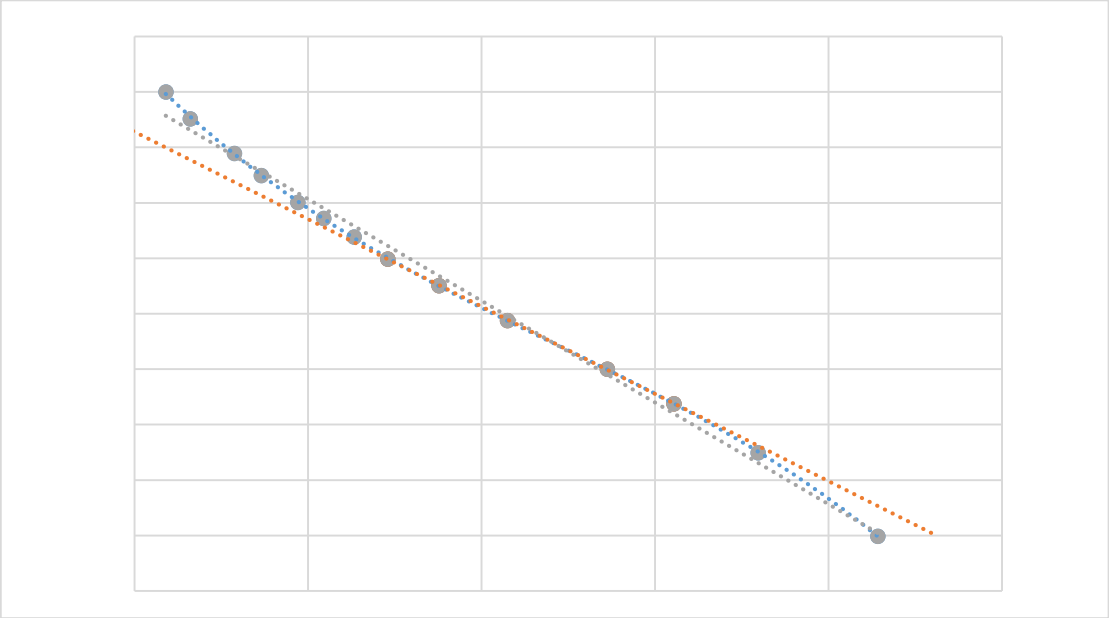
Estimating Sizes of λ-DNA fragments on an Agarose Gel with a Cubic function



Author: Feng Geng

Group No: 7

Tutor: Prof. Waksman

Lab Partner: Louise Nash

Abstract:

Isolating a piece of DNA into a replicative vector allows genome mapping and subsequent functional analysis. Here a genomic library is constructed against DNA of λ-bacteriophage for there is detailed knowledge about its genome. *Pst*I-digested λ DNA is inserted to pUC19 plasmid, and transformed into the DH5α strain of E. coli. Plasmid DNA is extracted from a transformed colony, of which a *Pst*I insert of around 2000bp was separated to produce a DIG-labelled probe. Hybridisation of the probe to a Southern Blot of λ DNA digestions with *Hind*III, *Eco* RI and *Mlu*I determined the genomic location of the insert to be 9777-11763bp. On encountering problem in interpreting the DNA migration on the gel, we discovered that cubic function is better at fitting log-transformed DNA size against migration distance. The difference between migration of size marker is also subject to interest.

Introduction:

Bacteriophage include different orders of viruses that specifically infect bacteria. The virus consists of protein capsid containing packed DNA, thus appears inadequate to replicate on its own. Upon infection, they usually follow a lytic pathway where bacterial molecules are metabolised to produce progeny viruses and the bacteria is lysed to allow further infection (1).

Bacteriophage λ is able to insert its DNA into bacterial chromosome and get passed along in this strain of bacteria. To establish such coordinated function, λ-phage employs a variety of gene regulation including transcriptional initiation, elongation and termination, cooperative binding and protein-nucleic acid interaction (2). This analysable complexity shed light in the early days of genomics with a transferable knowledge for all living organisms since they share a same set of genetic code.

Being temperate, a λ virus must choose between lysogenic cycle and lytic cycle upon infection. Previous study showed that cII protein plays a determinant role in that its activity is required to produce λ repressor that commit the virus to lysogenic phase (3). The level of cII is indicative of bacterial growth in that active growing bacteria produce more protease that would the susceptible cII protein. This sensing mechanism allows λ-phage to avoid over-lysis and to keep its lineage in a less aggressive manner. Understanding this regulation provides insight into how host-virus interaction can vary for different genetic system.

In the past, restriction mapping has been fundamental to characterise a genome and a restriction map is constructed by combination of partial digest, complete digest and double digest (4). λ-phage itself could be engineered to carry a fragment of up to 25kbp, compared to 15kbp by a plasmid, to serve as vector for a genomic library (5). Subsequent Southern blot against genomic library reveals location of a DNA of interest.

In this experiment, a genomic library of λ-DNA is constructed by transforming *Pst*I DNA fragment-bearing plasmids. Referencing to an established map of digestion

sites on λ-DNA (6), Southern blot of one *Pst*I fragment against 2 alternative digestion of λ-DNA reveals its location.

Multiple characterisations of DNA by electrophoresis are basic to DNA manipulation, assimilating any DNA mixture and reflecting action of any enzyme. Accurate and consistent estimate of DNA size is thus of importance. It was formulated that DNA mobility, which is proportional to drift speed in a uniform electrical filed, linearly decreases with increased log(DNA size/bp) (7). In practice, however, curvatures occur at the end of this linear relation. We subsequently developed a solution to better describe the relation between DNA migration, proportional to mobility, and log(DNA size/bp) , and to produce better prediction.

Results and Discussion:

Alkaline phosphatase (AP)-treated *Pst*I digest of pUC19 plasmid is hybridised to *Pst*I digests of λ-DNA followed by ligation. The *Pst*I enzyme is effective in both producing linear plasmid DNA and digesting λ-DNA into 14 visible (>500bp) pieces (2 bands for small DNA pieces are very faint), as confirmed by electrophoresis (see Figure 1). The uncut λ-DNA appeared to be pure and uncut plasmid is mostly in supercoiled form. The treatment of *Pst*I by AP reduces the chance of pUC19 religating to itself.

On the gel photo concerning the digestion, *Pst*I digest of λ-DNA shows a perfect linear relation (see Figure 2) between its log-transformed DNA length (obtained from literature and assigned by inspection (6)) and its migration distance, consistent on different lanes (see figure 1). The size marker, however, follow a linear trend at smaller pieces, with consistently smaller migrations (see Figure 2). That linear trend is violated for larger pieces, where its divergence from *Pst*I/λ-DNA is reduced. The unusual migration possibly results from the use of Gel-Red stain instead of ethidium bromide, as suggested by Prof. Shephard.

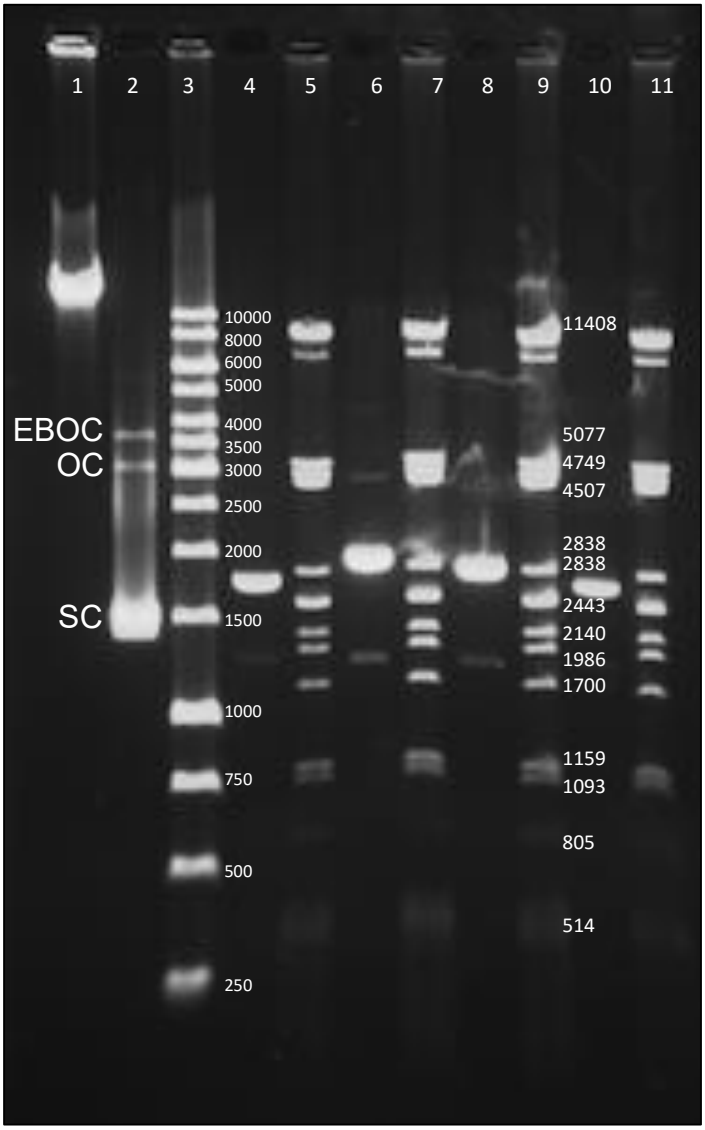


Figure 1: Electrophoresis concerning the digestion of DNA to ligate. Lane 1-Uncut λ-DNA; Lane 2-Uncut pUC19; Lane 3-DNA marker; Lane 4, 6, 8,10-PstI digestion of pUC19; Lane 5,7,9,11-PstI Digestion of λ-DNA. SC: supercoiled; OC: Opencircular plasmid; EBOC: enzyme-bound opencircular. Literature fragment Literature sizes (6) *are assigned to PstI digest of λ-DNA by inspection.*

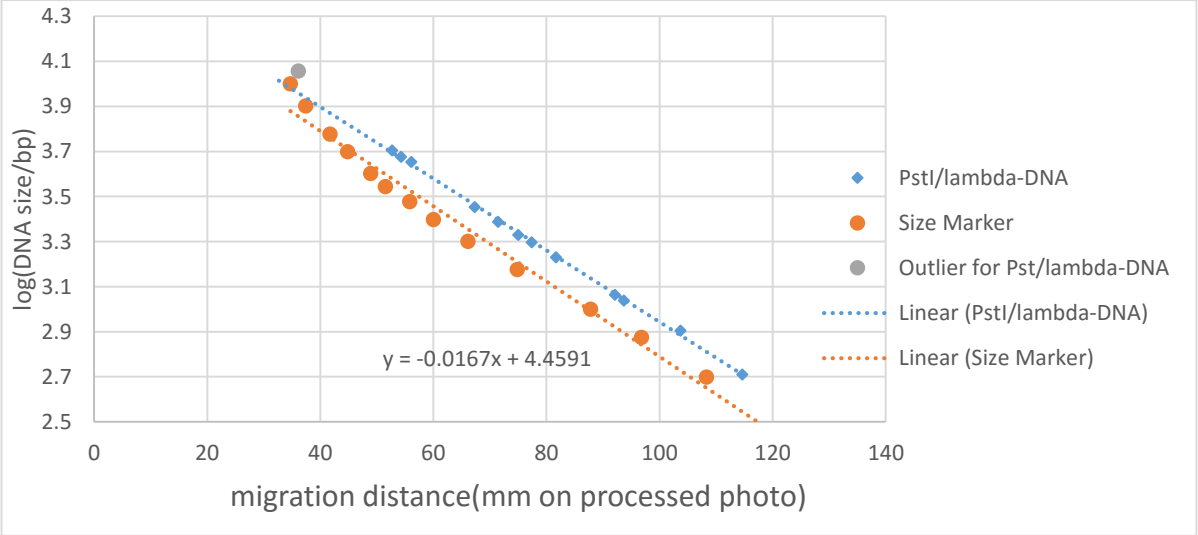


Figure 2 Calibration curves for size marker and PstI/λ-DNA on gel photo for PstI digestion. The outlier piece is too big (>10,000bp) to follow a linear relation.

The recombinant plasmids are then transformed to DH5α E. coli and plated on to nutrient agar containing ampicillin and IPTG, selecting for pUC19 expression and allowing differential expression of *lac* operon which results a blue colour for the colony containing plasmid with an intact α peptide-encoding *lac* operon and a white colour for the one with its *lac* operon disrupted by the insert. The blue colour is from the processed X-gal, which is present in the medium as well.

As for controls, transformation with uncut pUC19 shows that the E. coli cells are competent to take up and express pUC19 to acquire ampicillin resistance. Religated AP-treated plasmid yields a lower colony counts of around 10^3 per ug of plasmid, possibly due to transformation by undigested supercoiled pUC19 (see figure 1). The white colonies could be due to dimerised plasmid of linearized plasmid DNA. Overall, the action of did not result in a pure collection of linear plasmid DNAs. Transformation of plasmid/λ-DNA fragment ligation resulted in a higher colony count than that for religated pUC19, of around 10^4 per ug of plasmid added, with approximately 1:1 ratio for blue and white colonies.

pUC19 content (ng)	DNA transformed	Number of white colonies	Number of Blue colonies	Total Number of colonies per 1ug of pUC19 DNA
0.1	Uncut pUC19	0	103	1.03x10^6
45	*Pst*I cut, phosphatased and religated pUC19	2	127	2.87x10^3
4.5	*Pst*I cut, phosphatased pUC19, ligated to *Pst*I digest of λ	21	25	1.02x10^4

Table 1 Transformation result. Serial dilution is done to ensure a suitable number of colonies (not shown).

We extracted number 22 and number 39 white colonies (WC) and a blue colony(BC) supplied by the lab organiser for plasmid DNA following a mini-prep protocol that includes cell lysis, trapping DNA by silicon membrane, wash-off of chromosomal DNA and collection of plasmid DNA. Confirmed by electrophoresis see (figure 3), the extracted plasmids from WC22, WC39 and BC are of good purity. WC22 plasmid contains an insert of around 2000bp, and WC39 plasmid contains one around 2400bp, as confirmed by table 2 (see Appendix).

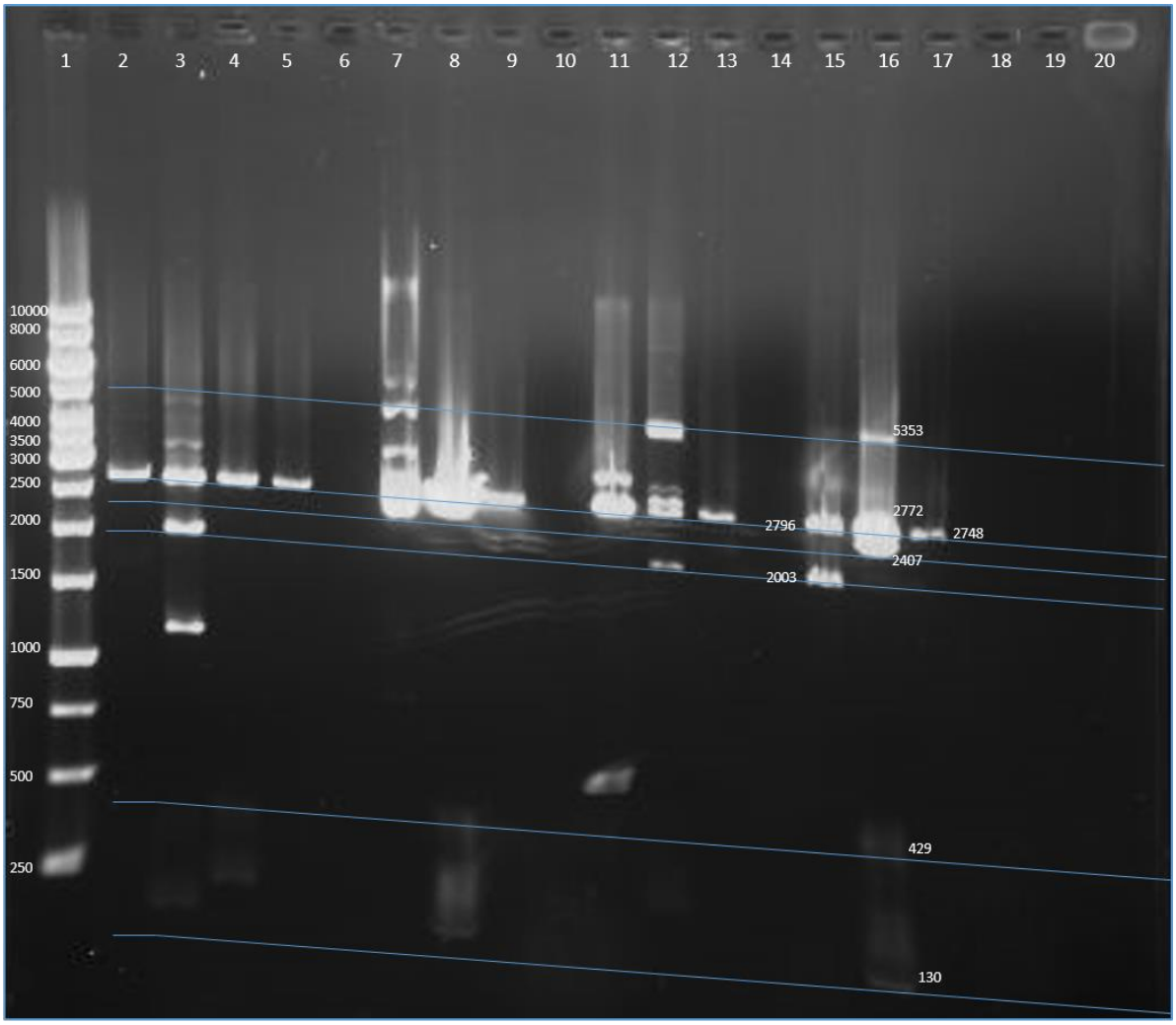


Figure 3 Gel photo for PstI digestion of plasmid extractions. The bands appear to sit at an angle to the axis determined by the alignment of the band front, possibly resulting from heterogeneity in either the gel concentration due to settlement or in electric potential. This is fixed by plotting a set of lines along the axis of inclination, and measuring the equivalent migration in lane 2. Content: Lane 1-Size Marker; Lane 2-linear pUC19. The rest lanes all run PstI digests of plasmid extraction from: WC22 for lane 15, WC39 for lane 16, unknown WC for lane 3, 4, 7, 8, 11, 12, BC for lane 5, 9, 13, 17. Size estimation for digests are produced by cubic fitting, see 'refined size' in table 2

We initially fitted a straight line to points to predict the size of DNA (see 'linear fit for full range' in figure 4). The straight line, however, does not fit the points perfectly, which are distributed over a curve by inspection (see figure 4). The error in fitted value for this method appears to be unacceptably high (>10%) between 3500bp and 2500bp (see 'Analysis of Error' in table 2). To improve the accuracy for this range, we plotted a standard curve using the central linear-distributed points by inspection, by which prediction yields acceptable error (<5%) only between 3500bp and 500bp (see 'Analysis of Error' in table 2). This improved prediction estimated the size of pUC19 and insert in WC22 accurately but failed to estimate outside this range.

We thus further examined the distribution of the error (or residual) produced by the central linear fit (CLF) by plotting the error against the migration distance (see figure 5), yielding a curve rather than following a normal distribution, indicating the error is systematic. A cubic polynomial fits perfectly with the error (see figure 5). A cubic function is thus applied to predict the systematic error by CLF and refined our prediction of DNA size (see 'refined size' in table 2). This improved prediction allowed me to conclude that in lane 16 (see figure 3), DNA for 5353 band is the combination of those for 2772 band and 2407 band, which would not be the case if sizes were estimated with a linear fit (see grey entries in table 2). Accurate size prediction is essential to identifying the fragment with confidence.

Considering the linear combination of the two functions CLF and error fixing, we successfully fitted a cubic function to the original distribution (see 'Cubic fit' in figure 4). On this discovery, we examined the standard curve for other electrophoreses and confirmed that cubic function fitted well with data depicted in figure 2 as well (see figure 5). The curvature at high DNA size could result from conformational change at high voltage (8), while we did not find any reason for the one at low DNA size.

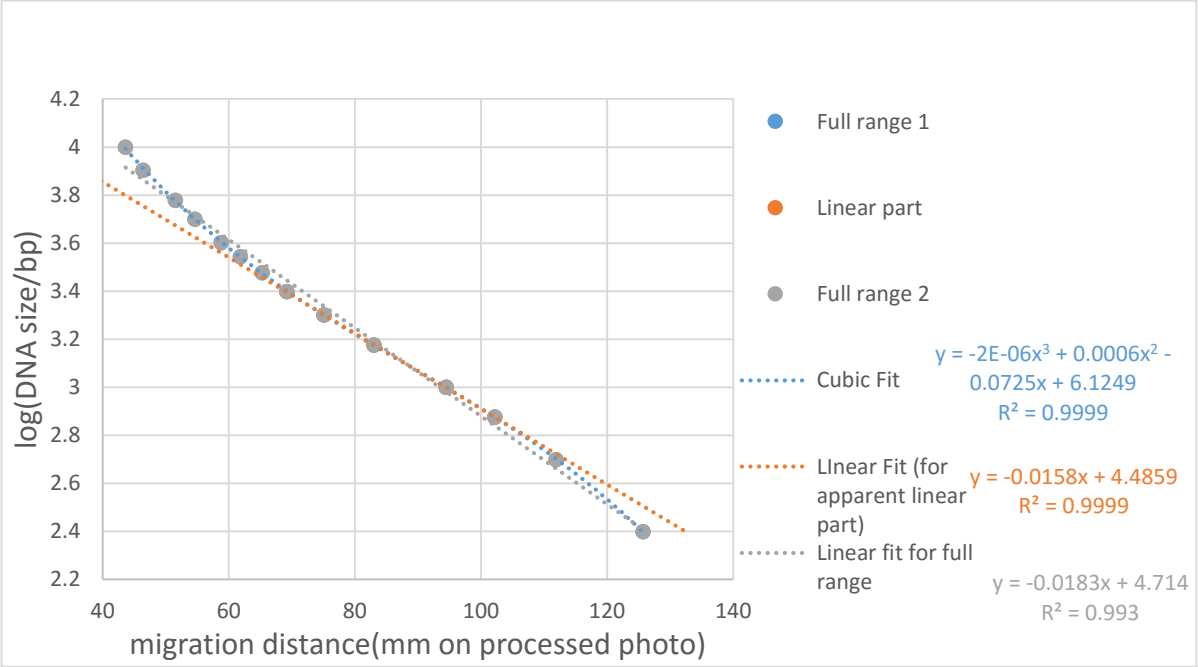


Figure 4 Comparison of different fits on the size prediction of the DNA ladder on figure 2, gel for plasmid extraction, where cubic function shows best fit.

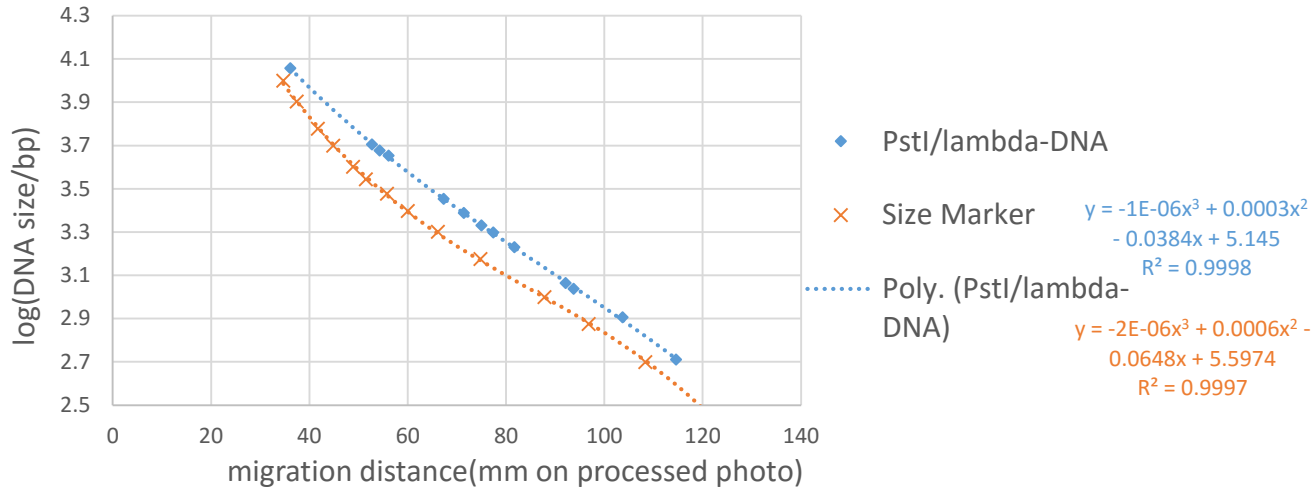
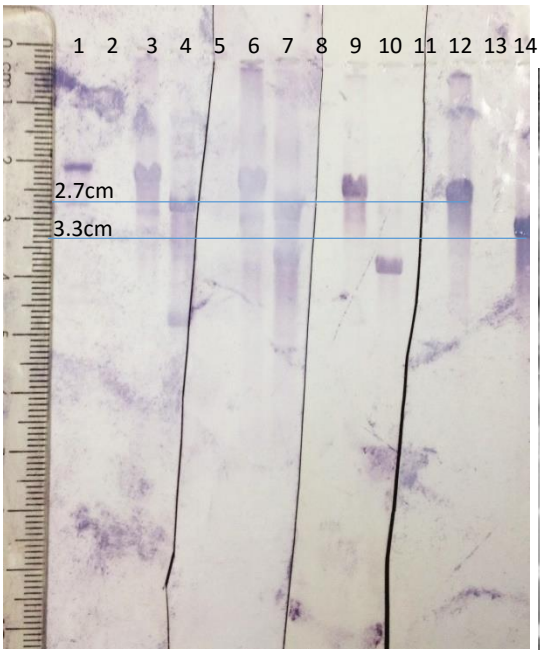


Figure 5 Using cubic function to refit figure 3. The PstI/λ-DNA curves less than size marker does, where a linear fit is enough to estimate the sizes.

We then prepared a DIG-labelled DNA probe from WC22 plasmid, and performed a hybridisation against a Southern blot of *Eco* RI and *Mlu*I digests of λ-DNA (see Figure 6). Purple colour is developed on the hybridised Southern blot by adding AP-bound anti-DIG antibody, and substrates including Nitro-blue tetrazolium (NBT) and 5-bromo-4-chloro-3 indolyl phosphate (BCIP) under alkaline condition. To identify the band of hybridisation, we find the correspondent band on gel by measuring migration on the developed blot (see figure 7). We then assigned literature fragment size to each band by inspection since the size marker is invalidated, migrating at a different scale (see Lane 1 in Figure 7) possibly due to Gel-Red. The assignment is examined by plot standard curves for each of 3 lanes including 1,12 and 14 (see Figure 8), good fit is obtained for all of 3 digests, while *Hind*III digest follows a distinct distribution. Colour localisation shows at 21226 fragment of *Eco* RI and 9824 fragment of *Mlu*I.

(A)



(B)

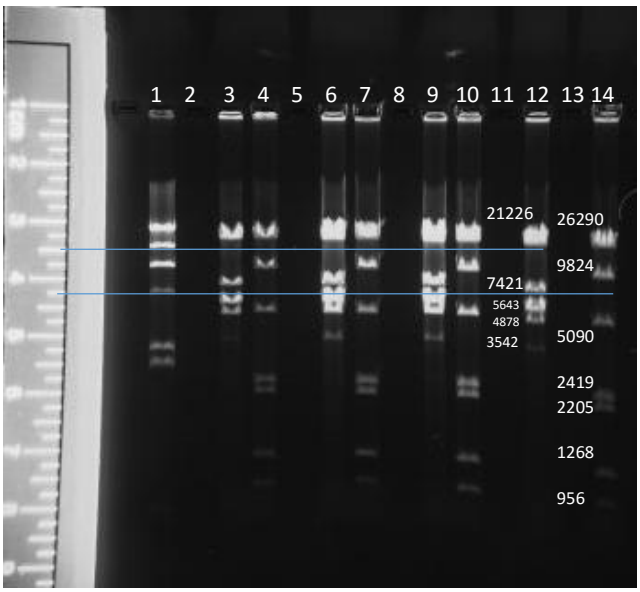


Figure 7 (A) Developed Southern blot (B) Gel photo for Southern blot. Lanes in (A) are numbered accordingly as in All lanes contain digests of λ-DNA. 1 contains HindIII digest; 3, 5,9,12 contain Eco RI digest; 4, 7, 10, 14 contain MluI digest. Literature sizes are assigned by inspection(6)*. (A) The colour bands are thick in lane 12 and 14, possibly due to late termination of the colour development reaction allowing the chromophore to diffuse. There is also random splash of colour, indicating unspecific binding of AP-bound antibody or DIG-labelled DNA.*

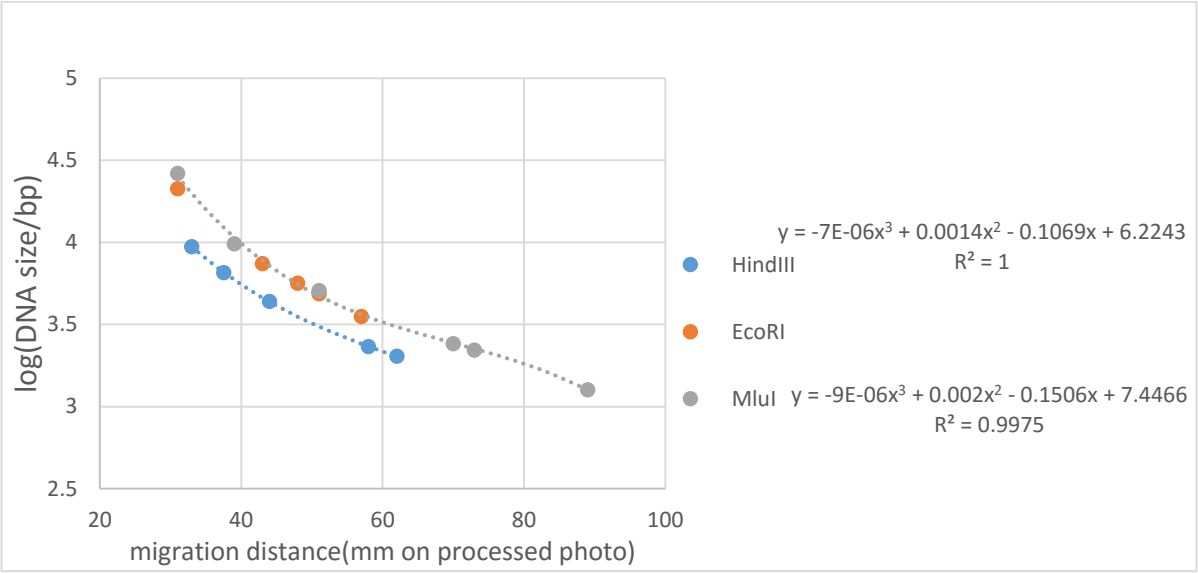


Figure 8 Standard curves for Southern blot gel.

On the genome of λ-DNA, 9824 fragment lies within the 21226 fragment (See table 3). Within that length of DNA, covering 5548-15372bp, only 1 *Pst*I fragment is about 2000bp, covering 9777-11763bp. This fragment is 1986bp long. According to an established genomic map (6), this piece is involved in synthesis and assembly of capsid components.

Conclusion:

We successfully transformed λ-genomic fragment into DH5α E. coli using pUC19 as a vector, at an acceptable colony density and able to extract recombinant plasmids from white colonies. The Southern blot yields extra information about the relative location on querying an established genomic map. Interestingly, the mobility of size marker is self-consistent if fitted by a cubic function despite showing reduced migration in 2 of the 3 gel, possibly due to choice of staining. This fragment identification applicable to nearly any DNA, thus essential to the analysis of the genome and any biochemical probing of a biological system.

References:

1. Bacteriophage page on Wikipedia [Internet]. [cited 2015 Nov 22]. Available from: https://en.wikipedia.org/wiki/Bacteriophage

2. Friedman DI, Court DL. Bacteriophage λ: alive and well and still doing its thing. Curr Opin Microbiol [Internet]. 2001 Apr [cited 2015 Nov 23];4(2):201–7. Available from:

http://www.sciencedirect.com/science/article/pii/S1369527400001892

3. Hyde D. Gene Expression: Control in Bacteria and Phages. In: Introduction to Genetic Principles. 1st ed. New York: McGraw-Hill; 2009. p. 543–51.

4. Hyde D. R*Eco*mbinant DNA Technology. In: Introduction to Genetic Principles. 1st ed. New York: McGraw-Hill; 2009. p. 372–9.

5. Genomic library [Internet]. [cited 2015 Nov 22]. Available from: https://en.wikipedia.org/wiki/Genomic_library

6. Unknown. Appendix. In: BIOC2001 Research Project Instructions. London: UCL Div of Biosciences; 2015. p. 48,47,43.

7. Stellwagen NC, Stellwagen E. Effect of the matrix on DNA electrophoretic mobility. 2010;1216(10):1917–29.

8. Southern EM. Measurement of DNA length by gel electrophoresis. Anal Biochem [Internet]. 1979 Dec [cited 2015 Nov 16];100(2):319–23. Available from: http://www.sciencedirect.com/science/article/pii/0003269779902355

Appendix:

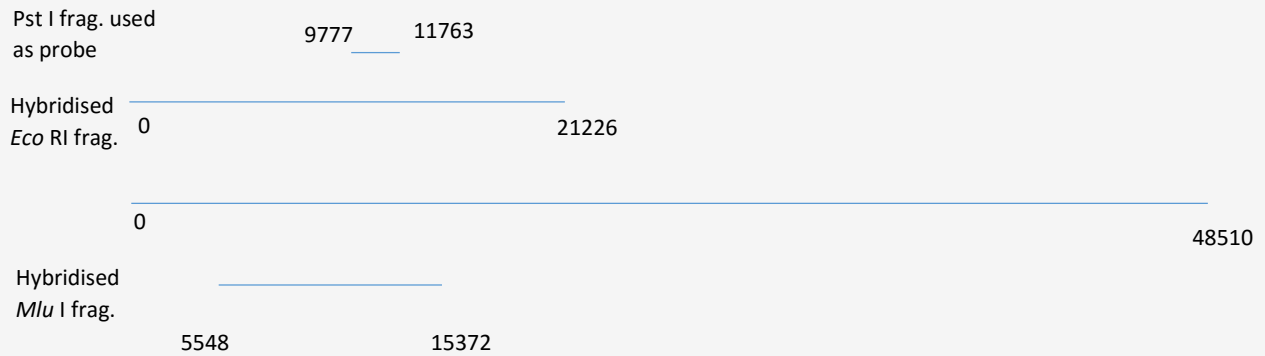
Table 2 Comparison of prediction accuracy by three different standard curves

		Central Linear fit	y = -0.0158x + 4.4859	Fitting the error		
Content	**migration distance(mm)**	**Fitted log(size/bp)**	**Estimated size(Da)**	**Actual size/bp**	**Actual log(size/Da)**	**Error in linear fit**
Size marker	43.6	3.80	6266	10000	4.00	0.203
	46.4	3.75	5660	8000	3.90	0.150
	51.5	3.67	4701	6000	3.78	0.106
	54.6	3.62	4200	5000	3.70	0.076
	58.8	3.56	3605	4000	3.60	0.045
	61.8	3.51	3232	3500	3.54	0.035
	65.3	3.45	2846	3000	3.48	0.023
	69.2	3.39	2469	2500	3.40	0.005
	75.1	3.30	1992	2000	3.30	0.002
	83	3.17	1495	1500	3.18	0.002
	94.5	2.99	984	1000	3.00	0.007
	102.2	2.87	743	750	2.88	0.004
	112.3	2.71	515	500	2.70	-0.013
	125.7	2.50	316	250	2.40	-0.102
WC22	66.7	3.43	2704	(pUC19) 2686	3.44	
	75	3.30	1999	(confirmed insert) 1986	3.30	
WC39	53.4	3.64	4387		3.65	
	66.9	3.43	2685	(pUC19) 268tw0	3.43	
	70.3	3.38	2372		3.38	
	115.5	2.66	458		2.65	
	135.6	2.34	221		2.32	
BC	67.1	3.43	2665		3.43	

Cubic fit function for error	$y = -2.4939E-06x^3 + 6.6055E-04x^2 - 5.7651E-02x + 1.6611E+00$		Full range linear fit	$y = -0.0183x + 4.714$	Analysis of Error		
Fitted error with cubic function	Refined log(size/bp)	Refined Size/bp	Fitted log(size/bp)	Fitted DNA size/bp	%error by linear fit	%error after refinement	%error for full range linear fit
0.196	3.994	9852	3.916	8244	-37.34%	-1.48%	-17.56%
0.159	3.912	8164	3.865	7326	-29.26%	2.04%	-8.42%
0.103	3.776	5964	3.772	5909	-21.65%	-0.59%	-1.51%
0.077	3.700	5010	3.715	5186	-16.01%	0.20%	3.72%
0.048	3.605	4026	3.638	4345	-9.88%	0.65%	8.62%
0.032	3.542	3483	3.583	3829	-7.66%	-0.50%	9.39%
0.019	3.473	2971	3.519	3304	-5.15%	-0.97%	10.13%
0.008	3.401	2517	3.448	2803	-1.24%	0.69%	12.12%
0.001	3.300	1995	3.340	2186	-0.39%	-0.23%	9.30%
0.001	3.175	1497	3.195	1567	-0.37%	-0.22%	4.47%
0.007	3.000	1000	2.985	965	-1.64%	0.03%	-3.47%
0.006	2.878	754	2.844	698	-0.90%	0.56%	-6.96%
-0.015	2.697	498	2.659	456	2.94%	-0.48%	-8.81%
-0.102	2.398	250	2.414	259	26.44%	0.02%	3.69%
0.014	3.446	2796	3.493	3115	0.68%	4.08%	15.95%
0.001	3.302	2003	3.342	2195	0.67%	0.85%	10.54%
0.086	3.729	5353	3.737	5455			
0.014	3.443	2772	3.490	3088			
0.006	3.381	2407	3.428	2676			
-0.028	2.633	429	2.600	398			
-0.229	2.115	130	2.233	171			
0.013	3.439	2748	3.486	3062			

Table 3 Position of PstI, EcoRI, HindIII sites in λ-DNA and sizes of fragments generated , excerpt from (6)

Pst I		**Eco R1**		**Hind III**		**Mlu1**	
Site	Size of Fragment	Site	Size of Fragment	Site	Size of Fragment	Site	Size of Fragment
0 L end		0 L end		0 L end		0 L end	
2556	2556 bp	21226	21226 bp	23130	23130 bp	458	458bp
2820	264	26104	4878	25157	2027	5548	5090
3625	805	31747	5643	27479	2322	15372	9824
3640	15	39168	7421	36895	9416	17791	2419
3856	216	44972	5804	37459	564	19996	2205
4370	514	48510 R end	3542	37584	125	20952	956
4709	339			44141	6557	22220	1268
4909	200			48510	4361	48510	26290
5120	211						
5214	94						
5682	468						
8520	2838						
9613	1093						
9777	164						
11763	1986						
11835	72						
14294	2459						
14381	87						
16081	1700						
16231	150						
17390	1159						
19833	2443						
20281	448						
22421	2140						
26928	4507						
32005	5077						
32252	247						
37001	4749						
48510 R end	11509						



Rescaled photo where DNA migrations are measured. Figure 1

Rescaled photo where DNA migrations are measured. Figure 3