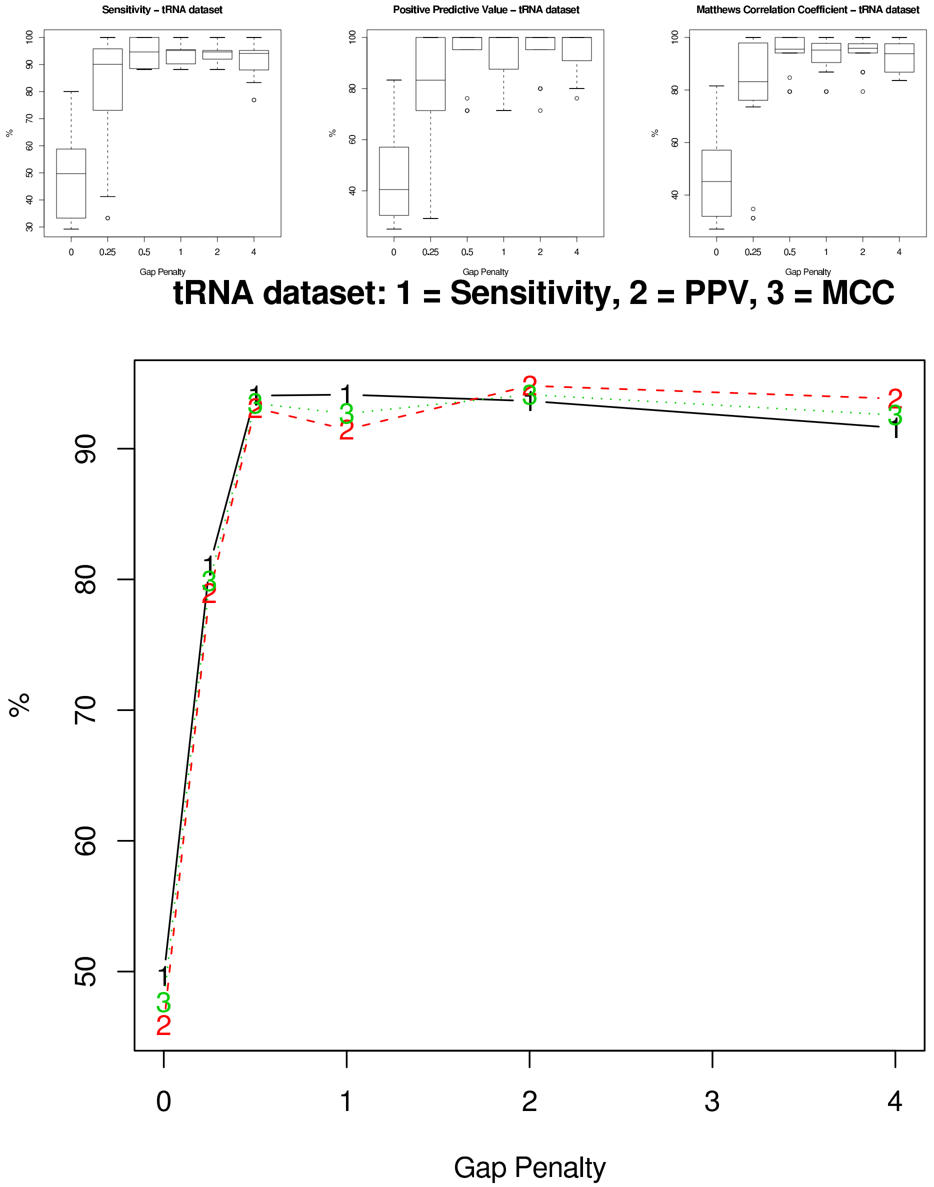


Fig. 1. Effect of various gap penalty scores on PPV, sensitivity and MCC for the tRNA dataset.

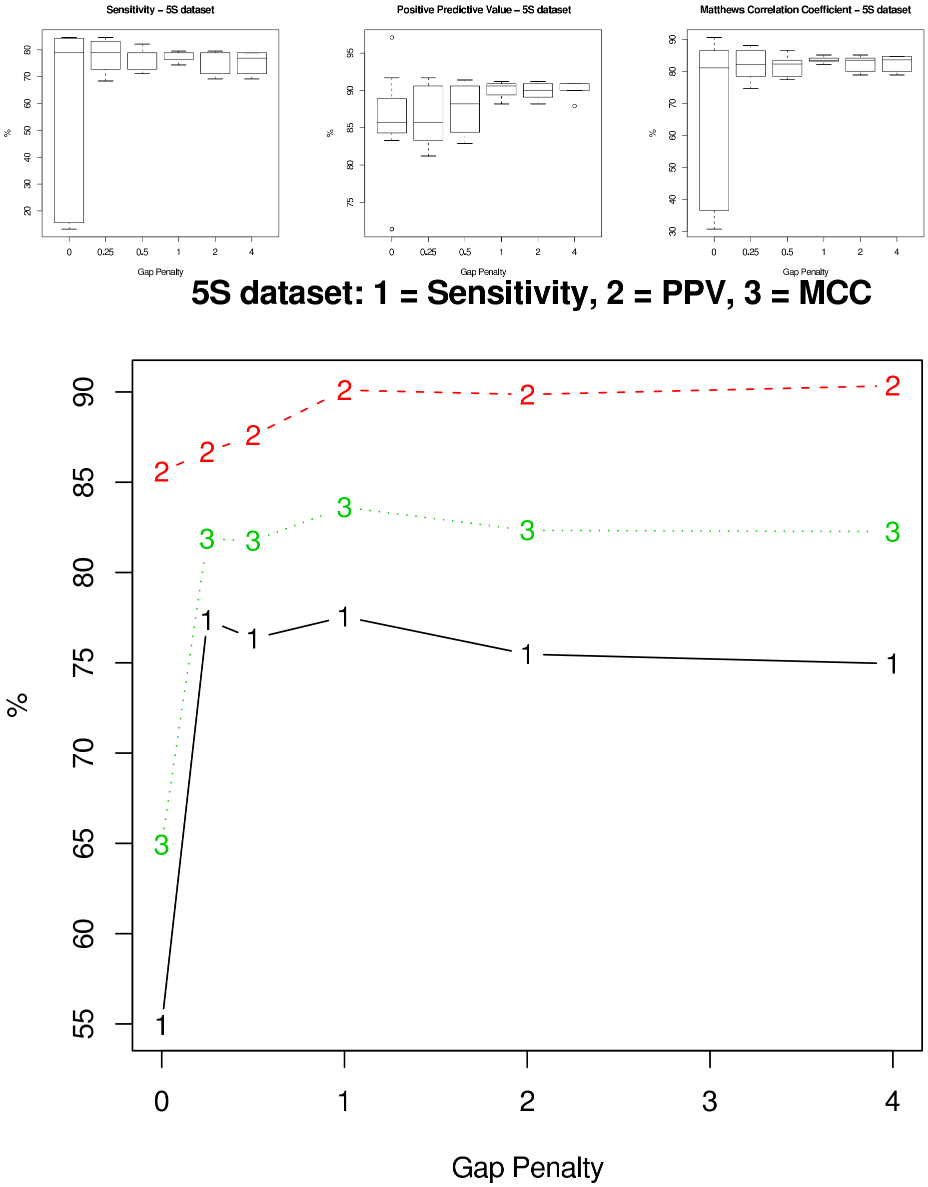
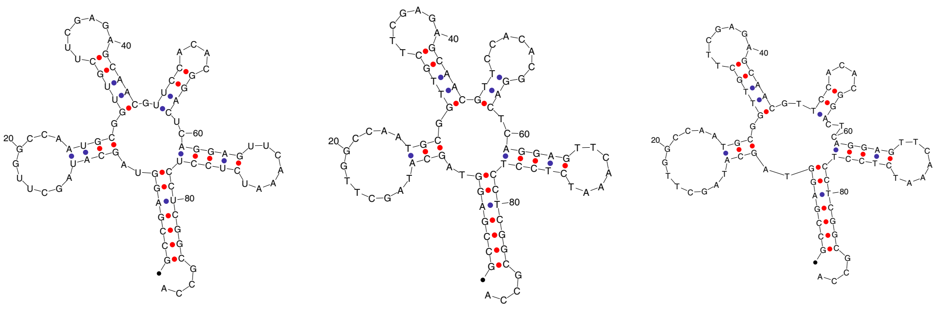


Fig. 2. Effect of various gap penalty scores on MCC, PPV and Sensitivity for the 5S dataset.

Table 4. MCC for the tRNA dataset.

Id	N$_{xd}$	N$_d$	Min$_{xd}$	Min$_d$	Max$_{xd}$	Max$_d$	Ave$_{xd}$	Ave$_d$
RD0260	4	5	97	67	100	100	99.4	93.0
RD0500	4	5	76	46	97	97	81.6	80.4
RD4800	5	5	97	67	100	100	99.5	93.5
RE2140	2	4	100	97	100	100	100.0	99.4
RE6781	2	4	100	79	100	100	100.0	94.8
RF6320	4	5	95	46	100	100	96.4	89.3
RL0503	1	2	97	97	97	97	97.9	97.9
RL1141	2	3	95	69	95	95	95.9	87.1
RS0380	1	2	95	81	95	83	95.9	82.5
RS1141	2	3	94	68	94	96	94.1	86.1

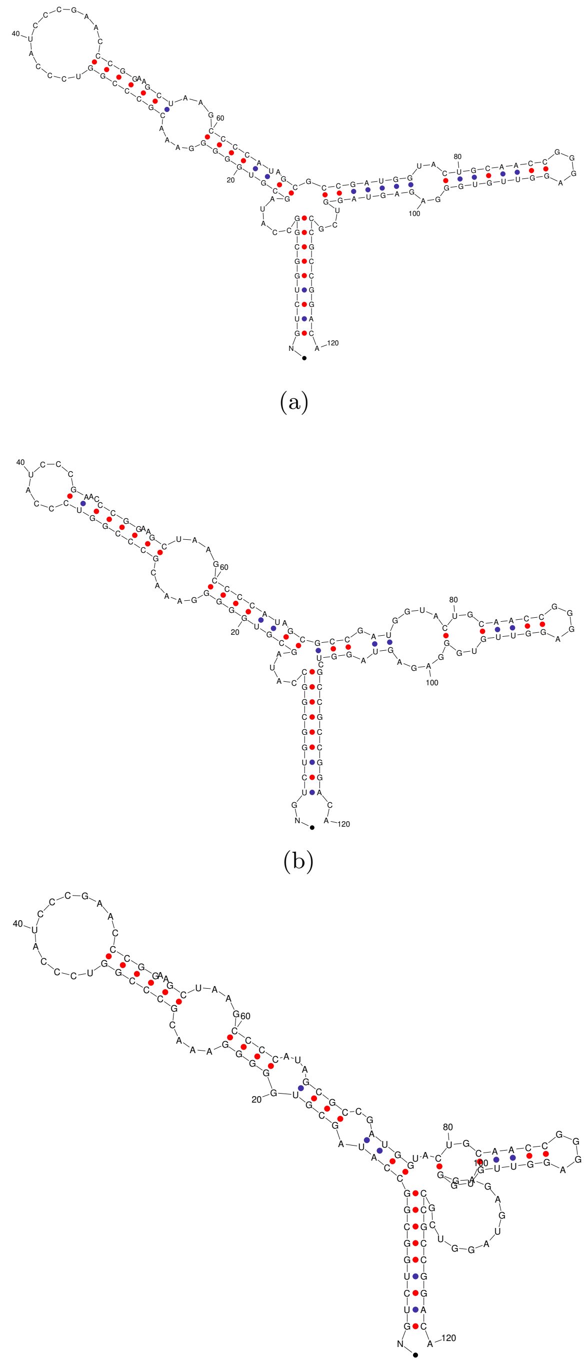


(a) **(b)** **(c)**

Fig. 3. Reference (a), Dynalign (b) and X-Dynalign (c) structures for the tRNA RS0380.

Table 5. Sensitivity for the 5S dataset.

Id	N$_{xd}$	N$_d$	Min$_{xd}$	Min$_d$	Max$_{xd}$	Max$_d$	Ave$_{xd}$	Ave$_d$
AJ131594	2	3	86	86	86	89	86.8	87.7
AJ251080	6	5	76	76	78	84	77.2	79.4
D11460	6	5	73	63	76	81	74.6	71.1
K02682	8	9	53	79	84	89	76.3	84.3
M10816	3	4	76	76	78	84	77.2	80.9
M16532	1	2	82	71	82	76	82.1	74.3
M25591	6	5	76	76	78	84	76.7	79.4
V00336	3	4	62	57	82	90	75.8	78.8
X02024	9	6	76	73	78	84	77.2	76.8
X02627	1	2	84	87	84	89	84.6	88.5
X04585	2	3	63	63	84	81	73.7	74.6
X08000	5	5	74	74	74	79	74.4	77.5
X08002	5	5	74	74	74	79	74.4	77.5



(a)

(b)

(c)

Fig. 4. Reference (a), Dynalign (b) and X-Dynalign (c) secondary structures for the 5S rRNA K02682.

Table 6. MCC for the 5S dataset.

Id	N_{xd}	N_d	Min_{xd}	Min_d	Max_{xd}	Max_d	Ave_{xd}	Ave_d
AJ131594	2	3	93	89	93	93	93.2	91.0
AJ251080	6	5	83	79	84	85	83.5	82.1
D11460	6	5	80	64	81	85	80.8	75.1
K02682	8	9	58	85	92	93	82.4	88.1
M10816	3	4	83	80	84	86	83.7	84.3
M16532	1	2	87	74	87	81	87.9	78.0
M25591	6	5	81	79	83	85	83.0	82.1
V00336	3	4	68	61	90	94	83.5	84.8
X02024	9	6	83	79	84	86	83.4	81.1
X02627	1	2	92	90	92	93	92.0	92.2
X04585	2	3	67	65	89	87	78.4	78.5
X08000	5	5	82	82	82	83	82.1	83.2
X08002	5	5	82	82	82	83	82.1	83.2