

SCURT ISTORIC AL SECVENȚIERII UNOR GENOMURI:

- 1976: Virusul ADN Φ -X174 (1976) - Frederick Sanger.
- 1995: Primul genom propriu-zis (celular): *Haemophilus influenzae* - TIGR.
- 1996: Primul genom eucariot: *Saccaromyces cerevisiae*
- 1998: Primul genom al unui organism pluricelular: *Caenorhabditis elegans*
- 2000: Modelul experimental *Drosophila melanogaster* – echipă internațională.
- 2001: Secvențe draft genom *Homo sapiens* - James Watson (consorțiu internațional) și Craig Venter.
- 2002: Genomul complet al *Mus musculus*
- 2003: Genomul complet al *H. sapiens*.

ETAPE GENERALE ALE SECVENȚIERII UNUI GENOM:

- izolarea și purificarea ADN-ului genomic de interes;
- fragmentarea randomică prin ultrasonicare a numeroaselor copii ale genomului;
- construirea de biblioteci de fragmente ADN genomic (*genomic libraries*), fie prin clonare în vectori cu care sunt transformate bacterii, fie prin adăugarea de adaptor pentru primeri universali;
- amplificarea fragmentelor ADN prin clonare (pentru WGS și HSS) sau prin PCR (pentru *Next Generation Sequencing* - NGS);
- secvențierea aleatorie și redundantă a fragmentelor (pentru a obține suprapunere de secvențe), utilizând metoda Sanger sau tehnologiile NGS.
- asamblarea *reads*-urilor (secvențe descifrate) în ordinea lor naturală, în *contig*-uri, *scaffold*-uri și super-contiguri.
- adnotarea și readnotarea genomului asamblat.

STRATEGIA SHOTGUN SEQUENCING

Implică fragmentarea **randomică** a unei molecule ADN țintă de mari dimensiuni în fragmente scurte, de câteva kilobaze (~2kb), urmată de clonarea și secvențierea randomică a acestor fragmente.

Are marele avantaj ca necesită doar primeri universali pentru secvențiere.

Etapele procedurii *shotgun sequencing*:

- Copiile moleculei de ADN țintă sunt fragmentate **randomic, prin ultrasonicare, în fragmente scurte de câteva kb, cu lungime medie cunoscută**, clonate ulterior în **vectori de clonare**, pentru a obține **clone library**;

- Doar unele clone (colonii de bacterii) sunt alese (**randomic**) pentru secvențiere;

- Fragmentele clonate sunt secvențiate **parțial, doar la capete**, prin metoda Sanger (dideoxi), cu primeri universali, specifici pentru vectorul de clonare utilizat; rezultă două *reads*-uri de circa 500 nucleotide/fragment clonat (**paired reads**), deci lungimea medie a segmentului intern ne-secvențiat este cunoscută; există și varianta de a secvenția doar la un capăt al fragmentului clonat, rezultând **unpaired read**.

- Cu ajutorul **bioinformaticii** sunt căutate identități parțiale (*overlapping regions*) între *reads*-uri obținute de la diferite clone (aceste suprapuneri sunt aproape inevitabile, datorită **faptului că fragmentarea ADN a fost randomică**);

- Diferitele grade de suprapunere (*overlap*) a *reads*-urilor permit, într-o situație ideală, obținerea unei secvențe continue (**contig**), care acoperă integral fragmentul de ADN țintă;

- De remarcat că prin *overlapping* se obține atât ordonarea fragmentelor clonate, cât și deducerea secvenței regiunilor interne care nu au fost secvențiate;

- *Shotgun sequencing* nu garantează acoperirea lineară, completă, a fragmentului țintă de ADN, astfel încât pot apărea *gap*-uri de secvențiere. De aceea, e posibil să fie nevoie de generarea unor *reads*-uri suplimentare, prin secvențierea altor clone din *clone library*;

- Metoda necesită utilizarea bioinformaticii pentru alinierea *reads*-urilor obținute.

STRATEGIA WGS versus STRATEGIA HSS

Metoda *shotgun sequencing* este implementată în cele **două strategii alternative de secvențiere** a genomurilor: **WGS** (*Whole Genome Sequencing*) și respectiv **HSS** (*Hierarchical Shotgun Sequencing*). Diferența dintre cele două strategii este fundamentală, după cum urmează:

1. Strategia **WGS**:

Se bazează pe fragmentarea tip *shotgun* a unui întreg genom, care este divizat în fragmente scurte, de câteva kb (2-5 kb/fragment). Aceste fragmente sunt clonate în vectori și secvențiate randomic și parțial cu primeri universali prin metoda Sanger, pentru a genera *paired-reads*. Rezultă un număr enorm de *reads*-uri, a căror asamblare corectă necesită software performante de

bioinformatică și **putere foarte mare de calcul**. Putem simboliza aceasta variantă de secvențiere cu acronimul WGSS (*Whole Genome Shotgun Sequencing*).

O alternativă la secvențierea Sanger este oferită de tehnologiile NGS. Astfel, la capetele fragmentelor scurte sunt adăugați adaptorii pentru primeri universali, care permit amplificarea prin PCR a fragmentelor de genom, precum și secvențierea ulterioară a acestora.

Colecția de numeroase **fragmente ADN scurte marcate cu adaptorii** poartă denumirea de **NGS genomic library**. Acest tip de *library* este un concept complet diferit de cel de *clone library*, care reprezintă o colecție de clone derivată dintr-o clonă BAC (vezi mai jos). Similitudinea între WGSS care implică clonare în vectori și secvențierea WGS-NGS care presupune clonare prin PCR, este că aceste strategii **NU** implică o ierarhizare prealabilă a fragmentelor genomice care urmează să fie secvențiate. În primul rând, aceste fragmente genomice obținute pentru WGS sunt oricum prea scurte pentru a putea fi ierarhizate cu ajutorul profilelor comune de restricție.

De remarcat că tehnologiile NGS evită clonarea în vectori, care necesită specialiști în inginerie genetică, costuri ridicate și timp mai îndelungat pentru manoperă. Astfel, secvențierea genomului prin strategia WGS-NGS este mai ieftină și mai rapidă decât strategiile WGSS și HSS, care implică clonare în vectori, urmată de secvențiere Sanger.

Strategia generală WGS este potrivită pentru secvențierea genomurilor bacteriene, care conțin foarte puține secvențe repetitive.

2. Strategia HSS:

Se procedează la fragmentarea inițială randomică în **segmente relativ mari, de cca 150-200 kb**. Fragmentele respective de ADN sunt clonate în vectori de tip BAC (*Bacterial Artificial Chromosome*). Înainte de secvențiere, clonele BAC sunt tratate cu enzime de restricție și apoi **IERARHIZATE** în *contig-uri* obținute pe baza **suprapunerii parțiale a profilurilor de restricție** a fragmentelor clonate. Pentru fiecare cromozom din genom sunt obținute unul sau mai multe *contig-uri* de clone BAC, astfel încât apartenența și ordinea acestor clone într-un cromozom de interes este cunoscută **înainte** de secvențierea propriu-zisă. Contig-urile clonelor BAC sunt contig-uri de restricție, iar obținerea acestora nu implică secvențierea.

Ulterior, se aplică procedura *shotgun sequencing* individual pentru clonele BAC din contig-urile de restricție. Evident, pentru fiecare clonă BAC, se procedează la ultrasonicare în fragmente de câteva kb, apoi se construiește *clone library*. Secvențierea randomică este realizată cu primeri universali pentru a obține *pair-reads-uri*, care sunt asamblate pentru generarea *contig-ului* ideal al clonei BAC respective.

Deoarece ordinea clonelor BAC in cromozom este deja cunoscută, asamblarea *contig*-urilor de secvența corespunzătoare contig-urilor de restricție este relativ facilă. Astfel, **HSS** solicită o putere de calcul semnificativ mai mică decât strategia alternativă **WGSS**.

Alegerea strategiei de secvențiere a unui genom depinde de:

- tipul de genom care este secvențiat: PK sau EK;
- scopul pentru care este secvențiat respectivul genom;
- disponibilitatea echipamentelor și a reactivilor;
- expertiza tehnică a membrilor echipei laboratorului.

De exemplu, se poate proceda relativ ieftin la secvențierea parțială a genomului. Exomul uman (care reprezintă doar o mică parte din genom) este secvențiat rapid și relativ ieftin prin strategia WGS-NGS (Illumina, Ion Torrent, etc.).

Dacă dorim să obținem o secvențiere și o asamblare de mare calitate a unui genom eucariot, strategia HSS reprezintă standardul în domeniu, chiar dacă este mai laborioasă și mai scumpă.