

Kako locirati mutacije koje izazivaju bolesti?

Uparivanje šablona

*Bioinformatics Algorithms:
an Active Learning Approach*

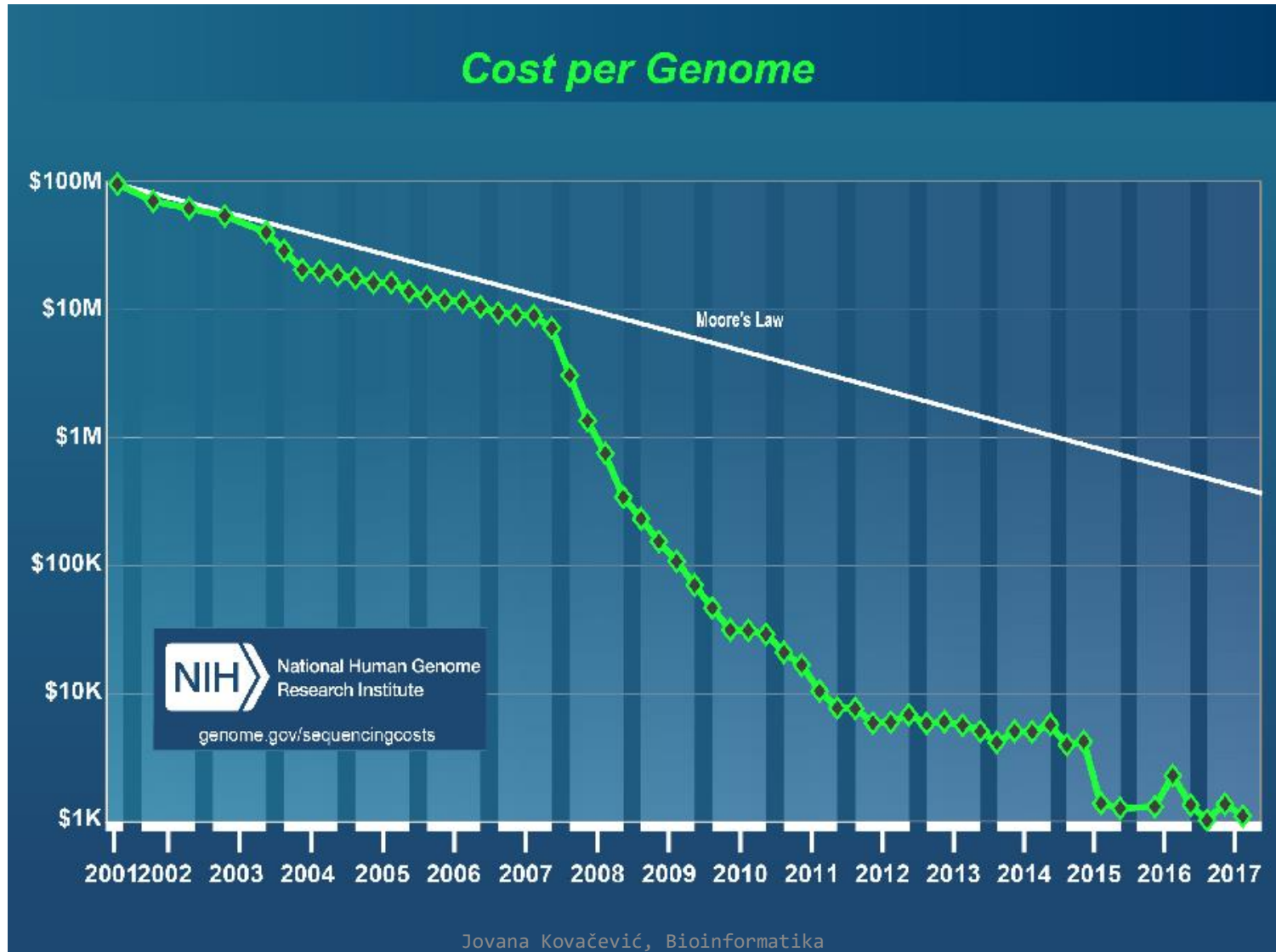
~Poglavlje 9~

Pregled

- **Mapiranje očitavanja**

- ...

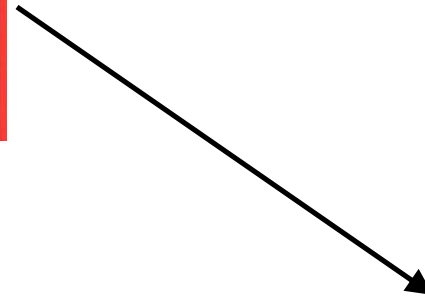
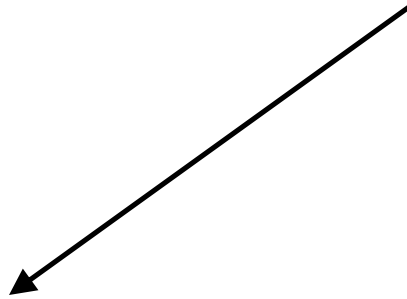
Troškovi sekvencioniranja



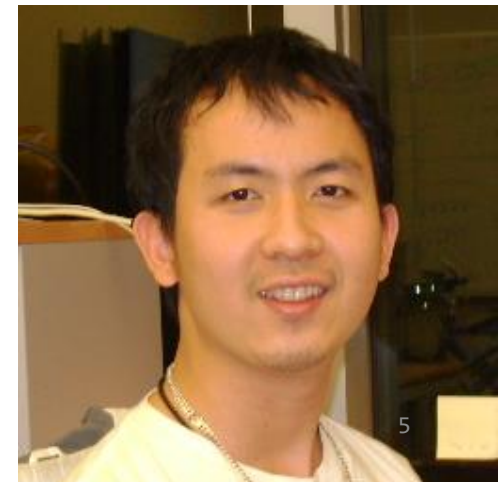
Od genoma vrsta do personalnih genoma



Od genoma vrsta do personalnih genoma



Jovana Kovačević, Bioinformatika



Od genoma vrsta do personalnih genoma

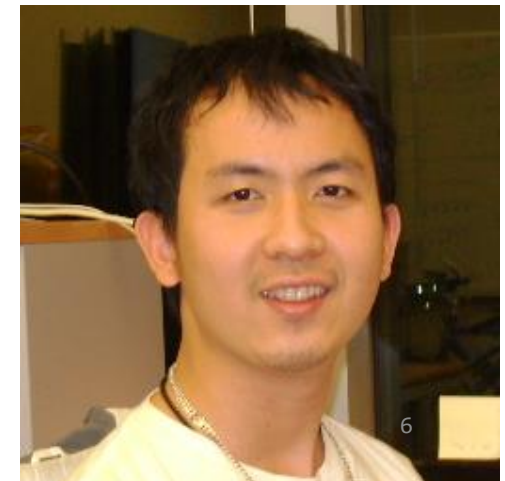
CTGATGATGGACTACGCTA
CTACTGCTAGCTGTATTAC
GATCAGCTAC**C**ACATCGTA
GCTACGATGCATTAGCAAG
CTATCGATCGATCGATCGA
TTATCTACGATCGATCGAT
CGATCACTATAACGAGCTAC
TACGTACGTACGATCGCG**G**
GACTATTATCGACTACAGA
TAAAACATGCTAGTACAAC
AGTATACATAGCTGCGGGA
TACGATTAGCTAATAGCTG
ACGATATCCGAT

CTGATGATGGACTACGCTA
CTACTGCTAGCTGTATTAC
GATCAGCTAC**A**ACATCGTA
GCTACGATGCATTAGCAAG
CTATCGATCGATCGATCGA
TTATCTACGATCGATCGAT
CGATCACTATAACGAGCTAC
TACGTACGTACGATCGCG**T**
GACTATTATCGACTACAGA
TGAAACATGCTAGTACAAC
AGTATACATAGCTGCGGGA
TACGATTAGCTAATAGCTG
ACGATATCCGAT

CTGATGATGGACTACGCTA
CTACTGCTAGCTGTATTAC
GATCAGCTAC**T**ACATCGTA
GCTACGATGCATTAGCAAG
CTATCGATCGATCGATCGA
TTATCTACGATCGATCGAT
CGATCACTATAACGAGCTAC
TACGTACGTACGATCGCG**A**
GACTATTATCGACTACAGA
TCAAACATGCTAGTACAAC
AGTATACATAGCTGCGGGA
TACGATTAGCTAATAGCTG
ACGATATCCGAT



Jovana Kovacević, Bioinformatika



Referentni i individualni genom

- **Referentni genom:** genom nastao na osnovu genoma 13 različitih pojedinaca.
- U proseku, razlika između individualnog i referentnog genoma je u oko 3 miliona mutacija
- **Pitanje:** Kako možemo efikasno sastaviti individualne genome koristeći referentni?

CTGATGATGGACTACGCTACTACTGCTAGCTGTAT

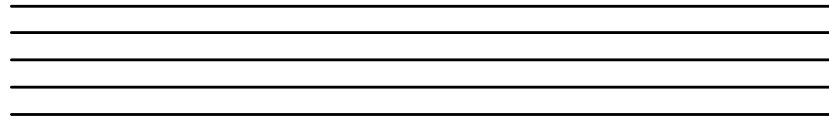
Individualni

CTGAGGATGGACTACGCTACTACTGATAGCTGTT

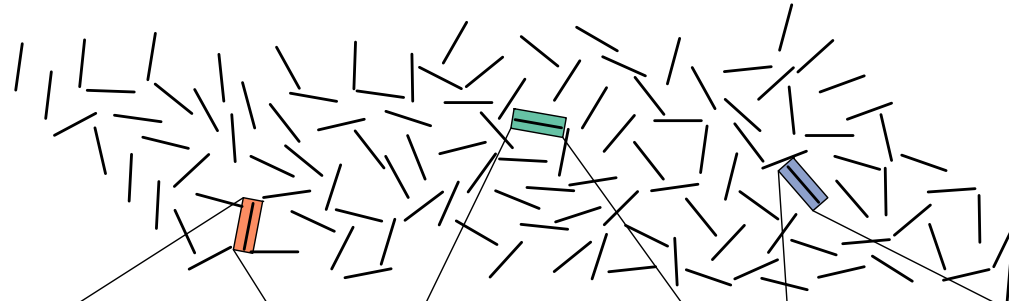
Referentni

Zašto ne bismo radili asembliranje?

Multiple copies of a genome



Shatter the genome into reads



Sequence the reads

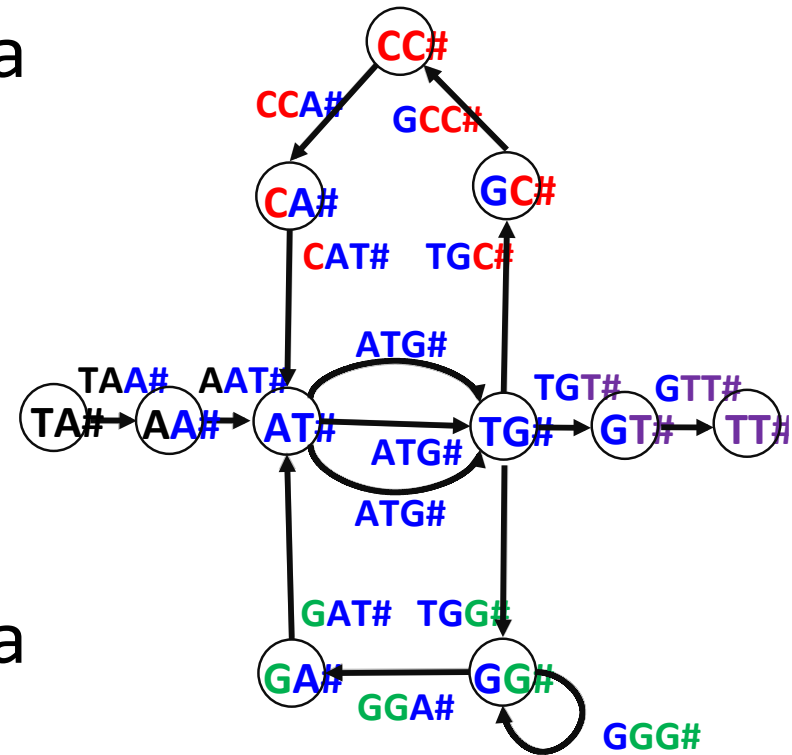


Assemble the genome with overlapping reads



Zašto ne bismo radili asembliranje?

- Konstrukcija de Brojnovog grafa zahteva mnogo memorije.
- Ideja: koristimo postojeću strukturu referentnog genoma kao pomoć u sekvencioniranju genoma pacijenta.



Mapiranje očitavanja

- **Mapiranje očitavanja:** odrediti pozicije u referentnom genomu sa kojima svako očitavanje ima visoku sličnost

CTGAGGATGGACTACGCTACTACTGATAGCTGTTT
GAGGA C**C**ACG TGA-A

Referentni
genom

Očitavanja
individualnog
genoma

Egzaktno uparivanje šablona

- Pitanje: gde se očitavanja *egzaktno* poklapaju sa referentnim genomom?
- **Problem jednostrukog uparivanja šablona:**
 - **Ulaz:** Niske *Pattern* i *Genome*.
 - **Izlaz:** Sve pozicije u niski *Genome* gde se niska *Pattern* pojavljuje kao podniska.

Egzaktno uparivanje šablona

- Pitanje: gde se očitavanja *egzaktno* poklapaju sa referentnim genomom?
- Problem **višestrukog** uparivanja šablona:
 - Ulaz: Kolekcija niski *Patterns* i niska *Genome*.
 - Izlaz: Sve pozicije u niski *Genome* gde se niske iz kolekcije *Patterns* pojavljuju kao podniske.

Rešavanje problema grubom silom

p a n a m a b a n a n a s

*Genome
Pattern*

Rešavanje problema grubom silom

p a n a m a b a n a n a s
n a n a

*Genome
Pattern*

Rešavanje problema grubom silom

p a n a m a b a n a n a s
n a n a

Genome
Pattern

Rešavanje problema grubom silom

p a n a m a b a n a n a s
n a n a

*Genome
Pattern*

Rešavanje problema grubom silom

p a **n** a m a b a n a n a s
 n a n a

*Genome
Pattern*

Rešavanje problema grubom silom

p a **n a** m a b a n a n a s
n a n a

*Genome
Pattern*

Rešavanje problema grubom silom

p a n a m a b a n a n a s
n a n a

*Genome
Pattern*

Rešavanje problema grubom silom

p a n **a** m a b a n a n a s
n a n a

*Genome
Pattern*

Rešavanje problema grubom silom

p a n a **m** a b a n a n a s
 n a n a

*Genome
Pattern*

Rešavanje problema grubom silom

p a n a m **a** b a n a n a s
 n a n a

Genome

Pattern

Rešavanje problema grubom silom

p a n a m a **b** a n a n a s
 n a n a

Genome

Pattern

Rešavanje problema grubom silom

p a n a m a b **a** n a n a s
n a n a

Genome

Pattern

Rešavanje problema grubom silom

p a n a m a b a **n** a n a s
 n a n a

Genome

Pattern

Rešavanje problema grubom silom

p a n a m a b a **n a** n a s
 n a n a

Genome

Pattern

Rešavanje problema grubom silom

p a n a m a b a **n a n** a s
 n a n a

Genome

Pattern

Rešavanje problema grubom silom

p a n a m a b a **n a n a** s
n a n a

Genome

Pattern

Rešavanje problema grubom silom

p a n a m a b a n **a** n a s
n a n a

Genome

Pattern

Rešavanje problema grubom silom

- Vreme izvršavanja:
 - Jednostruko *Pattern*: $O(|Genome| * |Pattern|)$
 - Višestruko *Patterns*: $O(|Genome| * |Patterns|)$
 - $|Patterns|$ = suma dužina *Patterns*
- Genomi mogu biti veoma dugi, na primer humani referentni genom zauzima 3GB a očitavanja individualnog genoma mogu dostići do 1TB

Pregled

- Mapiranje očitavanja
- **Sufiksna stabla**
- ...

Genome

p a n a m a b a n a n a s

Sufiksi

panamabananas\$

anamabananas\$

namabananas\$

amabananas\$

mabananas\$

abananas\$

bananas\$

ananas\$

nanas\$

anas\$

nas\$

as\$

s\$

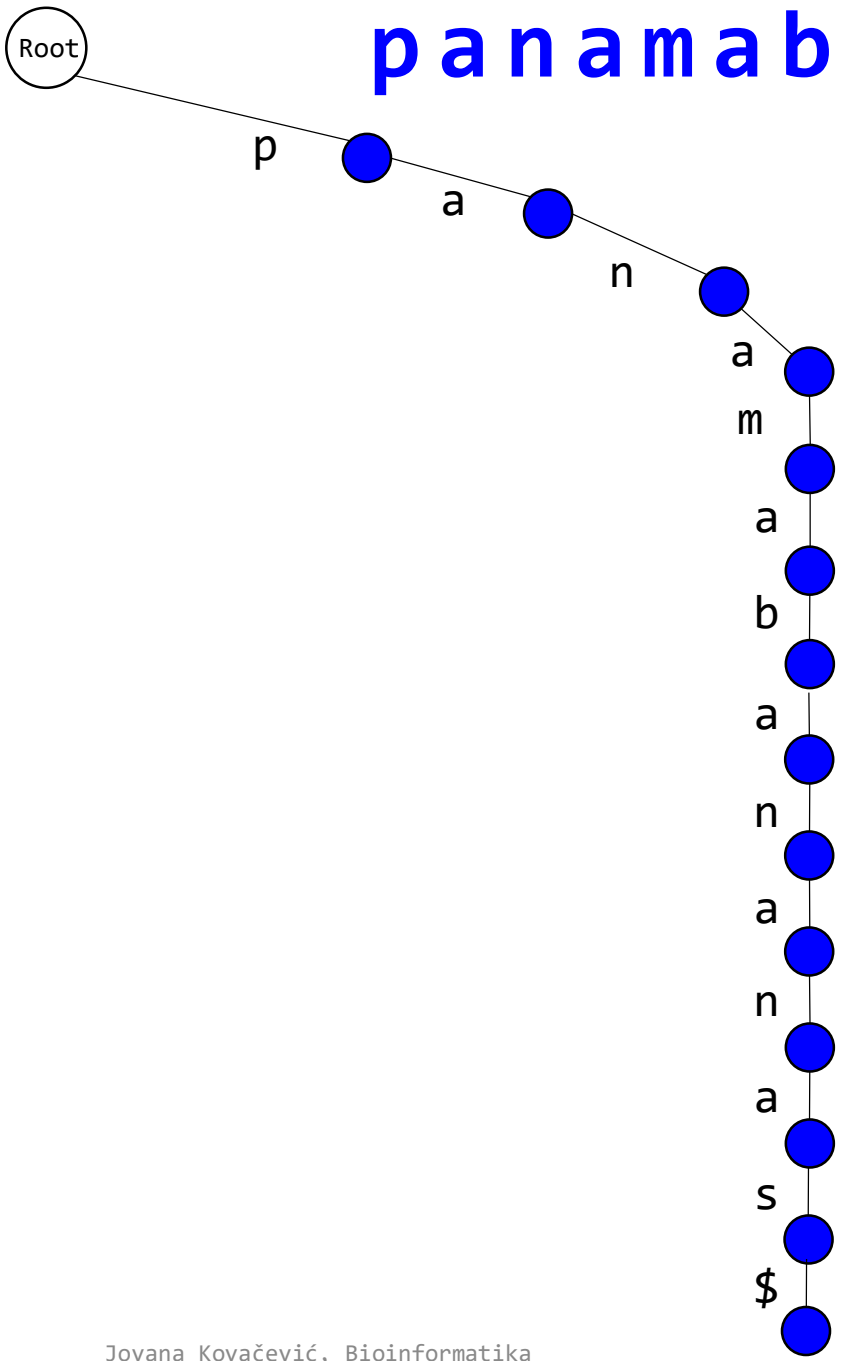
\$

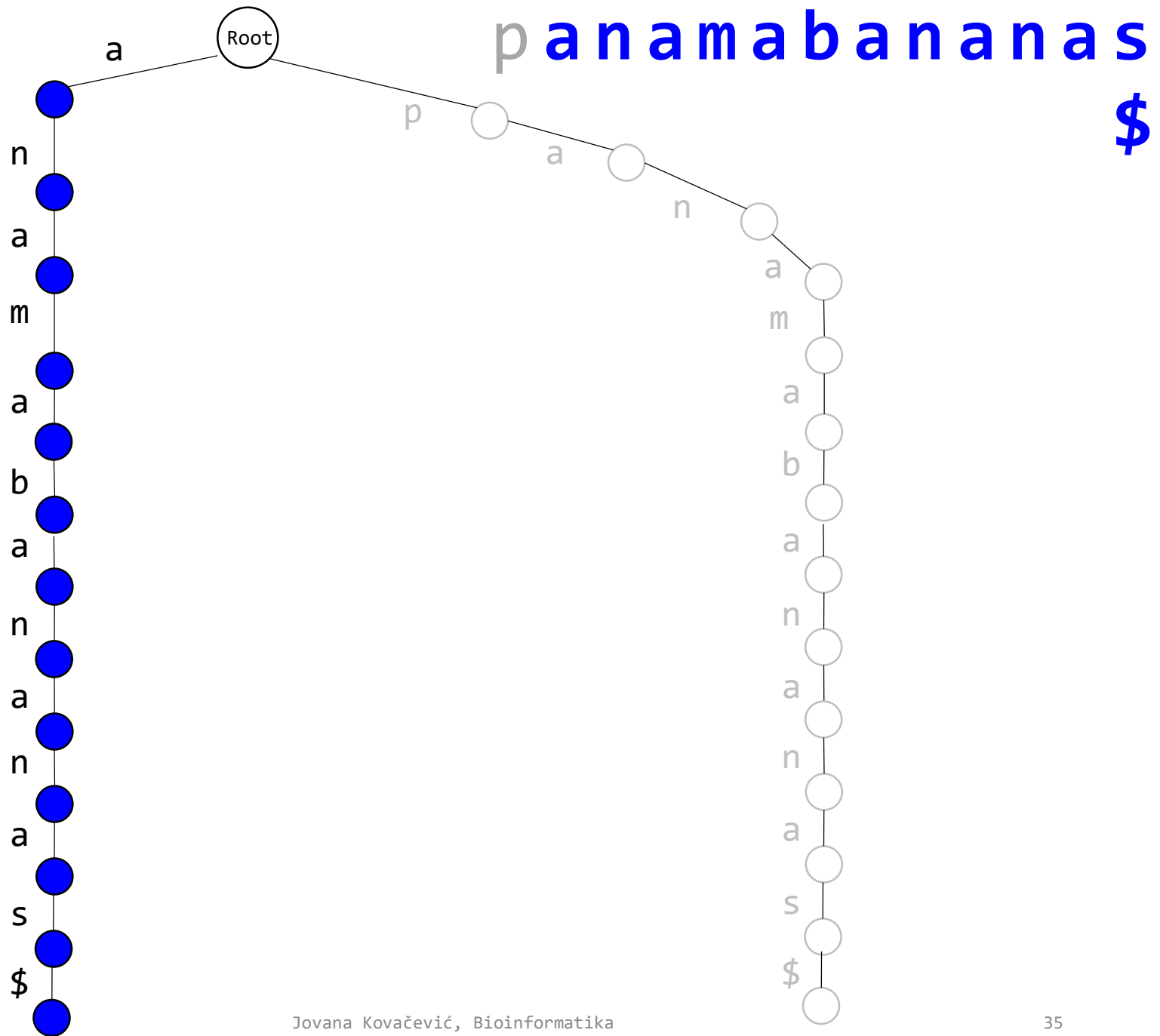
Root

panamabananas
\$

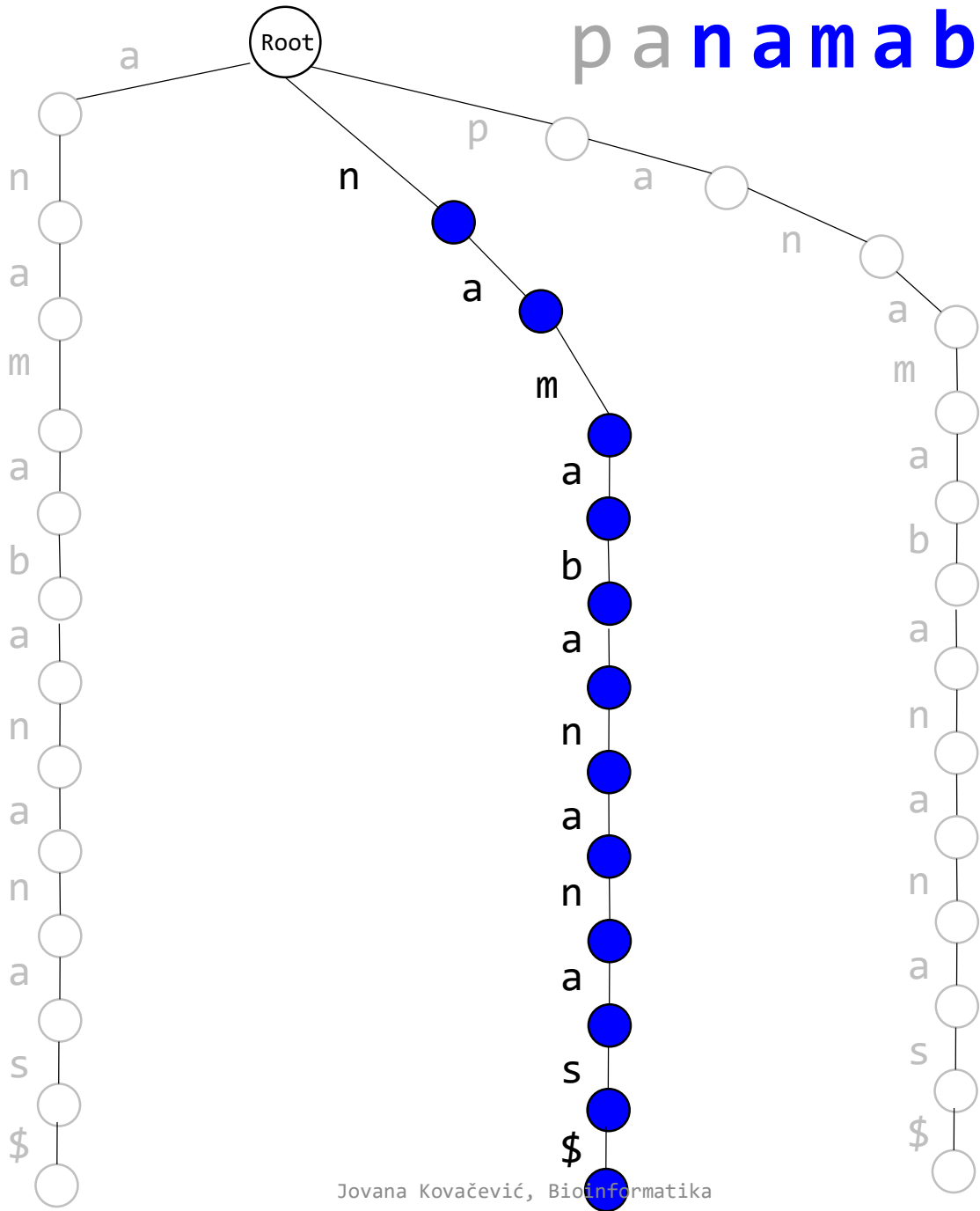
panamabananas

\$

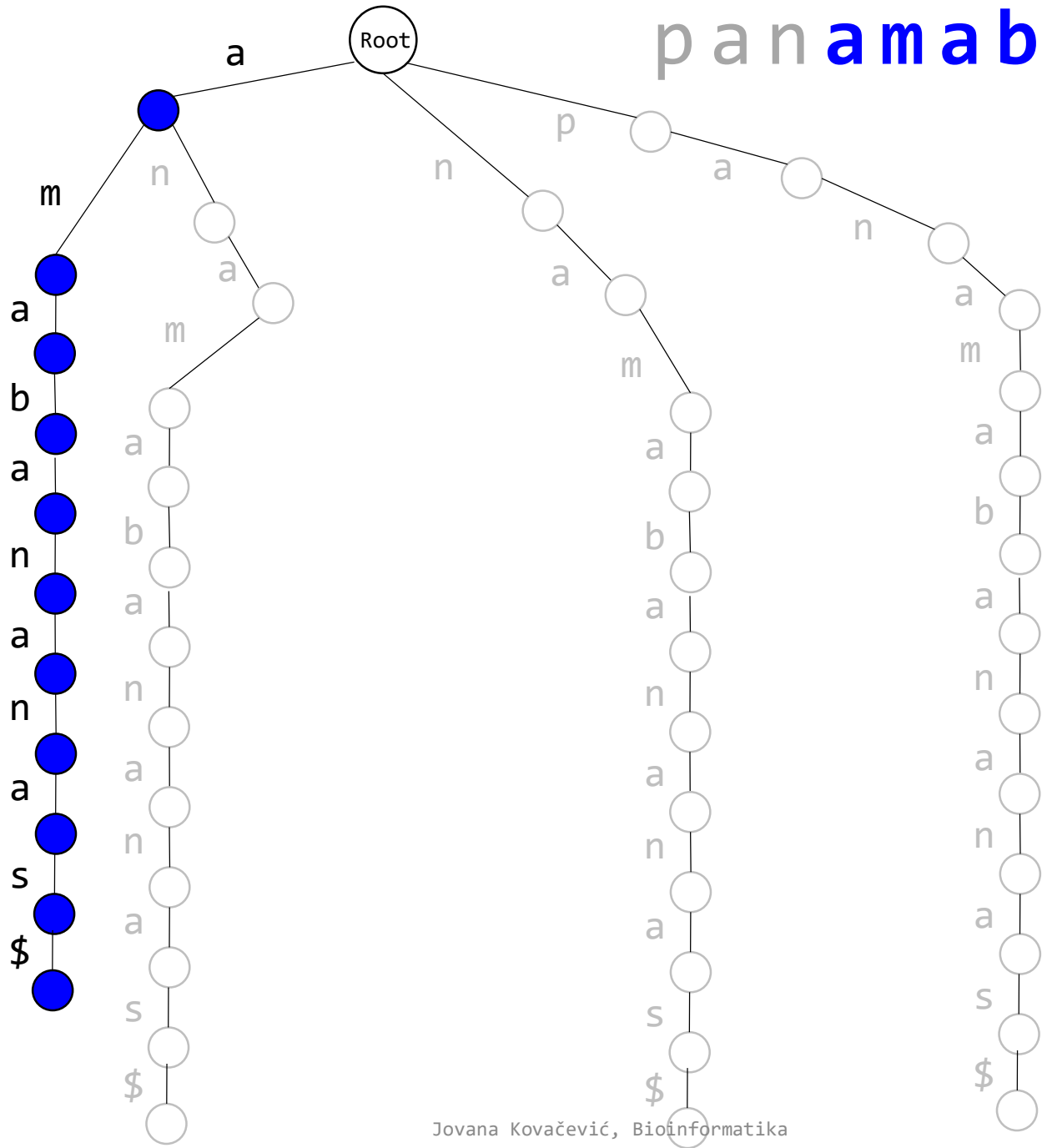




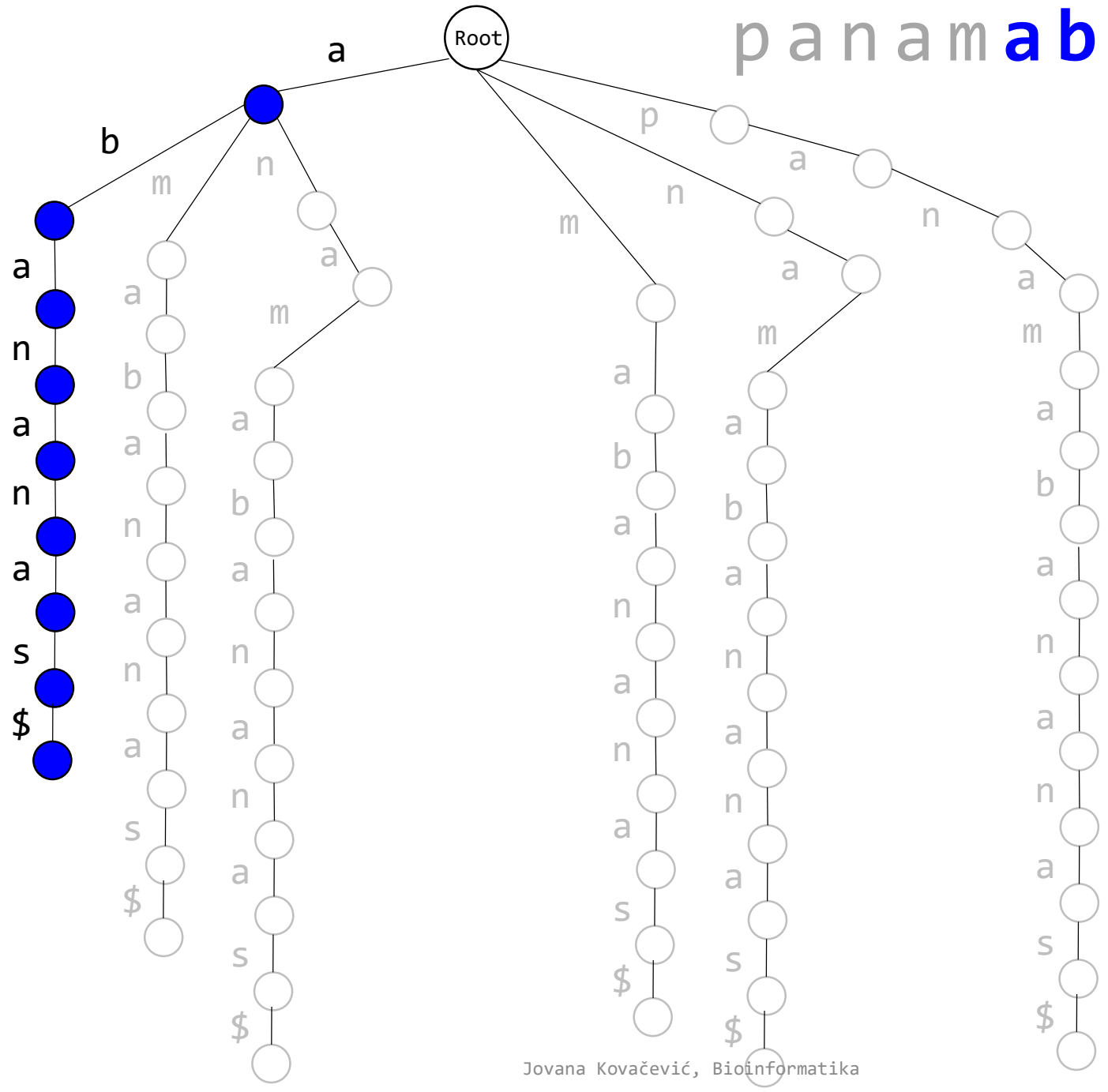
panamabananas
\$



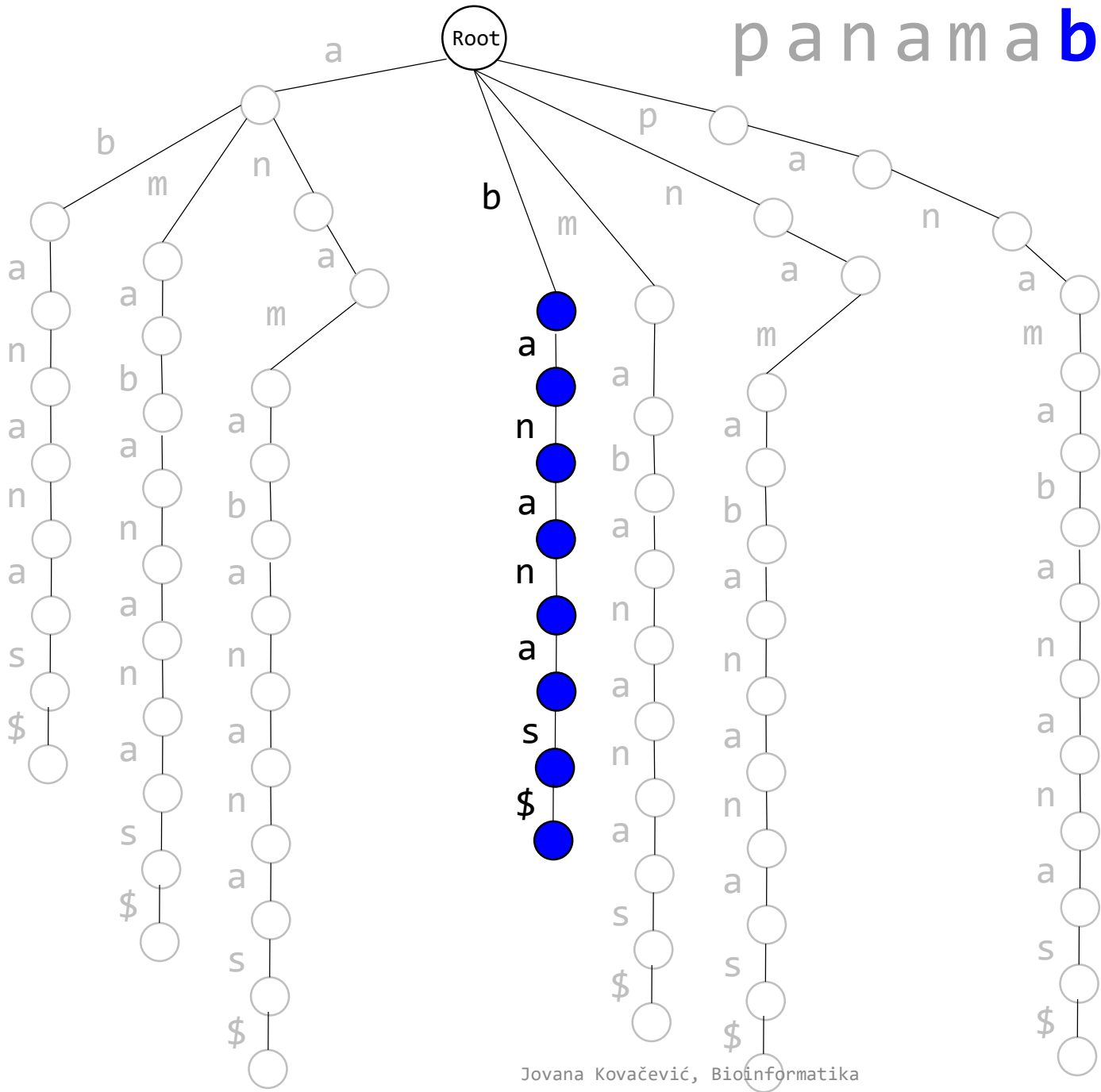
panamabananas
\$



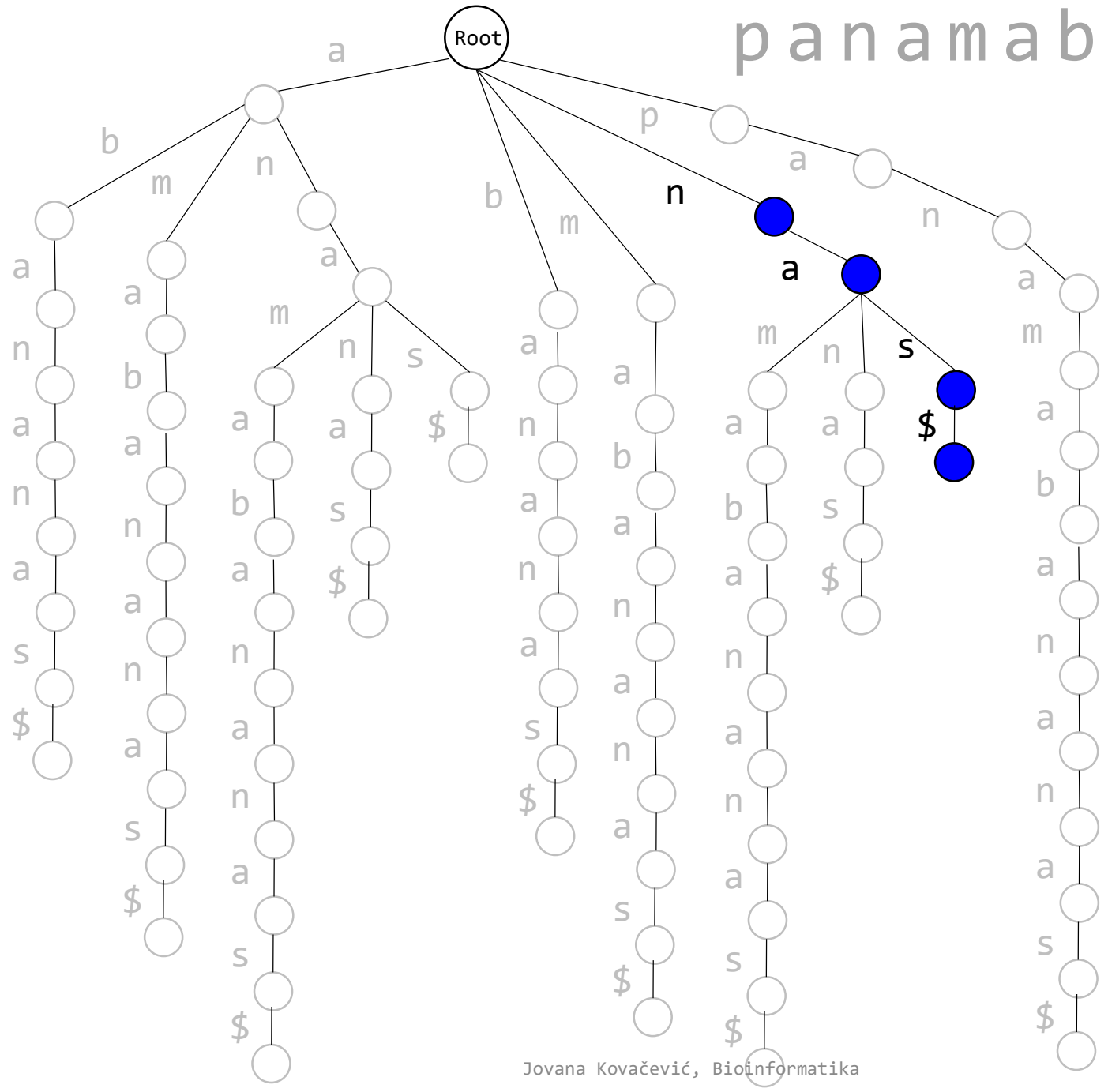
panamabanas
\$



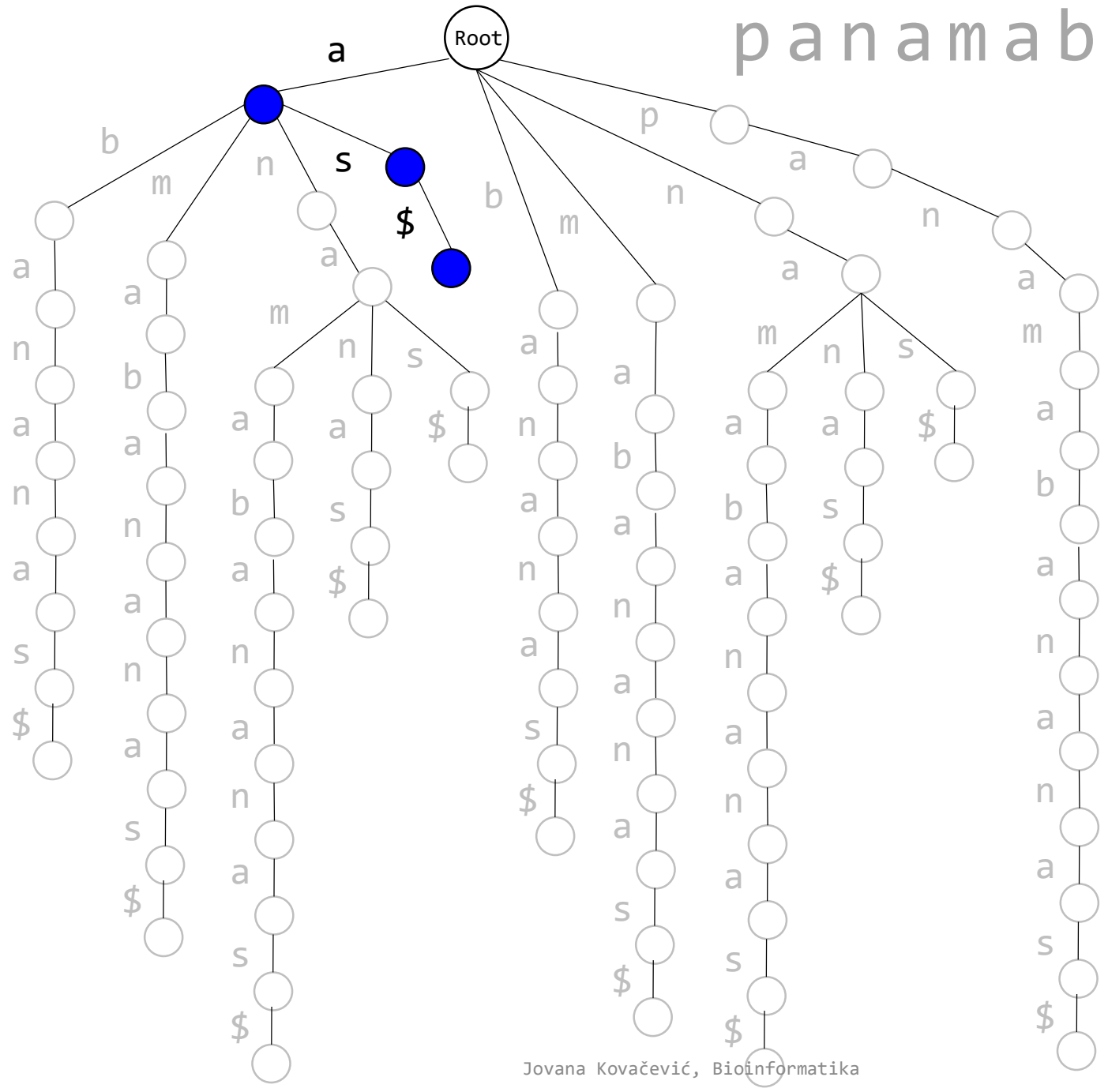
panama **bananas**
\$



panamabanas
\$

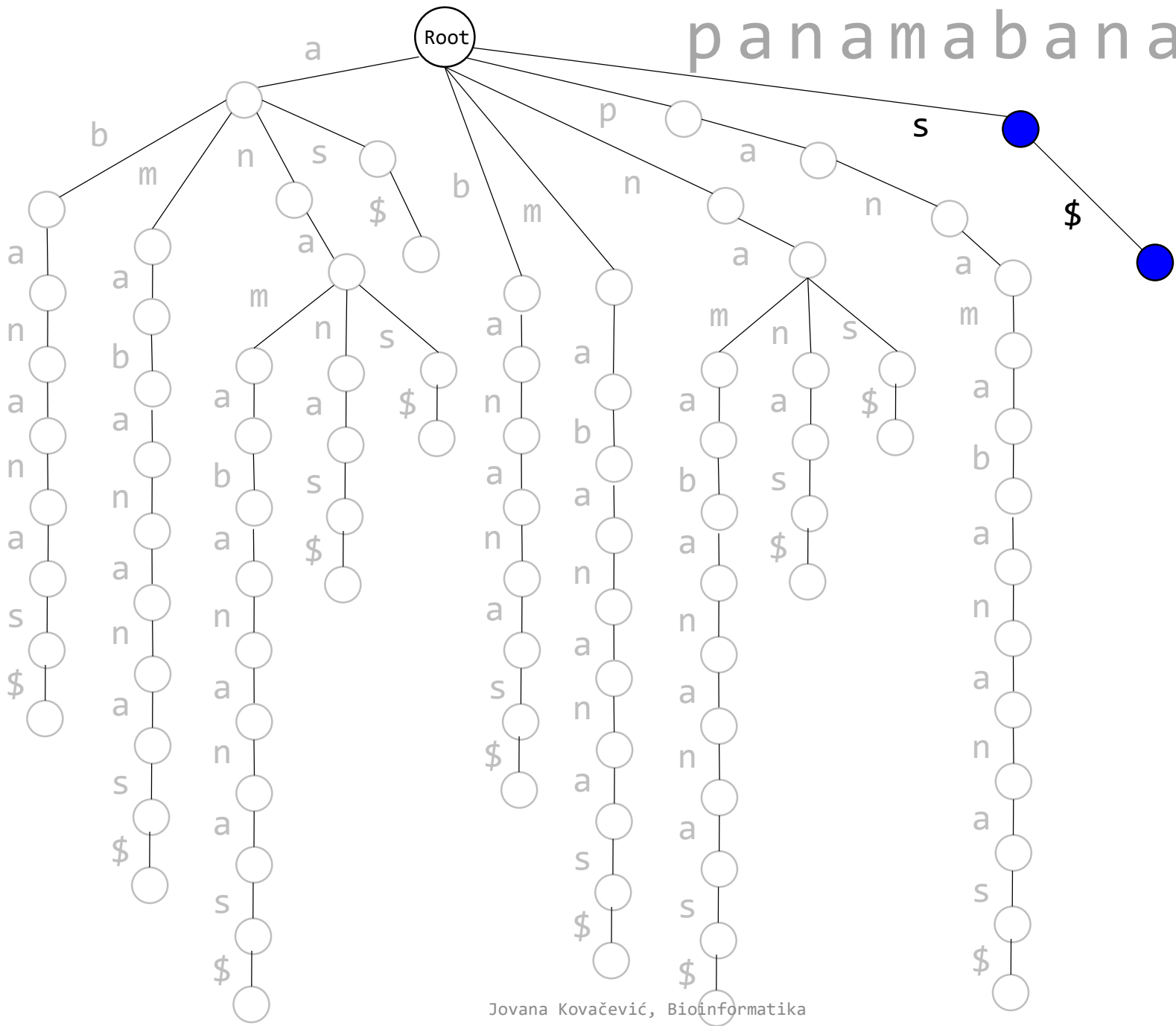


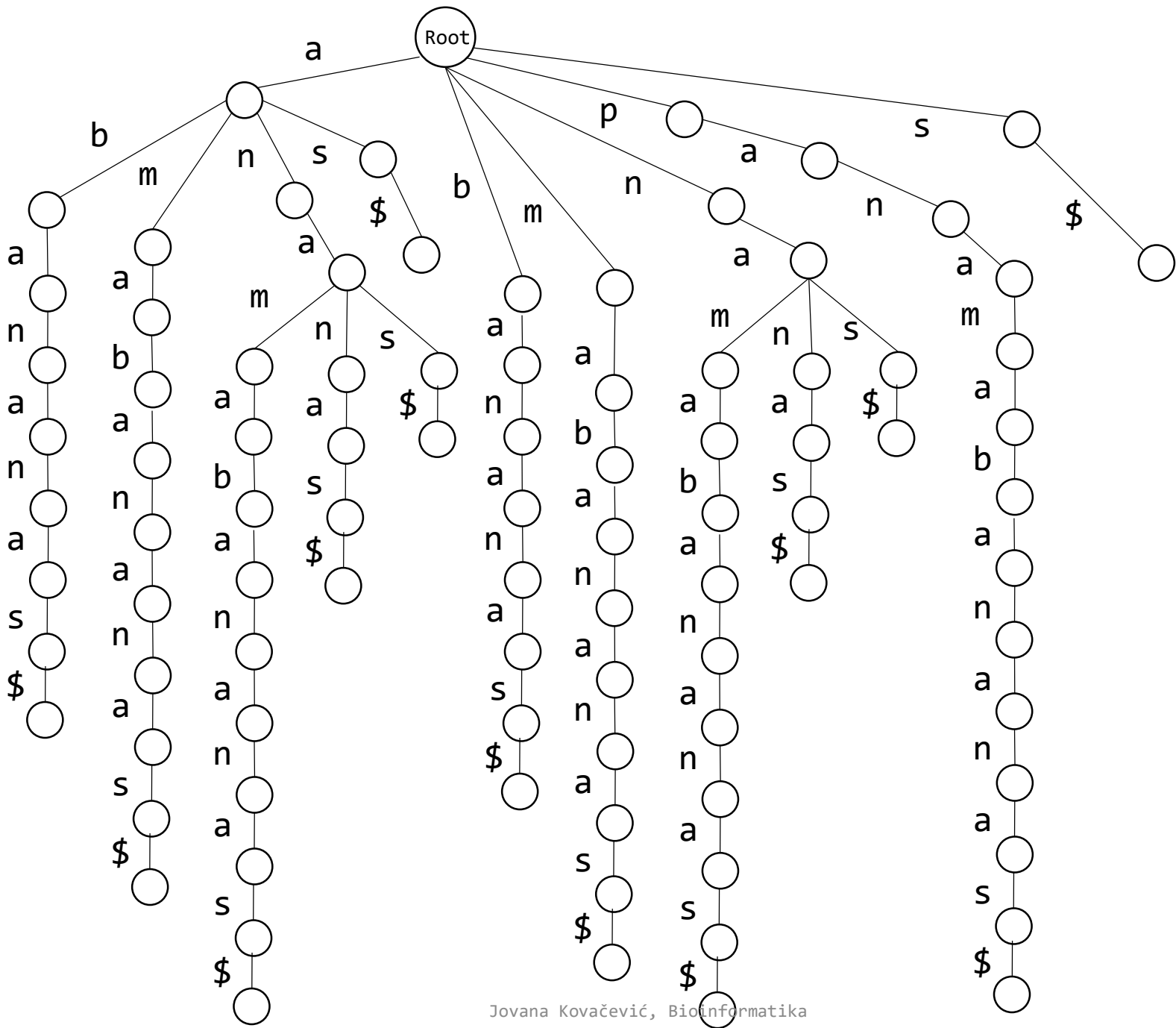
panamabanas
\$



panamabanaanasa

\$

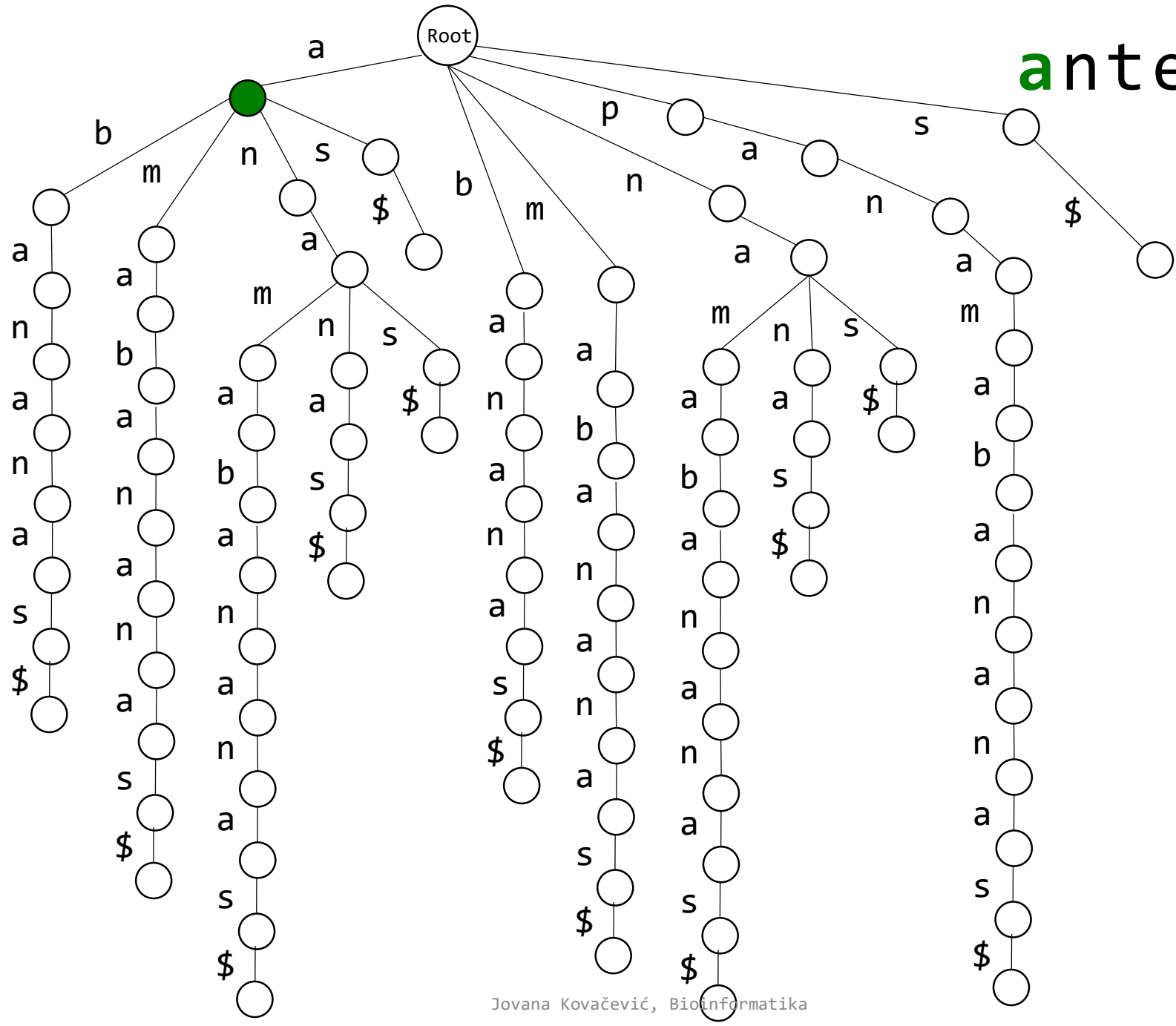




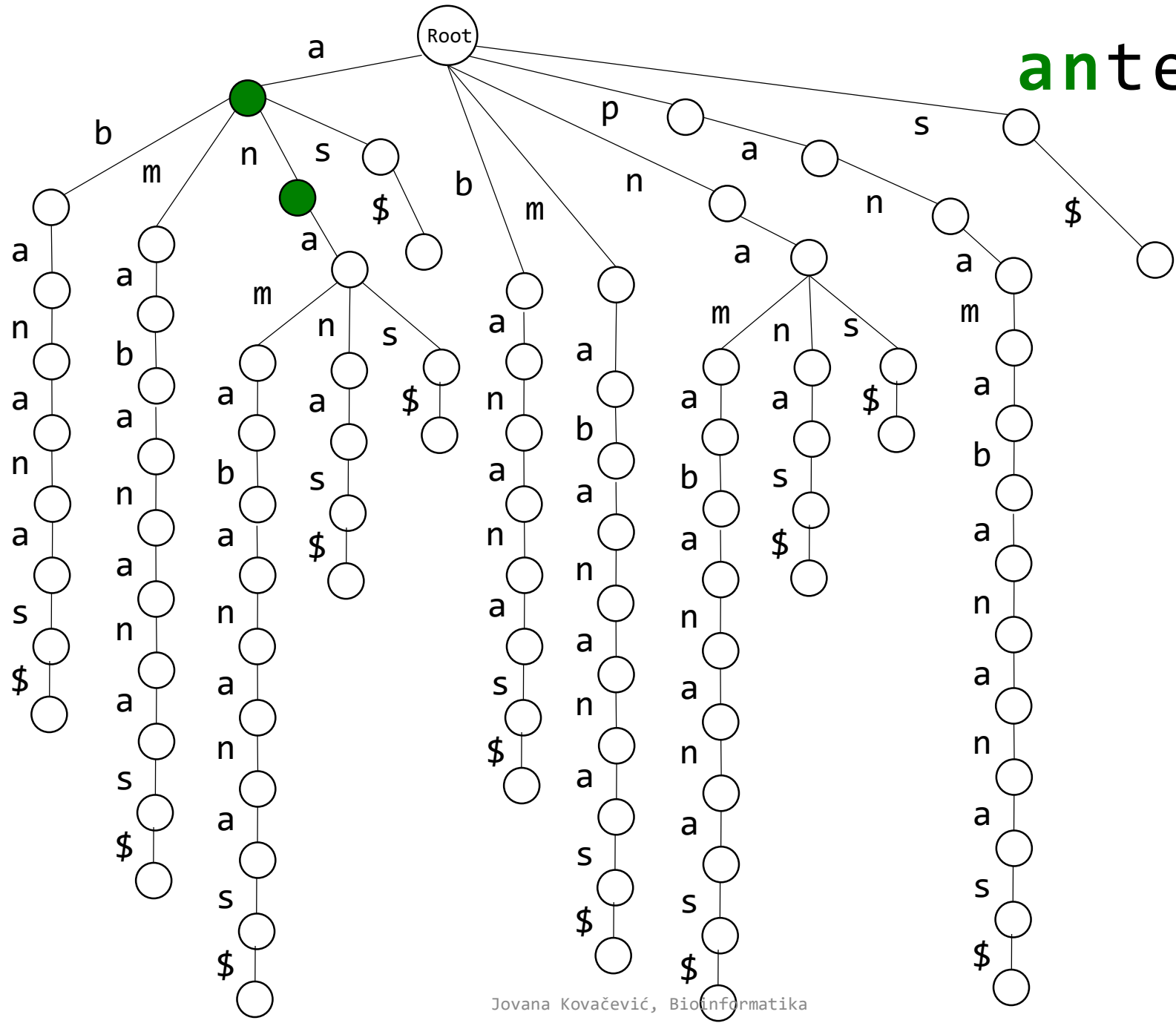
Nekompresovano sufiksno stablo i uparivanje šablona

- Za svaku nisku *Pattern*, proveriti da li se *Pattern* može pročitati od korena naniže u sufiksnom stablu
- Ne moramo da dođemo do listova sufiksnog stabla
- Moramo da potrošimo sve karaktere iz niske *Pattern*

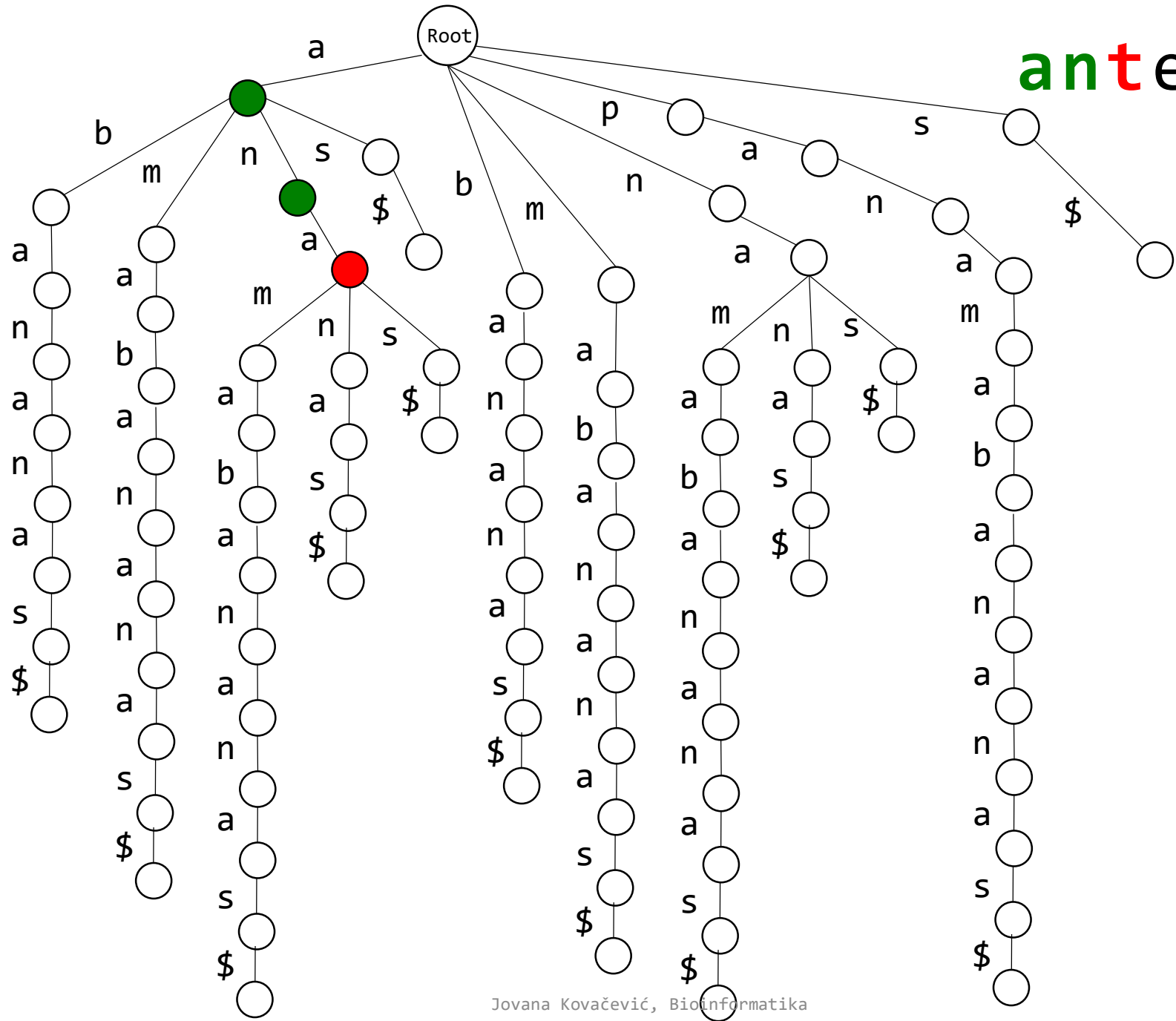
antenna



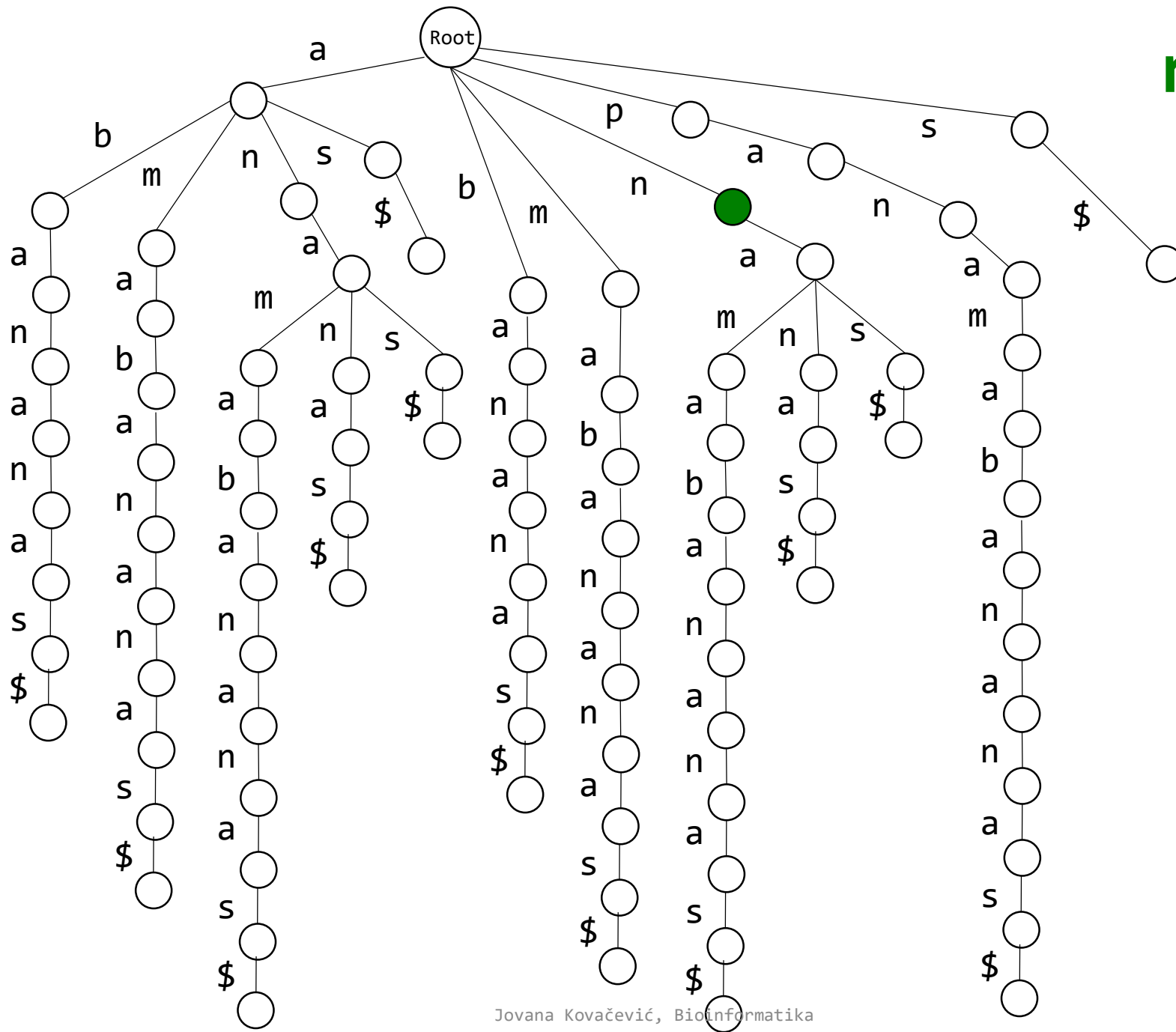
antenna



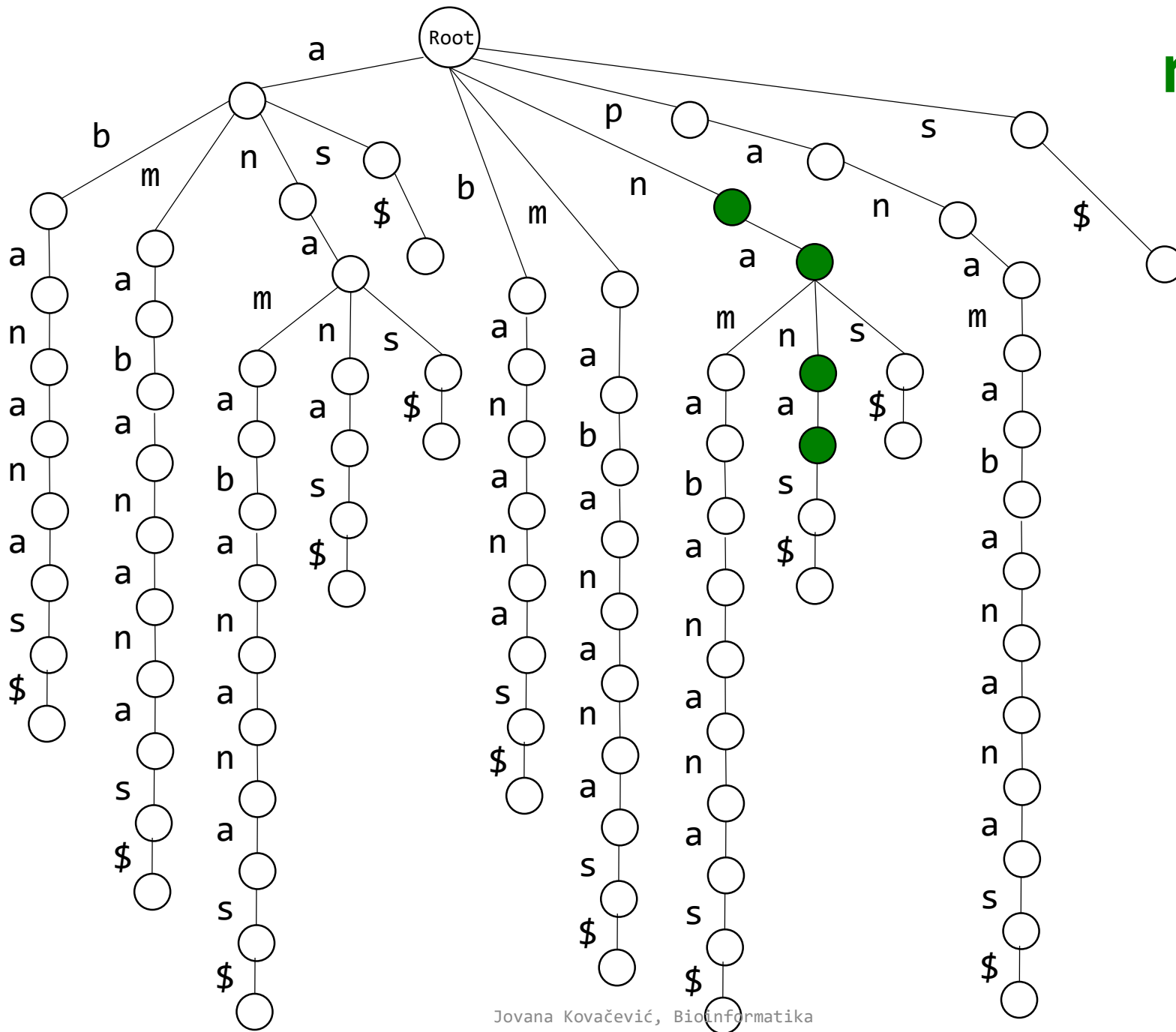
antenna



nana



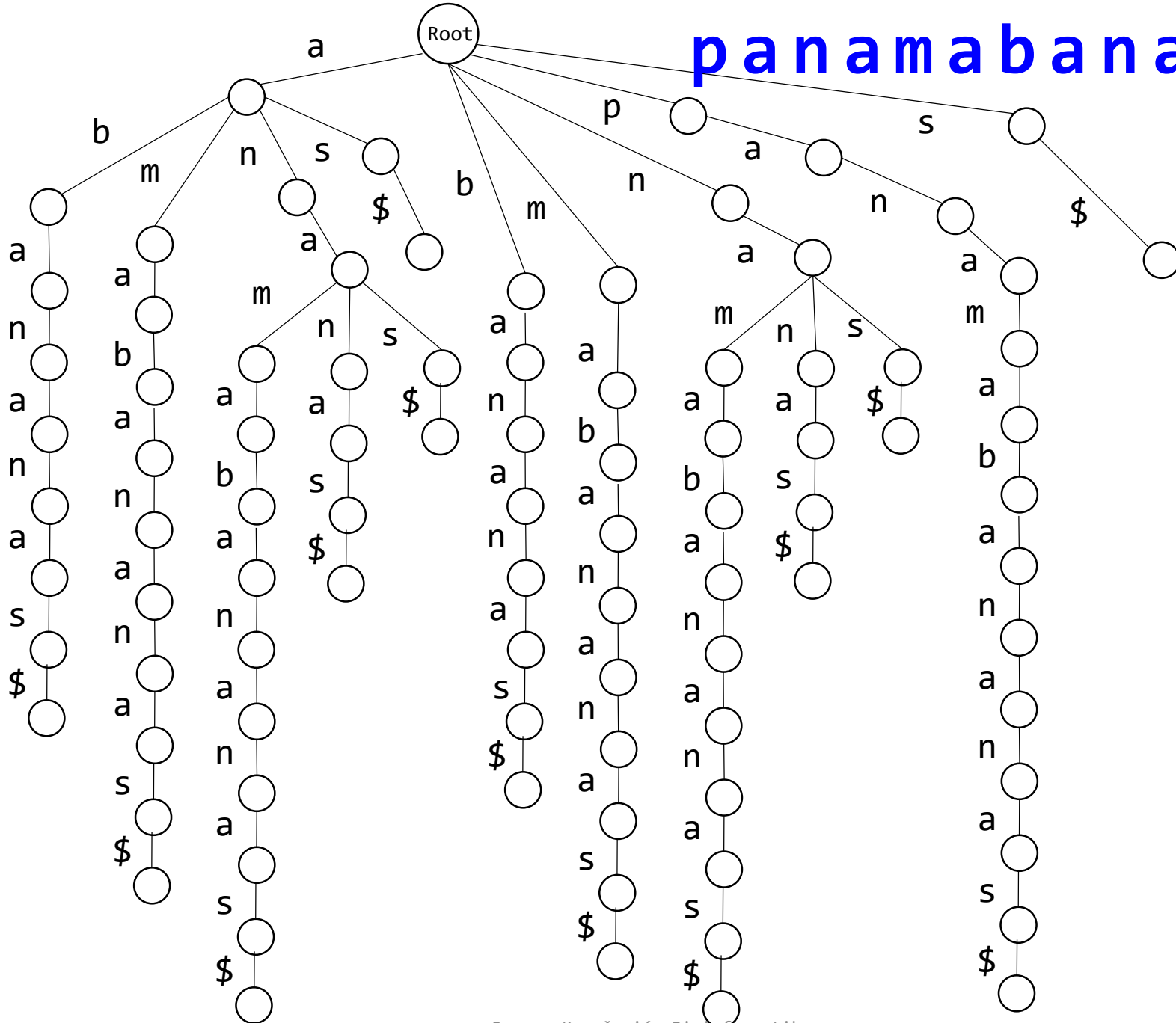
nana



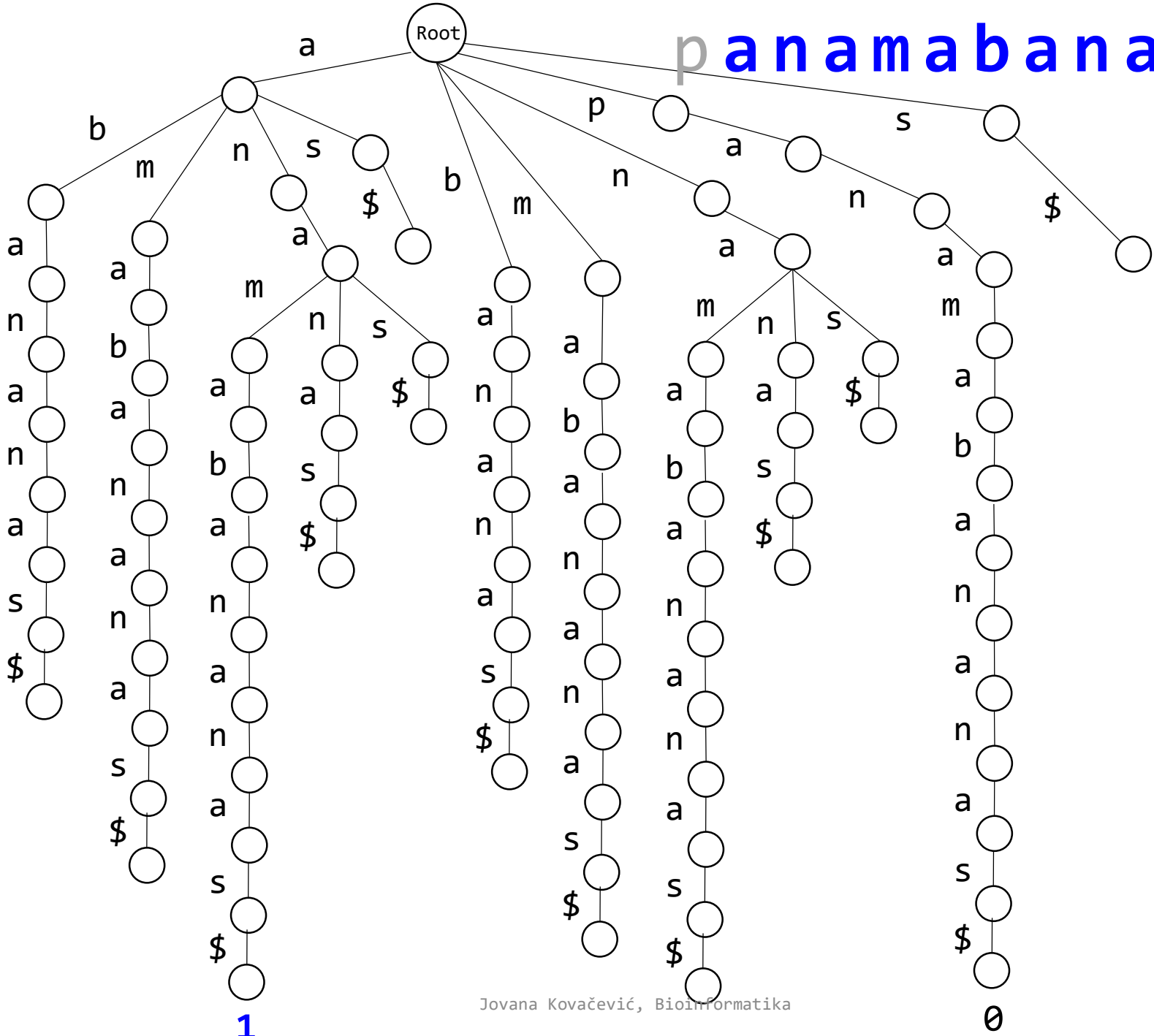
Gde su uparivanja?

- Opisanim pristupom možemo utvrditi da li se *Pattern* pojavljuje u *Genome*, ali ne i na kojoj poziciji
- Da bismo i to utvrdili, moramo dodati još informacija u sufiksno stablo.
- Na svakom listu (\$), dodajemo početnu poziciju u niski *Genome* sufiksa koji se završava u tom listu.

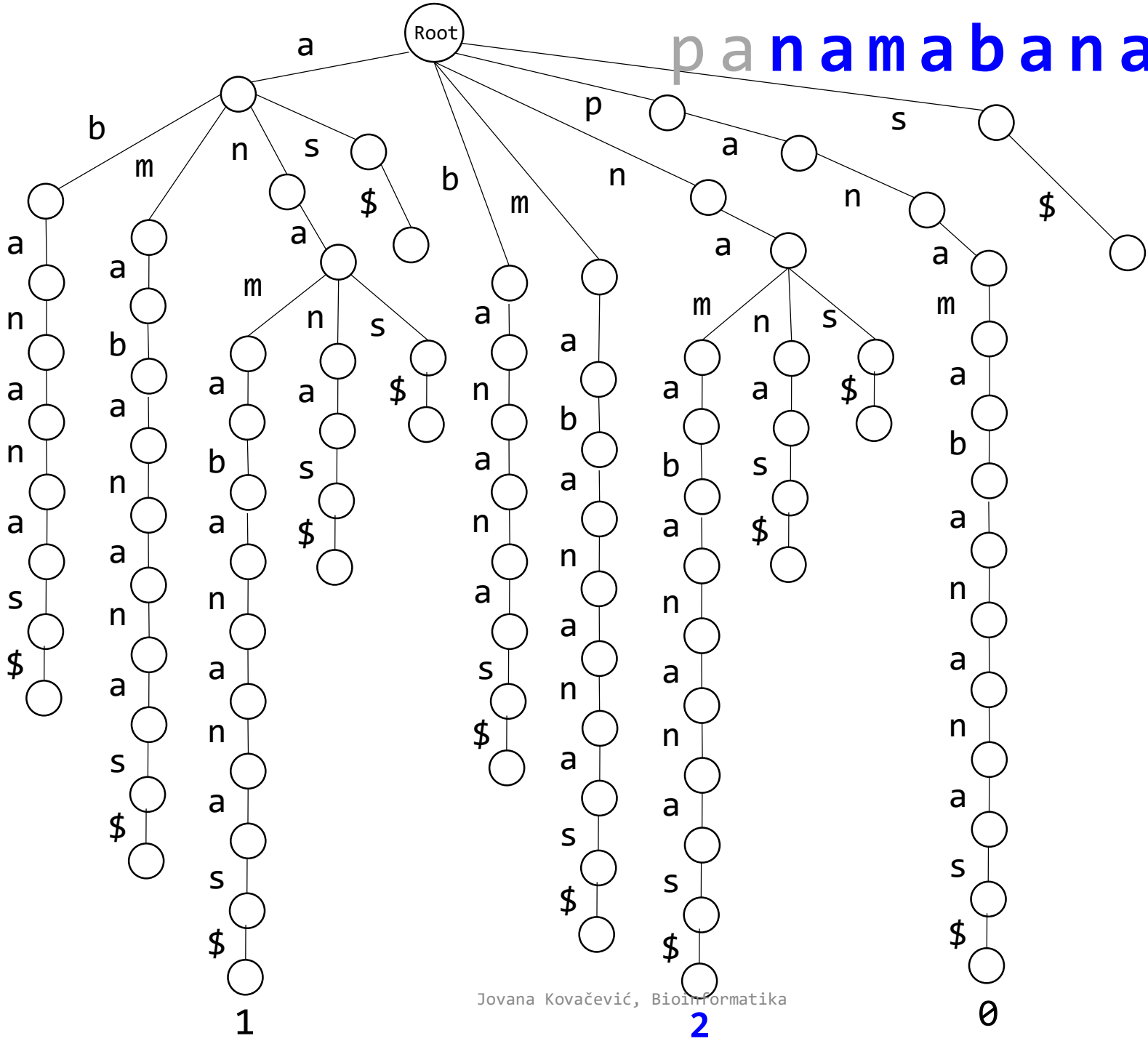
panamabananas



panamabananas



panamabananas

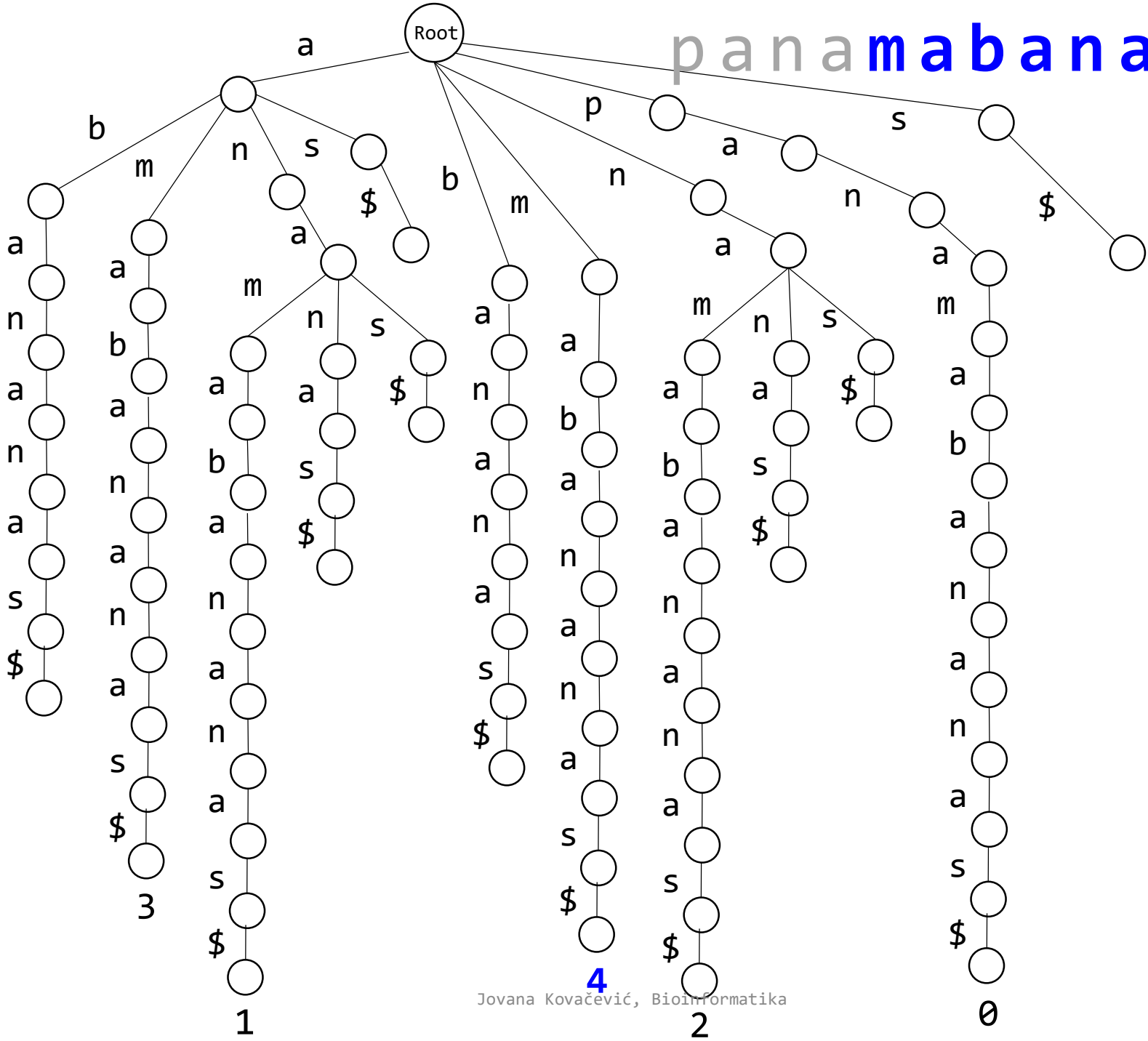


1

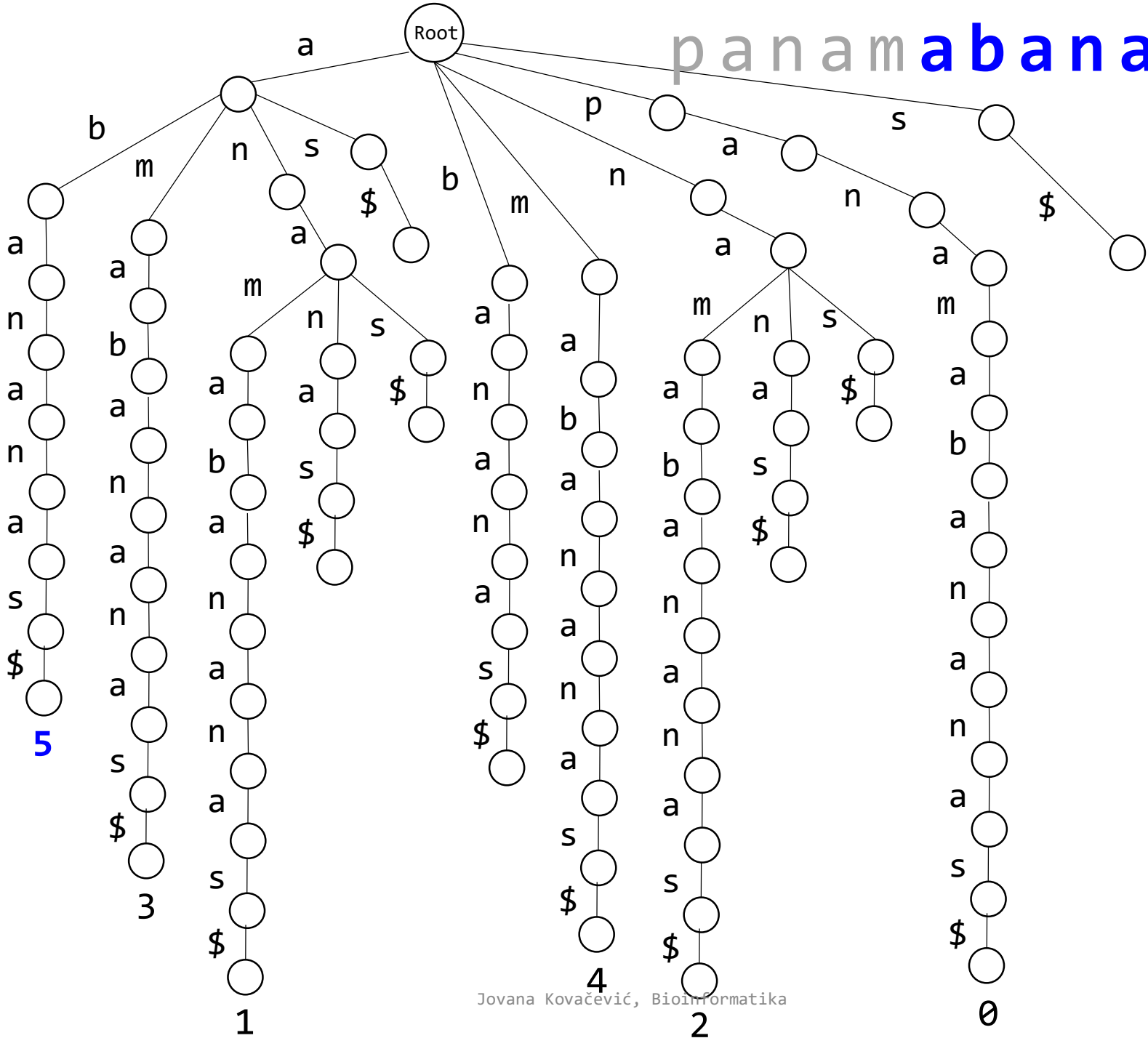
2

0

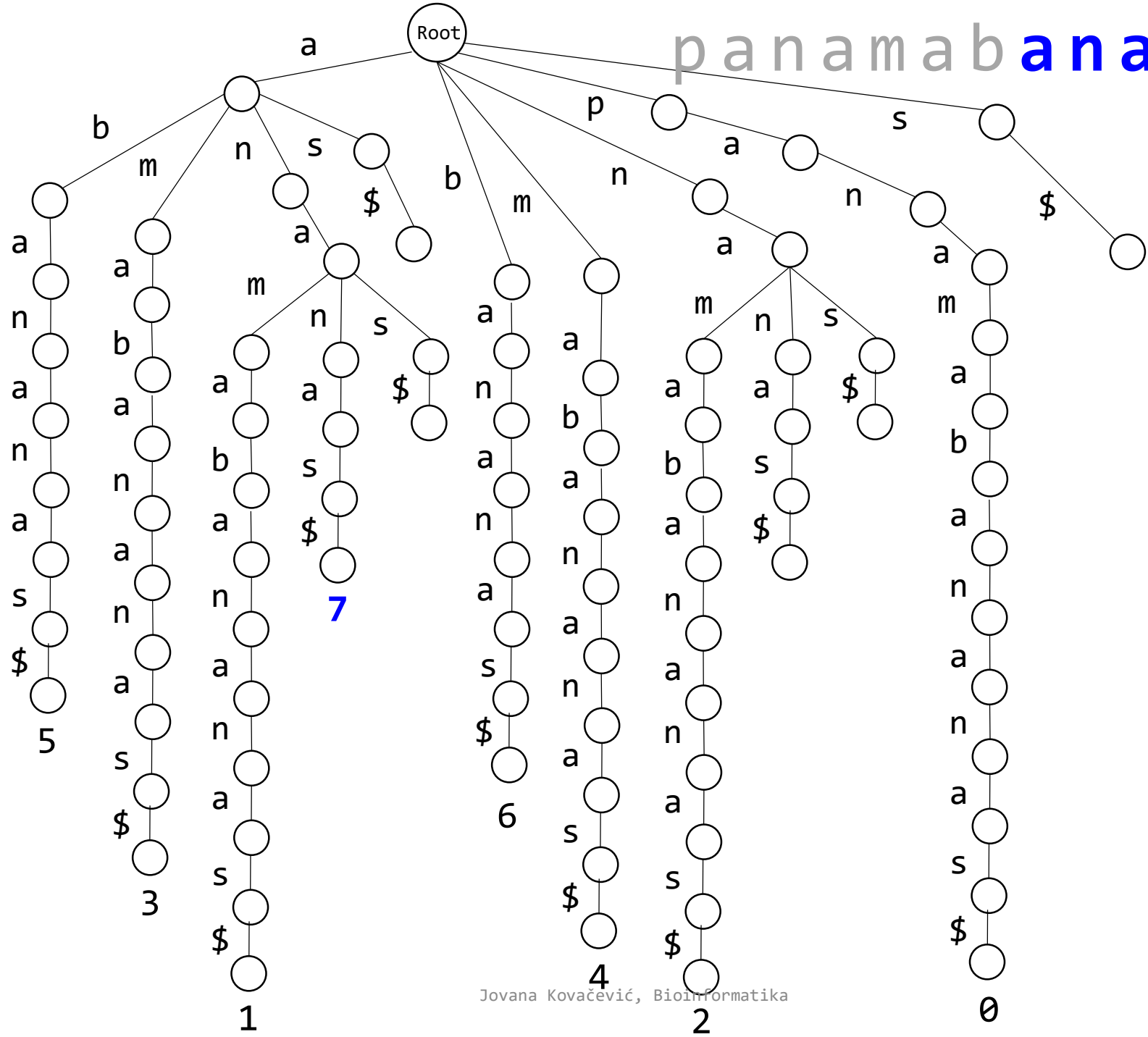
panamabananas



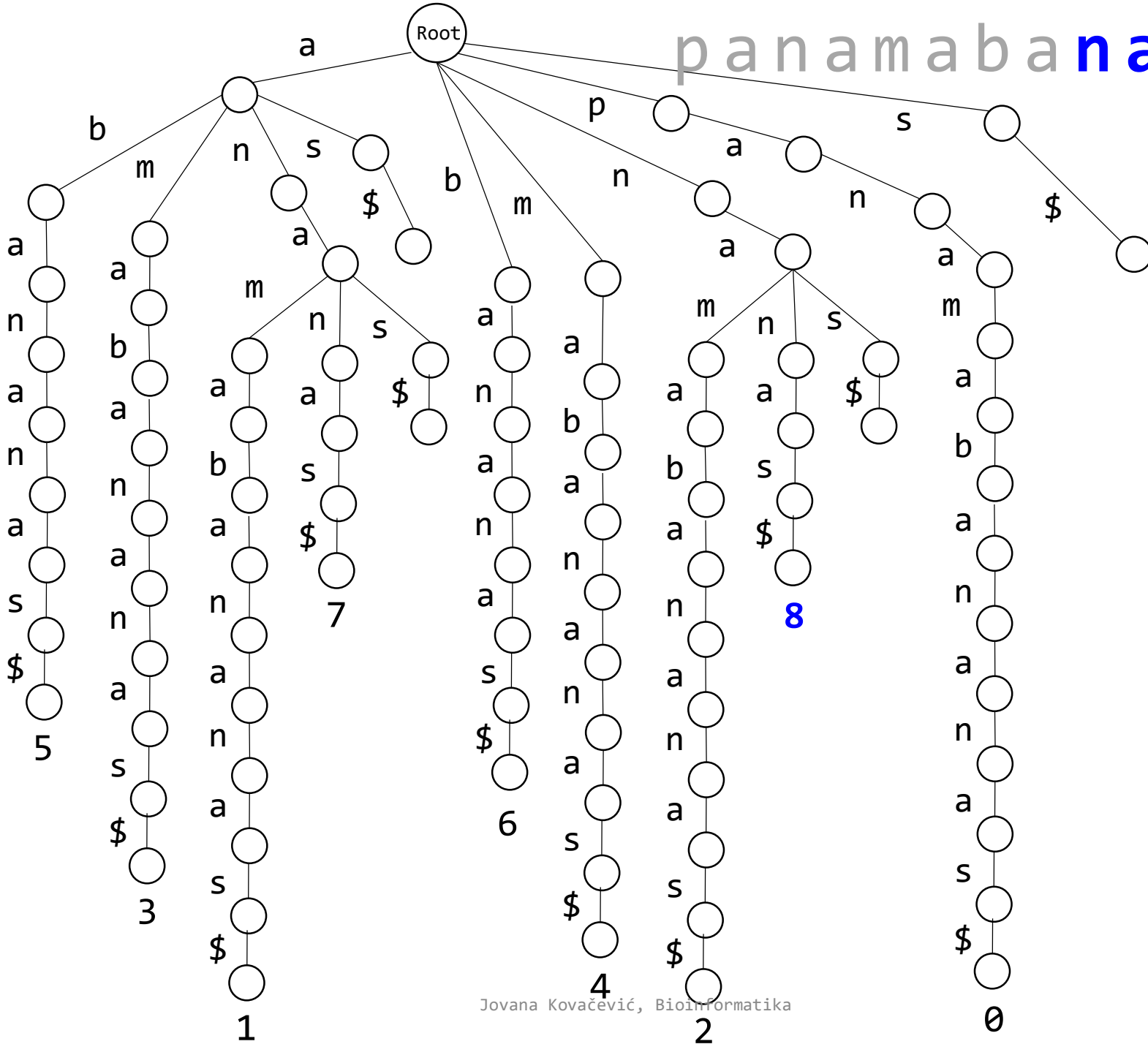
panamabanas
\$



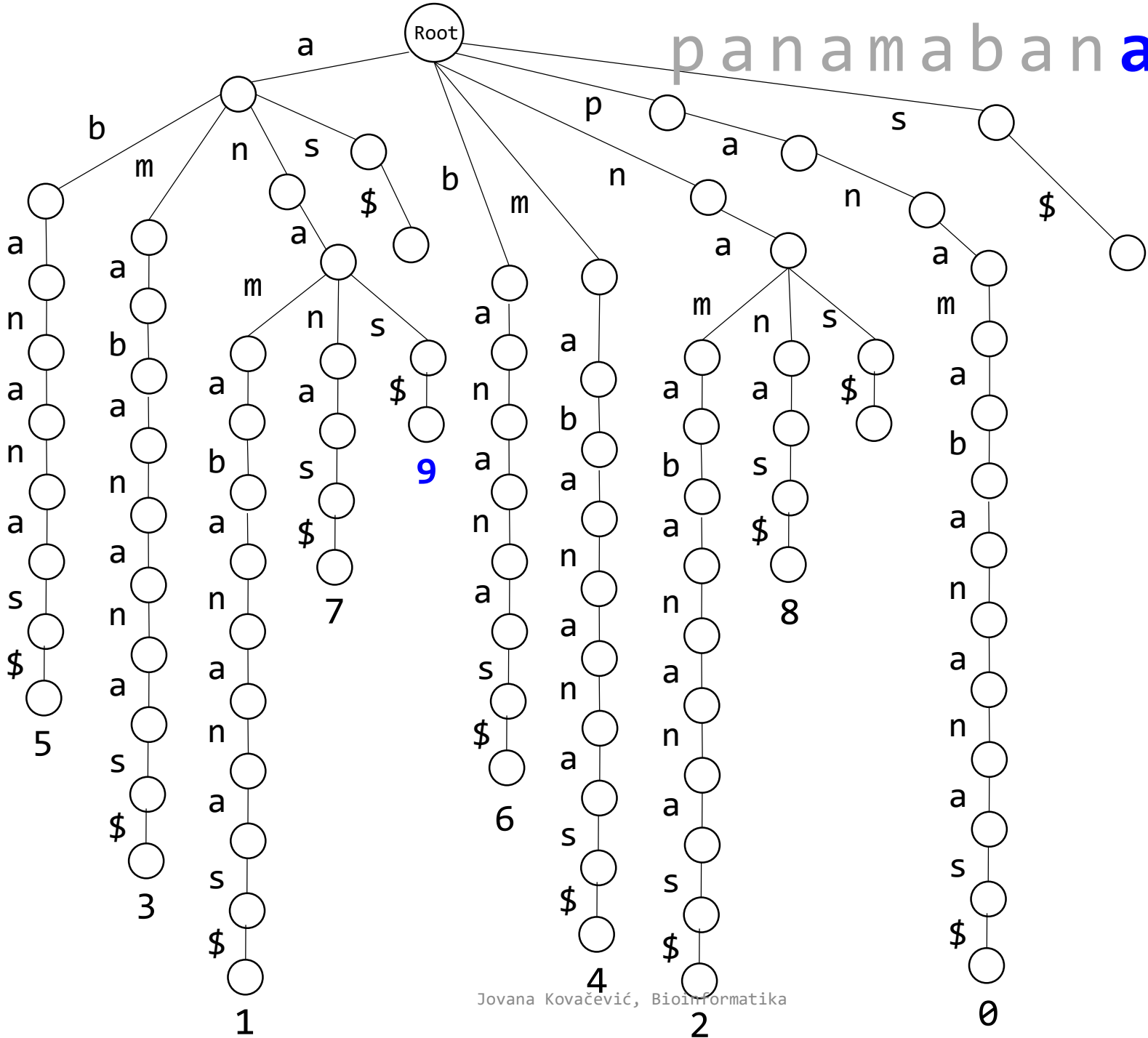
panamabanas
\$



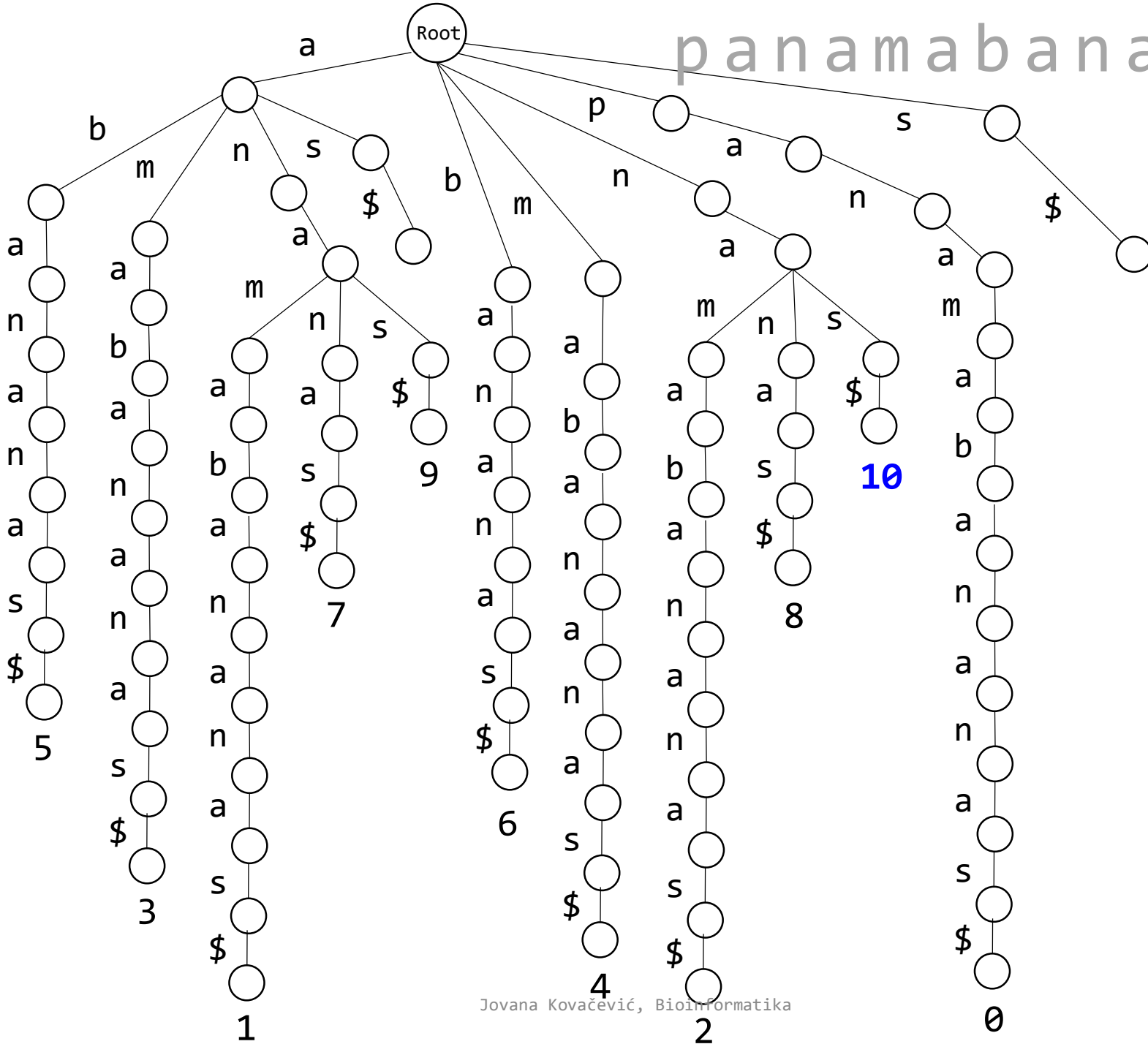
panamabananas
\$



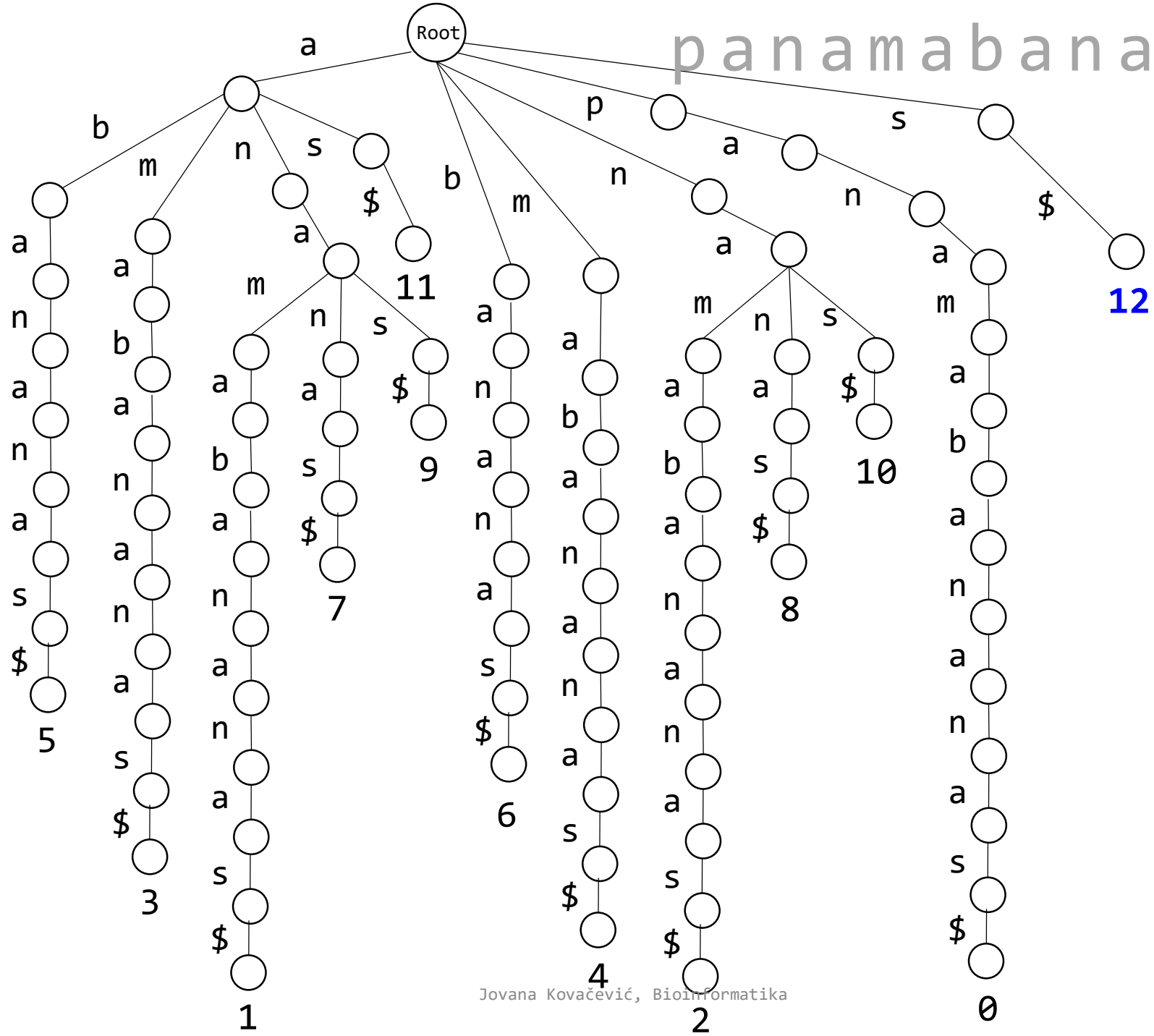
panamabanas
\$

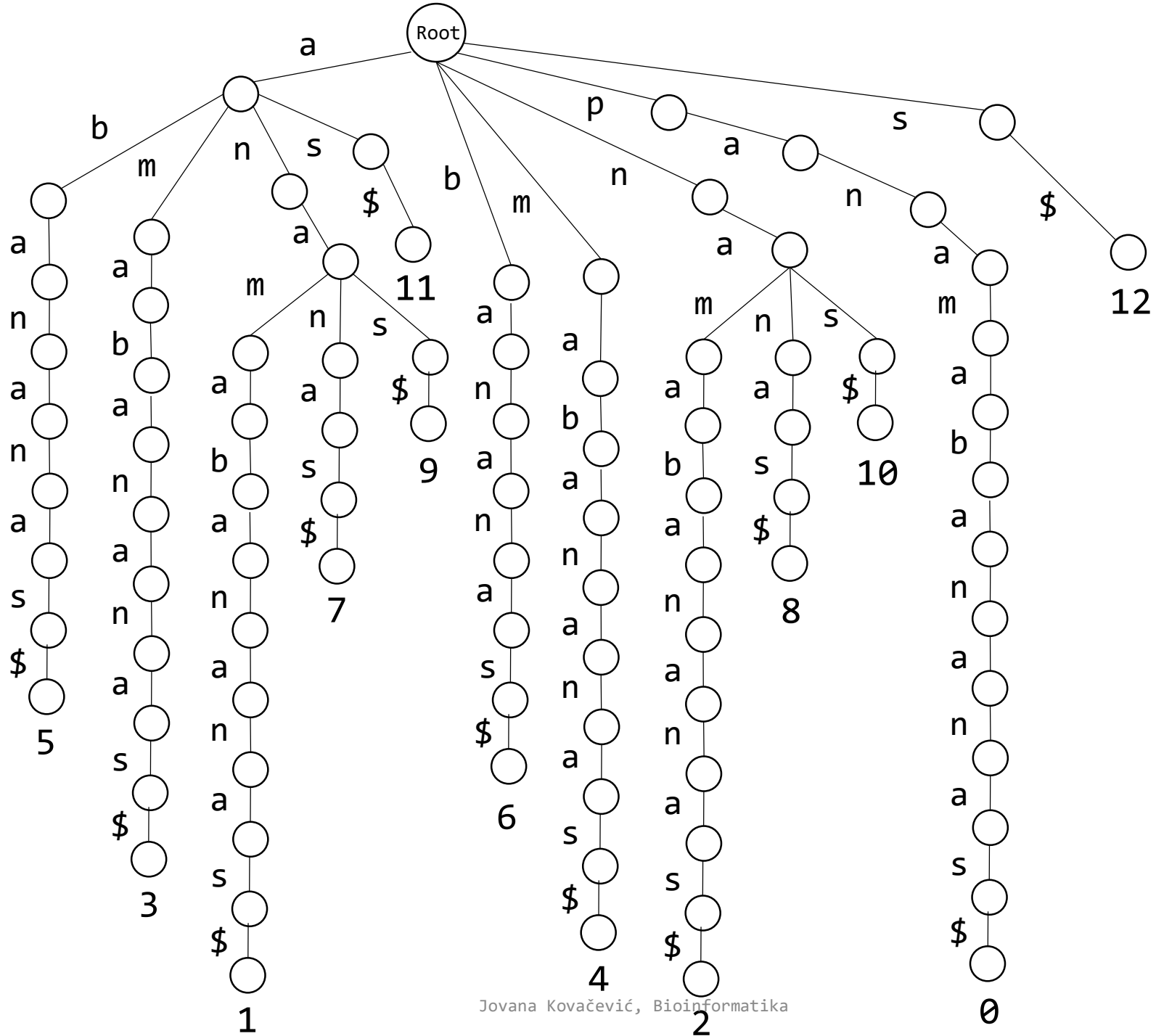


panamabanas
\$



panamabanaanasa

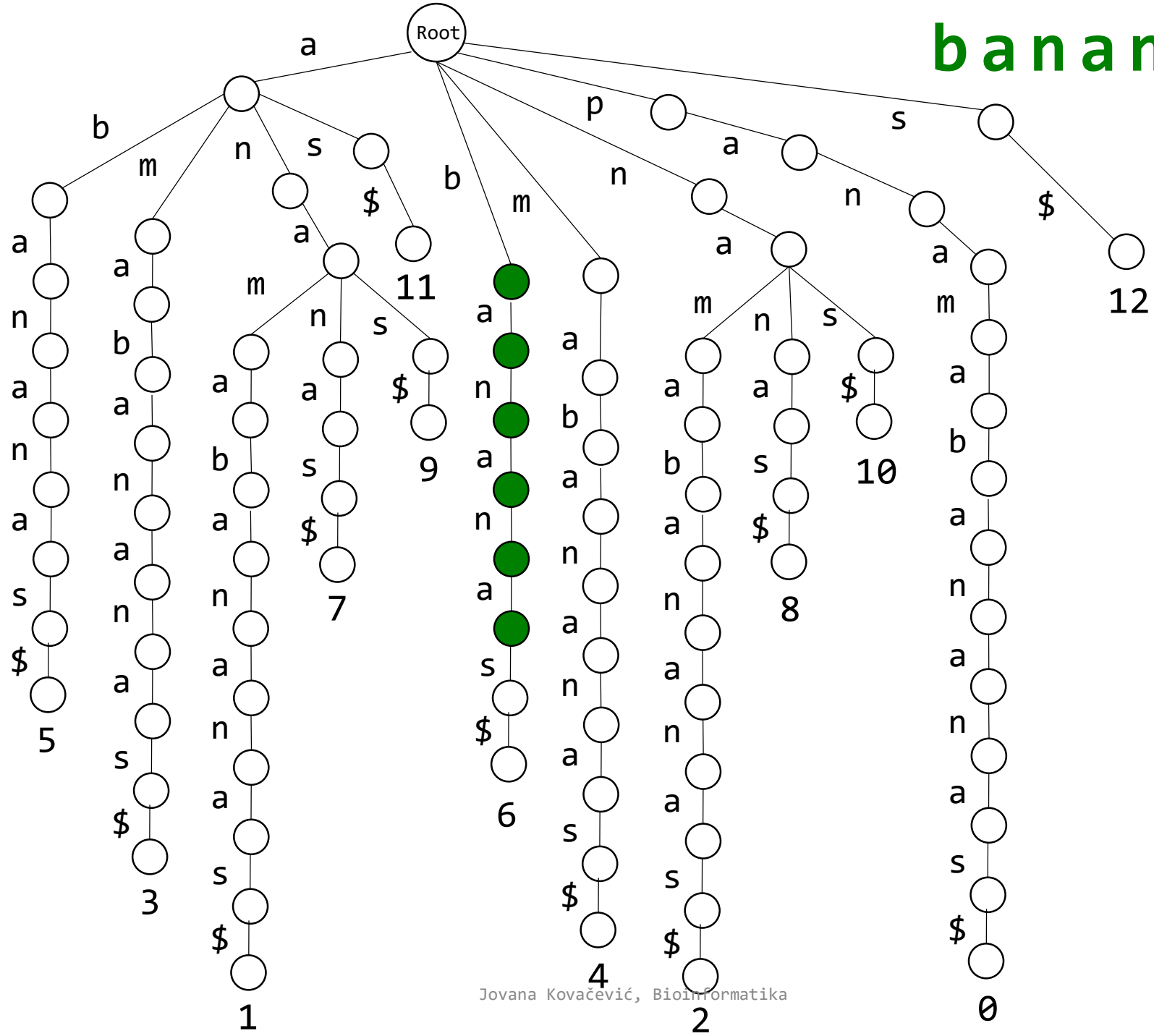




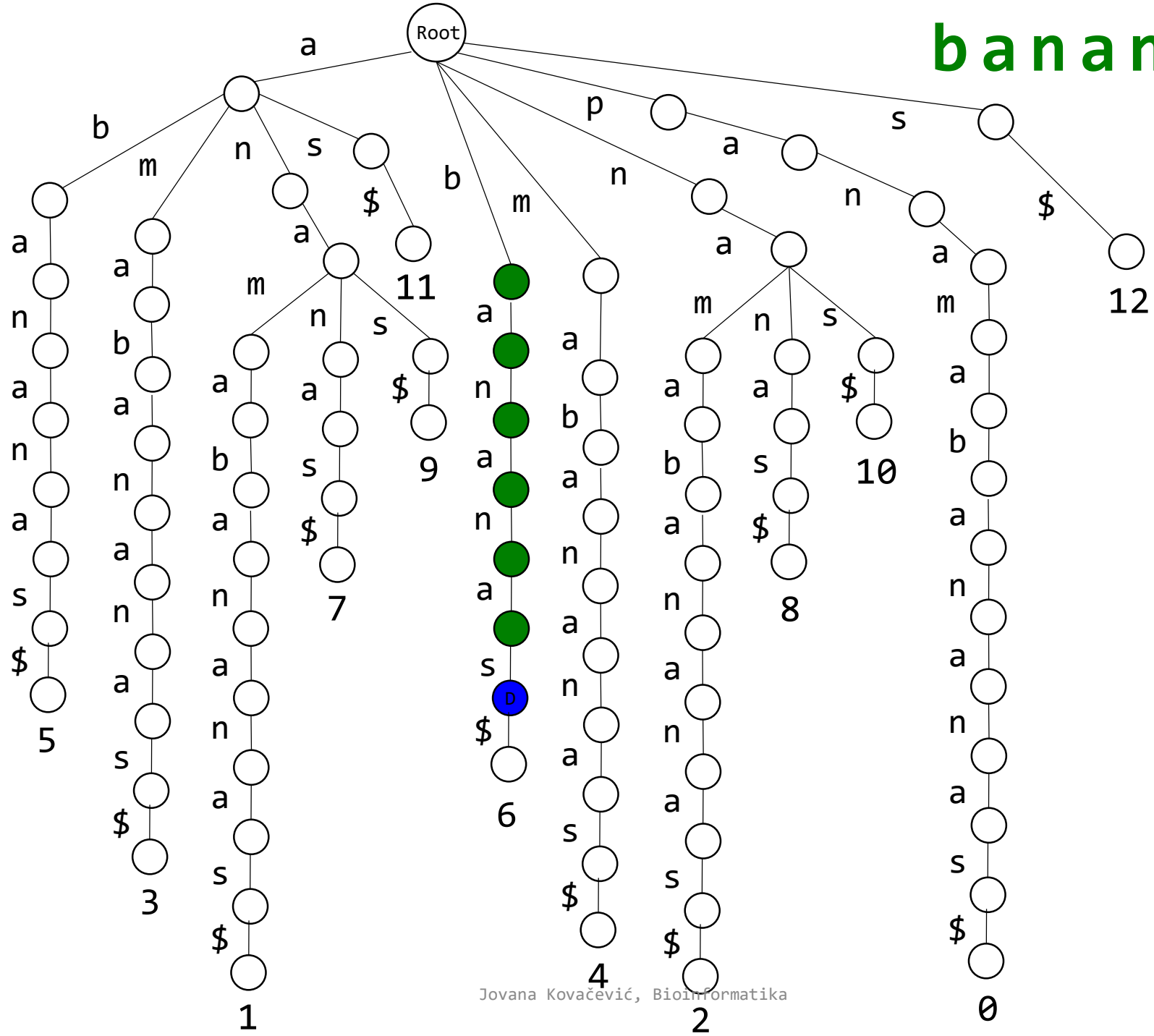
Gde su uparivanja?

- Kada pronađemo uparivanje, nastavimo sa kretanjem naniže do lista kako bismo pronašli poziciju odakle počinje pojavljivanje podniske.

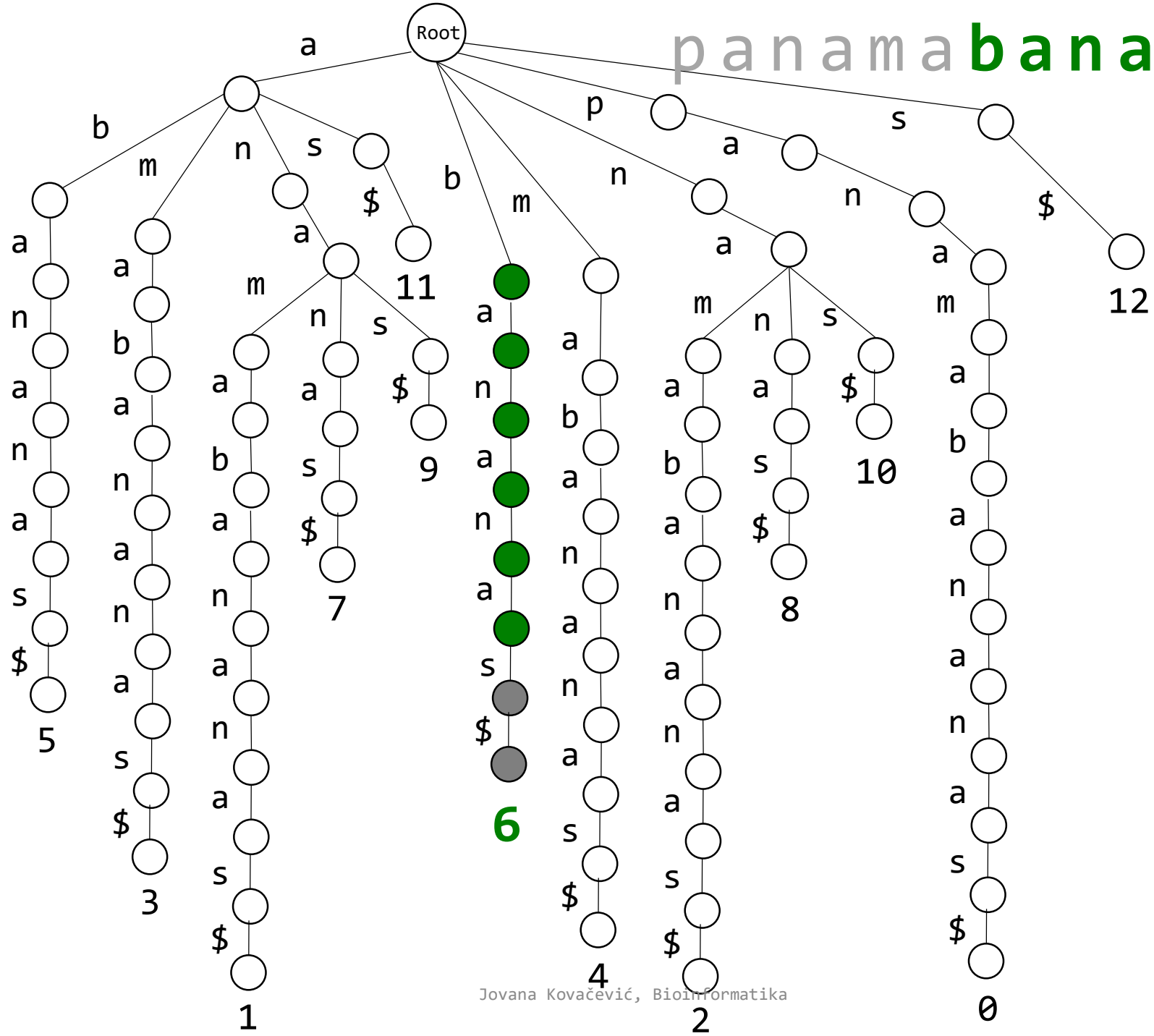
banana



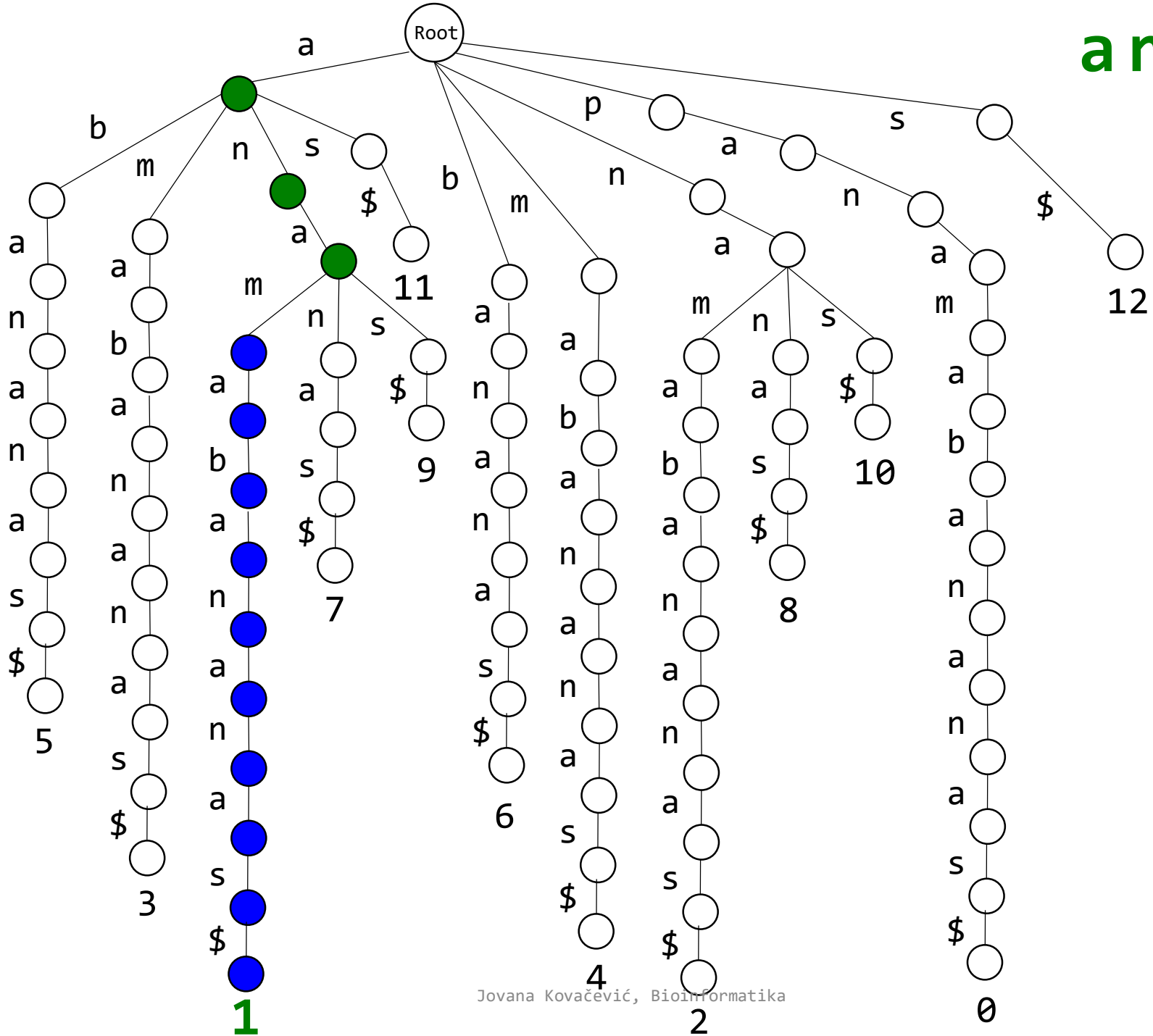
banana



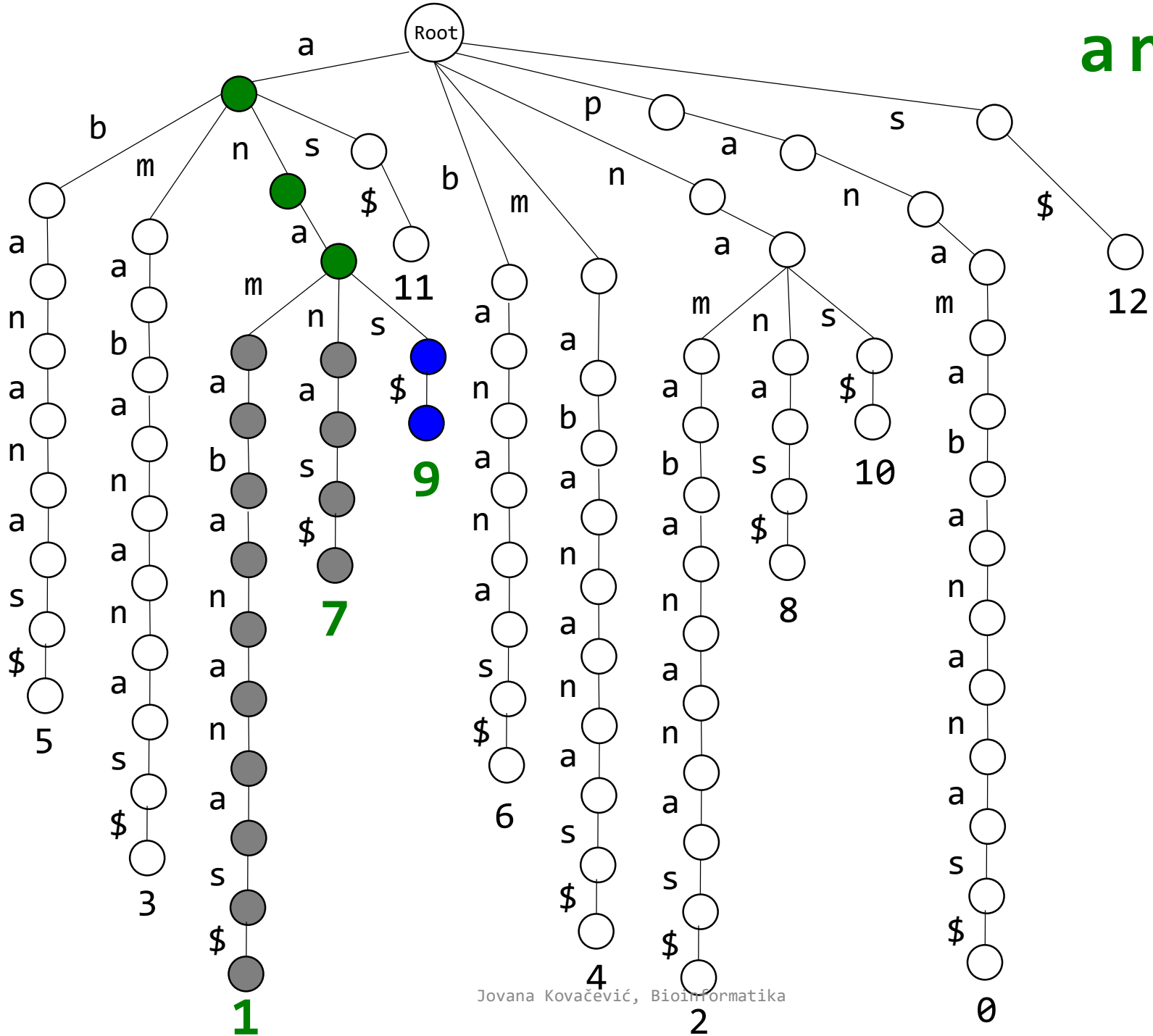
panama bananas \$



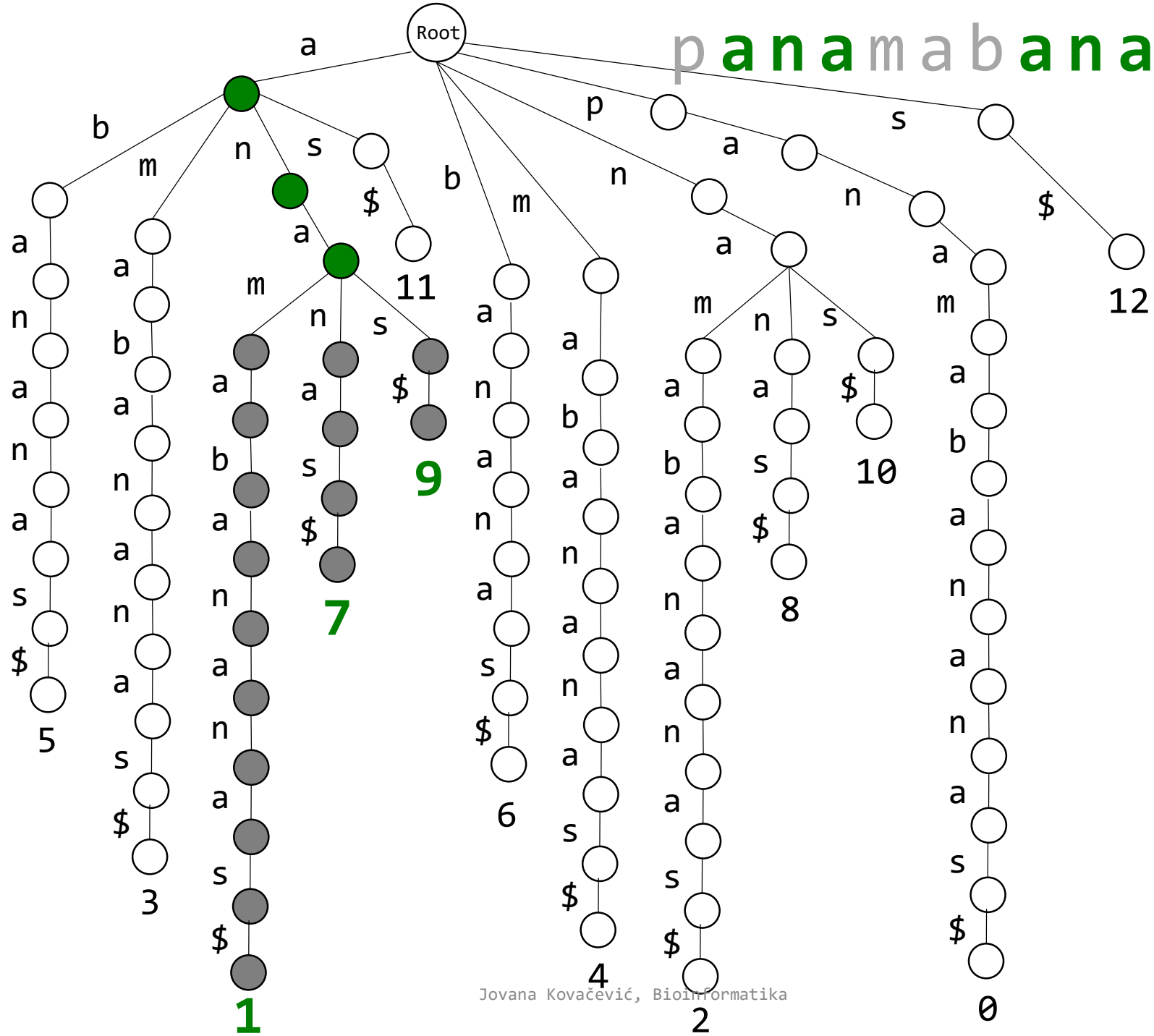
ana



ana



panamabananas \$



Problem sa prostornom složenošću

- Najgori slučaj: sufiksno stablo sadrži $O(|Suffixes|)$ čvorova.

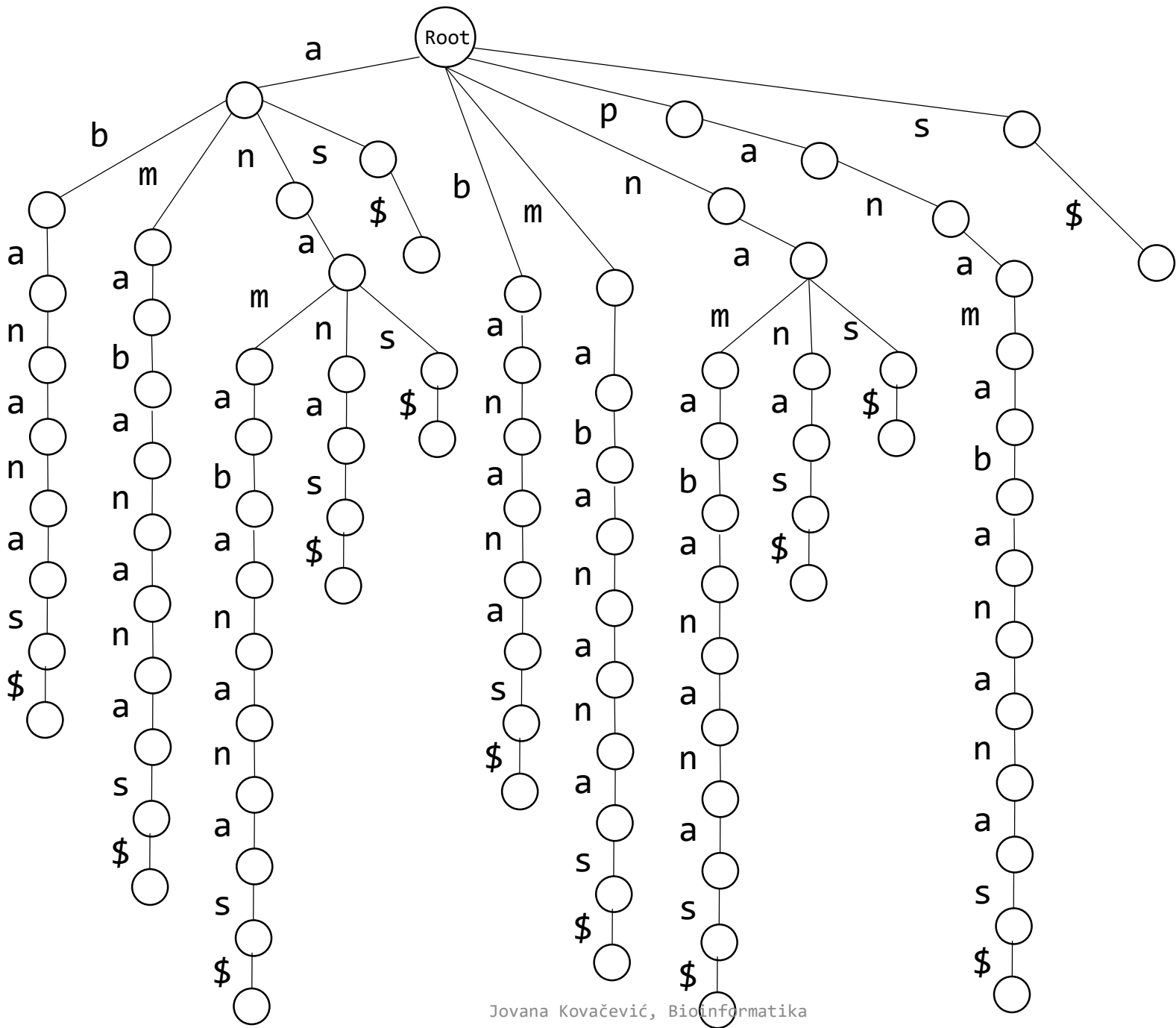
- Za genom dužine n ,
 $|Suffixes| = n(n - 1)/2 = O(n^2)$

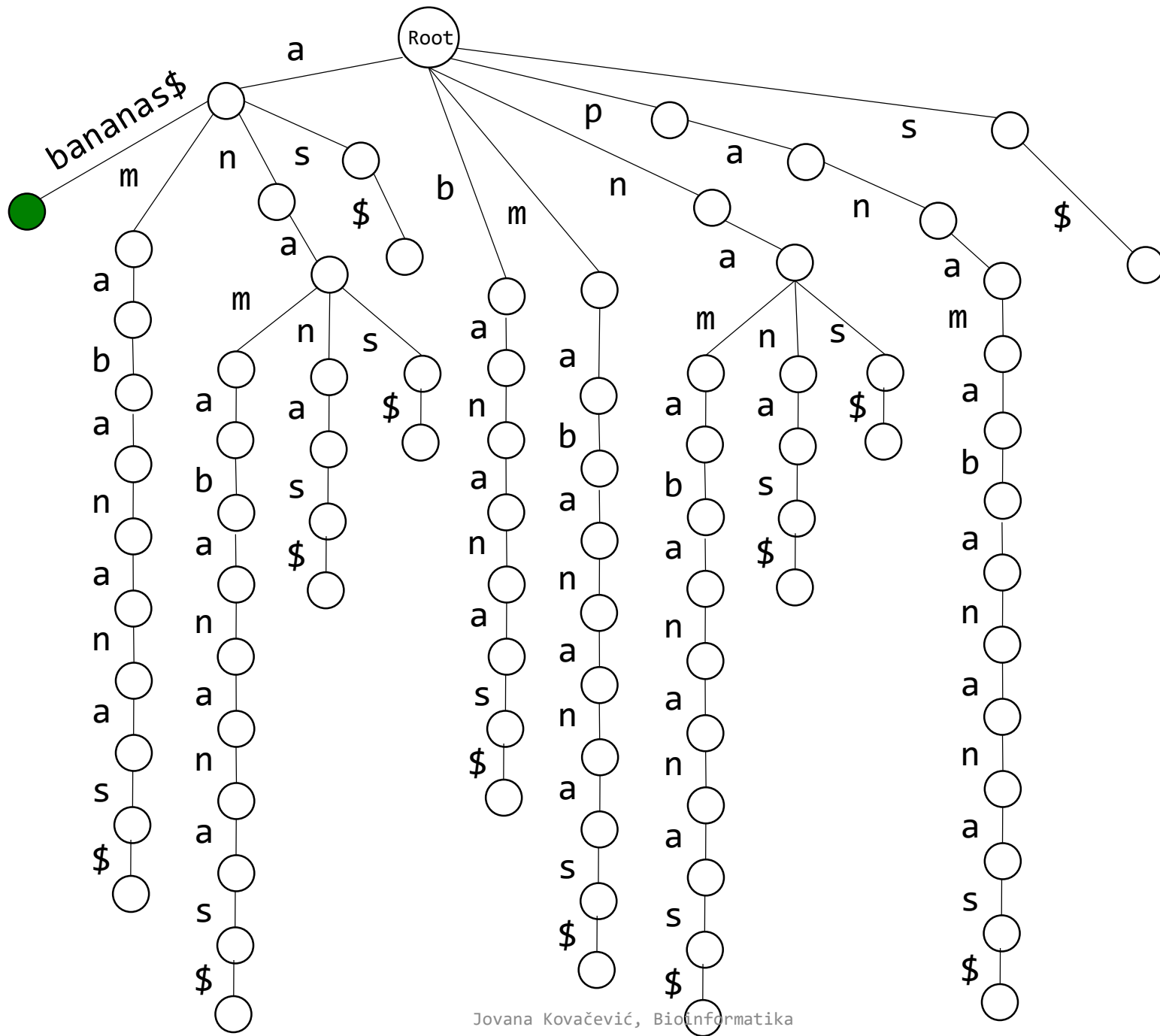
Suffixes

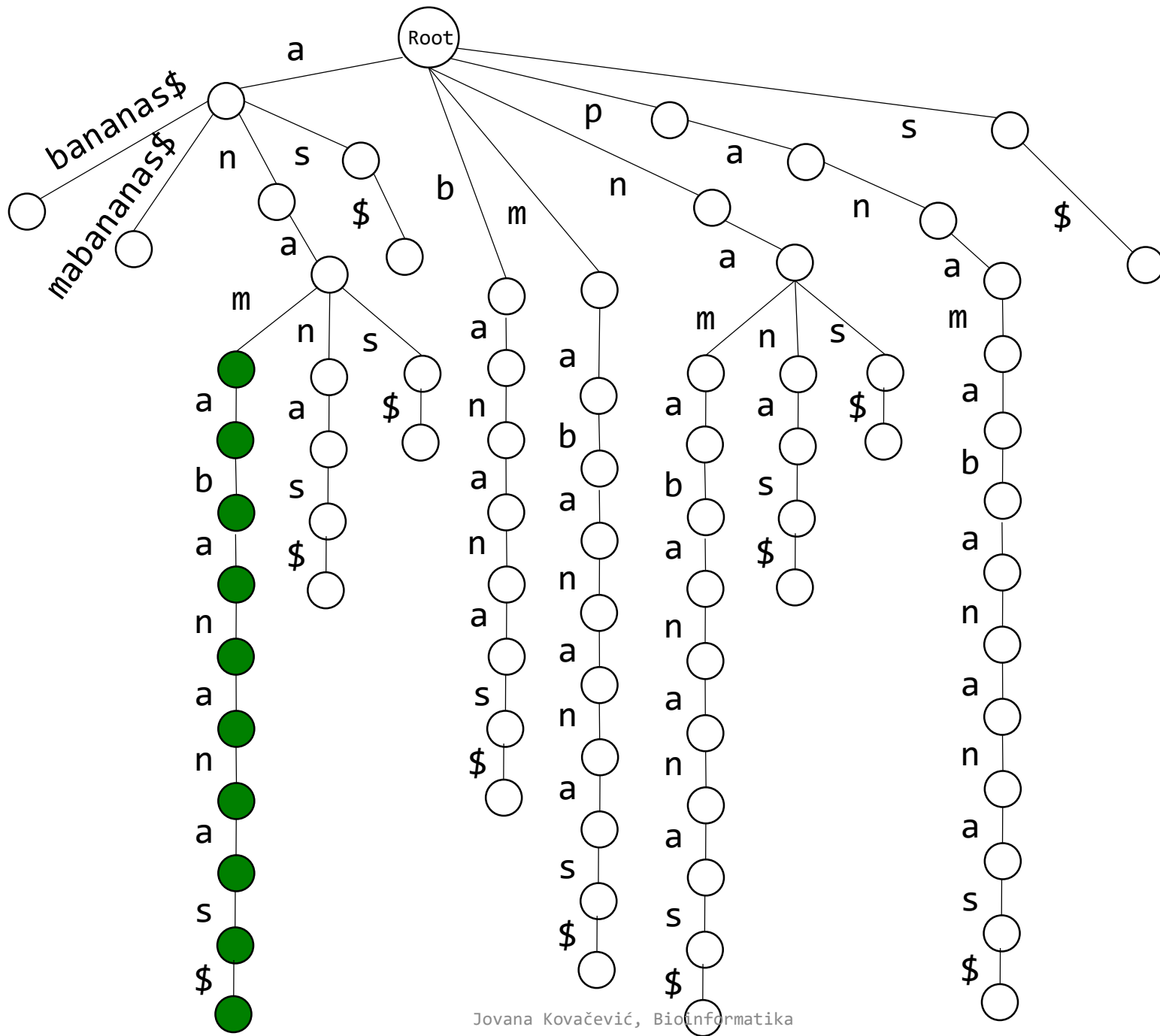
panamabananas\$
anamabananas\$
namabananas\$
amabananas\$
mabananas\$
abananas\$
bananas\$
ananas\$
nanas\$
anas\$
nas\$
as\$
s\$
\$

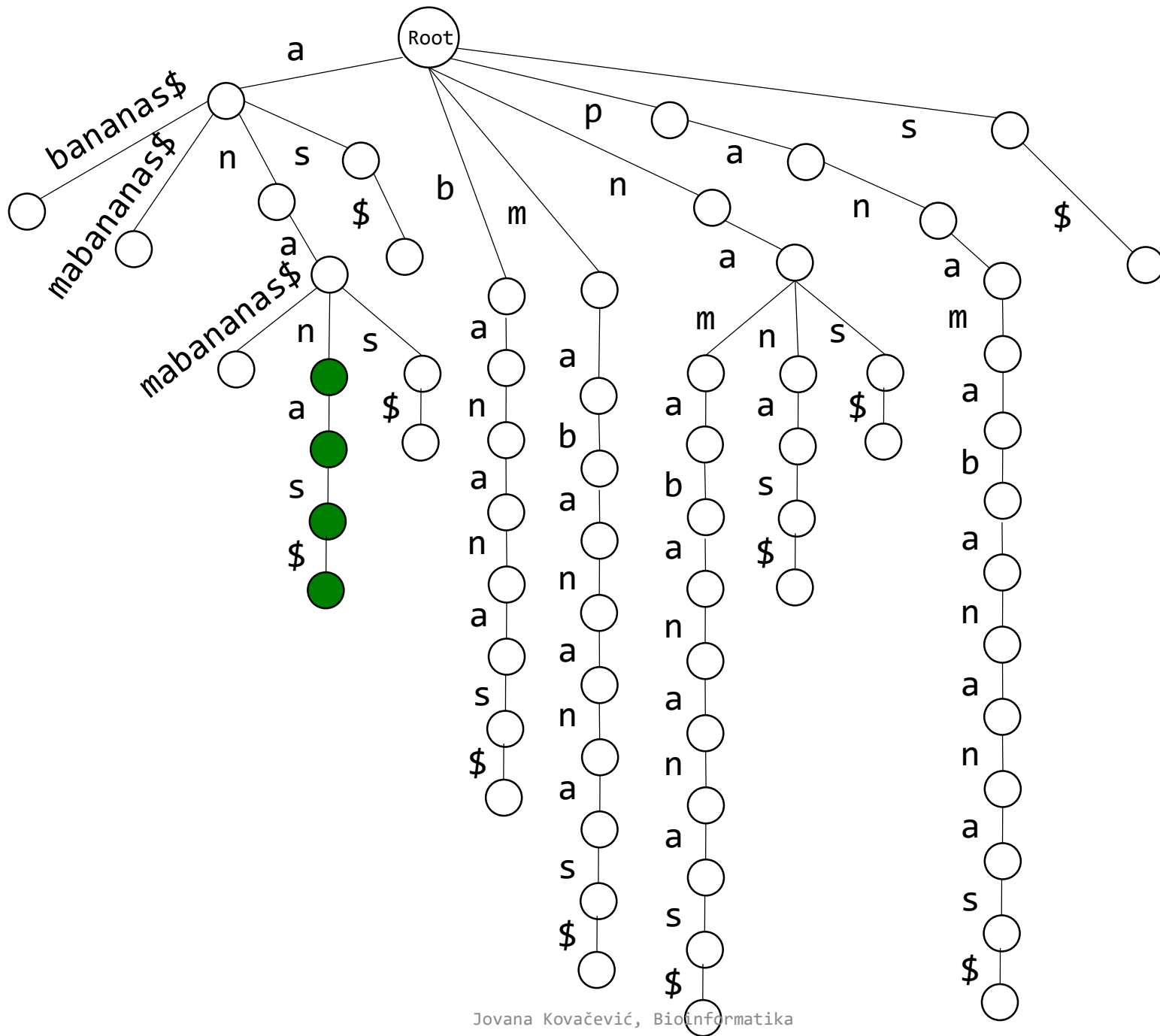
Kompresija

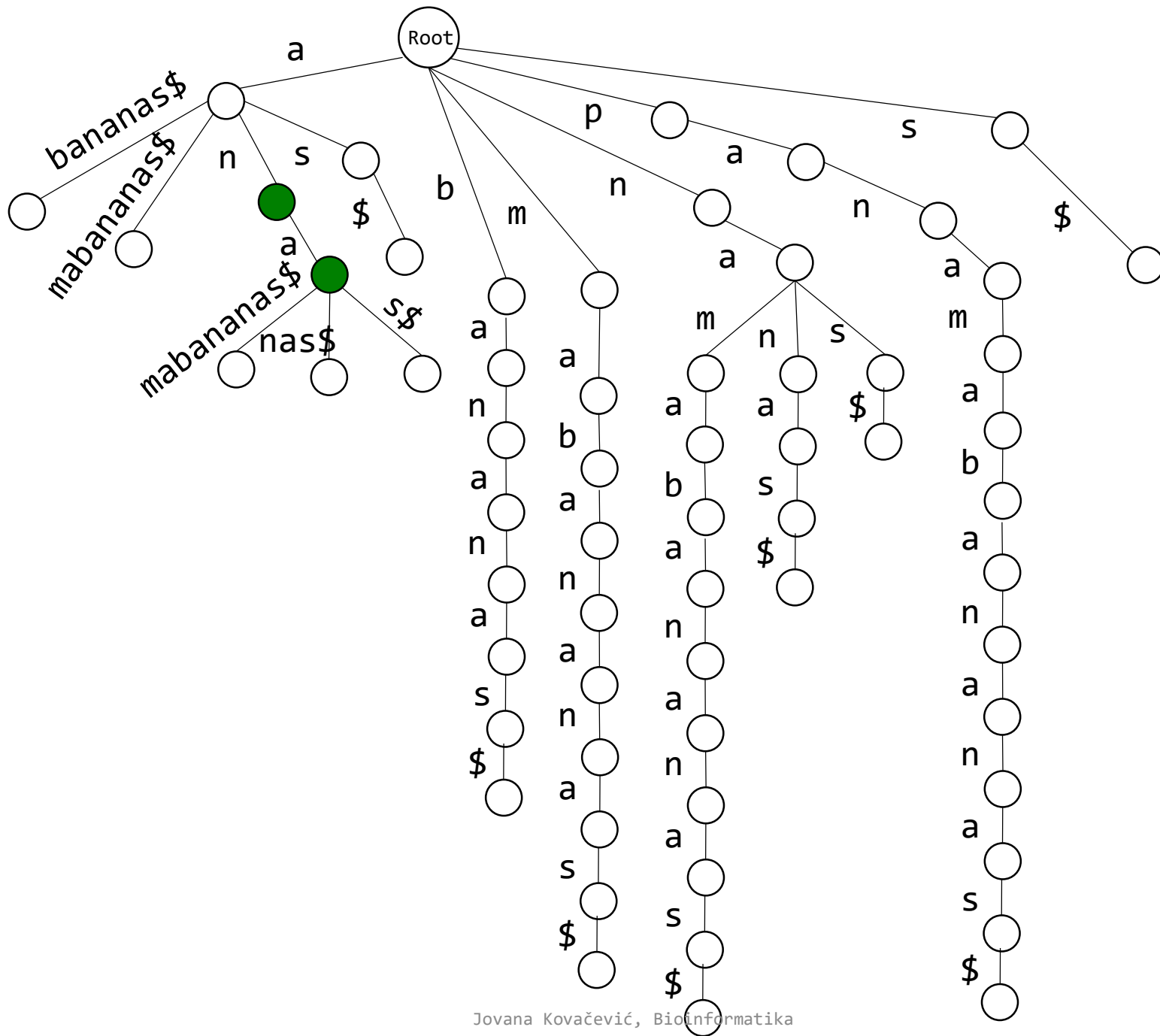
- Da bismo smanjili prostornu složenost, možemo kompresovati svaku putanju *koja se ne grana* u jednu granu.

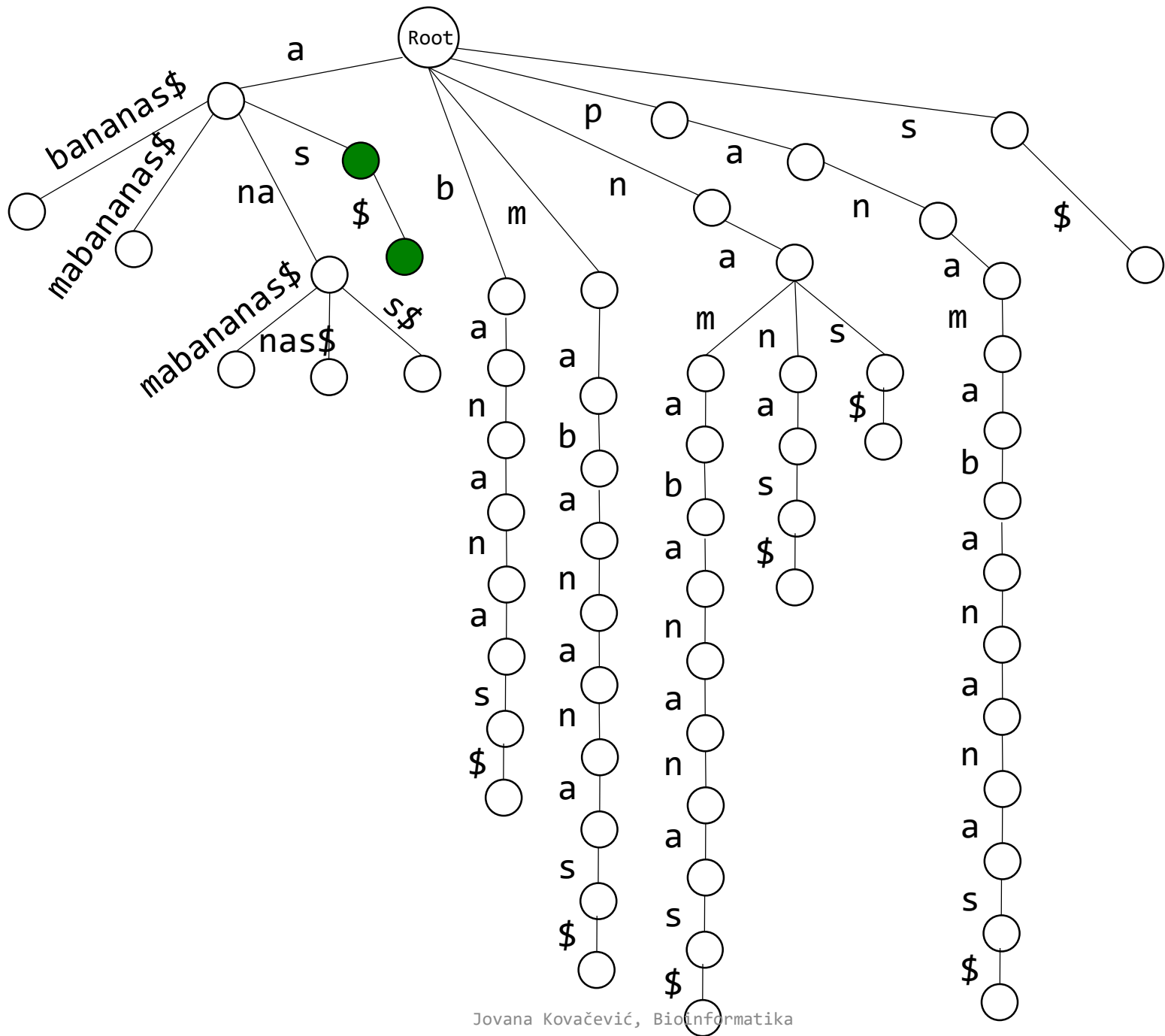


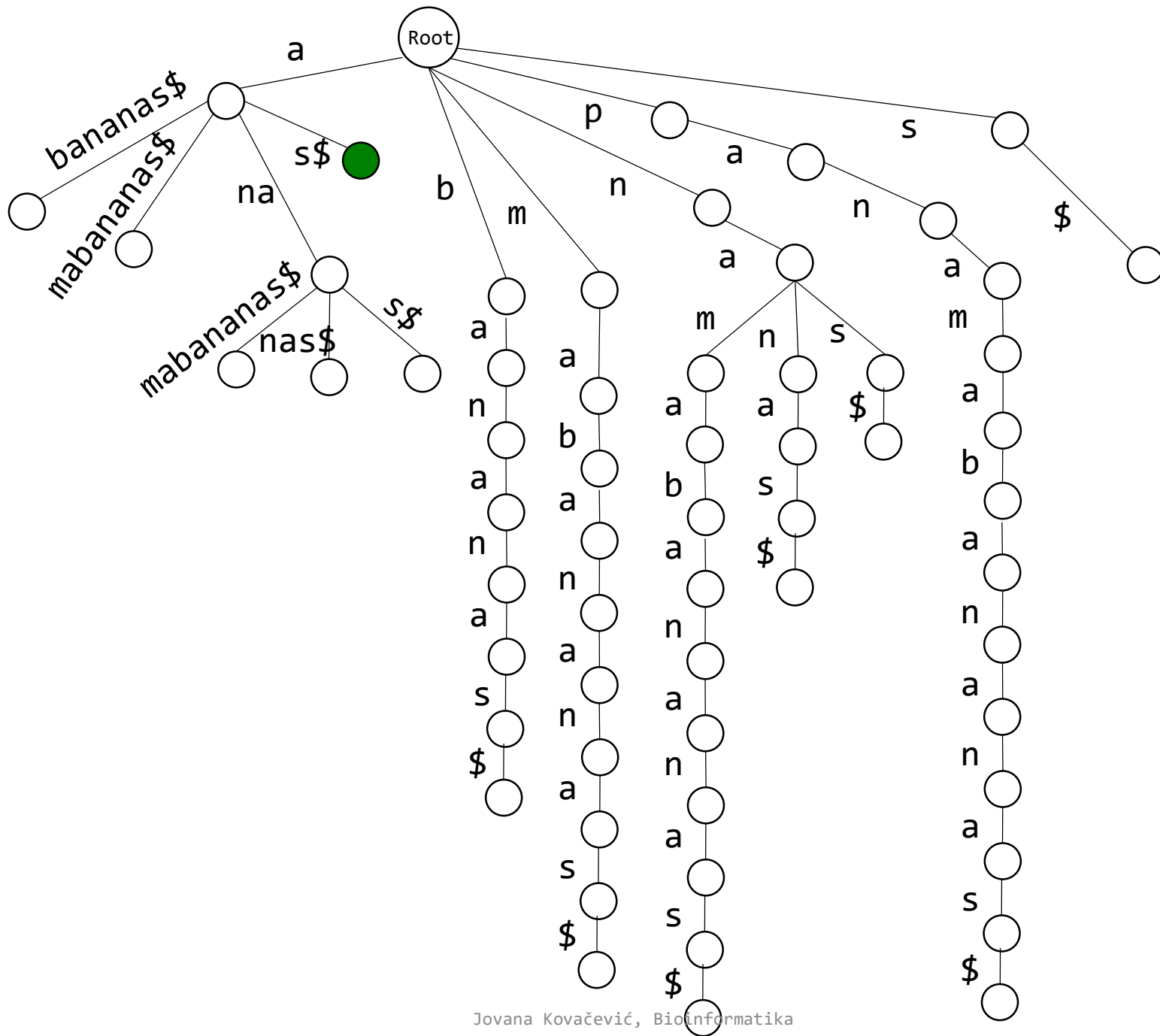


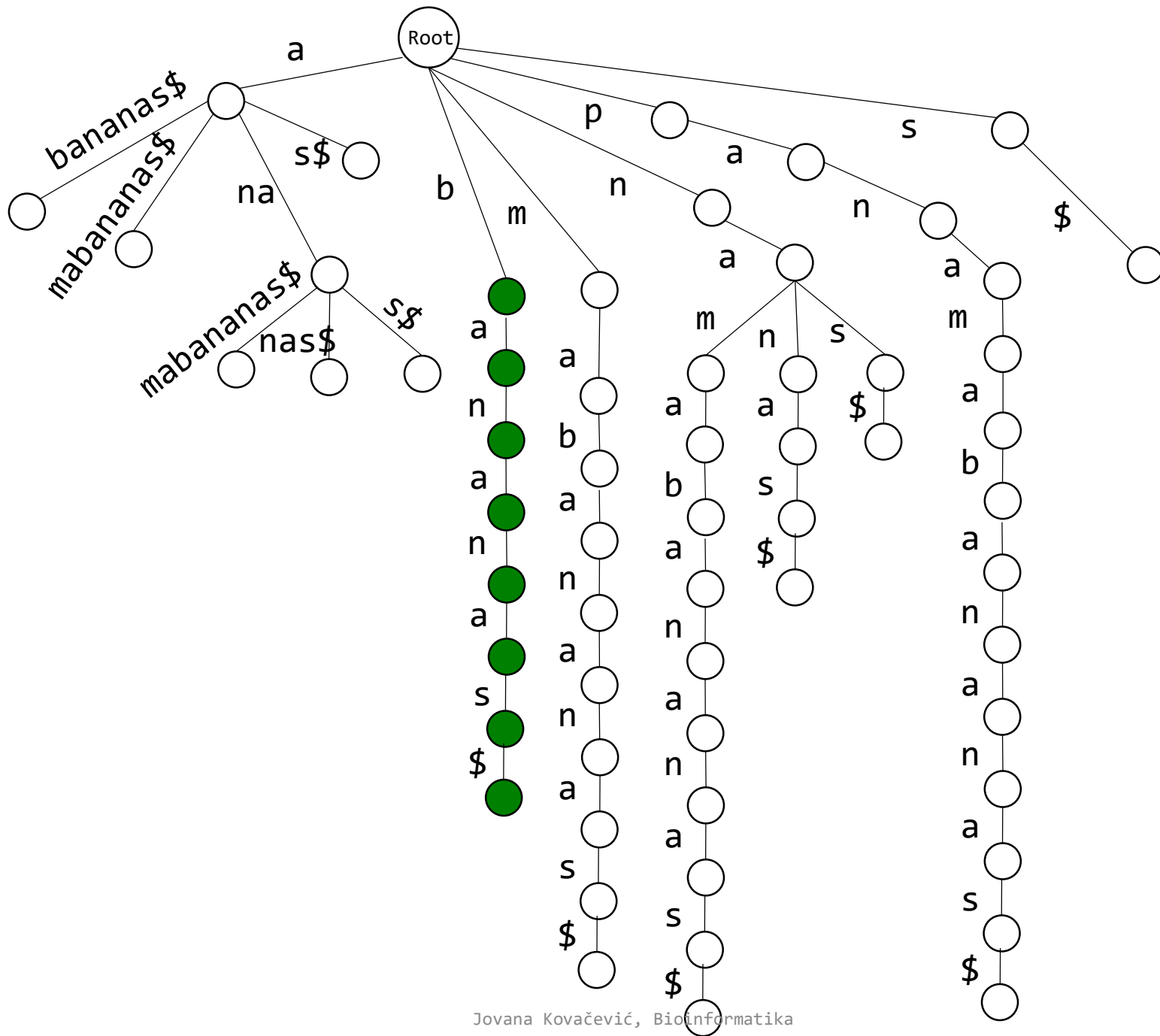


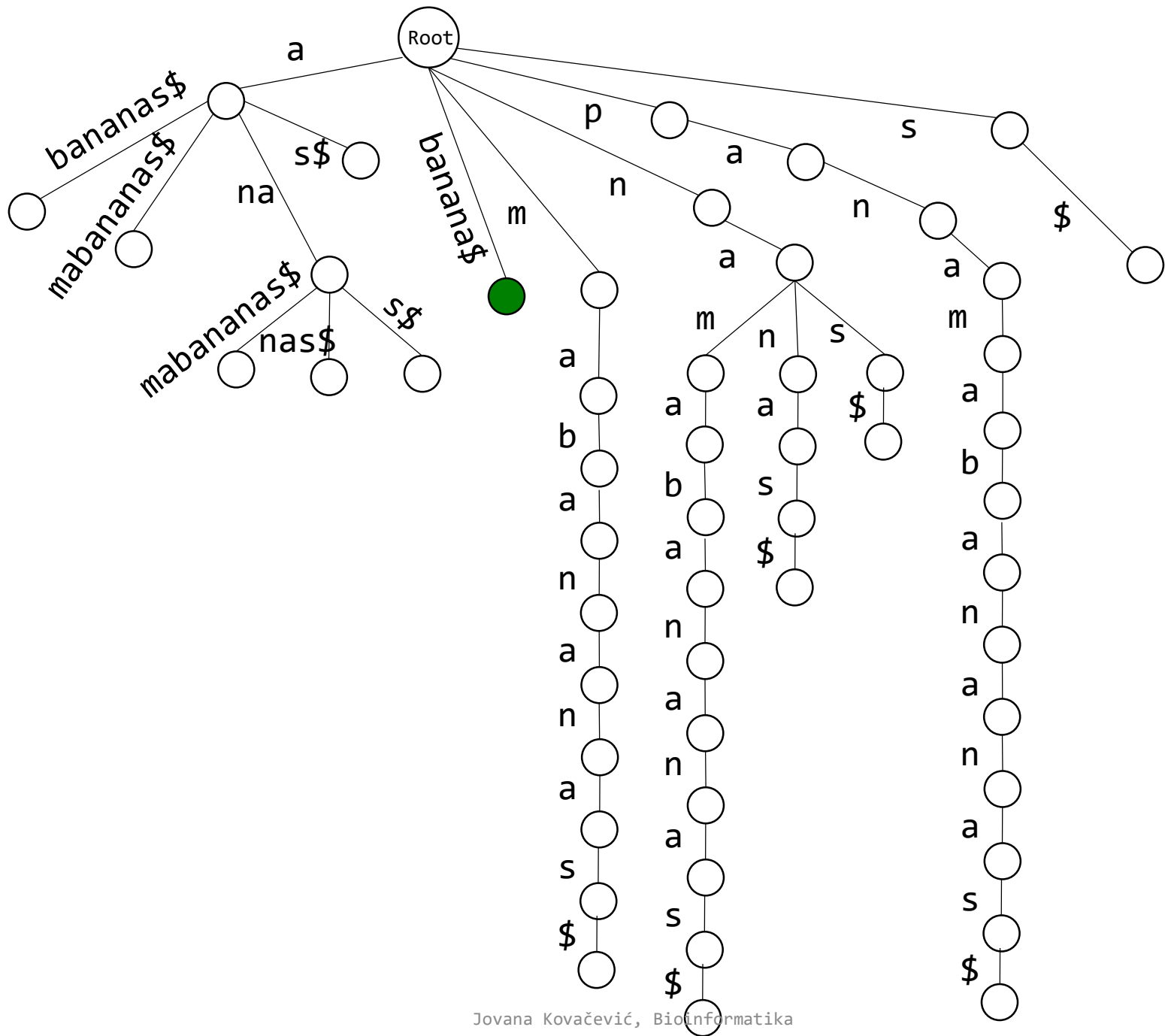


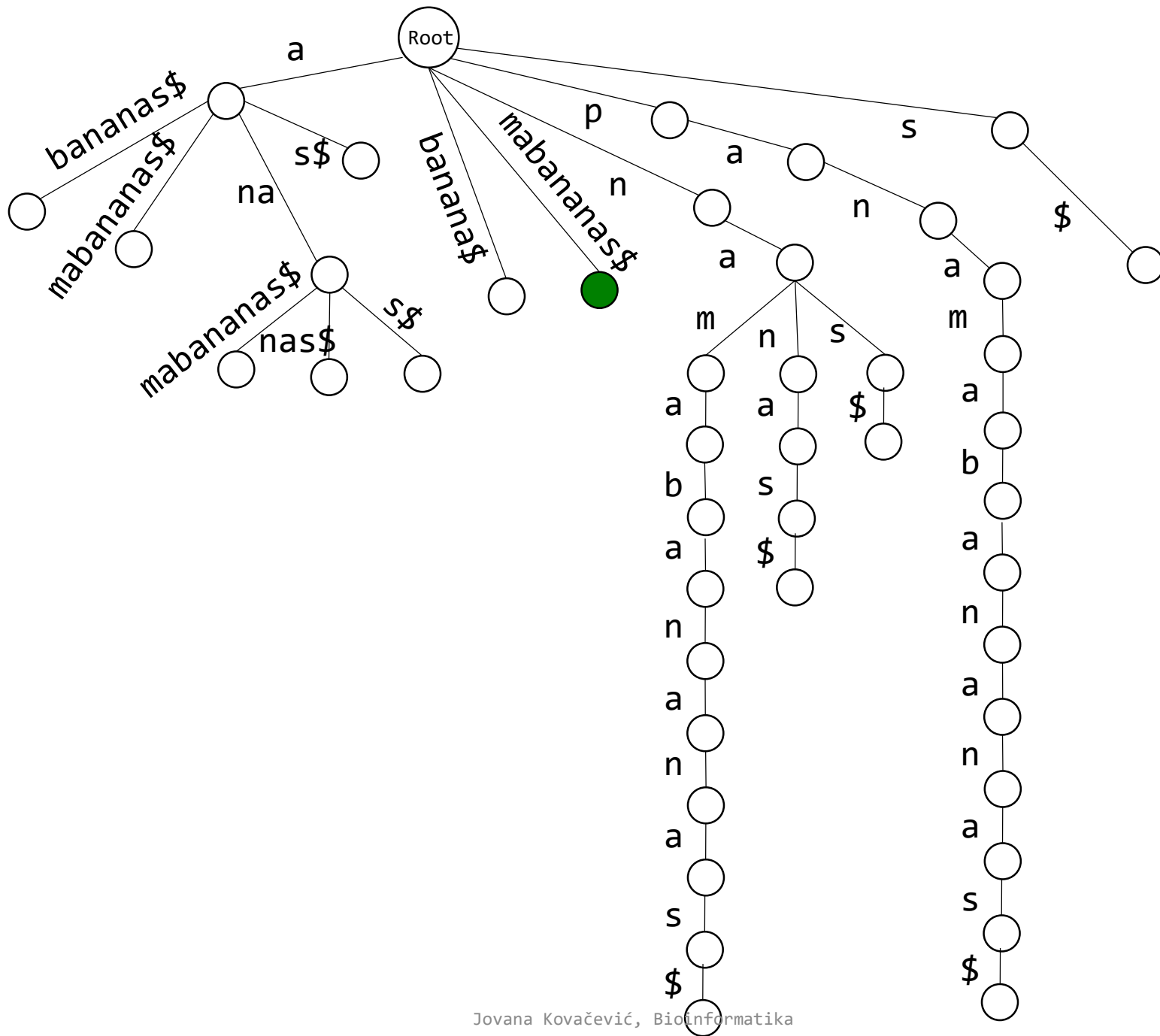


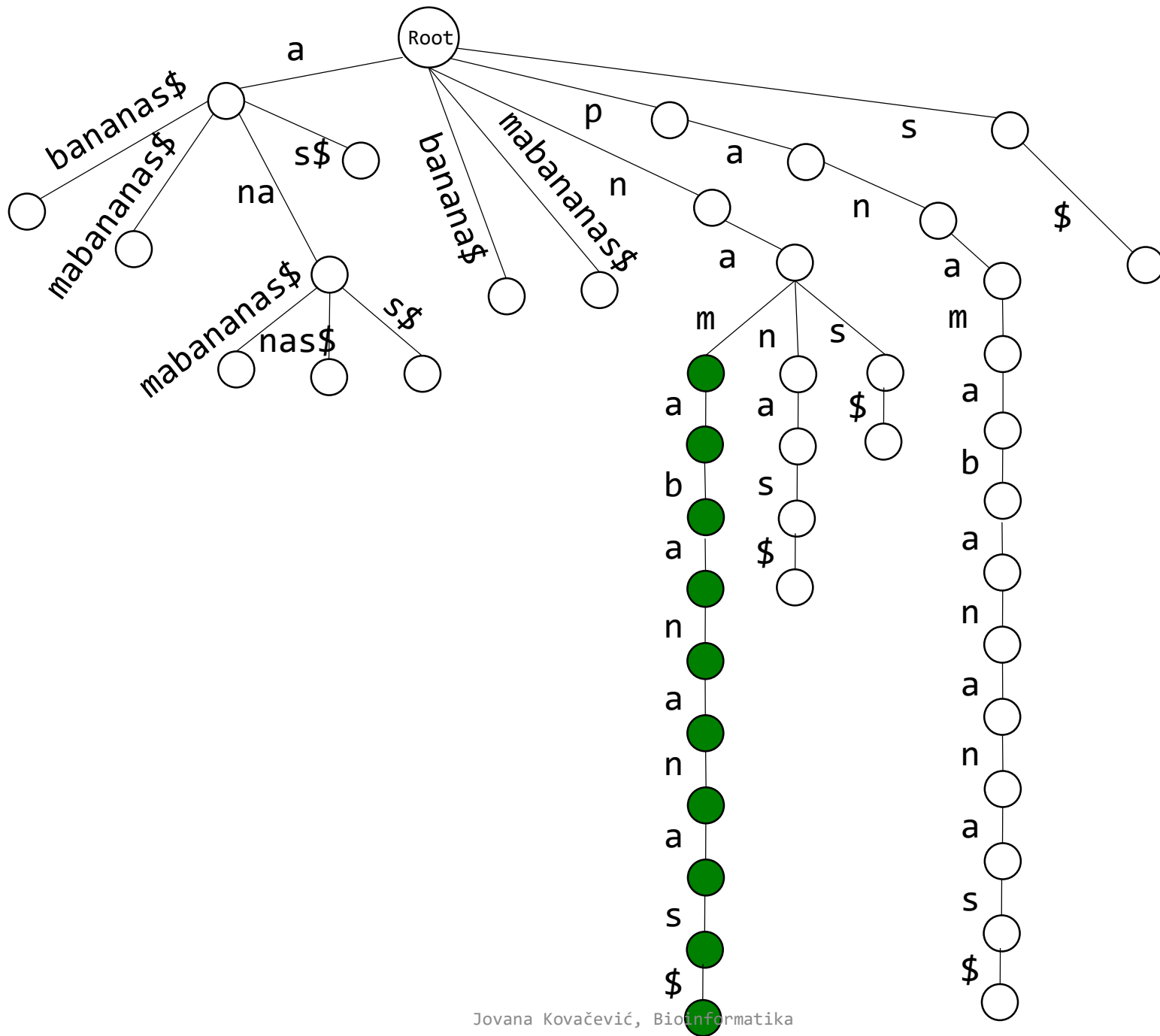


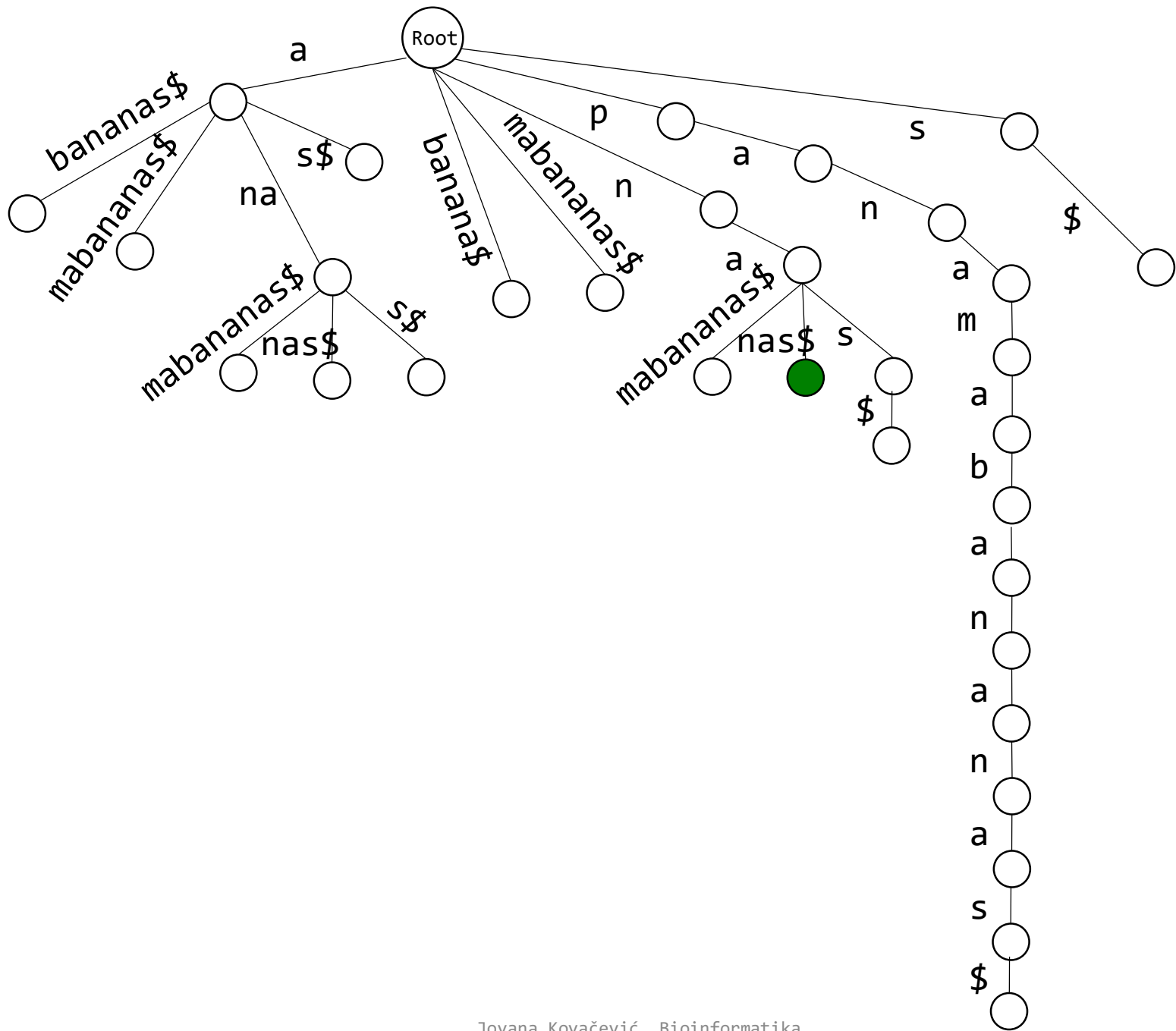


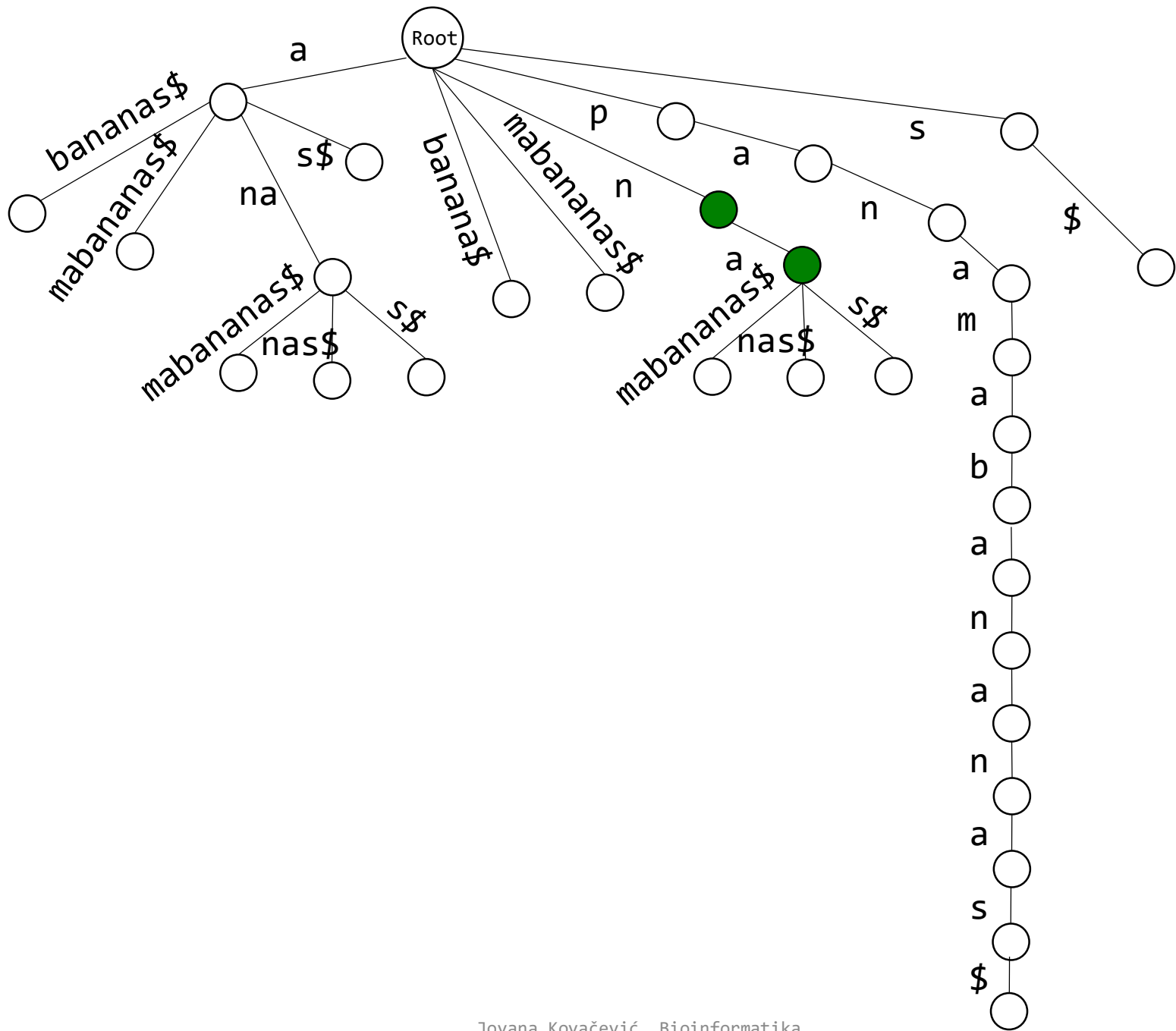


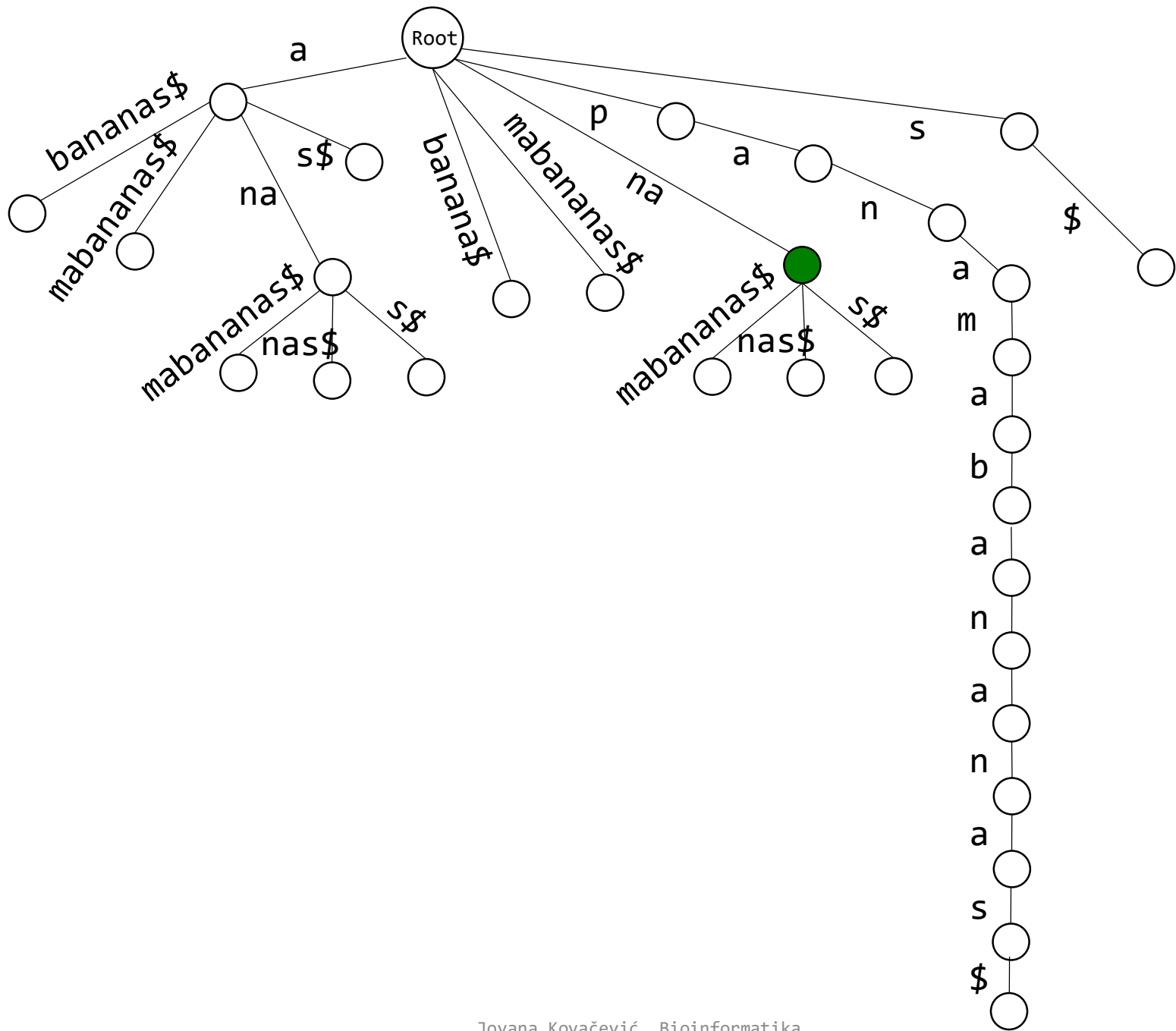


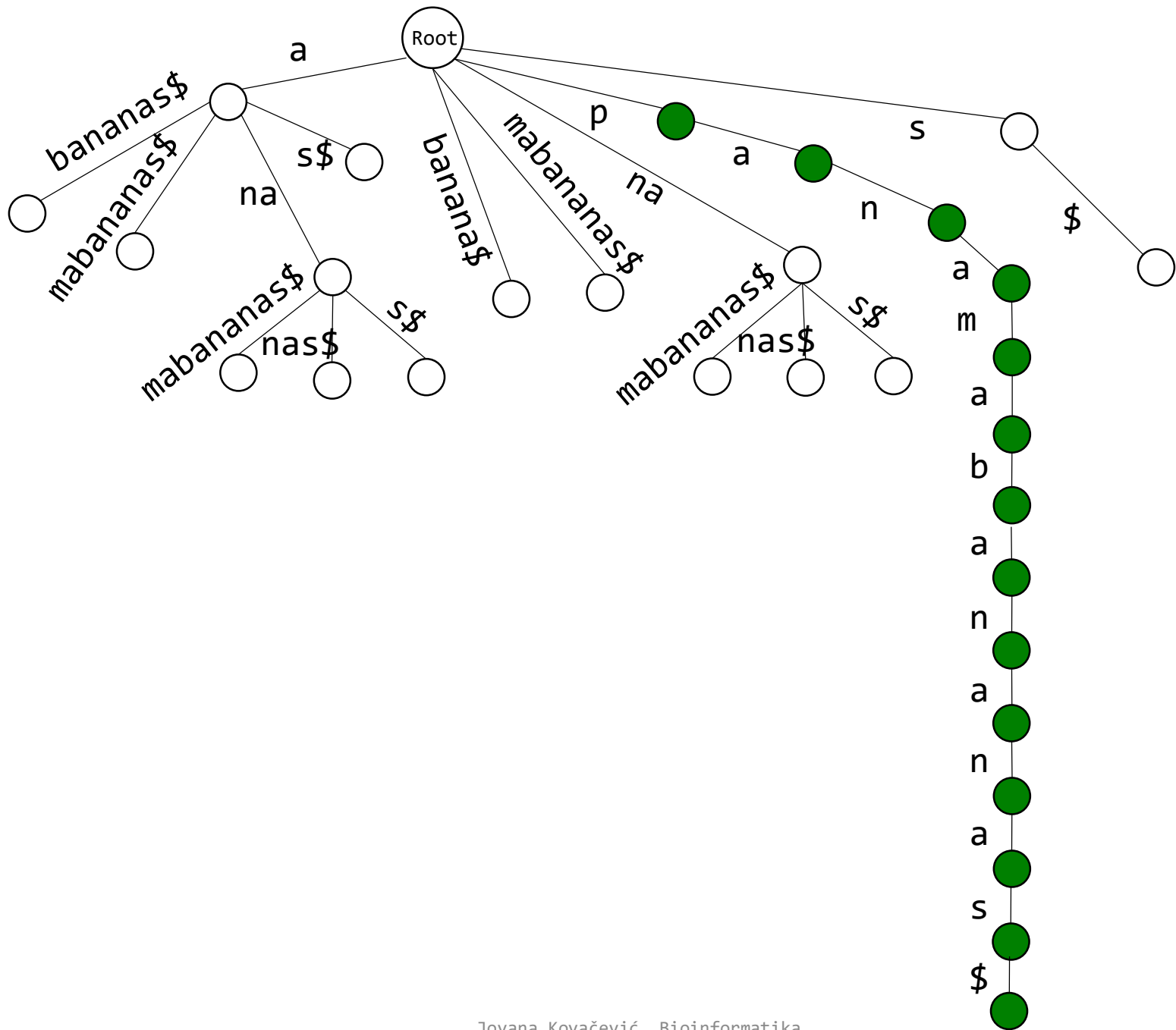


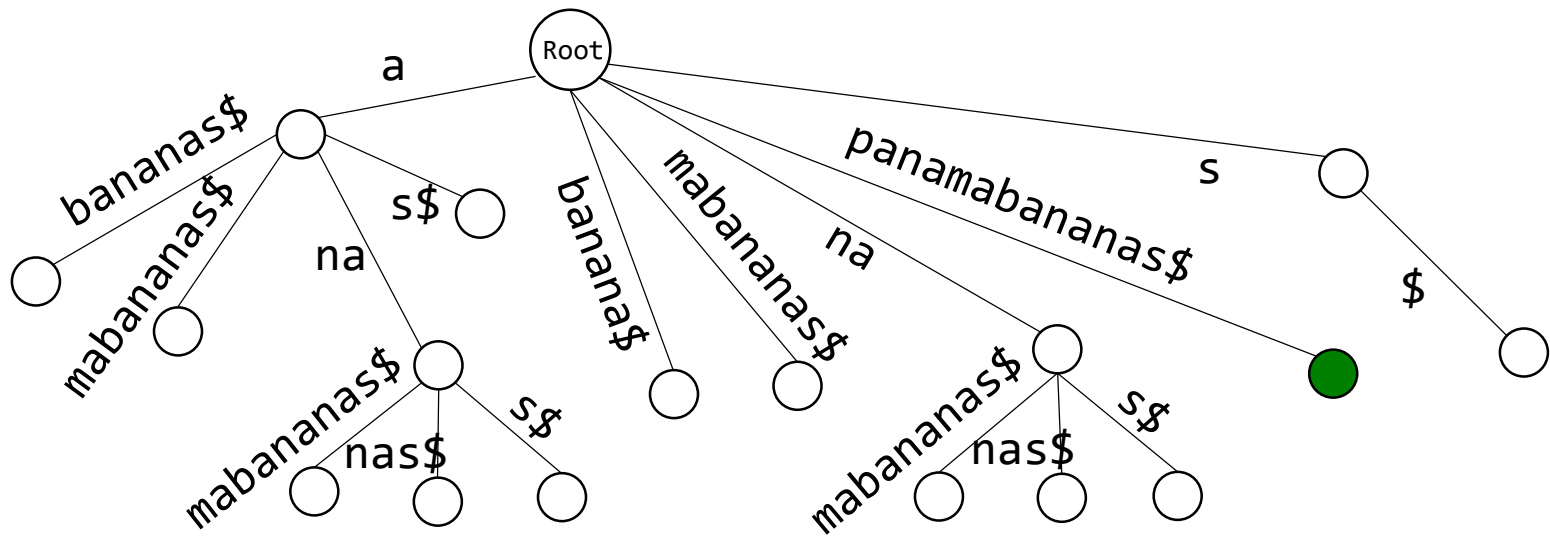


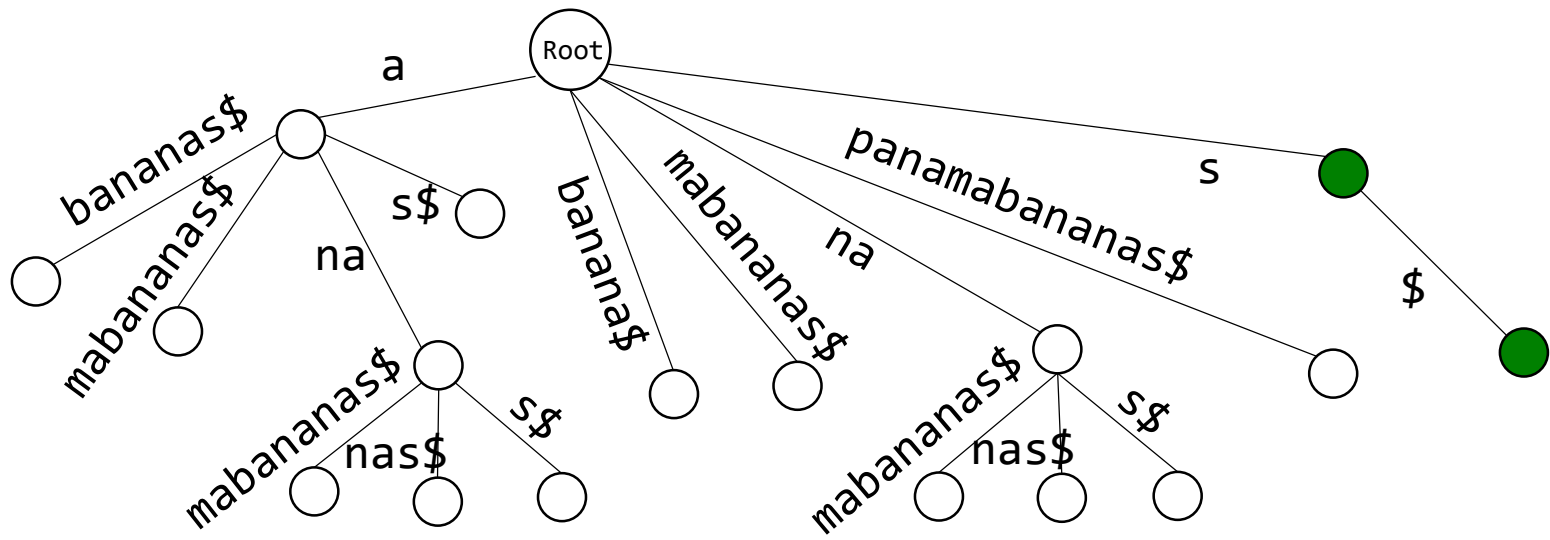


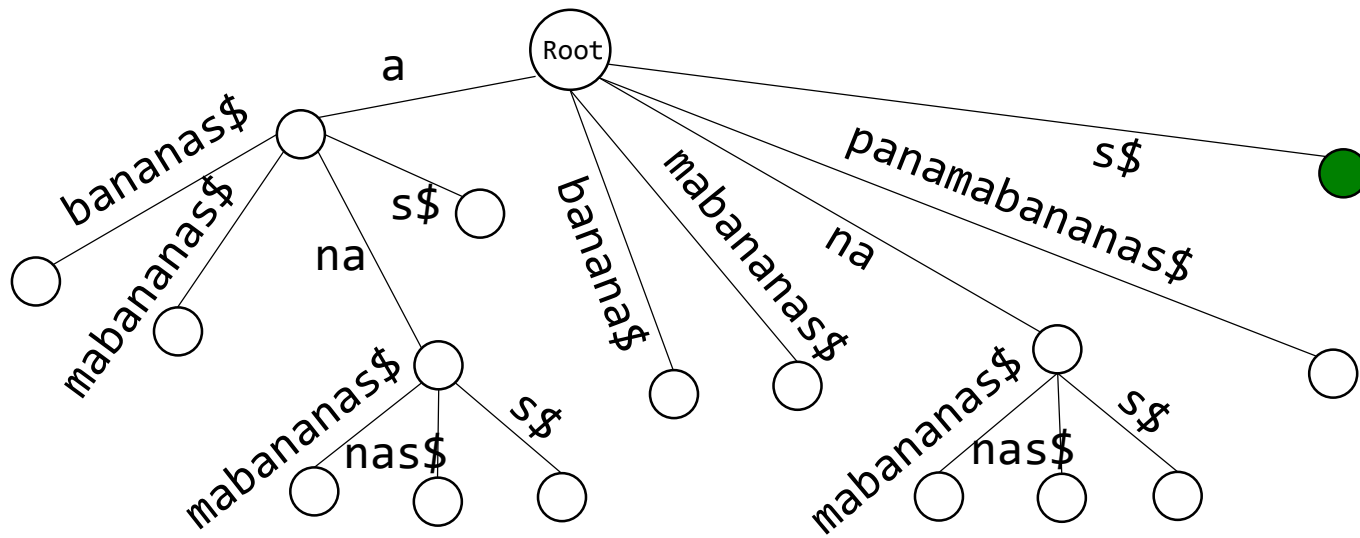


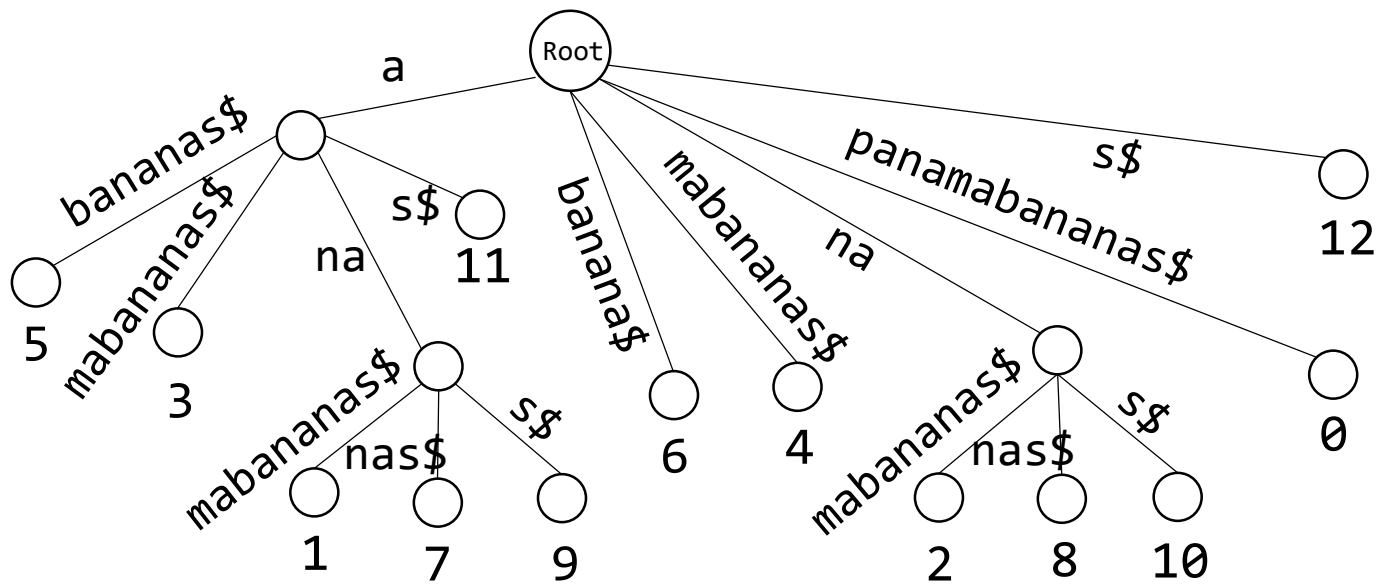




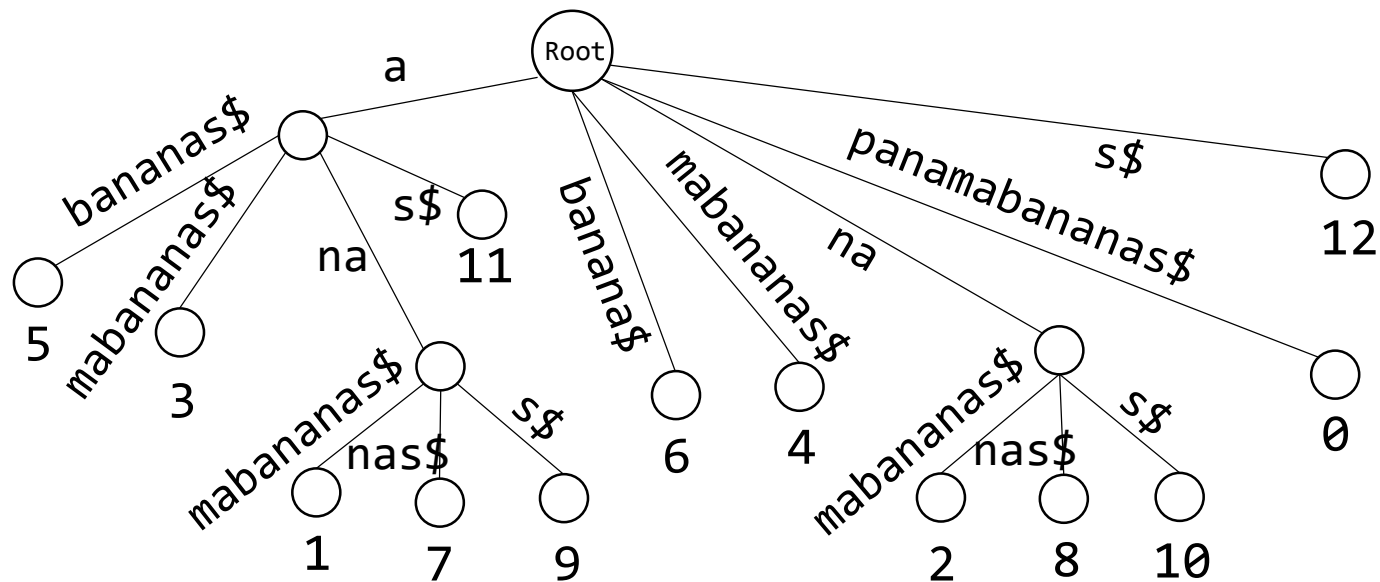




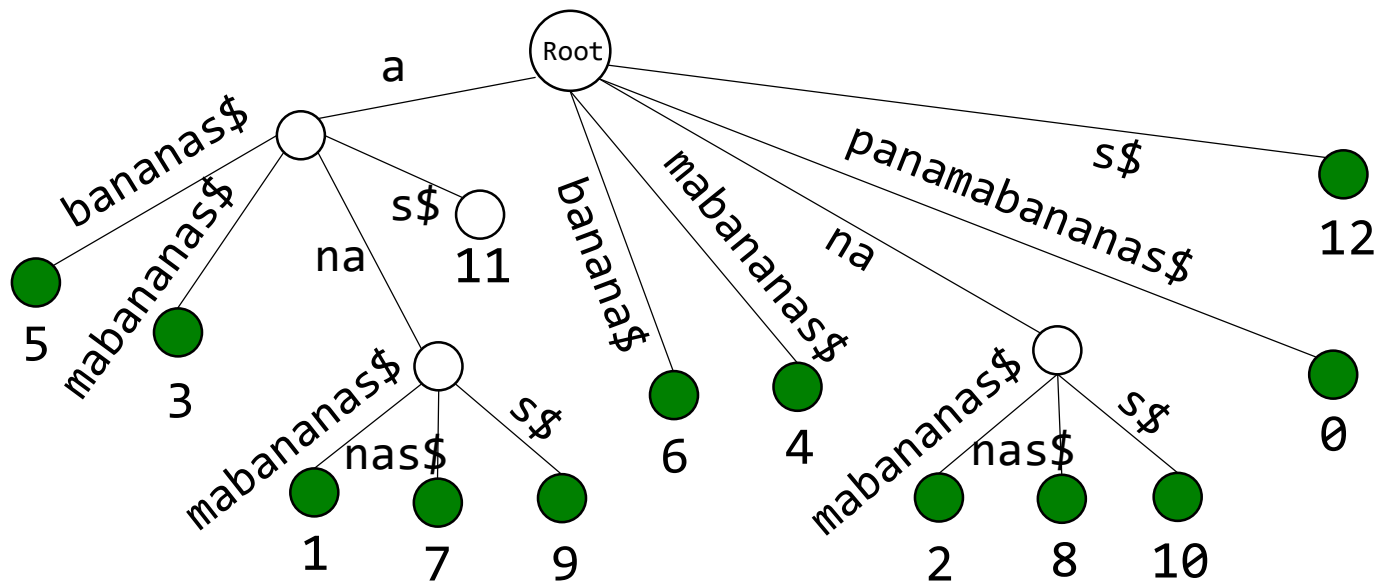




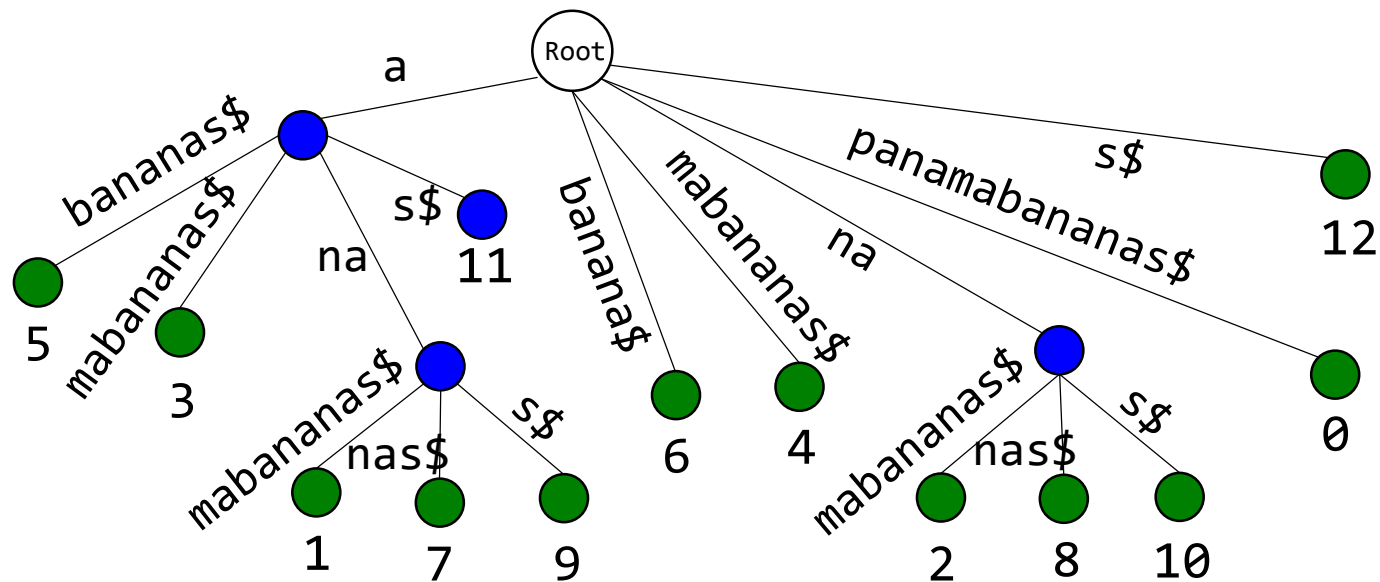
- Ovakva struktura podataka se naziva **sufiksno stablo**.



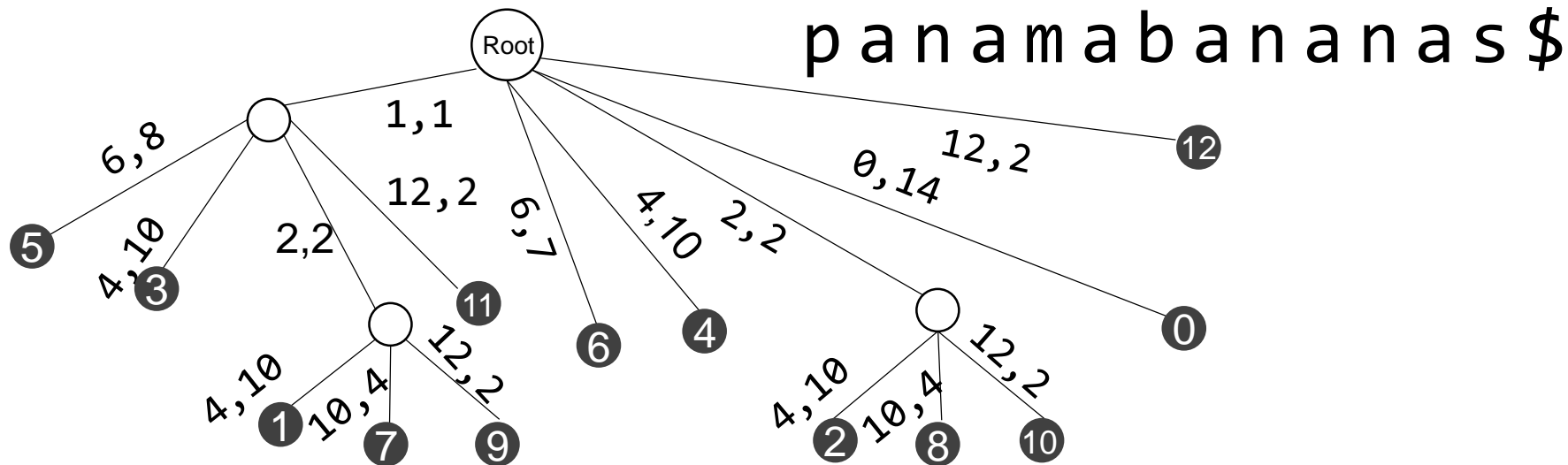
- Ovakva struktura podataka se naziva **sufiksno stablo**.
- Za svaku nisku *Genome*, $\# \text{ nodes} < 2|Genome| = O(|Genome|)$.



- Ovakva struktura podataka se naziva **sufiksno stablo**.
- Za svaku nisku *Genome*, $\# \text{ nodes} < 2|Genome|$.
 - **# leaves = |Genome|;**



- Ovakva struktura podataka se naziva **sufiksno stablo**.
- Za svaku nisku *Genome*, $\# \text{ nodes} < 2|Genome|$.
 - **# leaves** = $|Genome|$;
 - **# internal nodes** $\leq |Genome| - 1$



Umesto podniski, grane označimo uređenim parovima
 gde je prvi član indeks početka sufiksa a drugi
 njegova dužina. Uz to, čuvamo celo tekst
 panamabanas\$

Prostorna i vremenska složenost

- Vremenska složenost:
 - $O(|Genome|^2)$ za konstrukciju sufiksnog stabla tako što se prvo konstruiše nekompresovano sufiksno stablo
 - $O(|Patterns|)$ za nalaženje uparivanja.
- Prostorna složenost:
 - $O(|Genome|^2)$ za konstrukciju sufiksnog stabla tako što se prvo konstruiše nekompresovano sufiksno stablo
 - $O(|Genome|)$ za čuvanje sufiksnog stabla.

Prostorna i vremenska složenost

- Postoje algoritmi sa linearnom prostornom i vremenskom složenošću
- Vremenska složenost:
 - $O(|Genome|)$ za konstrukciju sufiksnog stabla *direktno*.
 - $O(|Patterns|)$ za nalaženje uparivanja.
- Prostorna složenost:
 - $O(|Genome|)$ za konstrukciju sufiksnog stabla *direktno*.
 - $O(|Genome|)$ za čuvanje sufiksnog stabla.

Prostorna i vremenska složenost

- Postoje algoritmi sa linearnom prostornom i vremenskom složenošću
- Vremenska složenost:
 - $O(|Genome|)$ za konstrukciju sufiksnog stabla *direktno*.
 - $O(|Patterns|)$ za nalaženje uparivanja.
 - Ukupno: $O(|Genome| + |Patterns|)$
- Prostorna složenost:
 - $O(|Genome|)$ za konstrukciju sufiksnog stabla *direktno*.
 - $O(|Genome|)$ za čuvanje sufiksnog stabla.
 - Ukupno: $O(|Genome|)$

Pregled

- Mapiranje očitavanja
- Sufiksna stabla
- **Kompresija niski i Barouz-Vilerova transformacija**
- ...

Prostor za unapređenje

- Složenost:
 - Vremenska: $O(|Genome| + |Patterns|)$
 - Prostorna: $O(|Genome|)$
- O-notacija ignoriše konstante!
 - Najpoznatija implementacija sufiksni stabala zahteva ~ 20 puta $|Genome|$
 - (npr. Veličina humanog genoma je 3GB \Rightarrow 60 GB; i dalje unapređenje u odnosu na 1TB).
 - Da li možemo da smanjimo faktor konstante?

Kompresija genoma

- Ideja: smanjiti količinu memorije neophodnu za čuvanje niske *Genome*.
- Potrebne su metode za **kompresiju** niske velike dužine što je naizgled sasvim drugačiji problem.

Kompresija genoma

Genome 1

GGGGGGGGGGCCCCCCCCCCCCAAAAAAATTTTTTTTTTTTTTTTTTTTCCCCCG

U ovom genomu imamo nekoliko uzastopnih ponavljanja jedne aminokiseline (ranovi, *runs*): prvo uzastopna ponavljanja aminokiseline G, pa C i tako dalje)

Kompresija genoma

Genome 2

GACGACGACGAC**CATTCATTCATCATT****ACGTAGACGTAGCACCCC**

U ovom genomu imamo nekoliko uzastopnih ponavljanja niza aminokiselina (*ripiti*, *repeats*): prvo uzastopna ponavljanja GAC, pa CATT i tako dalje

Genome 1

GGGGGGGGGGCCCCCCCCCCCCAAAAAATTTTTTTTTTTTTTTTTTTTCCCCCG

Genome 2

GACGACGACGACCATTCATTCATCATTACGTAGACGTAGCACCCC

Ideja #1: Kodiranje dužine ranova

- Kompresija ponavljanja n identičnih simbola.

Genome

GGGGGGGGGGCCCCCCCCCCAAAAAATTTTTTTTTTTTTTTTTTCCCCCG



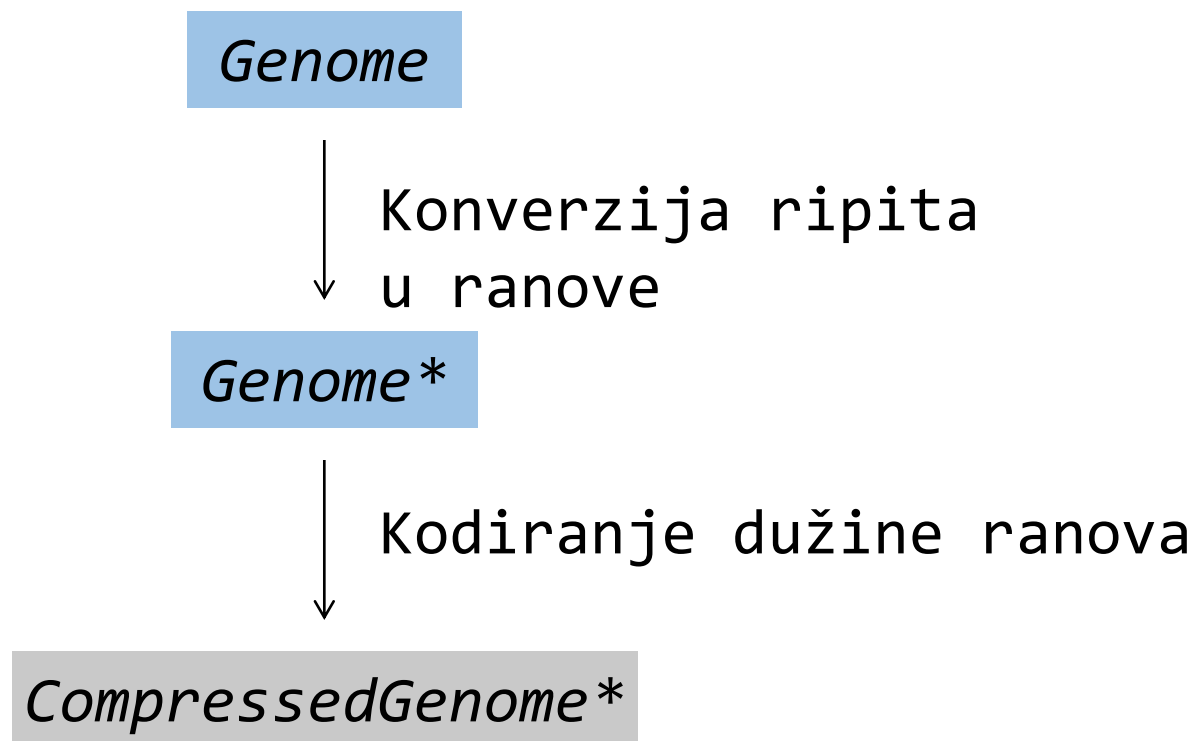
10G11C7A15T5C1G

Run-length encoding

- Problem: Kod genoma nema mnogo ranova

Konverzija ripita u ranove

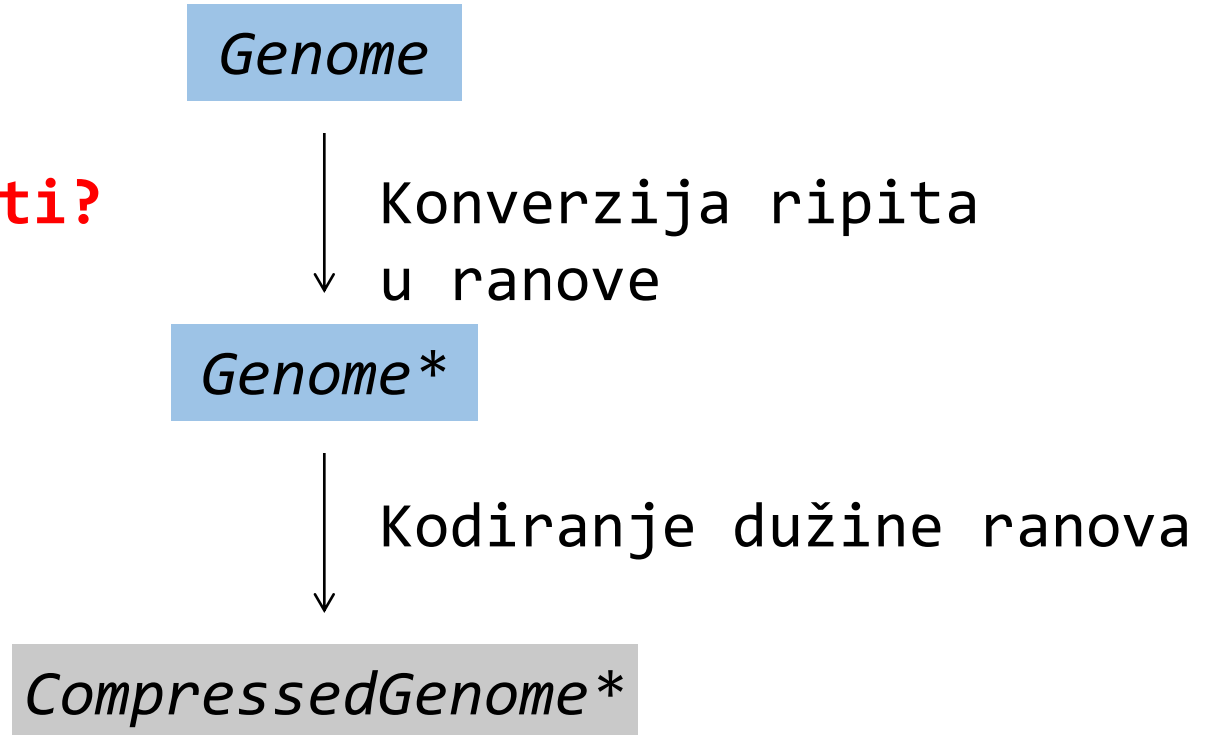
...ali ima mnogo ripita!



Konverzija ripita u ranove

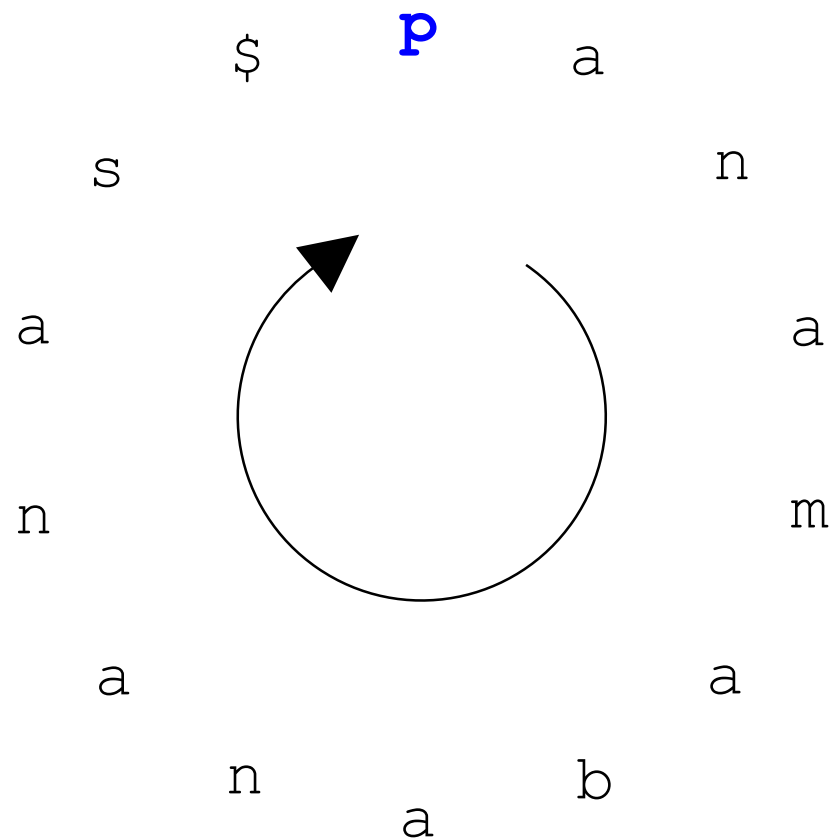
..ali ima mnogo ripita!

Kako ovo izvesti?



BWT – Burrows-Wheeler transformacija

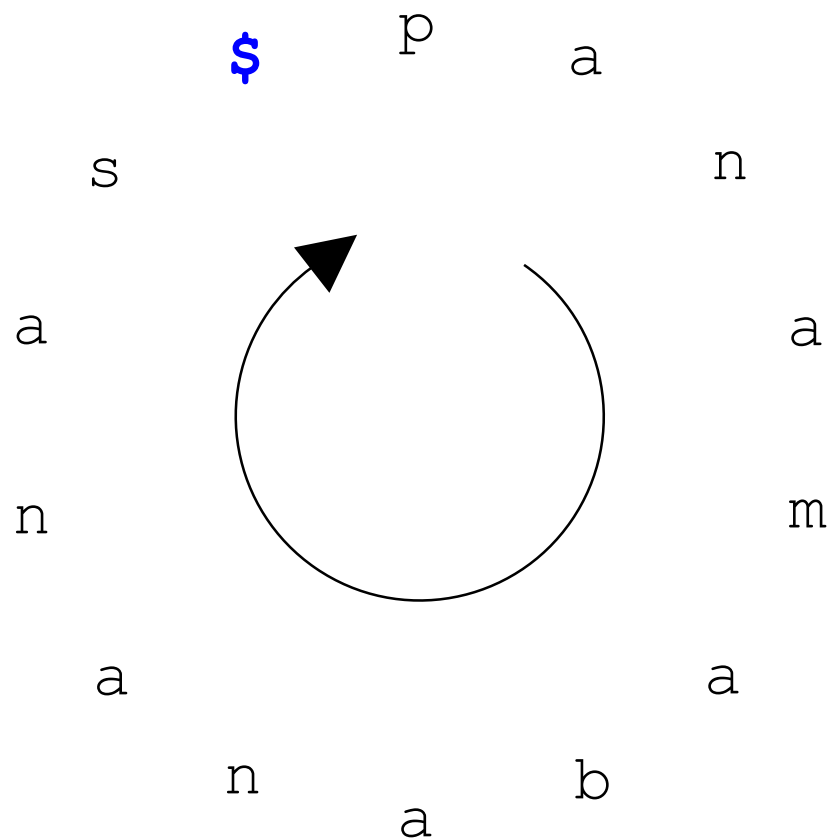
panamabananas\$



Formirati sve ciklične
rotacije niske
“panamabananas\$”

BWT – Burrows-Wheeler transformacija

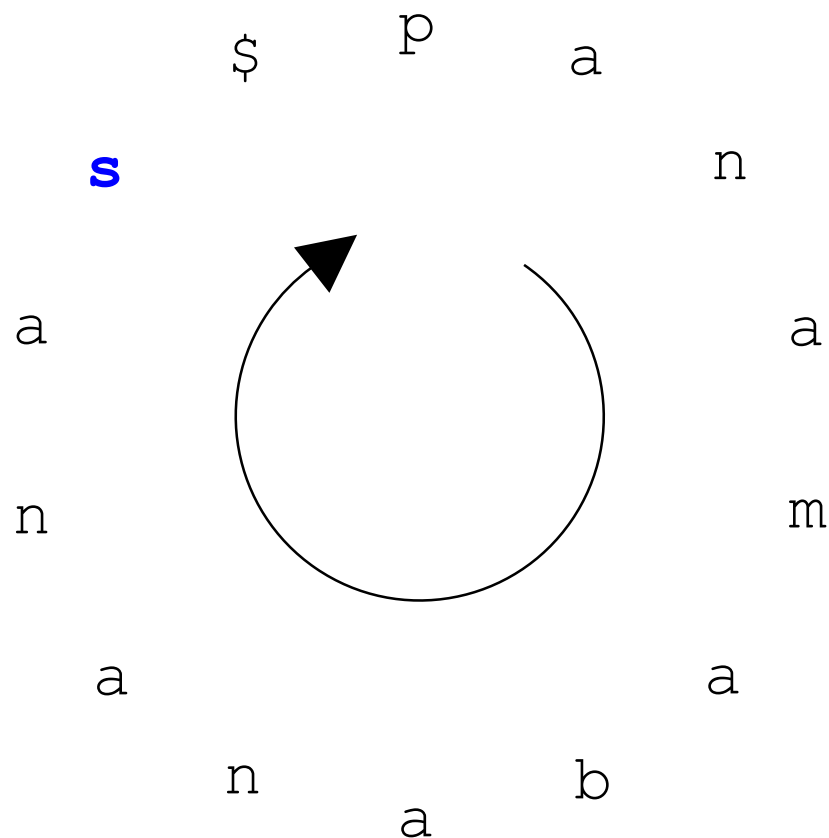
panamabananas\$
\$panamabananas



Formirati sve ciklične rotacije niske "panamabananas\$"

BWT – Burrows-Wheeler transformacija

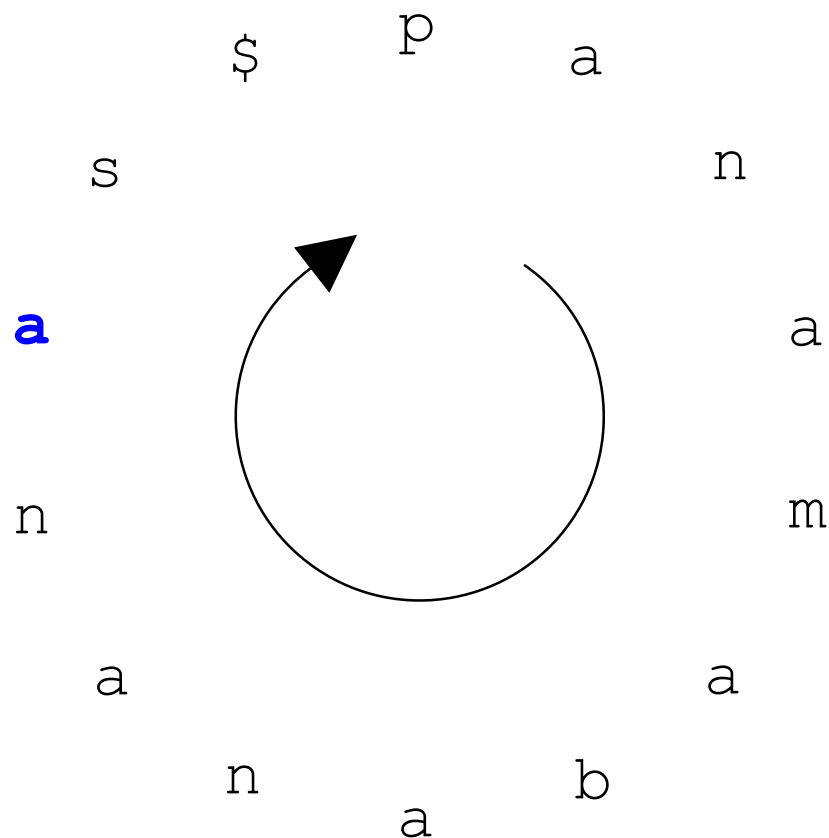
```
panamabananas$  
$panamabananas  
s$panamabana
```



Formirati sve ciklične rotacije niske
"panamabananas\$"

BWT – Burrows-Wheeler transformacija

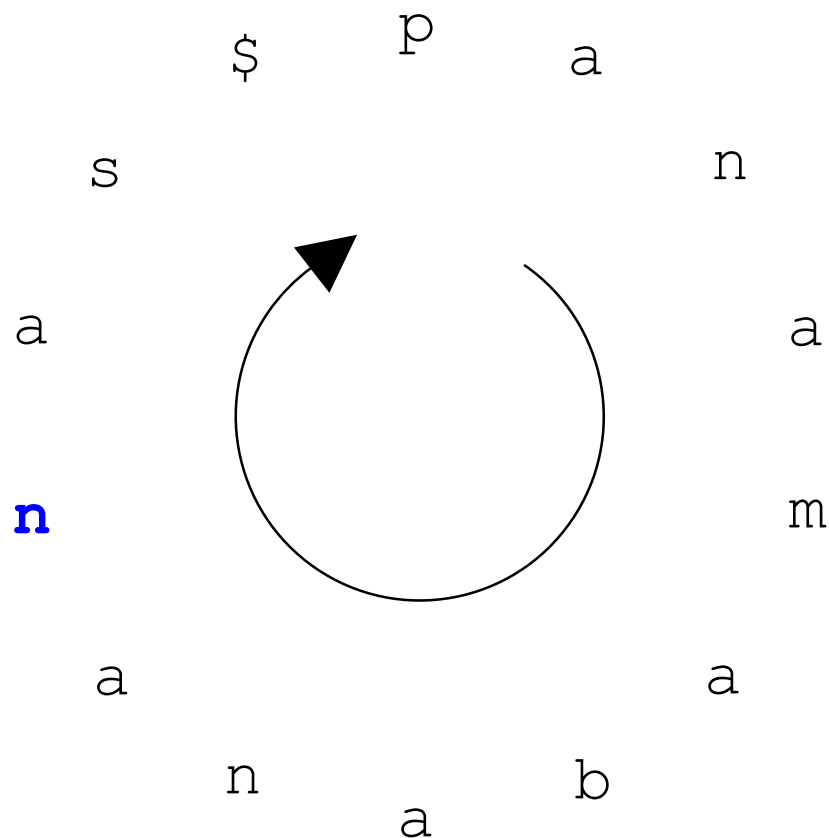
```
panamabananas$  
$panamabananas  
s$panamabanana  
as$panamaban
```



Formirati sve ciklične rotacije niske "panamabananas\$"

BWT – Burrows-Wheeler transformacija

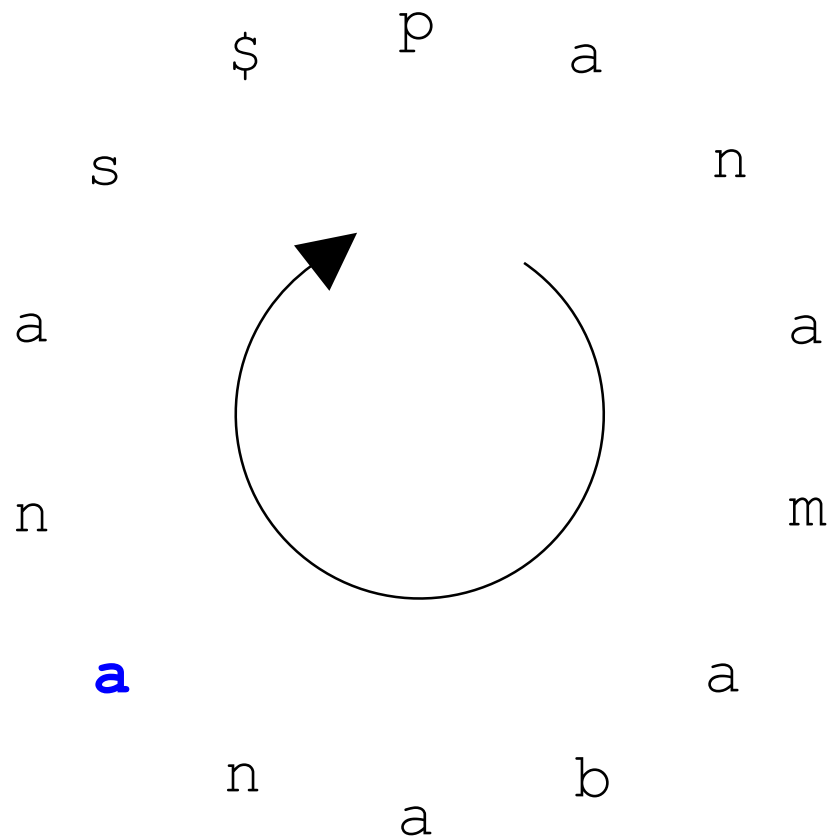
```
panamabananas$  
$panamabananas  
s$panamabanan  
as$panamabanan  
nas$panamabana
```



Formirati sve ciklične rotacije niske “panamabananas\$”

BWT – Barouz-Vilerova transformacija

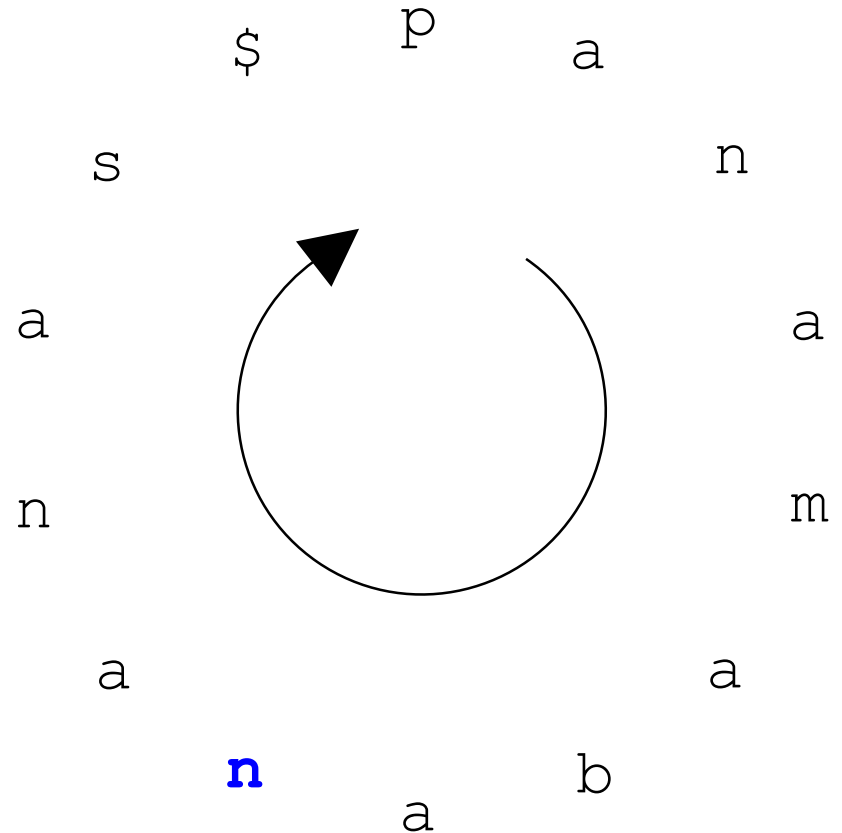
```
panamabananas$  
$panamabananas  
s$panamabanana  
as$panamabanana  
nas$panamabana  
anas$panamaban
```



Formirati sve ciklične rotacije niske
“panamabananas\$”

BWT – Barouz-Vilerova transformacija

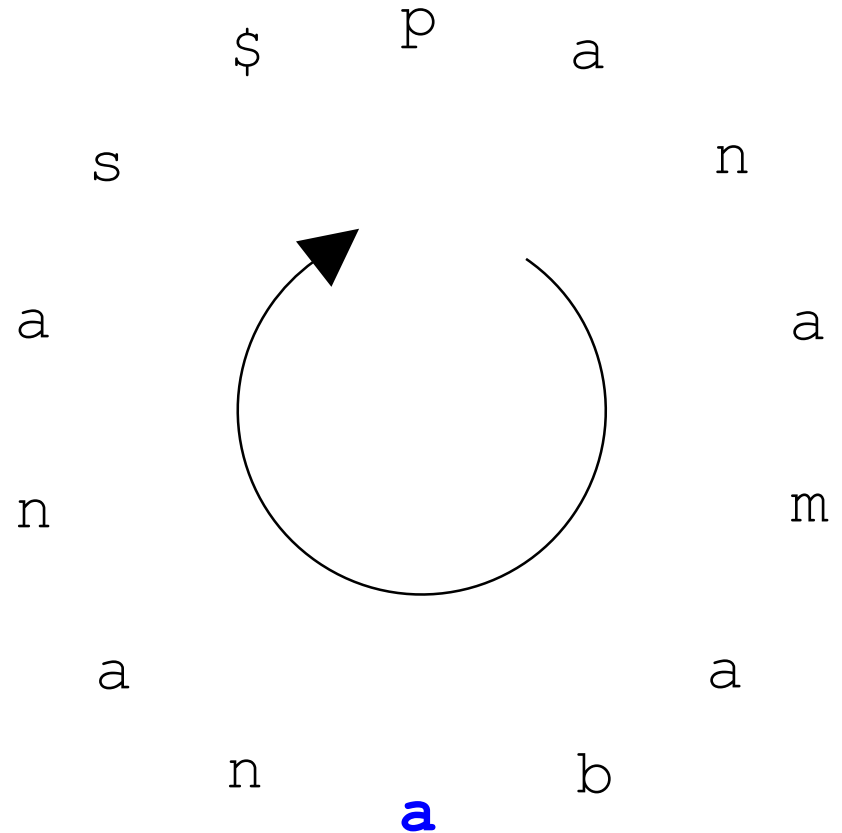
```
panamabananas$  
$panamabananas  
s$panamabanan  
as$panamabanan  
nas$panamabana  
anas$panamaban  
nanas$panamaba
```



Formirati sve ciklične rotacije niske "panamabananas\$"

BWT – Burrows-Wheeler transformacija

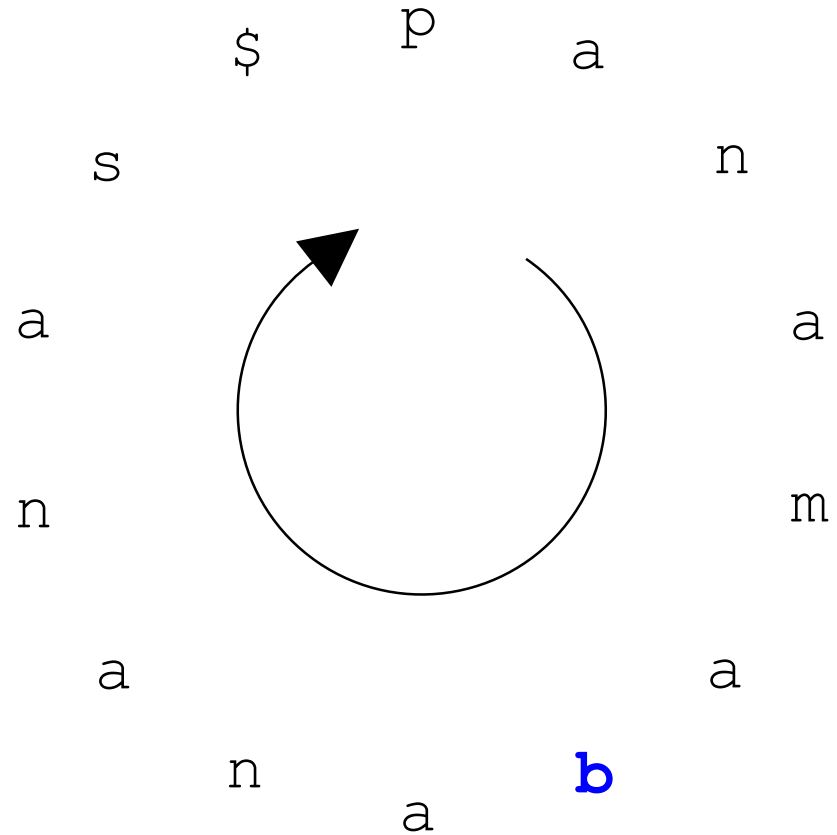
```
panamabananas$  
$panamabananas  
s$panamabanana  
as$panamabanana  
nas$panamabana  
anas$panamaban  
nanas$panamaba  
ananas$panamab
```



Formirati sve ciklične rotacije niske "panamabananas\$"

BWT – Burrows-Wheeler transformacija

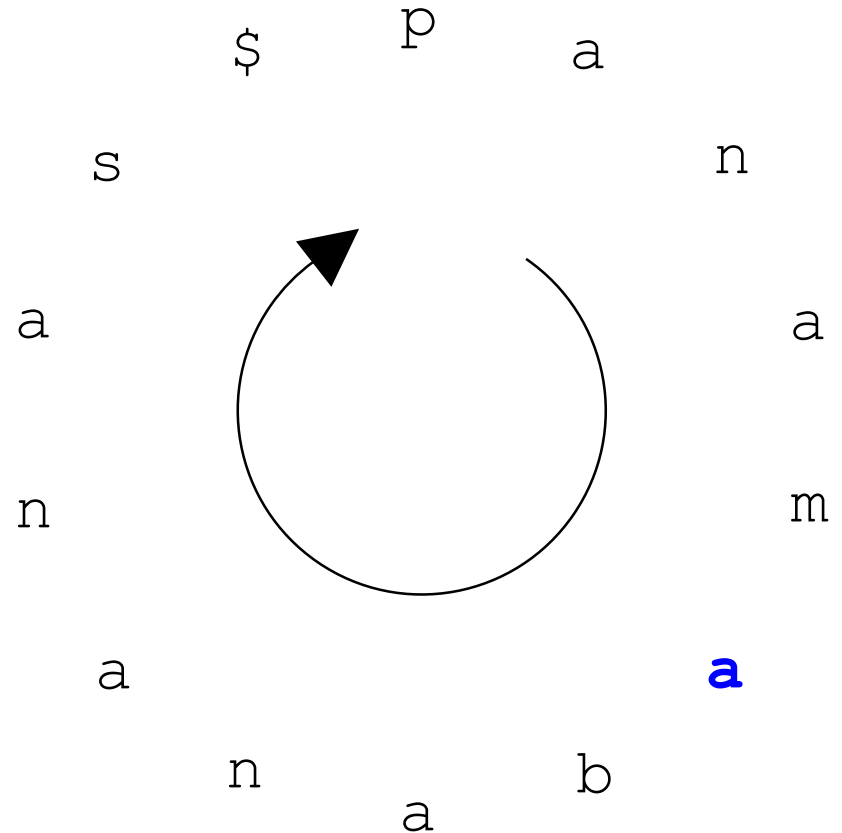
```
panamabananas$  
$panamabananas  
s$panamabanan  
as$panamabanan  
nas$panamabana  
anas$panamaban  
nanas$panamaba  
ananas$panamab  
bananas$panama
```



Formirati sve ciklične rotacije niske
“panamabananas\$”

BWT – Burrows-Wheeler transformacija

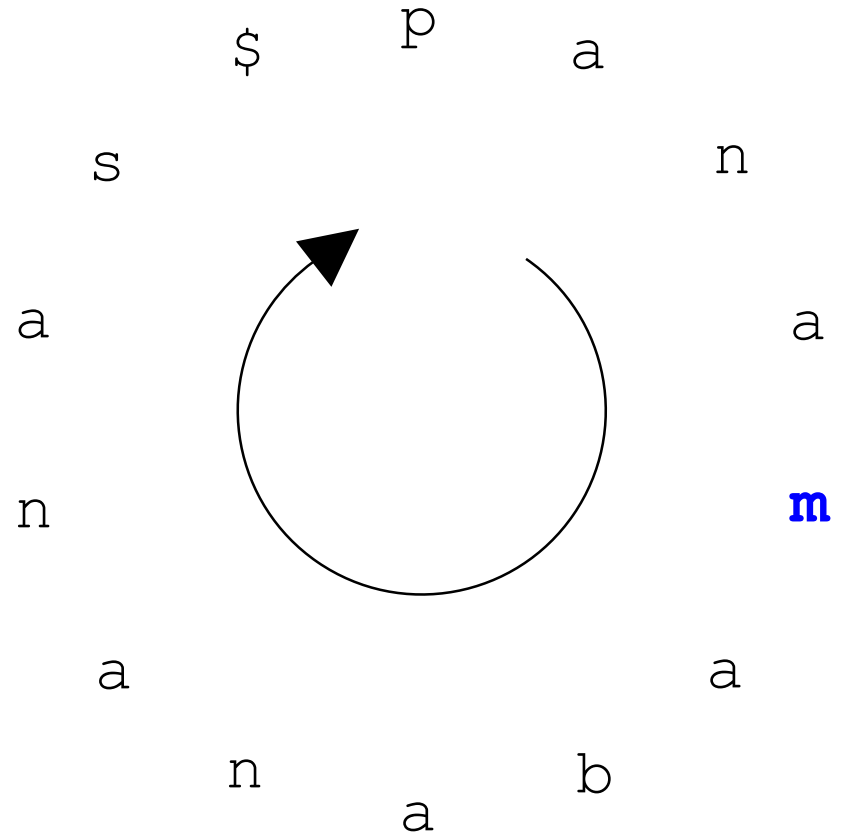
panamabananas\$
\$panamabananas
s\$panamabanan
as\$panamabanan
nas\$panamabana
anas\$panamaba
nanas\$panamaba
ananas\$panamab
bananas\$panama
abananas\$panam



Formirati sve ciklične
rotacije niske
“panamabananas\$”

BWT – Burrows-Wheeler transformacija

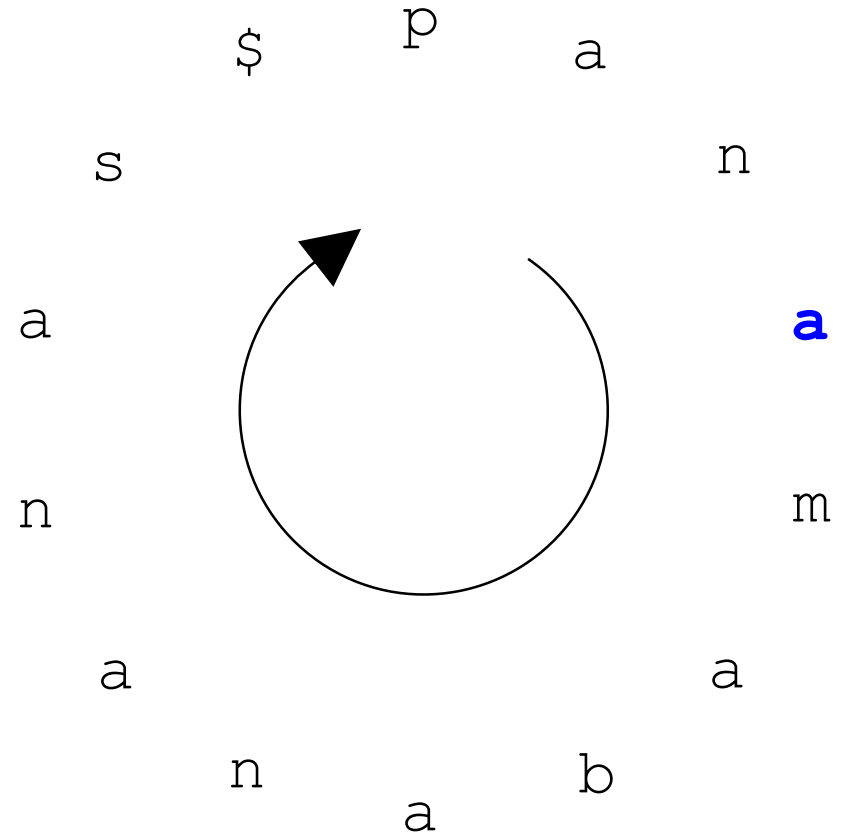
panamabananas\$
\$panamabananas
s\$panamabanan
as\$panamabanan
nas\$panamabana
anas\$panamaban
nanas\$panamaba
ananas\$panamab
bananas\$panama
abananas\$panam
mabananas\$pana



Formirati sve ciklične rotacije niske "panamabananas\$"

BWT – Burrows-Wheeler transformacija

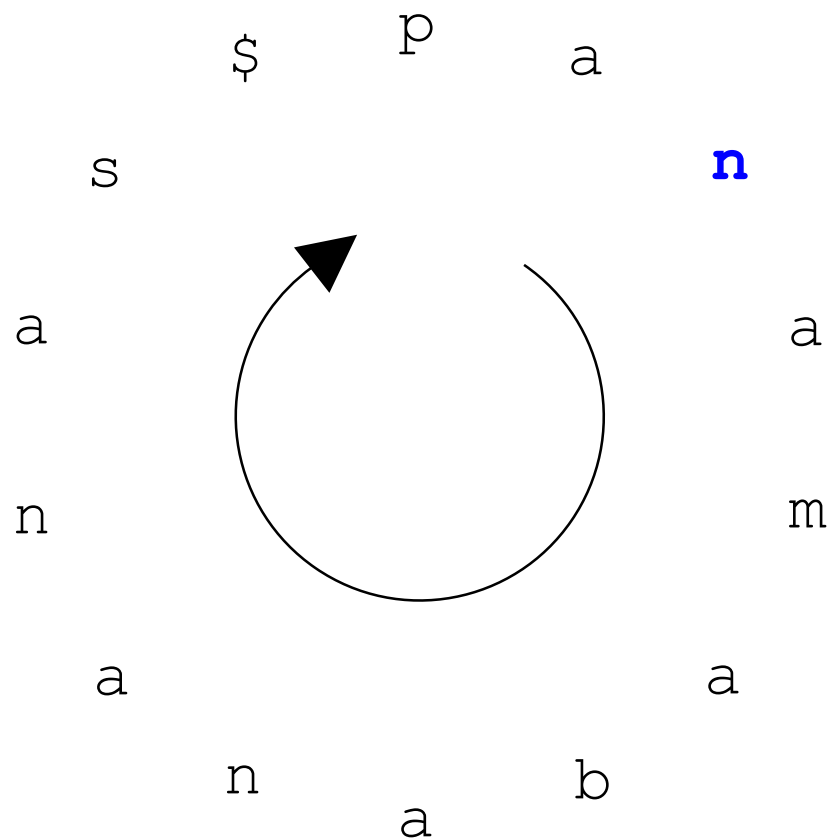
panamabananas\$
\$panamabananas
s\$panamabanan
as\$panamabanan
nas\$panamabana
anas\$panamaban
nanas\$panamaba
ananas\$panamab
bananas\$panama
abananas\$panam
mabananas\$pana
amabananas\$pan



Formirati sve ciklične
rotacije niske
“panamabananas\$”

BWT – Burrows-Wheeler transformacija

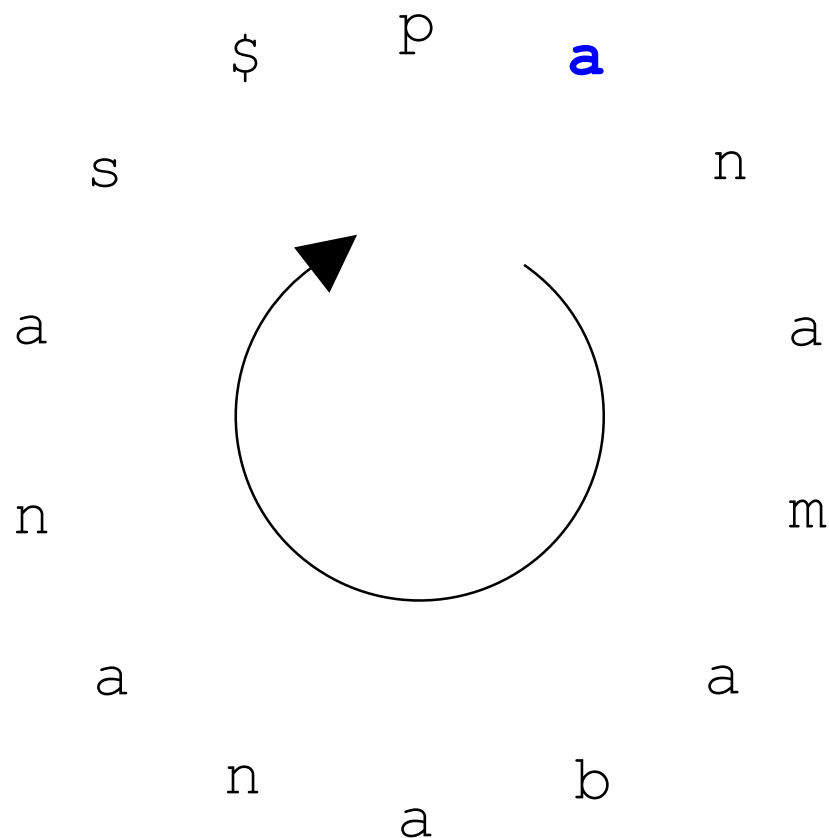
panamabananas\$
\$panamabananas
s\$panamabanana
as\$panamabanana
nas\$panamabana
anas\$panamaban
nanas\$panamaba
ananas\$panamab
bananas\$panama
abananas\$panam
mabananas\$pana
amabananas\$pan
namabananas\$pa



Formirati sve ciklične rotacije niske “panamabananas\$”

BWT – Burrows-Wheeler transformacija

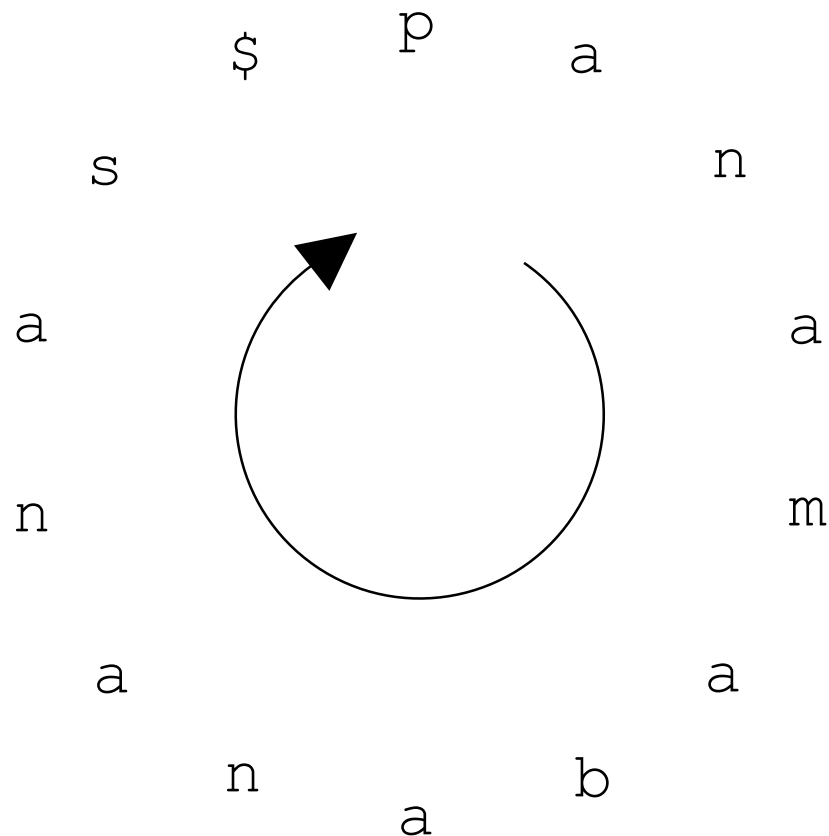
panamabananas\$
\$panamabananas
s\$panamabanana
as\$panamabanana
nas\$panamabana
anas\$panamaban
nanas\$panamaba
ananas\$panamab
bananas\$panama
abananas\$panam
mabananas\$pana
amabananas\$pan
namabananas\$pa
anamabananas\$p



Formirati sve ciklične rotacije niske “panamabananas\$”

BWT – Burrows-Wheeler transformacija

panamabananas\$
\$panamabananas
s\$panamabanana
as\$panamabanana
nas\$panamabana
anas\$panamaban
nanas\$panamaba
ananas\$panamab
bananas\$panama
abananas\$panam
mabananas\$pana
amabananas\$pan
namabananas\$pa
anamabananas\$p



Formirati sve ciklične
rotacije niske
“panamabananas\$”

BWT – Burrows-Wheeler transformacija

panamabananas\$
\$panamabananas
s\$panamabanana
as\$panamabanana
nas\$panamabana
anas\$panamaban
nanas\$panamaba
ananas\$panamab
bananas\$panama
abananas\$panam
mabananas\$pana
amabananas\$pan
namabananas\$pa
anamabananas\$p



\$panamabanana**s**
abananas\$panam
amabananas\$pan
anamabananas\$**p**
anas\$panamab
anas\$panamaban
as\$panamaban
bananas\$panama
mabananas\$pana
namabananas\$pa
nanas\$panamaba
nas\$panamabana
panamabananas\$
s\$panamabanana

Formirati sve ciklične rotacije niske “panamabananas\$”

Sortirati niske leksikografski (\$ je na prvom mestu)

BWT – Barouz-Vilerovala transformacija

panamabananas\$
\$panamabananas
s\$panamabanana
as\$panamabanana
nas\$panamabana
anas\$panamaban
nanas\$panamaba
nanas\$panamaba
nanas\$panamaba
nanas\$panama
abananas\$panama
abananas\$panama
mabananas\$pana
mabananas\$pana
mabananas\$pana
namabananas\$pa
namabananas\$pa



\$panamabananas
abananas\$panam
amabananas\$pan
anamabananas\$pa
nanas\$panamab
anas\$panamaban
anas\$panamaban
anas\$panamaban
nanas\$panama
mabananas\$pana
namabananas\$pa
nanas\$panamaba
nanas\$panamaba
panamabananas\$
s\$panamabanana

Formirati sve ciklične
rotacije niske
“panamabananas\$”

Sortirati niske
leksikografski
(\$ je na prvom mestu)

BWT – Barouz-Vilerova transformacija

panamabananas\$
\$panamabananas
s\$panamabanana
as\$panamabanana
nas\$panamabana
anas\$panamaban
nanas\$panamaba
nanas\$panamaba
nanas\$panamaba
nanas\$panama
bananas\$panama
abananas\$panam
mabananas\$pana
amabananas\$pan
namabananas\$pa
anamabananas\$p



\$panamabanana**s**
abananas\$pana**m**
amabananas\$pa**n**
anamabananas\$**p**
ananas\$panama**b**
anas\$panamaba**n**
anas\$panamaba**n**
anas\$panamaba**n**
bananas\$panama**a**
mabananas\$pana**a**
namabananas\$pa**a**
nanas\$panamaba**a**
nas\$panamabana**a**
panamabananas\$**s**
s\$panamabanana**a**

Formirati sve ciklične rotacije niske “panamabananas\$”

Barouz-Vilerova transformacija:
poslednja kolona = smnpbnnaaaaa\$a

Zašto poslednja kolona?

nd Corey (1). They kindly made their manuscript available a
nd criticism, especially on interatomic distances. We a
nd cytosine. The sequence of bases on a single chain d a
nd experimentally (3,4) that the ratio of the amounts o u
nd for this reason we shall not comment on it. We wish a
nd guanine (purine) with cytosine (pyrimidine). In oth a
nd ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin a
nd its water content is rather high. At lower water co a
nd pyrimidine bases. The planes of the bases are perpe a
nd stereochemical arguments. It has not escaped our no a
nd that only specific pairs of bases can bond together u
nd the atoms near it is close to Furberg's 'standard co a
nd the bases on the inside, linked together by hydrogen a
nd the bases on the outside. In our opinion, this stru a
nd the other a pyrimidine for bonding to occur. The hy a
nd the phosphates on the outside. The configuration of a
nd the ration of guanine to cytosine, are always very c a
nd the same axis (see diagram). We have made the usual u
nd their co-workers at King's College, London. One of a

Zašto poslednja kolona?

- Poslednja kolona sadrži veliki broj ranova.
 - Ali i prva kolona isto sadrži veliki broj ranova. Zašto ne uzmemo prvu?
- Da vidimo kako možemo da dekompresujemo nisku kompresovanu na ovaj način pa ćemo se onda vratiti na ovo važno pitanje

BWT: Konverzija ripita u ranove

**Barouz-Vilerova
transformacija**

Genome

Konverzija ripita
u ranove

BWT(Genome)

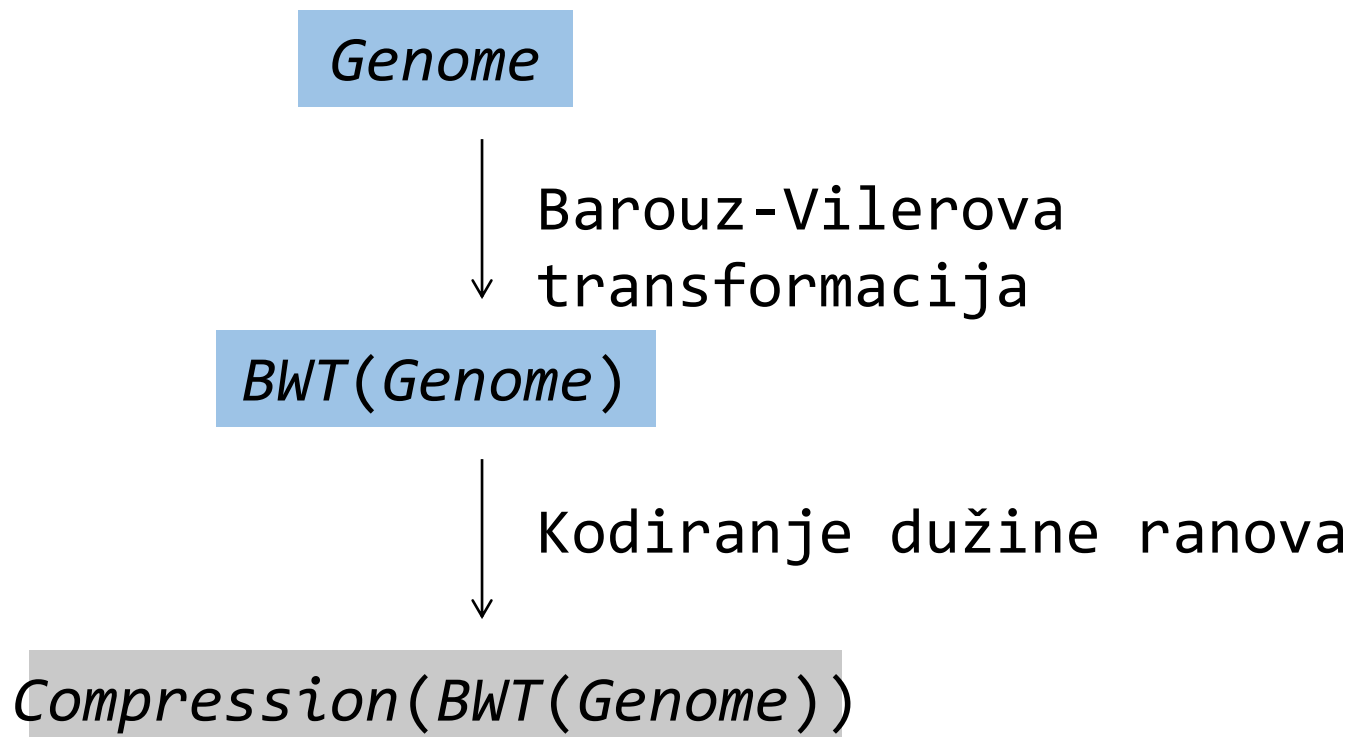
Kodiranje dužine ranova

Compression(BWT(Genome))

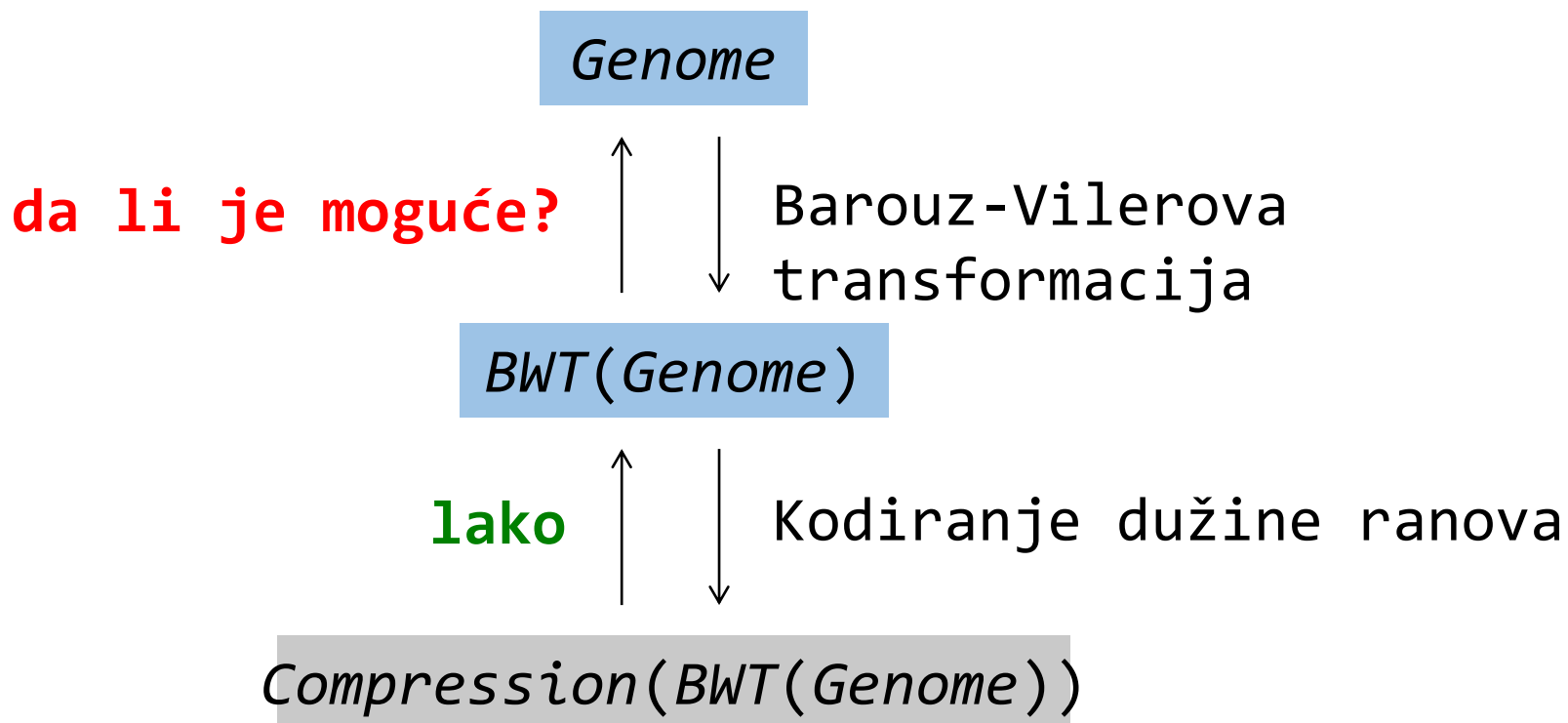
Pregled

- Mapiranje očitavanja
- Sufiksna stabla
- Kompresija niski i Burrows-Wheelerova transformacija
- **Inverzna BWT**
- ...

Kako izvršiti dekompresiju?



Kako izvršiti dekompresiju?



Rekonstrukcija niske banana

\$ b a n a n a **a**
a \$ b a n a **n**
a n a \$ b a **n**
a n a n a \$ **b**
b a n a n a \$
n a \$ b a n **a**
n a n a \$ b **a**

Rekonstrukcija niske banana

```
$ b a n a n a  
a $ b a n a n  
a n a $ b a n  
a n a n a $ b  
b a n a n a $  
n a $ b a n a  
n a n a $ b a
```

- Ako sortiramo karaktere poslednje kolone “annb\$aa”, dobićemo prvu kolonu matrice.

Rekonstrukcija niske banana

\$ b a n a n a		a \$
a \$ b a n a n		n a
a n a \$ b a n		n a
a n a n a \$ b	→	b a
b a n a n a \$	2-mers	\$ b
n a \$ b a n a		a n
n a n a \$ b a		a n

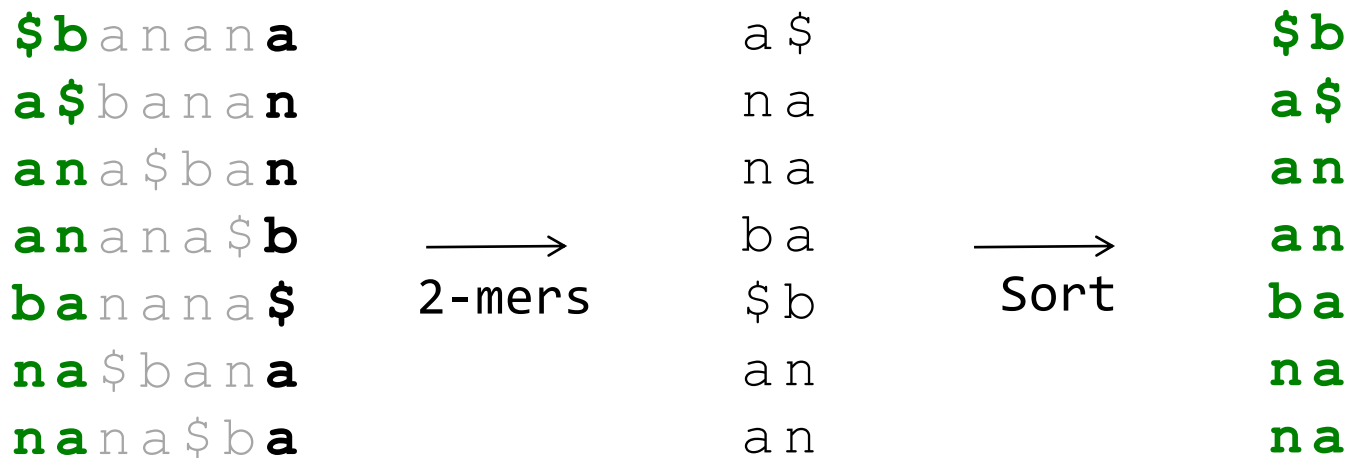
- Na osnovu toga znamo *2-gramski sastav* cirkularne niske banana\$

Rekonstrukcija niske banana

\$ b a n a n a		a \$		\$ b
a \$ b a n a n		n a		a \$
a n a \$ b a n		n a		a n
a n a n a \$ b	→	b a	→	a n
b a n a n a \$	2-mers	\$ b	Sort	b a
n a \$ b a n a		a n		n a
n a n a \$ b a		a n		n a

- Na osnovu toga znamo *2-gramski sastav* cirkularne niske banana\$
- Sortiranjem niski dobijamo prve dve kolone matrice

Rekonstrukcija niske banana



- Na osnovu toga znamo *2-gramski sastav* cirkularne niske banana\$
- Sortiranjem niski dobijamo prve dve kolone matrice

Rekonstrukcija niske banana

\$ b a n a n a
a **\$** b a n a n
a n a **\$** b a n
a n a n a **\$** b
b a n a n a **\$**
n a **\$** b a n a
n a n a **\$** b a

Rekonstrukcija niske banana

\$ b a n a n a		a \$ b
a \$ b a n a n		n a \$
a n a \$ b a n		n a n
a n a n a \$ b	→	b a n
b a n a n a \$	3-mers	\$ b a
n a \$ b a n a		a n a
n a n a \$ b a		a n a

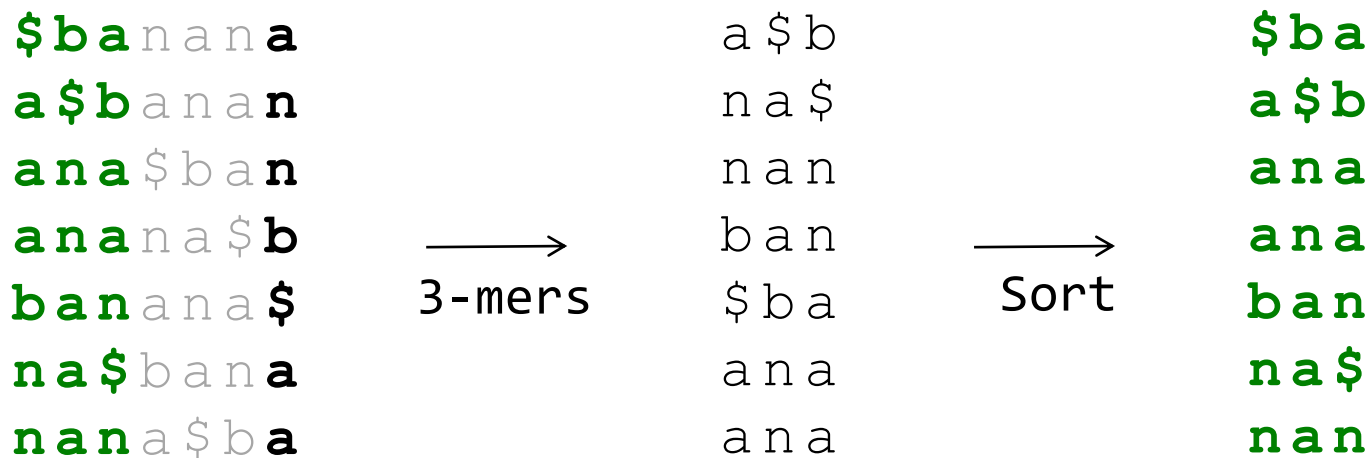
- Sada znamo *3-gramski sastav* cirkularne niske banana\$

Rekonstrukcija niske banana

\$ b a n a n a		a \$ b		\$ b a
a \$ b a n a n		n a \$		a \$ b
a n a \$ b a n		n a n		a n a
a n a n a \$ b	→	b a n	→	a n a
b a n a n a \$	3-mers	\$ b a	Sort	b a n
n a \$ b a n a		a n a		n a \$
n a n a \$ b a		a n a		n a n

- Sada znamo *3-gramski sastav* cirkularne niske banana\$
- Sortiranjem niski dobijamo prve 3 kolone matrice

Rekonstrukcija niske banana



- Sada znamo *3-gramski sastav* cirkularne niske banana\$
- Sortiranjem niski dobijamo prve 3 kolone matrice

Rekonstrukcija niske banana

\$ b a n a n a

a **\$** b a n a n

a n a **\$** b a n

a n a n a **\$** b

b a n a n a **\$**

n a **\$** b a n a

n a n a **\$** b a

Rekonstrukcija niske banana

\$ b a n a n a		a \$ b a
a \$ b a n a n		n a \$ b
a n a \$ b a n		n a n a
a n a n a \$ b	→	b a n a
b a n a n a \$	4-mers	\$ b a n
n a \$ b a n a		a n a \$
n a n a \$ b a		a n a n

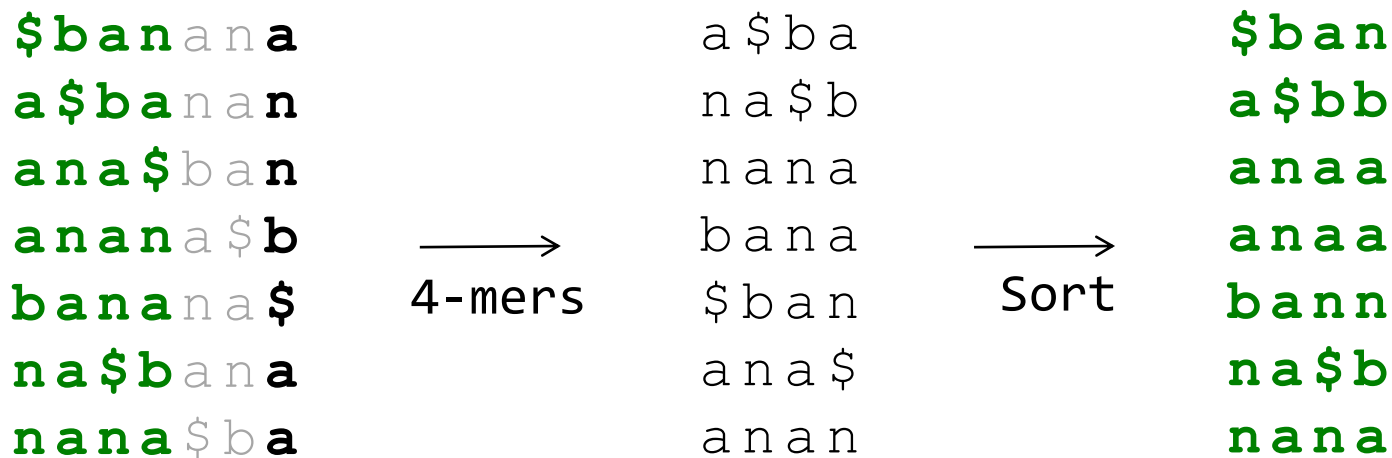
- Sada znamo *4-gramski sastav* cirkularne niske banana\$

Rekonstrukcija niske banana

\$ b a n a n a		a \$ b a		\$ b a n
a \$ b a n a n		n a \$ b		a \$ b b
a n a \$ b a n		n a n a		a n a a
a n a n a \$ b	→	b a n a	→	a n a a
b a n a n a \$	4-mers	\$ b a n	Sort	b a n n
n a \$ b a n a		a n a \$		n a \$ b
n a n a \$ b a		a n a n		n a n a

- Sada znamo *4-gramski sastav* cirkularne niske banana\$
- Sortiranjem niski dobijamo prve 4 kolone matrice

Rekonstrukcija niske banana



- Sada znamo *4-gramski sastav* cirkularne niske banana\$
- Sortiranjem niski dobijamo prve 4 kolone matrice

Rekonstrukcija niske banana

\$bana n a
a\$ban a n
ana\$ban
anana\$b
banana\$
na\$ban a
nana\$b a

Rekonstrukcija niske banana

\$ banana		a\$ban
a \$ banan		na\$ba
ana \$ ban		nana\$
anana \$	→	banan
banana \$	5-mers	\$bana
na \$ ban		ana\$b
nana \$ ba		anana

- Sada znamo *5-gramski sastav* cirkularne niske banana\$

Rekonstrukcija niske banana

\$ banana		a\$ban		\$bana
a \$ banan		na\$ba		a\$bbn
ana \$ ban		nana\$		anaab
anana \$	→	banan	→	anaaa
banana \$	5-mers	\$bana	Sort	bannn
na \$ banan		ana\$b		na\$ba
nana \$ ba		anana		nana\$

- Sada znamo *5-gramski sastav* cirkularne niske banana\$
- Sortiranjem niski dobijamo prve 5 kolone matrice

Rekonstrukcija niske banana

\$bana na		a\$ban		\$bana
a\$ban an		na\$ba		a\$bbn
ana\$b an		nana\$		anaab
anana \$b	→	banan	→	anaaa
banan a\$	5-mers	\$bana	Sort	bannn
na\$b ana		ana\$b		na\$ba
nana \$ba		anana		nana\$

- Sada znamo *5-gramski sastav* cirkularne niske banana\$
- Sortiranjem niski dobijamo prve 5 kolone matrice

Rekonstrukcija niske banana

\$bana n a
a\$ban a n
ana\$ba n
anana \$b
banan a \$
na\$ba n a
nana\$ba

Rekonstrukcija niske banana

\$bana	n	a		a	\$bana
a\$bana	n			n	a\$bana
ana\$bana				n	ana\$b
anana\$b			→	b	anana
banana\$b		6-mers		\$	banan
na\$bana				a	na\$b
nana\$b				a	nana\$b

- Sada znamo *6-gramski sastav* cirkularne niske banana\$

Rekonstrukcija niske banana

\$bana		a\$bana		\$banan
a\$bana		na\$bana		a\$bbna
ana\$bana		nana\$b		anaaba
anana\$b	→	banana	→	anaaa\$b
banana\$a	6-mers	\$banan	Sort	bannna
na\$bana		ana\$ba		na\$bana
nana\$ba		anana\$b		nana\$b

- Sada znamo *6-gramski sastav* cirkularne niske banana\$a
- Sortiranjem niski dobijamo prve 6 kolone matrice

Rekonstrukcija niske banana

\$banana		a\$banana		\$banan
a\$banan		na\$ban		a\$bbna
ana\$ban		nana\$b		anaaba
anana\$b	→	banana	→	anaaa\$
banana\$	6-mers	\$banan	Sort	bannna
na\$bana		ana\$ba		na\$ban
nana\$ba		anana\$		nana\$b

- Sada znamo *6-gramski sastav* cirkularne niske banana\$
- Sortiranjem niski dobijamo prve 6 kolone matrice

Rekonstrukcija niske banana

\$banana
a\$banan
ana\$ban
anana\$b
banana\$
na\$bana
nana\$ba

- Sada znamo celu matricu!

Rekonstrukcija niske banana

\$banana

a\$banana

ana\$ban

anana\$b

banana\$a

na\$banana

nana\$ba

- Sada znamo celu matricu!
- Svi elementi iz prvog reda (nakom simbola \$) daju polaznu nisku banana.

Prostorna složenost

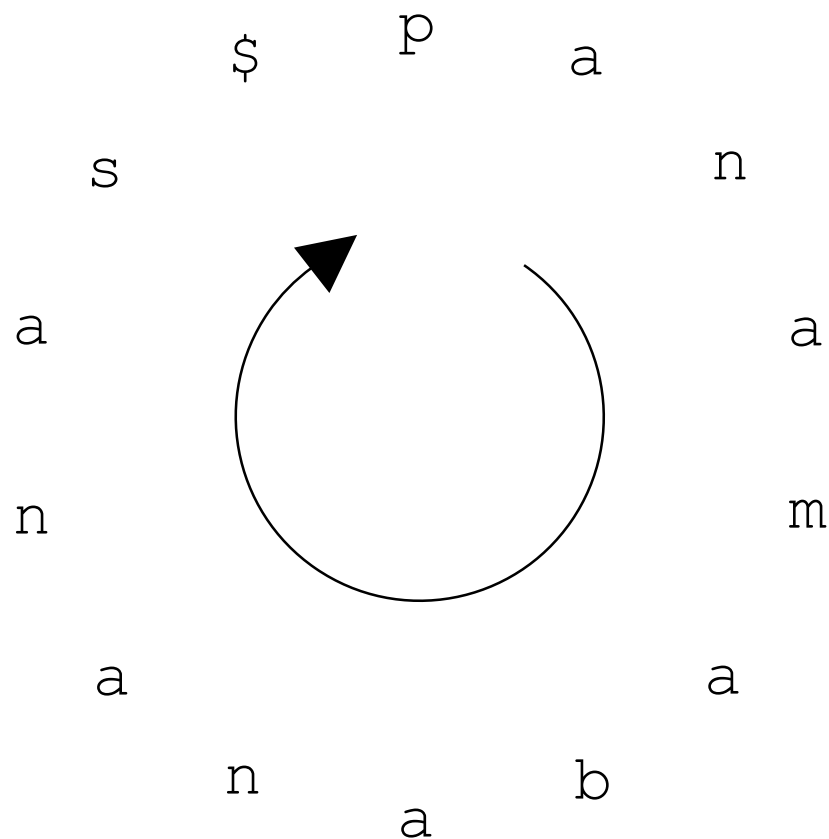
- Rekonstrukcija niske *Genome* na osnovu *BWT(Genome)* zahteva čuvanje $|Genome|$ kopija niske *Genome* (što iznosi $|Genome|^2$) .

```
$b a n a n a  
a $ b a n a n  
a n a $ b a n  
a n a n a $ b  
b a n a n a $  
n a $ b a n a  
n a n a $ b a
```

- Da li možemo invertovati BWT sa manje prostora?

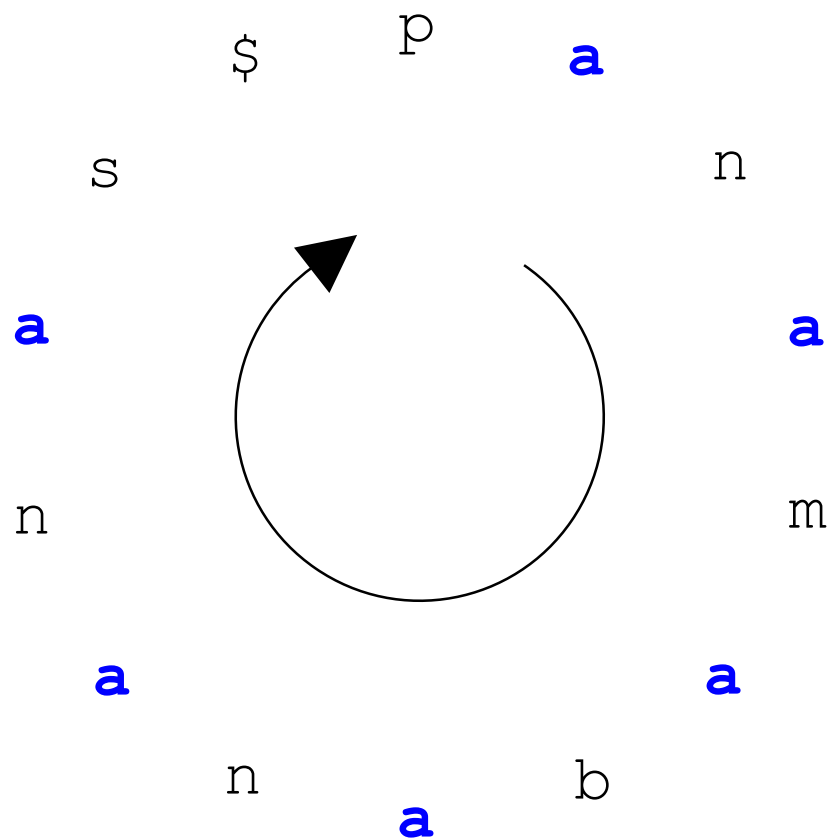
Neobična opservacija

\$panamabananas
abananas\$panam
amabananas\$pan
anamabananas\$p
ananas\$panamab
anas\$panamaban
as\$panamabanan
bananas\$panama
mabananas\$pana
namabananas\$pa
nanas\$panamaba
nas\$panamabana
panamabananas\$
s\$panamabana



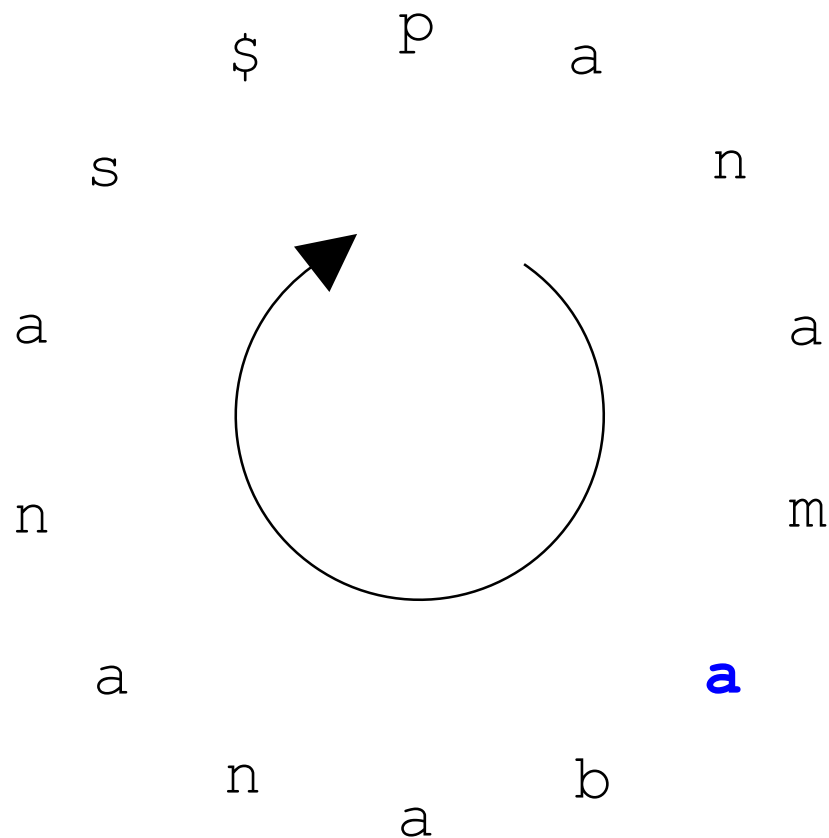
Neobična opservacija

\$panamabananas
abananas\$panam
amabananas\$pan
anamabananas\$p
ananas\$panamab
anas\$panamaban
as\$panamabanan
bananas\$panam**a**
mabananas\$pan**a**
namabananas\$p**a**
nanas\$panamab**a**
nas\$panamaban**a**
panamabananas\$
s\$panamabanan**a**



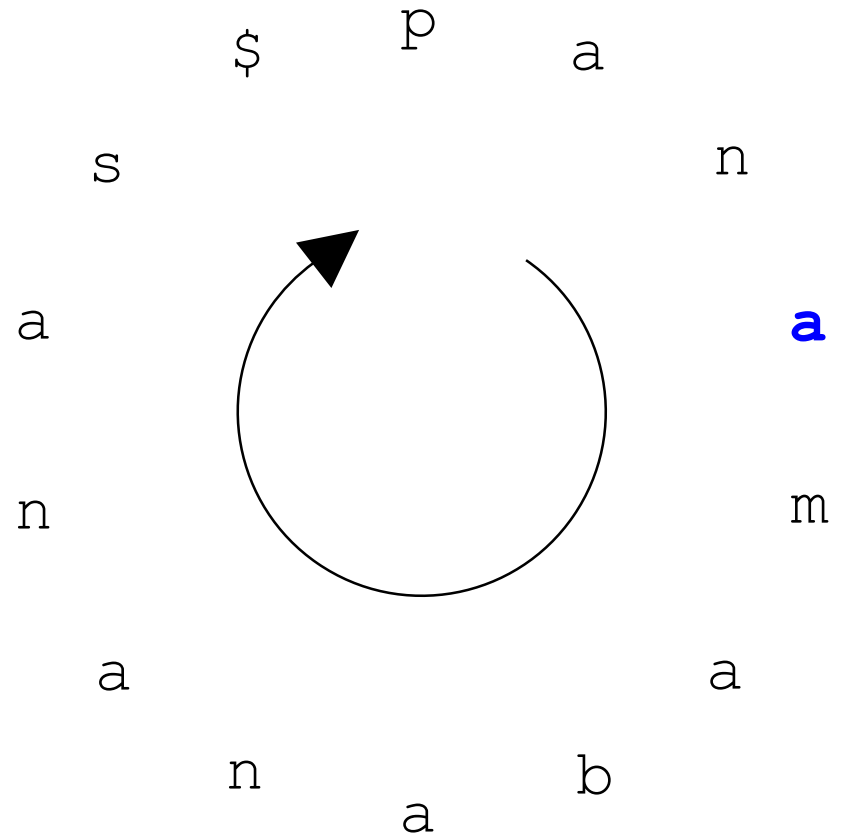
Neobična opservacija

\$panamabananas
abananas\$panam
amabananas\$pan
anamabananas\$p
ananas\$panamab
anas\$panamaban
as\$panamabanan
bananas\$panam**a**
mabananas\$pana
namabananas\$pa
nanas\$panamaba
nas\$panamabana
panamabananas\$
s\$panamabana



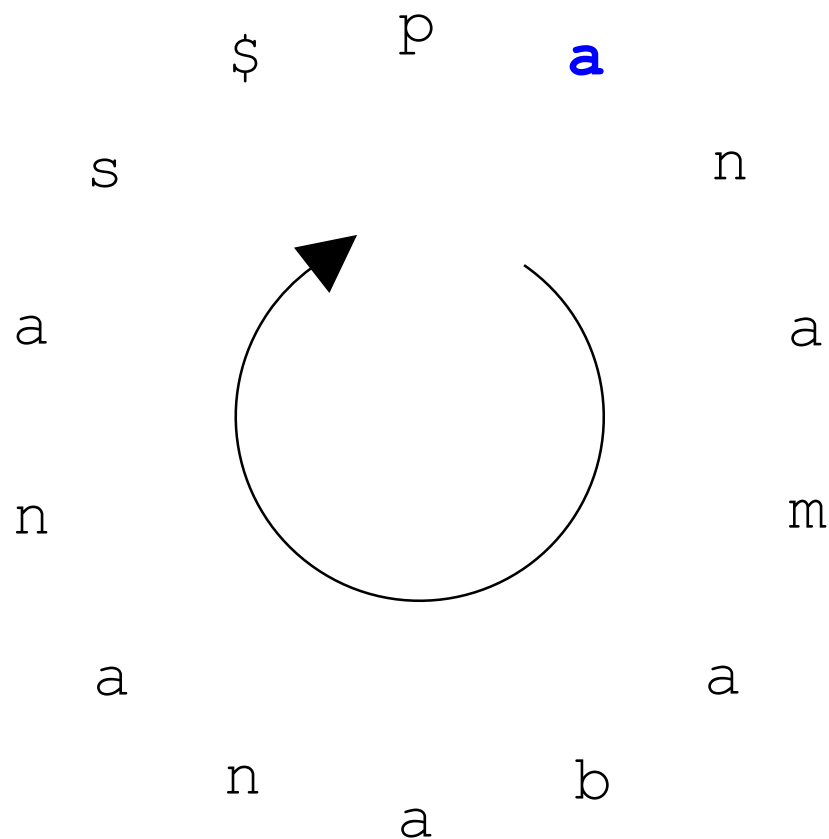
Neobična opservacija

\$panamabananas
abananas\$panam
amabananas\$pan
anamabananas\$p
ananas\$panamab
anas\$panamaban
as\$panamabanan
bananas\$panama
mabananas\$pan**a**
namabananas\$pa
nanas\$panamaba
nas\$panamabana
panamabananas\$
s\$panamabana



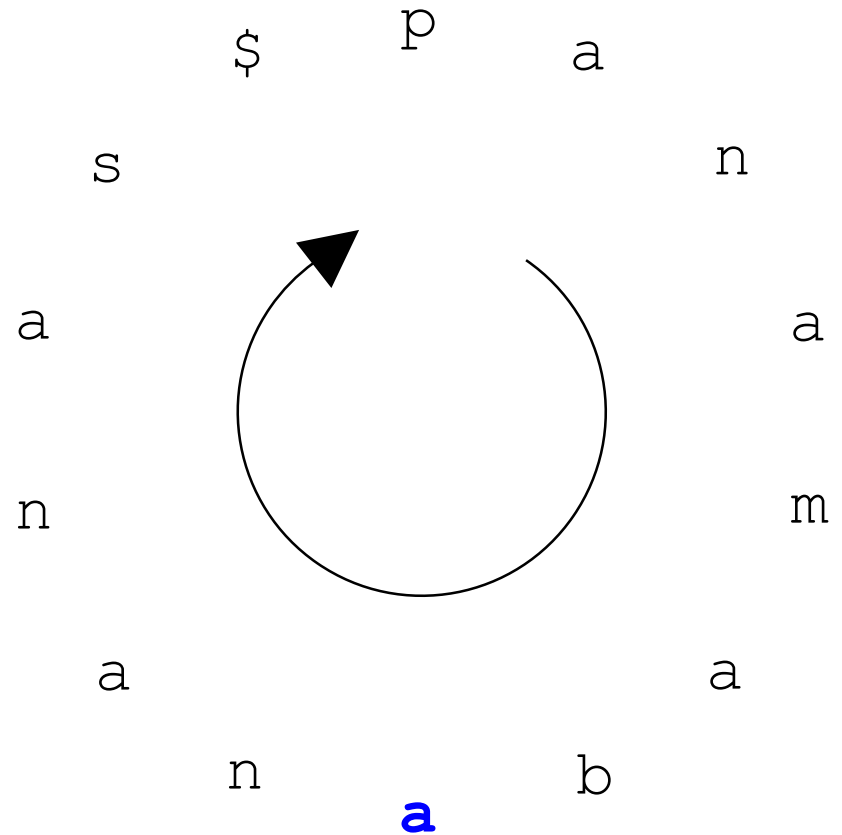
Neobična opservacija

\$panamabananas
abananas\$panam
amabananas\$pan
anamabananas\$
pananas\$panamab
anas\$panamaban
as\$panamabanan
bananas\$panama
mabananas\$pana
namabananas\$**p**
nanas\$panamaba
nas\$panamabana
panamabananas\$
s\$panamabana



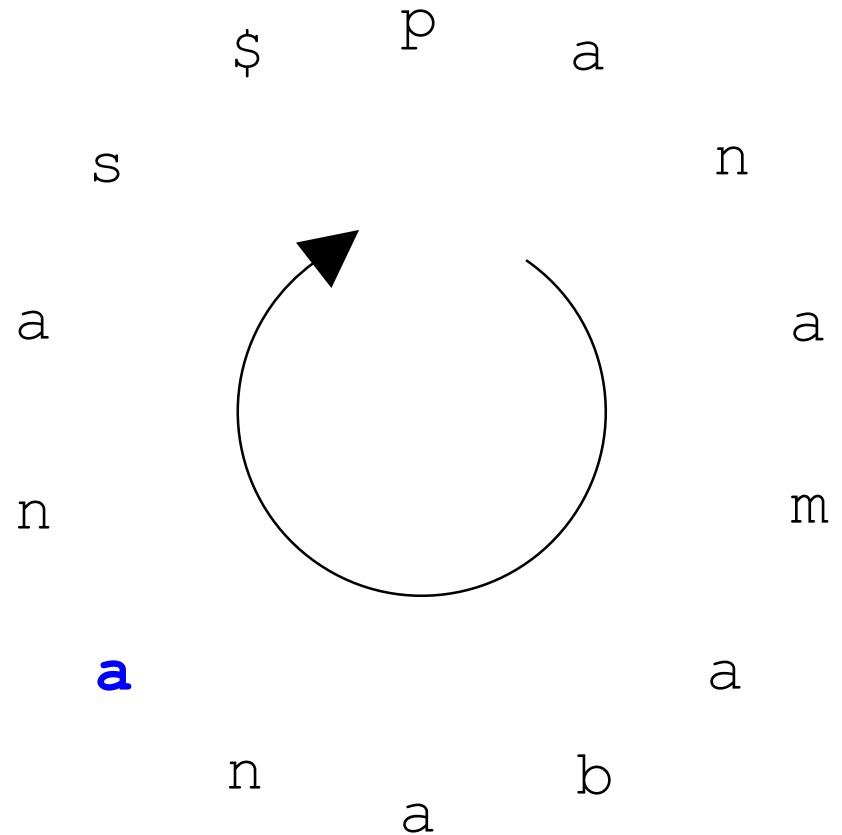
Neobična opservacija

\$panamabananas
abananas\$panam
amabananas\$pan
anamabananas\$p
ananas\$panamab
anas\$panamaban
as\$panamabanan
bananas\$panama
mabananas\$pana
namabananas\$pa
nanas\$panamab**a**
nas\$panamabana
panamabananas\$
s\$panamabana



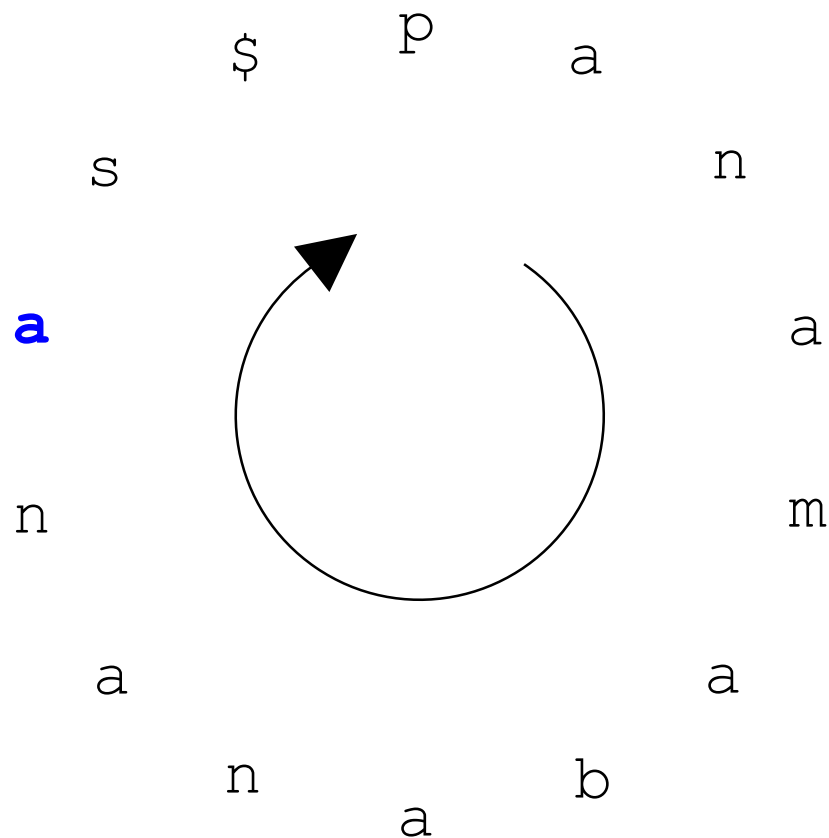
Neobična opservacija

\$panamabananas
abananas\$panam
amabananas\$pan
anamabananas\$p
ananas\$panamab
anas\$panamaban
as\$panamabanan
bananas\$panama
mabananas\$pana
namabananas\$pa
nanas\$panamaba
nas\$panamaban**a**
panamabananas\$
s\$panamabanana



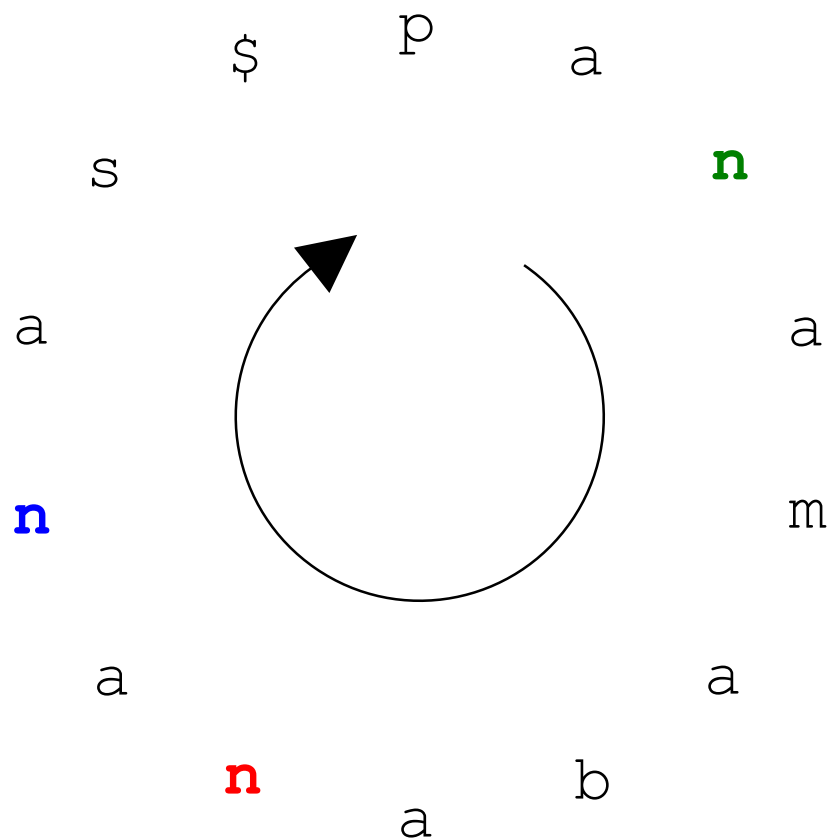
Neobična opservacija

\$panamabananas
abananas\$panam
amabananas\$pan
anamabananas\$p
ananas\$panamab
anas\$panamaban
as\$panamaban
bananas\$panama
mabananas\$pana
namabananas\$pa
nanas\$panamaba
nas\$panamabana
panamabananas\$
s\$panamaban**a**



Neobična opservacija

\$panamabananas
abananas\$panam
amabananas\$pa**n**
anamabananas\$p
ananas\$panamab
anas\$panamaba**n**
as\$panamabana**n**
bananas\$panama
mabananas\$pana
namabananas\$pa
nanas\$panamaba
nas\$panamabana
panamabananas\$
s\$panamabana



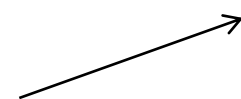
Da li ova opservacija uvek važi?

```
$panamabananas  
1 abananas$panam  
2 amabananas$pan  
3 anamabananas$p  
4 ananas$panamab  
5 anas$panamaban  
6 as$panamabanan  
bananas$panama  
mabananas$pana  
namabananas$pa  
nanas$panamaba  
nas$panamabana  
panamabananas$  
s$panamabana
```

Ove niske su sortirane

Da li ova opservacija uvek važi?

\$panamabananas
1 **a**bananas\$panam
2 **a**mabananas\$pan
3 **a**namabananas\$p
4 **a**nanas\$panamab
5 **a**nas\$panamaban
6 **a**s\$panamaban
bananas\$panama
mabananas\$pana
namabananas\$pa
nanas\$panamaba
nas\$panamabana
panamabananas\$
s\$panamabanana



Odsečemo **a**s\$panamaban

bananas\$panam
mabananas\$pan
namabananas\$p
nanas\$panamab
nas\$panamaban
s\$panamaban

Ove niske su sortirane

Da li ova opservacija uvek važi?

\$panamabananas
1 **a**bananas\$panam
2 **a**mabananas\$pan
3 **a**namabananas\$p
4 **a**nanas\$panamab
5 **a**nas\$panamaban
6 **a**s\$panamaban
bananas\$panama
mabananas\$pana
namabananas\$pa
nanas\$panamaba
nas\$panamabana
panamabananas\$
s\$panamabanana



Odsečemo **a**

bananas\$panam
mabananas\$pan
namabananas\$p
nanas\$panamab
nas\$panamaban
s\$panamabanan
I dalje su
sortirane

Ove niske su sortirane

Da li ova opservacija uvek važi?

\$panamabananas
1 **a**bananas\$panam
2 **a**mabananas\$pan
3 **a**namabananas\$p
4 **a**nanas\$panamab
5 **a**nas\$panamaban
6 **a**s\$panamaban
bananas\$panama
mabananas\$pana
namabananas\$pa
nanas\$panamaba
nas\$panamabana
panamabananas\$
s\$panamabanana



Odsečemo **a**

bananas\$panam
mabananas\$pan
namabananas\$p
nanas\$panamab
nas\$panamaban
s\$panamaban

I dalje su
sortirane

Dodamo **a**
na kraj

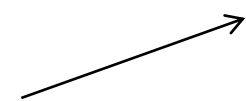
bananas\$panam**a**
mabananas\$pan**a**
namabananas\$pa**a**
nanas\$panamab**a**
nas\$panamaban**a**
s\$panamaban**a**

Ove niske su sortirane

Da li ova opservacija uvek važi?

```
$panamabananas  
1 a bananas$panam  
2 amabananas$pan  
3 anamabananas$p  
4 ananas$panamab  
5 anas$panamaban  
6 as$panamabanan  
bananas$panama  
mabananas$pana  
namabananas$pa  
nanas$panamaba  
nas$panamabana  
panamabananas$  
s$panamabanana
```

Ove niske su sortirane



Odsečemo **a**

```
bananas$panam  
mabananas$pan  
namabananas$p  
nanas$panamab  
nas$panamaban  
s$panamabanan
```



Dodamo **a**
na kraj

```
bananas$panama  
mabananas$pana  
namabananas$pa  
nanas$panamaba  
nas$panamabana  
s$panamabanan
```

I dalje su
sortirane

I dalje su
sortirane

Da li ova opservacija uvek važi?

```

$panamabananas
1 a bananas$panam
2 amabananas$pan
3 anabananas$p
4 ananas$panamab
5 anas$panamaban
6 as$panamabanan
bananas$panama 1
mabananas$pana 2
namabananas$pa 3
nanas$panamaba 4
nas$panamabana 5
panamabananas$
s$panamabanan a 6

```

Odsečemo a

Redosled
se ne
menja!

Ove niske su sortirane

```

bananas$panam
mabananas$pan
namabananas$p
nanas$panamab
nas$panamaban
s$panamabanan

```

I dalje su
sortirane

Dodamo a
na kraj

```

bananas$panama
mabananas$pana
namabananas$pa
nanas$panamaba
nas$panamabana
s$panamabanan a

```

I dalje su
sortirane

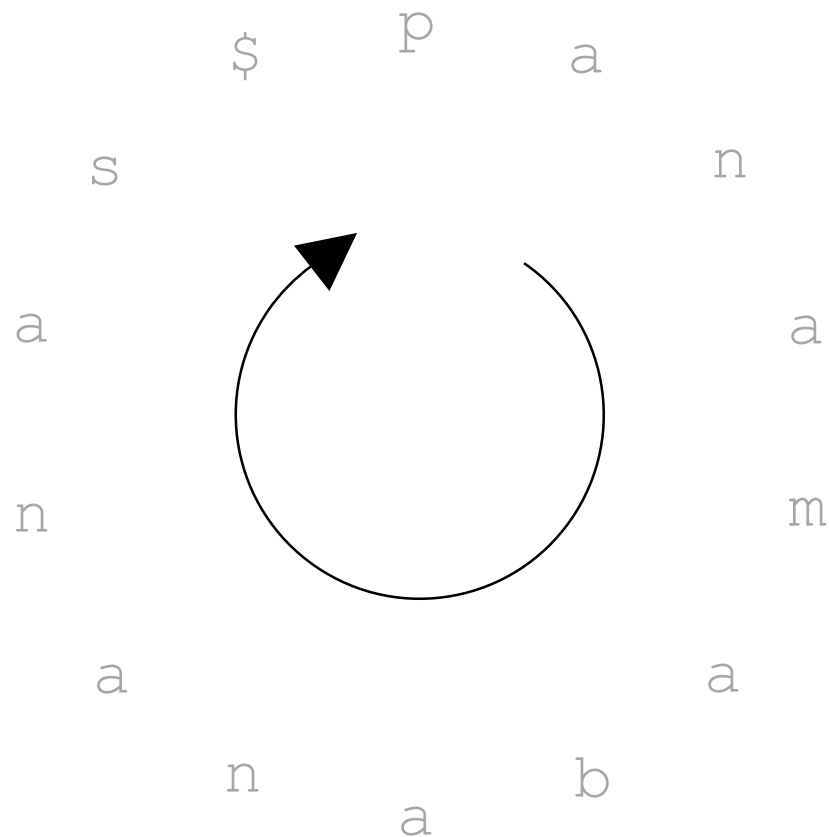
Da li ova opservacija uvek važi?

- **First-Last svojstvo:** k -to pojavljivanje simbola u *FirstColumn* i k -to pojavljivanje simbola u *LastColumn* odgovaraju istoj poziciji simbola u niski *Genome*.

s_1 panamabananas s_1
 a_1 bananas\$panam m_1
 a_2 mabananas\$pan n_1
 a_3 namabananas\$ p_1
 a_4 nanas\$panamab b_1
 a_5 nas\$panamaban n_2
 a_6 s\$panamaban n_3
 b_1 ananas\$panama a_1
 m_1 abananas\$pana a_2
 n_1 amabananas\$pa a_3
 n_2 anas\$panamaba a_4
 n_3 as\$panamabana a_5
 p_1 anamabananas\$ s_1
 s_1 \$panamabanana a_6

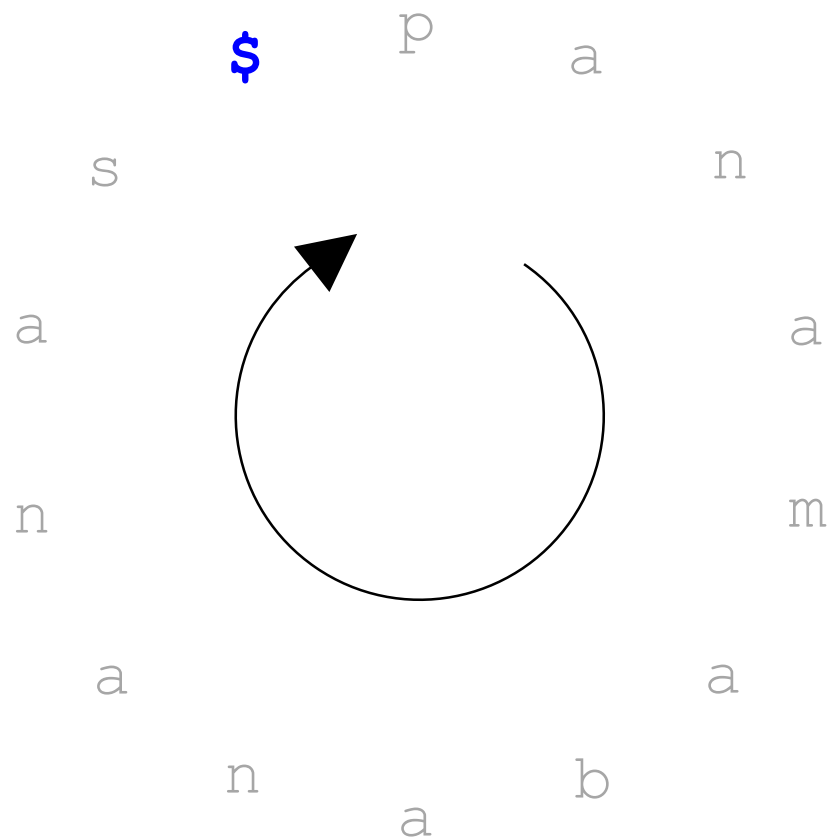
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



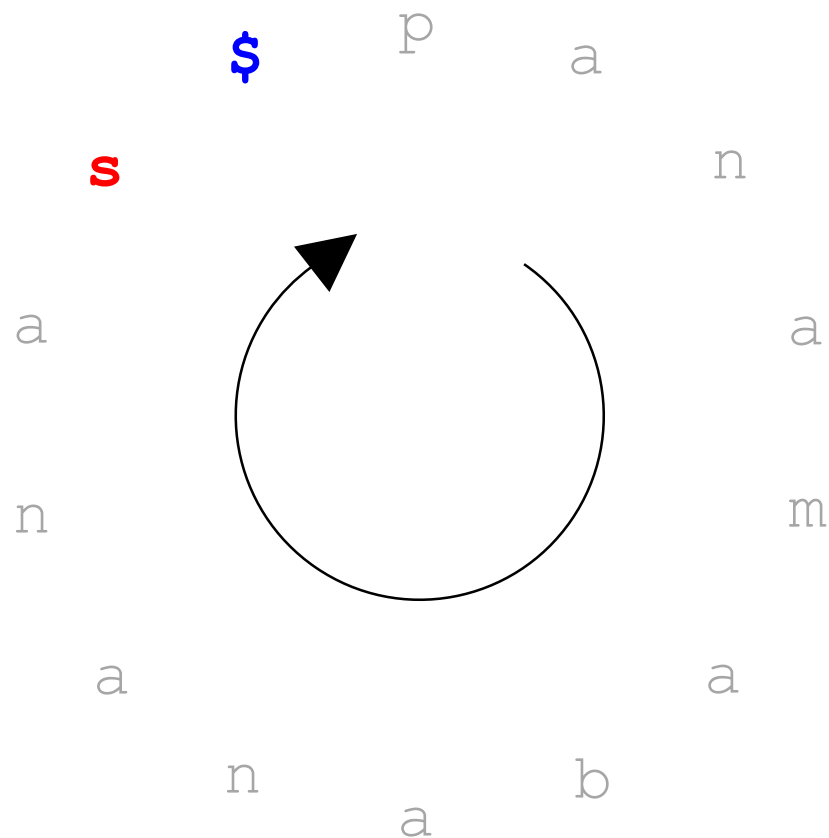
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



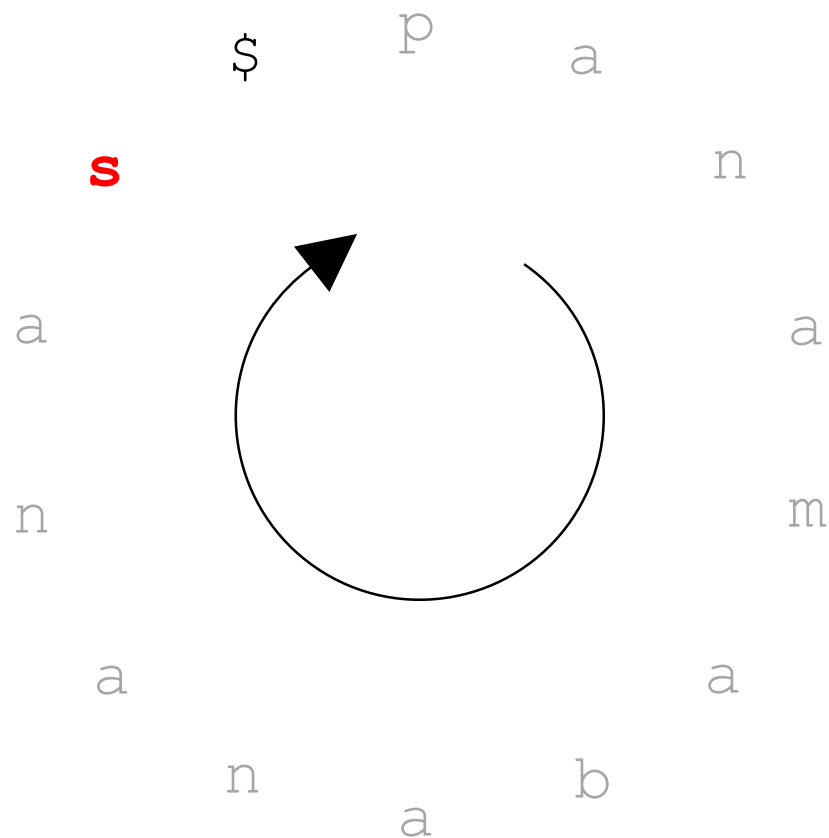
Efikasnija BWT dekompresija

\$₁panamabanana**s**₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanana₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



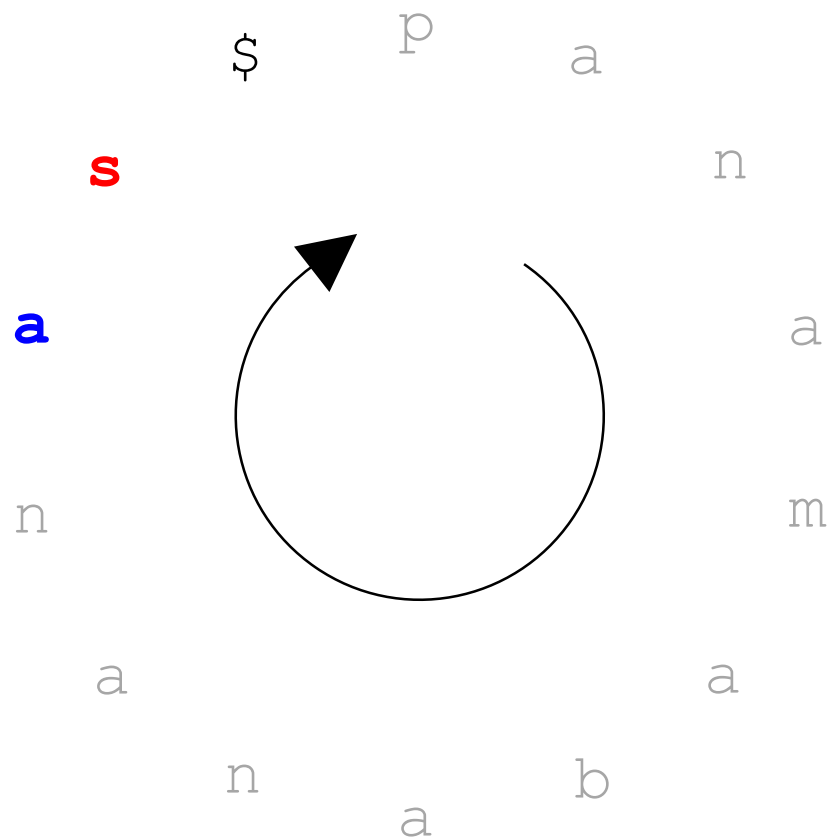
Efikasnija BWT dekompresija

\$₁panamabananas₁**s**₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanan₆



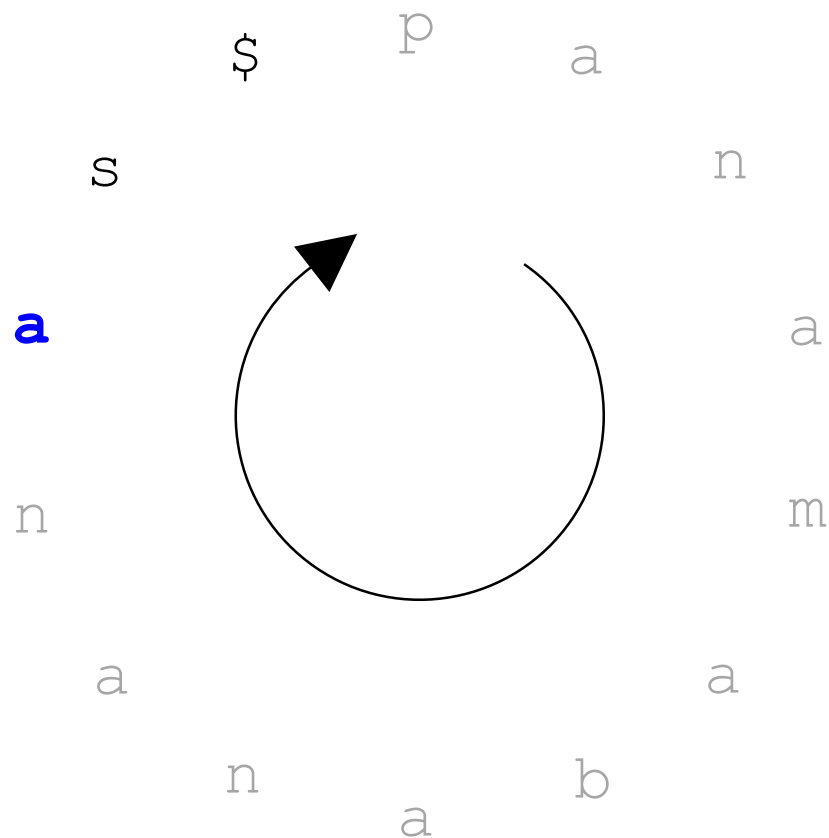
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanan**a**₆



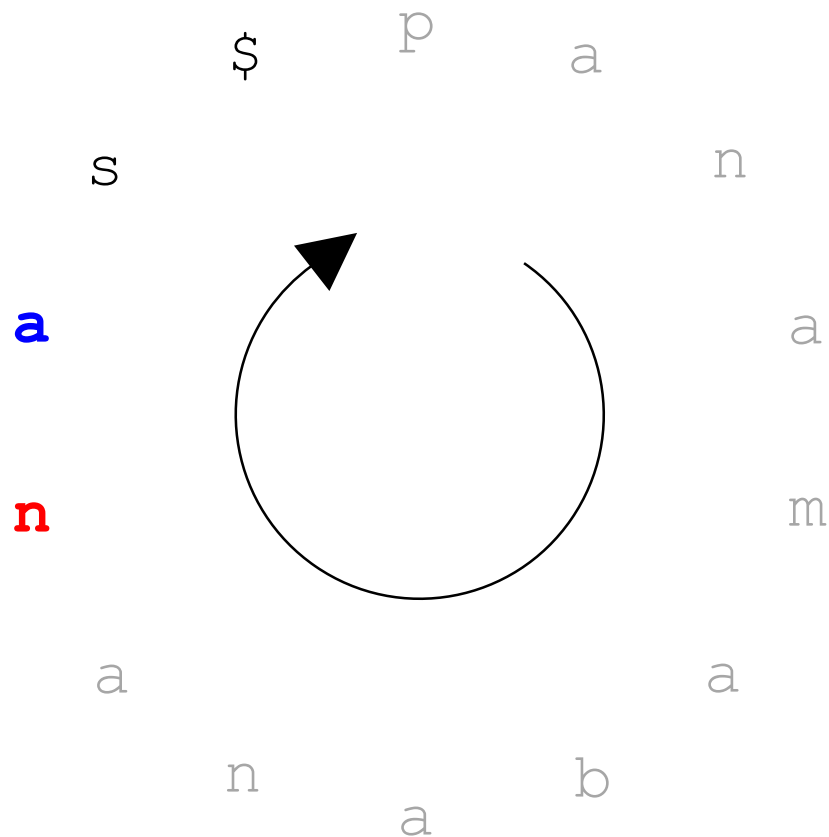
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana**a**₆



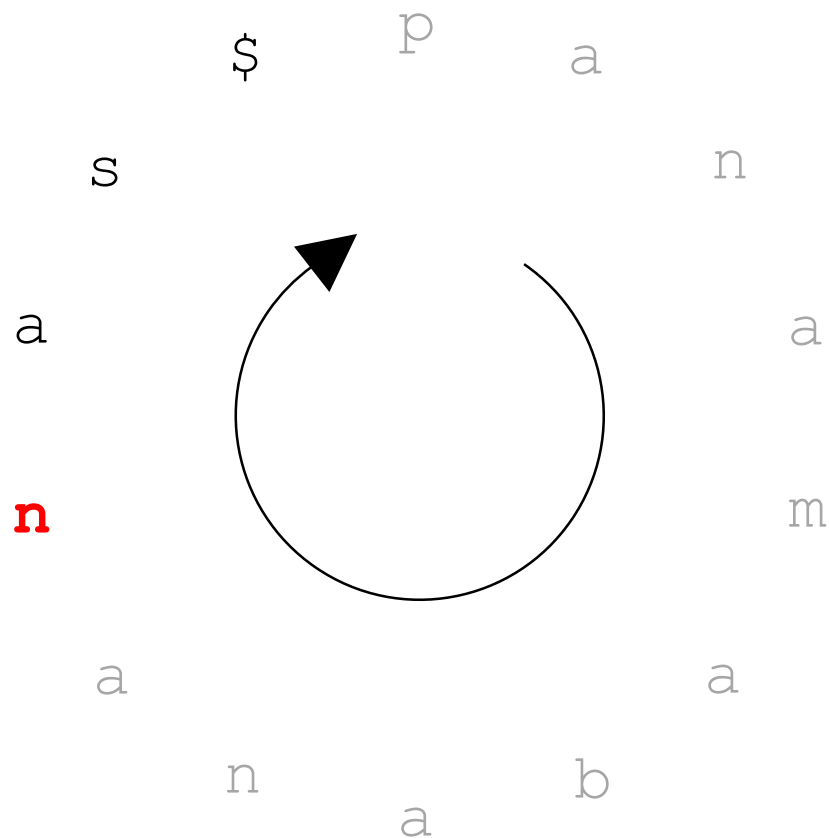
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabana**n**₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabananana₆



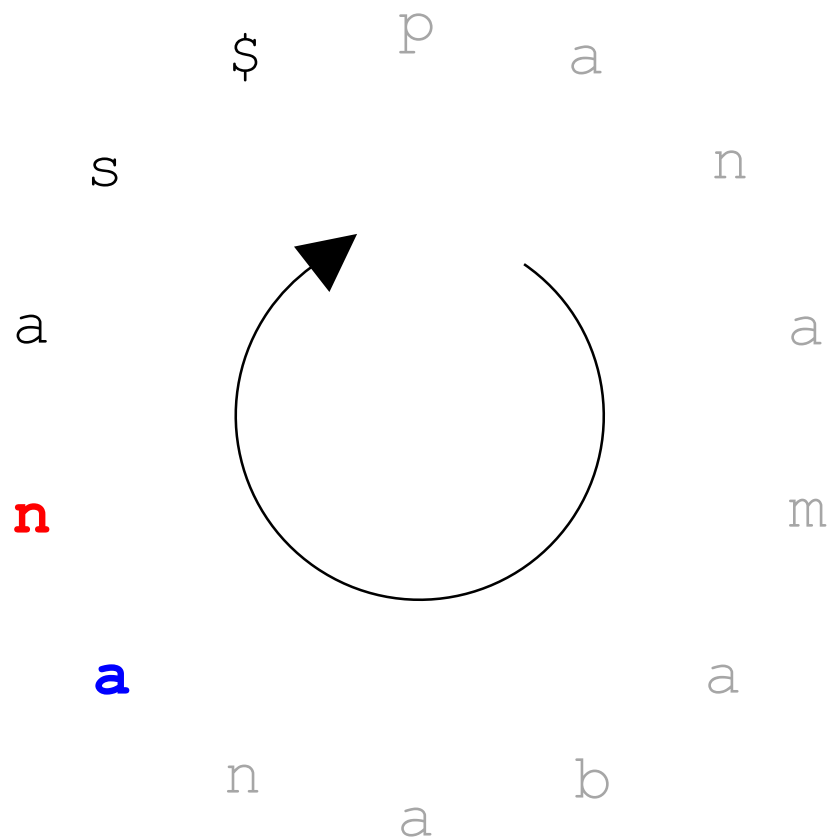
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabana**n**₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



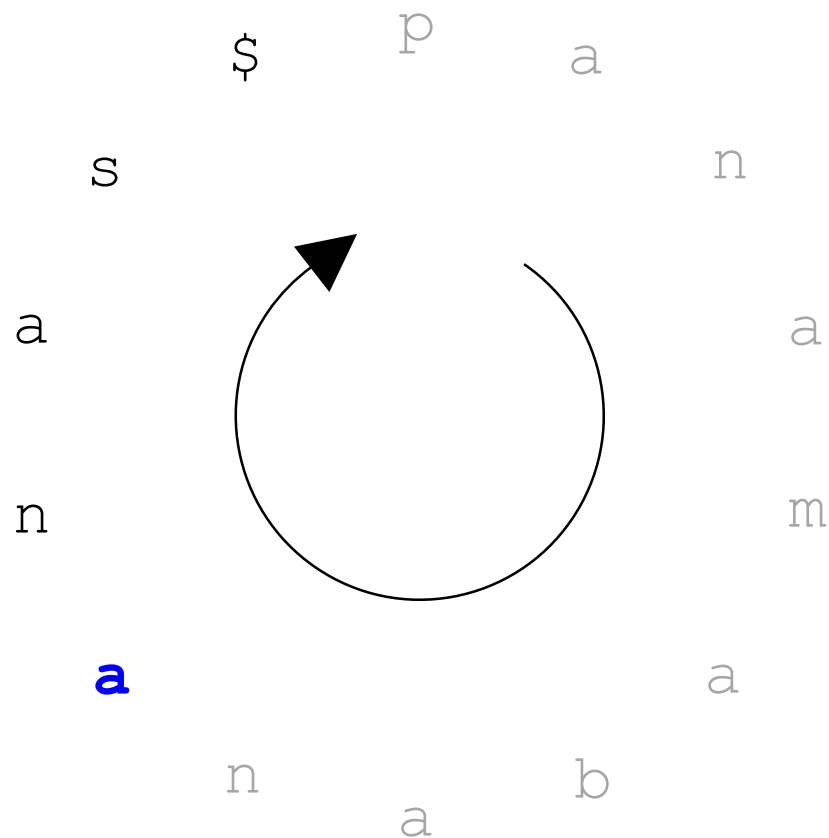
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamaban**a**₅
p₁anamabananas\$₁
s₁\$panamabanana₆



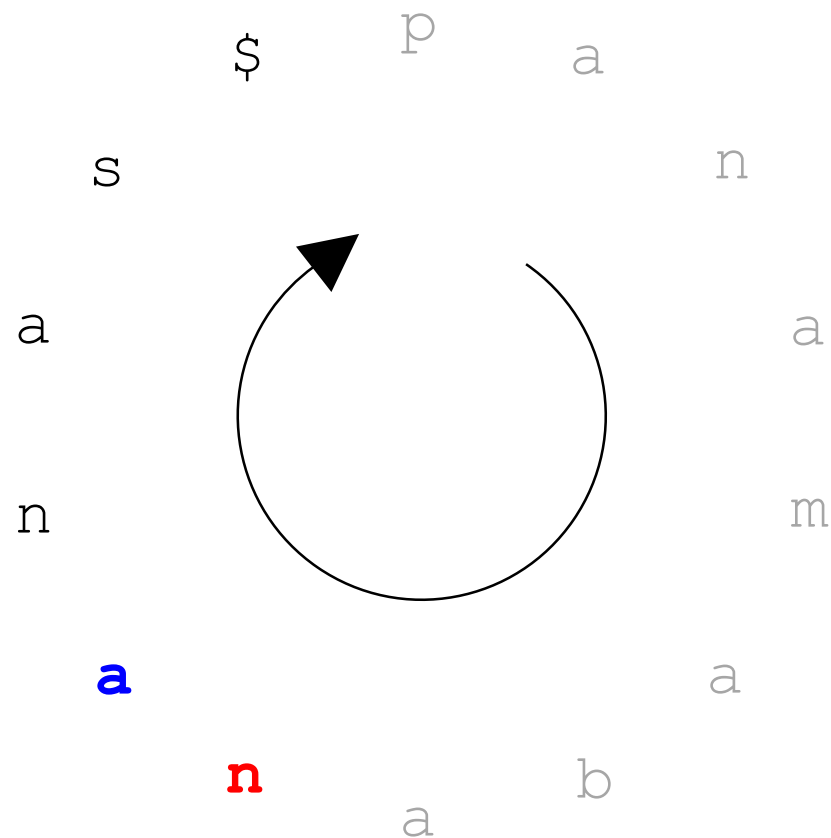
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamaban**a**₅
p₁anamabananas\$₁
s₁\$panamabanana₆



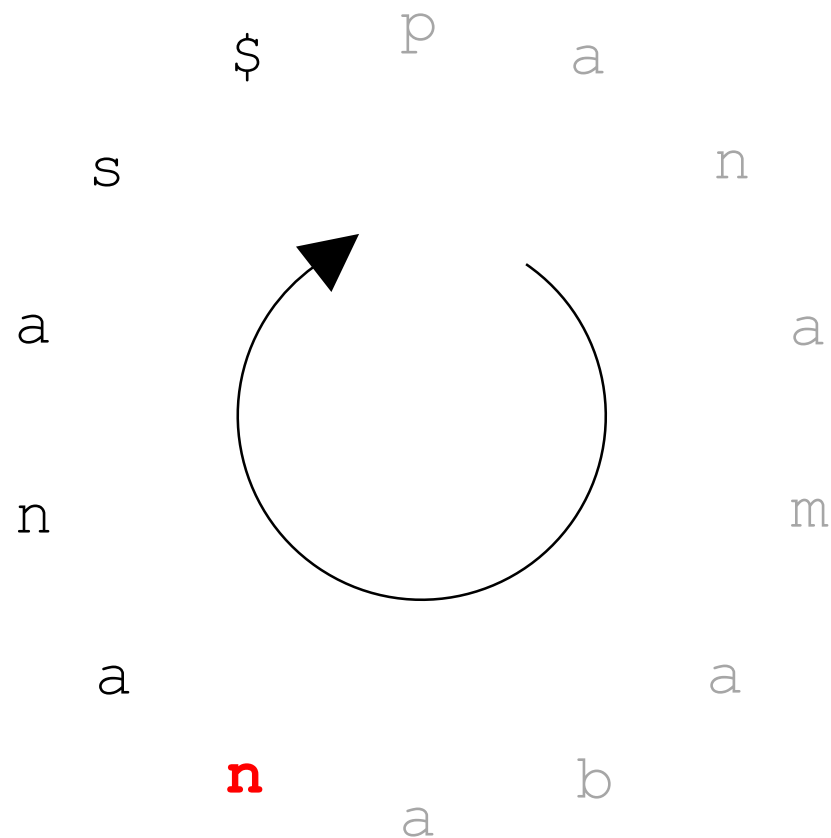
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaba**n**₂
a₆s\$panamabanana₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



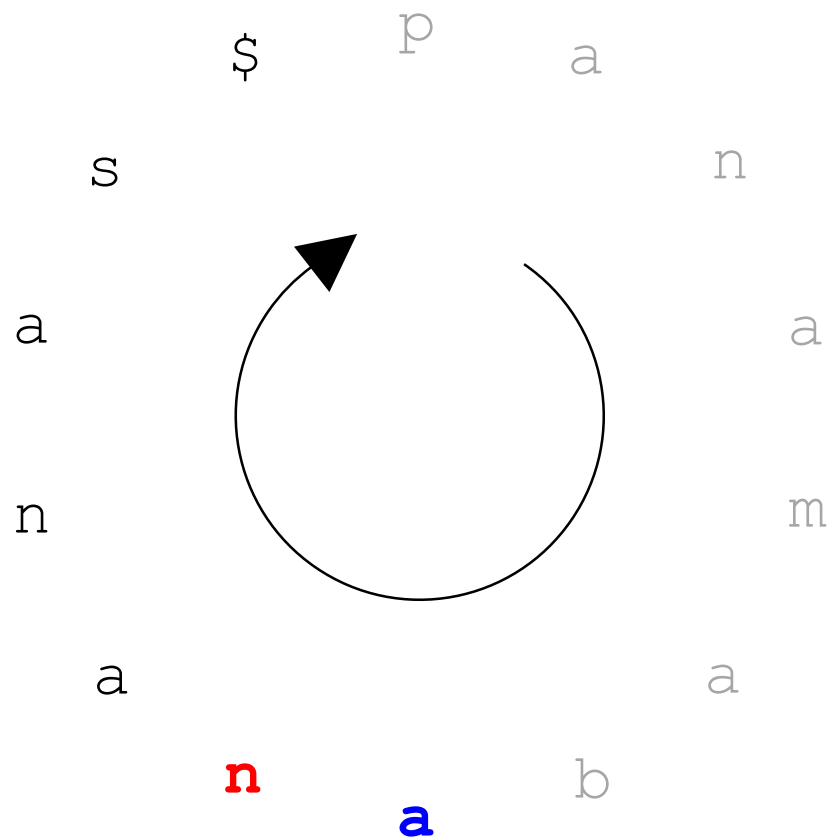
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaba**n**₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanan₆



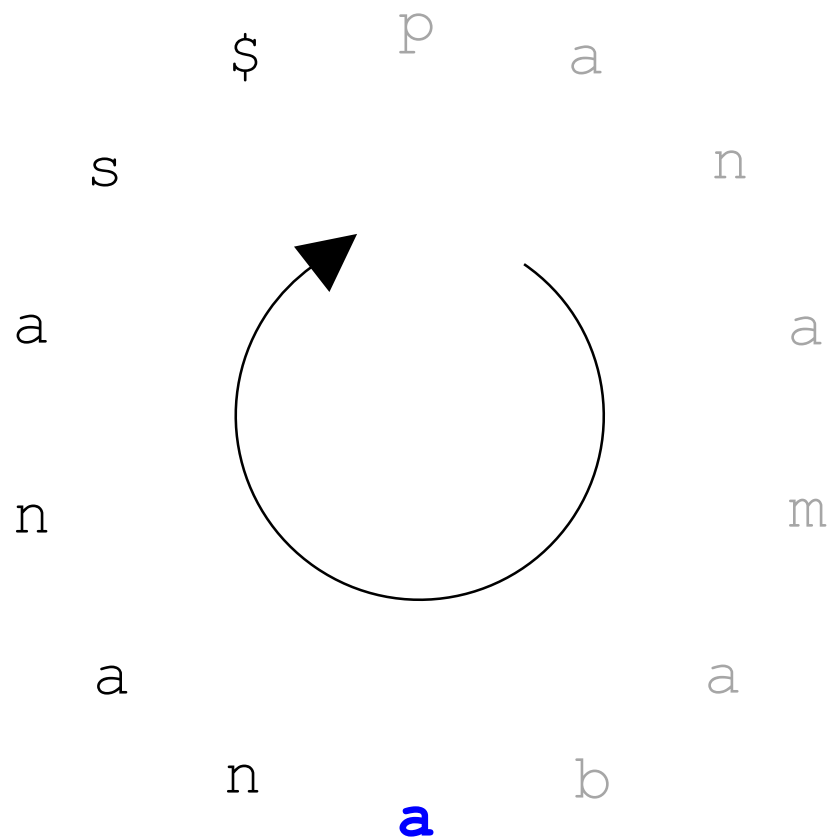
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamab**a**₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



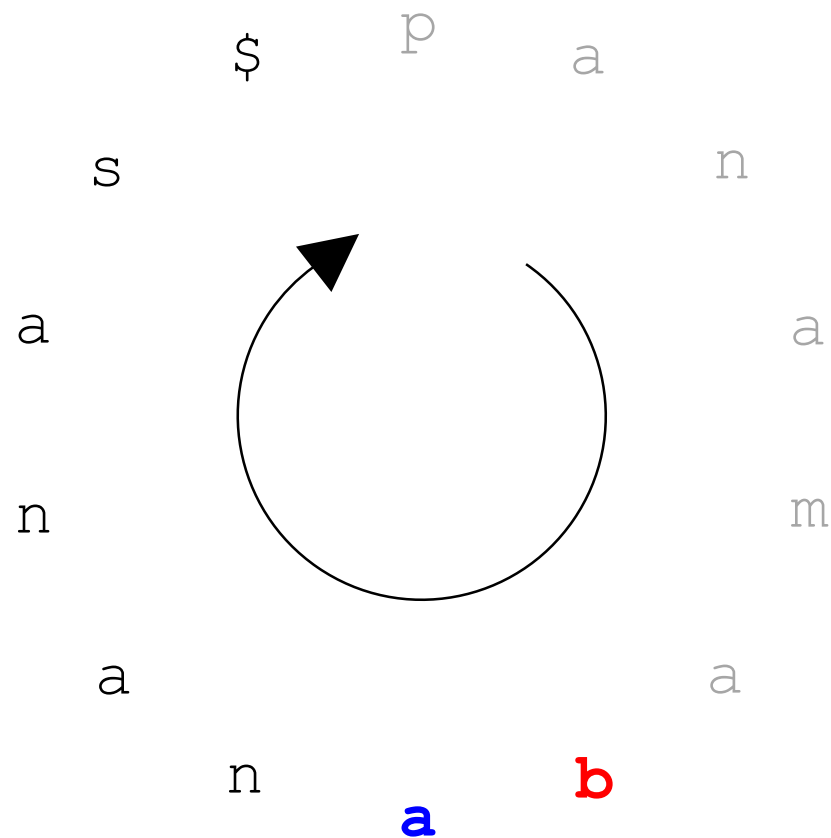
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamab**a**₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



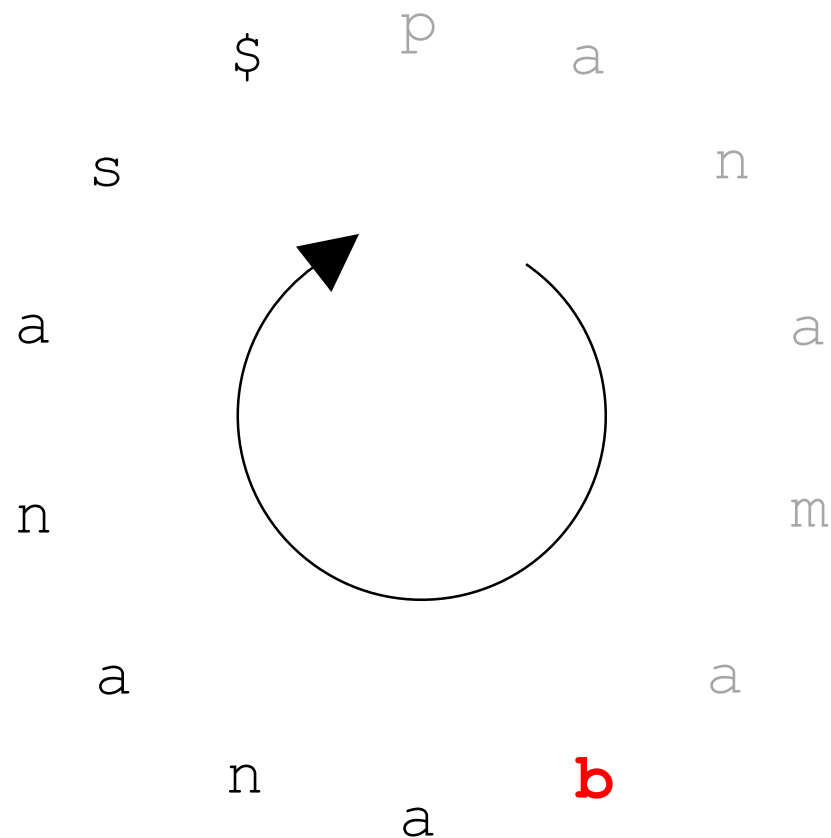
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panama**b**₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



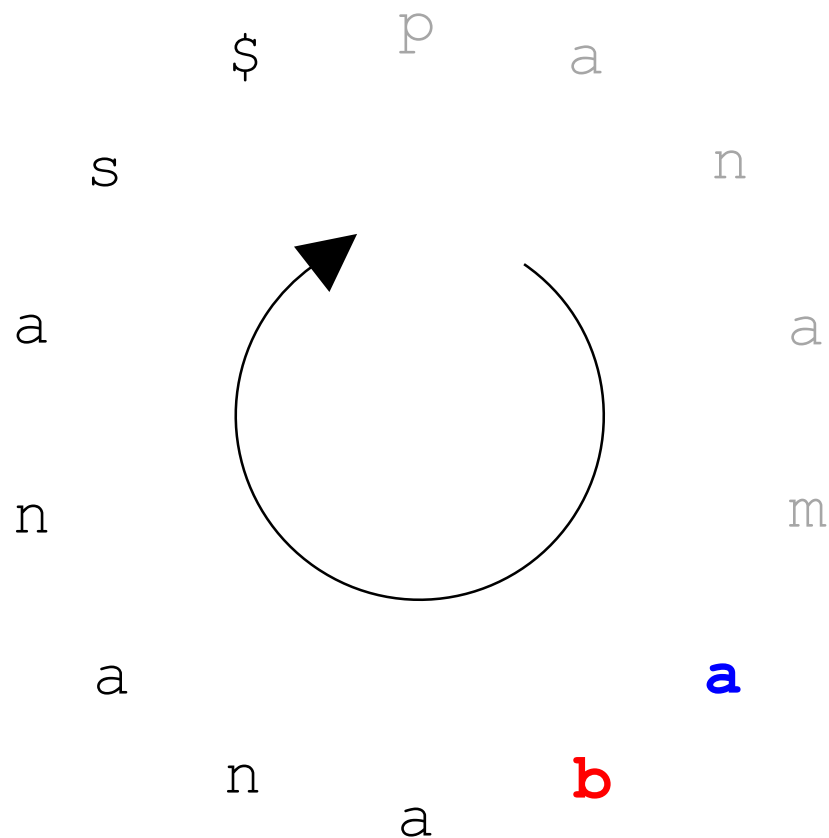
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panama**b**₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



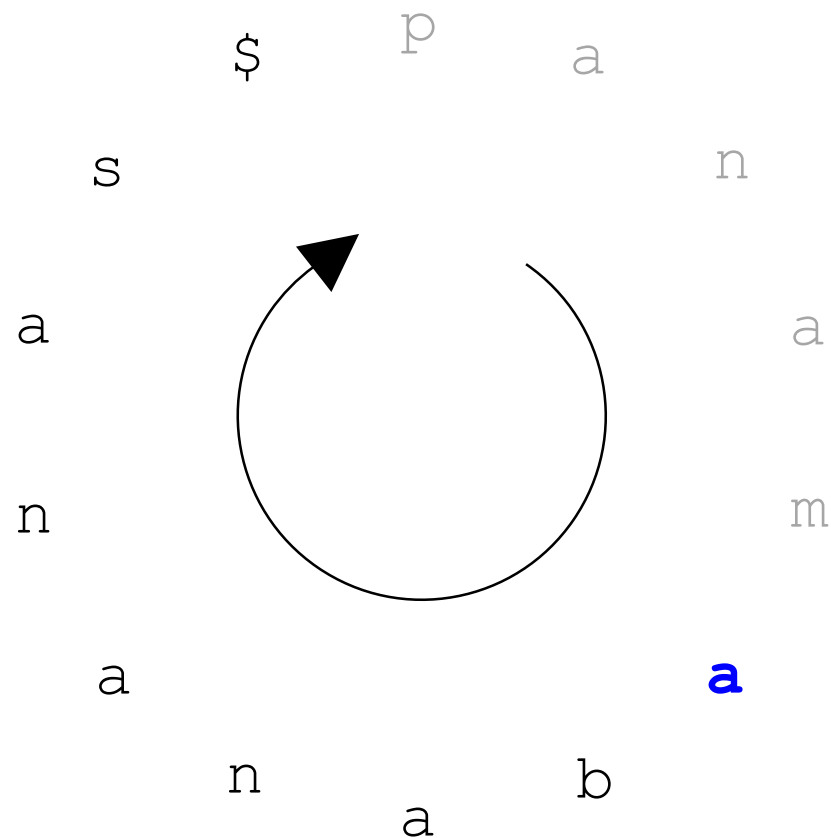
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panam**a**₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



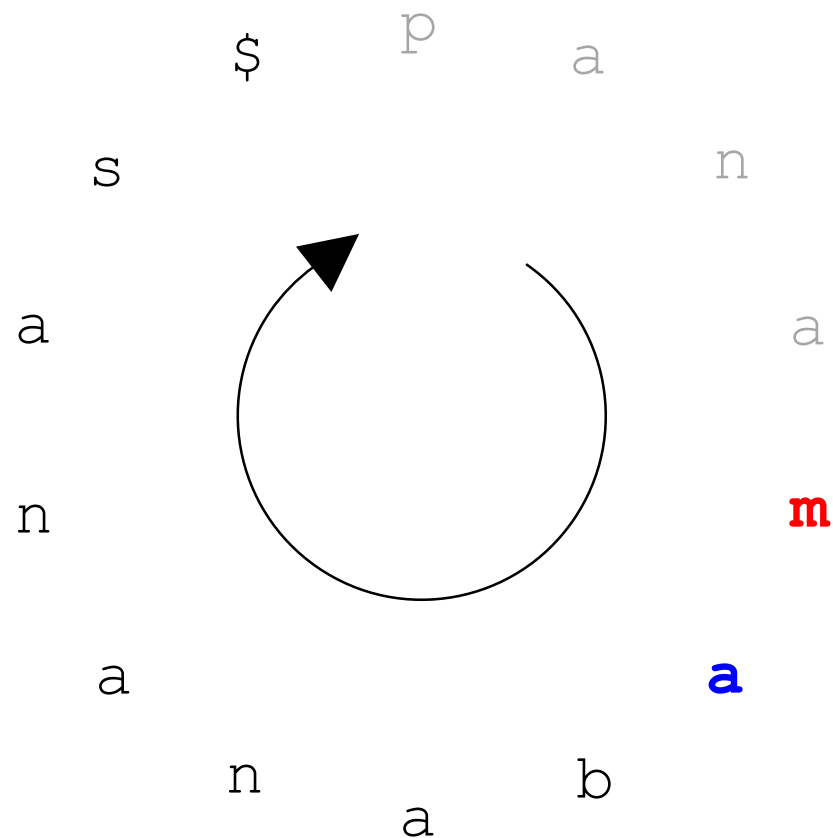
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panam**a**₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



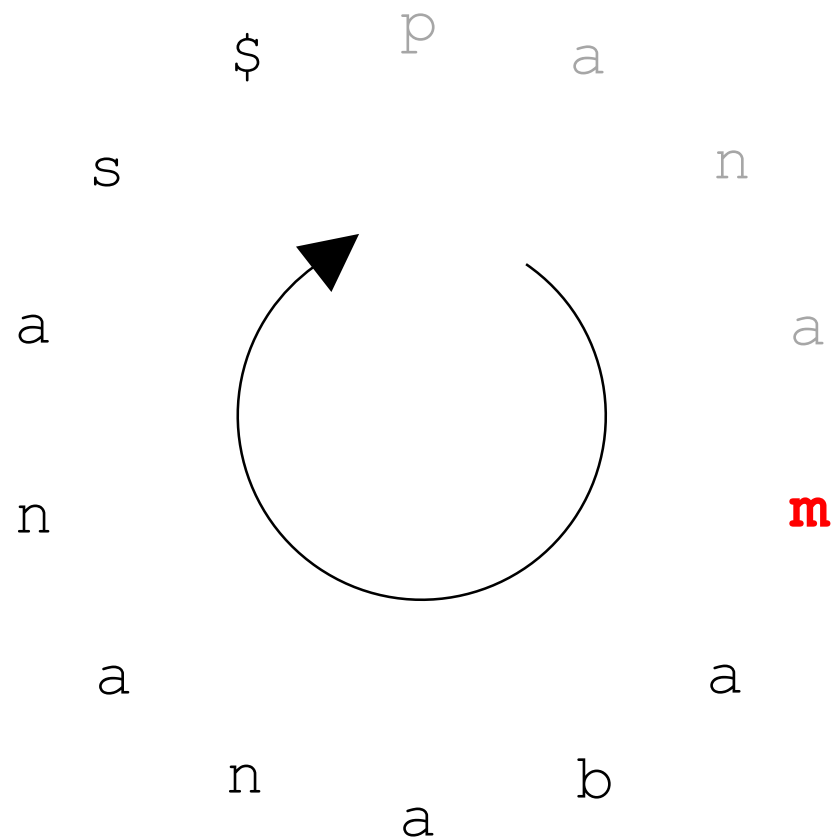
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$pana**m**₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$_₁
s₁\$panamabanana₆



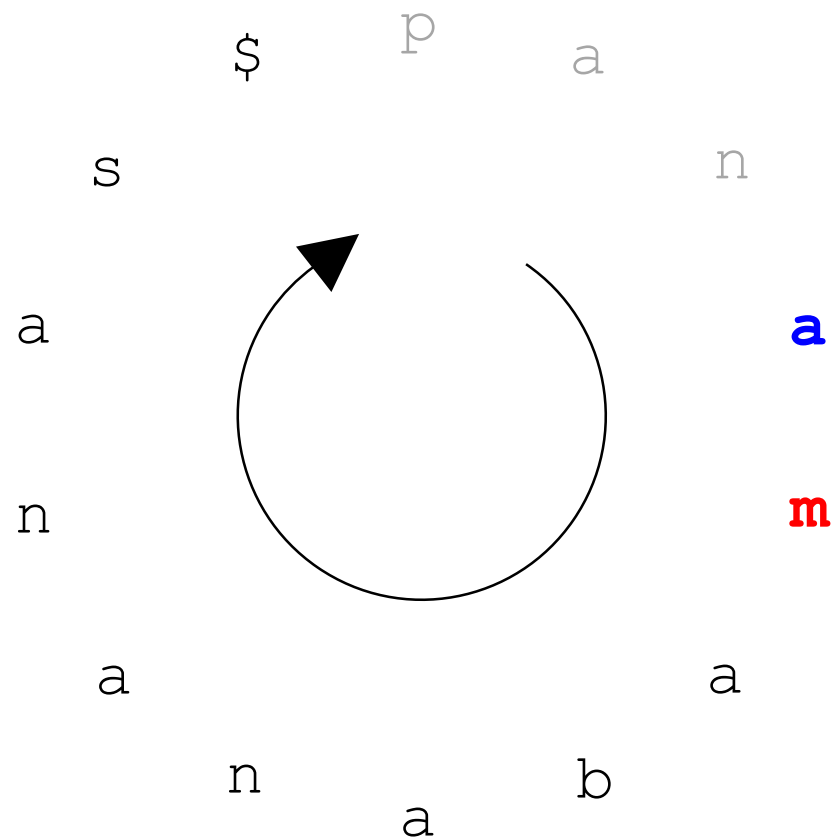
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$pana**m**₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$_₁
s₁\$panamabanana₆



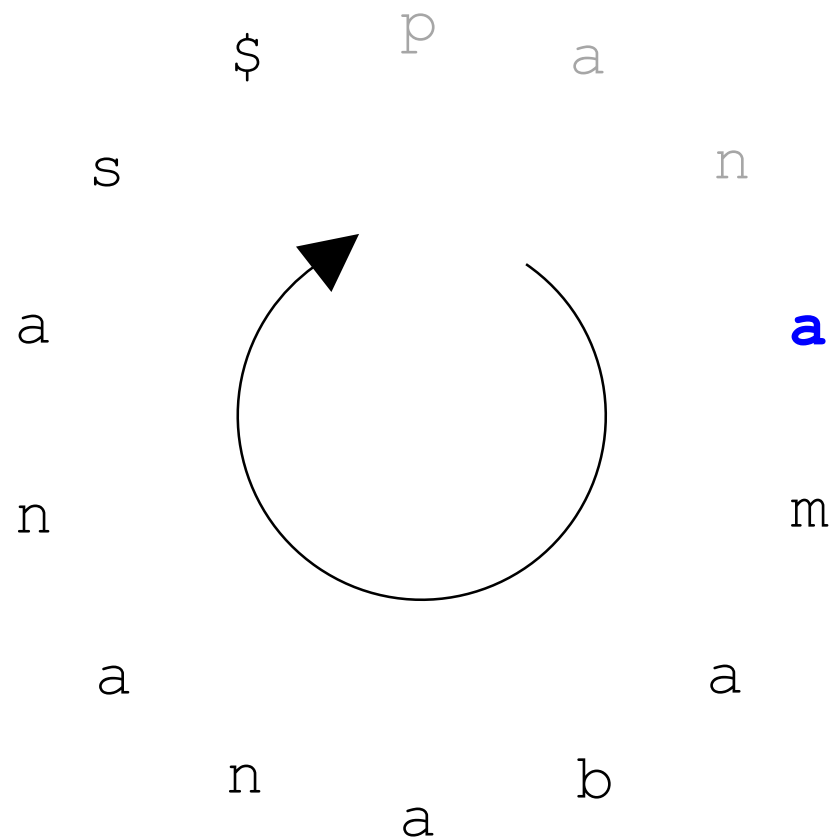
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pan**a**₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



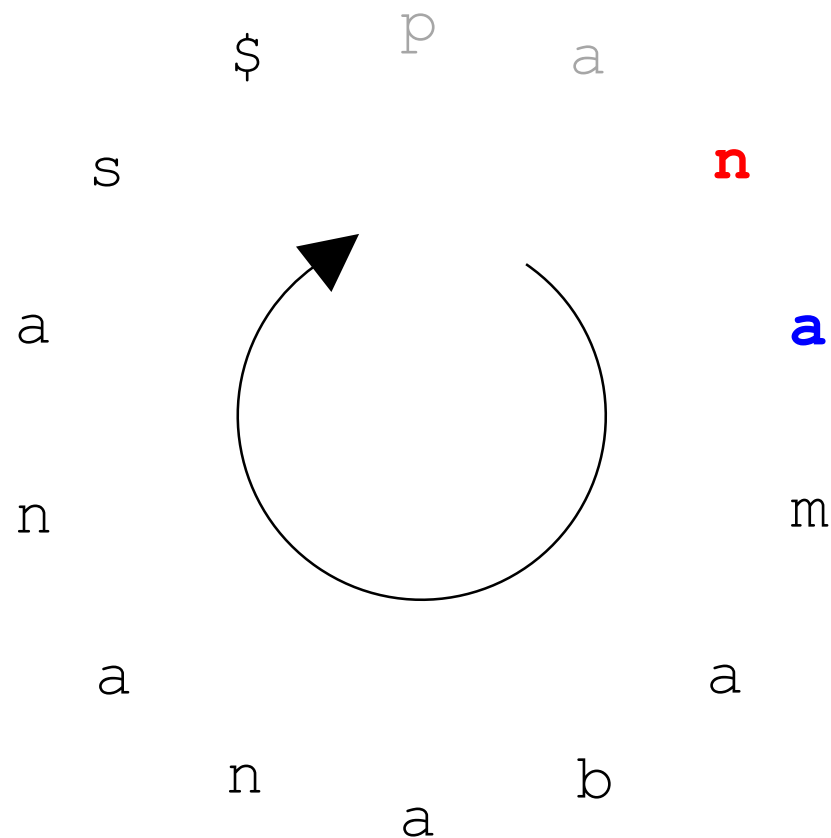
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pan**a**₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



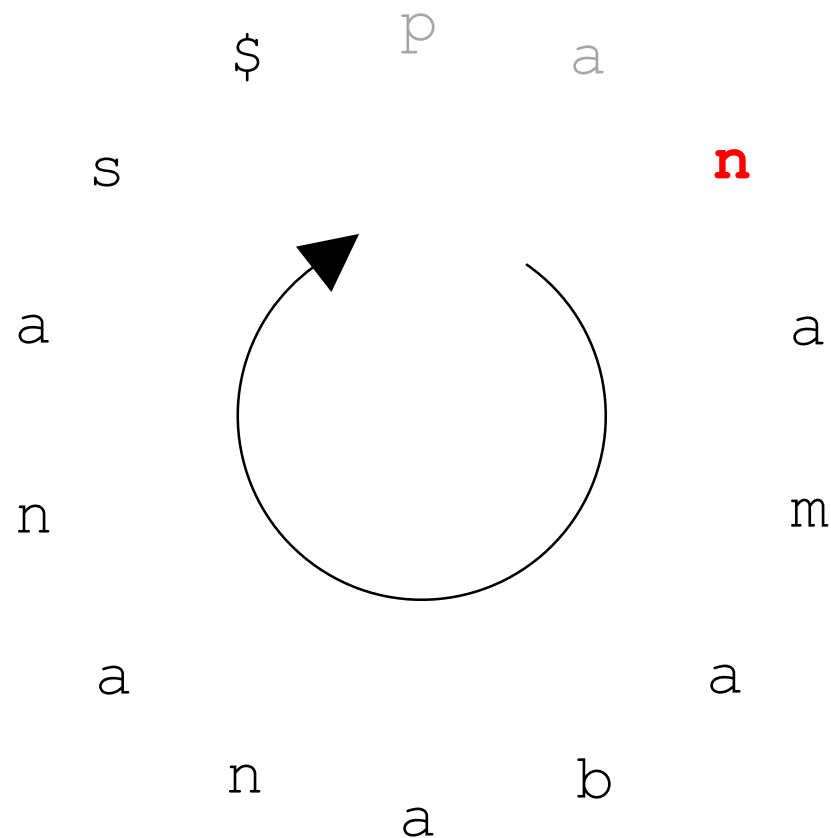
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pa**n**₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



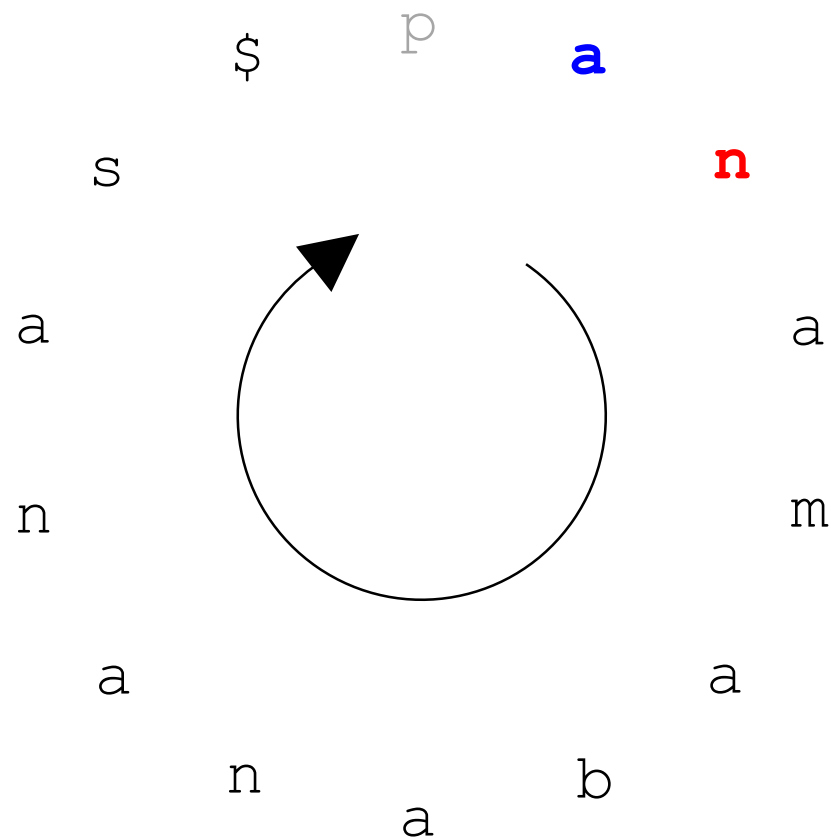
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁**n**₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$_₁
s₁\$panamabanana₆



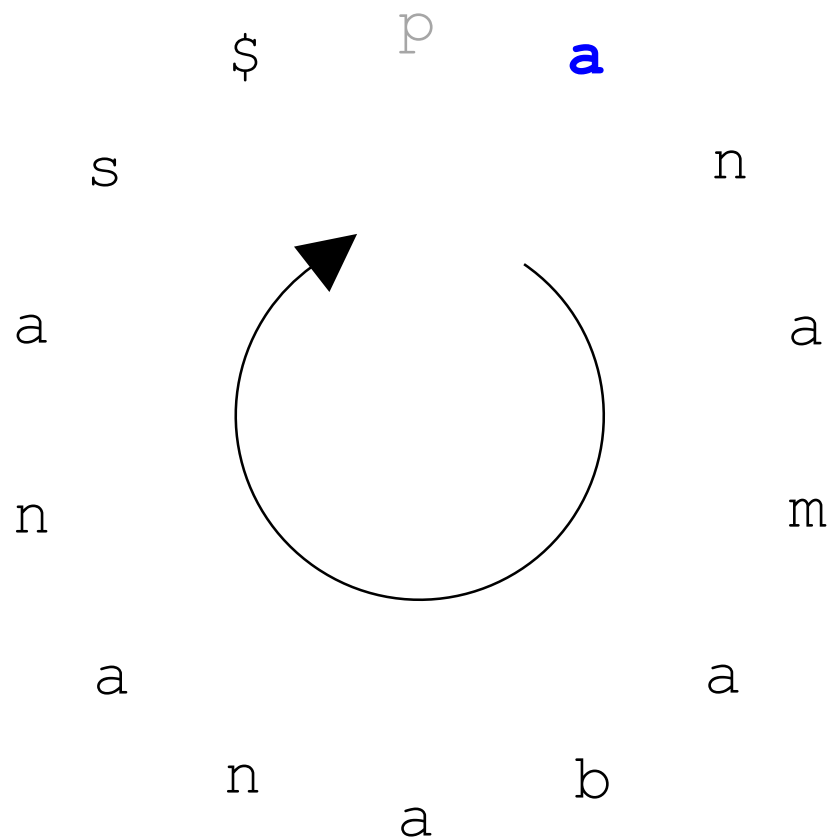
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$p**a**₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



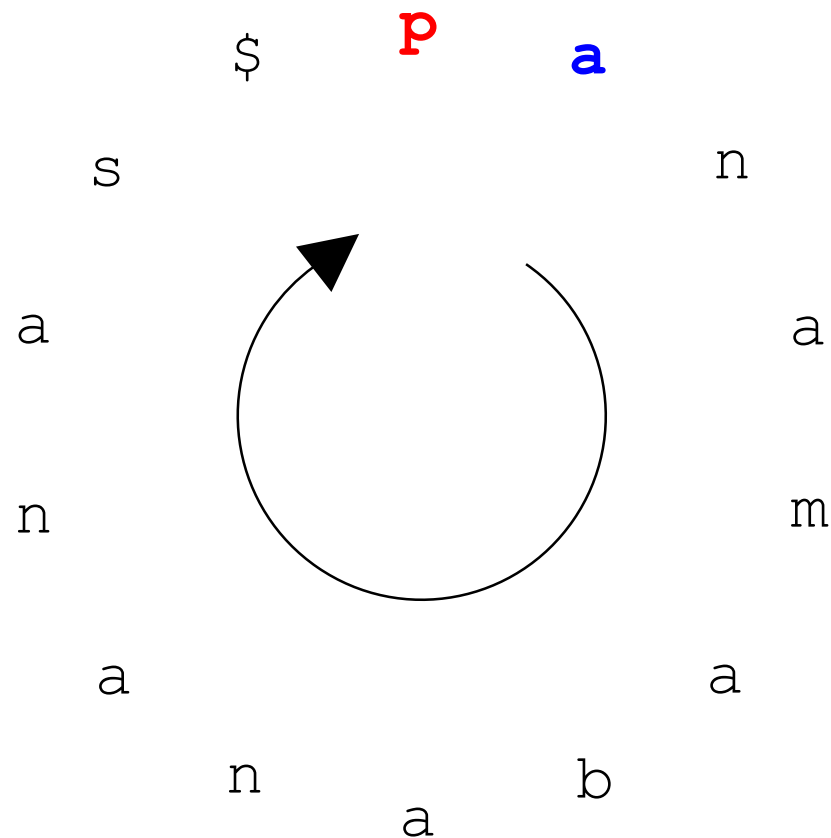
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$p**a**₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



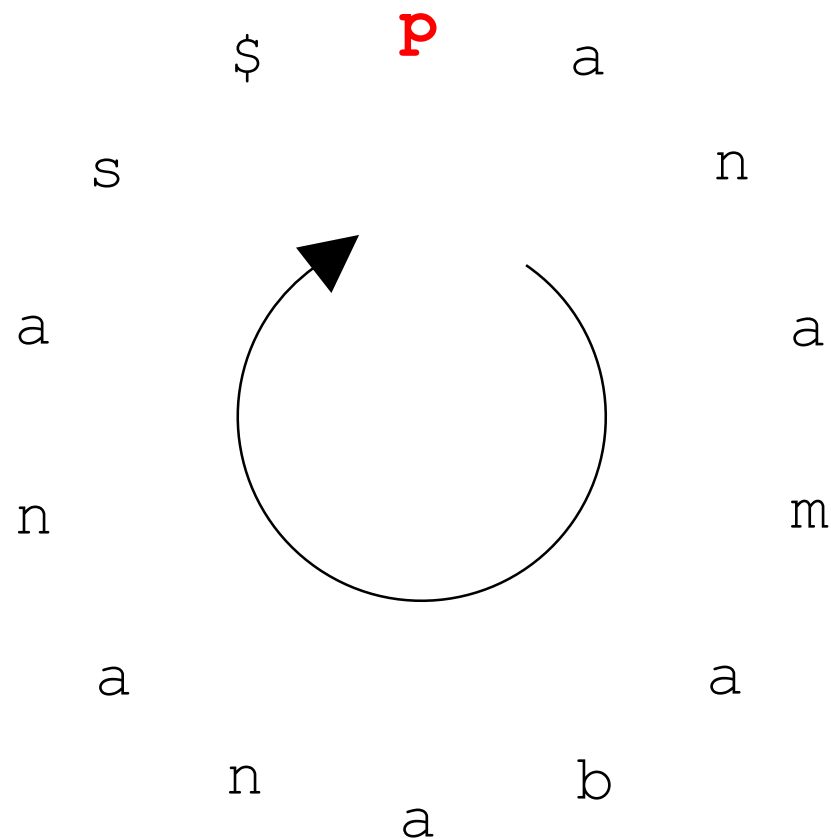
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$**p**₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆



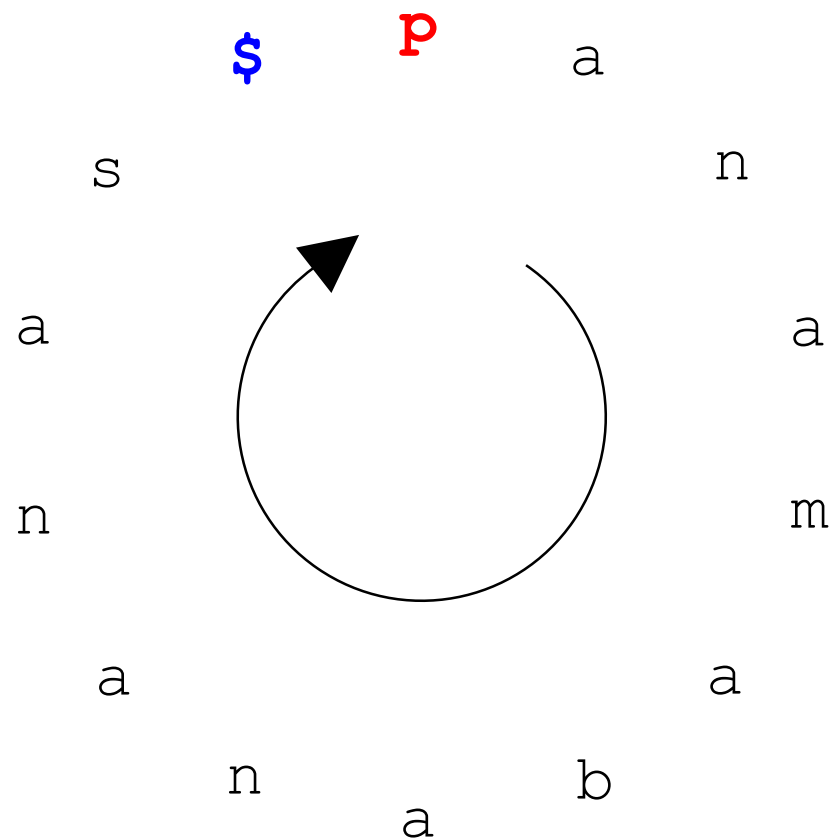
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$**P**₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamabanan₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
P₁anamabananas\$₁
s₁\$panamabanana₆



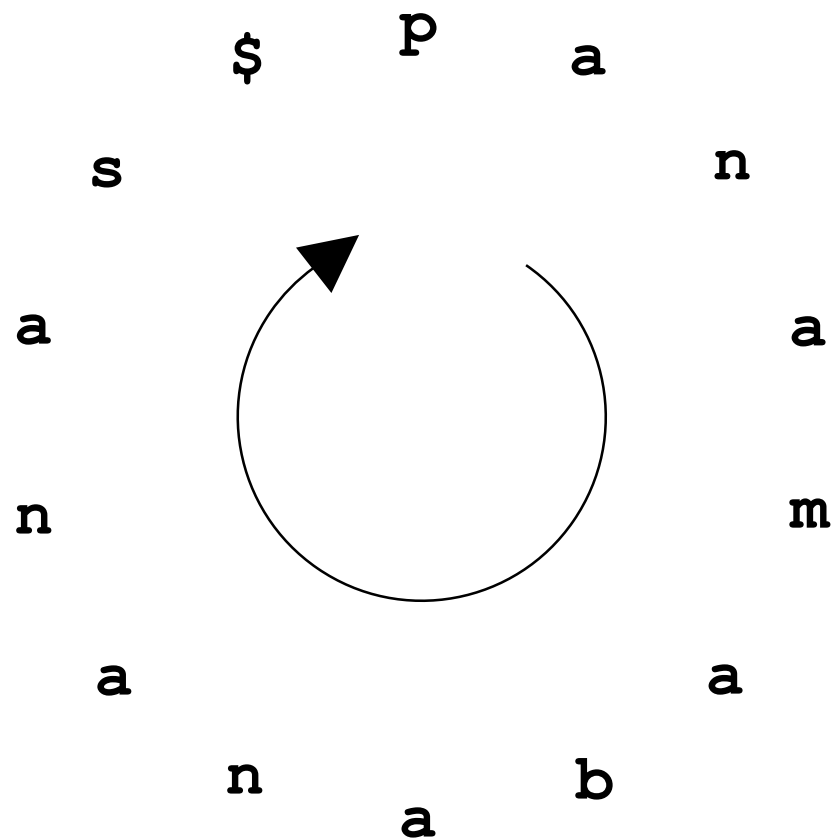
Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas**\$**₁
s₁\$panamabanana₆



Efikasnija BWT dekompresija

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$_₁
s₁\$panamabanana₆



- Prostorna složenost: $2|Genome| = O(|Genome|)$.

Pregled

- Mapiranje očitavanja
- Sufiksna stabla
- Kompresija niski i Burrows-Wheelerova transformacija
- Inverzna BWT
- **Korišćenje BWT za uparivanje šablona**

Podsećanje

- Uparivanje šablona pomoću sufiksni stabala:
 - Vremenska složenost: $O(|Genome| + |Patterns|)$
 - Prostorna složenost: $O(|Genome|)$
 - Problem: sufiksno stablo traži $2\theta \times |Genome|$ prostora
- Možemo li umesto sufiksnog stabla koristiti $BWT(Genome)$ kao strukturu podataka?

Uparivanje šablona preko BWT

- Traženje niske **ana** u niski **panamabananas**

```
$1panamabananas1
a1bananas$panam1
a2mabananas$pan1
a3namabananas$p1
a4nanas$panamab1
a5nas$panamaban2
a6s$panamaban3
b1anas$panama1
m1abananas$pana2
n1amabananas$pa3
n2anas$panamaba4
n3as$panamabana5
p1anamabananas$1
s1$panamabanana6
```

Uparivanje šablona preko BWT

- Traženje niske **ana u niski** panamabananas

```
$1panamabananas1  
a1bananas$panam1  
a2mabananas$pan1  
a3namabananas$p1  
a4nanas$panamab1  
a5nas$panamaban2  
a6s$panamaban3  
b1ananas$panama1  
m1abananas$pana2  
n1amabananas$pa3  
n2anas$panamaba4  
n3as$panamabana5  
p1anamabananas$1  
s1$panamabanana6
```

Uparivanje šablona preko BWT

- Traženje niske **ana** u niski panamabananas

```
$1panamabananas1  
a1bananas$panam1  
a2mabananas$pan1  
a3namabananas$p1  
a4nanas$panamab1  
a5nas$panamaban2  
a6s$panamaban3  
b1ananas$panama1  
m1abananas$pana2  
n1amabananas$pa3  
n2anas$panamaba4  
n3as$panamabana5  
p1anamabananas$1  
s1$panamabanana6
```

Uparivanje šablona preko BWT

- Traženje niske **ana** u niski panamabananas

```
$1panamabananas1  
a1bananas$panam1  
a2mabananas$pan1  
a3namabananas$p1  
a4nanas$panamab1  
a5nas$panamaban2  
a6s$panamabanan3  
b1anas$panama1  
m1abananas$pana2  
n1amabananas$pa3  
n2anas$panamaba4  
n3as$panamabana5  
p1anamabananas$1  
s1$panamabana6
```

Uparivanje šablona preko BWT

- Traženje niske **ana** u niski panamabananas

```
$1panamabananas1  
a1bananas$panam1  
a2mabananas$pan1  
a3namabananas$p1  
a4nanas$panamab1  
a5nas$panamaban2  
a6s$panamabanan3  
b1ananas$panama1  
m1abananas$pana2  
n1amabananas$pa3  
n2anas$panamaba4  
n3as$panamabana5  
p1anamabananas$1  
s1$panamabana6
```

Uparivanje šablona preko BWT

- Traženje niske **ana** u niski panamabananas

```
$1panamabananas1
a1bananas$panam1
a2mabananas$pan1
a3namabananas$p1
a4nanas$panamab1
a5nas$panamaban2
a6s$panamabanan3
b1anas$panama1
m1abananas$pana2
n1amabananas$pa3
n2anas$panamaba4
n3as$panamabana5
p1anamabananas$1
s1$panamabana6
```

Uparivanje šablona preko BWT

- Traženje niske **ana** u niski panamabananas

```
$1panamabananas1  
a1bananas$panam1  
a2mabananas$pan1  
a3namabananas$p1  
a4nanas$panamab1  
a5nas$panamaban2  
a6s$panamaban3  
b1ananas$panama1  
m1abananas$pana2  
n1amabananas$pa3  
n2anas$panamaba4  
n3as$panamabana5  
p1anamabananas$1  
s1$panamabanana6
```

Uparivanje šablona preko BWT

- Traženje niske **ana** u niski panamabananas

```
$1panamabananas1  
a1bananas$panam1  
a2mabananas$pan1  
a3namabananas$p1  
a4nanas$panamab1  
a5nas$panamaban2  
a6s$panamaban3  
b1ananas$panama1  
m1abananas$pana2  
n1amabananas$pa3  
n2anas$panamab4  
n3as$panamabana5  
p1anamabananas$1  
s1$panamabanana6
```


Uparivanje šablona preko BWT

- Traženje niske **ana** u niski panamabananas

```
$1panamabananas1  
a1bananas$panam1  
a2mabananas$pan1  
a3namabananas$p1  
a4nanas$panamab1  
a5nas$panamaban2  
a6s$panamaban3  
b1ananas$panama1  
m1abananas$pana2  
n1amabananas$pa3  
n2anas$panamaba4  
n3as$panamabana5  
p1anamabananas$1  
s1$panamabanana6
```

Pregled

- Mapiranje očitavanja
- Sufiksna stabla
- Kompresija niski i Burrows-Wheelerova transformacija
- Inverzna BWT
- Korišćenje BWT za uparivanje šablona
- **Pronalaženje uparenih šablona**

Gde su uparivanja?

- **Problem višestrukog uparivanja šablona:**
 - **Ulaz:** Kolekcija niski *Patterns* i niska *Genome*.
 - **Izlaz:** Sve pozicije u niski *Genome* gde se niske iz kolekcije *Patterns* pojavljuju kao podniske.

Gde su uparivanja?

- **Problem višestrukog uparivanja šablona:**
 - **Ulaz:** Kolekcija niski *Patterns* i niska *Genome*.
 - **Izlaz:** Sve **pozicije** u niski *Genome* gde se niske iz kolekcije *Patterns* pojavljuju kao podniske.
- Gde su **pozicije**? BWT nam ne daje odgovor na ovo pitanje.

Gde su uparivanja?

- Primer: Znamo da se niska **ana** pojavljuje 3 puta, ali na kojim pozicijama?

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

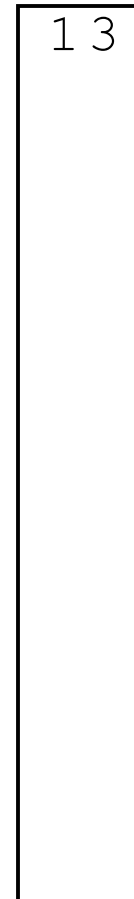
- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

```
$1panamabananas1
a1bananas$panam1
a2mabananas$pan1
a3namabananas$p1
a4nanas$panamab1
a5nas$panamaban2
a6s$panamaban3
b1ananas$panama1
m1abananas$pana2
n1amabananas$pa3
n2anas$panamaba4
n3as$panamabana5
p1anamabananas$1
s1$panamabanana6
```

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panamabananas\$



```

$1panamabananas1
a1bananas$panam1
a2mabananas$pan1
a3namabananas$p1
a4nanas$panamab1
a5nas$panamaban2
a6s$panamaban3
b1ananas$panama1
m1abananas$pana2
n1amabananas$pa3
n2anas$panamaba4
n3as$panamabana5
p1anamabananas$1
s1$panamabanana6
    
```

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panam**abananas**\$

1	3
5	

\$₁panamabananas₁
a₁**bananas**\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$pa₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

pan**amabananas**\$

1 3	\$ ₁ panamabananas ₁
5	a ₁ bananas\$panam ₁
3	a ₂ mabananas\$pan ₁
	a ₃ namabananas\$pa ₁
	a ₄ nanas\$panamab ₁
	a ₅ nas\$panamaban ₂
	a ₆ s\$panamaban ₃
	b ₁ ananas\$panama ₁
	m ₁ abananas\$pana ₂
	n ₁ amabananas\$pa ₃
	n ₂ anas\$panamaba ₄
	n ₃ as\$panamabana ₅
	p ₁ anamabananas\$ ₁
	s ₁ \$panamabanana ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

pan**a**mab**a**nan**a**s**\$**

1 3	\$ ₁ p	a	n	a	m	a	b	a	n	a	n	a	s	\$ ₁	
5	a ₁	b	a	n	a	n	a	s	\$	p	a	n	a	m	\$ ₁
3	a ₂	m	a	b	a	n	a	n	a	s	\$	p	a	n	\$ ₁
1	a ₃	n	a	m	a	b	a	n	a	n	a	s	\$	p	\$ ₁
	a ₄	n	a	n	a	s	\$	p	a	n	a	m	a	b	\$ ₁
	a ₅	n	a	s	\$	p	a	n	a	m	a	b	a	n	\$ ₂
	a ₆	s	\$	p	a	n	a	m	a	b	a	n	a	n	\$ ₃
	b ₁	a	n	a	n	a	s	\$	p	a	n	a	m	a	\$ ₁
	m ₁	a	b	a	n	a	n	a	s	\$	p	a	n	a	\$ ₂
	n ₁	a	m	a	b	a	n	a	n	a	s	\$	p	a	\$ ₃
	n ₂	a	n	a	s	\$	p	a	n	a	m	a	b	a	\$ ₄
	n ₃	a	s	\$	p	a	n	a	m	a	b	a	n	a	\$ ₅
	p ₁	a	n	a	m	a	b	a	n	a	n	a	s	\$	\$ ₁
	s ₁	\$	p	a	n	a	m	a	b	a	n	a	n	a	\$ ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panamab**ananas**\$

1 3	\$ ₁ panamabananas ₁
5	a ₁ bananas\$panam ₁
3	a ₂ mabananas\$pan ₁
1	a ₃ namabananas\$ _p ₁
7	a ₄ nanas\$panamab ₁
	a ₅ nas\$panamaban ₂
	a ₆ s\$panamabanan ₃
	b ₁ ananas\$panama ₁
	m ₁ abananas\$pana ₂
	n ₁ amabananas\$pa ₃
	n ₂ anas\$panamaba ₄
	n ₃ as\$panamabana ₅
	p ₁ anamabananas\$ ₁
	s ₁ \$panamabanana ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panamabanas\$

1 3	\$ ₁ panamabanas ₁
5	a ₁ bananas\$panam ₁
3	a ₂ mabanas\$pan ₁
1	a ₃ namabanas\$pa ₁
7	a ₄ nanas\$panamab ₁
9	a ₅ nas\$panamaban ₂
	a ₆ s\$panamabanan ₃
	b ₁ anas\$panama ₁
	m ₁ abanas\$pana ₂
	n ₁ amabanas\$pa ₃
	n ₂ anas\$panamaba ₄
	n ₃ as\$panamabana ₅
	p ₁ anamabanas\$ ₁
	s ₁ \$panamabana ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panamabanan**as**\$

1 3	\$ ₁ panamabananas ₁
5	a ₁ bananas\$panam ₁
3	a ₂ mabanas\$pan ₁
1	a ₃ namabanas\$ _p ₁
7	a ₄ nanas\$panamab ₁
9	a ₅ nas\$panamaban ₂
1 1	a ₆ s\$panamaban ₃
	b ₁ ananas\$panama ₁
	m ₁ abanas\$pana ₂
	n ₁ amabanas\$pa ₃
	n ₂ anas\$panamaba ₄
	n ₃ as\$panamabana ₅
	p ₁ anamabananas\$ ₁
	s ₁ \$panamabanana ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panama**bananas**\$

1 3	\$ ₁ panamabananas ₁
5	a ₁ bananas\$panam ₁
3	a ₂ mabananas\$pan ₁
1	a ₃ namabananas\$ _p ₁
7	a ₄ nanas\$panamab ₁
9	a ₅ nas\$panamaban ₂
1 1	a ₆ s\$panamaban ₃
6	b ₁ ananas\$panama ₁
	m ₁ abananas\$pana ₂
	n ₁ amabananas\$pa ₃
	n ₂ anas\$panamaba ₄
	n ₃ as\$panamabana ₅
	p ₁ anamabananas\$ ₁
	s ₁ \$panamabanana ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panam**abananas**\$

1 3	\$ ₁ panamabananas ₁
5	a ₁ bananas\$panam ₁
3	a ₂ mabananas\$pan ₁
1	a ₃ namabananas\$ _p ₁
7	a ₄ nanas\$panamab ₁
9	a ₅ nas\$panamaban ₂
1 1	a ₆ s\$panamaban ₃
6	b ₁ ananas\$panama ₁
4	m ₁ abananas\$pana ₂
	n ₁ amabananas\$pa ₃
	n ₂ anas\$panamaba ₄
	n ₃ as\$panamabana ₅
	p ₁ anamabananas\$ ₁
	s ₁ \$panamabanana ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panamabananas\$

1 3	\$ ₁ panamabananas ₁
5	a ₁ bananas\$panam ₁
3	a ₂ mabananas\$pan ₁
1	a ₃ namabananas\$pa ₁
7	a ₄ nanas\$panamab ₁
9	a ₅ nas\$panamaban ₂
1 1	a ₆ s\$panamaban ₃
6	b ₁ ananas\$panama ₁
4	m ₁ abananas\$pana ₂
2	n ₁ amabananas\$pa ₃
	n ₂ anas\$panamaba ₄
	n ₃ as\$panamabana ₅
	p ₁ anamabananas\$ ₁
	s ₁ \$panamabanana ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panamab**ananas**\$

1 3	\$ ₁ panamab ananas s ₁
5	a ₁ bananas \$panam ₁
3	a ₂ mabanas \$pan ₁
1	a ₃ namabanas \$p ₁
7	a ₄ nanas \$panamab ₁
9	a ₅ nas \$panamaban ₂
1 1	a ₆ s \$panamabanan ₃
6	b ₁ ananas \$panama ₁
4	m ₁ abanas \$pana ₂
2	n ₁ amabanas \$pa ₃
8	n ₂ anas \$panamaba ₄
	n ₃ as\$panamabana ₅
	p ₁ anamabanas\$ ₁
	s ₁ \$panamabana ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panamabanan**as**\$

1 3	\$ ₁ panamabananas ₁
5	a ₁ bananas\$panam ₁
3	a ₂ mabananas\$pan ₁
1	a ₃ namabananas\$pa ₁
7	a ₄ nanas\$panamab ₁
9	a ₅ nas\$panamaban ₂
1 1	a ₆ s\$panamabanan ₃
6	b ₁ ananas\$panama ₁
4	m ₁ abananas\$pana ₂
2	n ₁ amabananas\$pa ₃
8	n ₂ anas\$panamaba ₄
1 0	n ₃ as\$panamabana ₅
	p ₁ anamabananas\$ ₁
	s ₁ \$panamabanan ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panamabananas\$

1 3	\$ ₁ panamabananas ₁
5	a ₁ bananas\$panam ₁
3	a ₂ mabananas\$pan ₁
1	a ₃ namabananas\$pa ₁
7	a ₄ nanas\$panamab ₁
9	a ₅ nas\$panamaban ₂
1 1	a ₆ s\$panamabanana ₃
6	b ₁ ananas\$panama ₁
4	m ₁ abananas\$pana ₂
2	n ₁ amabananas\$pa ₃
8	n ₂ anas\$panamaba ₄
1 0	n ₃ as\$panamabana ₅
0	p ₁ anamabananas\$ ₁
	s ₁ \$panamabanana ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panamabanana**s**\$

1 3	\$ ₁ panamabananas ₁
5	a ₁ bananas\$panam ₁
3	a ₂ mabananas\$pan ₁
1	a ₃ namabananas\$ _p ₁
7	a ₄ nanas\$panamab ₁
9	a ₅ nas\$panamaban ₂
1 1	a ₆ s\$panamabanana ₃
6	b ₁ ananas\$panama ₁
4	m ₁ abananas\$pana ₂
2	n ₁ amabananas\$pa ₃
8	n ₂ anas\$panamaba ₄
1 0	n ₃ as\$panamabana ₅
0	p ₁ anamabananas\$ ₁
1 2	s ₁ \$panamabanana ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).

panamabananas\$

1 3	\$ ₁ panamabananas ₁
5	a ₁ bananas\$panam ₁
3	a ₂ mabananas\$pan ₁
1	a ₃ namabananas\$p ₁
7	a ₄ nanas\$panamab ₁
9	a ₅ nas\$panamaban ₂
1 1	a ₆ s\$panamaban ₃
6	b ₁ ananas\$panama ₁
4	m ₁ abananas\$pana ₂
2	n ₁ amabananas\$pa ₃
8	n ₂ anas\$panamaba ₄
1 0	n ₃ as\$panamabana ₅
0	p ₁ anamabananas\$ ₁
1 2	s ₁ \$panamabanana ₆

Nalaženje pozicija uparivanja pomoću sufiksnog niza

- **Sufiksni niz:** čuva početnu poziciju za svaki sufiks (niz karaktera u svakom redu matrice do simbola \$).
- Niska **ana** se pojavljuje na pozicijama **1, 7, 9** niske **p^{ana}mab^{anana}s\$**.



1 3	\$ ₁ p ^{ana} mab ^{anana} s ₁
5	a ₁ b ^{anana} s ₁ \$p ^{ana} m ₁
3	a ₂ m ^{abana} n ^{ana} s ₁ \$p ^{ana} n ₁
1	a₃n^{ana} m ^{abana} n ^{ana} s ₁ \$p ₁
7	a₄n^{ana} n ^{ana} s ₁ \$p ^{ana} m ^{ab} ₁
9	a₅n^{ana} s ₁ \$p ^{ana} m ^{aban} ₂
1 1	a ₆ s ₁ \$p ^{ana} m ^{abanana} ₃
6	b ₁ a ^{nana} s ₁ \$p ^{ana} m ^a ₁
4	m ₁ a ^{banana} s ₁ \$p ^{ana} ₂
2	n ₁ a ^{mabana} n ^{ana} s ₁ \$p ^a ₃
8	n ₂ a ^{na} s ₁ \$p ^{anamaba} ₄
1 0	n ₃ a ^s \$p ^{anamabana} ₅
0	p ₁ a ^{namabana} n ^{ana} s ₁
1 2	s ₁ \$p ^{anamabana} ₆

Sufiksni niz: prostorna složenost

- Prostorna složenost: $\sim 4 \times |Genome|$ (ako koristimo 4B cele brojeve kao elemente niza)
- Bolje nego $20 \times |Genome|$
- Postoji pristup sa $|Genome|$ prostornom složenošću (parcijalni sufiksni niz)

Pregled

- Mapiranje očitavanja
- Sufiksna stabla
- Kompresija niski i Burrows-Wheelerova transformacija
- Inverzna BWT
- Korišćenje BWT za uparivanje šablona
- Pronalaženje uparenih šablona
- **Približno preklapanje**

Približno preklapanje

- Ponekad je neophodno pronaći približna uparivanja šablona
- **Približno uparivanje šablona:**
 - **Ulaz:** Niska *Pattern*, niska *Genome*, ceo broj d .
 - **Izlaz:** Sve pozicije niske *Genome* gde se niska *Pattern* pojavljuje kao podniska sa najviše d razlika.

Približno preklapanje

- Ponekad je neophodno pronaći približna uparivanja šablona
- **Višestruko** približno uparivanje šablona:
 - **Ulaz:** Kolekcija niski *Patterns*, niska *Genome*, ceo broj d .
 - **Izlaz:** Sve pozicije niske *Genome* gde se niska iz *Patterns* pojavljuje kao podniska sa najviše d razlika.

BWT

- Podsećanje: traženje niske **ana u niski** panamabananas

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

\$₁panamabananas₁
a₁bananas\$panam₁
a₂mabananas\$pan₁
a₃namabananas\$p₁
a₄nanas\$panamab₁
a₅nas\$panamaban₂
a₆s\$panamaban₃
b₁ananas\$panama₁
m₁abananas\$pana₂
n₁amabananas\$pa₃
n₂anas\$panamaba₄
n₃as\$panamabana₅
p₁anamabananas\$₁
s₁\$panamabanana₆

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

```
$1panamabananas1  
a1bananas$panam1  
a2mabananas$pan1  
a3namabananas$p1  
a4nanas$panamab1  
a5nas$panamaban2  
a6s$panamabanan3  
b1anas$panama1  
m1abananas$pana2  
n1amabananas$pa3  
n2anas$panamaba4  
n3as$panamabana5  
p1anamabananas$1  
s1$panamabana6
```

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

Ako dozvolimo
jedan
mismatch,
onda moramo
zadržati
crvena slova.

```
$1panamabananas1  
a1bananas$panam1  
a2mabananas$pan1  
a3namabananas$p1  
a4nanas$panamab1  
a5nas$panamaban2  
a6s$panamabanan3  
b1anas$panama1  
m1abananas$pana2  
n1amabananas$pa3  
n2anas$panamaba4  
n3as$panamabana5  
p1anamabananas$1  
s1$panamabana6
```

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

Mismatches

Ako dozvolimo
jedan
mismatch,
onda moramo
zadržati
crvena slova.

```
$1panamabananas1  
a1bananas$panam1  
a2mabananas$pan1  
a3namabananas$p1  
a4nanas$panamab1  
a5nas$panamaban2  
a6s$panamabanan3  
b1ananas$panama1  
m1abananas$pana2  
n1amabananas$pa3  
n2anas$panamaba4  
n3as$panamabana5  
p1anamabananas$1  
s1$panamabannana6
```

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

Ako dozvolimo
jedan
mismatch,
onda moramo
zadržati
crvena slova.

	# Mismatches
\$ ₁ panamabananas ₁	
a ₁ bananas\$pana m ₁	1
a ₂ mabananas\$pa n ₁	0
a ₃ namabananas\$p p ₁	1
a ₄ nanas\$panama b ₁	1
a ₅ nas\$panamaba n ₂	0
a ₆ s\$panamabana n ₃	0
b ₁ ananas\$panama ₁	
m ₁ abananas\$pana ₂	
n ₁ amabananas\$pa ₃	
n ₂ anas\$panamaba ₄	
n ₃ as\$panamabana ₅	
p ₁ anamabananas\$ ₁	
s ₁ \$panamabanaa ₆	

Mismatches

\$ ₁ panamabananas ₁	
a ₁ bananas\$pana m ₁	1
a ₂ mabananas\$pa n ₁	0
a ₃ namabananas\$ p ₁	1
a ₄ nanas\$panama b ₁	1
a ₅ nas\$panamaba n ₂	0
a ₆ s\$panamabana n ₃	0
b ₁ ananas\$panama ₁	
m ₁ abananas\$pana ₂	
n ₁ amabananas\$pa ₃	
n ₂ anas\$panamaba ₄	
n ₃ as\$panamabana ₅	
p ₁ anamabananas\$ ₁	
s ₁ \$panamabanana ₆	

Mismatches

\$ ₁ panamabananas ₁	
a ₁ bananas\$panam ₁	
a ₂ mabananas\$pan ₁	
a ₃ namabananas\$ _p ₁	
a ₄ nanas\$panamab ₁	
a ₅ nas\$panamaban ₂	
a ₆ s\$panamabanan ₃	
b ₁ a ananas\$panama ₁	1
m ₁ a abananas\$pana ₂	1
n ₁ a mabananas\$pa ₃	0
n ₂ a anas\$panamaba ₄	0
n ₃ a s\$panamabana ₅	0
p ₁ a namabananas\$ ₁	1
s ₁ \$panamabanana ₆	

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

Proširujemo sve niske sa razlikom na najviše jednoj poziciji.

	# Mismatches
\$ ₁ panamabananas ₁	
a ₁ bananas\$pana m ₁	1
a ₂ mabananas\$pa n ₁	0
a ₃ namabananas\$ p ₁	1
a ₄ nanas\$panama b ₁	1
a ₅ nas\$panamaba n ₂	0
a ₆ s\$panamabana n ₃	0
b ₁ ananas\$panama ₁	
m ₁ abananas\$pana ₂	
n ₁ amabananas\$pa ₃	
n ₂ anas\$panamaba ₄	
n ₃ as\$panamabana ₅	
p ₁ anamabananas\$ ₁	
s ₁ \$panamabana ₆	

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

Proširujemo sve niske sa razlikom na najviše jednoj poziciji.

	# Mismatches
\$ ₁ panamabananas ₁	
a ₁ bananas\$pana m ₁	1
a ₂ mabananas\$pa n ₁	0
a ₃ namabananas\$p a ₁	1
a ₄ nanas\$panama b ₁	1
a ₅ nas\$panamaba n ₂	0
a ₆ s\$panamabana n ₃	0
b ₁ a nanas\$panama ₁	
m ₁ a bananas\$pana ₂	
n ₁ a mabananas\$pa ₃	
n ₂ a nas\$panamaba ₄	
n ₃ a s\$panamabana ₅	
p ₁ a namabananas\$ ₁	
s ₁ \$panamabana ₆	

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

Proširujemo sve niske sa razlikom na najviše jednoj poziciji.

	# Mismatches
\$ ₁ panamabananas ₁	
a ₁ bananas\$pana m ₁	1
a ₂ mabananas\$pa n ₁	0
a ₃ namabananas\$p a ₁	1
a ₄ nanas\$panama b ₁	1
a ₅ nas\$panamaba n ₂	0
a ₆ s\$panamabana n ₃	0
b ₁ a nanas\$panama ₁	
m ₁ a bananas\$pana ₂	
n ₁ a mabananas\$pa ₃	
n ₂ a nas\$panamaba ₄	
n ₃ a s\$panamabana ₅	
p ₁ a namabananas\$ ₁	
s ₁ \$panamabana ₆	

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

Proširujemo sve niske sa razlikom na najviše jednoj poziciji.

	# Mismatches
\$ ₁ panamabananas ₁	
a ₁ bananas\$panam ₁	
a ₂ mabananas\$pan ₁	
a ₃ namabananas\$p ₁	
a ₄ nanas\$panamab ₁	
a ₅ nas\$panamaban ₂	
a ₆ s\$panamaban ₃	
b ₁ a nanas\$panama ₁	1
m ₁ a bananas\$pana ₂	1
n ₁ a mabananas\$pa ₃	0
n ₂ a nas\$panamaba ₄	0
n ₃ a s\$panamabana ₅	0
p ₁ a namabananas\$ ₁	1
s ₁ \$panamabanana ₆	

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

Jedna niska
proizvodi
drugi *mismatch*
(\$), pa je
odbacujemo.

	# Mismatches
\$ ₁ panamabananas ₁	
a ₁ bananas\$panam ₁	
a ₂ mabananas\$pan ₁	
a ₃ namabananas\$p ₁	
a ₄ nanas\$panamab ₁	
a ₅ nas\$panamaban ₂	
a ₆ s\$panamaban ₃	
b ₁ ananas\$panam a ₁	1
m ₁ abananas\$pan a ₂	1
n ₁ amabananas\$pa a ₃	0
n ₂ anas\$panamab a ₄	0
n ₃ as\$panamaban a ₅	0
p ₁ anamabananas \$ ₁	2
s ₁ \$panamabanana ₆	

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

Na kraju,
 pronašli smo
 pet 3-grama sa
 najviše jednim
mismatch-om.

	# Mismatches
\$ ₁ panamabananas ₁	
a ₁ b ananas\$panam ₁	1
a ₂ m abananas\$pan ₁	1
a ₃ n amabananas\$p ₁	0
a ₄ n anas\$panamab ₁	0
a ₅ n as\$panamaban ₂	0
a ₆ s\$panamaban ₃	
b ₁ ananas\$panam a ₁	
m ₁ abananas\$pan a ₂	
n ₁ amabananas\$p a ₃	
n ₂ anas\$panamab a ₄	
n ₃ as\$panamaban a ₅	
p ₁ anamabananas\$ s ₁	
s ₁ \$panamabanana ₆	

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

Na kraju,
pronašli smo
pet 3-grama sa
najviše jednim
mismatch-om.

	# Mismatches
\$ ₁ panamabananas ₁	
a ₁ b ananas\$panam ₁	1
a ₂ m abananas\$pan ₁	1
a ₃ n amabananas\$p ₁	0
a ₄ n anas\$panamab ₁	0
a ₅ n as\$panamaban ₂	0
a ₆ s\$panamaban ₃	
b ₁ ananas\$panama ₁	
m ₁ abananas\$pana ₂	
n ₁ amabananas\$pa ₃	
n ₂ anas\$panamaba ₄	
n ₃ as\$panamabana ₅	
p ₁ anamabananas\$ ₁	
s ₁ \$panamabanana ₆	

BWT

- Podsećanje: traženje niske **ana** u niski panamabananas

Na kraju,
pronašli smo
pet 3-grama sa
najviše jednim
mismatch-om.

Suffix Array

\$ ₁ panamabananas ₁	
a ₁ b ananas\$panam ₁	5
a ₂ m abananas\$pan ₁	3
a ₃ n amabananas\$p ₁	1
a ₄ n anas\$panamab ₁	7
a ₅ n as\$panamaban ₂	9
a ₆ s\$panamaban ₃	
b ₁ ananas\$panama ₁	
m ₁ abananas\$pana ₂	
n ₁ amabananas\$pa ₃	
n ₂ anas\$panamaba ₄	
n ₃ as\$panamabana ₅	
p ₁ anamabananas\$ ₁	
s ₁ \$panamabanana ₆	

BWT

- Podsećanje: traženje niske **ana** u niski panam**ab**ananas

Na kraju,
pronašli smo
pet 3-grama sa
najviše jednim
mismatch-om.

Suffix Array

\$ ₁ panamabananas ₁	
a ₁ b ananas\$panam ₁	5
a ₂ mabananas\$pan ₁	3
a ₃ namabananas\$p ₁	1
a ₄ nanas\$panamab ₁	7
a ₅ nas\$panamaban ₂	9
a ₆ s\$panamabanan ₃	
b ₁ ananas\$panama ₁	
m ₁ abananas\$pana ₂	
n ₁ amabananas\$pa ₃	
n ₂ anas\$panamaba ₄	
n ₃ as\$panamabana ₅	
p ₁ anamabananas\$ ₁	
s ₁ \$panamabanan ₆	

BWT

- Podsećanje: traženje niske **ana** u niski pan**ma**bananas

Na kraju,
pronašli smo
pet 3-grama sa
najviše jednim
mismatch-om.

Suffix Array

\$ ₁ panamabananas	1	
a ₁ bananas\$panam	1	5
a₂ma bananas\$pan	1	3
a ₃ namabananas\$p	1	1
a ₄ nanas\$panamab	1	7
a ₅ nas\$panamaban	2	9
a ₆ s\$panamaban	3	
b ₁ ananas\$panama	1	
m ₁ abananas\$pana	2	
n ₁ amabananas\$pa	3	
n ₂ anas\$panamaba	4	
n ₃ as\$panamabana	5	
p ₁ anamabananas\$	1	
s ₁ \$panamabanana	6	

BWT

- Podsećanje: traženje niske **ana** u niski **panamabananas**

Na kraju,
pronašli smo
pet 3-grama sa
najviše jednim
mismatch-om.

Suffix Array

\$ ₁ panamabananas	1	
a ₁ bananas\$panam	1	5
a ₂ mabananas\$pan	1	3
a₃namabananas\$ p	1	1
a ₄ nanas\$panamab	1	7
a ₅ nas\$panamaban	2	9
a ₆ s\$panamaban	3	
b ₁ ananas\$panama	1	
m ₁ abananas\$pana	2	
n ₁ amabananas\$pa	3	
n ₂ anas\$panamaba	4	
n ₃ as\$panamabana	5	
p ₁ anamabananas\$	1	
s ₁ \$panamabanana	6	

BWT

- Podsećanje: traženje niske **ana** u niski panamab**ana**nas

Na kraju,
pronašli smo
pet 3-grama sa
najviše jednim
mismatch-om.

Suffix Array

\$ ₁ panamabananas ₁	
a ₁ bananas\$panam ₁	5
a ₂ mabananas\$pan ₁	3
a ₃ namabananas\$p ₁	1
a₄nanas\$panamab₁	7
a ₅ nas\$panamaban ₂	9
a ₆ s\$panamabanan ₃	
b ₁ ananas\$panama ₁	
m ₁ abananas\$pana ₂	
n ₁ amabananas\$pa ₃	
n ₂ anas\$panamaba ₄	
n ₃ as\$panamabana ₅	
p ₁ anamabananas\$ ₁	
s ₁ \$panamabanan ₆	

BWT

- Podsećanje: traženje niske **ana** u niski panamaban**ana**s

Na kraju,
pronašli smo
pet 3-grama sa
najviše jednim
mismatch-om.

Suffix Array

\$ ₁ panamabananas ₁	
a ₁ bananas\$panam ₁	5
a ₂ mabananas\$pan ₁	3
a ₃ namabananas\$p ₁	1
a ₄ nanas\$panamab ₁	7
a₅na s\$panamaban ₂	9
a ₆ s\$panamaban ₃	
b ₁ ananas\$panama ₁	
m ₁ abananas\$pana ₂	
n ₁ amabananas\$pa ₃	
n ₂ anas\$panamaba ₄	
n ₃ as\$panamabana ₅	
p ₁ anamabananas\$ ₁	
s ₁ \$panamabanana ₆	

- Slajdovi pokrivaju poglavlje 9 knjige *Bioinformatics Algorithms: an Active Learning Approach*
- Sadržaj slajdova je preuzet sa zvaničnih prezentacija autora i dodatno prilagođen