

# Suffix Arrays: min-LCP skipping

Ben Langmead



JOHNS HOPKINS

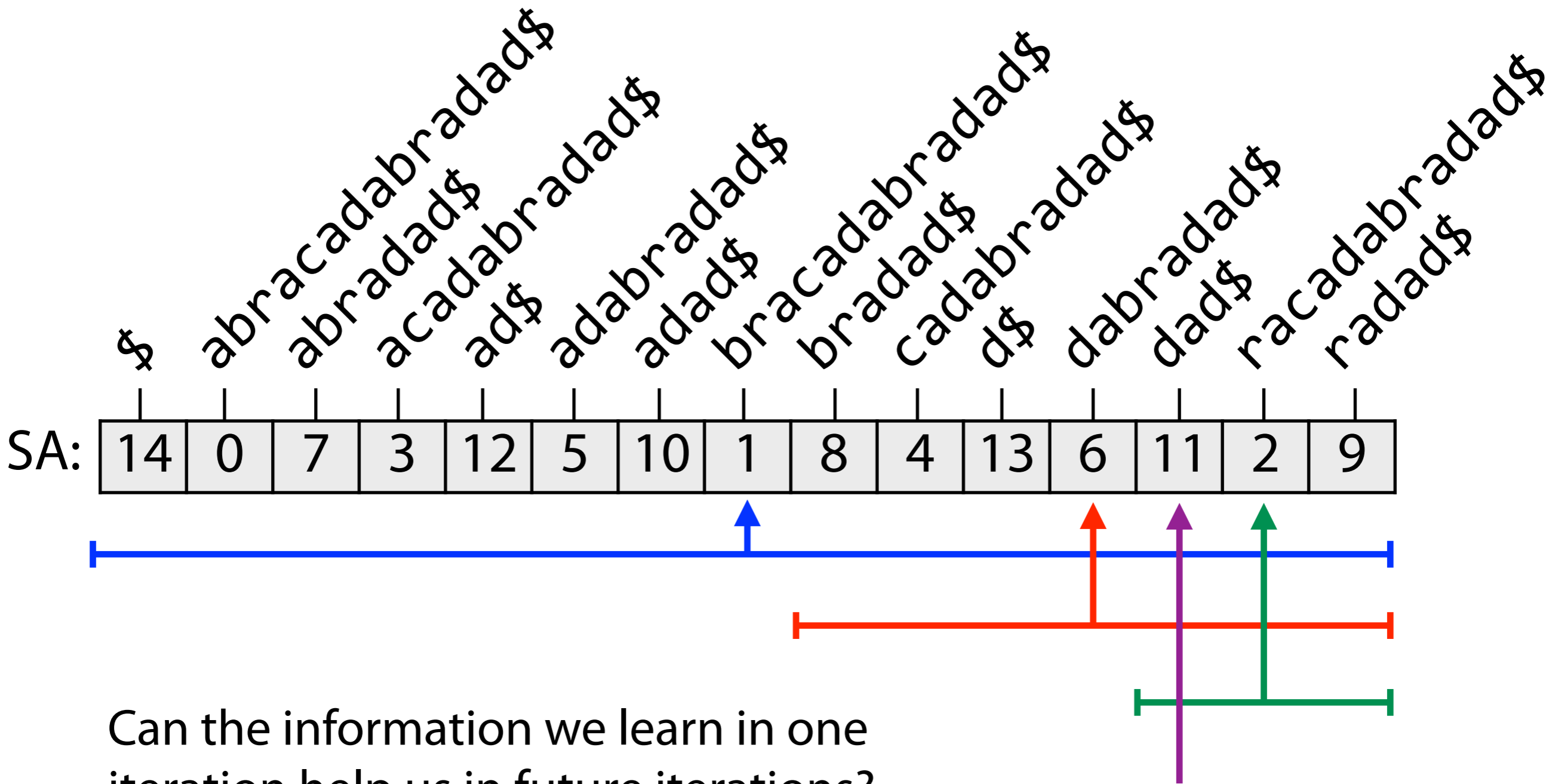
WHITING SCHOOL  
*of* ENGINEERING

Department of Computer Science



Please sign guestbook ([www.langmead-lab.org/teaching-materials](http://www.langmead-lab.org/teaching-materials)) to tell me briefly how you are using the slides. For original Keynote files, email me ([ben.langmead@gmail.com](mailto:ben.langmead@gmail.com)).

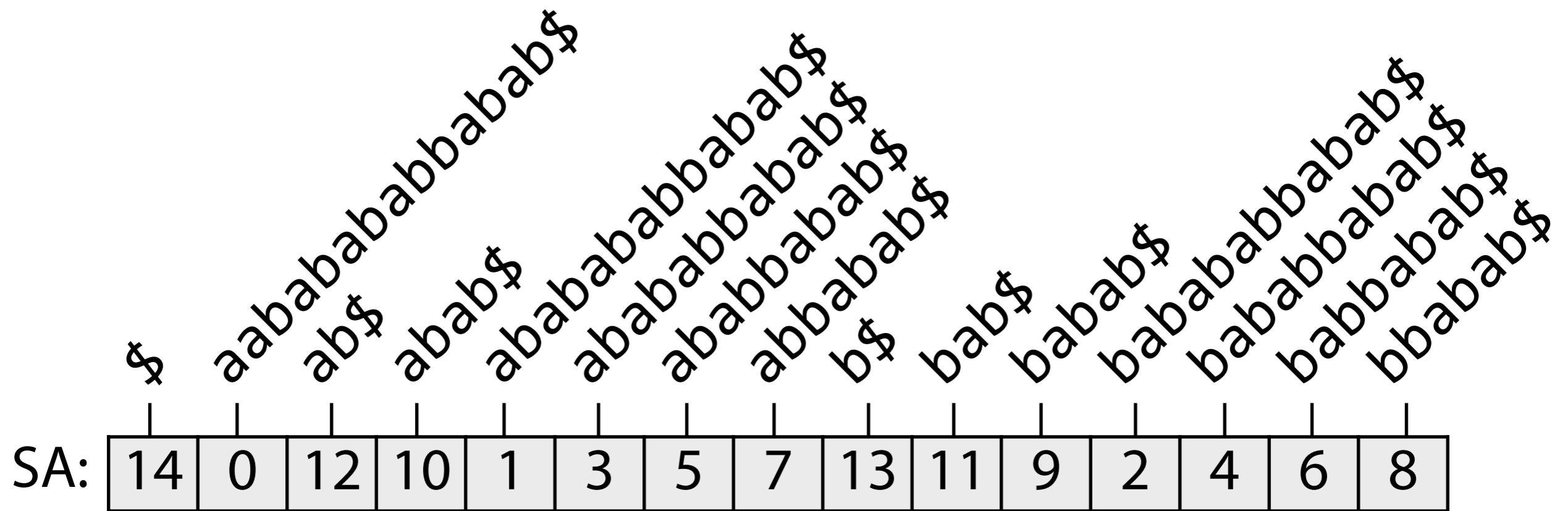
# Suffix array: querying



Can the information we learn in one iteration help us in future iterations?

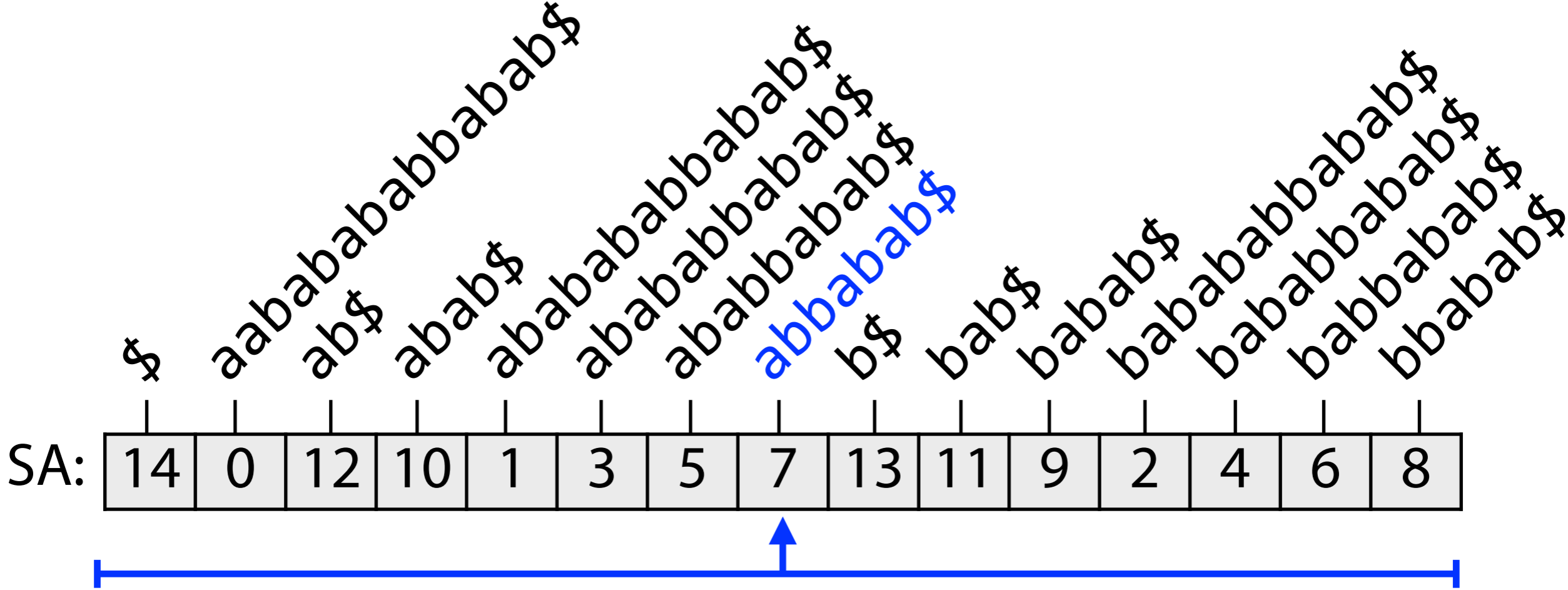
Specifically: information about **common prefixes**

# Suffix array



New example: `T = aababababbabab$`    `P = ababaa`

# Suffix array

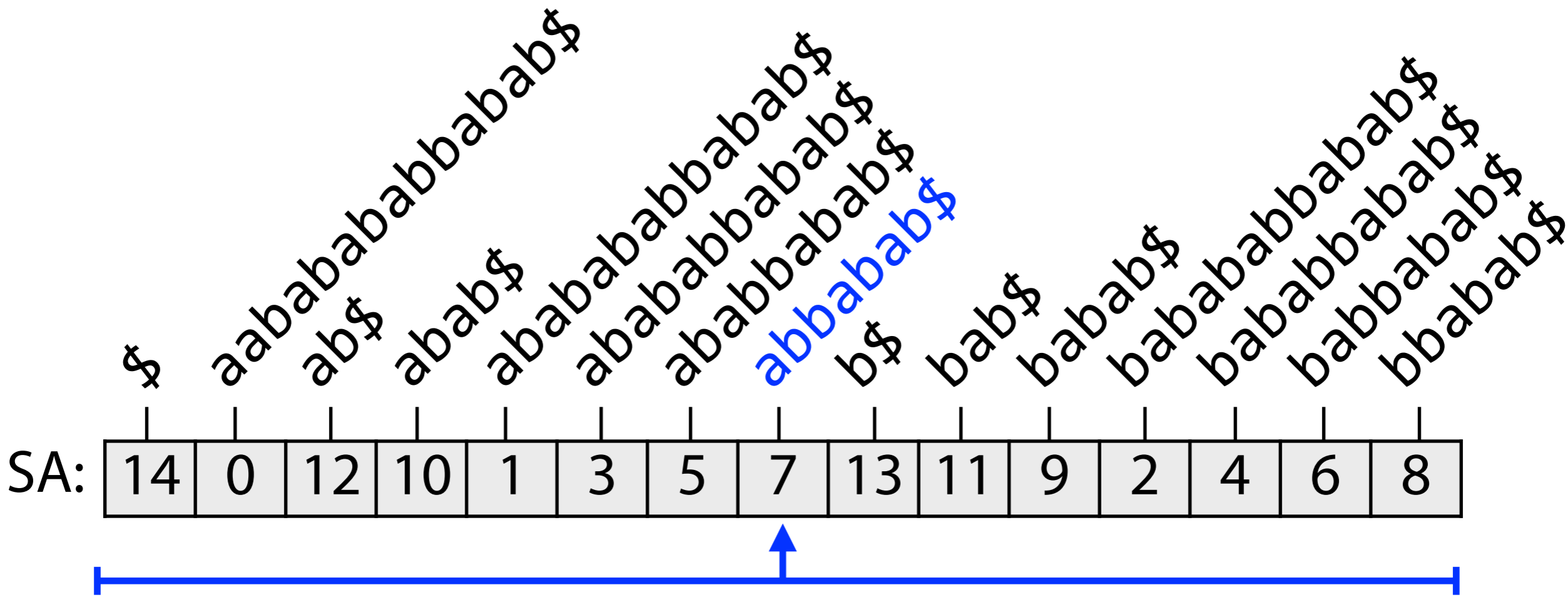


Query: a b a b a a

- - - - -

Pivot: a b b a b a b \$

# Suffix array



Query: a b a b a a

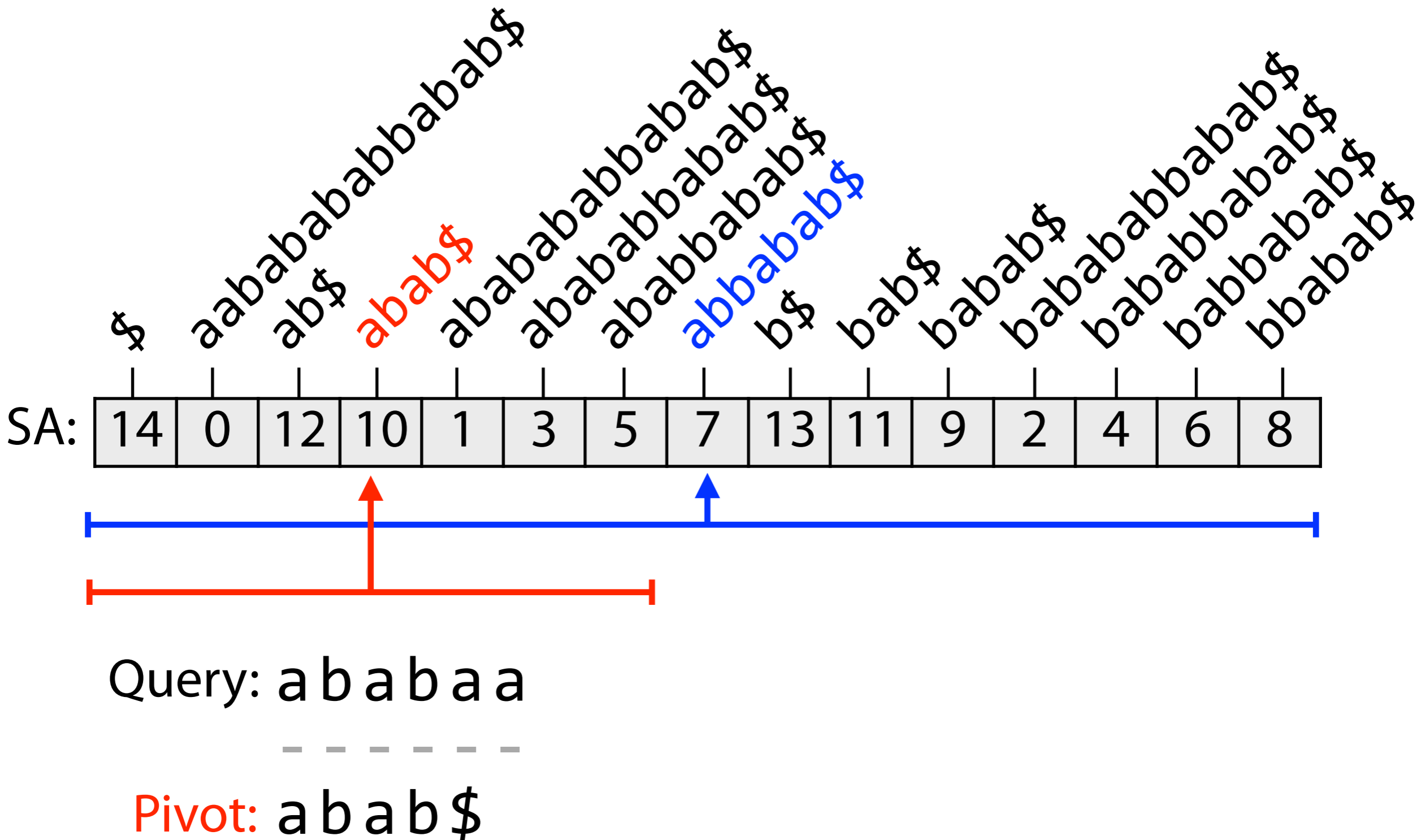
= = < - - -

Pivot: a b b a b a b \$

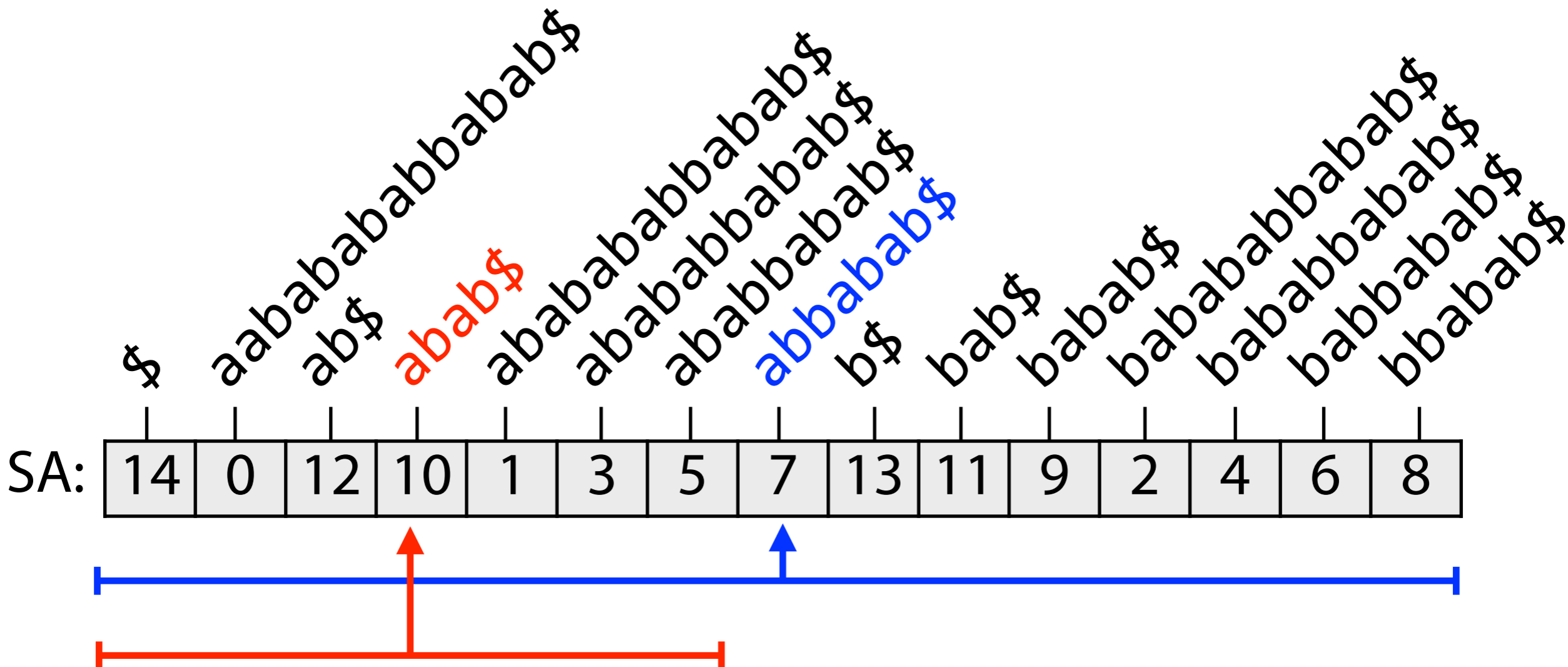
┌

Length of Common Prefix (LCP) = 2

# Suffix array

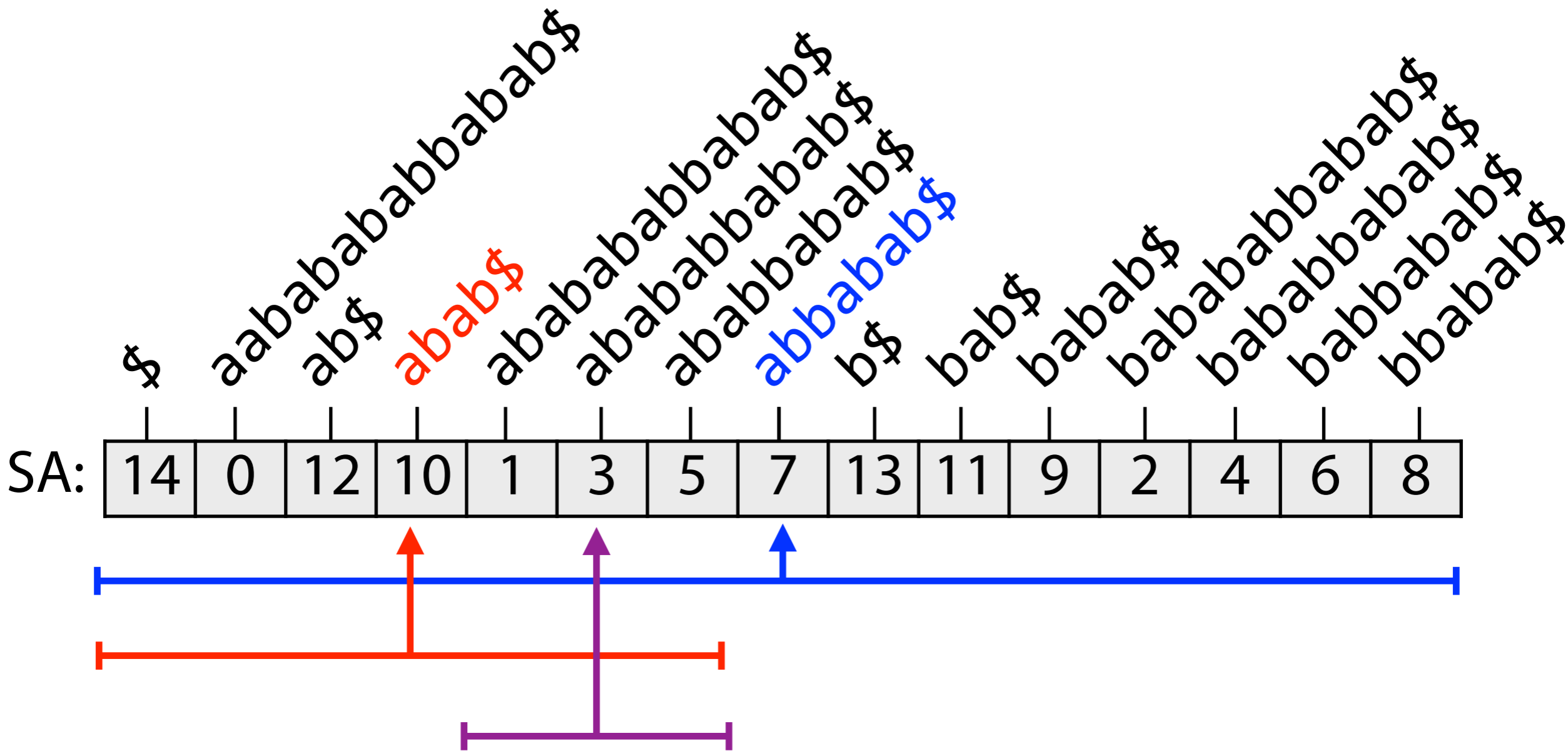


# Suffix array



Query: a b a b a a  
 = = = = > -  
 Pivot: a b a b \$  
 LCP = 4

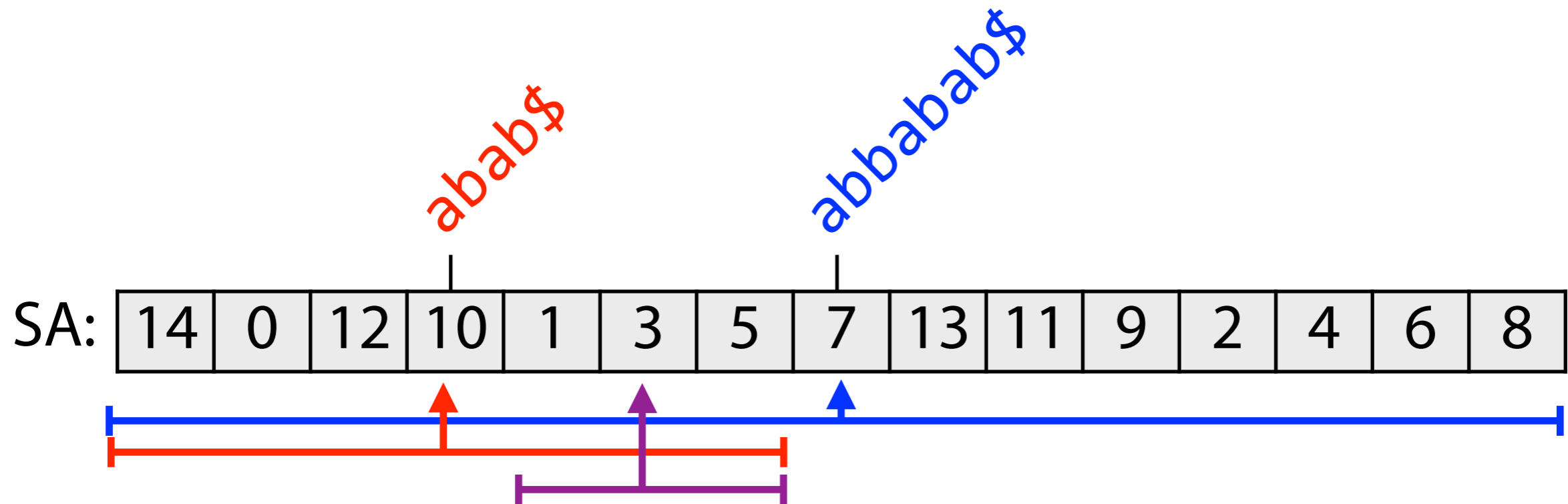
# Suffix array





# Suffix array

What did we learn about common prefixes in the last 2 rounds?



Query: a b a b a a

==== > -

Pivot: a b a b \$

-----

LCP = 4

Query: a b a b a a

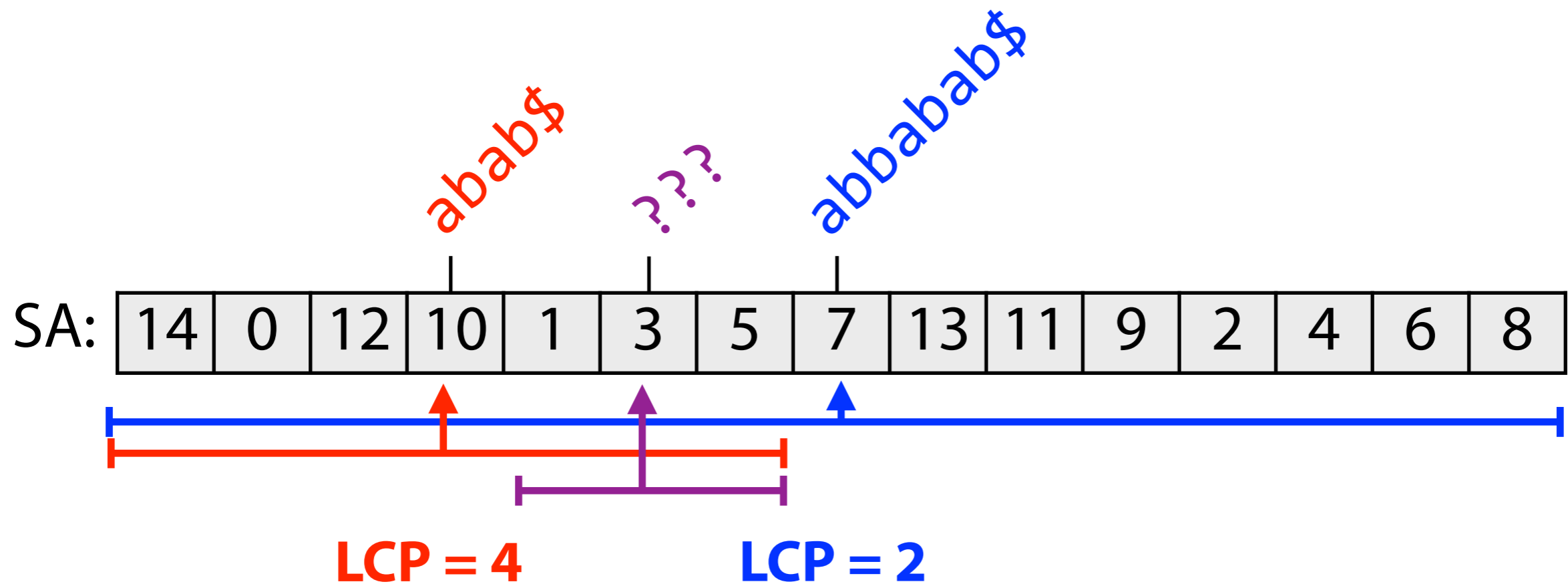
== < - - -

Pivot: a b b a b a b \$

-----

LCP = 2

# Suffix array



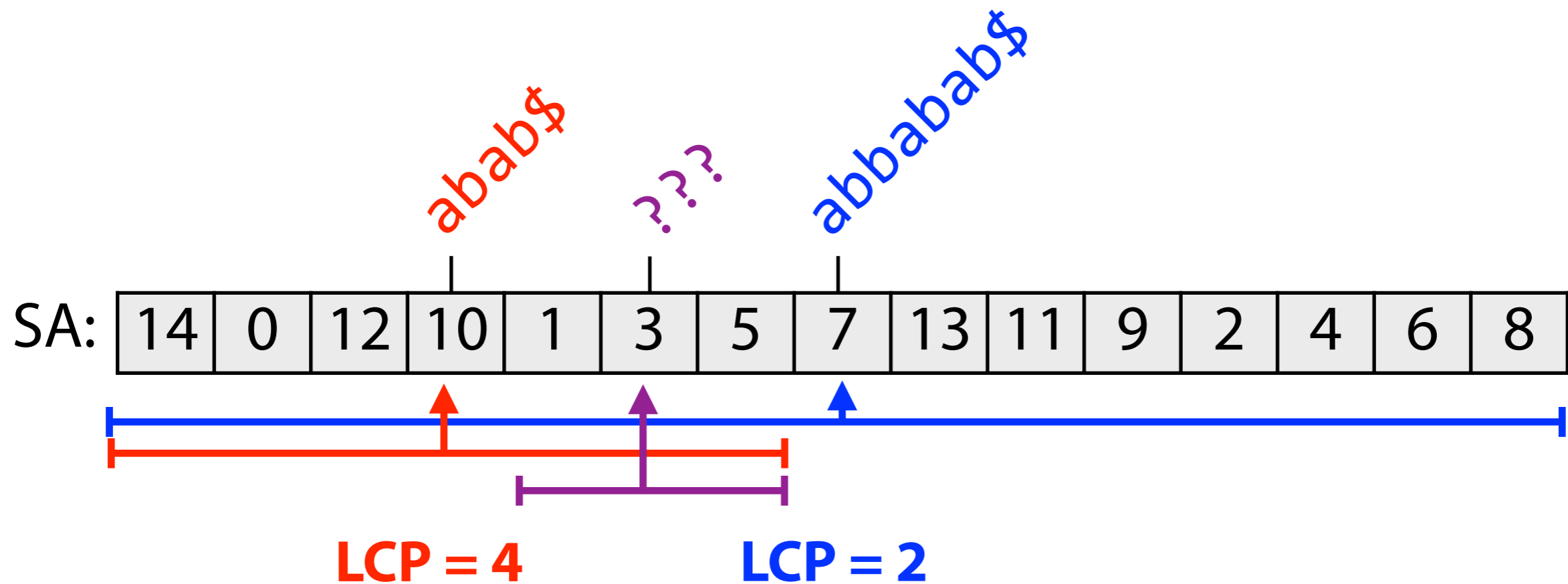
Combine facts:

LCP with **red** pivot is 4      LCP with **blue** pivot is 2

**Next** pivot is alphabetically between blue & red

LCP with next pivot is  $\geq$

# Suffix array



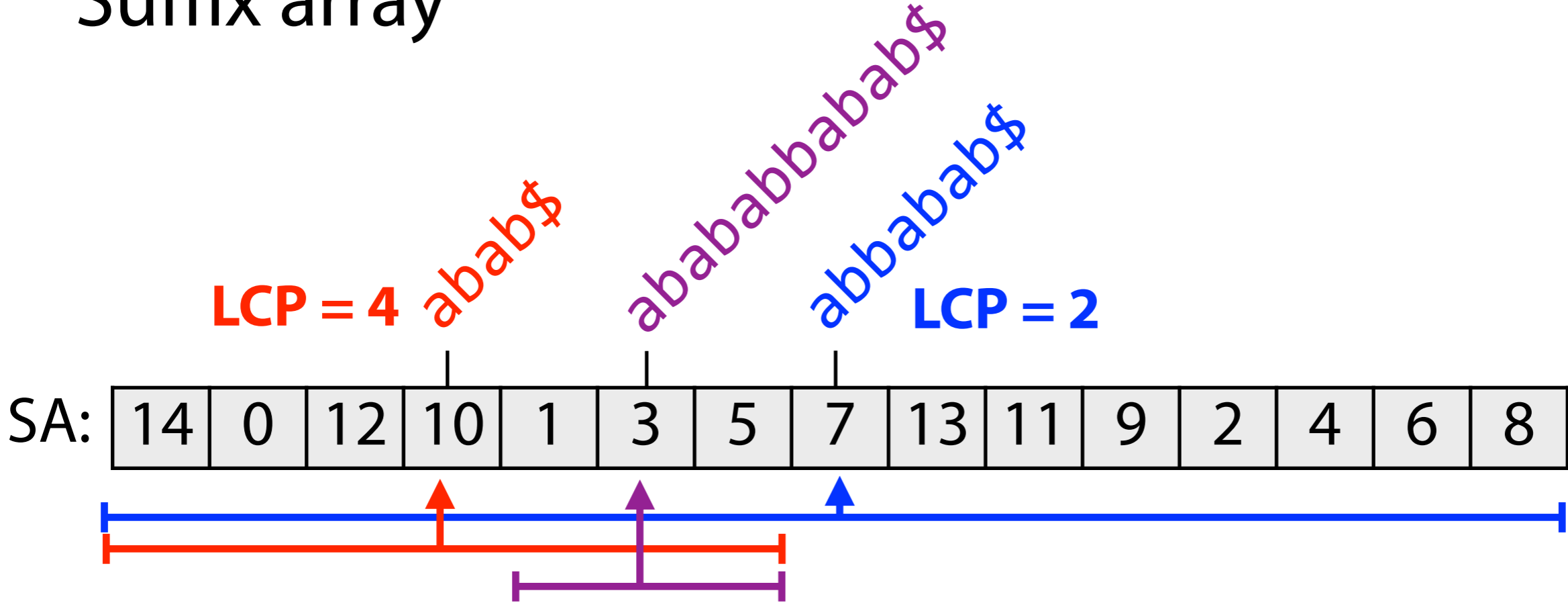
Combine facts:

LCP with **red** pivot is 4      LCP with **blue** pivot is 2

**Next** pivot is alphabetically between blue & red

LCP with next pivot is  $\geq \min(4, 2) = 2$

# Suffix array

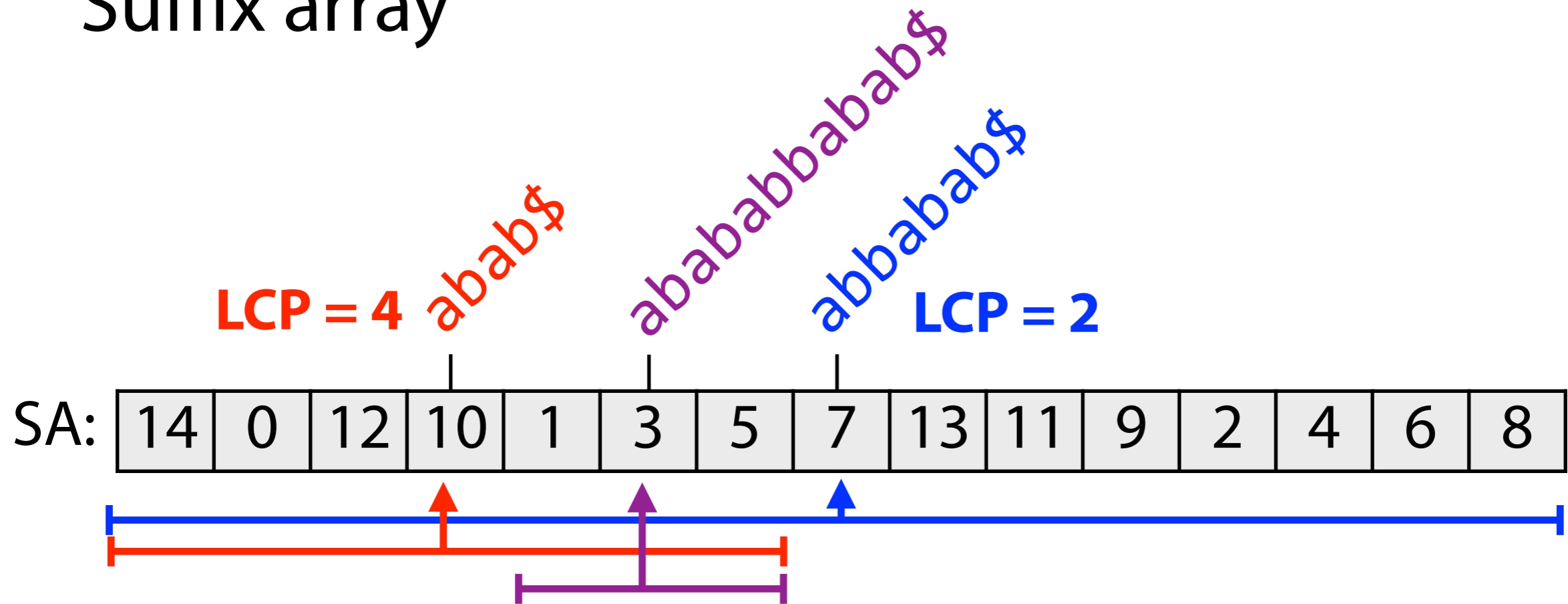


Query: a b a b a a

- - - - -

Pivot: a b a b a b b a b a b \$

# Suffix array



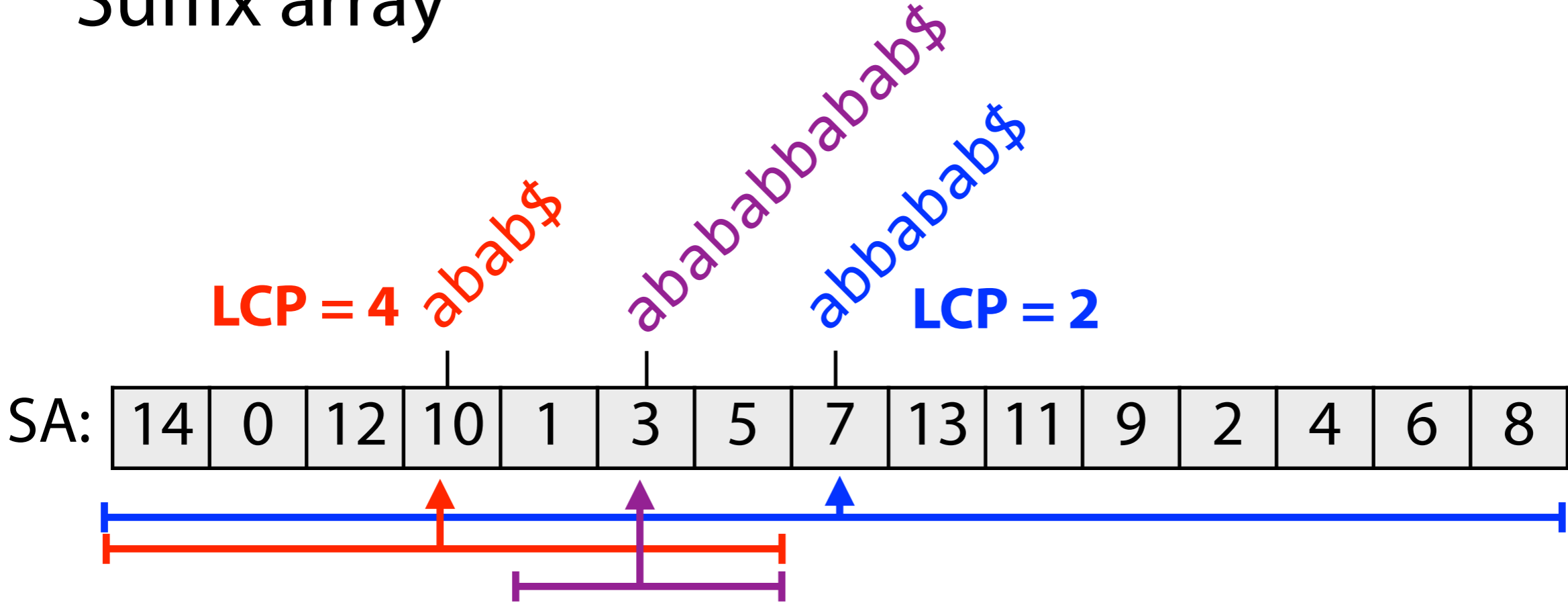
Query: a b a b a a

= = - - - -

Pivot: a b a b a b b a b a b \$

Skip!  
Skip!

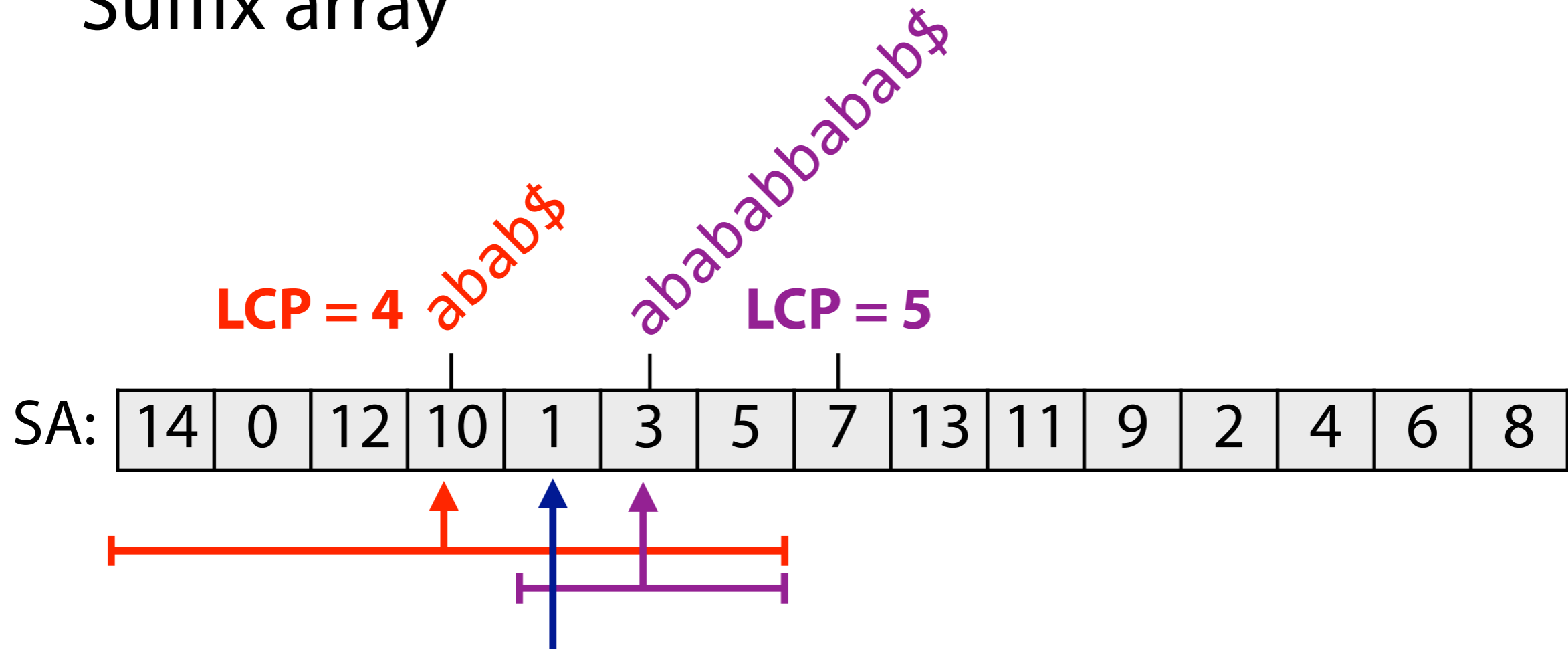
# Suffix array



Query: a b a b a a  
 = = = = = <

Pivot: a b a b a b b a b a b \$  
 Skip! Skip! LCP = 5

# Suffix array

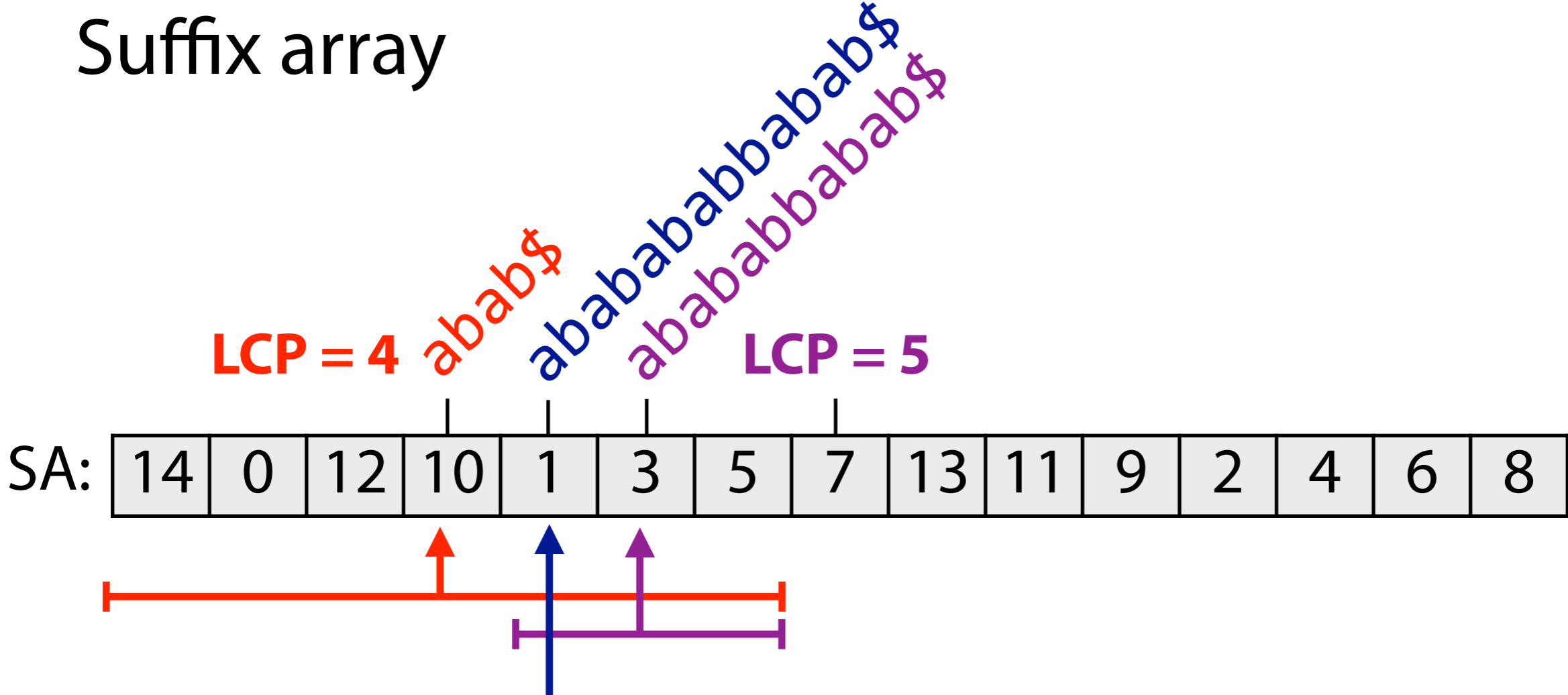


LCP with **red** pivot is 4      LCP with **plum** pivot is 5

**Next** pivot is alphabetically between red & plum

LCP with next pivot is  $\geq \min(4, 5) = 4$

# Suffix array



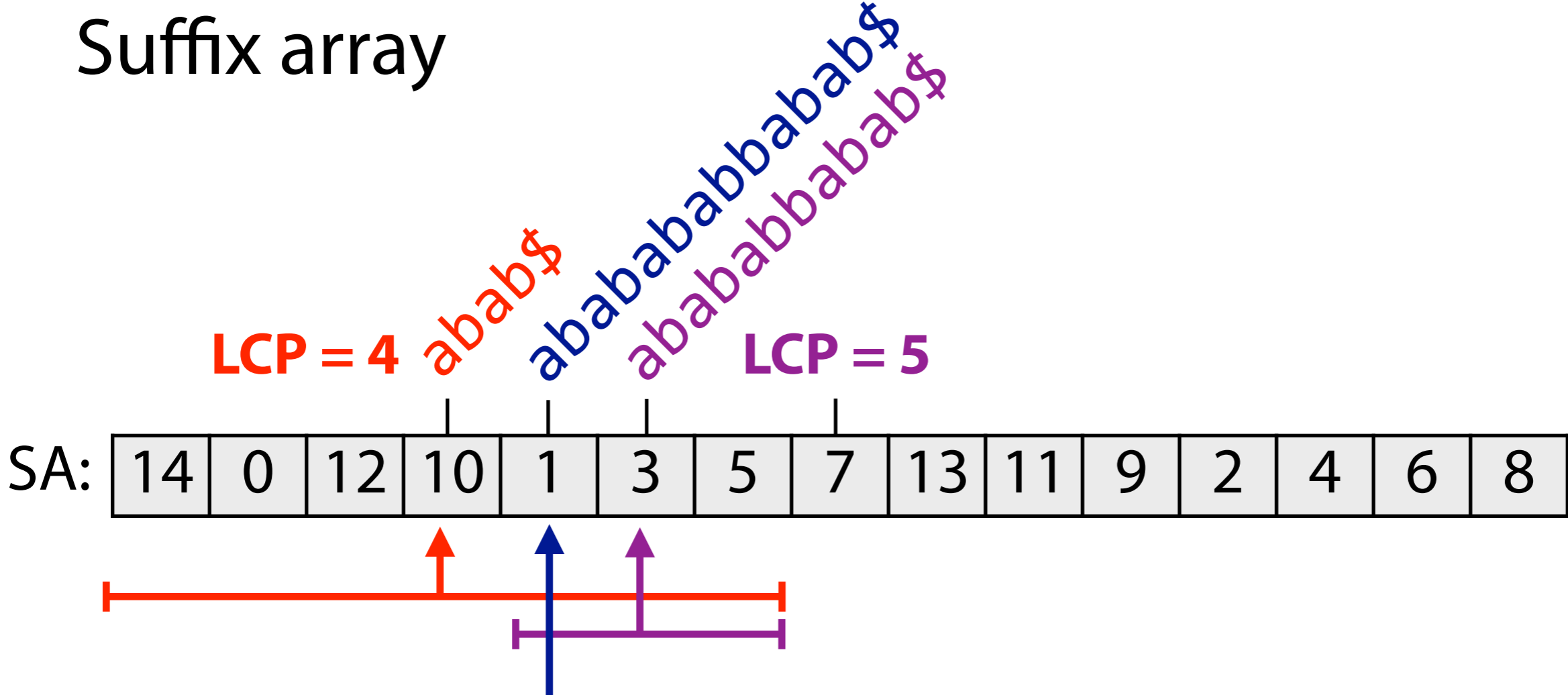
Query: a b a b a a

- - - - -

Pivot: a b a b a b b a b a b \$



# Suffix array

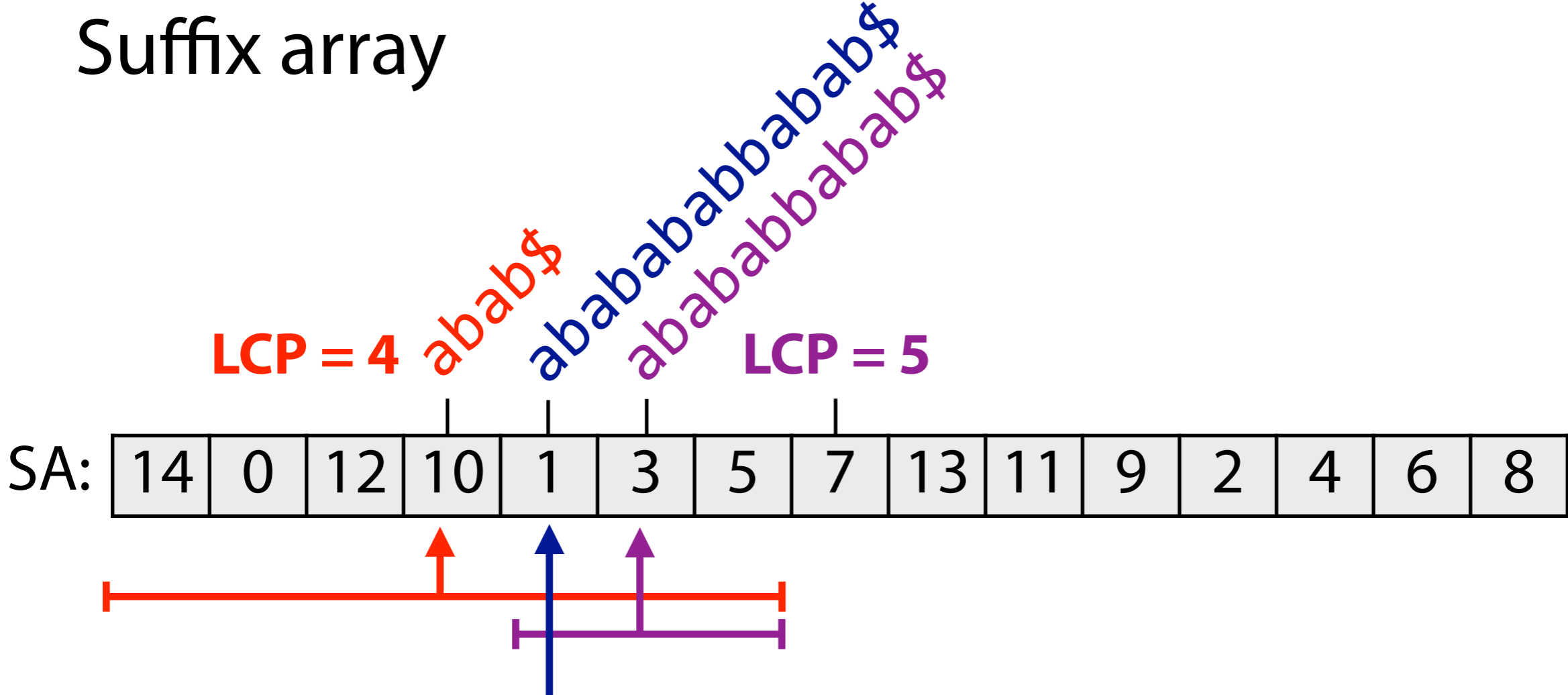


Query: a b a b a a  
 = = = = - -

Pivot: a b a b a b b a b a b \$

Skip!  
 Skip!  
 Skip!  
 Skip!

# Suffix array



Query: a b a b a a  
 = = = = = <

Pivot: a b a b a b a b b a b a b \$

Response is **No**: ababaa is not a substring

# Suffix array

## No skipping

a b a b a a  
== < - - -  
a b b a b a b \$

a b a b a a  
==== > -  
a b a b \$

a b a b a a  
===== <  
a b a b a b b a b a b \$

a b a b a a  
===== <  
a b a b a b a b b a b a b \$

## Min-LCP skipping

a b a b a a  
== < - - -  
a b b a b a b \$

a b a b a a  
==== > -  
a b a b \$

a b a b a a  
===== <  
a b a b a b b a b a b \$

a b a b a a  
===== <  
a b a b a b a b b a b a b \$