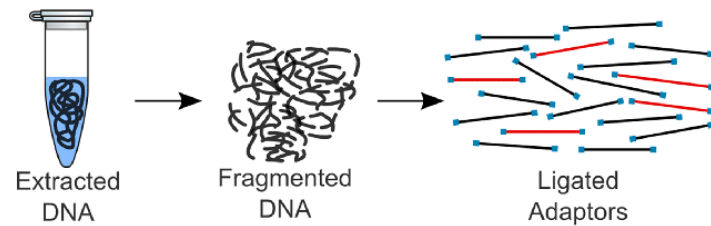


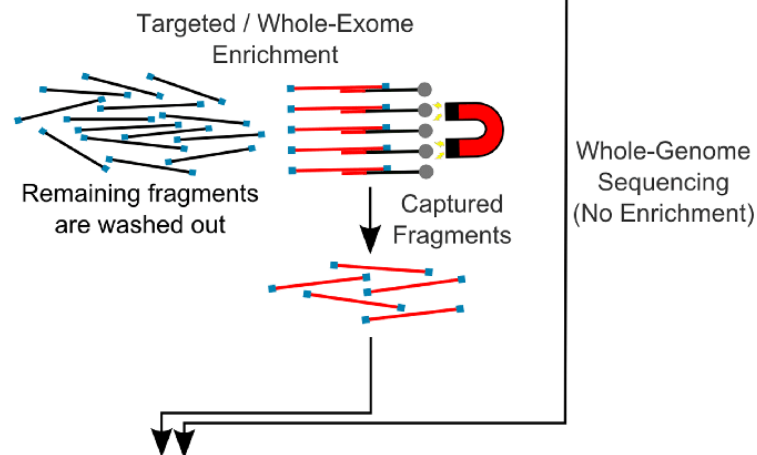
GENOMIK

Tugce Bilgin Sonay

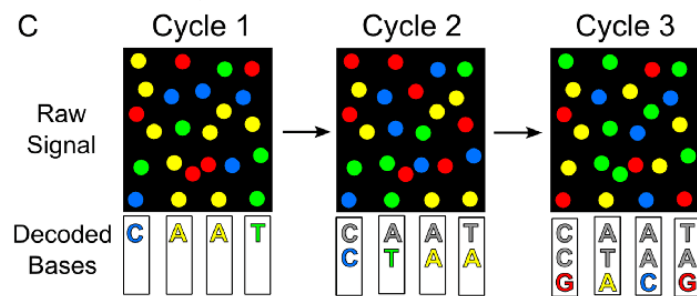
A



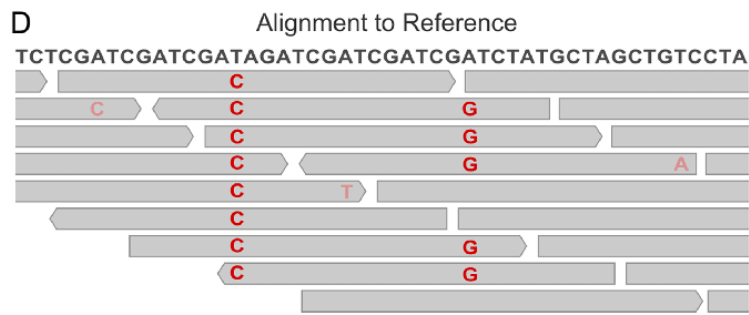
B



C



D





- İlk 35-400 baz çiftini diziliyoruz: okuma
- Bir seferde 6 milyar okuma elde edebiliriz
- Paired-end olursa iki taraftan okuruz.



```
GTTGAGGCTTGCGTTTTTGGTACGCTGGACTTTGT  
GTACTCGTCGCTGCGTTGAGGCTTGCGTTTTTGGT  
ATGGTACGCTGGACTTTGTAGGATACCCTCGCTTT  
TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC  
CTTGCGTTTATGGTACGCTGGACTTTGTAGGATAC  
TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC  
GCGTTTATGGTACGCTGGACTTTGTAGGATACCCT  
GAGGCTTGCGTTTATGGTACGCTGGACTTTGTAGG  
GCGTTGAGGCTTGCGTTTATGGTACGCTGGATTTT  
CGTTTATGGTACGCTGGACTTTGTAGGATACCCTC  
ATGGTACGCTGGACTTTGTAGGATACCCTCGCTTT  
GTTTATGGTACGCTGGACTTTGTAGGATACCCTCG  
TCTCGTCGCTCGTCGCTGCGTTGAGGCTTGCGTTTA  
TGCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTA  
GCTCGTCGCTGCGTTGAGGCTTGCGTTTATGGTAC  
TATGGTACGCTGGACTTTGTAGGATACCCTCGCTT  
TCGTGCTCGTCGCTGCGTTGAGGCTTGCGTTTTTG  
CGTCGCTGCGTTGAGGCTTGCGTTTATGGTACGCT  
GTTGAGGCTTGCGTTTATGGTACGCTGGGCTTTTT  
TTGCGTTTATGGTACGCTGGACTTTGTAGGATACC
```



SRA Short Read Archive

SRA

Access
Public (156)

Source
DNA (156)

[Clear all](#)


[Show additional filters](#)

Display Settings: ☒ Summary, 20 per page

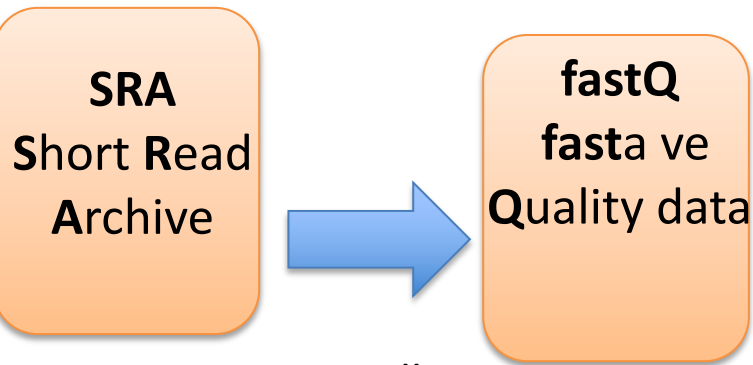
Send to: ☒

Results: 1 to 20 of 156

<< First < Prev Page of 8

 Showing SRA Experiments for SRP047378

- ☐ [full length mtDNA sequencing of child SC8C1: whole blood](#)
 1. 1 ILLUMINA (Illumina MiSeq) run: 1.4M spots, 597.8M bases, 382.3Mb downloads
Accession: SRX707999
- ☐ [full length mtDNA sequencing of mother SC8: whole blood](#)
 2. 1 ILLUMINA (Illumina MiSeq) run: 1.3M spots, 570.9M bases, 352.3Mb downloads
Accession: SRX707998



- Gene ID, uzunluk
- Fasta dizisi
- Gene ID, uzunluk
- Kalite verisi

```
@SRR1583063.1.1 1 length=225
GTATCCGACATCTGGTTCCTACTTCAGGGTCATAAAGCCTAAAT
+SRR1583063.1.1 1 length=225
?????BAADEEDEDDEDFGGGGGIIIIIIIIHIIIIIIHIIIIHIHI
@SRR1583063.2.1 2 length=251
GGTTCCTACTTCAGGGTCATAAAGCCTAAATAGCCCACACGTTC
+SRR1583063.2.1 2 length=251
<?????BBBDDDDDDDDGGGGGGIIIIIIIIHIIIIIIHHHHII
```

Dizinin kalite degerleri

kalite degeri: 30 -> baz %99.9 ihtimal dogru

ATGCAGGTTCGGANN



Dizinin kalite degerleri

kalite degeri: 30 -> baz %99.9 ihtimal dogru

kalite degeri: 10-
> baz %10 yanlis.

ATGCAGGTTCGGANN



Dizinin kalite degerleri

kalite degeri: 30 -> baz %99.9 ihtimal dogru

kalite degeri: 10 -> baz %10 yanlis.

ATGCAGGTTCGGANN



bazin ne oldugu
belirlenemediyse
N

$$Q = -10 \log_{10} P \quad \longrightarrow \quad P = 10^{\frac{-Q}{10}}$$

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%
50	1 in 100000	99.999%

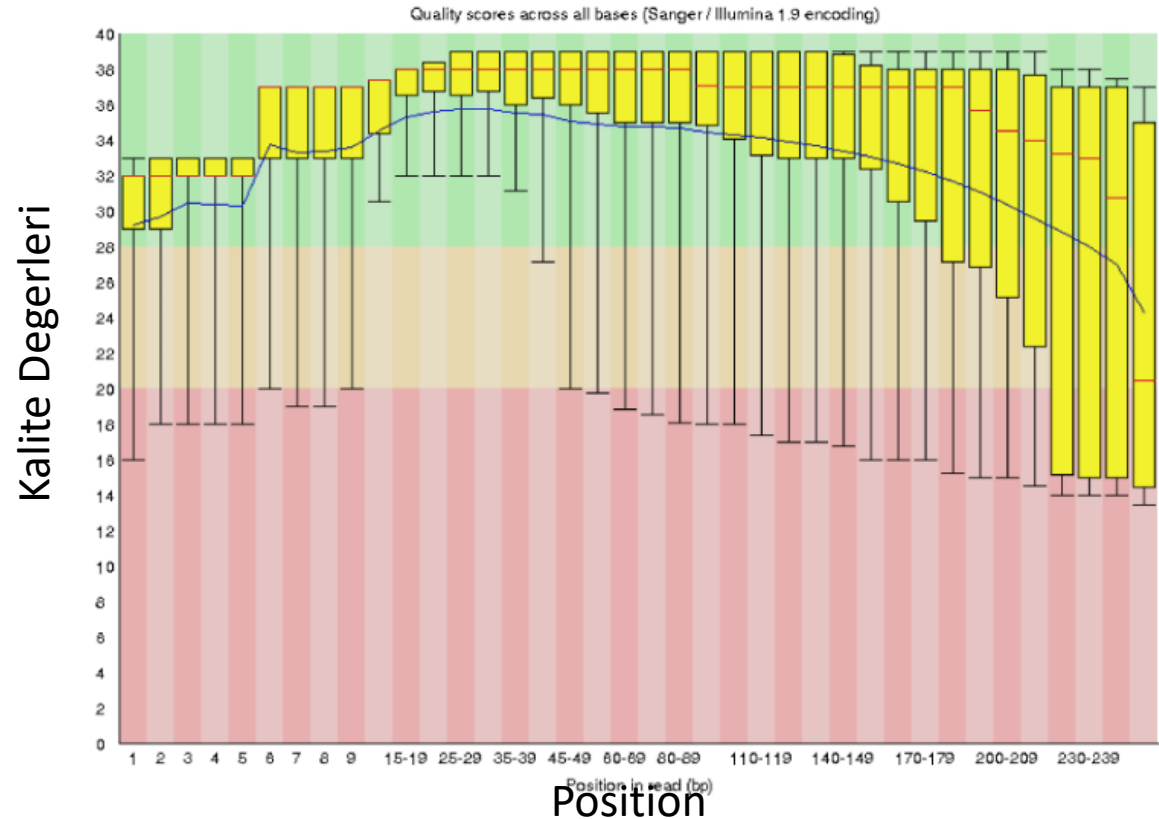
SRA
Short Read
Archive



fastQ
fasta ve
Quality data

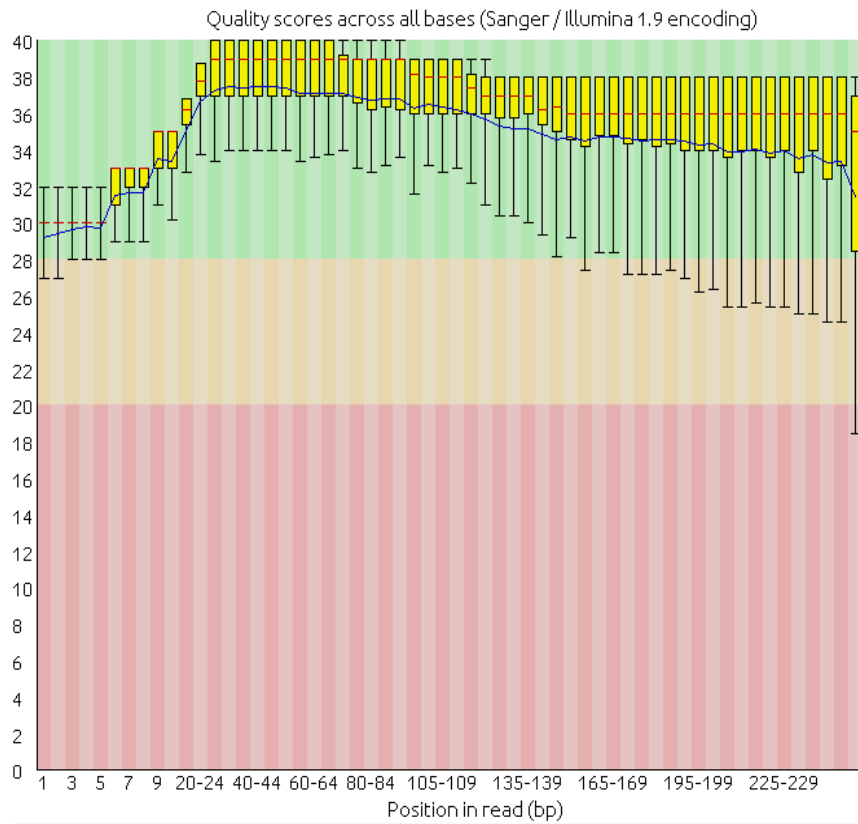
SRA toolkit

FastQC

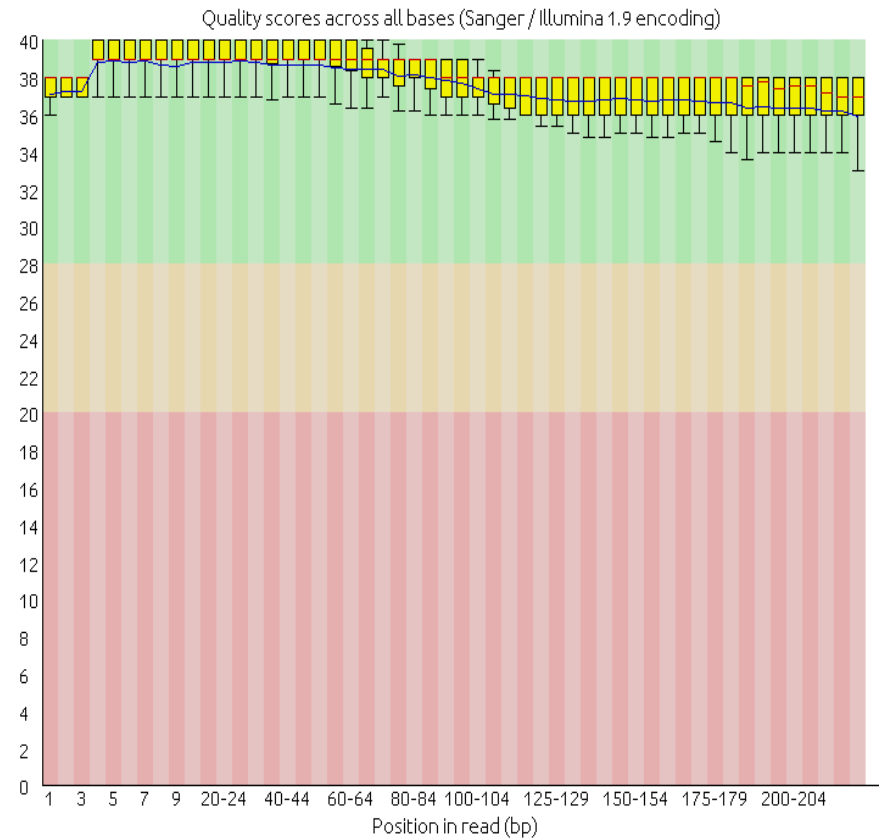


- Bastaki ve sondaki kotu kalite bazlari kes
- Aralardaki kotu kalite bazlari filtrele

Oncesi

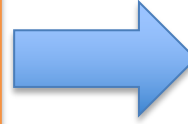


Sonrasi

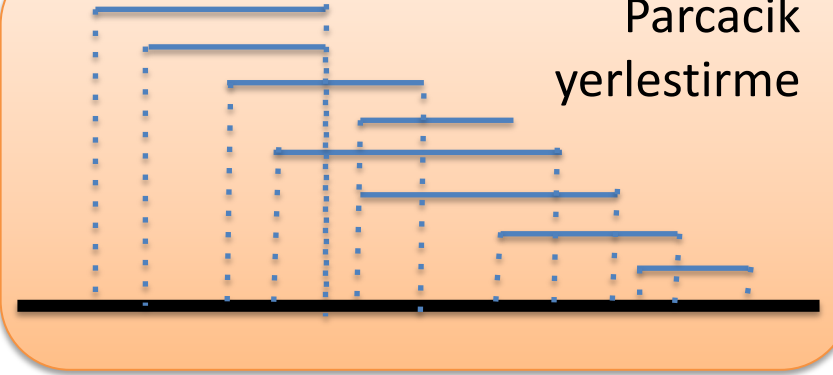




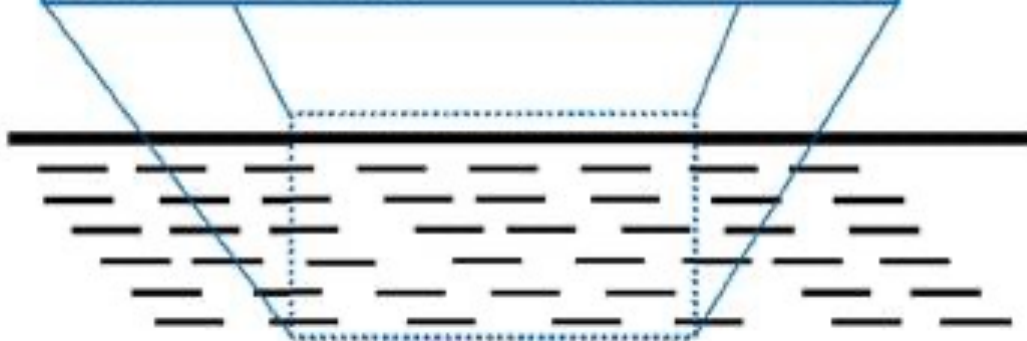
Kisa Diziler



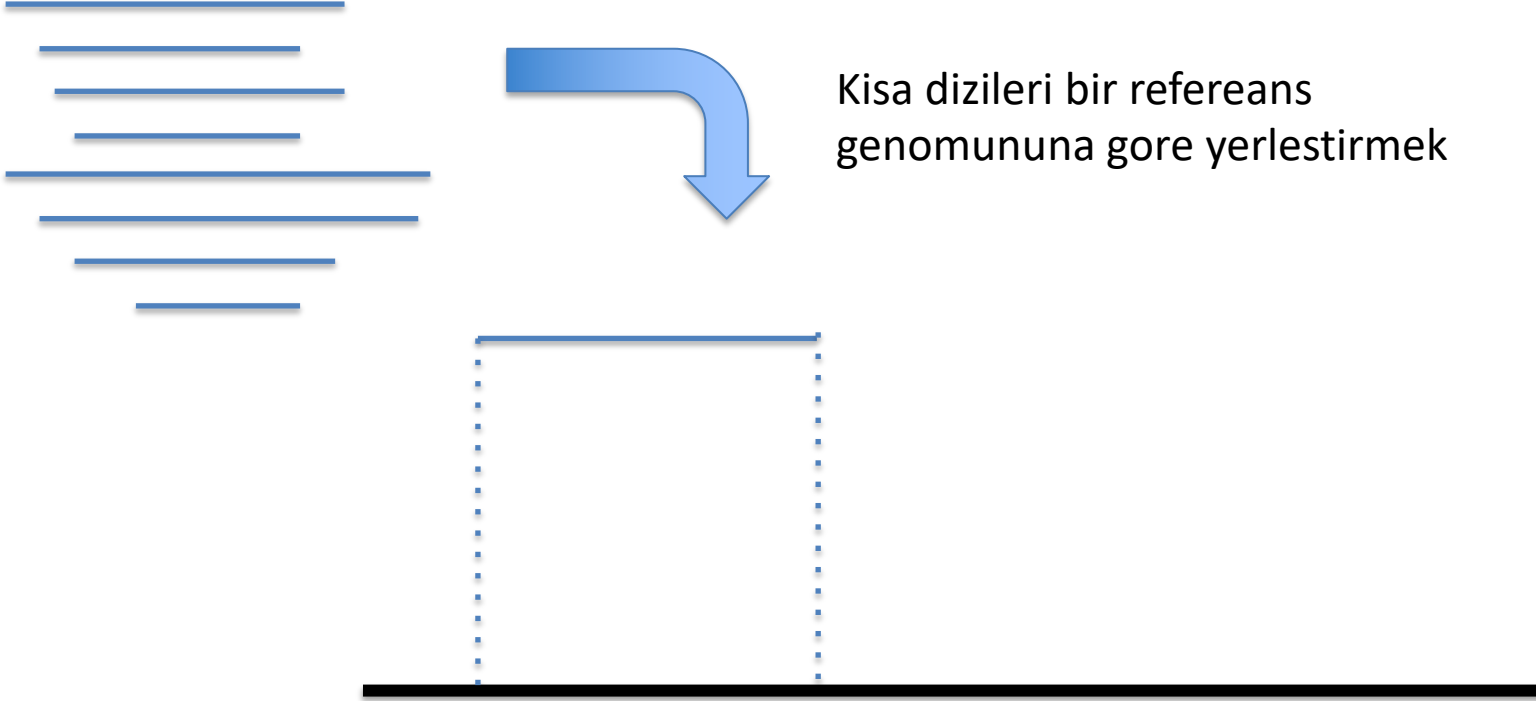
Parcacik
yerlestirme



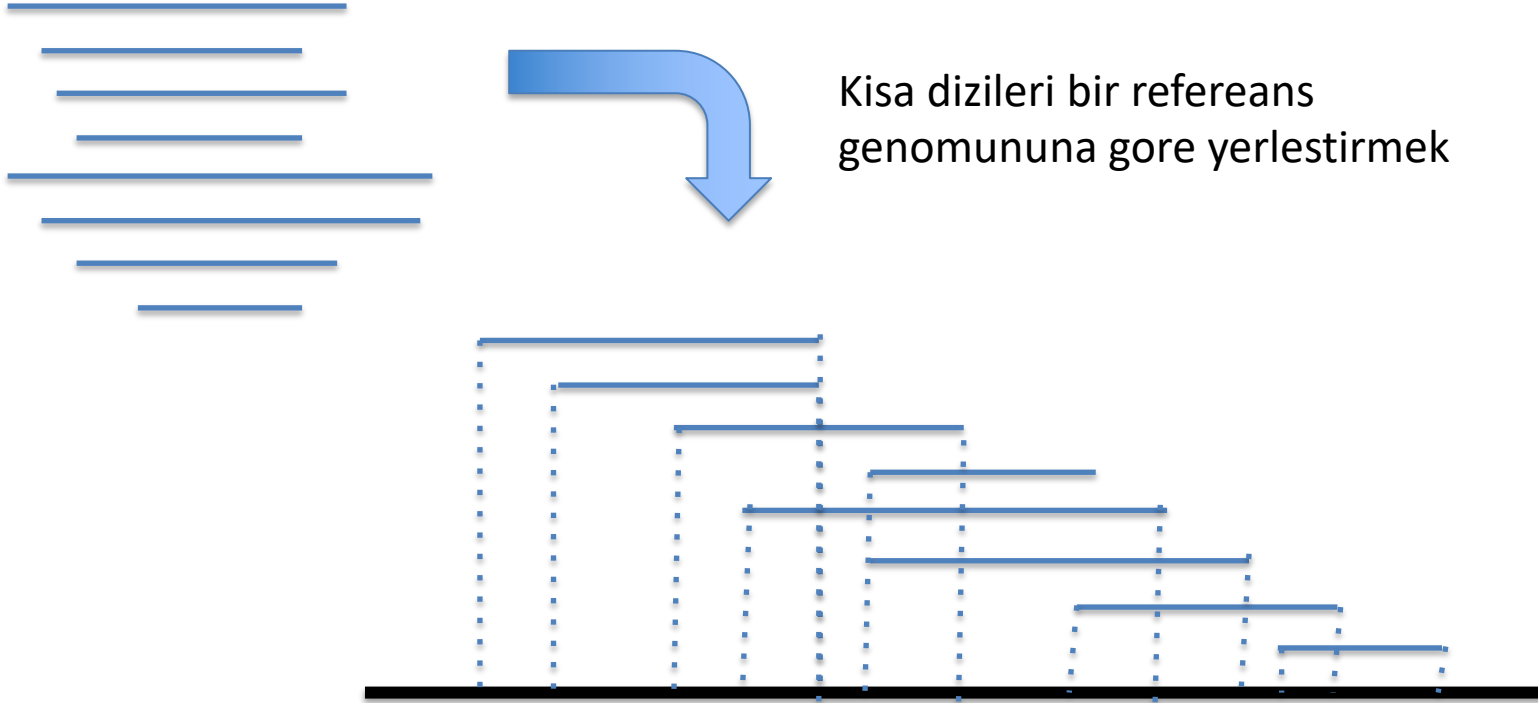
```
AGGGTTATTTACCCTACTGCGACTATCTAGT
AGG  ATTTAC  ACTGCG  TAGCTA T
AGGG  TTACCC  CTGCGA  AGCTAG
AGGGT      TCCTAC  GCGACT  GCTAGT
GGGTTA      CCTAC   ACTAGC  GT
AGG  ATTTAC  TACTGC   AGCTAG
AGGGT  TTACCC      CGACTAG  AGT
```



Parcacik Yerlestirme



Parcacik Yerlestirme



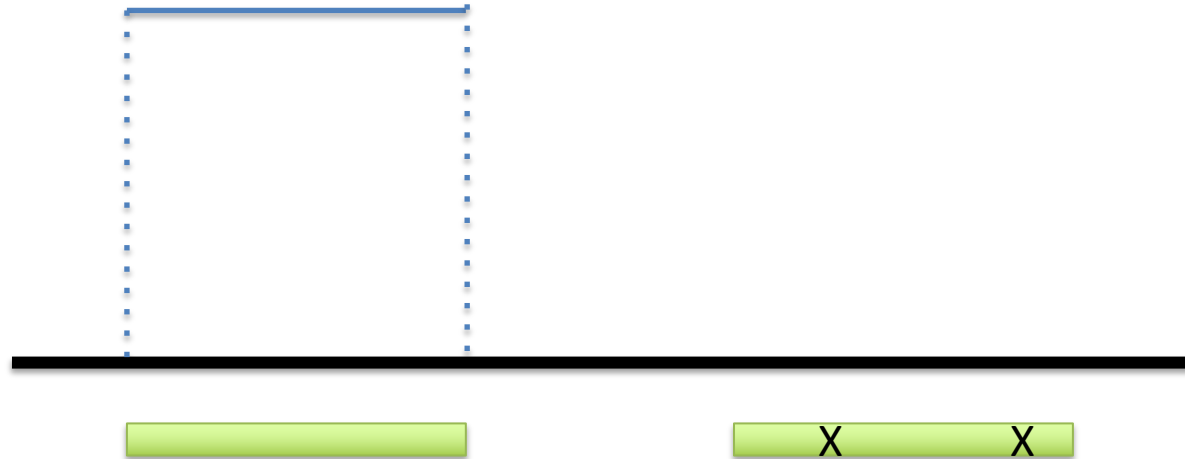
Parcacik Yerlestirme



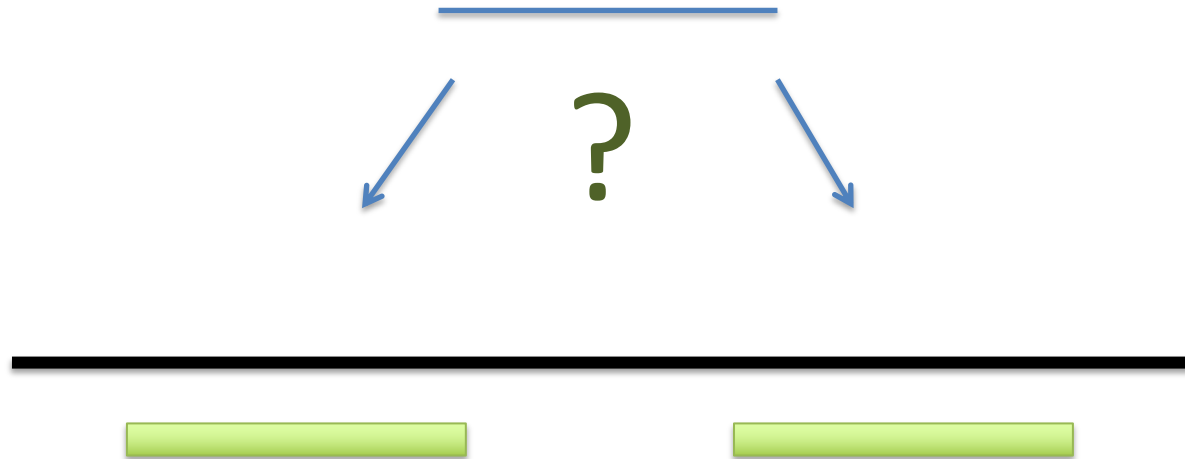
Parcacik Yerleştirme



Parcacik Yerleştirme

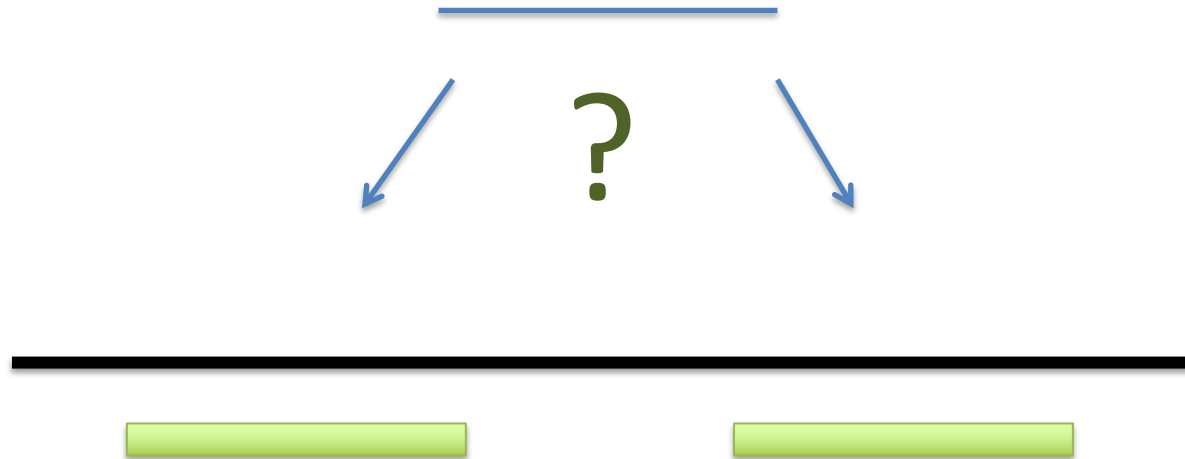


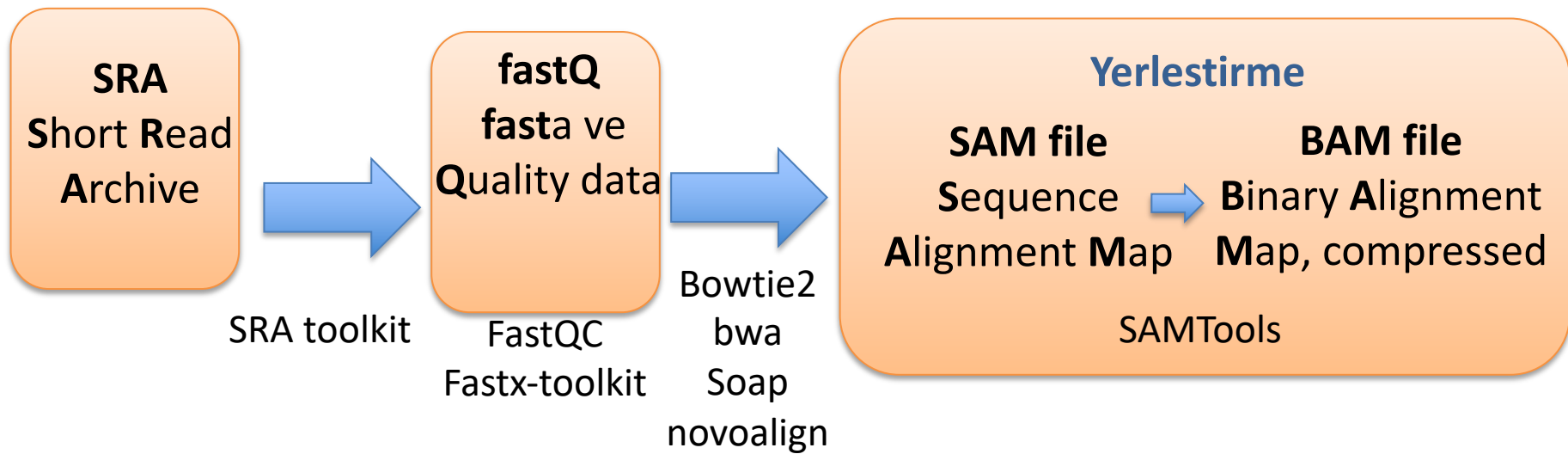
Parcacik Yerleştirme



Yerlestime Kalitesi

Bu parçacık genoma yerleştirilirken ne kadar kesin konuşabiliriz?





Sequence Alignment Map (SAM)

Dosyası

Sizce ne kadar büyüktür bu dosyalar?

A

Coor	10	20	30	40
ref	12345678901234	5678901234567890123456789012345		
	AGCATGTTAGATAA**GATAGCTGTGCTAGTAGGCAGTCAGCGCCAT			
+r001/1	TTAGATAAAGGATA*CTG			
+r002	aaaAGATAA*GGATA			
+r003	gcctaAGCTAA			
+r004	ATAGCT.....TCAGC			
-r003	ttagctTAGGC			
-r001/2			CAGCGGCAT	

B

Header section										QUAL (read quality; * meaning such information is not available)	
@HD VN:1.5 SO:coordinate											
@SQ SN:ref LN:45											
r001	99	ref	7	30	8M2I4M1D3M	=	37	39	TTAGATAAAGGATACTG	*	
r002	0	ref	9	30	3S6M1P1I4M	*	0	0	AAAAGATAAGGATA	*	
r003	0	ref	9	30	5S6M	*	0	0	GCCTAAGCTAA	*	SA:Z:ref,29,-,6H5M,17,0;
r004	0	ref	16	30	6M14N5M	*	0	0	ATAGCTTCAGC	*	
r003	2064	ref	29	17	6H5M	*	0	0	TAGGC	*	SA:Z:ref,9,+,5S6M,30,1;
r001	147	ref	37	30	9M	=	7	-39	CAGCGGCAT	*	NM:i:1

Alignment section

QNAME (query template name, aka. read ID)

FLAG (indicates alignment information about the read, e.g. paired, aligned, etc.)

RNAME (reference sequence name, e.g. chromosome /transcript id)

POS (1-based position)

MAPQ (mapping quality)

CIGAR (summary of alignment, e.g. insertion, deletion)

RNEXT (reference sequence name of the primary alignment of the NEXT read; for paired-end sequencing, NEXT read is the paired read; corresponding to the RNAME column)

PNEXT (Position of the primary alignment of the NEXT read in the template; corresponding to the POS column)

TLEN (the number of bases covered by the reads from the same fragment. In this particular case, it's 45 - 7 + 1 = 39 as highlighted in Panel A). Sign: plus for leftmost read, and minus for rightmost read

SEQ (read sequence)

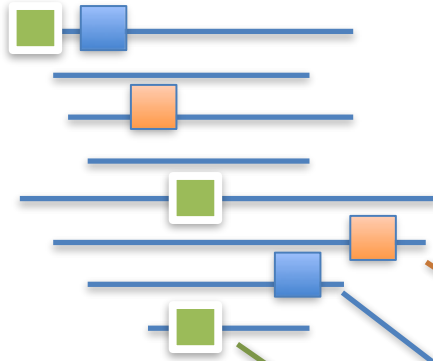
Optional fields in the format of TAG:TYPE:VALUE

Binary Alignment Map (BAM) Dosyasi

Bitwise Flags

Integer	Binary	Description (Paired Read Interpretation)
1	000000000001	template having multiple templates in sequencing (read is paired)
2	000000000010	each segment properly aligned according to the aligner (read mapped in proper pair)
4	000000000100	segment unmapped (read1 unmapped)
8	000000001000	next segment in the template unmapped (read2 unmapped)
16	000000010000	SEQ being reverse complemented (read1 reverse complemented)
32	000000100000	SEQ of the next segment in the template being reverse complemented (read2 reverse complemented)
64	000001000000	the first segment in the template (is read1)
128	000010000000	the last segment in the template (is read2)
256	000100000000	not primary alignment
512	001000000000	alignment fails quality checks
1024	010000000000	PCR or optical duplicate
2048	100000000000	supplementary alignment (e.g. aligner specific, could be a portion of a split read or a tied region)

Indexleme

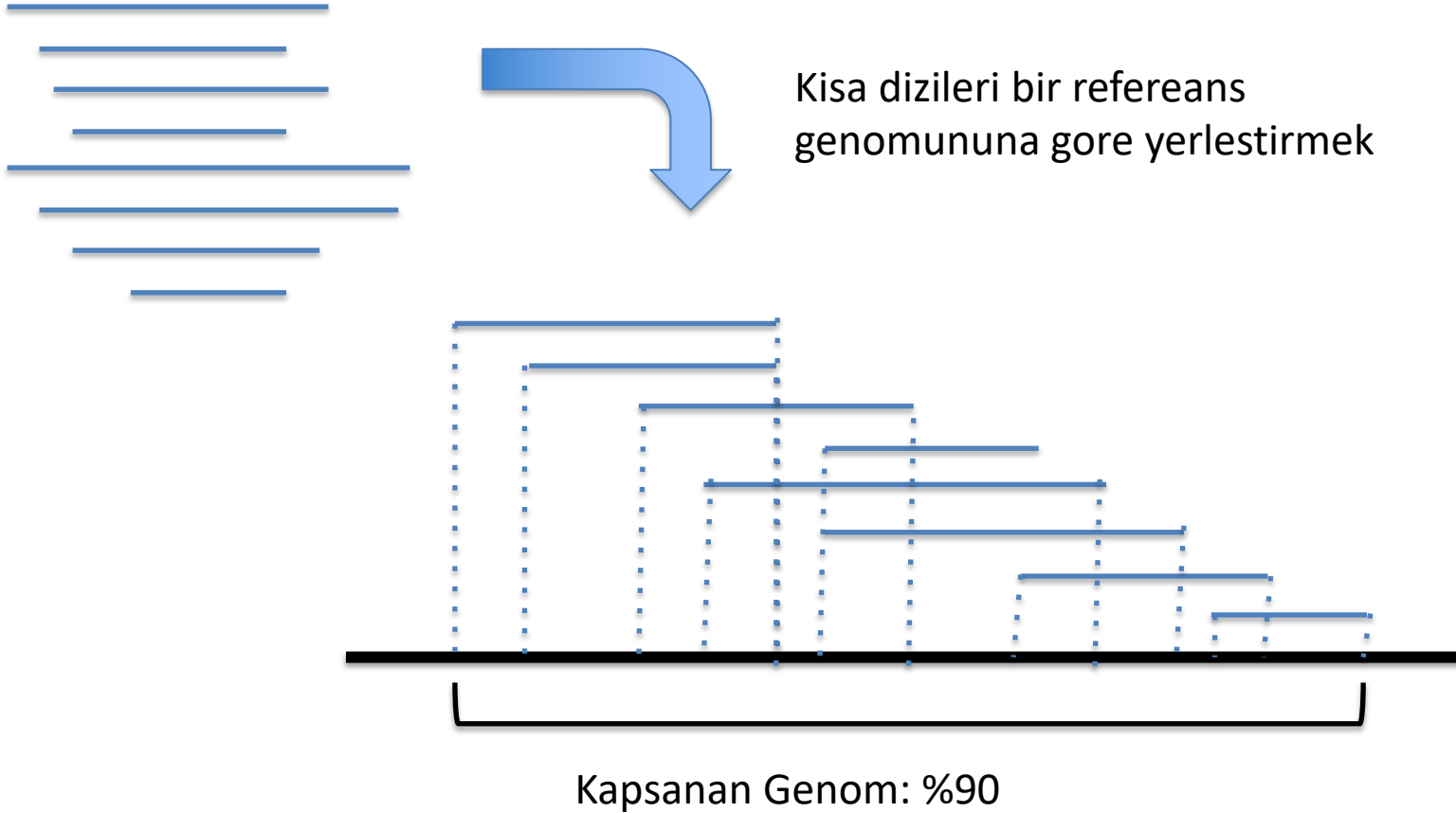


Her islem oncesinde kullandigimiz aracın kendi indeksleme komutuyla dizilerimizi indeksliyoruz.

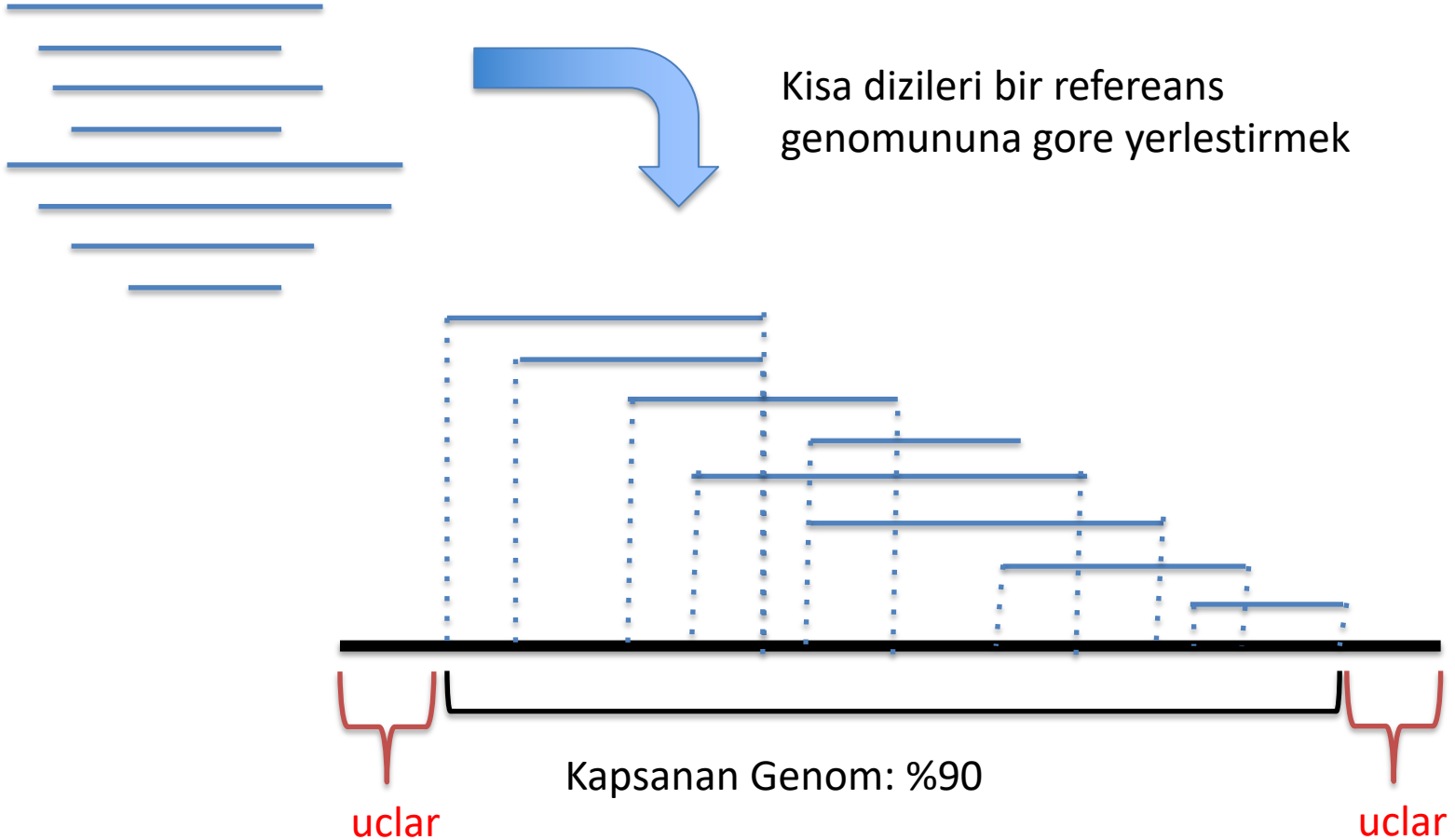
Kok	Pozisyon
AAAA	0,8
ACCT	4,16,24
GTGA	12,20
...	...

Bunun amaci okumaya hizlandirmek. Kitabın icindekiler kısmı gibi dusunebiliriz.

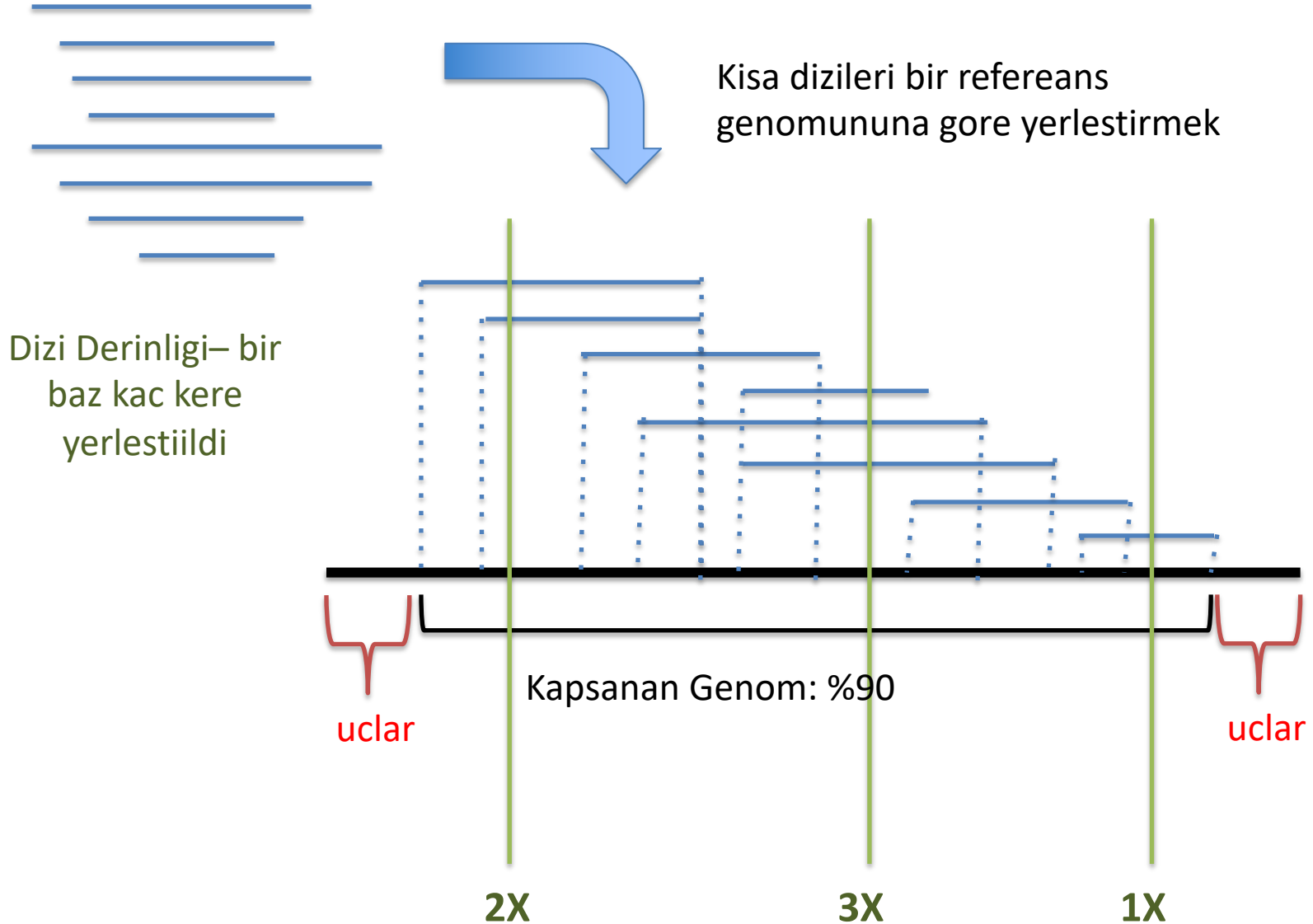
Parcacik Yerlestirme



Parcacik Yerlestirme



Parcacik Yerlestirme



Possible PCR clones

[illegible]

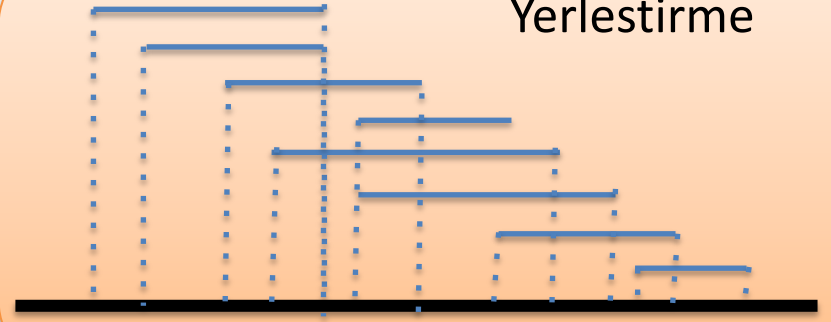
CTCTCGTGCTCGTCGCTGCGTTGAGGCTTGC GTTTATGGTACGCTGGACTTTGTAGGATACCCTCGCTTTC



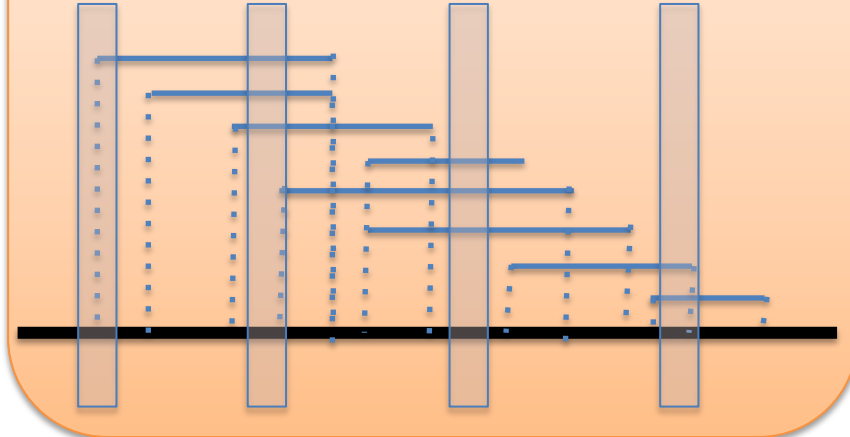
Kisa Diziler



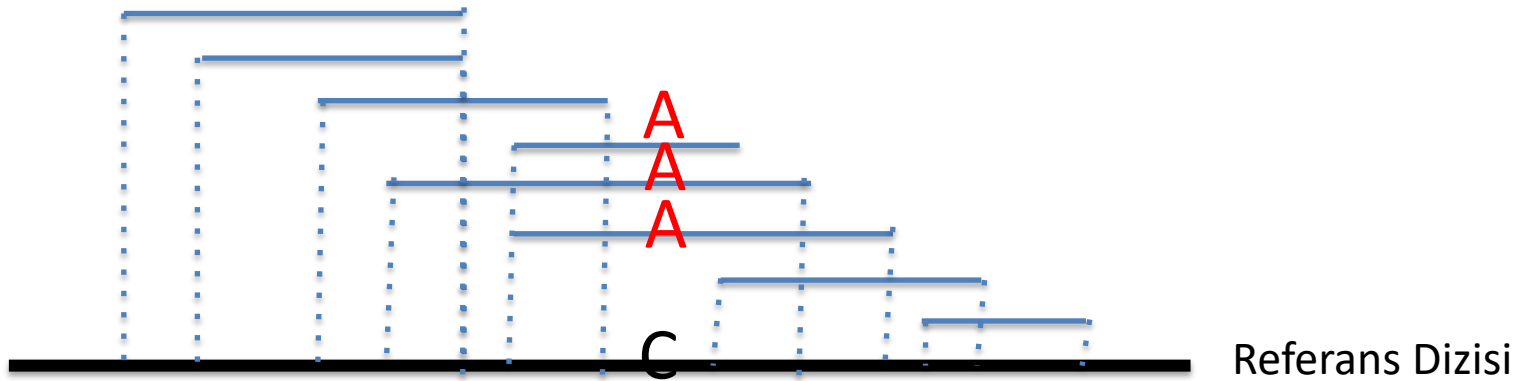
Yerleştirme



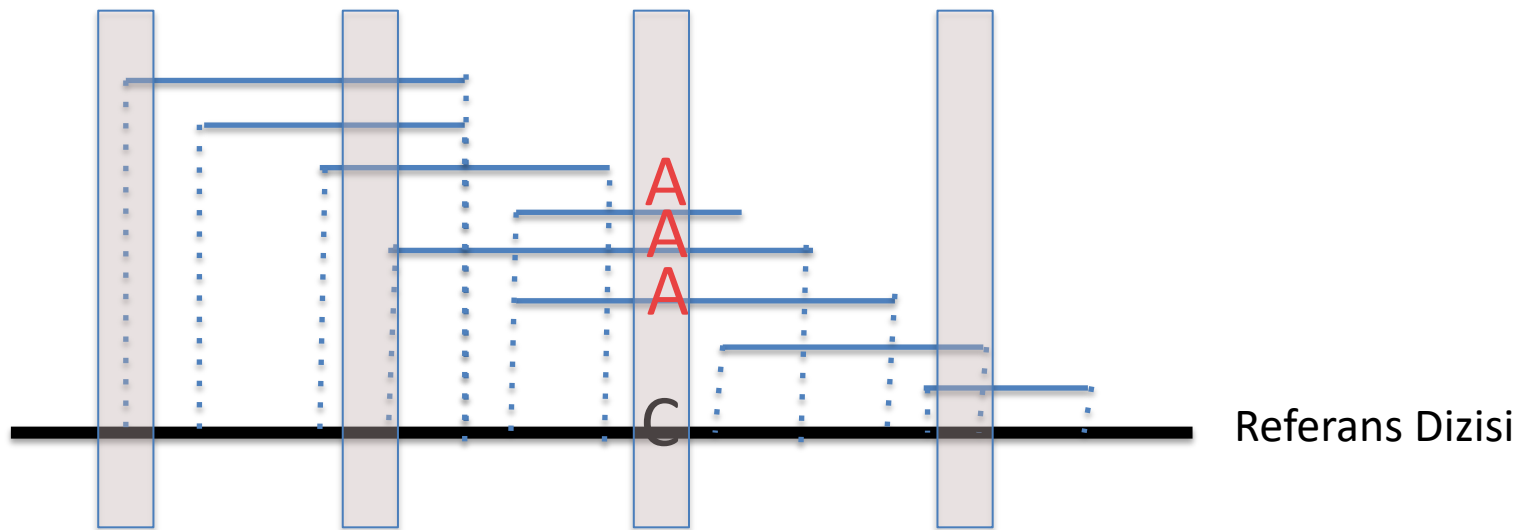
Varyant Tespiti



Varyant Tespiti



Varyant Tespiti



Variant Call File Format

Reference/
Alternate Allele

Upstream
pipeline Result

Genotype

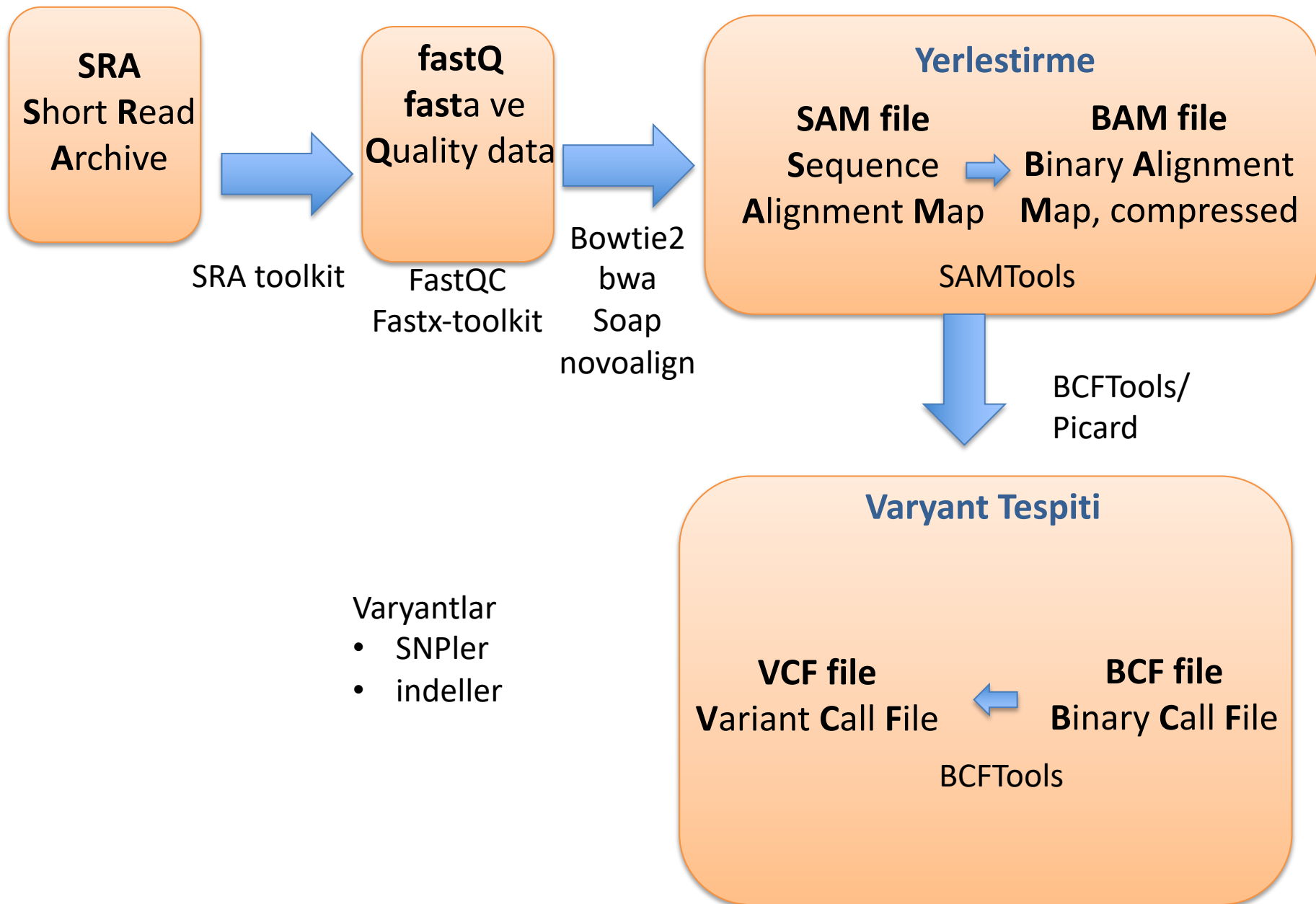
##fileformat=VCFv4.0									
#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	Sample
chr1	881627	2272757	G	A	46	PASS	DP=7	GT:DP	0/1:7
chr1	887801	.	A	ATC	35	PASS	DP=16	GT:DP	1/1:16
chr1	888659	3748597	T	C	28	PASS	DP=7	GT:DP	1/1:7
chr1	900286	.	AG	A	3	FAIL	DP=1	GT:DP	1/1:1
chr1	900505	28705211	G	C	0	PASS	DP=5	GT:DP	0/1:5
chr1	908823	28687780	G	A	29	PASS	DP=18	GT:DP	1/0:18
chr1	948921	15842	T	C	.	PASS	DP=28	GT:DP	1/.:28
chr1	949608	1921	G	A	44	PASS	DP=7	GT:DP	0/1:7
chr1	949654	8997	A	G	31	PASS	DP=16	GT:DP	1/1:16
chr1	980460	.	G	A	2	FAIL	DP=2	GT:DP	./.:2
chr1	982941	3128102	T	C	8	PASS	DP=15	GT:DP	1/1:15
chr1	1158631	.	AG	CC	75	PASS	DP=21	GT:DP	1/1:21
chr1	1225612	.	G	A	5	FAIL	DP=4	GT:DP	1/1:4
chr1	1226063	.	G	A	26	PASS	DP=6	GT:DP	0/1:6

Chromosome

Position

Call Quality

Read Depth either in INFO or within each samples



Websiteleri

SAMtools: <http://www.htslib.org>

Picard tools: <https://broadinstitute.github.io/picard/>

BWA: <http://bio-bwa.sourceforge.net>

Bowtie: <http://bowtie-bio.sourceforge.net/index.shtml>

SOAP: <http://soap.genomics.org.cn/index.html>

Novoalign: <http://www.novocraft.com/products/novoalign/>

FASTX-toolkit: http://hannonlab.cshl.edu/fastx_toolkit/

FastQC: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>



- 156 kisinin mitokondriyal DNAsi
- Agizdan veya kandan örnek
- 20,000X dizi derinligi



Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA

Boris Rebolledo-Jaramillo^{a,1}, Marcia Shu-Wei Su^{b,1}, Nicholas Stoler^a, Jennifer A. McElhoe^c, Benjamin Dickins^d, Daniel Blankenberg^a, Thorfinn S. Korneliussen^{e,f}, Francesca Chiaromonte^g, Rasmus Nielsen^e, Mitchell M. Holland^c, Ian M. Paul^h, Anton Nekrutenko^{a,2}, and Kateryna D. Makova^{b,2}

Departments of ^aBiochemistry and Molecular Biology, ^bBiology, and ^gStatistics, ^cForensic Science Program, Pennsylvania State University, University Park, PA 16802; ^dSchool of Science and Technology, Nottingham Trent University, Nottingham NG1 4BU, United Kingdom; ^eDepartment of Integrative Biology, University of California, Berkeley, CA 94720; ^fCentre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, DK-1350 Copenhagen, Denmark; and ^hDepartment of Pediatrics, College of Medicine, Pennsylvania State University, Hershey, PA 17033

Edited by Michael Lynch, Indiana University, Bloomington, IN, and approved September 8, 2014 (received for review May 20, 2014)