

Univerza v Ljubljani
Fakulteta za računalništvo
in informatiko



Sintetični biološki preklopni sistemi kot procesna platforma prihodnosti - motivacijski del

prof.dr.Miha Mraz

<http://lrss.fri.uni-lj.si/bio/index.html>

21. november
2016



Motivi za povezovanje računalništva in sintezne biologije

- **Kartiranje** in **anotiranje** človeškega genoma (C.Venter, 2001)
- **DNA osnove : računalniške osnove**
- **Adlemanov eksperiment** – reševanje NP polnih problemov v DNA materialu (L.Adleman, 1994)
- Padanje cene **oslikave genoma** (junij 2013, The Guardian: 5.000 USD)
- **iGEM** (International Genetically Engineered Machines) fundacija



1. Kartiranje in anotiranje genoma – C. Venter

- Na prelomu med 20. in 21. stoletjem C.Venter objavi „karto“ človeškega genoma (2001)
- Knjiga: C.Venter – Genom mojega življenja
- Veliko pomoč pri kartiranju (določanju zapisa) nudi računalništvo z vse bolj zmogljivimi postopki analize velikih količin podatkov



2. DNA osnove : računalniške osnove

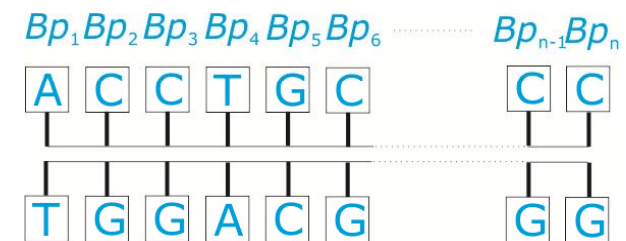
- Temelj: molekula deoksiribonukleinske kisline (DNA), prelep dveh molekul v dvojno vijačnico
- Hrani gensko zasnovo vsakega živega organizma (angl. *blue print*)
- Vsaka vijačnica sestavljena iz poljubnega dolgega niza, ki ga tvorijo Adenin, Gvanin, Citozin, Timin (baze ali nukleotidi A, C, G, T)
- 1.podobnost: **Obstoj kodnega sistema**
 - Narava: {A,C,G,T} – kvaternarno kodiranje
 - Računalništvo: {0,1} – binarno kodiranje



- Gostota zapisa (razmaki med posameznimi bazami so približno 0,35 nm):
 - DNA:
 - 18 Mbit/inch (inch – dolžina niza)
 - 10^6 Gbit/inch² (zvijanje niza na površini)
 - Trenutne tehnologije trdih magnetnih diskov:
 - 10 Gbitov/inch²
- 2.primjerjava: **Izredno velika gostota zapisa – možnost hrambe – trajni pomnilnik**
- DNA se danes sintetizira po naročilu, a cena je zaenkrat še visoka (okvirna cena nekaj 100 USD na cca 10^6 integriranih baz (dolžina niza))



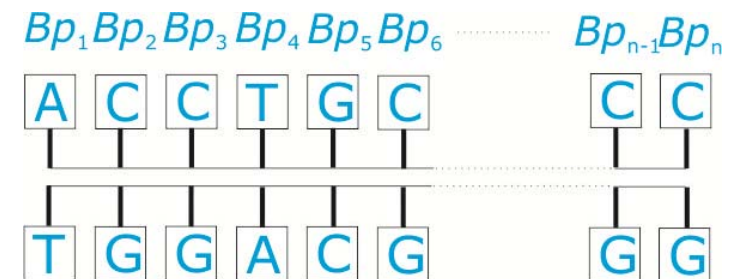
- Če imamo na eni od vijačnic bazo X, leži na istoležni poziciji na drugi vijačnici njej komplementarna baza Y (par imenujemo za bazni par, *bp*)
- Komplementarna para (Watson-Crick komplement:
 - A-T (T-A), C-G (G-C)
- Primer:
 - Vijačnica1: ATTACGTCG
 - Vijačnica2: TAATGCAGC
- 3.primerja: **Redundanca zapisa** (možnost detekcije napak???)

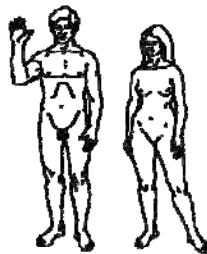


Sintetični biološki preklopni sistemi
kot procesna platforma prihodnosti
- motivacijski del



- Vsak bazni par ima indeks svoje lokacije
- 4.primjerjava: **Linearnost naslavljanja pomnilnika (obstoj naslovnega prostora)**
- Okvara niza (koda): bodisi napako naredijo že pri tvorbi encimi ($p < 10^{-9}$), ali pa do napake pride zaradi vpliva zunanjega okolja (UV svetloba, T, itd.)





Homo Sapiens,
 $3.1 \cdot 10^9$ BP.



Mus, $3.1 \cdot 10^9$ BP.



Drosophila melanogaster, $139 \cdot 10^6$ BP.



Escherichia coli, $5 \cdot 10^6$ BP.

Sintetični biološki preklonni sistemi
kot procesna platforma prihodnosti
- motivacijski del



- Genom: celotni dedni zapis živega organizma (v kontekstu naših potreb ga enačimo s sekvenco baznih parov)
- Genom: sestavljen iz množice genov (podzaporedij baznih parov)
- Gen = nosilec posamezne značilnosti organizma (človeški genom sestavljen iz približno 25.000 genov)
- Primeri genov: barva las, dedna podvrženost neki bolezni (npr. celiakija), itd.
- 5.primjerjava: **Gen = program, Genom = množica programov**
- **Intermezzo: patentiranje genoma ali posameznega gena??** – slabe plati patentiranja se prenašajo tudi na to področje (patentiranje funkcij programske opreme, patentiranje brezžičnega kuhanja, itd.)



- Aktualni vedi področja DNA procesiranja:
 - Sistemska biologija
 - Sintezna biologija
- Obe vedi za razvoj na svojem področju nujno potrebujeta pomoč s strani računalništva (očitno imamo sedaj najustreznejše sosede)
- Ostale vede, prisotne na tem področju:
 - Kemijska tehnologija, Biotehnologija, Farmacija, Fizika, Medicina (skratka vede o življenju)



- **Sistemska biologija (po Gruden, Baebler (NIB)):**
 - Interdisciplinarna veda, ki omogoča razumevanje dinamike procesov v bioloških sistemih
 - Cilj: razumevanje delovanja obstoječih organizmov
 - Opazovanje in spoznanja na različnih nivojih gledano od spodaj navzgor (geni, proteini, metabolne signalne poti, subcelični mehanizmi, celice, tkiva, organi, organizmi, interakcije med enakimi in različnimi organizmi)
 - Primer področja: funkcionalna genomika
 - Pristop: mehanistični

Sintetični biološki preklonni sistemi
kot procesna platforma prihodnosti
- motivacijski del



- Sintezna biologija:
 - Veda, ki v biologijo uvaja inženirske postopke, ki celico spremenijo tako, da novonastala celica opravlja za človeka koristno funkcijo, ki jo predhodno ni opravljala
 - Zgodovina: genetsko inženirstvo (križanje sort, itd.)
 - Koncept: „bottom-up“
- Podpora obem biologijam s strani računalništva:
 - Bioinformatika (analiza velikih količin podatkov, kartiranje genoma, iskanje relacij med podsekvencami genov, itd.)
 - Računska biologija (angl. *computational biology*): računske metode kot pomoč snovanju novih organizmov (<http://lrss.fri.uni-lj.si/bio/>)



- Kja sta danes obe biologiji:
 - Na začetku
 - Napredek odvisen od razvoja tehnologij obvladovanja bioloških procesov
 - Kje se pričakuje napredek: lepljenje, rezanje popravljjanje, sintetiziranje genoma, izboljšanje infrastrukture

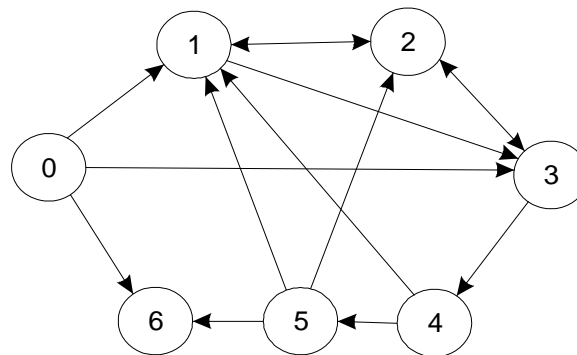


3. Adlemanov eksperiment

- 1994: potrditev ideje (angl. *proof of the concept*), da je mogoče z visokim paralelizmom procesiranja nad DNK učinkovito reševati NP-polne probleme
- Primer reševanega problema:
 - Hamiltonove poti po usmerjenem grafu
 - Imamo n vozlišč, med katerimi vodi poljubno število eno ali dvosmernih povezav;
 - Problem: ali je ob danem izhodiščnem in ciljnem vozlišču možno prepotovati vsa vozlišča natanko enkrat;
- Ni nujno, da rešitev obstaja, če pa obstaja, kompleksnost njenega iskanja strmo raste z višanjem števila mest



- Konkretni problem ob robnih pogojih, ki ga rešuje Adleman, predvideva, da se pot začne v 0 in konča v 6; rešitev predstavlja sekvenca obiskov 0, 1, 2, 3, 4, 5, 6;





- Ideja za reševanje problema: prostorska potratnost nadomesti procesno;
- Psevdo algoritem:
 - Vhod: usmerjeni graf G z n vozlišči in izvornim ter ponornim vozliščem
 - 1.korak: generiraj potencialne poti v grafu G naključno in v velikih količinah
 - 2.korak: zavrži vse poti, ki nimajo ustreznega ponora in izvora
 - 3.korak: zavrži vse poti, ki ne obišejo natanko n mest
 - 4.korak: ponavljaj za vsako od n vozlišč: preglej vse generirane poti in posamezno pot zavrži, če ne vsebuje trenutno opazovanega vozlišča;
 - Izhod: če je v množici poti še kakšna "ustrezna" odgovori pozitivno, drugače negativno – problem nima rešitve;



- Hipoteza:
 - obstoj platforme, ki za posamezni korak algoritma potrebuje le en “procesni korak” dolžina katerega je linearno odvisna od n ;
 - Koraki 1, 2, 3: enkratni koraki (konstantna odvisnost)
 - Korak 4: linearno odvisen od števila n
- Veljavnost omenjene hipoteze omogočata nedeterminizem in visok paralelizem procesiranja v mediju DNA



- Adlemanov pristop k kodiranju:
 - posamezno od 7 vozlišč zakodiramo z 20 baznimi pari (A,C,T,G)
 - kodiranje naključno (z dolžino zagotovljena dovoljšnja "razdalja" kodov)
- zgledi kodiranja vozlišč (niz 10+10):
 - s2=TATCGGATCG GTATATCCGA
 - s3=GCTATTCGAG CTTAAAGCTA
 - s4=GGCTAGGTAC CAGCATGCTT



- Cilj iskanja: možne poti po grafu
- Kodiranje poti:
 - 20 kodni niz = [izvor[11..20], ponor[1-10]]
 - Koda poti: $\text{neg}(\text{stik}(\text{izvor}[11-20], \text{ponor}[1-10]))$
- Dolžina niza kode prehoda kot dela poti: 20 nukleotidnih baz = 20 znakov
- Niz prehoda negiramo na osnovi že znanega Watson-Crickovega komplementa



- Tehnologija izvedbe psevd algoritma:
 - 1.faza: Sinteza ob prisotnosti encimov (ligaz)
 - Posledica: "negacije" – prehodi na 2.vijačnici
 - 2.faza: PCR (angl. *Polymerase Chain Reaction*) – izločanje verig z neustreznim izvorom in ponorom
 - 3.faza: GE (angl. *Gel Electrophoresis*) – izločanje neustrezno dolgih verig
- Slabosti eksperimenta:
 - velika poraba materije (200 vozlišč – $3 \cdot 10^{25}$ kg DNK materije – teža planeta???)
 - stohastičnost procesa (veliko število iteracij ???)



4. Oslikava genoma

- Junij 2013, The Guardian: 5.000 USD
- http://en.wikipedia.org/wiki/DNA_sequencing:
 - Do 1 USD na 10^6 bp -> cca 3.000 USD na kompletno oslikavo (cene nekaterih komercialnih ponudnikov so dosti manjše)
 - Potrebno je biti pozoren tudi na faktor zanesljivosti (angl. *accuracy*) oslikave
 - Oslikovanje (identifikacija) posameznih genov dosti cenejša zaradi parcialnosti oslikav in sestavljanja dokončnega vzorca;



5. iGEM tekmovanje

- iGEM – www.igem.org
- Vsakoletno tekmovanje študentskih ekip iz sintezne biologije
- Najprej kontinentalne kvalifikacije, nato svetovni finale v Bostonu na MIT
- Posamezna ekipa nastopi s svojim sinteznim projektom in se ocenjuje na osnovi naslednjih kriterijev:
 - Ocena projekta (projekt je potrebno teden pred tekmovanjem naložiti na ustrezno Wiki stran)
 - Zagovor projekta
 - Posterska predstavitev
 - Ocena Wiki strani



- Rezultati projektov (posamezni novi geni, itd.) se shranjujejo v javno dostopni **repozitorij genskih bioloških gradnikov – BioBrick Registry of Standard Biological parts** (za nas gradniki za izvajanje logičnih funkcij, oscilatorji, pomnilni elementi, itd.)
- Repozitorij si bomo ogledali na koncu poglavja o DNA procesiranju



- Ekipe UL (običajno 10 študentov je v preteklosti vodil prof.dr.Roman Jerala, Kemijski inštitut:
 - Ekipa UL 2010 1. v svetovnem merilu (od 10 študentov štirje člani FRI)
 - Ekipa UL 2012 2. v svetovnem merilu (od 10 študentov trije člani FRI): najboljši Wiki, najboljši medicinski projekt, najboljši modeli - <http://2012.igem.org/Jamborees>
- Naloge članov FRI:
 - Tehnični del: Wiki
 - Znanstveni del: Računalniško modeliranje dinamike ekspresije gena -> napovedovanje odzivanja genskega materiala – dinamike v celici
 - Fizično delo v laboratoriju;
 - Trajanje projekta: od januarja, do konca novembra



- Ekipe leta 2012 in njihovi projekti:
http://igem.org/Team_Wikis?year=2012



6. Literatura motivacijskega dela

- M.Moškon, M.Mraz: Modeling as the essential step in construction of biological computer structures (glej spletno učilnico)
- L.M.Adleman: Computing with DNA, Scientific American, August 1998 (glej spletno učilnico).