

Bioinformatika a DNA

Kolek Jan · Informačné technológie, Medicína

29.08.2012



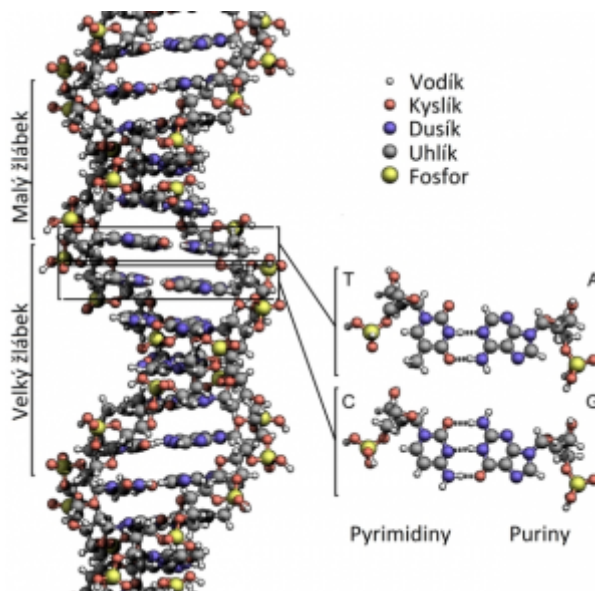
Informatika je viední disciplínou, ktorá v dnešnej dobe zažíva obrovský rozvoj. Je dnes používaná takmer vo všetkých oblastiach ľudskej činnosti. Napríklad vo vzdelávaní, bankovníctve, kriminalistike, priemysle a mnohá, opravdu mnohá ďalších. Informatika ale preniká i do mnoha ďalších viedních odborů, kde bychom to nečakali. Jedným z týchto odborů je klasická

biologie. Spojením biologie a informatiky vznikl nový viední obor - bioinformatika.

Bioinformatika sa zaoberá navrhovaním metód pre shromažďovanie, analýzu a vizualizáciu rozsáhlých súborů predvážne biologických dát. Bioinformatika úzko súvisí s tzv. genomikou, čo je jedna z častí genetiky, ktorá sa zaoberá predovšetkým získavaním sekvencií DNA rôznych organizmov a práce s nimi (výpočetná genomika). Trendom poslednej doby je navyše použitie tzv. umelého intelligenca (súčasť informatiky), ktorá patrí v dnešnej dobe medzi najviac sa rozvíjajúce oblasti ľudskej činnosti. Snaha o to, aby počítače, ale i ďalšie stroje dokázali nahradiť človeka tam, kde je potreba „inteligentného prístupu“ vedie k tvorbe algoritmov, metód a postupů, ktoré majú obsahovať „umelú intelligenca“.

1. DNA

Deoxyribonukleová kyselina - DNA je v dnešnej dobe veľmi často používaným slovným spojením vo viednom svete i bežnom živote. Jej využitie napríklad pre určenie otcovstva v biomedicíne, pre výskum pôvodu nemoci v lekárstve, pre osvedčenie pachateľů trestných činů v kriminalistike, pri šľachtení nových plodín v poľnohospodárstve, alebo pre určovanie pôvodu človeka v antropológii a tak tiež v mnoha ďalších oblastiach ľudskej činnosti je už známe a používané dlhšiu dobu. Taká presnosť a nediskutovateľnosť výsledkov analýzy a spracovania DNA ju stavia do popredia viedneho zájmu. Nie je preto žiadnym divom, že sú stále skúmané nové a nové metódy pre jej dokonalejšiu a presnejšiu analýzu, separáciu, izoláciu, farbenie, sekvenovanie, porovnávanie i pre jej umelú syntézu.



Obr. 1.: Dvoušroubovice DNA

DNA je nositelkou základní genetické informace všech živých organismů na Zemi a je uložena v jádrech buněk téměř všech buněčných organismů. Výjimkou jsou buňky, které nemají jádro, a tam je DNA uložena volně přímo v cytoplasmě. DNA je v podstatě stavebním návodem, nebo programem každého živého organismu na Zemi uloženého v jeho buňkách a předurčuje tak další vývoj tohoto organismu a vymezuje jeho základní vlastnosti.

1.1. Historie DNA

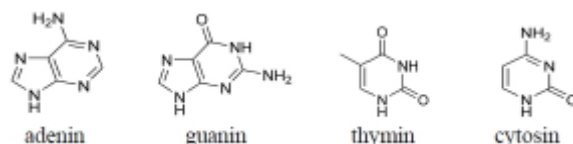
Deoxyribonukleová kyselina (DNA) byla objevena a popsána již v roce 1869 švýcarským lékařem Friedrichem Miescherem, který zkoumal hnis na nemocničních obvazech a objevil látky, které nazval nuklein. V roce 1919 americký biochemik Phoebus Levene popsal její složení z fosfátových skupin, cukrů a nukleových bází. Ale stále nebylo známo nic o její funkci až do roku 1943, kdy Colin MacLeod, Maclyn McCarty a Oswald Avery provedli test, který prokázal, že deoxyribonukleová kyselina je nositelkou genetické informace všech buněk. Nejslavnějším počinem ve výzkumu DNA však bylo objevení její trojrozměrné struktury. Dvoušroubovici, tak jak ji známe dnes představili v roce 1953 v časopise Nature britský molekulový biolog Francis Crick a americký biolog James Dewey Watson na základě rentgenové difrakční analýzy provedené v roce 1952 Rosalindou Franklinovou a Raymondem Goslingem.

V roce 1957 popsal Francis Crick vztahy DNA, RNA a proteiny dnes nazývané jako „centrální dogma molekulární biologie“. Dalším krokem byl objev replikace DNA v buňkách dvěma američany, molekulárním biologem Matthewem Meselsonem a genetikem Franklinem Stahlem. Zatím posledním významnějším krokem v oblasti výzkumu DNA bylo rozluštění genetického kódu americkými biochemiky Robertem Williamem Halleyem, Marshalllem Warrenem Nirenbergem a Harem Gobind Horanem.

1.2. Stavba DNA

DNA je biologická makromolekula a skládá se z fosfátové skupiny, cukru, deoxyribózy a jedné ze čtyř nukleových bází, souhrnně nazývaných nukleidy. Nositelem genetické informace jsou především tyto čtyři nukleové báze dělící se na puriny: adenin (A) a

guanin (G) a pyrimidiny: cytosin (C) a thimyn (T).

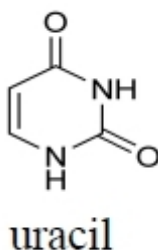


Obr. 2.: Nukleové báze DNA

Tyto čtyři nukleové báze ve velkém počtu a pořadí tvoří různé řetězce vypadající jako šroubovice a jejich uspořádání je unikátní pro každého člověka na Zemi. Navíc dochází ke spojování dvou šroubovic k sobě, kde se pomocí dvou vodíkových vazeb (můstků) k sobě vážou adenin (A) z jedné šroubovice a thimyn (T) ze druhé šroubovice, nebo pomocí tří vodíkových vazeb guanin (G) z jedné šroubovice a cytosin (C) ze druhé šroubovice a vzniká tak hodně známá dvoušroubovice DNA.

1.3. RNA

Ribonukleová kyselina (RNA) slouží ke zkopírování genetické informace z DNA a přenos na místo, kde dochází k její translaci na výsledný protein. Může také plnit funkci katalyzátoru. Její složení je obdobné jako u DNA pouze s tím rozdílem, že v RNA jsou obsaženy hydroxylové skupiny a tím, že neobsahuje thimyn (T), ale ten je nahrazen uracilem (U). Vědci tuto rozdílnost vysvětlují tím, že uracil (U) je energeticky méně náročný. Také se od DNA liší v tom, že tvoří pouze jednoduchá kratší vlákna (dvoušroubovice pouze velmi výjimečně).



Obr. 3.: Uracil (U)

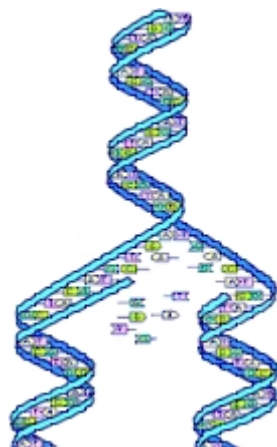
2. LIDSKÝ GENOM

Lidský genom nese veškerou genetickou informaci obsaženou v DNA. Je to vlastně popis toho, jak se má vlastní organismus vyvíjet a určuje základní i sekundární rysy organismu. Jednotky obsažené v genomu popisující jednotlivé rysy jsou nazývány geny. Snaha o rozluštění lidského genomu je vlastně snahou rozluštit funkce jednotlivých genů. Tedy určit, kterou vlastnost určuje který gen. S tím souvisí i pořadí jednotlivých nukleotidů a dalších složek samotné DNA.

Studium a porovnávání genomu má význam v mnoha ohledech, např. je možné určit postupný vývoj a evoluci člověka. Další možností je např. sledování závislosti výskytu dědičných nemocí na některých úsecích DNA. Vzhledem k „nekonečné“ délce DNA je právě tato oblast pro ruční práci nevhodná. Dochází proto k nasazení počítačů, algoritmů pro vyhodnocování lidského genomu a využití dalších možností Bioinformatiky.

2.1. Analýza sekvencí DNA

Při analýze sekvencí DNA se věda zaměřuje na mnoho aspektů a vlastností, které DNA nese. Patří mezi ně genetická podstata konkrétních proteinů, dále mutace, polymorfismy, podobnost DNA mezi druhy organismů, koncentraci, čistotu, chemické vazby na proteiny a mnoho dalších vlastností. Existuje několik způsobů analýzy DNA, ale nejpoužívanější je metoda navržená americkým vědcem Frederikem Sangerem z Cambridge.



Obr. 4.: Replikace DNA

Jeho metoda spočívá v rozdělení fragmentů DNA na samostatné šroubovice a ve vystavení působení volných nukleotidů na obě šroubovice. Tímto způsobem dochází k prodlužování obou samostatných šroubovic. Každý terminující nukleotid je označen přidáním atomu, jenž při ozáření laserem svítí. Každý ze čtyř nukleotidů svítí specifickým způsobem, takže je možno rozpoznat je podle toho. Následně dojde působením elektrického napětí k separaci jednotlivých fragmentů DNA podle jejich délky. Podle barvy koncového nukleotidu se určí, o který jde. V dalším cyklu se DNA rozštěpí na jiném místě a pokračuje se dál stejným způsobem.

2.2. Porovnávání sekvencí DNA

Porovnávání sekvencí DNA spočívá v porovnání shody mezi jednotlivými prvky každé sekvence. Základním předpokladem pro porovnání sekvencí je zajištění optimálního zarovnání porovnávaných sekvencí. Pro zarovnání dvou sekvencí před porovnáváním se zavádí tzv. skórovací funkce. Při evoluci organismu dochází u každého druhu ke třem různým změnám - mutaci, inserci a deleci.

1. Mutace - záměna jednotlivých nukleotidů.
2. Inzerce - vložení jednoho nebo více nových nukleotidů do původní struktury.
3. Delece - odebrání jednoho či více nukleotidů.

U DNA sekvence bez insercí a delecí je porovnání realizováno pouhým srovnáním a vyhodnocením shody, popřípadě neshody porovnávaných řetězců. Pokud ovšem nastane inserce, nebo delece je nutno zavést pojem mezera, která tyto změny v sekvenci DNA zohledňuje. Během vyhodnocování sekvencí, které obsahují mezery, je třeba do skórovací funkce přidat tzv. hodnotu penalizace za zarovnání znaku jedné sekvence s mezerou druhé sekvence. Z důvodu různé pravděpodobnosti výskytu substituce znaků je nutné, aby velikost penalizace této pravděpodobnosti odpovídala.

Následující obrázek ukazuje některé z možností zarovnání sekvencí AATCTATA a AAGATA.

```

AATCTATA      AATCTATA      AATCTATA
AAG-AT-A      AA-G-ATA      AA--GATA

```

Obr. 5.: Ukázka zarovnání sekvencí DNA

2.3. Porovnávání sekvencí DNA pomocí algoritmu BLAST

Algoritmus BLAST patří mezi představitele algoritmů pro porovnávání sekvencí založené na technice heuristické analýzy. Jeho princip spočívá ve vyřazení nevhodných sekvencí ihned po začátku porovnávání. V tuto dobu se snaží algoritmus co nejdříve lokalizovat podobné úseky sekvence bez mezer mezi analyzovanou sekvencí a sekvencemi z existující databáze. Algoritmus obsahuje tři kroky - předzpracování dotazu, generování zásahů a prodloužení zásahů.

1. Předzpracování dotazu - mějme abecedu znaků sekvencí A. Mějme D, sekvenci databáze nad abecedou A. Mějme Q, sekvenci dotazů nad abecedou A. V prvním kroku jsou nejdříve vygenerována všechna slova nad abecedou A, která mají stanovenou délku W (parametr programu). Nato se postupně každé slovo s délkou W zarovnává s každým podřetězcem sekvence Q stejné délky (W). Dojde k asociaci každé pozice sekvence Q se seznamem slov, jenž při porovnání s podřetězcem sekvence Q získaly skóre vyšší jak stanovený práh T (parametr programu).
2. Generování zásahů - nyní je v Q obsažen seznam slov. Každá pozice Q je porovnávána s každým podřetězcem sekvence D, zdali není jedno ze slov na pozici v sekvenci, Q identické slovu z D. Pokud se tak stane, je zaznamenána shoda.
3. Prodloužení zásahů - každá shoda nalezená v předešlém kroku se teď rozšiřuje v obou směrech bez mezer s cílem určit, zda tato shoda může být částí většího úseku podobnosti. Toto rozlišování se zastaví, pokud skóre rozšířené shody klesne o větší míru, než je stanovena parametrem X, který je stanoven vůči nejvyšší dosažené hodnotě skóre daného zásahu. Každá takto rozšířená dvojice, jenž má skóre S (parametr programu), jenž již nelze zlepšit prodloužením je uchována a nazvána jako HSP (Hight scoring Segment pair). Sekvence HSP lze označit za nejvíce podobné úseky sekvenci Q.

3. VYUŽITÍ DNA V PRAXI

Jak už bylo popsáno výše DNA se dnes používá v mnoha oblastech běžného života, např.:

1. pro určení otcovství.
2. v genetické daktyloskopii.
3. v lékařství pro výzkum dědičných nemocí.
4. při šlechtění nových plodin.
5. v mnoha dalších oborech.



Obr. 6.: Rozdíly v DNA markerech 6 lidí

Poděkování

Tento článek vznikl za podpory Interní grantové agentury UTB ve Zlíně v rámci projektu číslo IGA/FAI/2012/041.

Literatura

1. Wikipedie [online]. 2012 [cit. 2012-08-06]. DNA. Dostupné z WWW: <http://cs.wikipedia.org/wiki/DNA>
2. MOTÝL, Ivo. Bioinformatika a kvantové zpracování informace. Zlín, 2009. 20 s. Semestrální práce. Univerzita Tomáše Bati ve Zlíně.
3. GURECKÝ, Libor. Studium stability genomické DNA rodu *Borrelia*. Brno, 2011. 49 s. Bakalářská práce. Masarykova Univerzita.
4. VACULKA, Tomáš. Identifikace osob pomocí DNA. Zlín, 2010. 52 s. Bakalářská práce. Univerzita Tomáše Bati ve Zlíně.
5. Wikipedia [online]. 2012 [cit. 2012-08-06]. DNA replication. Dostupné z WWW: http://en.wikipedia.org/wiki/DNA_replication