

HMM

Sept 21, 2015

ANNOUNCEMENTS

HIDDEN MARKOV MODEL

HIDDEN MARKOV MODEL

- (a)
- | | |
|-------|-------|
| 1D8U | HAMSV |
| 1OJ6A | HIRKV |
| 2hhbB | HGKKV |
| 1FSL | HAEKL |
| 2MM1 | HGATV |

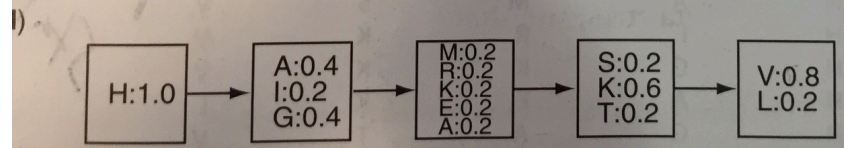
(b)

Probability	position				
	1	2	3	4	5
p(H)	1.0				
p(A)		0.4			
p(I)		0.2			
p(G)		0.4			
p(M)			0.2		
p(R)			0.2		
p(K)			0.2		
p(E)			0.2		
p(A)			0.2		
p(S)				0.2	
p(K)				0.6	
p(T)				0.2	
p(V)					0.8
p(L)					0.2

(c)

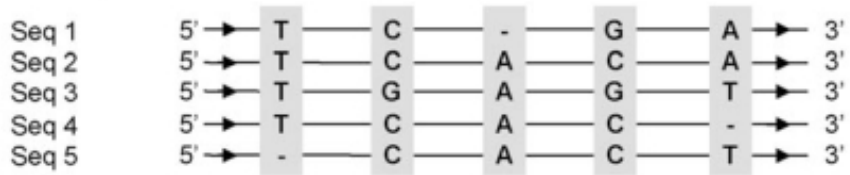
$$p(\text{HARTV}) = (1.0)(0.4)(0.2)(0.2)(0.8) = 0.0128$$

$$\text{Log odds score} = \ln(1.0) + \ln(0.4) + \ln(0.2) + \ln(0.2) + \ln(0.8) = -4.4$$

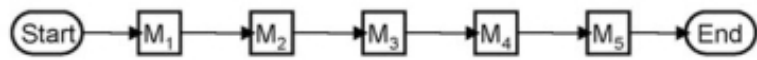


HIDDEN MARKOV MODEL

(a) Sequence Alignment



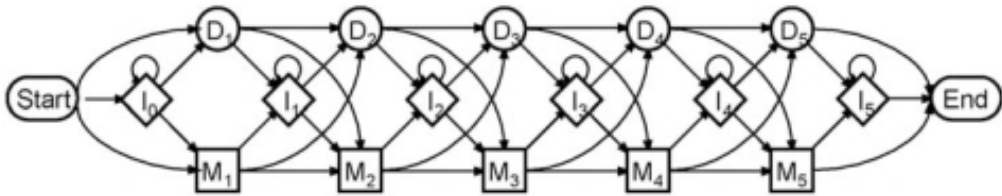
(b) Ungapped HMM



M_k Match states



(c) Profile-HMM

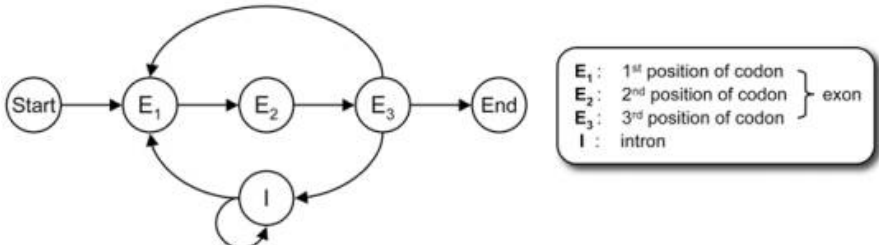


M_k Match states

I_k Insert states

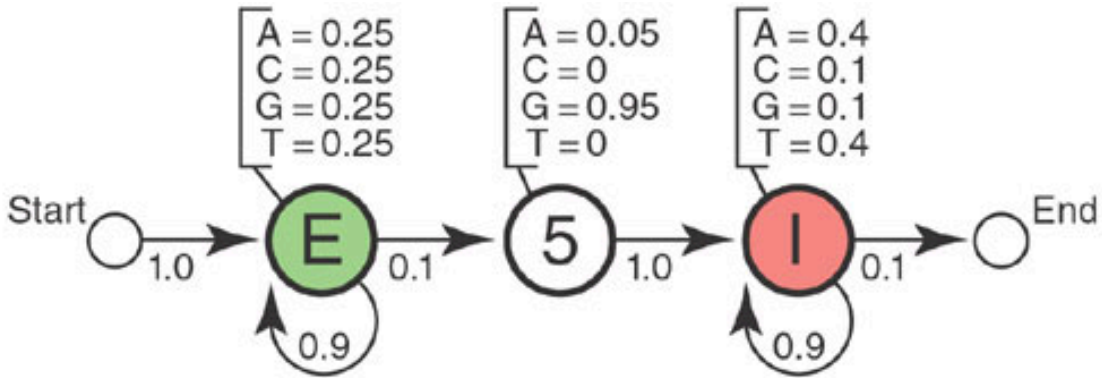
D_k Delete states

HIDDEN MARKOV MODEL

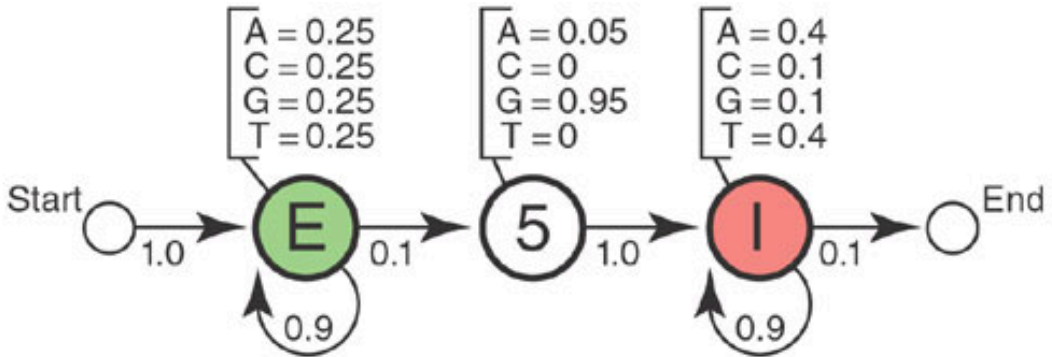


x	A	T	G	C	G	A	C	T	G	C	A	T	A	G	C	A	C	T	T	observed symbols
y	E ₁	E ₂	E ₃	E ₁	E ₂	E ₃	E ₁	E ₂	E ₃	I	I	I	I	E ₁	E ₂	E ₃	E ₁	E ₂	E ₃	hidden states
	exon									intron				exon						

HIDDEN MARKOV MODEL

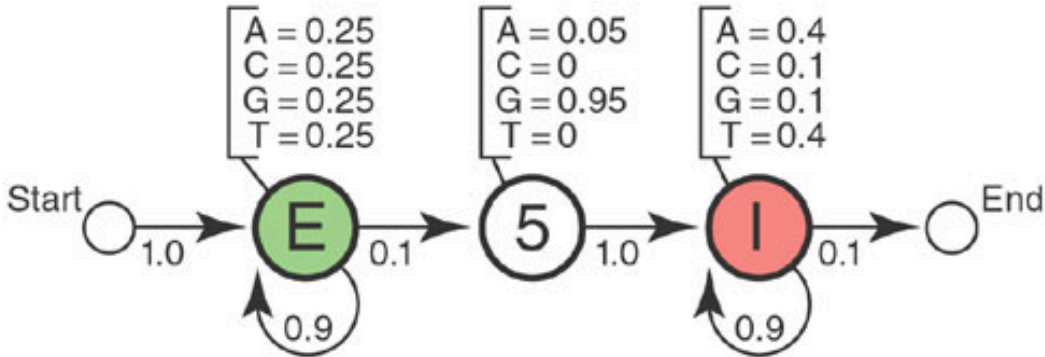


HIDDEN MARKOV MODEL



Sequence: C T T C A T G T G A A A G C A G A C G T A A G T C A

HIDDEN MARKOV MODEL



Sequence: **CTTCATGTGAAAGCAGACGTAAGTCA**



fastQ format and Illumina sequence data

fastQ

Description

Uses

fastQ

```
@HSQ-7001360:67:H88RHADXX:1:1101:1448:2158 1:N:0:CAGATC  
ATCTATCTGAGACTGATACGCCTTCGGCTTAATTTATAACAAG  
+  
BBBFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
```

fastQ

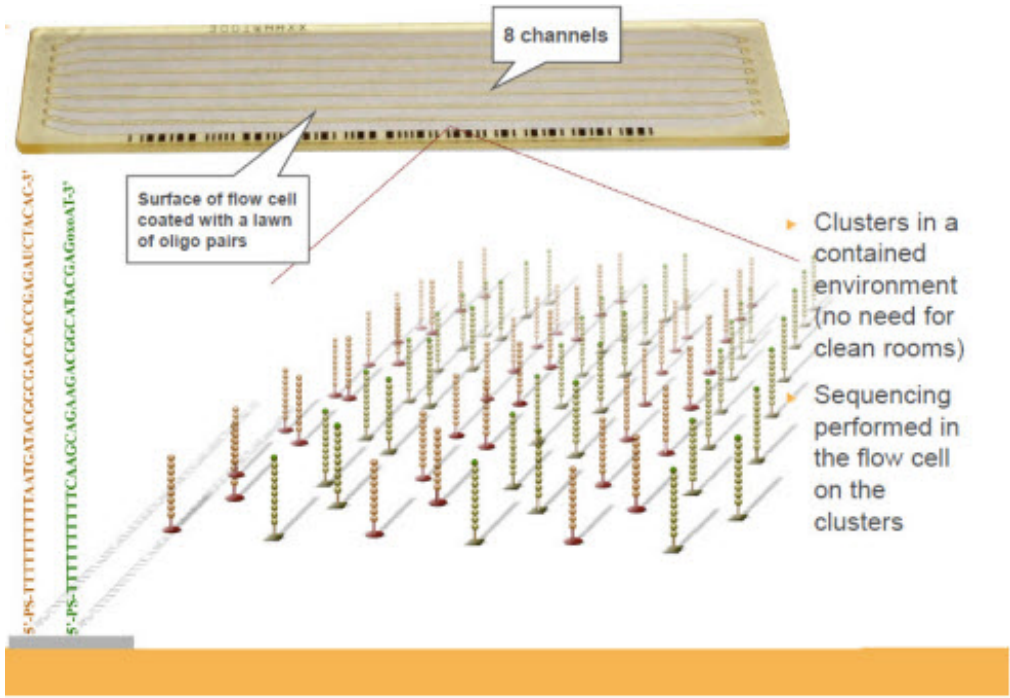
@HSQ-7001360:67:H88RHADXX:1:1101:1448:2158 1:N:0:CAGATC

- HSQ-7001360= Instrument name
- 67= run ID
- H88RHADXX=Flowcell ID
- 1=lane 1
- 1101=tile number
- 1448= x coordinate
- 2158= y coordinate
- 1=left read
- N=not filtered
- 0=control bit -> (not used anymore)
- CAGATC= adapter sequence

fastQ -> Illumina Seq

<https://youtu.be/womKfikWlxM>

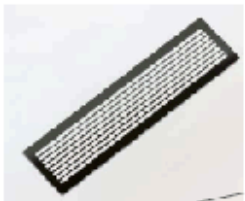
Illumina Seq



Illumina Seq

Technology Overview - GAI

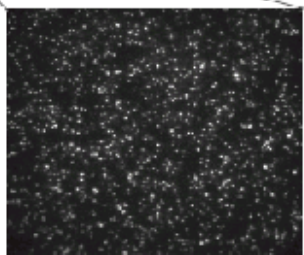
flow cell



A flow cell contains eight lanes



Tile



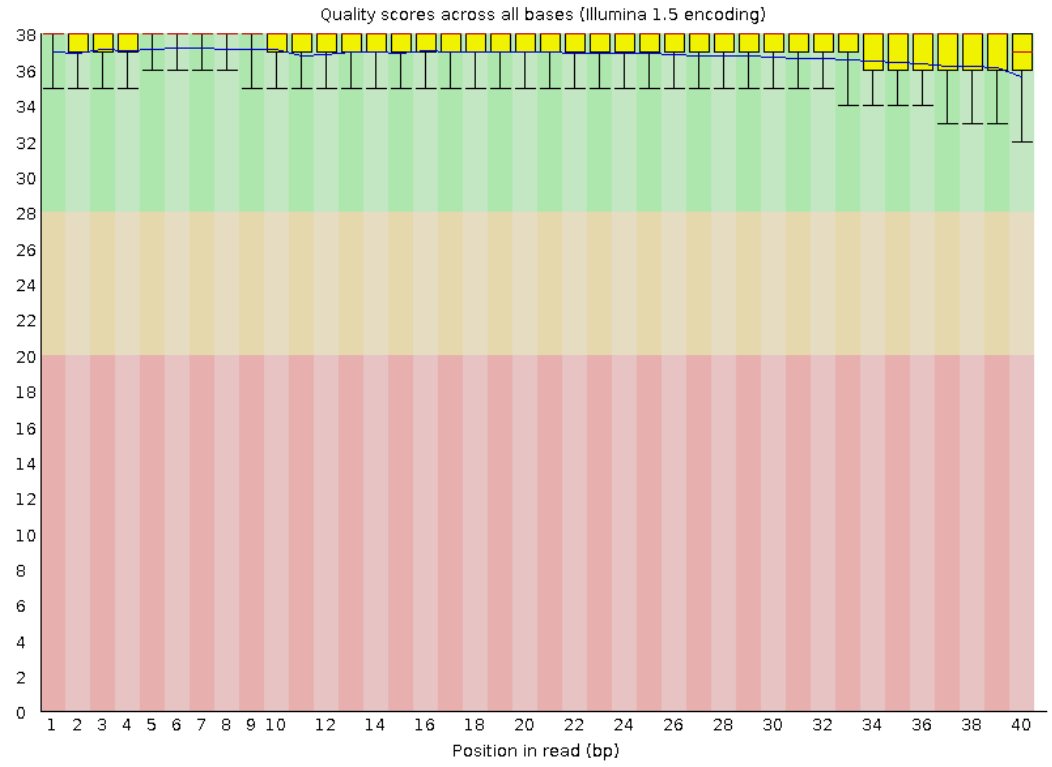
Illumina Seq

@HSQ-7001360:67:H88RHADXX:1:1101:1448:2158 1:N:0:CAGATC

- 1101=tile number
- 1448= x coordinate
- 2158= y coordinate

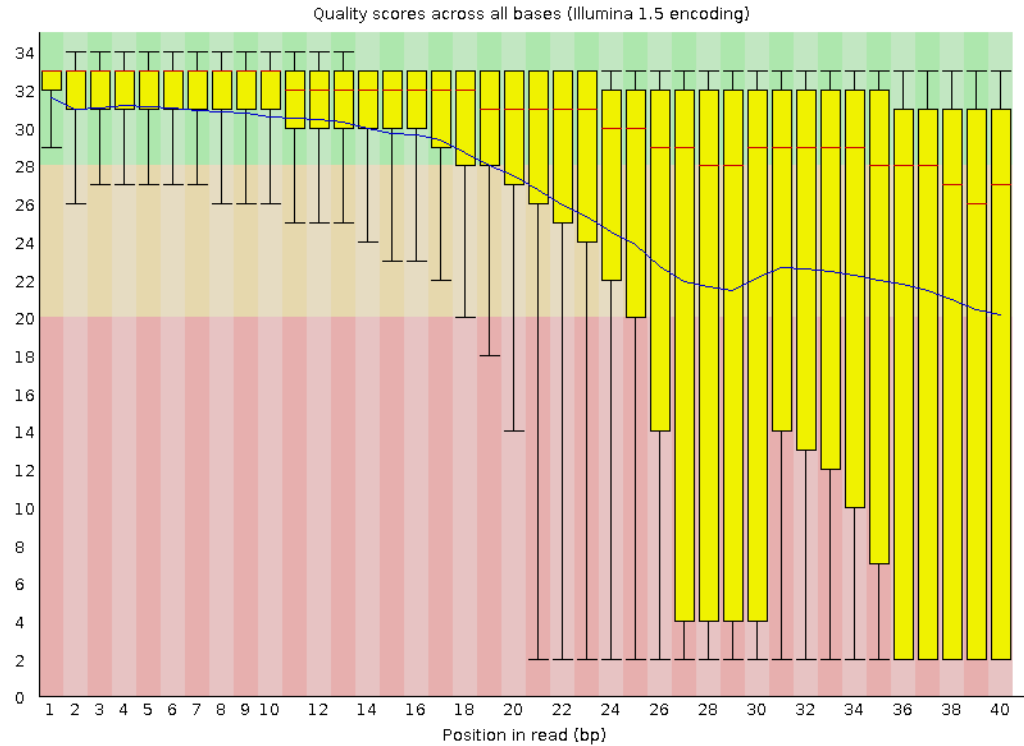
Illumina Seq

BBBFF



Illumina Seq

BBBFF



Illumina Seq

— Figure 4. Paired-End Sequencing and Alignment —

Paired-End Reads

