

İLERİ ARAŞTIRMA YÖNTEMLERİ 56901007

2020-21 BAHAR

DR. GÜNSELİ ÇUBUKÇUOĞLU DENİZ

	Roche 454 GS FLX	Illumina Genome Analyzer	Applied Biosystems SOLiD	Sanger
Sequencing method	Pyrosequencing	Reversible dye terminators	Sequencing by ligation	Dye terminators
Read lengths	400 bases	36 bases	35 bases	800 bp
Sequencing run time	10 h	2.5 days	6 days	3 h
Total bases per run	500 Mb	1.5 Gb	4 Gb	800 bp

VERİ ANALİZİ

- Shotgun: tüm genomun rasgele fragmentasyonu

Input: GGCGTCTATATCTCGGCTCTAGGCCCTCAATTTTTT

Copy: GGCGTCTATATCTCGGCTCTAGGCCCTCAATTTTTT
GGCGTCTATATCTCGGCTCTAGGCCCTCAATTTTTT
GGCGTCTATATCTCGGCTCTAGGCCCTCAATTTTTT
GGCGTCTATATCTCGGCTCTAGGCCCTCAATTTTTT

Fragment: GGCGTCTA TATCTCGG CTCTAGGCCCTC AATTTTTT
GGC GTCTATAT CTCGGCTCTAGGCCCTCA TTTTTT
GGCGTC TATATCT CGGCTCTAGGCCCT CAAATTTTTT
GGCGTCTAT ATCTCGGCTCTAG GCCCTCA TTTTTT

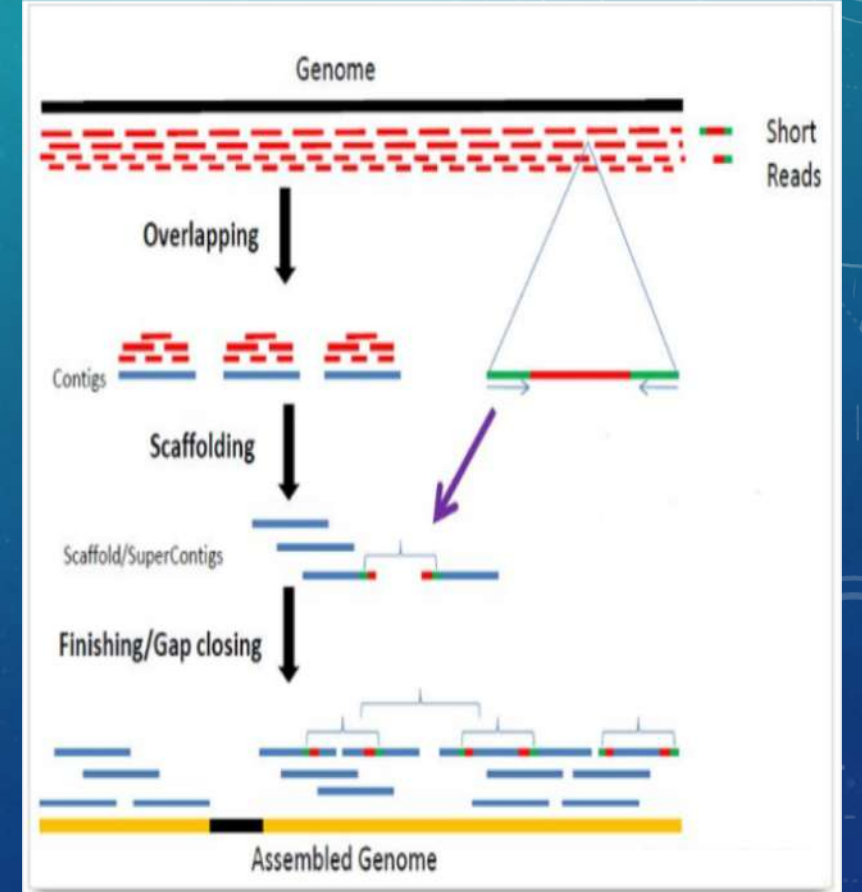
Reconstruct
this

CTAGGCCCTCAATTTTTT
CTCTAGGCCCTCAATTTTTT
GGCTCTAGGCCCTCAATTTTTT
CTCGGCTCTAGGCCCTCAATTTTTT
TATCTCGACTCTAGGCCCTCA
TATCTCGACTCTAGGCC
TCTATATCTCGGCTCTAGG
GGCGTCTATATCTCG
GGCGTCGATATCT
GGCGTCTATATCT
→ GGCGTCTATATCTCGGCTCTAGGCCCTCAATTTTTT

From these

VERİ ANALİZİ

- kapsam= average coverage= genomdaki bir pozisyonu kapsayan ortalama okuma sayısı
- overlap= uç kısımları örtüşen
- contig= overlap sekansların bir kısmı
- scaffolding= boşluklu contiglerin yanyana getirilmesi



STANDART FASTQ FORMAT

@9WV6Z:791:946

GCTCTTCCGATCTATGGATGCACCAAGATATATGAC

+

774399222034221774399222034221774399

DE NOVO GENOM BİRLEŐTİRME

- DNA sekansı, uzunluđu, kompozisyonu ile ilgili önceki bilgiler yok
- Genellikle 35-150 bp kısa sekanslar üretilir
- Çiftlenmiş uçlar daha uzun sekanslar oluşturur (\approx 250-500 bp)
- Çiftlenmiş uçlardan ortaya doğru okunur ve bir çift okuma elde edilir.

REFERANS GENOM HARİTALAMA

- Önceden sekanslanıp biraraya getirilmiş genom referans alınır
- Okumalar bağımsız olarak bu referans genomda hizalanır.
- Her okuma en uygun pozisyonuna yerleştirilir.
- Indel deęişiklikleri hizalamanın daha kompleks hale gelmesine neden olur.