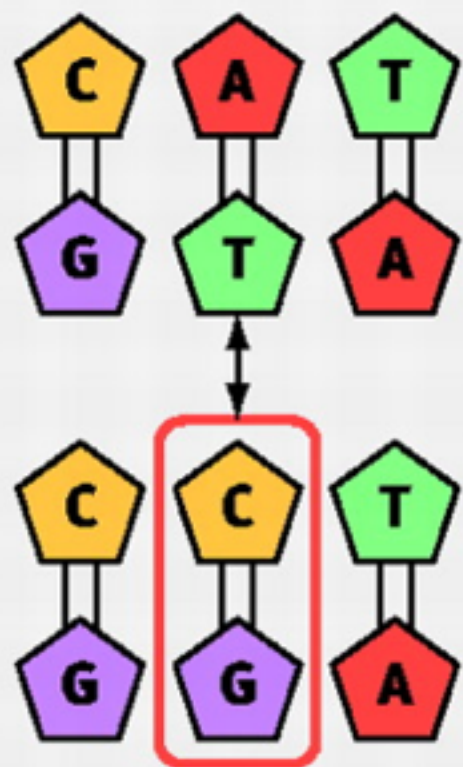
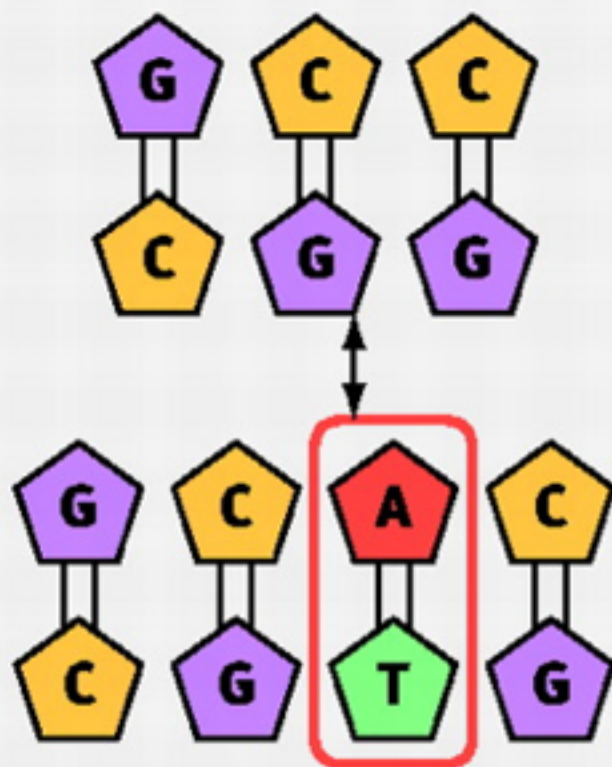


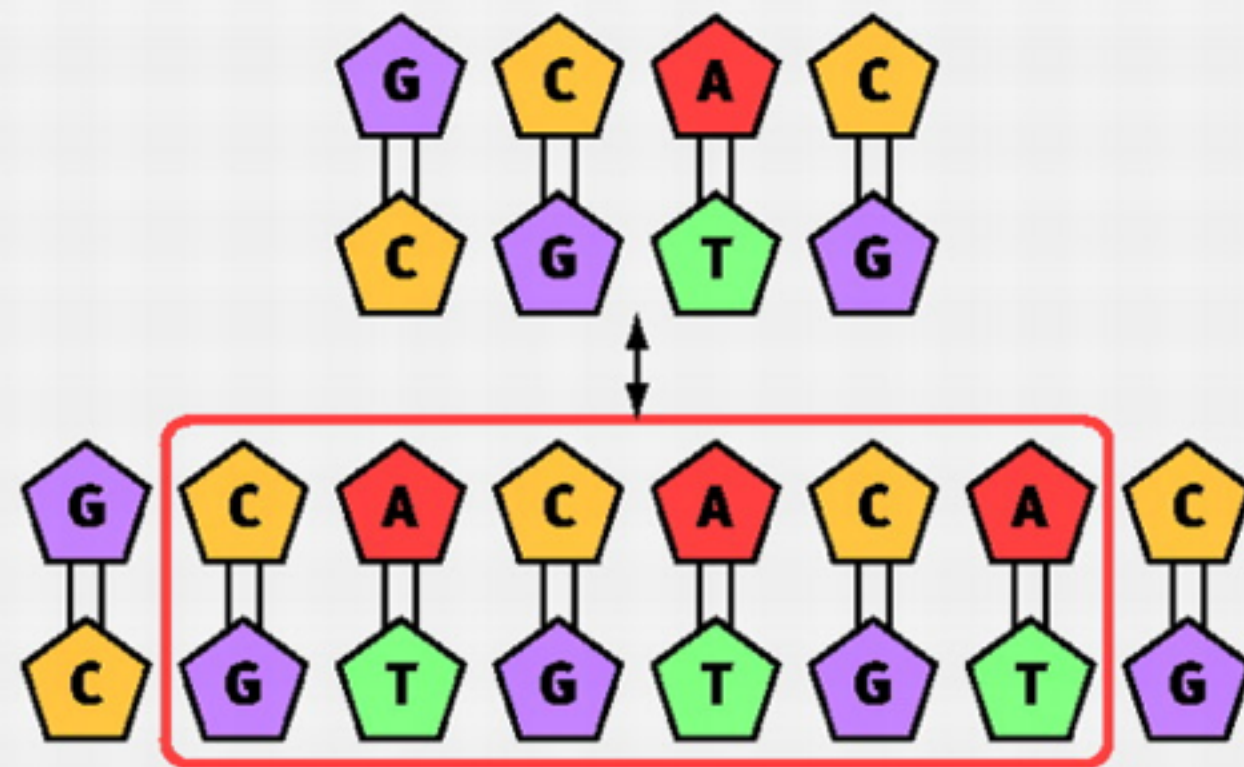
Single nucleotide polymorphism (SNP)



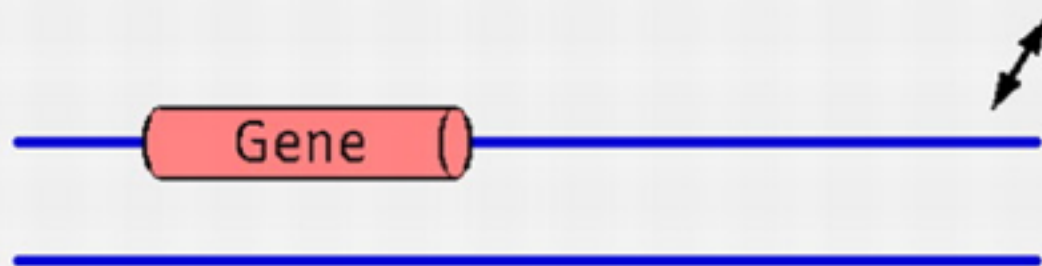
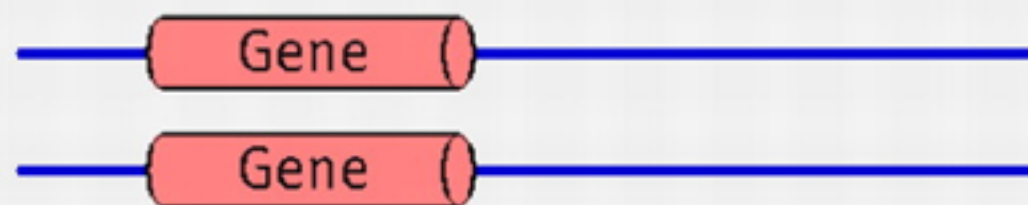
Insertion and deletion polymorphism (indel)



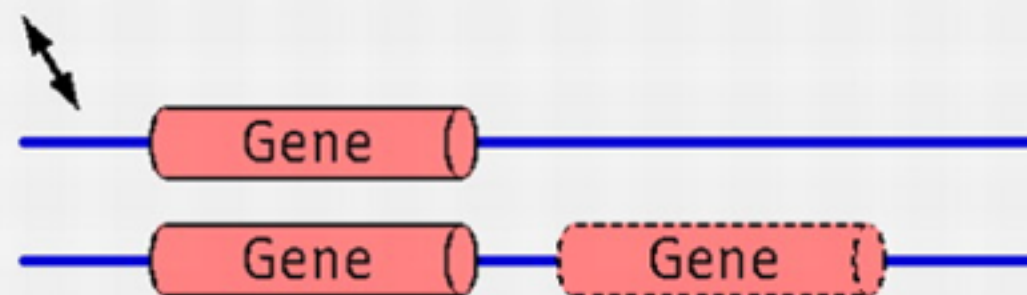
Nucleotide repeat polymorphism



Copy number variation



Deletion



Duplication

Геномные перестройки

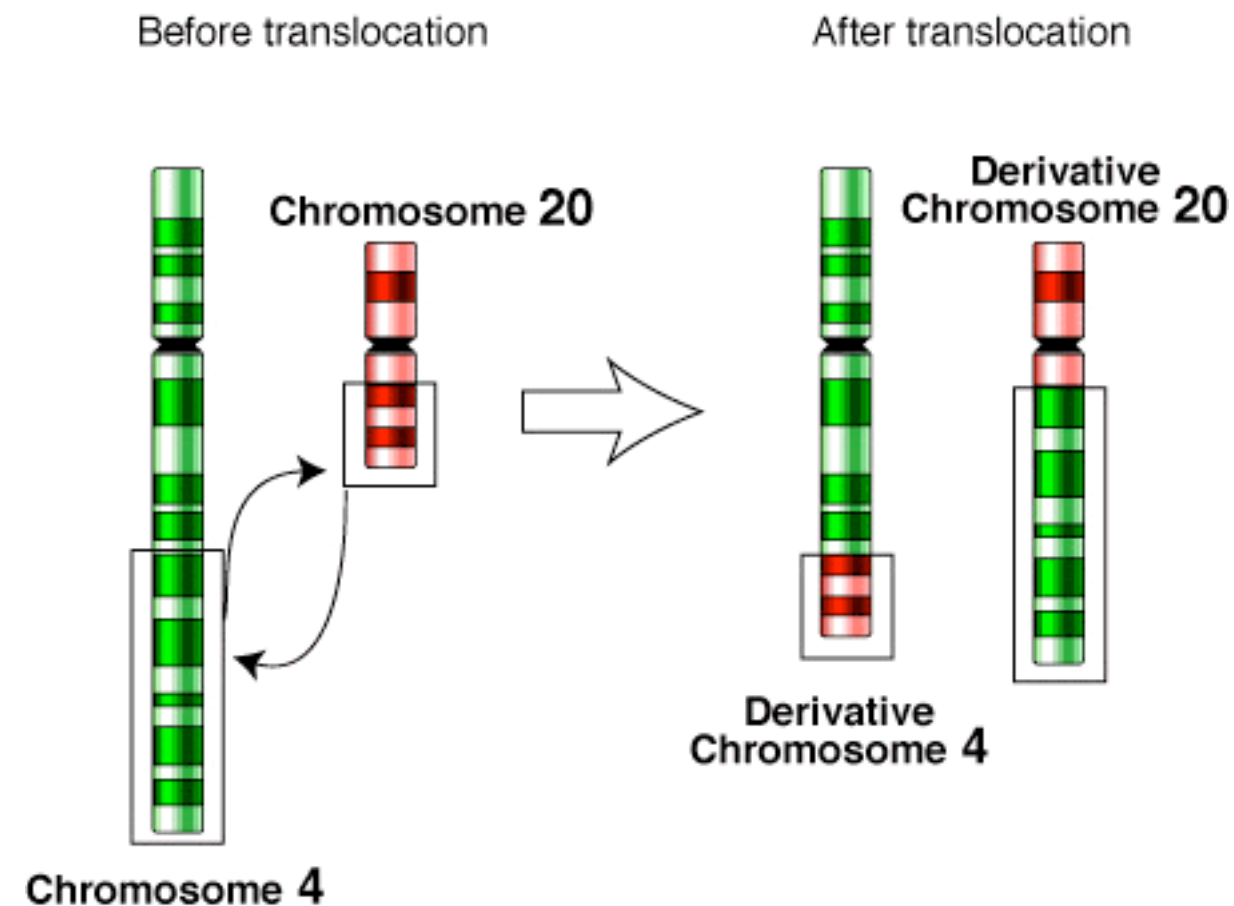
Genome Rearrangements:

реверсия (инверсия)

транслокация

слияние

расщепление



ОТВЕТЫ НА КВИЗ

1. Single Nucleotide Polymorphism, indels, короткие повторы, Copy Number Variations (дубликации, делеции), геномные перестройки (реверсии, транслокации, слияния, расщепления).
2. Геномные перестройки
3. $O(N^3N!)$
4. Понимание эволюции, функций, рака
5. International Union of Pure and Applied Chemistry
6. Локальное выравнивание.

Symbol	Description	Bases represented				
A	adenosine	A				1
C	cytidine		C			
G	guanosine			G		
T	thymidine				T	
U	uridine				U	
W	weak	A			T	2
S	strong		C	G		
M	amino	A	C			
K	keto			G	T	
R	purine	A		G		
Y	pyrimidine		C		T	3
B	not A (B comes after A)		C	G	T	
D	not C (D comes after C)	A		G	T	
H	not G (H comes after G)	A	C		T	
V	not T (V comes after T and U)	A	C	G		4
N or -	any base (not a gap)	A	C	G	T	

Сегодня

Наивный алгоритм

Кнут-Моррис-Пратт

Рабин-Карп

Бойер-Мурр

Неточный поиск с заменами

Задача

Поиск подстроки в строке:

Даны:

шаблон P длины N ,

текст T длины M .

Найти все позиции вхождения P в T .

Наивный алгоритм

АТСА

АТАТGAACTGAGATCAAT

Наивный алгоритм

АТСА

АТАТГААСТGAGATCAAT

Наивный алгоритм

ATCA

ATATGAAC TGAGATCAAT

Наивный алгоритм

АТСА

АТАТГААСТGAGATCAAT

Наивный алгоритм

АТСА

АТАТГААСТGAG**АТСА**АТ

Кнут—Моррис—Пратт

Префикс функция — длина наибольшего префикса строки, который не совпадает с этой строкой и одновременно является её суффиксом.

$S = \text{abacaba}$

$\pi(S) = 0010123$

Кнут—Моррис—Пратт

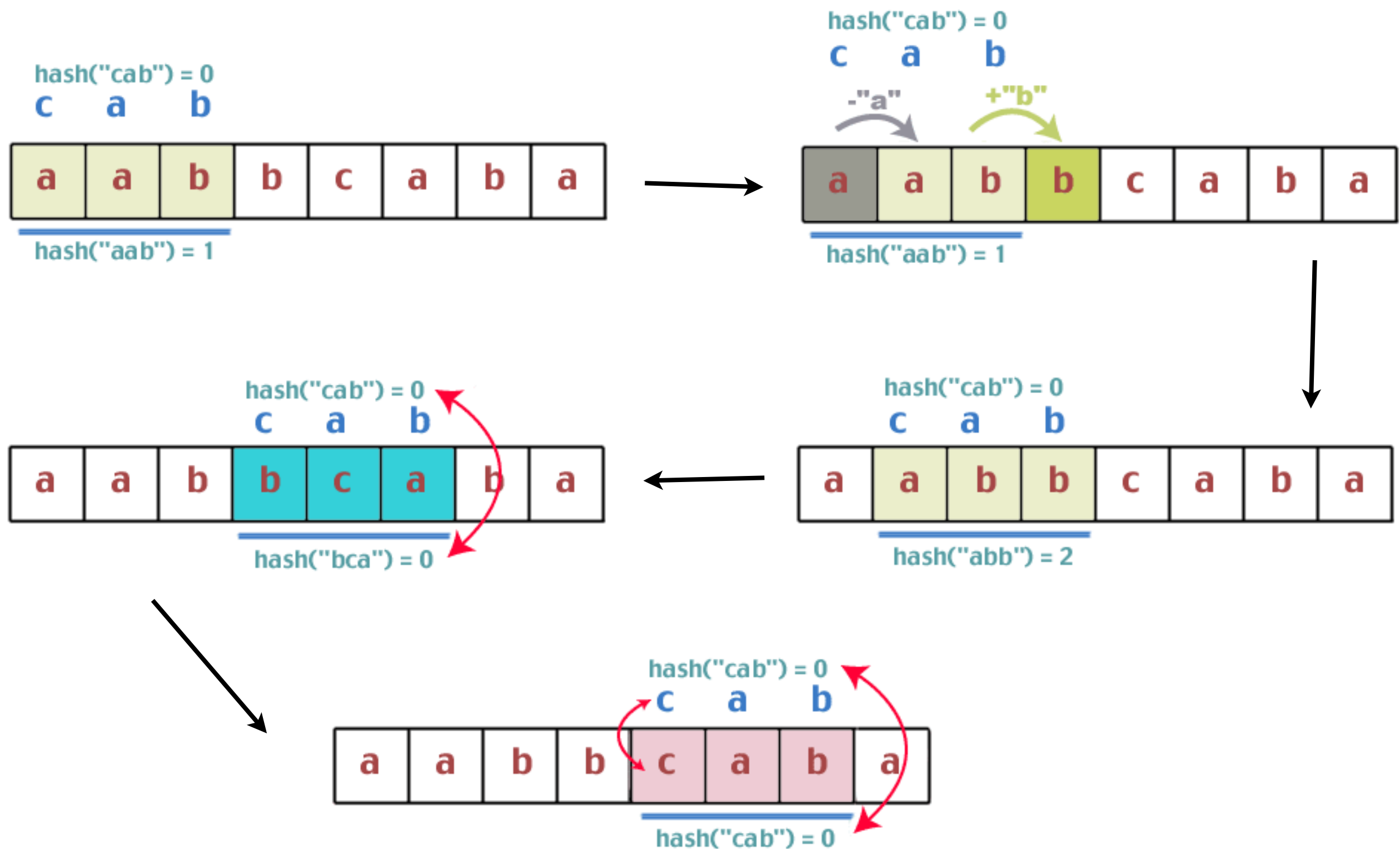
Префикс функция — длина наибольшего префикса строки, который не совпадает с этой строкой и одновременно является её суффиксом.

$S = \text{abacaba}$

$\pi(S) = 0010123$

Построим префикс функцию от P\$T...

Рабин–Карп



Бойер–Мурр

Сканирование слева направо.

Сравнение справа налево.

Бойер–Мурр

Сканирование слева направо.

Сравнение справа налево.

The Bad Character Rule

(Эвристика стоп-символа)

The Good Suffix Rule

(Эвристика совпавшего суффикса)

The Bad Character Rule

A N P A **N** M A N A M -
- N **N** A A M A N - - -
- - - N **N** A A M A N -

The Bad Character Rule

A N P A N M A N A M -
- N N A A M A N - - -
- - - N N A A M A N -

Препроцессинг: таблица стоп-символов.

The Bad Character Rule

A N P A **N** M A N A M -
- N **N** A A M A N - - -
- - - N **N** A A M A N -

Препроцессинг: таблица стоп-символов.

N	A	M
1	5	4

Бойер–Мур–Хорспул

Изменённая эвристика стоп-символа.

A	N	P	A	N	M	A	N	A	M	-
-	N	N	A	A	M	A	N	-	-	-
-	-	-	-	-	-	N	N	A	A	M

The Good Suffix Rule

M A N P A N A M A N A P -
A N A M P N A M - - - - -
- - - - A N A M P N A M -

The Good Suffix Rule

M A N P A N A M A N A P -
A N A M P N A M - - - - -
- - - - A N A M P N A M -

Препроцессинг: таблица суффиксов

The Good Suffix Rule

M A N P A N A M A N A P -
A N A M P N A M - - - - -
- - - - A N A M P N A M -

Препроцессинг: таблица суффиксов

abacaba
4444462