

Tablet – Next Generation Sequence Assembly Visualization

Jakub Šalagovič
IV105, FI MUNI, Brno

Sequence assembly

Zostavenie sekvencie

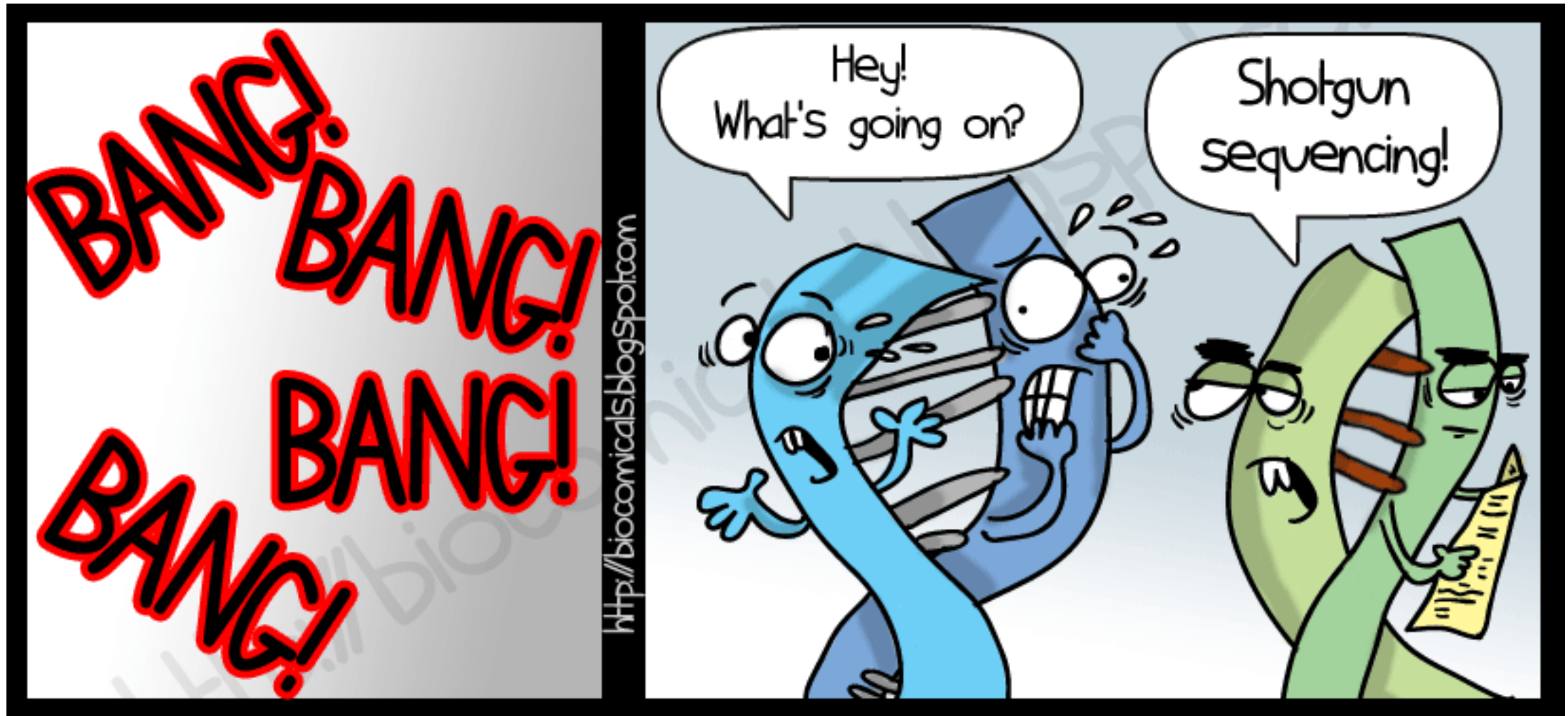
Sekvenovaním DNA nevieme prečítať celý genóm – len krátke úseky (reads)

Read – krátky úsek (rádovo 10^1 - 10^4 bp) získaný jedným prechodom

Úlohou je zostaviť sekvenciu na základe týchto krátkych úsekov

Metódy sekvenovania: Primer walking, Shotgun sequencing, Next-generation sequencing...

Shotgun sequencing



Shotgun sequencing



Shotgun sequencing



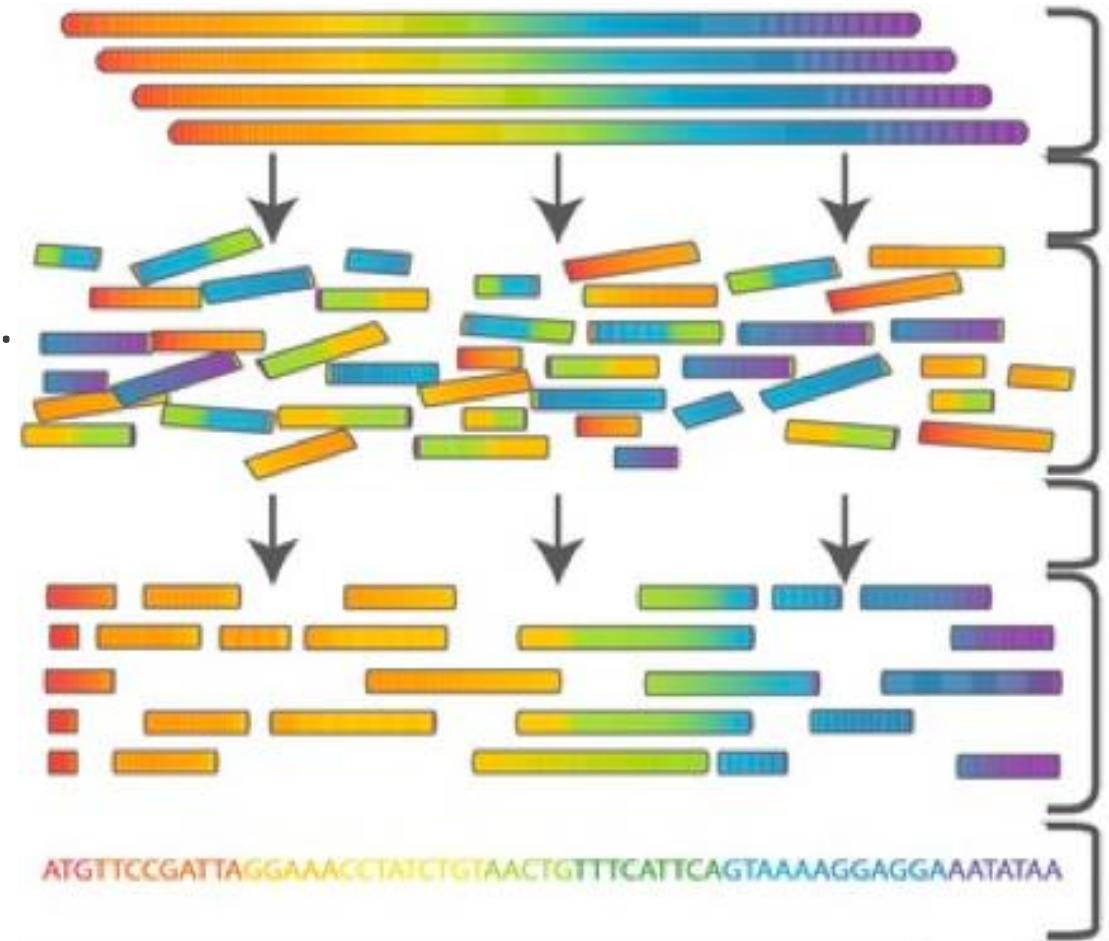
Shotgun sequencing

Pomenované na základe kvázi-náhodného vzoru rozsiahleho rozstrelu brokov

Proces: zväčšenie množstva klonovaním a rozdelenie DNA do veľkého počtu náhodných krátkych úsekov (restrikčné enz.). Následné paralelné sekvenovanie každého úseku zvlášť Sangerovou metódou.

Výsledkom je množstvo krátkych úsekov DNA

Na základe prekrytia zostavený reťazec nukleotidov



Next-generation sequencing

Viac prístupov: 454, Illumina (Solexa), SOLiD...

V princípe podobné ako shotgun sequencing

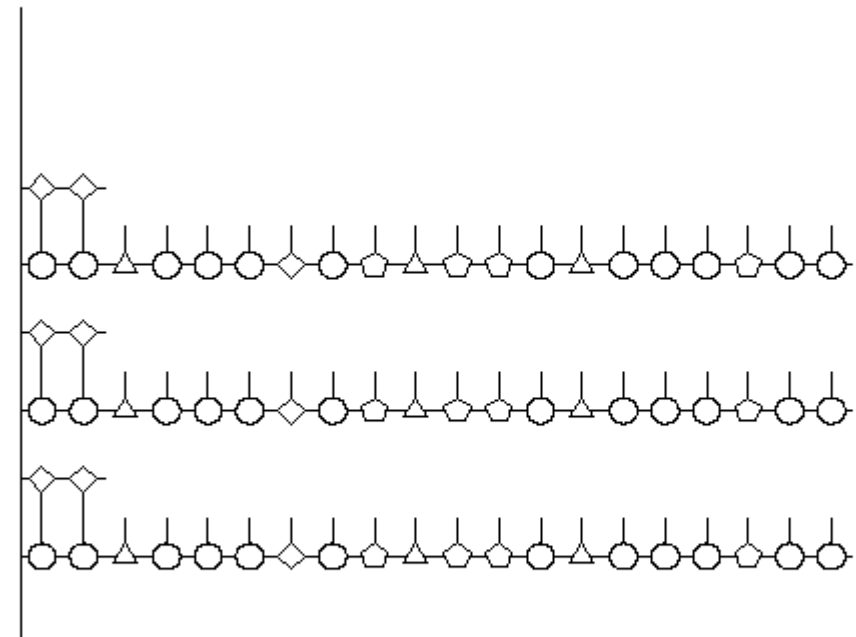
Rozmnoženie a rozdelenie na krátke úseky

Náhodne rozmiestnené na pevnú podložku

Postupné pripájanie fluorescenčných bází

Rádovo cca miliarda sekvencií zároveň

Relatívne rýchle a lacné

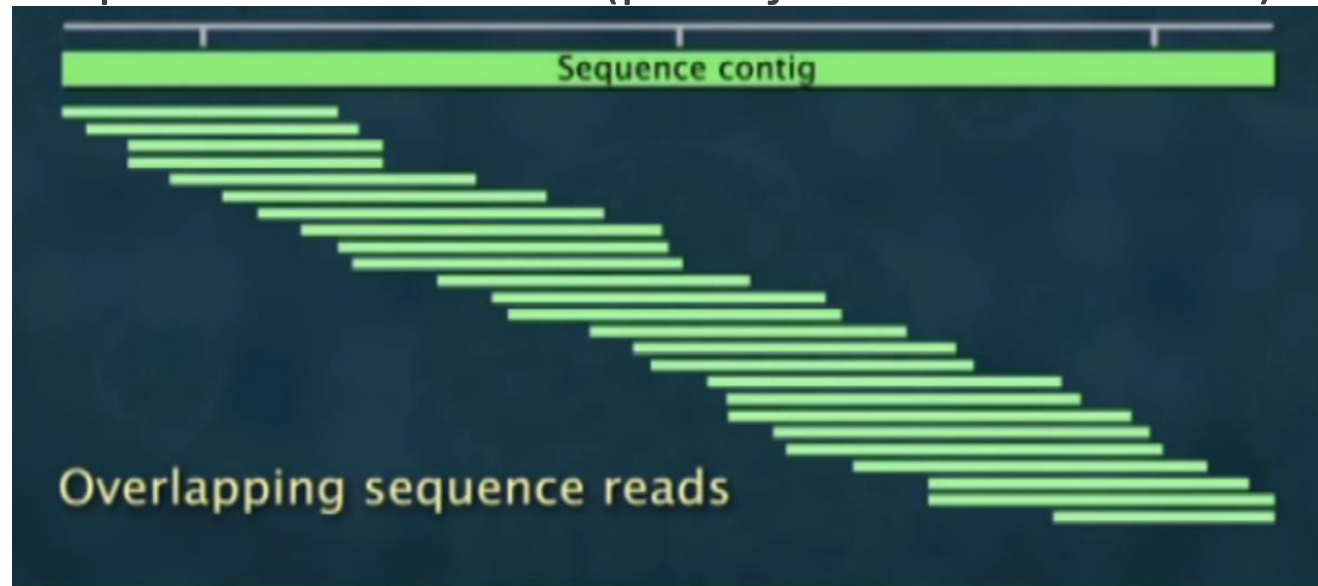


Sequence assembly

Množstvo krátkych úsekov -> súvislá sekvencia

Hlavne na základe prekrývajúcich sa úsekov

Výpočtovo veľmi náročné aj pri súčasnej technike – v súčasnosti trvá tento proces dlhšie ako práca v laboratóriu (pri najnovších metódach)



Zobrazenie získaných dát

Presun z veľkých centier do laboratórií

Potreba užívateľsky prívetivého prostredia

Plynulé zobrazenie s nižšími hardvérovými nárokmi

Podpora viacerých operačných systémov

Program podporujúci výstupy z viacerých typov sekvenovania

Tablet

The James Hutton Institute, Škótsko

„Lightweight, high-performance graphical viewer for next generation sequence assemblies and alignments.“

Java (1.6)

Podpora 32 aj 64 bit

Windows, Linux, OS X, prípadne Java Web Start

Tablet - výkon

Jednoduchá a rýchla inštalácia

122 MB na disku

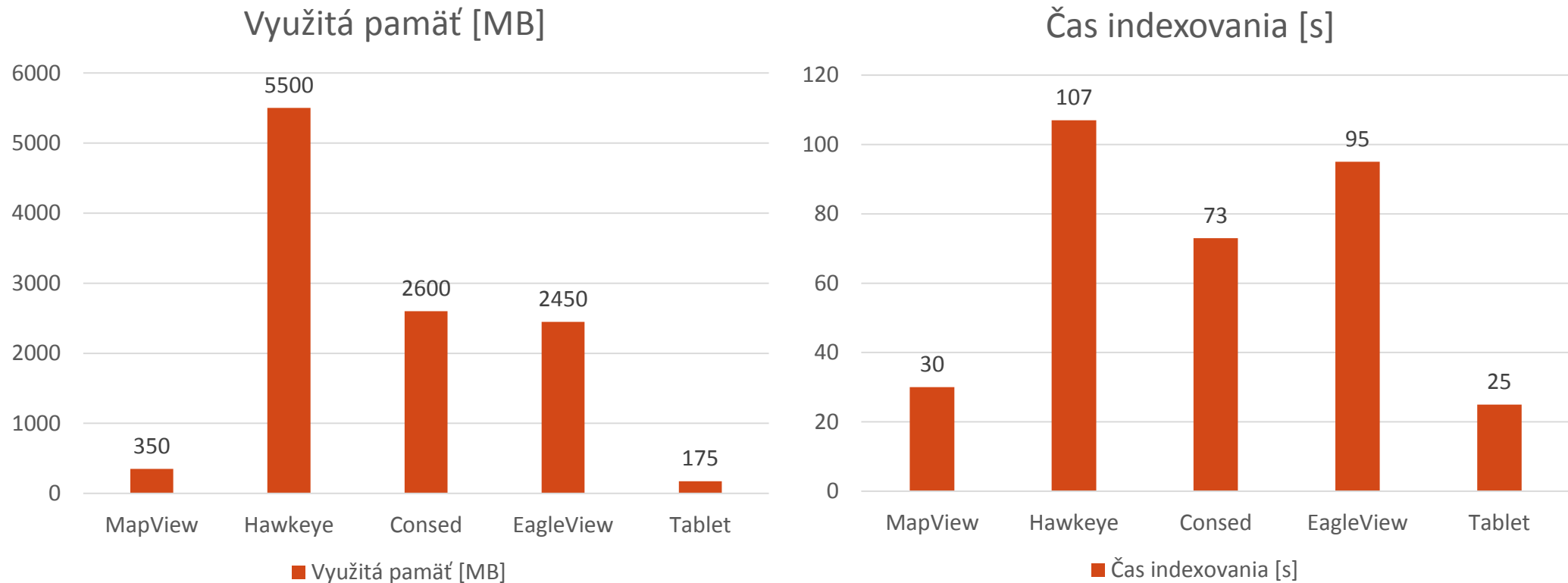
Využívanie viacerých jadier (ak dostupné)

Rozumná práca s operačnou pamäťou (zobrazovaný aktuálny stav)

Podporuje Sanger aj Next-generation sequencing (Illumina, 454,...)

Tablet – operačná pamäť

Cache based: MapView; Memory based: Hawkeye, Consed, Eagle View



Tablet – podporované formáty

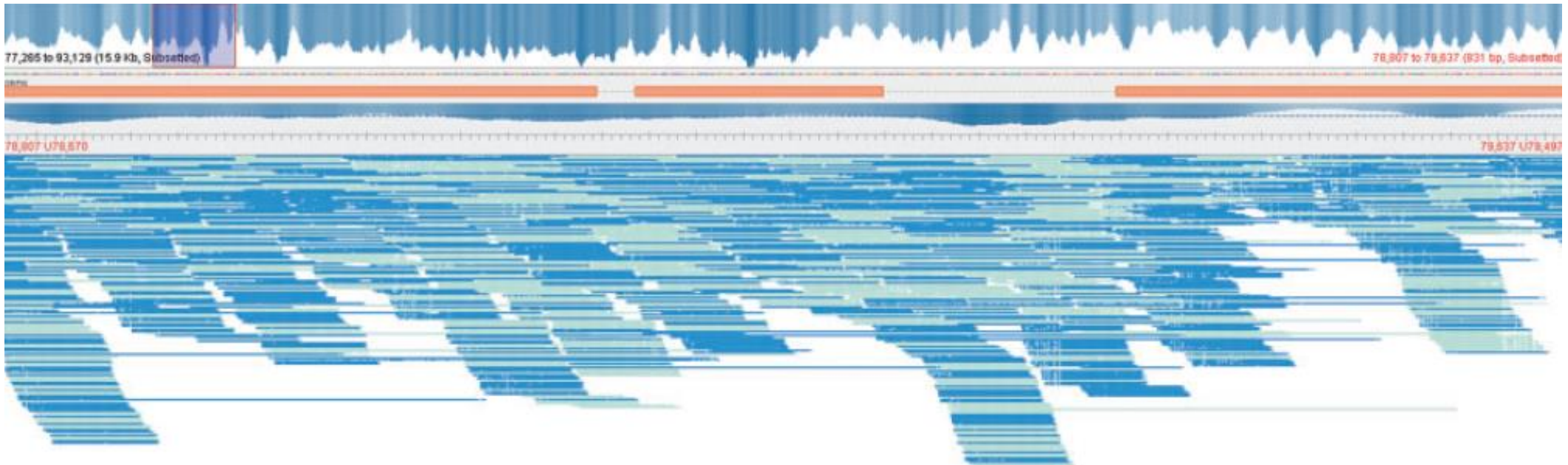
Input file	Format	Type
ACE	Text	Read + Reference/Consensus
AFG	Text	Read + Reference/Consensus
MAQ	Binary/text ^c	Mapping
SOAP	Text	Mapping
SAM	Text	Mapping
BAM	Binary	Mapping
FASTA	Text	Reference/Consensus
FASTQ	Text	Reference/Consensus
GFF3	Text	Feature Annotations

Tablet - funkcie

Vizuálna validácia

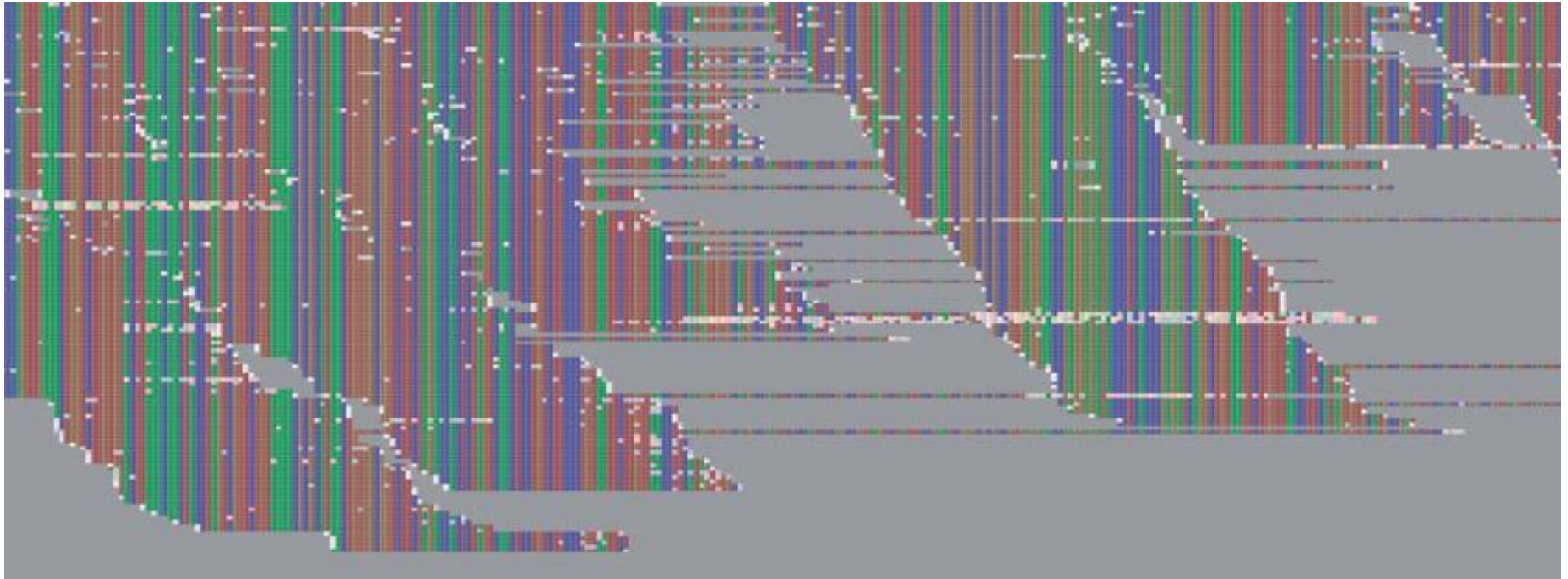
Po celogenómovom sekvenovaní je potrebné de-novo sekvencie skontrolovať

Oblasti vysokého a nízkeho prekrytia



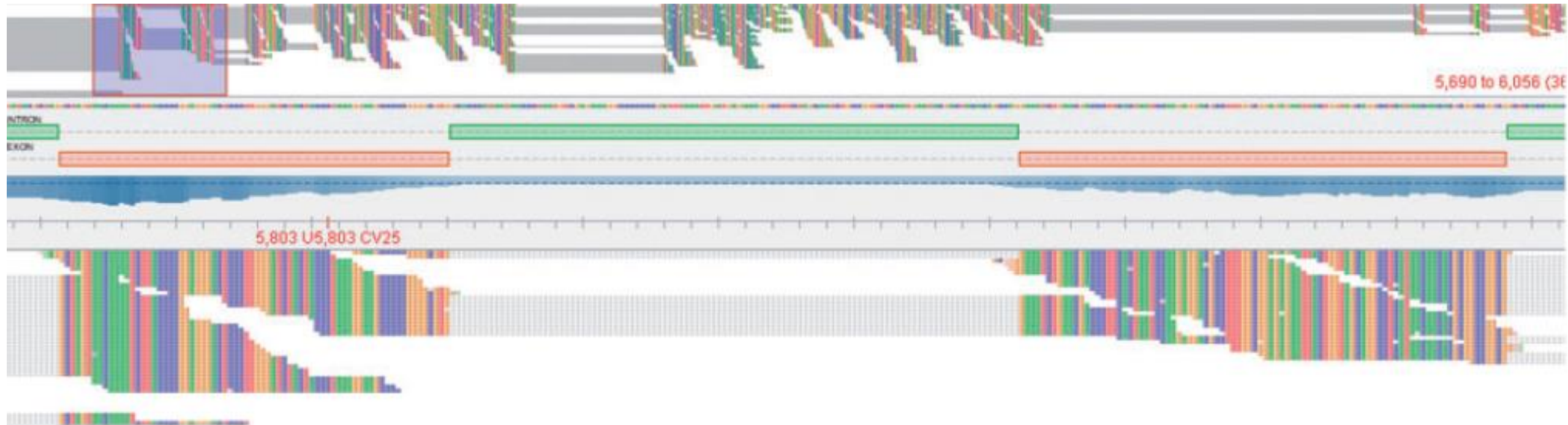
Tablet - funkcie

Niekedy vhodné použiť viac assemblerov, resp. ich rôzne parametrizácie



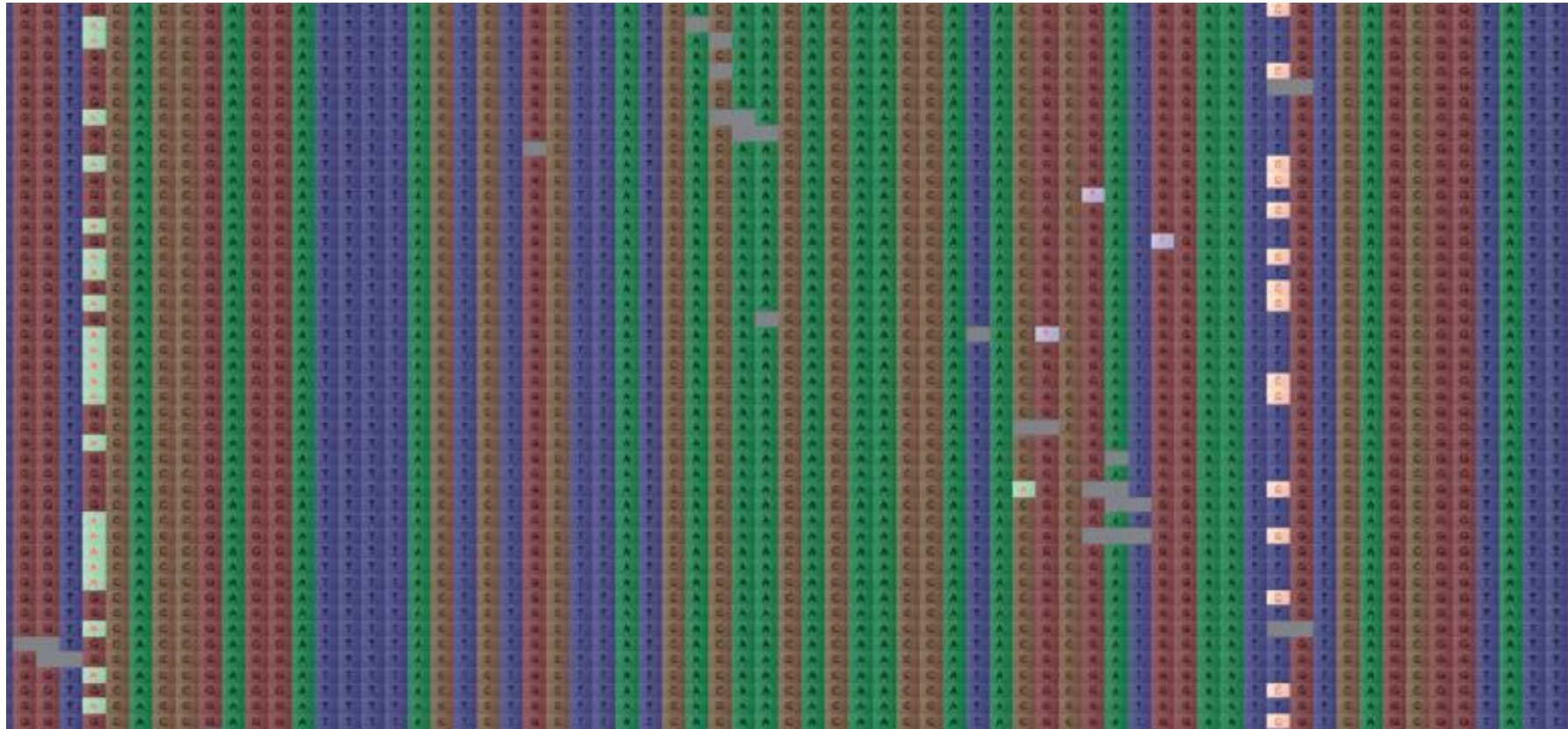
Tablet - funkcie

Kontrola s príslušnou sekvenciou RNA



Tablet - funkcie

Odlíšenie náhodných chýb a SNP



Tablet - funkcie

Export- formát FASTA

- jednotlivé ready
- skupiny readov
- celá sekvencia
- vizuálne formáty -> export obrazovky

Tablet - technológie

- Dobrá čitateľnosť:
- výber vhodných farieb (možno zmeniť)
 - zobrazovanie nukleotidov len od určitého priblíženia
 - antialiasing
- Hybridný prístup:
- niečo medzi pamäťovo a diskovo zameranými
 - v pamäti držané len ID readu, jeho pozícia a dĺžka
 - ostatné dáta na disku
- Vlastná abeceda:
- 16 znakov
 - dve bázy na byte
 - lepšia kompresia, rýchlejšie I/O

Zdroje

Milne I, Bayer M, Cardle L, et al. Tablet—next generation sequence assembly visualization. *Bioinformatics* 2010; 26: 401–2.

MILNE, Iain, et al. *Using Tablet for visual exploration of second-generation sequencing data. Briefings in bioinformatics.* 24. 3. 2012, vol. 14, no. 2, s. 193-202.

The James Hutton Institute, Tablet. Retrieved 18:49, November 5, 2014, <http://ics.hutton.ac.uk/tablet/>

Sequence assembly. (2014, November 3). In *Wikipedia, The Free Encyclopedia*. Retrieved 17:43, November 5, 2014, from http://en.wikipedia.org/w/index.php?title=Sequence_assembly&oldid=632332574

DNA sequencing. (2014, October 31). In *Wikipedia, The Free Encyclopedia*. Retrieved 17:43, November 5, 2014, from http://en.wikipedia.org/w/index.php?title=DNA_sequencing&oldid=631881291

Howard Hughes Medical Institute, DNA Sequence Assembly. Retrieved 18:51, November 5, 2014, http://media.hhmi.org/biointeractive/click/Sequence_Assembly

Dáta: http://genome.crg.es/~jlagarde/encode/pre-DCC/wgEncodeCshlMouseLongRnaSeq/20121121_mouse/SID38132_BC19DCACXX_6.bam

<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwRepliSeq/wgEncodeUwRepliSeqBg02esG1bAlnRep1.bam>