



BESCHLEUNIGUNG DES SABA-ALIGNMENTS

Zwischenvortrag zur Diplomarbeit

Frank Hoffmann

Dresden, 03.04.2014

Free-Shift-Alignment

Performance

TODO

01 Free-Shift-Alignment

Berechnet das optimale Alignment zweier Sequenzen (db,read).
 Ein Alignment besteht aus Editieroperationen, welche mit Kosten behaftet sind. Für die Berechnung wird eine Matrix benutzt, in welche die benötigten Kosten eingetragen werden.

		db							
		A	A	C	G	G	T	T	A
read	A	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	C	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	G	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	T	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
	T	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

01 Free-Shift-Alignment

Berechnet das optimale Alignment zweier Sequenzen (db,read).

Ein Alignment besteht aus Editieroperationen, welche mit Kosten behaftet sind. Für die Berechnung wird eine Matrix benutzt, in welche die benötigten Kosten eingetragen werden.

$$SI(x, y) = \max \left\{ \begin{array}{l} SI(x, y - 1) + CIE \\ S(x, y - 1) + CIB \end{array} \right. \quad (1)$$

$$SD(x, y) = \max \left\{ \begin{array}{l} SD(x - 1, y) + CDE \\ S(x - 1, y) + CDB \end{array} \right. \quad (2)$$

$$S(x, y) = \max \left\{ \begin{array}{l} SI(x, y) \\ SD(x, y) \\ S(x - 1, y - 1) + \left\{ \begin{array}{ll} CM & DB(x) = Read(y) \\ CX & DB(x) \neq Read(y) \end{array} \right. \end{array} \right. \quad (3)$$

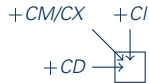
01 Free-Shift-Alignment

Berechnet das optimale Alignment zweier Sequenzen (db,read).

Ein Alignment besteht aus Editieroperationen, welche mit Kosten behaftet sind. Für die Berechnung wird eine Matrix benutzt, in welche die benötigten Kosten eingetragen werden.

		db							
		A	A	C	G	G	T	T	A
read		0	0	0	0	0	0	0	0
	A	1							
	C	2							
	G	3							
	T	4							
	T	5							

Vereinfachung:
alle Kosten, außer CM , sind 1



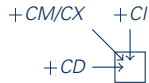
01 Free-Shift-Alignment

Berechnet das optimale Alignment zweier Sequenzen (db,read).

Ein Alignment besteht aus Editieroperationen, welche mit Kosten behaftet sind. Für die Berechnung wird eine Matrix benutzt, in welche die benötigten Kosten eingetragen werden.

		db							
		A	A	C	G	G	T	T	A
read		0	0	0	0	0	0	0	0
	A	1	0	0	1	1	1	1	0
	C	2							
	G	3							
	T	4							
	T	5							

Vereinfachung:
alle Kosten, außer CM , sind 1

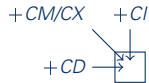


01 Free-Shift-Alignment

Berechnet das optimale Alignment zweier Sequenzen (db,read).
 Ein Alignment besteht aus Editieroperationen, welche mit Kosten behaftet sind. Für die Berechnung wird eine Matrix benutzt, in welche die benötigten Kosten eingetragen werden.

		db							
		A	A	C	G	G	T	T	A
read		0	0	0	0	0	0	0	0
	A	1	0	0	1	1	1	1	0
	C	2	1	1	0	1	2	2	1
	G	3							
	T	4							
	T	5							

Vereinfachung:
 alle Kosten, außer CM , sind 1



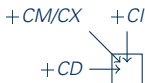
01 Free-Shift-Alignment

Berechnet das optimale Alignment zweier Sequenzen (db,read).

Ein Alignment besteht aus Editieroperationen, welche mit Kosten behaftet sind. Für die Berechnung wird eine Matrix benutzt, in welche die benötigten Kosten eingetragen werden.

		db							
		A	A	C	G	G	T	T	A
read		0	0	0	0	0	0	0	0
	A	1	0	0	1	1	1	1	0
	C	2	1	1	0	1	2	2	1
	G	3	2	2	1	0	1	2	2
	T	4	3	3	2	1	2	1	2
	T	5	4	4	3	2	2	1	2

Vereinfachung:
alle Kosten, außer CM , sind 1



01 Free-Shift-Alignment

Berechnet das optimale Alignment zweier Sequenzen (db,read).

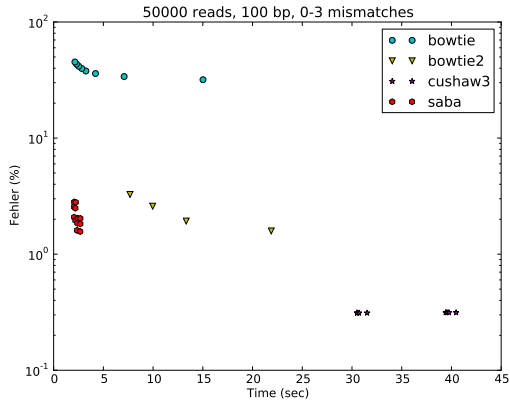
Ein Alignment besteht aus Editieroperationen, welche mit Kosten behaftet sind. Für die Berechnung wird eine Matrix benutzt, in welche die benötigten Kosten eingetragen werden.

		db								
		A	A	C	G	G	T	T	A	
read		0	0	0	0	0	0	0	0	
	A	1	0	0	1	1	1	1	1	0
	C	2	1	1	0	1	2	2	2	1
	G	3	2	2	1	0	1	2	3	2
	T	4	3	3	2	1	2	1	2	3
	T	5	4	4	3	2	2	2	1	2

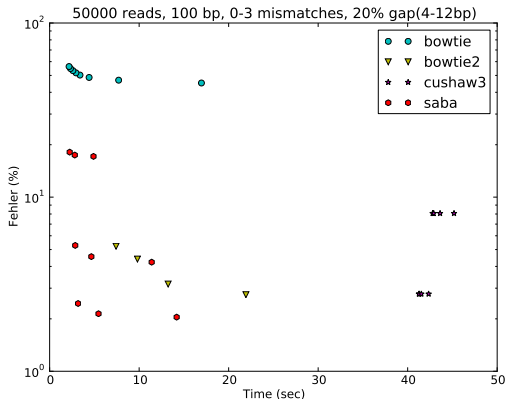
Vereinfachung:
alle Kosten, außer CM , sind 1

Ergebniss:
 $2=1D3=$

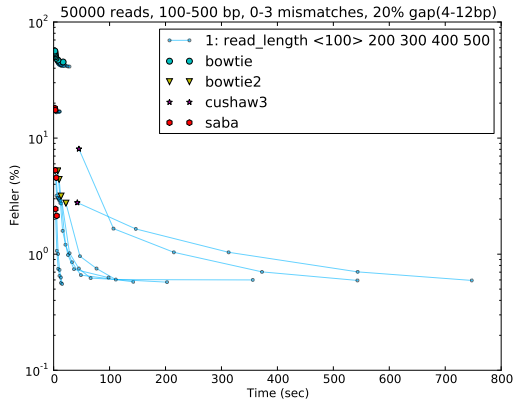
02 Performance



02 Performance



02 Performance



Performance counter stats for:

```
bin/saba-at --db saba-db/chr1 \
    --fa reads/chr1-50000-100--x-p3--prob-20-g-4p12.fa \
    -s 40 --depth 4
```

2981.612914	task-clock	#	0.999 CPUs utilized	
251	context-switches	#	0.000 M/sec	
0	CPU-migrations	#	0.000 M/sec	
498	page-faults	#	0.000 M/sec	
11,276,135,286	cycles	#	3.782 GHz	[83.24%]
6,146,309,393	stalled-cycles-frontend	#	54.51% frontend cycles idle	[83.27%]
4,285,896,208	stalled-cycles-backend	#	38.01% backend cycles idle	[66.74%]
13,746,294,092	instructions	#	1.22 insns per cycle	
		#	0.45 stalled cycles per insn	[83.38%]
1,325,538,374	branches	#	444.571 M/sec	[83.40%]
46,515,090	branch-misses	#	3.51% of all branches	[83.37%]

2.985705701 seconds time elapsed

Performance counter stats for:

```
bin/saba-at-prefetch --db saba-db/chr1 \
    --fa reads/chr1-50000-100--x-p3--prob-20-g-4p12.fa \
    -s 40 --depth 4
```

2805.581417	task-clock	#	0.999 CPUs utilized	
237	context-switches	#	0.000 M/sec	
0	CPU-migrations	#	0.000 M/sec	
497	page-faults	#	0.000 M/sec	
10,635,131,267	cycles	#	3.791 GHz	[83.33%]
5,516,687,220	stalled-cycles-frontend	#	51.87% frontend cycles idle	[83.33%]
3,695,984,513	stalled-cycles-backend	#	34.75% backend cycles idle	[66.69%]
13,810,761,878	instructions	#	1.30 insns per cycle	
		#	0.40 stalled cycles per insn	[83.33%]
1,326,373,072	branches	#	472.762 M/sec	[83.33%]
46,451,183	branch-misses	#	3.50% of all branches	[83.40%]

2.809510533 seconds time elapsed

03 TODO

- Ausgabe von mehr als einem Alignment (Qualität).
- MPI
- Local-Alignment



»Wissen schafft Brücken.«