

Asembliranje genoma

Vladimir Mijić

27. mart 2024.

attaaaggtt	tataccttcc	caggtaacaa	acaaaccaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gctgtcaactc	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgagggc
tgcttacggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgaccgaaag	gtaagatgga	gagccttgtc	cctgggttca	acgagaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggctttgg	agactcogtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagatg	gcacttgtyg	cttagtagaa	gttgaaaaag	gcgtttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctcgaaactg	caactcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtaacggtc	gtagtgggtga
gacacttggt	gtccttgctcc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	tacggcgccg	atctaaaagtc	atttgactta	ggcgacgagc	ttggcactga	tccttatgaa	gattttcaag	aaaactgga	cactaaacac

Sekvenciranje genoma - osnovni pojmovi

- Genom - genetski materijal nekog organizma
 - genom se nalazi u DNK ili RNK molekulima

attaaggtt	tataccttcc	caggttaacaa	acaaaccaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttta	aatctgtgtg	gctgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggc
tgcttacggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgaccgaaag	gtaagatgga	gagccttgtc	cctgggttca	acgagaaaa	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggctttgg	agactccgtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagatg	gcacttgctg	cttagtagaa	gttgaaaaag	gcgtttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctogaactg	caactcatgg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggctc	gtagtgggtga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttcgttaag	aacggttaata	aaggagctgg	tgcccatagt	tacggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

Sekvenciranje genoma - osnovni pojmovi

- Genom - genetski materijal nekog organizma
 - genom se nalazi u DNK ili RNK molekulima
- Nukleotid - osnovni gradivni molekul DNK i RNK
 - adenin, citozin, guanin, timin (DNK) ili uracil (RNK)

attaaagggt tataccttcc caggtaacaa accaaccacc ttctgatctc ttgtagatct gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggtctcatgc ttagtgcact cagcgagtat aattaataac taattactgt cgttgacagg acacgagtaa ctgtctatcc ttctgaggcc
tgcttaagggt ttgctcogtg ttgcagccga tcctcagcac atctaggttt cgtccgggtg tgacgcgaaag gtaagatgga gagccttgtc cctggtttca acagagaaaa acacgtccaa ctacgtttgc ctgtttttaca ggttcgcgac gtgctcgtac gtgcttttg gtgactcogtg
tatcagaggo acgtcaacat cttaaagatg gcacttgttg cttagtagaa gttgaaaaag tcaacttgaa cagccctatg tgttcaccaa acgttcggat gctogaactg caacctcctg tcactgttat gttagagctg tagcagaact cgaaggcact cagtaagggt gtagtggtga
gacacttggt gtctctgtcc ctcatgtggg cgaataacca gtggttacc gcaaggttct tcttctgaag aacggttaata aaggagctgg tggccatagt tacggcgccg atctaaaagtc atttgactta ggcgcgcgagc ttggcactga tccttatgaa gattttcaag aaaaactgaa cactaaacca

Sekvenciranje genoma - osnovni pojmovi

- Genom - genetski materijal nekog organizma
 - genom se nalazi u DNK ili RNK molekulima
- Nukleotid - osnovni gradivni molekul DNK i RNK
 - adenin, citozin, guanin, timin (DNK) ili uracil (RNK)
- Ljudski genom sadrži oko 3 milijarde nukleotida

attaaagggtt tataccttcc caggtaacaa accaaccac accatgacac atctaggttt cgtccgggtg tgacggaaa gtaagatgga gagccttgtc cctgggttca acgagaaaa acagtcocaa ctoagtttgc ctggttttaca ggttcgcgac gtgctcgtac gtggttttg gtggttttg agactcgtg gaggagggtc tctcgcaggc
tgctaacggt ttgctcgggt ttgcagccga tcatcagcac atctaggttt cgtccgggtg tgacggaaa gtaagatgga gagccttgtc cctgggttca acgagaaaa acagtcocaa ctoagtttgc ctggttttaca ggttcgcgac gtgctcgtac gtggttttg agactcgtg gaggagggtc tctcgcaggc
tatcagaggo acgtcaacat cttaaagatg gcacttggtg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccctatg tgttccatca acgttcggat gctogaactg cactcctatg tcatgttatg gttgagctg tagcagaact cgaaggcatt cagtaagggtc gtagtggtga
gacacttggt gtcccttgcc ctcctgtggg cgaataacca gtggttacc gcaaggttct tcttcgttaag aacggttaata aaggagctgg tggccatagt tacggcgccg atctaaagt c atttgaacta ggccagcaga ttggcactga tcttatgaa gattttcaag aaaaactgaa cactaaaca

Sekvenciranje genoma - osnovni pojmovi

- Genom - genetski materijal nekog organizma
 - genom se nalazi u DNK ili RNK molekulima
- Nukleotid - osnovni gradivni molekul DNK i RNK
 - adenin, citozin, guanin, timin (DNK) ili uracil (RNK)
- Ljudski genom sadrži oko 3 milijarde nukleotida
- Sekvenciranje genoma predstavlja proces otkrivanja sastava genoma

attaaagggtt tataccttcc caggttaacaa accaaccacac ttctgatctc ttgtagatct gttctctaaa cgaactttta aatctgtgtg gctgtcaact ggctgcctgc ttagtgcact cagcagttat aattaataac taattactgt cgttgacagg acacagatga ctgtctctatc ttctgaaggc
tgcttaacgt ttgttcogtg ttgcagccga tcatcagcac atctaggttt cgtccgggtg tgacggaaa gacgcttgct cctggtttca acgagaaaaa ctaagtttgc ctgtttttaca ggttcogac gtgctcgtac gtgcttttg agactcogtg gaggaggtct
tatcagaggo acgtcaacat cttaaagatg gcacttgtyg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccctatg tgttcaatca acgttcoggt gctogaactg atctaaagtc atttgactta ggccagcago ttggcaacta tctttatga gatittcaag aaaaactgaa cactaaaca
gacacttggt gtctttgtcc ctcatgtggg cgaataacca gtggcttacc gcaaggttct tctttgtaag aacggttaata aaggagctgg tggccatagt taaggcgcgc atctaaagtc atttgactta ggccagcago ttggcaacta tctttatga gatittcaag aaaaactgaa cactaaaca

Sekvenciranje genoma - istorijski pregled

- 1977. - Prve metode za sekvenciranja DNK
 - **Sangerova metoda** - Frederick Sanger
 - **Maxam - Gilbert metoda** - Allan Maxam i Walter Gilbert

attaaggtt	tataccttcc	caggttaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cggtgacagg	acaagagtaa	ctcgtctatc	ttctgcaggg
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcatacgcac	atctaggttt	cgtccgggtg	tgacccgaaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggcttttg	agactccgtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaaaagatg	gcacttgtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gttcgaactg	cacctcatgg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggtc	gtagtggtag
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttcgttaag	aacggtaata	aaggagctgg	tgcccatagt	tacggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

Sekvenciranje genoma - istorijski pregled

- 1977. - Prve metode za sekvenciranje DNK
 - **Sangerova metoda** - Frederick Sanger
 - **Maxam - Gilbert metoda** - Allan Maxam i Walter Gilbert
- Problem?

attaaggtt	tataccttcc	caggttaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacccgaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaa	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcogac	gtgctcgtac	gtggctttgg	agactccgtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagatg	gcacttggtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gttcgaactg	caactcattg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggctc	gtagtgggtga
gacacttggt	gtccttgctc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtaag	aacggtaata	aaggagctgg	tgcccatagt	tacggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

Sekvenciranje genoma - istorijski pregled

- 1977. - Prve metode za sekvenciranja DNK
 - **Sangerova metoda** - Frederick Sanger
 - **Maxam - Gilbert metoda** - Allan Maxam i Walter Gilbert
- Problem?
 - Predložene metode su bile izuzetno skupe - \$3 milijarde dolara za sekvenciranje ljudskog genoma

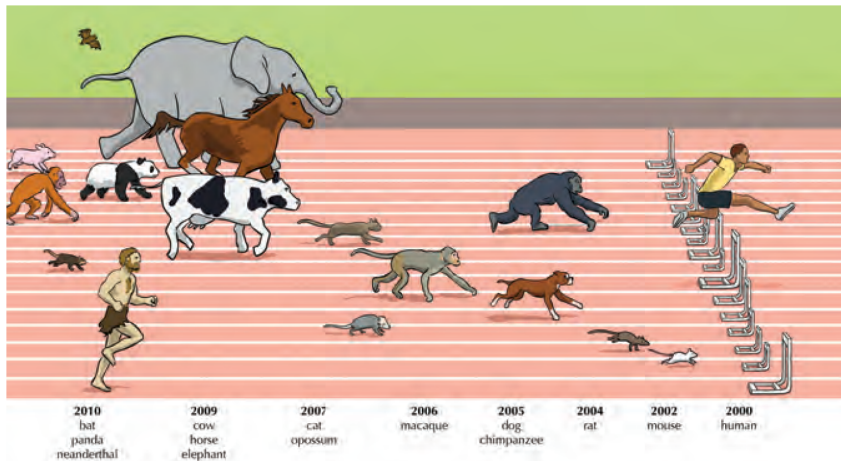
attaaggtt	tataccttcc	caggtaacaa	accaaccaac	tttogatctc	ttgtagatct	gttctotaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cggtgacagg	acaagagtaa	ctcgtctatc	ttctgaggcc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcatacgcac	atctaggttt	cgtccgggtg	tgacccgaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaa	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggcttttg	agactccgtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagatg	gcacttgtyg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctcgaactg	cacctcatgg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcoatt	cagtacggtc	gtagtgggtg
gacacttggt	gtccttgctc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttgtaag	aacggtaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

Sekvenciranje genoma - istorijski pregled

- 1977. - Prve metode za sekvenciranja DNK
 - **Sangerova metoda** - Frederick Sanger
 - **Maxam - Gilbert metoda** - Allan Maxam i Walter Gilbert
- Problem?
 - Predložene metode su bile izuzetno skupe - \$3 milijarde dolara za sekvenciranje ljudskog genoma
- 1990. - Početak projekta **Human Genome Project**
 - Cilj: sekvenciranje ljudskog genoma
 - Prvi rezultati objavljeni 2001. godine
 - Završen 2003. godine

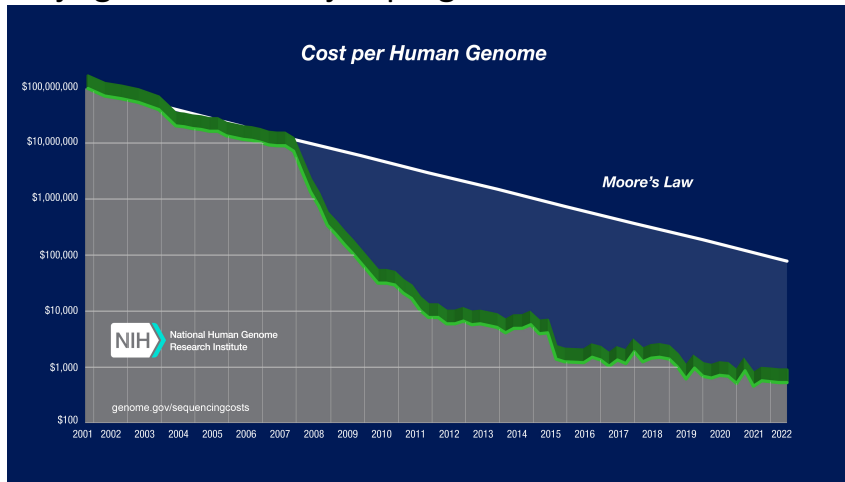
attaaagggtt tataccttcc caggttaacaa accaaccacac ttctgatctc ttgtagatct gttctctaaa cgaactttaa aatctgtgtg gctgtcaact ggctgcatgc ttagtgcact cagcgagtat aattaataac taattactgt cgttgacagg acacagagtaa ctgtctatcc ttctgaggcc
tgcttaacgtt ttgttcogtg ttgcagcaga tcatcagcac atctaggttt cgtccgggtg tgacagaaag gtaagatgga gagccttgtc cctggtttca acagagaaaa acacgtccaa ctacgtttgc ctgtttttca ggtttcgac gtgtcgttac gtggttttg agactcogtg gaggaggtct
tatcagaggo acgtcaacat cttaaagatg gcactttgtg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccctatg tgttcaatca acgttoggat gctogaactg caactccttg tcatgttatg gttgagctgg tagcagaact cgaaggcaat cagtaagggtc gtagtggtga
gacacttggt gtctttgtcc ctcatgtggg cgaataacca gtggtttacc gcaaggttct tctttgtaag aacggttaata aaggagctgg tggccatagt taaggcgccg atctaaagtc atttgactta ggccagcaga ttggcaactg tctttatgaa gattttcaag aaaaactgaa cactaaacac

Sekvenciranje genoma - istorijski pregled



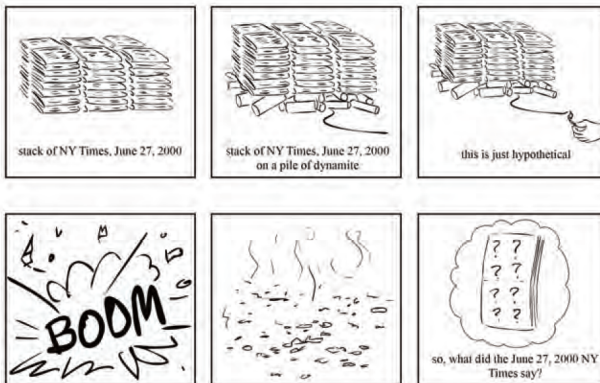
attaaaggtt tataccttcc caggtaacaa accaaccac tttgatcttc ttgtagatct gttctotaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcactc ttagtgcact cagcgagtat aattaataac taattactgt cgttgacagg acacagagtaa ctgctctatc ttctgaagga
 tgcttaacgtt ttgttcogtg ttgcagcoga tcatcagcac atctaggttt cgtccgggtg tgacogaaag gtaagatgga gagccttgto cctgttttca acagagaaac ctagtgcata cgtgttttca ggtgttcac
 tatcagagga acgtcaacat cttaaagatg gcacttgtyg cttagtagaa ggtttttgoc tcaacttgaa cagccttatg gctgttcacg atctaaagtc atttgactta ggcagcagag ttggcaactg taagcagaact gattttcaag aaaactgcaa cactaaace
 gacacttggt gtctttgtcc ctcattgtgg cgaaatacca gtggtttacc gaaaggttct tctttgttaag aacggtaata aaggagctgg tggcactagt taaggcgcgc atctaaagtc atttgactta ggcagcagag ttggcaactg taagcagaact gattttcaag aaaactgcaa cactaaace

Sekvenciranje genoma - istorijski pregled



attaaagggtt tataccttcc caggttaacaa accaaccacac tttagatctc ttgtagatct gtctctataa cgaactttaa aatctgtgtg gctgtcacctc ggctgcacgc ttagtgcaact caccgcagtat aattaataac taattactgt cgttgacagg acacagagtaa ctgctctatc ttctgacggc
tgcttaacgtt ttgttcogtg ttgcagccga tcatcagcac atctaggttt cgtccgggtg tgacccgaa gtaagatgga gagccttgtc cctgggtttca acagagaaac acacgtocaa ctacagtttg cctgttttaca ggttcgcgac gtgctcgtac gtggttttg agactcogtg gaggaggtct
tatcagaggo acgtcaacat cttaaagatg gcacttgtyg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccctatg tgttcatcaa gcttcggat gctcgaactg atctaaagtc atttgactta ggcgcagcagc gttgagctgg tagcagaact cgaaggcaatt cagtaacggtc gtagtggtga
gacacttggt gtccctgtcc ctcatgtggg cgaataacca gtggtttacc gcaaggttct totttogtaag aacggtaata aaggagctgg tggccatagt tacggcgccg atctaaagtc atttgactta ggcgcagcagc gttgacactga tcttatgaa gattttcaag aaaactgcaa cactaaacac

Analogija - problem eksplozije novina



Slika: Problem eksplozije novina će poslužiti u razumijevanju problema asembliranja genoma

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttagatctc	ttgtagatct	gtttctotaaa	cgaacttttaa	aattctgtgtg	gctgtcaactc	ggctgcactgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acacagagtaa	ctcgtctatc	ttctgagggc
tgcttaacgt	ttgtccogtg	ttgcagccga	tcatacagcac	atctaggttt	cgtccgggtg	tgacccgaag	gtaagatgga	gagccttgct	cctggtttcca	acgagaaaaac	acacgtccaa	ctcagttttgc	ctgtttttaca	ggtttcgcagc	gtgctcgttac	gtggcttttg	agactcogtg	gaggaggtct
tatcagaggo	acgtcaacat	cttaaaagatg	gcacttgctg	cttagtagaa	gttgaaaaag	gggtttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	ctctagaactg	cacctcctatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcoatt	cagtaacgttc	gtagtggtga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggtccatagt	taaggcgccg	atctaaaagtc	atttgactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactgga	cactaaacac

Analogija - problem eksplozije novina



Slika: Spajanje dijelova koji se preklapaju

attaaagggtt tataccttcc caggtaacaa accaaccac ttctgatctc ttgtagatct gttctotaaa cgaactttaa aatctgtgtg gctgtcactc ggtctcatgc ttagtgcact cagcgagtat aattaataac taattactgt cgttgacagg acacagataa ctctctatct ttctgaaggc
tgcttacggt ttgtccgtg ttgcagccga tcatcagcac atctagggtt cgtccgggtg tgacccgaag gtaagatgga gagccttgtc cctggtttca acagagaaac acacgtccaa ctacgtttgc gaggaggtct
tatcagaggo acgtcaacat cttaagatg gcacttgtyg cttagtagaa gttgaaaaag ggtttttgoc tcaacttgaa cagccctatg tgttcatcaa acgttcggat gctogaactg cactctatgg tcatgttatg gttgagctgg tagcagaact cgaaggcaatt cagtacggtc gtagtgggta
gacacttggt gtccctgtcc ctcatgtggg cgaataacca gtggtttacc gcaagggtct totttgttaag aacggtaata aaggagctgg tggccatagt taaggcgocg atctaaagtc atttgactta ggccagcago ttggcaactga tctttatgaa gattttcaag aaaactgcaa cactaaacac

Analogija - problem eksplozije novina

- Kopije novina - odgovaraju uzorcima DNK

attaaggtt	tataccttcc	caggttaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cggtgacagg	acaagagtaa	ctcgtctatc	ttctgaggcc
tgcttacggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgacccgaa	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaa	acaagtcaca	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggctttgg	agactccgtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagatg	gcacttgtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tggtcatcaa	acgttcggat	gctcgaaact	cacctcatgg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtacggct	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttcgttaag	aacggtaata	aaggagctgg	tgcccatagt	tacggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcaactg	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

Analogija - problem eksplozije novina

- Kopije novina - odgovaraju uzorcima DNK
- Dijelovi novina - odgovaraju očitanjima sekvenciranja

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gotgtcactc	ggctgcctgc	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacgagtaa	ctcgtctatc	ttctgcaggc
tgcttaagggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacgcgaaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggttttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaagatg	gcacttgtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctogaactg	caacctcatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtc
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaaagtc	atttgaacta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacca

Analogija - problem eksplozije novina

- Kopije novina - odgovaraju uzorcima DNK
- Dijelovi novina - odgovaraju očitanjima sekvenciranja
 - Sekvencer *sječe* genom na manje dijelove koji se nazivaju **očitanja** (eng. *reads*)
 - Neki dijelovi će biti izgubljeni

attaaggtt	tataccttcc	caggttaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gctgtcaact	ggctgcactg	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acacagataa	ctcgtctatc	ttctgcaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcatacgcac	attcaggttt	cgtccgggtg	tgacggaaa	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaa	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtgcttttgg	agactccgtg	gaggaggctc
tatcagaggo	acgtcaacat	cttaaaagatg	gcacttgctg	cttagtagaa	gttgaaaaag	gogttttgoc	tcacacttgaa	cagccctatg	tgttcaatcaa	acgttcggat	gctogaactg	caactccatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtc
gacacttggt	gtccttggcc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttcgttaag	aacggtaata	aaggagctgg	tggtccatagt	taaggcgccg	atctaaaagtc	atttgaacta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactggaa	cactaaacac

Analogija - problem eksplozije novina

- Kopije novina - odgovaraju uzorcima DNK
- Dijelovi novina - odgovaraju očitanjima sekvenciranja
 - Sekvencer *sječe* genom na manje dijelove koji se nazivaju **očitanja** (eng. *reads*)
 - Neki dijelovi će biti izgubljeni
- Cilj - spojiti dijelove u jednu cjelinu

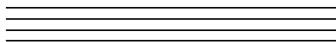


Slika: Nije isto kao slaganje slagalice!

attaaagggtt tataccttcc caggttaacaa accaaccacac ttctgatctc ttgtagatct gttctctaaa cgaactttaa aatctgtgtg gctgtcaact ggctgcatgc ttagtgcaact cagcgagtat aattataaac taattactgt cgttgacagg acacagatga ctgctctatc ttctgaggc
tgcttaagggt ttgttcogtg ttgcagccga tcatcagcac atctaggttt cgtccgggtg tgacggaaa gtaagatgga gagccttgtc cctggtttca acagagaaaa acacgtccaa ctacagtttg cttgctgtac gtgctgttac gtgcttttg agactcogtg gaggagggtc
tatcagaggo acgtcaacat cttaaagatg gcacttggtg cttagtagaa gttgaaaaa tcaacttgaa cagccctatg tgttcatcaa acgttoggat gctogaactg atctaaagtc atttgactta ggcgcagago ttggcactga tctttatgaa gattttcaag aaaaactgaa cactaaacac
gacacttggt gtccctgtcc ctcatgtggg cgaatataca gtggcttacc gcaaggttct tcttctgaag aacggttaata aaggagctgg tggccatagt taaggcgccg atctaaagtc atttgactta ggcgcagago ttggcactga tctttatgaa gattttcaag aaaaactgaa cactaaacac

Ilustracija problema

Multiple identical
copies of a genome



Shatter the genome
into reads



Sequence the reads

AGAATATCA TGAGAATAT GAGAATATC

Assemble the
genome using
overlapping reads

AGAATATCA
GAGAATATC
TGAGAATAT
...TGAGAATATCA...

attaaagggt tataccttcc caggtaacaa accaaccac tttagatctc tttagatctc gttctctaaa cgaactttaa aatctgtgtg gctgtcaact gctgtcaact gctgtcaact tttagtgcact cagcgagtat aattaataac taattactgt cgttgacagg acacagatga ctctgtctatc ttctgacagg
tgcttaacgt ttgttcogtg ttgcagcoga tcacacagac atctaggttt cgtccgggtg tgacogaaag gtaagatgga gagccttgct cctggtttca acgagaaaac acacgtocaa ctacagtttg ctgtttttaca ggttcogac gtgctogtac gtggttttg agactcogtg gaggaggtct
tatcagagga acgtcaacat cttaaagatg gcacttggtg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccttatg tgttcaatca acgttcoggt gctogaactg caacttcag tcactgttatg gttgagctgg tagcagaact cgaaggcaat cagtacggct gtagtggtga
gacacttggt gtctttgtcc ctcatgtggg cgaataacca gtggottacc gcaaggttct totttogtaag aacggtaata aaggagctgg tggccatagt taaggcgocg atctaaagtc atttgactta ggcagcagag ttggcaactga tccttatgaa gattttcaag aaaactgaa caactaaaca

Problem rekonstrukcije niske

Asembliranje genoma je težak problem.

Prije uvođenja formulisanja računarskog problema, navešćemo neke ključne karakteristike problema:

- DNK je dvostruki lanac nukleotida - nije poznato iz kog lanca proizilazi koje očitavanje

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttta	aatctgtgtg	gctgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacgagtaa	ctcgtctatc	ttctgacggc
tgcttaacgt	tttgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacgcgaaag	gtaagatgga	gagccttgtc	cctggtttca	acgcagaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggttttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaagatg	gcacttgtgg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctogaactg	caacctcatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtgc	atttgaacta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactggaa	cactaaacac

Problem rekonstrukcije niske

Asembliranje genoma je težak problem.

Prije uvođenja formulisanja računarskog problema, navešćemo neke ključne karakteristike problema:

- DNK je dvostruki lanac nukleotida - nije poznato iz kog lanca proizilazi koje očitavanje
- Očitavanja mogu sadržati greške

attaaagggt	tataccttcc	caggtaacaa	accaaccaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gctgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacagataa	ctcgtctatc	ttctgacggc
tgcttaacgt	tttgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacgcgaaag	gtaagatgga	gagccttgtc	cctggtttca	acgcagaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtgcttttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaaaagatg	gcacttgttg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctcgaactg	caactccatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaaagtc	atttgaacta	ggcgcagcgc	ttggcaactga	tccttatgaa	gattttcaag	aaaactggaa	cactaaacac

Problem rekonstrukcije niske

Asembliranje genoma je težak problem.

Prije uvođenja formulisanja računarskog problema, navešćemo neke ključne karakteristike problema:

- DNK je dvostruki lanac nukleotida - nije poznato iz kog lanca proizilazi koje očitavanje
- Očitavanja mogu sadržati greške
- Očitavanja ne pokrivaju cijeli genom

attaaagggtt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttta	aattctgtgtg	gctgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aatttaataac	taattactgt	cgttgacagg	acacagatga	ctgtctctatc	ttctgacaggc
tgcttaacgtt	tttgtccgtg	ttgcagccga	tcactcagcac	atctaggttt	cgtccgggtg	tgacgcgaaag	gtaagatgga	gagccttgct	cctgtgttca	acgcagaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgtcgtgtac	gtggttttgg	agactccgtg	gaggagggtct
tatcagaggo	acgtcaacat	cttaaaagatg	gcacttgtgg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctogaactg	caacctcatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtgc	atttgaacta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactggaa	cactaaacaca

Problem rekonstrukcije niske

Asembliranje genoma je težak problem.

Prije uvođenja formulisanja računarskog problema, navešćemo neke ključne karakteristike problema:

- DNK je dvostruki lanac nukleotida - nije poznato iz kog lanca proizilazi koje očitavanje
- Očitavanja mogu sadržati greške
- Očitavanja ne pokrivaju cijeli genom

Očitavanja su iste dužine, možemo pretpostaviti da su očitavanja svi k-grami (eng. k-mer, podniska dužine k) za neko k.

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttta	aattctgtgtg	gctgtcaact	ggctgcattc	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacagataa	ctgtctatcc	ttctgcaggc
tgcttaacgt	tttgtccgtg	ttgcagccga	tcactcagcc	atctaggttt	cgtccgggtg	tgacggaaa	gtaagatgga	gagccttgct	cctggtttca	acgagaaaa	acaagtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtgcttttgg	agactccgtg	gaggaggtct
tatcagaggo	acgtcaacat	cttaagatg	gcacttggg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaccaa	acgttoggat	gctogaactg	caactccatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggtttacc	gcaaggttct	tctttgttaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtcc	atttgactta	ggcgacgagc	ttggcaactga	tccttatgaa	gatttttaag	aaaactggaa	cactaaacaa

Problem rekonstrukcije niske

Računarska formulacija problema asembliranja genoma:

Za dati string *Text*, njegova k-gram kompozicija $COMPOSITION_3(Text)$ je kolekcija svih k-gram podniski od *Text*. Primjer:

$$COMPOSITION_3(TATGGGGTGC) = \{ATG, GGG, GGG, GGT, GTG, TAT, TGC, TGG\}$$

k-grami su prikazani u leksikografskom poretku, zato što tačan redoslijed očitavanja nije poznat kada se generišu.

attaaggtt	tataccttcc	caggttaacaa	accaaccaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttta	aatctgtgtg	gctgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcacagcac	atctaggttt	cgtccgggtg	tgacogaaag	gtaagatgga	gagccttgct	cctgggttca	acgagaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggctttgg	agactccgtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaaaagatg	gcacttgctg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctogaactg	cacctcatgg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtaaggctc	gtagtgggtga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgaag	aacggtaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgacgago	ttggcactga	tccttatgaa	gattttcaag	aaaactgga	cactaaacac

Problem rekonstrukcije niske

Rekonstruisati nisku iz njenih k-grama.

Ulaz: Vrijednost k i kolekcija k-grama *Patterns*.

Izlaz: Niska *Text* koja sadrži sve k-grame iz *Patterns*, ako takva niska postoji.

attaaaggtt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gotgtcactc	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggcc
tgcttaacgtt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgacccgaag	gtaagatgga	gagccttgtc	cctgggtttca	acgagaaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggctttgg	agactccgtg	gaggaggtct
tatcagaggcc	acgtcaacat	cttaagatg	gcacttggtg	cttagtagaa	gttgaaaaag	gcgtttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctogaactg	cacctcatgg	tcctgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggtc	gtagtgggtga
gacacttggt	gtccttgctc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtaag	aacggtaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgacgago	ttggcactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

Problem rekonstrukcije niske - primjer 1.

Neka je zadata 3-gram kompozicija

AAT ATG GTT TAA TGT

attaaaggtt	tataccttcc	caggttaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcactg	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacgagtaa	ctcgtctatc	ttctgcaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacggaaa	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgcctgtac	gtggttttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaaaagatg	gcacttggtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctogaactg	cacctcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacgtc	gtagtgggtg
gacacttggt	gtccttgctc	ctcactgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtta	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagt	atttgaacta	ggcgacgagc	ttggcaactg	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacaa

Problem rekonstrukcije niske - primjer 1.

Neka je zadata 3-gram kompozicija

AAT ATG GTT TAA TGT

Zadatak: povežimo parove k-grama ako se preklapaju u k-1 simbola.

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gotgtcactc	ggctgcctgc	ttagtgcact	cacgcagtat	aattataaac	taattactgt	cgttgacagg	acacagataa	ctcgtctatc	ttctgaaggc
tgctaacggt	tttgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgacggaaaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggttttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaaaagatg	gcacttgttg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctogaactg	caacctcatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaagggtc	gtagtgggtc
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggtccatagt	taaggcgccg	atctaaaagtc	atttgaacta	ggcgcagcgc	ttggcaactga	tccttatgaa	gatttttcaag	aaaaactgaa	cactaaacac

Problem rekonstrukcije niske - primjer 1.

Neka je zadata 3-gram kompozicija

AAT ATG GTT TAA TGT

Zadatak: povežimo parove k-grama ako se preklapaju u k-1 simbola.

TAA

AAT

ATG

TGT

GTT

TAAATGTT

Slika: Rekonstruisana niska je *TAATGTT*

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatoto	ttgtagatct	gtttototaaa	cgaacttttaa	aattotgtgtg	gotgtcaactc	ggctgcactgc	ttagtgcact	cacgcagtat	aattataaac	taattactgt	cgttgacagg	acacagataa	ctgtctatct	ttctgaggc
tgcttaacgt	tttgtccgtg	ttgcagccga	tcactcagcac	atctaggttt	cgtccgggtg	tgacogaaaag	gtaagatgga	gagccttgct	cctggtttcca	acgagaaaaac	acacgtccaa	ctcagtttgc	ctgtttttaca	ggtttccgac	gtgctcgtac	gtgcttttg	agactccgtg	gaggaggct
tatcagaggo	acgtcaacat	cttaaaagatg	gcactttgtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatcaa	acgttoggat	gotogaactg	cacctccatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tctttgttaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgcgcg	atctaaaagtc	atttgaacta	ggcgcagago	ttggcaactga	tccttatgaa	gattttcaag	aaaactggaa	cactaaacac

Problem rekonstrukcije niske - primjer 2.

Neka je zadata 3-gram kompozicija

AAT ATG ATG ATG CAT CCA GAT GCC GGA GGG GTT TAA TGC TGG TGT

attaaaggtt	tataccttcc	caggtaacaa	accaaccaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gctgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggcc
tgcttaacgtt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgaccgaaag	gtaagatgga	gagccttgtc	cctgggttca	acgagaaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggctttgg	agactccgtg	gaggaggtct
tatcagaggc	acgtcaacat	cttaagatg	gcacttgtyg	cttagtagaa	gttgaaaaag	gcgttttgcc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctcgaaactg	cacctcatgg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggtc	gtagtgggtga
gacacttggt	gtccttgctc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttcgttaag	aacggttaata	aaggagctgg	tggccatagt	tacgcgcgcg	atctaaagtc	atttgactta	ggcgacgagc	ttggcactga	tccttatgaa	gattttcaag	aaaactgga	cactaaacac

Problem rekonstrukcije niske - primjer 2.

AAT ATG ATG ATG CAT CCA GAT GCC GGA GGG GTT TAA TGC TGG TGT

TAA
AAT
ATG
TAATG

Slika: ATG može biti povezan sa TGC, ili TGG, ili TGT

attaaggtt	tataccttcc	caggtacaaa	acaaacaaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttaa	aatctgtgtg	gctgtcaactc	ggctgcattgc	ttagtgcact	cacgcagtat	aattataaac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggc
tgctaacggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgacogaaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggcttttg	agactccgtg	gaggaggtct
tatcagaggo	acgtcaacat	cttaagatg	gcacttgtyg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatcaa	acgttcggat	gctogaactg	cacctcattg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtaacggtc	gtagtgggtga
gacacttggt	gtccttgtoo	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	tacggcgocg	atctaaaagtc	atttgactta	ggcgaecgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

Problem rekonstrukcije niske - primjer 2.

AAT ATG ATG ATG CAT CCA GAT GCC GGA GGG GTT TAA TGC TGG TGT

TAA
AAT
ATG
TGT
TAATGT

Slika: Kako smo izabrali TGT, jedini sljedeći izbor je GTT

attaaggtt	tataccttcc	caggtacaaa	acaaacaaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttta	aatctgtgtg	gotgtcaact	ggctgcattg	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggc
tgcttacggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgacogaaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaa	acaagtcoc	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggcttttg	agactccgtg	gaggaggtct
tatcagaggo	acgtcaacat	cttaagatg	gcacttggtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccttatg	tggtcatcaa	acgttcggat	gctogaactg	cacctcatgg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaaggctc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggtaata	aaggagctgg	tgcccatagt	taaggcgocg	atctaaagtc	atttgactta	ggcgcagcgc	ttggcaactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

Problem rekonstrukcije niske - primjer 2.

AAT ATG ATG ATG CAT CCA GAT GCC GGA GGG GTT TAA TGC TGG TGT

TAA
AAT
ATG
TGT
GTT
TAATGTT

Slika: Ne možemo nastaviti dalje jer nema k-grama koji počinju sa TT

attaaggtt	tataccttcc	caggttaacaa	acaaacaaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttta	aatctgtgtg	gotgtcaact	ggctgcattg	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cggtgacagg	acaagagtaa	ctgtctatcc	ttctgaaggg
tgctaacgtt	ttgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacogaaag	gtaagatgga	gagccttgct	cctgttttca	acgagaaaa	acaagtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggcttttg	agactccgtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagatg	gcacttggtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tggtcatcca	acgttcggat	gtctgaactg	cacctcatgg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtacggct	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tctttgttaag	aacggtaata	aaggagctgg	tggtccatagt	tacggcgocg	atctaaagtc	atttgactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactgga	cactaaacac

Problem rekonstrukcije niske - primjer 2.

AAT ATG ATG ATG CAT CCA GAT GCC GGA GGG GTT TAA TGC TGG TGT

TAA
AAT
ATG
TGC
TAA TGC

Slika: Vraćamo se na k-gram ATG i nastavljamo dalje sa TGC

attaaggtt	tataccttcc	caggtacaaa	acaaacaaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttta	aattctgtgtg	gotgtcaactc	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggc
tgctaacggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgacccgaaag	gtaagatgga	gagccttgctc	cctggtttcca	acgagaaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggcttttg	agactccgtg	gaggaggtct
tatcagaggo	acgtcaacat	cttaagatg	gcacttgctg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatcaa	acgttcgggat	gttcgaactg	cacctcctatg	tcctgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggctc	gtagtgggtg
gacacttggt	gtccttgctc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggtaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactgga	cactaaacac

Problem rekonstrukcije niske - primjer 2.

*AAT ATG ATG ATG CAT CCA GAT GCC GGA **GGG** GTT TAA TGC TGG TGT*

TAA
AAT
ATG
TGC
GCC
CCA
CAT
ATG
TGG
GGA
GAT
ATG
TGT
GTT
TAATGCCATGGATGTT

Slika: Nastavljajući dalje, dobili smo *TAATGCCATGGATGTT* ali nismo iskoristili k-gram GGG

attaagggtt	tataccttcc	caggtaacaa	acaaacaaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattataaac	taattactgt	cgttgacagg	acaagagtaa	ctgtctatcc	ttctgaaggg
tgctaacggt	ttgtccggtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacogaaag	gtaagatgga	gagccttgct	cctggtttca	acagaaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	gggttcggac	gtgctcgtac	gtggcttttg	agactcogtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagatg	gcacttggtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tggtcatcaa	acgttcggat	gtctogaactg	cacctcatgg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggctc	gtagtgggtga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tctttogtaag	aacggtaata	aaggagctgg	tggtccatagt	atctaaagtc	atttgactta	ggcgaagcago	ttggcaactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac	

Problem rekonstrukcije niske - primjer 2.

AAT ATG ATG ATG CAT CCA GAT GCC GGA GGG GTT TAA TGC TGG TGT

TAA
AAT
ATG
TGC
GCC
CCA
CAT
ATG
TGG
GGG
GGA
GAT
ATG
TGT
GTT
TAATGCCATGGGATGTT

Slika: Ovo je jedno od mogućih rješenja, korištene su različite boje za interval između pojavljivanja ATG

attaaagggt tataccttcc caggtaacaa accaaccacac tttagatctc ttgtagatct gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggtctcatgc ttagtgcaact cagcgagtat aattaataac taattactgt cgttgacagg acacagagtaa ctctctctac ttctgcaggc
tgcttaacgt ttgtccgttg ttgcagccga tcatcagcac atctagggtt cgtccgggtg tgaccgaaag gtaagatgga gagccttgtc cctggtttca acagagaaac acacgtccaa ctacgtttgc ctgtttttaca ggttcgcgac gtgtcgttac gtggcttttg agactcogtg gaggaggtct
tatcagaggo acgtcaacat cttaaagatg gcacttgtyg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccctatg tggctcatca acgttcggat gctcgaaactg cactctcatg tcatgttatg gttgagctgg tagcagaact cgaaggcaatt cagtacggtc gtagtgggtg
gacacttggt gtccttgtoo ctcatgtggg cgaataacca gtggottacc gcaaggttct totttogtaag aacggtaata aaggagctgg tggccatagt taaggcgocg atctaaagtc atttgactta ggcgacgago ttggcaactga tccttatgaa gattttcaag aaaactgcaa cactaaacac

Problem rekonstrukcije niske - grafovska reprezentacija

- Čvorovi grafa su k-grami

attaaaggtt	tataccttcc	caggtaacaa	acaaaccaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttta	aatctgtgtg	gctgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgcaggg
tgcttacggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgacccgaag	gtaagatgga	gagccttgtc	cctgggttca	acgagaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggctttgg	agactccgtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaaaagatg	gcacttgctg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctcgaaactg	cacctcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggctc	gtagtgggtga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttcgttaag	aacggtaata	aaggagctgg	tggccatagt	tacggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

Problem rekonstrukcije niske - grafovska reprezentacija

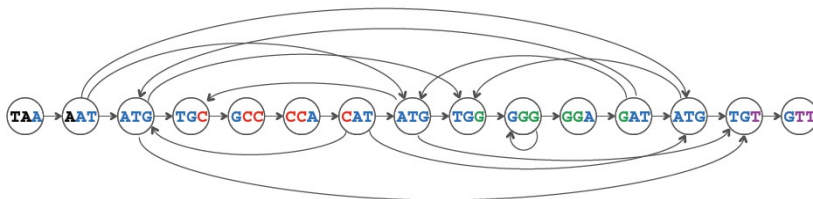
- Čvorovi grafa su k-grami
- Grane su usmjerene, od čvora a do čvora b postoji grana ako:
 - $\text{suffix}(a) == \text{prefix}(b)$



Slika: Genomska putanja TAATGCCATGGGATGTT

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gtttctataa	cgaactttaa	aattctgtgtg	gctgtcaactc	ggctgcactgc	ttagtgcact	cacgcagtat	aattataaac	taattactgt	cgttgacagg	acacagataa	ctcgtctatc	ttctgaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcatacagac	atctaggttt	cgtccgggtg	tgacggaaaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtcgtcgtac	gtgcctttgg	agactccgtg	gaggaggctc
tatcagaggo	acgtcaacat	cttaaagatg	gcacttgttg	cttagtagaa	gttgaaaaaag	gogttttgoc	tcacacttgaa	cagccctatg	tgttcaatcaa	acgttoggat	gctogaactg	caactcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggta
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tctttgtaag	aacggtaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtc	atttgaacta	ggcgcagcgc	ttggcaactga	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacac

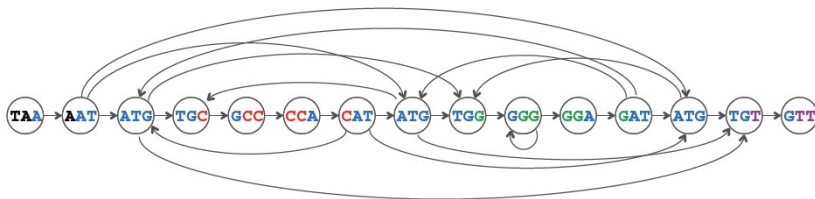
Problem rekonstrukcije niske - grafovska reprezentacija



Slika: Graf na osnovu k-gramskog sastava

attaaggtt	tataccttcc	caggtacaaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttta	aatctgtgtg	gotgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattataaac	taattactgt	cgttgacagg	acaagagtaa	ctctctatcc	ttctgcaggg
tgctaacggt	ttgtccggtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacccgaa	gtaagatgga	gagccttctc	cctgttttca	acgagaaaa	acacgtccaa	ctcagtttgc	ctgtttttca	ggtttccgac	gtgctcgtac	gtggttttgg	agactccgtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagagtg	gcacttctgg	cttagtagaa	gttgaaaaag	gcgtttttgc	tcaacttgaa	cagccttatg	tgttcatcaa	acgttcggat	gctcgaaactg	caactccatg	tcctgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggct	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggctttac	gcaaggttct	tctttgttaag	aacggttaata	aaggagctgg	tggtccatagt	taaggcgcgc	atctaaagtc	atttgactta	ggcgacgagc	ttggcaactg	tccttatgaa	gattttcaag	aaaactgga	cactaaacac

Problem rekonstrukcije niske - grafovska reprezentacija

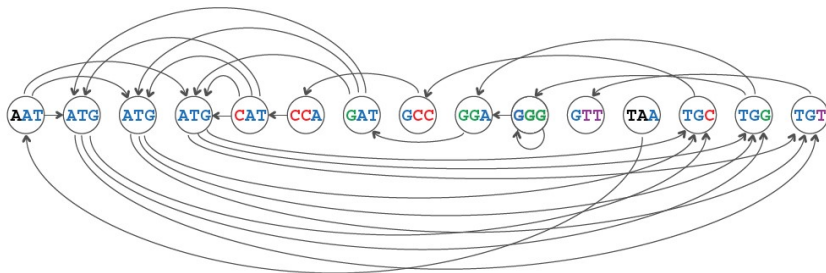


Slika: Graf na osnovu k-gramskog sastava

Možemo li pronaći genomsku putanju u ovom grafu?

attaaggtt	tataccttcc	caggtaacaa	acaaacaaac	tttagatctc	tttagatctc	gttctctaaa	cgaactttta	aattctgtgtg	gctgtcaactc	ggctgcactgc	tttagtgcact	cacgcagttat	aatttaataac	taattactgt	cgttgacagg	acacagagtaa	ctcgtctatc	ttctgaggc
tgcttaacgtt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacccgaag	gtaagatgga	gagccttctc	cctgttttca	acgagaaaaac	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcagc	gtgctcgtac	gtggttttgc	agactccgtg	gaggaggtct
tatcagagggc	acgtcaacat	cttaagatg	gcacttctg	cttagtagaa	gttgaaaaag	gggttttgc	tcacactgaa	cagccctatg	tgcttcaatca	acgttcggat	gctcgaactg	cacctccatg	tcctgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggctc	gtagtgggtga
gacacttggt	gtccttctcc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtaag	aacggtaata	aaggagctgg	tgcccatagt	taaggcgcgcg	atctaaagtc	atttgactta	ggcgcagcagc	ttggcactga	tccttatgaa	gattttcaag	aaaactgaaa	cactaaacac

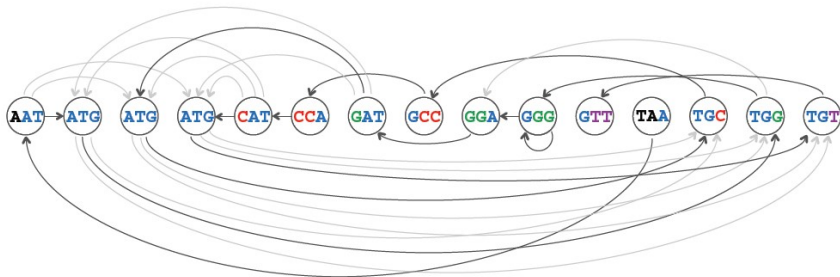
Problem rekonstrukcije niske - grafovska reprezentacija



Slika: Graf na osnovu k-gramskog sastava, sortiran po leksikografskom poretku

attaaggtt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gtttctotaaa	cgaacttttaa	aattctgtgtg	gotgtcaactc	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	cttgtctatc	ttctgaaggg
tgcttaacgtt	ttgtccogtg	ttgcagccga	tcatacagcac	atctaggttt	cgtccgggtg	tgacccgaaag	gtaagatgga	gagccttctc	cctggtttcca	acgagaaaaac	acacgtccaa	ctcagttttg	ctgtttttaca	ggtttccgac	gtgtcgtgtac	gtggttttgg	agactccgtg	gaggaggtct
tatcagagggc	acgtcaacat	cttaagagtg	gcacttctg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tggttcaatca	acgttcggat	gctcgaaactg	cacctccatg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcoatt	cagtacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tctttgttaag	aacggtaata	aaggagctgg	tggtccatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactgga	cactaaacac

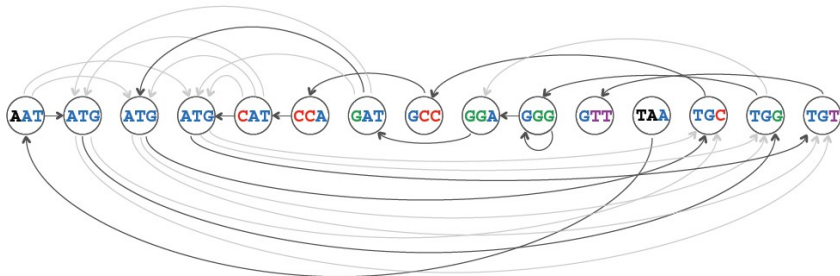
Problem rekonstrukcije niske - grafovska reprezentacija



Slika: Genomska putanja TAATGCCATGGGATGTT

attaaggtt	tataccttcc	caggttaacaa	acaaacaaac	tttogatctc	ttgtagatct	gtttctotaaa	cgaacttttaa	aattctgtgtg	gotgtcaactc	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaaggg
tgcttaacgtt	ttcgtccgtg	ttgcagccga	tcacagcac	atctaggttt	cgtccgggtg	tgacogaaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaaac	acacgtocaa	ctcagtttgc	ctgtttttaca	ggttcogac	gtcctcgtac	gtggctttgg	agactcogtg	gaggaggtct
tatcagagggc	acgtcaacat	cttaagatg	gcacttctg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccttatg	tgttcaatcaa	acgttcggat	gctogaactg	cacctcctag	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggtc	gtagtgggta
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tgggcaatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactgaaa	cactaaacac

Problem rekonstrukcije niske - grafovska reprezentacija



Slika: Genomska putanja TAATGCCATGGGATGTT

Da li je ovo jedino rješenje?

attaaggtt	tataccttcc	caggtacaaa	acaaacaaac	tttogatoto	ttgtagatct	gtttototaaa	cgaacttttaa	aatctgtgtg	gotgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acacagatga	cttgtctatc	ttctgaagga
tgctaacggt	ttgttcogtg	ttgcagocga	tcatacagcac	atctaggttt	cgtcoggggtg	tgacogaaaag	gtaagatgga	gagocctgtc	cctggttttca	acgagaaaac	acacgtocaa	ctcagtttgc	ctgtttttaca	ggttcogac	gtgctogtac	gtggttttgg	agactcogtg	gaggaggtct
tatcagagga	acgtcaacat	cttaagatg	gcacttgtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagocctatg	tggttcaatca	acgttcoggt	gctogaactg	cacctcatgg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcoatt	cagtacgggc	gtagtgggga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggotttacc	gcaaggttct	tcttctgaag	aacggtaata	aaggagctgg	tggtccatagt	taaggcgocg	atctaaagtc	atttgactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactgaaa	cactaaacac

Problem rekonstrukcije niske - problem Hamiltonove i Ojlerove putanje

Hamiltonova putanja

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gotgtcactc	ggctgcctgc	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacgagtaa	ctcgtctatc	ttctgcaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacgcgaaag	gtaagatgga	gagccttgtc	cctggtttca	acgcgaaaaac	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggtcttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaaagatg	gcacttgtgg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctcgaaactg	cacctcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaaagtc	atttgaactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacca

Problem rekonstrukcije niske - problem Hamiltonove i Ojlerove putanje

Hamiltonova putanja - putanja koja prolazi kroz sve **čvorove** grafa tačno jednom.

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttta	aatctgtgtg	gotgtcaact	ggctgcactg	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacgagtaa	ctcgtctatc	ttctgcaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacgcgaaag	gtaagatgga	gagccttgtc	cctggtttca	acgcagaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggtcttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaaaagatg	gcacttgttg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctcgaaactg	caacctcatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaaagtc	atttgactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacca

Problem rekonstrukcije niske - problem Hamiltonove i Ojlerove putanje

Hamiltonova putanja - putanja koja prolazi kroz sve **čvorove** grafa tačno jednom.

- Problem pronalaska Hamiltonove putanje je NP-težak - nema efikasnog algoritma za rješavanje

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttta	aatctgtgtg	gotgtcaact	ggctgcactg	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacgagtaa	ctgtctctac	ttctgaggc
tgcttaagg	tttgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgacgcgaag	gtaagatgga	gagccttgct	cctggtttca	acgcagaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgtcgtgac	gtggttttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaaaagatg	gcacttgtgg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctogaactg	cacctcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtct	atttgaactta	ggcgcagcago	ttggcaactga	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacac

Problem rekonstrukcije niske - problem Hamiltonove i Ojlerove putanje

Hamiltonova putanja - putanja koja prolazi kroz sve **čvorove** grafa tačno jednom.

- Problem pronalaska Hamiltonove putanje je NP-težak - nema efikasnog algoritma za rješavanje

Ojlerova putanja

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcactg	ttagtgcact	cacgcagtat	aattataaac	taattactgt	cgttgacagg	acacagataa	ctctctatcc	ttctgaggcc
tgcttaacgt	tttgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacgcgaaag	gtaagatgga	gagccttgtc	cctggtttca	acgcagaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggttttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaaaagatg	gcacttgttg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tggtccatcaa	acgttcggat	gctcgaaactg	caacctcatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtaacggtc	gtagtgggtc
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaaagtc	atttgaactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactggaa	cactaaacaa

Problem rekonstrukcije niske - problem Hamiltonove i Ojlerove putanje

Hamiltonova putanja - putanja koja prolazi kroz sve **čvorove** grafa tačno jednom.

- Problem pronalaska Hamiltonove putanje je NP-težak - nema efikasnog algoritma za rješavanje

Ojlerova putanja - putanja koja prolazi kroz sve **grane** grafa tačno jednom.

attaaagggt tataccttcc caggtaacaa accaaccacc ttctgatctc ttgtagatct gttctotaaa cgaactttaa aatctgtgtg gotgtcactc ggtgtcatgc ttagtgcact cagcgagtat aattaataac taattactgt cgttgacagg acacagatga ctgtctatct ttctgaggc
tgcttaagggt ttgtccogtg ttgcagccga tcctcagcac atctagggtt cgtccgggtg tgacogaaa gtaagatgga gagcctgtgc cctggtttca acagaaaaac acacgtccaa ctacgtttgc ctgttttaca ggttcogac gtgtcgttac gtgctttgg agactcogtg gaggaggtc
tatcagaggo acgtcaacat cttaaagatg gcacttgttg cttagtagaa gttgaaaaag toaacttgaa cagccctatg tgttcaatca acgttcoggt gctogaactg cactctatgg tcatgttatg gttgagctgg tagcagaact cgaaggcaat cagtaagggt gtagtggtg
gacacttggt gtcttctgac ctcatgtggg cgaataacca gtggtttacc gcaaggttct tcttctgaag aacggttaata aaggagctgg tggccatagt taaggcgccg atctaaaagtc atttgactta ggcgcagago ttggcactga tctttatgaa gattttcaag aaaactgga cactaaaca

Problem rekonstrukcije niske - problem Hamiltonove i Ojlerove putanje

Hamiltonova putanja - putanja koja prolazi kroz sve **čvorove** grafa tačno jednom.

- Problem pronalaska Hamiltonove putanje je NP-težak - nema efikasnog algoritma za rješavanje

Ojlerova putanja - putanja koja prolazi kroz sve **grane** grafa tačno jednom.

- Problem pronalaska Ojlerove putanje nije NP-težak - postoji efikasan algoritam za rješavanje

attaaagggtt	tataccttcc	caggtaacaa	acaaacaaac	tttogatoto	ttttagatct	gttctotaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcattg	ttagtgcact	cacgcagtat	aattataaac	taattactgt	cgttgacagg	acacagataa	ctgtctatcc	ttctgaggcc
tgcttaacgtt	tttgtccgtg	ttgcagccga	tcctcagcac	attcaggttt	cgtccgggtg	tgacgaaaag	gtaagatgga	gagccttgct	cctggtttca	acgagaaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtgcttttgg	agactccgtg	gaggaggtct
tatcagaggo	acgtcaacat	cttaaaagtg	gcacttgttg	cttagtagaa	gttgaaaaag	gogttttgoc	tcacacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctogaactg	cacctcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtc
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tctttgttaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaaagtc	atttgactta	ggcgcagcgc	ttggcaactga	tccttatgaa	gattttcaag	aaaactggaa	cactaaacaa

Problem rekonstrukcije niske - problem Hamiltonove i Ojlerove putanje

Hamiltonova putanja - putanja koja prolazi kroz sve **čvorove** grafa tačno jednom.

- Problem pronalaska Hamiltonove putanje je NP-težak - nema efikasnog algoritma za rješavanje

Ojlerova putanja - putanja koja prolazi kroz sve **grane** grafa tačno jednom.

- Problem pronalaska Ojlerove putanje nije NP-težak - postoji efikasan algoritam za rješavanje
- Potrebno je drugačije definisati graf

attaaagggtt	tataccttcc	caggtaacaa	acaaacaaac	tttccatctc	tttccatctc	gttctctaaa	cgaactttta	aattctgtgtg	gttctctaaa	ggctgcactc	ttagtgcact	caggcagtat	aatttaataac	taattactgt	cgttgacagg	acacagagtaa	ctcgtctatc	ttctgcaggc
tgcttaacgtt	ttcgtccgtg	ttgcagccga	tcctcagcac	attcaggttt	cgtccgggtg	tgacgcagaa	gtaagatgga	gagccttgct	cctggtttca	acgcagaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcagc	gtgcctgtac	gtgccttttg	agactccgtg	gaggaggtct
tatcagaggo	acgtcaacat	cttaagagtg	gcactttgtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcacacttgaa	cagccctatg	tgcttcacaa	acgttcggat	gctgcagact	cacccctatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcact	cagtaacggtc	gtagtggtga
gacacttggt	gtccttgctc	ctcatgtggg	cgaataacca	gtgccttacc	gcaaggttct	ttcttcttaag	aacggttaata	aaggagctgg	tgccctatgt	taaggcgcgc	atctaaagtc	atttgactta	ggcgcagcgc	ttggcactga	tccttatgaa	gatttttaag	aaaactggaa	cactaaacaa

Problem rekonstrukcije niske - problem Ojlerove putanje

Predstavljanje k-gramskog sastava na drugačiji način:

- Čvorovi grafa su k-1-grami
 - Izlazni čvor će biti prefiks k-grama
 - Ulazni čvor će biti sufiks k-grama

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aatttaataac	taattactgt	cgttgacagg	acacagataa	ctcgtctatc	ttctgacggc
tgcttaacgt	tttgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacgcgaaag	gtaagatgga	gagccttgtc	cctgggttca	acgcagaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggctttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaaaagtg	gcacttgtyg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctogaactg	caacctcatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgctc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tctttogtaag	aacggtaata	aaggagctgg	tgcccatagt	taaggcgccg	atctaaagtc	atttgaacta	ggcgacgagc	ttggcaactg	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacac

Problem rekonstrukcije niske - problem Ojlerove putanje

Predstavljanje k-gramskog sastava na drugačiji način:

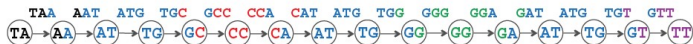
- Čvorovi grafa su k-1-grami
 - Izlazni čvor će biti prefiks k-grama
 - Ulazni čvor će biti sufiks k-grama
- Grane su označene k-gramom

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcactg	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacagataa	ctcgtctatc	ttctgaaggc
tgcttaagggt	tttgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgacogaaaag	gtaagatgga	gagccttgct	cctgggttca	acgagaaaaac	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggctttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaaaagtg	gcacttggg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctogaactg	caacctcatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaagggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaaagtc	atttgaacta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacaa

Problem rekonstrukcije niske - problem Ojlerove putanje

Predstavljanje k-gramskog sastava na drugačiji način:

- Čvorovi grafa su k-1-grami
 - Izlazni čvor će biti prefiks k-grama
 - Ulazni čvor će biti sufiks k-grama
- Grane su označene k-gramom



Slika: Grafovska reprezentacija za TAATGCCATGGGATGTT

3-grami su grane, 2-grami su čvorovi

attaaggtt	tataccttcc	caggtacaaa	acaaacaaac	tttccatctc	ttgtagatct	gtttctataa	cgaactttta	aattctgtgt	gctgtcaact	ggctgcattc	ttagtgcact	cacgcagtat	aattataaac	taattactgt	cgttgacagg	acacagataa	ctcgtctatc	ttctgcaggc
tgcttaaggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacgaaaag	gtaagatgga	gagccttgct	cctggtttca	acgagaaaaa	acacgtccaa	ctcagtttgc	ctgtttttca	ggttcgcgac	gtcctcgtac	gtgcttttgc	agactccgtg	gaggaggtct
tatcagaggo	acgtcaacat	cttaaaagtg	gcactttgtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcacacttga	cagccctatg	tggttcataa	acgttcggat	gctcgaactg	cacctccatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtc
gacacttggt	gtccttgtcc	ctcatgtggg	cgaatatacca	gtggcttacc	gcaaggttct	tctttgttaag	aacggttaata	aaggagctgg	tggtccatag	taaggcgccg	atctaaagtc	atttgactta	ggcgacgago	ttggcactga	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacac

Problem rekonstrukcije niske - problem Ojlerove putanje

Predstavljanje k-gramskog sastava na drugačiji način:

- Čvorovi grafa su k-1-grami
 - Izlazni čvor će biti prefiks k-grama
 - Ulazni čvor će biti sufiks k-grama
- Grane su označene k-gramom



Slika: Grafovska reprezentacija za TAATGCCATGGGATGTT

Neki čvorovi imaju istu oznaku, možemo li ih spojiti?

attaagggtt	tataccttcc	caggtaacaa	acaaacaaac	tttogatoto	ttgtagatct	gtttototaaa	cgaacttttaa	aatctgtgtg	gotgtcaactc	ggctgcattgc	ttagtgcactc	cacgcagttat	aatttaataac	taattactgt	cgttgacagg	acacagatgaa	ctcgtctatc	ttctgcaggc
tgottaagggt	tttgtccgtg	ttgcagccga	tcatacagcac	atctaggttt	cgtccgggtg	tgacgaaaag	gtaagatgga	gagccttgtc	cctggttttca	acgagaaaaac	acacgttccaa	ctcagttttgc	ctgtttttaca	ggtttccgac	gtgctcgtac	gtggttttgg	agactccgtg	gaggaggtct
tatcagaggo	acgtcaacat	cttaaaagatg	gcactttgtg	cttagtagaa	gttgaaaaag	gogtttttgc	tcacacttgaa	cagccctatg	tgttccatcaa	acgttoggat	gctogaactg	cacctccatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaatatacca	gtggottacc	gcaaggttct	tctttgttaag	aacggttaata	aaggagctgg	tggtccatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgacgago	ttggcaactga	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacaa

Problem rekonstrukcije niske - problem Ojlerove putanje

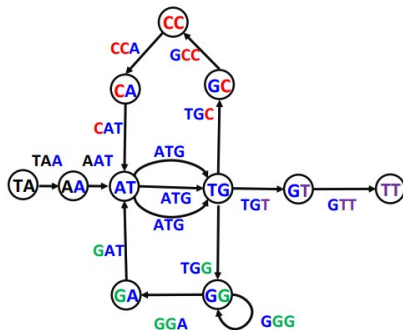
Sve čvorove koji imaju istu oznaku je moguće spojiti u jedan čvor, zadržavajući sve grane.
Takav graf se naziva **De Bruijnov graf**.

attaaaggtt	tataccttcc	caggttaacaa	accaaccaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttta	aatctgtgtg	gotgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggc
tgcttacggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgaccgaaag	gtaagatgga	gagccttgtc	cctgggttca	acgagaaaa	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtcctcgtac	gtggcttttg	agactccgtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaaaagatg	gcacttgttg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctcgaaactg	cacctcatgg	tcctgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggctc	gtagtgggtga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttcgttaag	aacggtaata	aaggagctgg	tggccatagt	tacggcgccg	atctaaagtc	atttgactta	ggcgacgago	ttggcactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

Problem rekonstrukcije niske - problem Ojlerove putanje

Sve čvorove koji imaju istu oznaku je moguće spojiti u jedan čvor, zadržavajući sve grane.

Takav graf se naziva **De Bruijnov graf**.

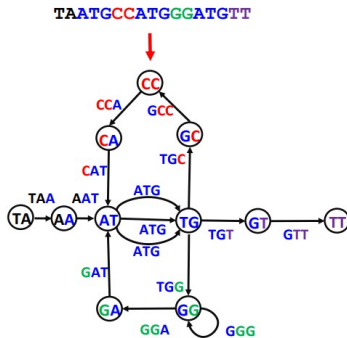


Slika: De Bruijnov graf za TAATGCCATGGGATGTT

attaaggtt	tataccttcc	caggtacaaa	acaaacaaac	tttogatoto	ttgtagatct	gtttototaaa	cgaacttttaa	aattotgtgtg	gotgtcaactc	ggctgcactgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	cttgtotatc	ttctgaaggg
tgcttaacgtt	tttgtccgtg	ttgcagccga	tcatacagcac	atctaggttt	cgtccgggtg	tgacogaaag	gtaagatgga	gagocctgtc	cctgggttcca	acgagaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggtttcogac	gtgtcgttac	gtggctttgg	agactcogtg	gaggaggtct
tatcagagggc	acgtcaacat	cttaagatg	gcacttgtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcacattgaa	cagocctatg	tgttcaatcaa	acgttcggat	gctogaactg	cacctcatgg	tcattgtatg	gttgagctgg	tagcagaact	cgaaggcoatt	cagtaacggtc	gtagtgggtga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tottogtaag	aacggttaata	aaggagctgg	tgccaatagt	taaggcgocg	atctaaagtc	atttgactta	ggcgacgagc	ttggcaactga	tcottatgaa	gattttcoag	aaaactggee	cactaaacac

Problem rekonstrukcije niske - De Bruijnov graf od k-grama

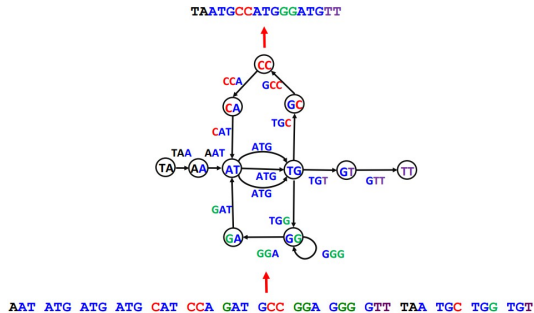
Vidjeli smo kako od zadate niske (genoma) možemo konstruisati De Bruijnov graf. U realnim primjenama, potrebno je konstruisati graf od k-grama (očitanja).



Slika: Genom → De Bruijnov graf

attaaggtt	tataccttcc	caggtacaaa	acaaacaaac	tttogatoto	ttgtagatct	gtttototaaa	cgaacttttaa	aatctgtgtg	gotgtcaactc	ggctgcactgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	cttgtotatc	ttctgaagga
tgottaagg	tttgtccgtg	ttgcagcoga	tcatacagcac	atctaggttt	cgtccgggtg	tgacogaaag	gtaagatgga	gagcctttgc	cctgttttca	acgagaaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcogcac	gtgtcgtgac	gtggttttgg	agactcogtg	gaggaggtct
tatcagagga	acgtcaacat	cttaagatg	gcacttgtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccttatg	ttgtcatcaa	acgttcggat	gctogaactg	cacctcatg	tcactgtatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtaacggtc	gtagtgggta
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tottogtaag	aacggttaata	aaggagctgg	ttggcattag	taaggcgocg	atctaaagtc	atttgactta	ggcagcagag	ttggcaactg	tcottatgaa	gattttcaag	aaaactgaaa	cactaaacac

Problem rekonstrukcije niske - De Bruijinov graf od k-grama



Slika: Očitavanja \rightarrow De Bruijnov graf \rightarrow genom

atttaaaaggtt tataccttcc caggtaacaa accaaccaac ttctgatctc ttgtagatct gttctotaaa cgaactttaa aatctgtgtg gctgtcactc ggtcgtatcg ttatgtgaact caagcagtat aattaataac taattactgt cyttgaaggy acacagatga ctctgtctat tctgcaggc

tgcttaacggt ttctgtccgtg ttgcagccga tcatacagac atctaggattt ggtccgggtg tgaccgaaag tcaagtatga gagccttgtc cctgggtttca acgagaaaaa acagctccaa ctacgtttgc ctgttttata ggttcgcac gtctcgtac gtggcttgg agactccgtg gaggaggttg

tatcacagagc agctcaacat cttaaaagatg ccaacttgttg cttagattaga gttgttttgc gtaactttga cagccctatg ttctctatcaa acgtttccgat gctcgaactg caactcatga ctactcatga gttgacatg tagcagaat cgaagggcat cactacggct gactgtgtga

gaactcttgt atctctctcc ctctaatatgg caaataacca cttagcttacc gcaagatttct tctgtttatga aaagctgtatg ttgcttcatac acgttgccttg atttcagatt atttcaatga ggcacagac ttgtctgaa ttgactgaa cgttttcaa aaaaattgaa ctaataacaa ctaataacaa

Problem rekonstrukcije niske - De Bruijnov graf od k-grama

Postupak:

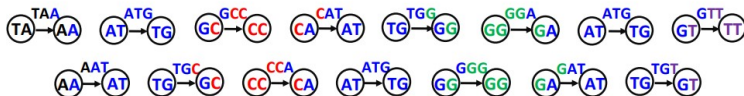
- Za svaki k-gram kreiramo dva čvora i jednu granu, jedan čvor predstavlja prefiks, a drugi sufix k-grama

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttta	aatctgtgtg	gotgtcaact	ggctgcactc	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacgagtaa	ctcgtctatc	ttctgcaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgacgcgaaag	gtaagatgga	gagccttgtc	cctggtttca	acgcagaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggttttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaaaagatg	gcacttgttg	cttagtagaa	gttgaaaaag	gogttttgoc	tcacactgaa	cagccctatg	tgttcaatca	acgttcggat	gctcgaaactg	cacctcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtc	atttgaactta	ggcgacgagc	ttggcaactga	tccttatgaa	gatttttaag	aaaaactgaa	cactaaacac

Problem rekonstrukcije niske - De Bruijnov graf od k-grama

Postupak:

- Za svaki k-gram kreiramo dva čvora i jednu granu, jedan čvor predstavlja prefiks, a drugi sufix k-grama



- Čvorove sa istom oznakom spajamo u jedan dok ne dobijemo graf sa jedinstvenim čvorovima

attaaagggt tacaacttcc caggttaacaa accaaccacac tttagatctc tttagatctc tttagatctc cgaactttta aatctgtgtg gctgtcaact gctgtcaact ttagtgcaact cagcagttat aattataaac taattactgt cgttgacagg acacagttaa ctgctctatc ttctgaaggc
tgcttaacgt ttgttcogtg ttgcagccga tcatcagcac atctaggttt cgtccgggtg tgacgaaaag gtaagatgga gagcctgtgc cctgtgttca acagaaaaac acacgtccaa ctacgttttg cgttttttga ggttcogac gtgtcgttac gtgctgttac gtaggctttg agactcogtg gaggaggtct
tatcagaggo acgtcaacat cttaaagatg gcacttgttg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccctatg tgttcaatca acgttoggat gctogaactg cactctatgg tcatgttatg gttgagctgg tagcagaact cgaaggcact cagtaocgtc gtagtggtga
gacacttggt gtctttgtcc ctcatgtggg cgaataacca gtggtttacc gcaaggttct tcttctgaag aacgttaata aaggagctgg tggccatagt taaggcgcgc atctaaagtc atttgactta ggccagcago ttggcaactg tctttatgaa gattttcaag aaaaactgaa cactaaacac

Problem rekonstrukcije niske - De Bruijnov graf od k-grama

Postupak:

- Za svaki k-gram kreiramo dva čvora i jednu granu, jedan čvor predstavlja prefiks, a drugi sufix k-grama
- Čvorove sa istom oznakom spajamo u jedan dok ne dobijemo graf sa jedinstvenim čvorovima

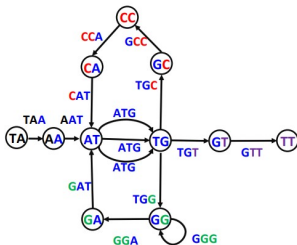


attaaagggtt tatacctttcc caggtaacaa accaaccacac tttagatctc tttagatctc gttctctaaa cgaactttaa aatctgtgtg gotgtcaact ggctgcactc ttagtgcaact cagcagttat aattataaac taattactgt cgttgacagg acacagatga ctgtctctac ttctgaaggc
tgcttaacgtt ttgttcogtg ttgcagcaga tcatcagcac atctaggttt cgtccgggtg tgacagaaa gtagatgga gagcctgtgc cctggtttca acagagaaac acacgtccaa ctacagtttg cgtttttaca ggttcogac gtgtcgttac gtggttttg agactcogtg gaggaggtct
tatcagaggo acgtcaacat cttaaagatg gcacttgtyg cttagtagaa ggtttttgoc tcaacttgaa cagccctatg tgttcaatca acgttoggat gotogaactg cactctatgg toatgttatg gttgagctgg tagcagaact cgaaggcact cagtaaggto gtagtggtga
gacacttggt gtctttgtcc ctcatgtggg cgaataacca gtggtttacc gcaaggttct tctttgttaag aacggttaata aaggagctgg tggccatagt taaggcgocg atctaaagtc atttgactta ggcgcagago ttggcaactga tctttatgaa gattttcaag aaaaactgaa caactaaaca

Problem rekonstrukcije niske - De Brujinov graf od k-grama

Postupak:

- Za svaki k-gram kreiramo dva čvora i jednu granu, jedan čvor predstavlja prefiks, a drugi sufixs k-grama
- Čvorove sa istom oznakom spajamo u jedan dok ne dobijemo graf sa jedinstvenim čvorovima



Slika: De Bruijnov graf za TAATGCCATGGGATGTT

attaaagggtt tatacctttcc caggtaacaa accaaccacac tttagatctc tttagatctc gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcctgc ttagtgcact cagcgagtat aattataaac taattactgt cgttgacagg acacagatga ctgctctatc ttctgaaggc
tgcttaagggt ttgctcogtg ttgcagcaga tcatacagcac atctaggttt cgtccgggtg tgacgcagaa gtaagatgga gagcctgtgc cctgtgttca acagagaaaa acagctgcac ctgagtttgc ggttcogac gtgctcgtac gtagcctttg agactcogtg gaggaggtct
tatcagaggo acgtcaacat cttaaagatg gcacttcttg cttagtagaa gttgaaaaag gogttttgoc taaacttgaa cagccctatg ttttcaatca acgttoggat gctogaactg caactctatg tcatgttatg gttgagctg tagcagaact cgaaggcaat cagtaacggtc gtagtggtga
gacacttggt gtccttctcc ctcctgtggg cgaataacca gtggtttacc gcaaggttct tctttgttaag aacggttaata aaggagctgg tggccatagt taaggcgocg atctaaagtc atttgactta ggccagcago ttggcaactga tctttatgaa gattttcaag aaaaactgaa caactaaaca

Ojlerova teorema

Svaki Ojlerov graf je balansiran.

Svaki povezan graf i balansiran graf je Ojlerov.

Povezan graf je graf u kojem postoji put između svaka dva čvora.

EULERIANCYCLE(*Graph*)

form a cycle *Cycle* by randomly walking in *Graph* (don't visit the same edge twice!)

while there are unexplored edges in *Graph*

select a node *newStart* in *Cycle* with still unexplored edges

form *Cycle'* by traversing *Cycle* (starting at *newStart*) and then randomly walking

Cycle ← *Cycle'*

return *Cycle*

Slika: Pseudokod algoritma EulerianCycle

attaaagggtt tatacctttcc caggtaacaa accaaccacac ttctgatctc ttgtagatct gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcactc ttagtgcaact cagcgagtat aattaataac taattactgt cgttgacagg acacagagtaa ctgctctatc ttctgagggc
tgcttaacggt ttgctcogtg ttgcagocga tcatcagcac atctaggttt cgtccgggtg tgacogaaag gtaagatgga gagccttgtc cctggtttca acagaaaaac acacgtocaa ctacgtttgc ctgcttttaca ggttcogac gtgctogtac gtggctttgg agactcogtg gaggaggtct
tatcagaggo acgtcaacat cttaaagatg gcacttgtyg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccctatg tgttcatcaa acgttcggat gctogaactg cactctatgy tcatgttatg gttgagctgy tagcagaact cgaaggcoatt cagtacggtc gtagtggtga
gacacttggt gtccttgtcc ctcatgtggg cgaataacca gtggtttacc gcaaggttct tottctgaag aacggttaata aaggagctgy tggccatagt taaggcgcog atctaaagtc atttgactta ggcgacgago ttggcaactga tccttatgaa gattttcaag aaaactgcaa cactaaacac

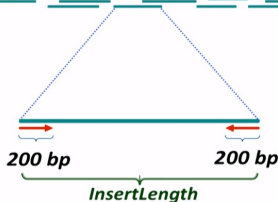
Sastavljanje parova očitavanja

- **Problem:** Postojanje više Ojlerovih putanja u De Bruijnovom grafu
- **Rješenje:** Sekvenciranje sa parovima očitavanja

Multiple identical copies of genome



↓ Randomly cut genomes into large equally sized fragments of size *InsertLength*



Generate **read-pairs**:
two reads from the
ends of each fragment
(separated by a fixed
distance)

attaaagggt tataccttcc caggtaacaa accaaccac tttagatctc ttgtagatct gttctotaaa cgaactttaa aatctgtgtg gctgtcacct ggctgcatgc ttagtgcaact cagcgagtat aattaataac taattactgt cgttgacagg acacagatga ctgctctatc ttctgagggc
tgcttaacgt ttgtccgtg ttgcagccga tcatcagcac atctagggtt cgtccgggtg tgacccgaa gtaagatgga gagccttgct cctgggttca acagagaaac acacgtccaa ctacagtttg cttgctgtac gtgctgttac gtggctttgg agactcogtg gaggaggtct
tatcagagga acgtcaacat cttaaagatg gcacttgtyg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccctatg tgttcatcaa acgttcggat gctogaactg cacttccttg tcatgttatg gttgagctgg tagcagaact cgaaggcaatt cagtaacggtc gtagtgggtga
gacacttggt gtctttgtcc ctcatgtggg cgaataacca gtggtttacc gcaaggttct totttogtaag aacggtaata aaggagctgg tggccatagt taaggcgocg atctaaagtc atttgactta ggccagcago ttggcaactga tccttatgaa gattttcaag aaaactgga cactaaacac

Sastavljanje parova očitavanja

Upareni k-gram predstavlja par k-grama na fiksiranom rastojanju d u genomu.



Slika: TCA i TCC na rastojanju $d=11$ čine jedan upareni 3-gram

attaaggtt	tataccttcc	caggtacaaa	acaaacaaac	tttogatoto	ttgtagatct	gtttototaaa	cgaacttttaa	aattotgtgtg	gotgtcaactc	ggctgcactgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggc
tgottaaggt	ttcgtccgtg	ttgcagccga	tcatacagcac	atctagggtt	cgtccgggtg	tgacccgaag	gtaagatgga	gagccttgtc	cctggttttca	acgagaaaaac	acacgtocaa	ctcagtttgc	ctgtttttaca	ggttcggcac	gtgctcgtac	gtggctttgg	agactcogtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagatg	gcacttggtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctcgaactg	cacctcatgg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtaacggtc	gtagtgggtga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	totttctaag	aacggtaata	aaggagctgg	tgcccatagt	taaggcgcgcg	atctaaagtc	atttgactta	ggcgacgago	ttggcaactga	tccttatgaa	gattttcaag	aaaactgga	cactaaacac

Sastavljanje parova očitavanja

Upareni k-gramski sastav $PairedComposition_d(Text)$ se sastoji od svih k-grama niske $Text$ i njihovih parova.

$PairedComposition(TAATGCCATGGGATGTT)$

TAA GCC
AAT CCA
ATG CAT
TGC ATG
GCC TGG
CCA GGG
CAT GGA
ATG GAT
TGG ATG
GGG TGT
GGA GTT

Slika: $PairedComposition_3(TAATGCCATGGGATGTT)$

attaaggtt	tataccttcc	caggtacaaa	acaaaccaac	tttogatctc	ttgtagatct	gttctotaaa	cgaactttta	aatctgtgtg	gctgtcaact	ggctgcctgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcatacgcac	atctaggttt	cgtccgggtg	tgacccgaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtcctcgtac	gtggttttgg	agactccgtg	gaggaggtct
tatcagaggc	acgtcaacat	cttaagatg	gcacttgttg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatcaa	acgttcggat	gctcgaactg	caactcattg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggctc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgaagago	ttggcaactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

Sastavljanje parova očitavanja

TAA	AAT	ATG	TGC	GCC	CCA	CAT	ATG	TGG	GGG	GGA
GCC	CCA	CAT	ATG	TGG	GGG	GGA	GAT	ATG	TGT	GTT
AAT	ATG	ATG	CAT	CCA	GCC	GGA	GGG	TAA	TGC	TGG
CCA	CAT	GAT	GGA	GGG	TGG	GTT	TGT	GCC	ATG	ATG

Slika: Leksikografski poredak kolekcije $\text{PairedComposition}_3$ niske TAATGCCATGGGATGTT

Problem rekonstrukcije niske na osnovu parova očitavanja:

- **Ulaz:** Kolekcija uparenih k-grama
- **Izlaz:** Niska *Text* takva da je $\text{PairedComposition}_k(\text{Text})$ jednak kolekciji uparenih k-grama

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttaa	aattctgtgt	gtgtgcactc	ggctgcactc	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acaagagtaa	cttgtctatc	ttctgaaggg
tgottaagggt	tttgtccgtg	ttgcagccga	tcatacgcac	atctagggtt	cgtccgggtg	tgacccgaa	gtaagatgga	gagccttgct	cctggtttca	acgagaaaa	acaagtcoc	ctcagtttgc	ctgtttttca	ggtttcgcac	gtgtcgtgac	gtggcttttg	agaactccgt	gaggaggtct
tatcagaggg	acgtcaacat	cttaagagtg	gcacttggtg	cttagtagaa	gttgaaaaag	gggtttttgc	tcaacttgaa	cagccctatg	tggttcacaa	acgttcggat	gttcgaactg	cacctcatg	tcattgttat	gttgagctgg	tagcagaact	cgaaggcaat	cagtacggct	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tctttgtta	aacggttaata	aaggagctgg	tgcccatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgacgago	ttggcaactg	tccttatgaa	gattttcaag	aaaactgaaa	cactaaacac

Sastavljanje parova očitavanja



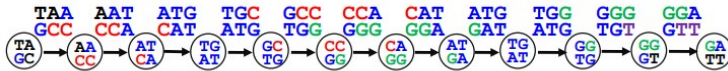
Slika: Genom TAATGCCATGGGATGTT predstavljen kao putanja

attaaggtt	tataccttcc	caggttaacaa	acaaacaaac	tttogatctc	ttgtagatct	gtttctotaaa	cgaacttttaa	aattctgtgtg	gctgtcaactc	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcatacgcac	atctaggttt	cgtccgggtg	tgacccgaag	gtaagatgga	gagccttctc	cctggtttca	acgagaaaa	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtcctcgtac	gtggctttgg	agactccgtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagatg	gcacttctg	cttagtagaa	gttgaaaaag	gcgtttttgoc	tcaacttgaa	cagccctatg	tgttcaatcaa	acgttcggat	gctcgaaactg	caactcattg	tcattgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggctc	gtagtgggtga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tctttgttaag	aacggtaata	aaggagctgg	tgcccatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactgga	cactaaacac

Sastavljanje parova očitavanja



Slika: Genom TAATGCCATGGGATGTT predstavljen kao putanja

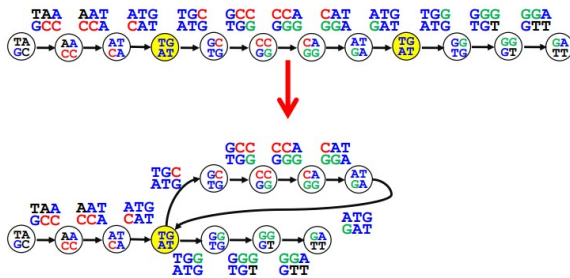


Slika: Graf koji odgovara uparenom 3-gramskom sastavu TAATGCCATGGGATGTT

Grana - upareni 3-gram, čvor - prefiks i sufiks uparenog 3-grama

attaaggtt	tataccttcc	caggtaacaa	accaaccaac	tttogatoto	ttgtagatct	gttctotaaa	cgaactttaa	aatctgtgtg	gctgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	cttgtctatc	ttctgaaggc
tgottaaggt	tttgtccgtg	ttgcagccga	tcatacgcac	atctaggttt	cgtccgggtg	tgacccgaag	gtaagatgga	gagccttgct	cctggtttca	acagagaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	gggttcgcac	gtgtcgttac	gtggctttgg	agactccgtg	gaggaggtct
tatcagaggo	acgtcaacat	cttaagatg	gcacttgtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctogaactg	cacctcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcoatt	cagtacggctc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tctttogtaag	aacggttaata	aaggagctgg	tgcccatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

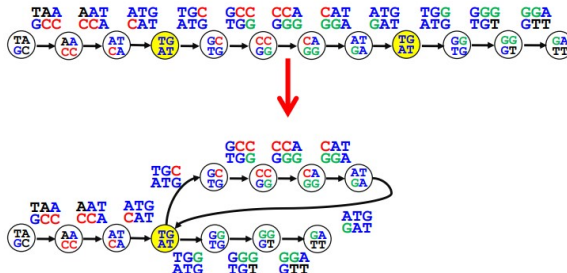
Sastavljanje parova očitavanja



Slika: Upareni De Bruijnov graf za TAATGCCATGGGATGTT

Dva čvora (TG AT) su spojeni u jedan čvor

Sastavljanje parova očitavanja



Slika: Upareni De Bruijinov graf za TAATGCCATGGGATGTT

Dva čvora (TG AT) su spojeni u jedan čvor **Kako konstruisati upareni De Bruijinov graf na osnovu uparenih k-grama?**

attaaggtt	tataccttcc	caggttaacaa	acaaacaaac	tttogatoto	ttgtagatct	gtttototaaa	cgaacttttaa	aattgtgtgtg	gotgtcaactc	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acacagagtaa	cttgtotatc	ttctgagggc
tgottaaggt	ttgtccgtg	ttgcagcoga	tcatacagcac	atctaggttt	cgtccgggtg	tgacogaaag	gtaagatgga	gagccttgtc	cctgttttcaa	acagaaaaac	acacgtocaa	ctcagtttgc	ctgtttttaca	ggtttogcag	gtgtcgtgac	gtggcttttg	agactcogtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagagatg	gcacttgtg	cttagtagaa	gttgaaaaag	gogttttgoc	tcacactgaa	cagccctatg	tgttcaatcaa	acgttcggat	gctogaacgt	cacctcctag	tcactgttatg	gttgagctg	tagcagaact	cgaaggcaatt	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacaa	gtggottacc	gcaaggttct	tcttgtagaag	aacggtaata	aaggagctgg	tggtccatagt	taaggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactgaaa	cactaaacac

Sastavljanje parova očitavanja

Upareni De Bruijnov graf na osnovu kolekcije uparenih k-grama:

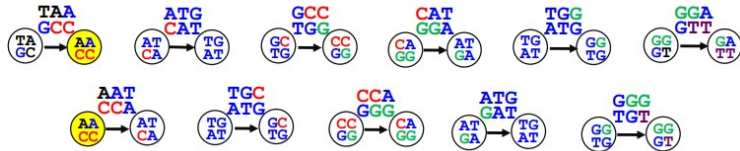
- Svaka grana je označena jednim uparenim k-gramom
- Svaki čvor je označen prefiksom/sufiksom izlazne/ulazne grane
- Čvorovi sa istom oznakom se spajaju u jedan čvor

attaaggtt	tataccttcc	caggtaacaa	accaaccaac	tttogatctc	ttgtagatct	gttctotaaa	cgaactttta	aatctgtgtg	gotgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaaggg
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgaccgaaag	gtaagatgga	gagccttgtc	cctgggttca	acgagaaaac	acaagtcocaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggctttgg	agactcogtg	gaggaggtct
tatcagaggg	acgtcaacat	cttaagagtg	gcacttgctg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctcgaaactg	cacctcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggctc	gtagtgggtga
gacacttggt	gtccttgctc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttcgttaag	aacggtaata	aaggagctgg	tgcccatagt	tacggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcactga	tccttatgaa	gattttcaag	aaaactgga	cactaaacac

Sastavljanje parova očitavanja

Upareni De Bruijnov graf na osnovu kolekcije uparenih k-grama:

- Svaka grana je označena jednim uparenim k-gramom
- Svaki čvor je označen prefiksom/sufiksom izlazne/ulazne grane
- Čvorovi sa istom oznakom se spajaju u jedan čvor



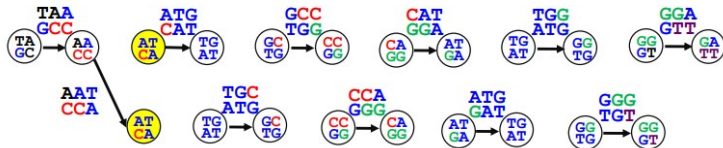
Slika: Konstrukcija uparenog De Bruijnovog grafa na osnovu uparenih k-grama

attaaagggtt tataccttcc caggtaacaa accaaccac tttogatatc ttgtagatct gttctotaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcactc ttagtgcact cagcgagtat aattaataac taattactgt cgttgacagg acacagatga ctctgtatcc ttctgaagga
tgcttaacgtt ttgttcogtg ttgcagccga tcatcagcac atctaggttt cgtccgggtg tgacogaaag gtaagatgga gagccttgct cctgtgttca acagagaaac acacgtccaa ctacagttgc gttgtcagga
tatcagagga acgtcaacat cttaaagatg gcacttgctg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccctatg tgttcaatca acgttcggat gctogaactg cacttcactg tagcagaact cgaaggcaat cagtaacggtc gtagtggtga
gacacttggt gtccttgctcc ctcatgtggg cgaataacca gtggottacc gcaaggttct tcttctgaag aacggttaata aaggagctgg tggccatagt taaggcgocg atctaaagtc atttgactta ggcgacgagc ttggcaactga tccttatgaa gattttcaag aaaactgga cactaaacac

Sastavljanje parova očitavanja

Upareni De Bruijnov graf na osnovu kolekcije uparenih k-grama:

- Svaka grana je označena jednim uparenim k-gramom
- Svaki čvor je označen prefiksom/sufiksom izlazne/ulazne grane
- Čvorovi sa istom oznakom se spajaju u jedan čvor



Slika: Konstrukcija uparenog De Bruijinovog grafa na osnovu uparenih k-grama

attaaagggtt tatacctttcc caggtaacaa accaaccac tttgatcttc ttgtagatct gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcactc ttagtgcact cagcagttat aattataaac taattactgt cgttgacagg acacagatga ctgctctatc ttctgaaggc
tgcttaacgtt ttgttcogtg ttgcagcaga tcatcagcac atctaggttt cgtccgggtg tgacccgaag gtaagatgga gagcctgtgc cctgtgttca acagaaaaac acacgtccaa ctacagtttg cgttttttca ggttcogcac gtcctcgtac gtcgcttttg agactcogtg gaggaggtct
tatcagaggo acgtcaacat cttaaagatg gcactttgtg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagcctatg tgttcatcaa acgttcggat gctcgaactg tcatgttatg ttgcagctg tagcagaact cgaaggcaat cagtaacggtc gtagtggtga
gacacttggt gtcccttgtcc ctcattgtgg cgaataacca gtggtttacc gcaaggttct tcttgtaag aacggttaata aaggagctgg tggccatagt taaggcgcgc atctaaagtc atttgactta atttgactta ggcgcagagc ttggaactga tcttatgaa gattttcaag aaaactgaa cactaaacac

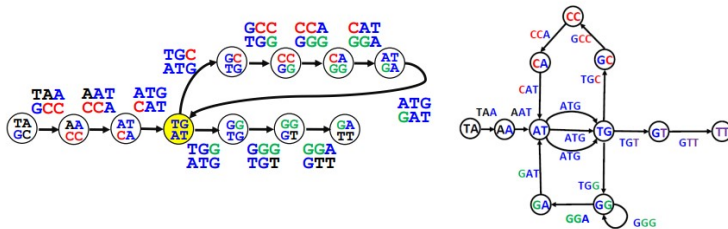
attaaaaggtt tataccttcc caggtaacaa accaaccac ttctgatctc ttgtagatct gttctctaaa cgaactttaa aatctgtgtg gctgtcactc gggtgcgtgc ttagtgaact cagcagtat aattaataac taattactgt cyttgacag cacaagtaa ctctgtctatc ttctgcagg
 tgcctacggt acgttcoggt ttgcagcgca tcactgacac atctagagtt cgttcggggtg tgacagaaag tgaagatgga gagcctgttc cctgtgttca acgagaaaac acatgcctcaa ctacagttgc ttgttttata ggttcgcgac tgcctgtctac gtgcggcttg agactcgtg gagaggttg
 tactagagag cagctcaacat cttaaaagat gcacttgttg cttagtagaa cttgtaaaaa tcaactttaa gacgcctatg ttgttcatac actcctggat gcctgaactg cactctcatg ttgtgagctg tagcagaact cgaagcgatt gtagcggctc gtagaggttg
 gacacttggt gtctctgtcc ctctatgtgg cgaataacca gtggtctacc gaaaggtctc tctgtttaag acaggtttaa aagagcgttg tggccatagt tacggtgcgc attcaaaagc atttgactga ggcacagcag ttgcagctga tccattagaa gattttcaag aaaaactgaa cactaaacac

Sastavljanje parova očitavanja

TAATGCCATGGATGTT

TAATGCCATGGATGTT

TAATGGGATGCCATGTT



Slika: Poređenje uparenog De Brujinovog grafa i običnog De Brujinovog grafa

Lijevo - jedinstvena rekonstrukcija genoma, desno - više mogućih rekonstrukcija genoma

attaaagggtt tataccttcc caggtaacaa accaaccacac tttagatctc ttgtagatct gttctotaaa cgaactttaa aatctgtgtg gctgtcaact ggctgcactg ttagtgcaact cagcgagtat aattaataac taattactgt cgttgacagg acacagagtaa ctgctctatc ttctgagggc
tgcttaacggt ttgttcogtg ttgcagccga tcatcagcac atctaggttt cgtccgggtg tgacccgaag gtaagatgga gagccttgct cctggtttca acagagaaac acacgtccaa ctacagtttg cgtgttttgg gttgagctgg
tatcagagggc acgtcaacat cttaaaagatg gcacttgtg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccctatg tgttcatcaa acgttcggat gctogaaact cactctatg tcatgttatg tagcagaact cgaaggcaatt cagtaacggtc gtagtggtga
gacacttggt gtcccttgctc ctcatgtggg cgaataacca gtggottacc gcaaggttct tottctgaag aacggtaata aaggagctgg tggccatagt taacggcgoc atctaaagtc atttgactta ggccagcago ttccattgaa gattttcaag aaaactgcaa cactaaacac

U realnosti

Do sada smo pretpostavljali da su očitavanja savršena i da ne sadrže greške. Ali, u realnosti...

attaaggtt	tataccttcc	caggttaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttta	aatctgtgtg	gctgtcaactc	ggctgcatgc	ttagtgcact	cacgcagtat	aattaataac	taattactgt	cgttgacagg	acaagagtaa	ctcgtctatc	ttctgaggc
tgcttacggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctagggtt	cgtccgggtg	tgacccgaaag	gtaagatgga	gagccttgtc	cctgggttcca	acgagaaaaac	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggcttttg	agactccgtg	gaggaggtct
tatcagaggc	acgtcaacat	cttaaaagatg	gcacttggtg	cttagtagaa	gttgaaaaag	gcgtttttgoc	tcaacttgaa	cagccctatg	tgttcatcaa	acgttcggat	gctcgaaactg	cacctcctgg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaatt	cagtacggtc	gtagtgggtga
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttcgttaag	aacggtaata	aaggagctgg	tggccatagt	tacggcgccg	atctaaagtc	atttgactta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaactgcaa	cactaaacac

U realnosti

- Očitavanja **sadrž**e greške

attaaaggtt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcatgc	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacgagtaa	ctcgtctatc	ttctgcaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcatacgcac	atctaggttt	cgtccgggtg	tgacccgaaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaaac	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggtcttgg	agactccgtg	gaggaggctc
tatcagaggo	acgtcaacat	cttaaaagatg	gcacttgtgg	cttagtagaa	gttgaaaaag	gogttttgoc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gtctgaactg	cacctcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggta
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggtaata	aaggagctgg	tggccatagt	tacggcgccg	atctaaaagtc	atttgaacta	ggcgacgagc	ttggcaactga	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacca

U realnosti

- Očitavanja **sadrže** greške

atgccgtatggacaacgact
atgccgtatg
gccgtatgga
gtatggacaa
gacaacgact

atgccgtatggacaacgact
atgcc
tgccg
gccgt
ccgta
cgtat
gtatg
tatgg
atgga
tggac
ggaca
gacaa
acaac
caacg
aacga
acgac
cgact

attaaagggt tataccttcc caggtaacaa accaaccac tttagatctc ttgtagatct gttctctaaa cgaactttaa aatctgtgtg gctgtcactc gctgtcactc ttagtgcaact caccgagtat aattaataac taattactgt cgttgacagg acacgagtaa ctgctctatc ttctgcaggc
tgcttaacgt ttgctcgtg ttgcagccga tcatcagcac atctaggttt cgtccgggtg tgacgcgaaag gagccttgtc cctggtttca acgagaaaaac acacgtccaa ctacgttttg cttgttttca ggttcgcgac gttgcgtgac gtgctcgtac taggctttg agactccgtg gaggaggctc
tatcagaggo acgtcaacat cttaaagatg gcacttgtyg cttagtagaa gttgaaaaag gogttttgoc tcaacttgaa cagccctatg tgttccatca acgttcggat gctogaactg caacctatgg tcatgttatg ttgcagaaact cagtaacggtc gtagtggtga
gacacttggt gtctttgtcc ctcatgtggg cgaataacca gtggtttacc gcaaggttct tctttgtaag aacggttaata aaggagctgg tggccatagt tacggcgccg atctaaagtc atttgactta ggccagcagc ttggcaactga tctttatgaa gattttcaag aaaaactgaa cactaaacac

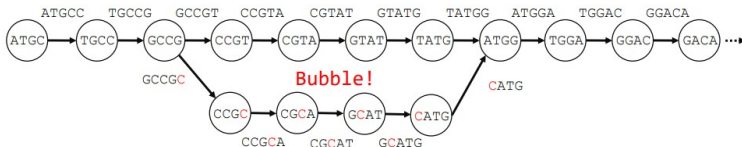
U realnosti

- Očitavanja **sadrž**e greške
- Rastojanja između očitavanja u parovima **nisu tačna**

attaaaggtt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gttctctaaa	cgaactttaa	aatctgtgtg	gotgtcaact	ggctgcactg	ttagtgcact	cacgcagtat	aatttaatac	taattactgt	cgttgacagg	acacgagtaa	ctcgtctatc	ttctgcaggc
tgcttacggt	ttcgtccgtg	ttgcagccga	tcctcagcac	atctaggttt	cgtccgggtg	tgacgcgaaag	gtaagatgga	gagccttgtc	cctggtttca	acgagaaaaac	acaagtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgctcgtac	gtggtcttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaagatg	gcacttgctg	cttagtagaa	gttgaaaaag	gcgtttttgc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctogaactg	cacctcactg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaaggct	gtagtgggta
gacacttggt	gtccttgctc	ctcatgtggg	cgaataacca	gtggcttacc	gcaaggttct	tcttcgttaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgccg	atctaaagtgc	atttgaacta	ggcgcagcgc	ttggcaactga	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacaa

U realnosti

- Očitavanja **sadrže** greške
- Rastojanja između očitavanja u parovima **nisu tačna**



attaaagggtt tataccttcc caggtaacaa accaaccacc tttagatctc tttagatctc gttctctaaa cgaactttaa aatctgtgtg gctgtcaact gctgtcaact gctgtcaact ttagtgcaat cagcagtat aattaataac taattactgt cgttgacagg acacagatga ctgtctatcc ttctgacagg
tgcttaacgt ttgttcogtg ttgcagccga tcatcagcac atctaggttt cgttcoggtg tgacgcgaag gtaagatgga gagcctgtgc cctgtgttca acagataaac atcagtttgc ctcagtttgc cttgttcacg ggttcogac gttgcogtac gttgcogtac agactcogtg gaggaggtct
tatcagaggo acgtcaacat cttaaagatg gcaactgttg cttagtagaa ggtgaaaaag gogttttgoc taaacttgaa cagcctatg ttttcaatca acgttcoggt gctogaactg cacttcactg tcatgttatg gttgagctgg tagcagaact cgaaggcact gattttcaag aaaaactgaa cactaaacac
gacacttggt gtccttgtcc ctcatgtggg cgaataacca gtggottacc gcaaggttct tctttgttaag aacggttaata aaggagctgg tggccatagt taaggcgcgc atctaaagtc atttgactta ggccagcago ttggcactga tctttatgaa

U realnosti

- Očitavanja **sadrže** greške
- Rastojanja između očitavanja u parovima **nisu tačna**
- **Nesavršena** pokrivenost genoma očitavanjima

attaaagggt	tataccttcc	caggtaacaa	acaaacaaac	tttogatctc	ttgtagatct	gtctctataa	cgaactttta	aatctgtgtg	gctgtcaact	ggctgcactg	ttagtgcact	cacgcagtat	aattataaac	taattactgt	cgttgacagg	acacgagtaa	ctcgtctatc	ttctgcaggc
tgcttaacgt	ttcgtccgtg	ttgcagccga	tcatacagac	attcaggttt	cgtccgggtg	tgacgcgaaag	gtaagatgga	gagccttgtc	cctggtttca	acgcagaaaa	acacgtccaa	ctcagtttgc	ctgtttttaca	ggttcgcgac	gtgcctgtac	gtggttttgg	agactccgtg	gaggagggtc
tatcagaggo	acgtcaacat	cttaaagatg	gcacttgctg	cttagtagaa	gttgaaaaag	gcgtttttgc	tcaacttgaa	cagccctatg	tgttcaatca	acgttcggat	gctogaactg	caacctcatg	tcactgttatg	gttgagctgg	tagcagaact	cgaaggcaat	cagtaacggtc	gtagtgggtg
gacacttggt	gtccttgtcc	ctcatgtggg	cgaataacca	gtggottacc	gcaaggttct	tcttctgtaag	aacggttaata	aaggagctgg	tggccatagt	taaggcgcgc	atctaaagtgc	atttgaactta	ggcgcagcgc	ttggcaactga	tccttatgaa	gattttcaag	aaaaactgaa	cactaaacaa

Hvala na pažnji!

Reference

- Phillip Compeau, Pavel Pevzner, Bioinformatics Algorithms 2nd Edition, 2015.
- Bioinformatics Algorithms video predavanja
- Predavanja iz kursa Uvod u bioinformatiku, prof. dr Jovana Kovačević - MATF

attaaagggtt tatacctttcc caggtaacaa accaaccacac tttagatctc tttagatctc gttctctaaa cgaactttaa aatctgtgtg gctgtcactc ggctgcctgc ttagtgcact cagcagttat aattaataac taattactgt cgttgacagg acacagagtaa ctgctctatc ttctgcaggc
tgcttaacggt ttgctcogtg ttgcagccga tcatcagcac atctagggtt cgtccgggtg tgacgcgaag gtaagatgga gagccttgtc cctggtttca acagagaaaa acacgtccaa ctacgctttg ctgcttttaca ggttcgcgac gtgctcgtac gtggttttg agactcogtg gaggaggtct
tatcagaggo acgtcaacat cttaaagatg gcactttgtg cttagtagaa gttgaaaaag tcaacttgaa cagccctatg tgttcaatca acgttcggat gctogaactg caacctcatg tcatgttatg gttgagctgg tagcagaact cgaaggcatt cagtaagggtc gtagtggtga
gacacttggt gtctttgtcc ctcatgtggg cgaataacca gtggtttacc gcaaggttct tcttctgaag aacggttaata aaggagctgg tggccatagt taaggcgcgg atctaaagtc atttgactta ggccagcago ttggcaactga tctttatgaa gattttcaag aaaaactgaa caactaaaca