COMP9319 Web Data Compression and Search

An Occ Implementation, RLFM (Compressed FM Index) Revisit

An example Occ implementation

1

FM Index (when reversing from L[5])

3

4

FM Index (when reversing from L[5])

FM Index $(L(x) \neq c)$

	2	<u>F</u>	<u>L</u> i	<u>i m p s</u>	<u>C</u> # 0
()			1000	
-	1	i	р	1010	i 1
2	2	i	s	1011	m 5
3	3	i	s	1012	p 6
4	1	i	m	1112	s 8
į	5	m	#	1112	
(5	p	p	1 1 2 2	
	7	р	i	2 1 2 2	
8	3	S	s	2 1 2 3	
9	9	S	s	2 1 2 4	
	10	S	i	3 1 2 4	
	11	S	i	4 1 2 4	6

5

FM Ir	ndex (c	on't)			
pssi ⇒ Fst=1 Lst=4	F 0 i i 2 i 3 i i 5 m 6 p 7 p 8 s 9 s s 10 s s 11 s	Lipsm#pissii	i m p s 1 0 0 0 1 0 1 0 1 0 1 1 1 0 1 2 1 1 1 2 1 1 1 2 1 1 2 2 2 1 2 2 2 1 2 3 2 1 2 4 3 1 2 4 4 1 2 4	C # 0 i 1 m 5 p 6 s 8	7

FM In	dex	(con't)			
→	F 0 ##1 1 i 2 i 3 i 44 i 55 mm 66 p 77 p 88 s 99 s 110 s	p	i m r 1 0 0 1 0 1 1 0 1 1 1 1 1 1 1 2 1 2 2 1 2 2 1 2 3 1 2 4 1 2	0 0 1 2 2 2 2 2 2 3 4 4	E # 0 i 1 m 5 p 6 s 8 Fst=8+0 Lst=(8+2)-1

7

FM In	dex (c	on't)		
→	E # 1 i 2 i 3 i 4 i 5 m 6 p 7 p 8 s 9 s 10 s 11 s	Lipssm#pissii	i m p s 1 0 0 0 1 0 1 0 1 0 1 2 1 1 1 2 1 1 1 2 1 1 2 2 2 1 2 2 2 1 2 3 2 1 2 4 3 1 2 4 4 1 2 4	E # 0 i 1 m 5 p 6 s 8 Fst=8+0 Lst=(8+2)-1

FM Ir	ndex (d	con't)		
p <u>ssi</u> ⇒	F 0 # 1 i 2 i 3 i 4 i 5 m 6 P 7 P 8 s 9 s 10 s 11 s	Li pssm# pissii	i m p s 1 0 0 0 0 1 0 1 0 1 0 1 1 1 0 1 2 1 1 1 2 1 1 2 2 2 1 2 2 2 1 2 3 2 1 2 4 3 1 2 4 4 1 2 4	C # 0 i 1 m 5 p 6 s 8 'st=8+2 st=(8+4)-1

9 10

FM I	ndex (c	on't)		
p <u>ssi</u>	F 0 # 1 i 2 i 3 i 4 i 5 m 6 p 7 p 8 s 9 s 10 s 11 s	Lipssm#pissii	i m p s 1 0 0 0 1 0 1 0 1 0 1 2 1 1 1 2 1 1 1 2 2 1 2 2 2 1 2 3 2 1 2 4 3 1 2 4 4 1 2 4	E # 0 i 1 m 5 p 6 s 8 Fst=8+2 Lst=(8+4)-1

		•	con't)		
		F	<u>L</u> i	<u>imps</u>	<u>C</u> # 0
<u>pssi</u>	0	#	i	1000	# O
	1	i	p	1010	i 1
	2	i	s	1011	m 5
	3	i	s	1012	p 6
	4	i	m	1112	s 8
	5	m	#	1112	
	6	р	p	1 1 2 2	
	7	p	i	2 1 2 2	
	8	S	s	2 1 2 3	
	9	S	s	2 1 2 4	Fst=6+2
	10	S	i	3 1 2 4	Lst=(6+2)-1
	11	S	i	4 1 2 4	

11 12

7/12/20

FM Ir	ndex (d	con't)		
<u>pssi</u>	E 1	Li pssm# pissii	i m p s 1 0 0 0 1 0 1 0 1 0 1 1 1 0 1 2 1 1 1 2 1 1 1 2 2 1 2 2 2 1 2 2 2 1 2 3 2 1 2 4 3 1 2 4 4 1 2 4	E # 0 i 1 m 5 p 6 s 8 Fst=6+2 Lst=(6+2)-1 Fst>Lst => No match

FM Index (con't) <u>C</u> # 0 i 1 <u>L</u> <u>imps</u> <u>pssi</u> 0 1 р m 5 p 6 s 8 1 0 1 2 i m m p i р 7 2 1 2 2 р 10 s 11 s 4 1 2 4

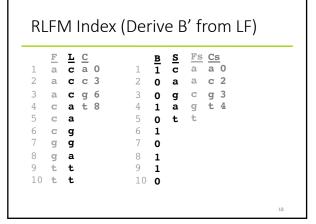
13 14

FM Inc	dex (c	on't)		
_	i i i i i i i i i i i i i i i i i i i	Lipssm#pissii	1 0 1 2 2 1 2 2 4 1 2 4	© # 0 i 1 m 5 p 6 s 8

Similar when L(x) = c<u>C</u> # 0 i 1 <u>L</u> <u>imps</u> 0 р m 5 i i 1 0 1 2 p 6 m 6 р р р 2 1 2 2 To reduce space 10 s 4 1 2 4 11 s

15 16

RLFM Index (Revisit)



17 18

RLFM Index (con't from the prev lecture) <u>L</u> <u>C</u> c a 0 c c 3 $\frac{Fs}{a} \frac{Cs}{a} 0$ F a <u>s</u> <u>B</u> <u>B′</u> 1 ī c 2 а 0 а а a/c g 6 c a t 8 c a g 3 t 4 3 C 0 g 4 5 1 g 1 а 0 t 0 6 7 Lc 1 g g g 0 8 g а 1 9 t 1 10 t 10 0

RLFM Index (con't from the prev lecture) $\frac{\mathbf{F}}{1} \quad \frac{\mathbf{L}}{a} \quad \mathbf{C} \quad \mathbf{C} \quad \mathbf{B} \quad \mathbf{S} \quad \frac{\mathbf{F}}{3} \quad \mathbf{C} \quad \mathbf{S} \quad \mathbf{B}'$

c 2 а **c** c 3 2 а 0 а 0 g 3 t 4 g 6 С 3 0 g t 8 а g 1 1 а а 0 t 0 g g g 0 8 g а t 10 0

19

20

```
RLFM Index (con't from the prev
lecture)
      <u>L</u> <u>C</u> a
           0
                                a 0
                             а
                                      1
      c c 3
                   2
                      0
                         а
                                c 2
                                      0
                   3
                                g 3
      c g 6
                      0
                             C
                   4
5
                                  4
      а
           8
                                t
                      0
                                      0
                   6
6
   C
                      1
                                      0
      g
   g
                      0
                                      1
                   8
   g
      а
                                      0
                      1
                   9
      t
                      1
10
                   10
      t
```

RLFM Index (con't from the prev lecture)

 $\frac{\mathbf{L}}{\mathbf{c}} \stackrel{\mathbf{C}}{=} 0$ <u>s</u> <u>B′</u> 1 a 0 а **c** c 3 0 а c 2 0 g 3 3 С С g 6 0 g а 1 g t а 0 g 0 1 g g 0 1 g t а 0 1 t 1 10 t t 10

21

22

```
RLFM Index (no L & F, nor LF)
                                 <u>B</u>
                                      <u>s</u>
                                                         <u>B′</u>
                            2
                                 0
                                     а
                            3
4
                                 0
1
                                      g
                            5
                                 0
                            6
                                 0
                            8
                                 1
                            10 0
     If only B and S are stored and given... then how \ref{eq:Barter}
```

RLFM Index (no L & F, nor LF)

Fs <u>В</u> 1 <u>B′</u> а C а 0 а 0 1 С g g t а 0 0 8 1 10 0 If only B and S are stored and given... then how $\ref{eq:Barter}$

23

RLFM Index (no L & F, nor LF) $\frac{Fs}{a} \frac{Cs}{a} 0$ <u>s</u> <u>B</u> <u>B′</u> 1 a c 2 0 а c g 3 g t 4 0 g 4 5 6 7 1 а t 0 1 0 8 1 1 10 0

RLFM Index (No LF mapping) Fs Cs a 0 <u>s</u> <u>B</u> <u>B′</u> a c 2 0 а c g 3 g t 4 0 g g 1 а 0 t 0 1 10 0

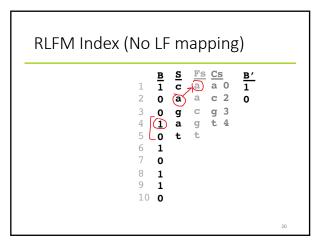
25 26

RLFM Index (No LF mapping) Fs Cs a 0 <u>В</u> 1 <u>B′</u> 2 0 (a) a c 2 **g** c g 3 0 1 t 4 4 5 6 7 0 0 8 1 1 9 10 0

RLFM Index (No LF mapping)

| B | S | F | CS | B' |
| 1 | 1 | c | a | a | c | 2
| 3 | 0 | g | c | g | 3
| 4 | 1 | a | g | t | 4
| 5 | 0 | t | t
| 6 | 1 | 7 | 0
| 8 | 1 | 9 | 1
| 10 | 0

27 28



29 30

```
RLFM Index (No LF mapping)

| B | S | F | CS | B' | 1 | 2 | 0 | a | a | c | 2 | 0 | 3 | 0 | g | c | g | 3 | 4 | 1 | (a) g | t | 4 | 5 | 0 | t | t | 6 | 1 | 7 | 0 | 8 | 1 | 9 | 1 | 10 | 0 | 0
```

RLFM Index (No LF mapping) Fs <u>Cs</u> a a 0 (a) c 2 <u>s</u> <u>B</u> <u>B'</u> 0 а 0 g 3 t 4 0 1 0 1 C g 1 a g t 7 **0** 8 **1** 9 **1** 10 **0**

31 32

```
RLFM Index (No LF mapping)
                                   \frac{Fs}{a} \frac{Cs}{a}_0
                           <u>В</u>
1
                                              <u>B'</u>
                       2
                           0
                               а
                                   а
                                      c 2
                                              0
                              g 3
t 4
                       3
                           0
                                   g
t
                       4
5
6
7
                           1
                           0
                           0
                       8
                           1
1
                       9
                       10 0
```

33 34

```
RLFM Index (No LF mapping)
                                      \frac{Fs}{a} \frac{Cs}{a} 0
                              <u>B</u>
                                                  <u>B'</u>
                                  C
                                      a c 2
                              0 a
                                                   0
                             0 (g)
1 a
                                      c
G
                                          g 3
t 4
                                                  1
1
                         3
4
5
6
7
                             0
                                                   0
                             0
                         8
                              1
1
                         10 0
```

35 36

RLFM Index (No LF mapping) $\frac{Fs}{a} \frac{Cs}{a}_0$ <u>B′</u> <u>B</u> ī C 1 2 c 2 0 а а 0 g 3 t 4 0 g c 1 4 5 6 7 **(g)** 1 а 1 0 t t 0 **(i)** 0 0 1 8 1 n 1 10 0

RLFM Index (No LF mapping) $\frac{Fs}{a} \frac{Cs}{a} 0$ <u>B′</u> <u>B</u> ī С 1 c 2 0 а а 0 g 3 C 0 g 1 t 4 g 1 а 1 ćŧ 0 (t 0 1 8 0 9 1 1 10 o

37 38

Now we have B, S, **B'**Let's **reverse (decode)** using LF mapping

CHANGES TO FORMULAS

- o Recall that we need to compute $C_T[c]+\frac{rank_c(L.i)}{rank_c(L.i)}$ in the backward search.
- o Theorem: $C[c]+rank_c(L_i)$ is equivalent to $select_1(B,C_s[c]+1+rank_c(S,rank_1(B_i)))-1$, when $L[i]\neq c$ (e.g., when backward search), and otherwise (e.g., when reverse, sometimes backward search too) to $select_1(B',C_s[c]+rank_c(S,rank_1(B_i)))+i-select_1(B,rank_1(B_i)).$

You can apply these formulas to do reversing & backward search.

40

39

CHANGES TO FORMULAS

- o Recall that we need to compute $C_T[c] + \underbrace{rank_c(L,i)}_{} \text{ in the backward search.}$
- o Theorem: C[c]+rank_c(L,i) is equivalent to select₁(B',C_S[c]+1+rank_c(S,rank₁(B,i)))-1, when L[i]≠c (e.g., when backward search), and otherwise (e.g., when reverse, sometimes backward search too) to select₁(B',C_S[c]+rank_c(S,rank₁(B,i)))+ i-select₁(B,rank₁(B,i)).

But I promised that I would explain why/how these formulas actually work

41

RLFM Index (con't from the prev lecture)

```
<u>L</u> C
                              Fs Cs
      c a 0
                              a a 0
   а
                       ī
                                        ī
                           C
      c c 3
                              a c 2
   a
                   2
                       0
                          а
                                        0
                                 g 3
t 4
                              С
   a,
      c g 6
                       0
                          g
                                        1
   c a t 8
                       1
                              g
                                        1
                           а
5
   С
      а
                       0
                                        0
      g
   g
                       0
      g
                                        1
8
   g
      a
                    8
                       1
                                        0
9
       t
                    9
10 t t
                    10 0
                                        0
Suppose reverse from L[8]
```

41 42

RLFM Index (con't from the prev lecture) $\frac{Fs}{a} \frac{Cs}{a} 0$ <u>B</u> <u>s</u> <u>B′</u> 1 c 2 0 а 0 g t 0 g 1 1 $rank_{\underline{a}}(S, rank_1(B, 8)) = 2$ 0 8 1 1 1 10 0 0

RLFM Index (con't from the prev lecture) Fs Cs a 0 <u>В</u> <u>s</u> <u>B′</u> 1 c 2 а 0 а 0 0 g 1 1 ₄a 0 $rank_a(S, rank_1(B, 8))$ = 2 0/ 1 1 10 0 0

43 44

RLFM Index (con't from the prev lecture) a a 0 $\overline{1}$ 0 c 2 0 g 3 0 1 0 1 6 0 0 1 8 0 9 1 10 select₁(B', C_s[a]+rank_a(S, rank₁(B, 8)))

RLFM Index (con't from the prev lecture) $\frac{\mathbf{L}}{\mathbf{c}} \stackrel{\mathbf{C}}{=} 0$ <u>B′</u> 1 a 0 а **c** c 3 0 c 2 0 g 3 c g 6 (1) 0 g g 1 g t a t 1 10 t 10 $select_1(B', C_s[a]+rank_a(S, rank_1(B, 8))) = 3$

45 46

Good, but not good enough

```
RLFM Index (con't from the prev
lecture)
                                    Fs Cs
                                               <u>B'</u>
                            <u>В</u>
                                    a
                                       a 0
                               C
                                   а
                                       c 2
                            0
                               а
                                               0
                                       g
t
                            0
                               g
                                   g
t
                            1
                               а
                            0
                       10 0
                 select_1(B', C_s[c]+rank_c(S, rank_1(B, 3)))
                 =select<sub>1</sub>(B', 2 + 1)=4
```

47 48

RLFM Index (con't from the prev lecture) $\frac{\mathbf{L}}{\mathbf{c}} \stackrel{\mathbf{C}}{\mathbf{a}}$ $\frac{Fs}{a} \frac{Cs}{a}_0$ <u>B</u> 0 а ī C 1 2 c 2 а С C 3 0 а а 0 g 3 t 4 3 C g 6 0 g 1 8 C а g 1 а C а 0 t C g g g 0 1 8 8 g а 0 9 t 10 0 $select_1(B', C_s[c]+rank_c(S, rank_1(B, 3)))$ =select₁(B', 2 + 1)=4

RLFM Index (con't from the prev lecture) <u>L</u> <u>C</u> a 0 $\frac{Fs}{a} \frac{Cs}{a} 0$ <u>B</u> <u>B′</u> ī 1 ď c 2 а **c** c 3 0 а а 0 3 C g 6 g 0 g 1 t 8 g a t 1 а 1 а 0 t g g g 0 1 8 g а 8 1 0 t 1 t 10 0 $select_1(B', C_s[c]+rank_c(S, rank_1(B, 3)))$ =select₁(B', 2 + 1)=4 + 2

50

52

49

RLFM Index (con't from the prev lecture) <u>L</u> <u>C</u> a <u>В</u> 1 a 0 а a 0 1 **c** c 3 2 c 2 0 a 0 3 g 3 c g 6 lo C 8 1 t 5 0 c 6 7 6 0 g 1 g 0 1 g а 0 1 t 9 1 10 10 0 t $select_1(B', C_s[c]+rank_c(S, rank_1(B, 3)))$ =select₁(B', 2 + 1)= $4 + (i - rank_1(B, i))$

Another example, LF[5] = ?

51

RLFM Index (con't from the prev lecture) Fs Cs <u>В</u> 1 a a 0 c 1 c 2 2 0 ₄a а 0 g 3 С 0 1 g t а

1/ g 5 0 1 0 1 8 9 10 0 $select_1(B', C_s[a]+rank_a(S, rank_1(B, 5)))$ =select₁(B', 0 + 1)=1 + (i -rank₁(B, i))

RLFM Index (con't from the prev lecture) Fs Cs a 0 а ī 1 C c 2 а 0 ₄a 0 g t С 3 0 g 1 1/ g а 0 0 0 $select_1(B', C_s[a]+rank_1(S, rank_1(B, 5)))$ =select₁(B', 0 + 1)=1 + (i -rank₁(B, i))

54 53

7/12/20

RLFM Index (con't from the prev lecture) $\frac{Fs}{a} \frac{Cs}{a} 0$ <u>B</u> <u>B′</u> C 1 1 c 2 0 ₄a а 0 g 3 t 4 C 0 g 1 g 1/ а o t 1 0 1 8 0 10 0 $select_1(B', C_s[a]+rank_a(S, rank_1(B, 5)))$ =select₁(B', 0 + 1)=1 + (i -rank₁(B, i)) 55

RLFM Index (con't from the prev lecture) Fs Cs a a 0 <u>B</u> <u>B′</u> c 1 1 c 2 0 ₁a а 0 g 3 0 g 1 g t 1/ а 0 t 1 0 1 0 10 0
$$\begin{split} & select_1(B', C_s[a] + rank_{\underline{a}}(S, rank_1(B, 5))) \\ & = select_1(B', O + 1) = 1 + (i - select_1(B, rank_1(B, i))) \end{split} \label{eq:select_1}$$

56

RLFM Index (con't from the prev lecture) B' а a 0 1 c 2 0 ₄a 0 g 3 0 C 1 g t 4 0 6 1 0 0 O 9 1 10 0 $select_1(B', C_s[a]+rank_a(S, rank_1(B, 5)))$ =select₁(B', 0 + 1)=1 + (i -select₁(B, rank₁(B, i))) ₅₇

RLFM Index (con't from the prev lecture) <u>L</u> <u>C</u> a В' a 0 а a 0 <u>1</u> **c** c 3 c 2 0 ₄a 0 **c** g 6 0 g C g 3 1/ g t 0 g 1 g g 0 g t а 0 1 t 1 1 t t 10
$$\begin{split} & select_1(B', C_s[a] + rank_1(S, rank_1(B, 5))) \\ & = select_1(B', O + 1) = 1 + \left(\underbrace{i - select_1(B, rank_1(B, i))} \right) \end{split}$$

57 58

RLFM Index (con't from the prev lecture) **CHANGES TO FORMULAS** o Recall that we need to compute $C_T[c] + rank_c(L,i)$ in the backward search. $\begin{tabular}{ll} \begin{tabular}{ll} \be$ when L[i] c (e.g., when backward search), and otherwise (e.g., when reverse, sometimes backward search too) to select,(B,Cs;[c]+rank,(S,rank,(B,i)))+ i-select,(B,rank,(B,i)). $select_1(B', C_s[a] + rank_a(S, rank_1(B, 5)))$ =select₁(B', 0 + 1)=1 + (i -select₁(B, rank₁(B, i))) **Backward Search**

60 59

7/12/20

Backward sear	rch ·	for	"ci'	,		
Dackward Scar	CH	101	ار خ			
	В	S	F.	С	B′	
1	<u>B</u>	<u>s</u> i	<u>F</u> s #	<u>C</u> # 0	<u>B′</u> 1	
2	1	р	i	i 1	1	
3	1	s	i	m 4	1	
4	0	m	i	p 5	1	
5	1	#	m	s 7	0	
6	_	-	р		1	
7	_	i	р		1	
8	_	s	s		1	
9	_	i	s		1	
1	.0 0				0	
1	.1 1				Ţ	
1	.2 0				0	61

Backward search for "si" <u>В</u> 1 <u>s</u> C B' 2 i i 1 1 1 р i i 1 s 0 m 1 # p i р p 1 8 1 9 **1** 10 **0** 1 11 **1** 1 12 **0** 0

61 62

Backward search for "si" <u>s</u> i C B' 1 p i i 1 1 1 m 4 1 s i р 5 s 7 i m 1 # m 1 p p 1 i p # m 0 **1** c = i 1 Lst = 5 8 1 1 1 s s 1 10 0 0 1 11 12 **0**

Backward search for "si" <u>C</u> # 0 <u>C</u> 0 1 р i 1 2 1 i i 1 -1 р m 5 3 1 i s s p 6 s 8 р 5 s 7 0 5 6 7 i m 1 # m 0 1 m # р **1** c=i р 1 Fst = 2 1 i p p s 1 1 s i 8 s s 10 0 10 s 0 i 11 11 s 1 12 0

63 64

Backward search for "si" <u>C</u> # 0 i 1 <u>C</u> # 0 <u>В</u> 1 <u>s</u> <u>B'</u> 1 <u>L</u> 2 i i 1 -1 р р m 4 p 5 1 1 s m s s m 5 3 i 1 p 6 4 5 6 0 i i s 8 1 # m s 7 0 m m # 1 рр 1 1 i p Fst = 1 р р 8 C[c] + Occ(c, 1 1 8 i р s s Fst - 1) + 1 = ? 1 1 10 s s 0 0 11 s i 11 **1** 12 **0** 0

Backward search for "si" <u>F</u>s # i <u>B'</u> 1 <u>C</u> # 0 i 1 <u>s</u> i <u>C</u> # 0 i 1 <u>L</u> <u>В</u> 1 0 р р -1 s i m i s m 5 3 1 **p** 6 p 5 1 4 0 Fr. 2 # i s 8 1 m 0 c = i Fst = 2 Lst = 5 m 1 р 1 7 1 i р р р 8 8 i 1 р s s 9 1 s 10 0 11 s i 11 **1** 12 s **i** 12 **0** 0

65 66

	F	<u>L</u> i	<u>C</u> # 0		<u>В</u> 1	<u>s</u> i	<u>F</u> s #	<u>C</u> # 0	<u>B′</u>	
1	#	i	# O	1	1	i	#	# 0	1	
2	i	р	i 1	2	1-	p	i	i 1	1	
3	i	s	m 5	3	1	s	i	m 4	1	
4	i	s	p 6	4	0	- m	i	p 5	1	
5	i	m	s 8	5	1	#	m	s 7	0	
6	m	#		6	1	р	р		1	c = s
7	р	р		7	1	i	p		1	Fst = ??
8	р	i		8	1	s	s		1	
9	s	s		9	1	i	s		1	
10	s	s		10	0				0	
11	S	i		11	1				1	
12	S	i		12	0				0	

Backward search for "si" <u>L</u> <u>C</u> # 0 <u>В</u> 1 <u>s</u> i F_s # i <u>C</u> # 0 i 1 <u>B'</u> 1 i 1 2 1 → p 1 1 1 i i 3 m 4 s m 5 1 s 4 **p** 6 $0 \nearrow m$ s р i s 8 # 1 m 0 p i 6 7 m 1 р 1 p 1 1 р р 8 1 1 р i s s ranks(S, rank1(B,2-1)) S s 9 1 i 10 s 10 0 s i 11 1 11 s 12 **s** 12 0 0

67 68

```
Backward search for "si"
            <u>C</u>
# 0
                                      <u>C</u>
#
        <u>L</u>
                            ī
                                         0
                                            1
        р
            i 1
                     2
                         1 → p
                                 i
                                      i 1
                                            1
                     3
                         1
                                      m 4
                                            1
        s
            m 5
                            s
                                 i
            p 6
                                      p 5
        s
                     4
                         0
                                 i
                            🖈 m
    i
       m
                     5
                         1
                             #
                                 m
                                      s 7
                                            0
                                р
   m
                         1
                     6
        #
                                            1 c = s
                             р
                                               Fst
Occ of s:
                                            1
    р
       р
                             i
                                 р
                         1
    p
s
        i
                     8
                                            1
8
                             s
                                 s
                                               ranks(S,
rank1(B,2-1))
                     9
                                             1
        s
                                             0
10 s
        s
i
                     10 0
                                                selecti(B',7+
                                             1
11 s
                     11
                         1
                     12 0
                                               So Fst = 9
```

Backward search for "si" <u>C</u> # 0 <u>C</u> # 0 i 1 <u>s</u> <u>L</u> 1 i p i 1 2 1 → p i 1 i m 5 3 1 i m 1 s s p 6 s 8 p 5 s 7 i s 0 🥕 m i i m 5 6 7 1 # m 0 m 1 # р **1** c = s р 1 1 <u>Fst</u> Occ of s: р р i р s i p s i 8 1 s rank_s(S, rank₁(B,2-1)) 9 1 s 10 s 10 0 0 s selecti(B',7+ i 11 11 S 1 1+0) So Fst = 9 12 **s** 12 0

69 70

```
Backward search for "si"
                                          <u>C</u>
# 0
i 1
                                                B'
1
1 c=s
1 Lst
             <u>C</u>
# 0
i 1
                                    <u>F</u>s
#
i
        <u>L</u>
                            <u>В</u>
1
                                i
                           1 → p
                       2
    i
         р
                                          m 4
p 5
                                                1
1
         s
s
              m 5
                       3
                            1 s
                                    i
                                                     ranks(S,
ranks(B,5))
             p 6
                       4
                                    i
    i
                            0
                              💉 m
                       5
                                                 0
    i
                            1
                                #
                                          s 7
         m
              s 8
                                    m
    m
         #
                       6
                            1
                                р
                                    р
                                                 1
                            1
                                i
                                                 1
    р
         р
                                    р
                            1
8
                       8
                                                 1
    р
         i
                                s
                                     s
                       9
9
                            1
                                                 1
    S
         s
                       10
                            0
                                                 0
11 s
        i
                       11
                           1
                                                 1
12 s i
                       12 0
                                                 0
```

```
Backward search for "si"
                                                                         <u>B'</u>
                                               <u>s</u>
i
                                                       <u>F</u>s
#
i
                                                               <u>C</u>
# 0
i 1
             <u>L</u>
                    <u>C</u>
# 0
i 1
                                         <u>В</u>
1
                                                                         1
      i
                                  2
                                          \mathbf{1}^{\longrightarrow}\,\mathbf{p}
             р
                                                      i
i
             s
s
                    m 5
                                   3
                                                                m 4
                                          1 s
                                                                                rank<sub>s</sub>(S,
rank<sub>1</sub>(B,5))
                    p 6
                                   4
                                                                p 5
                                                                         1
      i
                                          0
                                              , m
                                   5
      i
                     s 8
                                          1
                                                #
                                                                          0
                                                       m
                                                                s
                                                                               = 1
select<sub>1</sub>(B',7+
1+1) = 11
11 - 1 = 10
So Lst = 10
      m
                                          1
                                                р
                                                       р
                                                                          1
                                   7
                                          1
                                                i
      р
                                                       р
             р
                                   8
8
                                         1
      р
             i
                                                s
                                                       s
                                   9
                                          1
                                                                               -1: since
inclusively,
e.g., Lst-Fst+1
= #matches
      S
             s
                                   10
                                         0
      S
             i
                                   11
                                         1
12 s
             i
                                   12 0
                                                                          0
                                                                                       72
```

71 72

Backward search for "ssi"										
1 2 3 4 5 6 7 8 9 10 11	F#:i:i:m ppsssss	Li pssm # pissii	C # 0 i 1 m 5 6 s 8	1 2 3 4 5 6 7 8 9 10 11	B 1 1 1 0 1 1 1 0 1 0 0	Sip s m # p i s s i	F # i i i m p p s s	C # 0 i 1 m 4 p 5 s 7	B' 1 1 1 1 0 1 1 1 0 1 0 1	c = s <u>Fst</u> Occ of s: rank ₄ (S, rank ₁ (B,9-1)) = 1 select ₁ (B',7+ 1-1) So Fst = 11

 C
 B'

 # 0 1 c = s

 i 1 1 Lst

 m 4 1 Occ of s:

 rank.(S,

 p 5 1 rank.(B,10))

 j 7 0 = 1

 1 since L(i]=c,

 select.(B',Cs[c]+

 rank.(S,rank.(B,

 1 i)))+i

 select.(B,rank.(

 B,i)).

 o select.(B',7+2)

 1 = 11

 0 I1 + 1 = 12

 So Lst = ½2
 Backward search for "ssi" <u>L</u> <u>C</u> # 0 <u>В</u> <u>s</u> i <u>F</u>s F # 2 i i 1 2 i р 1 р i m 5 3 i s 1 s i i p 6 0 s m i s 8 5 1 # m p i m 1 р 1 р р р 8 8 1 р i S 9 9 1 i s 10 10 0 S S i 11 11 s 1 12 **s** 12 0

73 74

Backward search for "issi" <u>C</u> # 0 <u>B′</u> 1 <u>L</u> i <u>s</u> <u>C</u> # ī 0 i i 1 2 1 i **i** 1 1 р р 3 1 m 4 1 i s m 5 s i p 5 p 6 0 i m i m s 8 5 1 # m s 7 0 Fst Occ of i: 6 1 6 m # 1 р р 1 р р i р 1 rank₅(S. ranks(S, ranks(B,11-1)) = 2 selects(B',1+ 8 8 1 1 р i s s 9 1 1 S s 10 s 10 0 0 s i 1+2) So Fst = 4 11 S 11 1 1 12 **s** 12 0

Backward search for "issi" <u>C</u> # 0 <u>s</u> <u>C</u> <u>L</u> 0 1 c=i 1 Lst Occ of i: 1 ranks(S, 1 ranks(B,12)) i i 1 2 1 i i 1 р р i m 5 3 i m 4 s 1 s p 5 s 7 p 6 s 0 i m 0 = 3 1 Since L[i]=c, selecti(B, Cs[c]+ 1 rankc(S, rankı(B, 1 i)))+ i-selecti(B, rankı(i m 5 1 # m 6 m 1 # 6 р р р р 1 i р 8 8 1 i s р s select₁(B,rank₁(B,i)). 9 1 S s 0 selecti(B',1+3) 1 = 4 0 4+1=5 So Lst = 5 10 0 10 s s 11 11 S i 1 12 S 12

75 76

Therefore ...

CHANGES TO FORMULAS

• Recall that we need to compute

CT[C]+rankc(L,i) in the backward search.
• Theorem: C[C]+rankc(L,i) is equivalent to

select, (B',Cs[C]+1+rankc(S,rank₁(B,i)))-1

when L[i]≠ c (e.g., when backward search), and
otherwise (e.g., when reverse, sometimes

backward search too) to

select₁(B',Cs[C]+rankc(S,rank₁(B,i)))+

i-select₁(B,rank₁(B,i)).