Epidem. Inf. (1988), **100**, 467–479
Printed in Great Britain

Computer matching of oligonucleotide patterns on electrophoretic gels: an application to the epidemiology of cytomegalovirus

By A. E. ADES,[1] A. J. GARRETT,[2] J. COWELL,[3] K. S. CHIN[3] and C. S. PECKHAM[1]

[1] *Department of Paediatric Epidemiology, Institute of Child Health, 30 Guilford St, London WC1N 1EH*

[2] *National Institute for Biological Standards and Control, Holly Hill, London NW3 6RB*

[3] *Department of Child Health, Charing Cross Hospital Medical School, London W6 8RP*

(*Accepted 2 December 1987*)

SUMMARY

A computer program was written to analyse oligonucleotide patterns displayed by gel electrophoresis following restriction endonuclease digestion of human cytomegaloviral DNA, and was applied to an epidemiological study of the transmission of infection in a hospital special care baby unit, with regard to infant-to-infant and mother-to-infant transmission.

The program calculates the molecular weight of oligonucleotides from their mobilities, using a cubic spline curve based on the mobilities of oligonucleotides from the AD169 strain. A matching algorithm then calculates the number of unmatched fragments for each pair of viral isolates. This was used as a similarity measure which successfully distinguished mother and infant isolate pairs from epidemiologically unrelated pairs.

The program is not intended to provide fully automatic matching, but could be recommended as a screening device to pick out pairs of strains which are sufficiently similar to suggest a common source of infection, and which may warrant closer comparison. Other applications are discussed, and the possible use of densitometers to automate data entry is considered.

INTRODUCTION

Restriction endonuclease analysis of cytomegalovirus (CMV) DNA, first developed by Huang and co-workers (Huang *et al.* 1976; Huang *et al.* 1980), has been used to characterize viral isolates and has provided information on patterns of viral transmission (Wilfert, Huang & Stagno, 1982; Yow *et al.* 1982; Dworsky, Lakeman & Stagno, 1984; Adler, 1985; Grillner & Strangert 1986). Garrett & Warren (1985) described a modification of this technique which is simpler, requires

Address for correspondence: Dr A. E. Ades, Department of Paediatric Epidemiology, Institute of Child Health, 30 Guilford Street, London WC1N 1EH.

less virus, and which has been used by Peckham *et al.* (1986) to compare cultures of viral isolates from several epidemiologically related groups.

Viral DNA isolated from urine or throat swabs is labelled with ^{32}P and then digested with restriction endonuclease. The resulting oligonucleotides are separated by electrophoresis on slab gels, and made visible as banding patterns by autoradiography. These patterns can then be used to compare strains and thereby provide evidence for common source infections.

Comparison of the banding patterns on the gels in the study by Peckham *et al.* (1986) was done by eye, a long and tedious process, and the need for a computer program which could select potential matching pairs of isolates became apparent. In a study with only 25 isolates, for example, there are 300 pairings to be inspected for each restriction enzyme, and 1225 pairings for 50 isolates.

This paper describes a program that estimates the molecular weight (MW) of nucleic acid fragments from their gel mobility, using fragments of the laboratory strain AD169 run on the same gel as a reference. The second part of the program matches each pattern of MWs to every other, and generates a measure of similarity that can be used to help decide whether two samples should be run on the same gel to confirm a match.

MATERIALS AND METHODS

Source of isolates

Twelve CMV isolates were grown from throat swabs, urine or breast milk, as part of a study on possible horizontal spread of infection within a Special Care Baby Unit (SCBU) in a West London Hospital. Some of the isolates were included in the study by Peckham *et al.* (1986). Ten isolates were from infants in the SCBU, including one who was congenitally infected, and two were from mothers of infected infants. The results analysed in this paper were obtained from digests of 12 isolates with *Eco* R1 and with *Bgl* II restriction endonucleases. The resulting cleavage patterns had earlier been examined 'blind' by one of the authors (A.J.G.) who had found that only two pairs of isolates matched, and these had originated from mother and infant pairs.

Duplicates of two of the isolates digested with *Eco* R1 were run on different gels. There was therefore a total of 14 patterns of fragments following *Eco* R1 digestion run on 9 separate gels, and 12 following *Bgl* II run on 6 separate gels. Each gel included a track of the digested fragments from the laboratory strain AD169.

Biochemical technique

Briefly, the cytomegalovirus strains were propagated in human embryonic lung cells in the presence of ^{32}P. A crude extract of DNA was digested with restriction endonuclease, and the resulting oligonucleotides were separated by agarose gel electrophoresis. The gel was dried and the oligonucleotide patterns were visualized on an autoradiograph. A detailed description has already been published (Garrett & Warren, 1985).

Table 1. *Sequence of processing steps for the AD169 pattern on track 1 of Figure 1*

Band	Mobility* (mm)	Code†	MW‡	MW§	Log scale¶
1	30·5	FG	8·3	8·28	11·95
2	33·0	HI	7·4	7·42	16·01
3	35·5	J	6·8	6·78	19·32
4	37·5	K	6·3	6·34	21·82
5	48·0	MN	5·0	4·92	31·21
6	52·0	OP	4·5	4·49	34·53
7	54·5	Q	4·3	4·32	36·01
8	60·0	R	3·9	3·95	39·29
9	68·5	ST	3·45	3·46	44·20
10	79·5	U	2·85	2·93	50·37
11	82·5	VW	2·8	2·81	51·92
12	89·5	X	2·6	2·52	55·90
13	91·0	Y	2·5	2·47	56·63
14	93·0	Z	2·4	2·39	57·82
15	96·5	a	2·3	2·27	59·82
16	107·0	b	1·9	1·93	65·69
17	118·0	c	1·65	1·63	71·97
18	140·5	d	1·1	1·12	85·97
19	145·0	e	1·0	1·02	89·14
20	150·0	f	0·95	0·94	92·49

* Distance of band from origin in millimetres.

† Letter code for MW used by Spector, Hock & Tamashiro (1982).

‡ Molecular weight given by Spector, Hock & Tamashiro, in Daltons ($\times 10^{-6}$). Band pairs F and G, H and I, M and N, S and T, V and W could not be distinguished on any of our autoradiographs. Each was therefore taken as a single band and ascribed the average MW.

§ Average fitted molecular weight over all AD169 patterns.

¶ Log scale transformation of molecular weight (see text for details).

Data entry

The distance of every band from the origin, its mobility, was measured with a ruler to the nearest 0·5 mm. In the case of AD169 patterns, each band was given a single letter code which identified the MW of the fragment it represented. The letter codes and MWs given by Spector, Hock & Tamashiro (1982, table 1) were used.

Visual examination of the gels suggested that many of the fragments found on test patterns were apparently homologous to fragments on the AD169 track. If both of the authors (A.A and J.C) who measured the gels judged individual fragments in the test track to be homologous with AD169 fragments, they were given the appropriate letter code. Table 1 shows typical data for an AD169 pattern and Table 2 a test pattern from the same gel. Column 2 of Tables 1 and 2 lists the mobility data derived from tracks 1 and 2 in Fig 1 in the form used for computer analysis, and column 3 shows the letter codes.

Bands at the extreme ends of the pattern that were poorly separated or indistinct due to over- or under-exposure were not entered: the program does not impose limits on the MW scale where data must begin or end.

Table 2. *Processing steps for the test pattern on track 2 of Figure 1*

Band	Mobility* (mm)	Code†	MW‡	MW§	Log scale¶
1	31·0	FG	30·5	8·28	11·95
2	33·0		32·5	7·57	15·29
3	34·5		34·0	7·15	17·38
4	36·0	J	35·5	6·78	19·32
5	41·5		41·2	5·71	25·71
6	44·0		43·8	5·37	27·98
7	48·0	MN	48·0	4·92	31·21
8	53·0		52·8	4·47	34·76
9	54·5		54·2	4·34	35·83
10	60·5	R	60·0	3·95	39·29
11	63·5		63·0	3·77	41·00
12	69·5	ST	68·5	3·46	44·20
13	80·0		79·5	2·93	50·37
14	82·5		82·0	2·83	51·65
15	91·5		91·0	2·47	56·63
16	93·5	Z	93·0	2·39	57·82
17	108·0	b	107·0	1·93	65·69
18	118·5	c	118·0	1·63	71·97
19	133·5		130·0	1·28	80·98
20	141·0	d	140·5	1·12	85·97
21	145·0		144·5	1·03	88·79
22	150·0		149·5	0·94	92·17

* Distance of band from origin in millimetres.

† Letter code for MW used by Spector, Hock & Tamashiro (1982).

‡ Corrected mobility, so that bands of known MW match up with the distance from origin of their counterparts in the AD169 (see Table 1), and intervening segments are linearly stretched or compressed.

§ MW in Daltons ($\times 10^{-6}$) obtained by solving a cubic spline equation, relating the mobilities to the MWs of the AD169 fragments in this gel, at the corrected mobilities in column 3.

¶ Log scale transformation of molecular weight (see text for details).

Establishing a molecular weight scale

The molecular weights for the AD169 *Eco* R1 fragments estimated by Spector, Hock & Tamashiro (1982) are given in column 4 of Table 1. Graphs of these MW against mobility for AD169 suggested that their relationship was generally reciprocal, but there was significant variation from this form of curve. Therefore MW was fitted by polynomial curves in 1/mobility. No appreciable improvement in fit was found beyond a fourth degree polynomial. For *Eco* R1, for instance, the curve fitting yielded for each of the nine AD169 patterns a set of 'expected' MWs, which are the MWs predicted from the polynomial equation applied to the mobility. The variation among these predicted MWs within each fragment was used to assess the accuracy of the measurement system.

The standard deviation of measurement, calculated from the nine AD169 tracks was $0{\cdot}019 \times 10^{-6}$ Daltons for *Eco* R1 and $0{\cdot}028 \times 10^{-6}$ for *Bgl* II. This implies that 95% of fragments will be located within about $0{\cdot}05 \times 10^{-6}$ Daltons of their true value. At the same time it was noticed that the average predicted MW of some fragments showed statistically significant deviation from the original values of Spector, Hock & Tamashiro (1982). This variation appeared to be random, in the

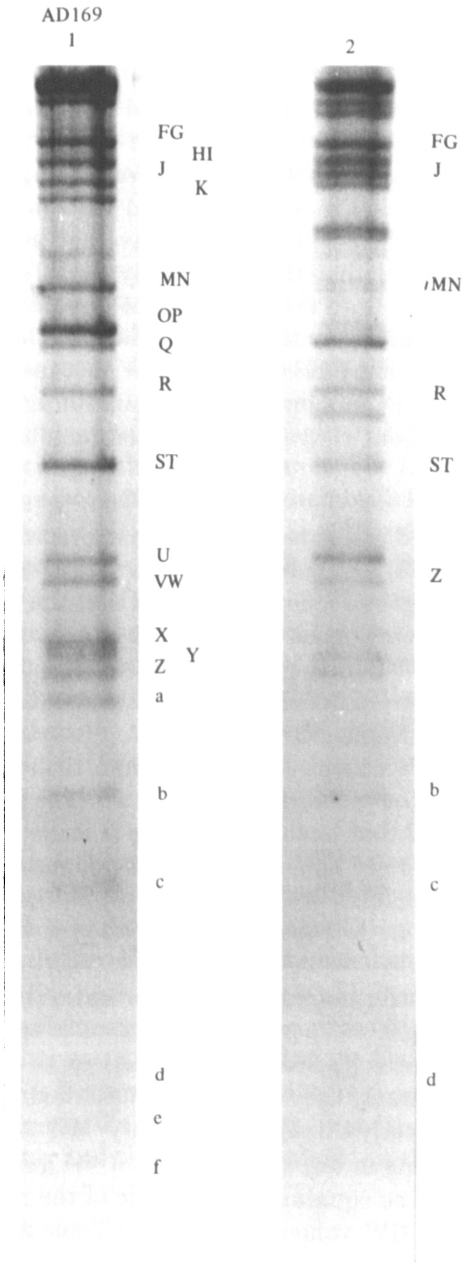

Fig. 1. Autoradiogram of an electrophoretic gel showing the AD169 (track 1) and a test isolate (track 2). The AD169 fragments are labelled according to Spector, Hock & Tamashiro (1982, Table 1). In the test isolate, only fragments that are judged homologous to AD169 fragments are labelled. (Bands d, e and f were visible in the original autoradiograph.)

https://doi.org/10.1017/S0950268800067200 Published online by Cambridge University Press

sense that it could not be removed by fitting more complex polynomials, and may largely be due to rounding errors in the MWs quoted by Spector, Hock & Tamashiro (1982).

In the subsequent processing, the MWs found by Spector, Hock & Tamashiro (1982) for AD169 were replaced by the average of the MWs predicted from the fourth-degree polynomials in 1/mobility. (The *Eco* R1 fragment which would have been labelled L in the series (Fig. 1) was not used as its mobility was inconsistent, and it often lay far from the fitted curve.) The average predicted MWs are shown in Table 1 where it is clear that they differ only slightly from those of Spector, Hock & Tamashiro (1982). Both versions of the MW scale gave very similar results on the subsequent computer matching of strains. In fact, satisfactory results might be obtained with many rescalings of MW, because the method described below for transforming the mobilities of test isolate fragments into MWs makes very few assumptions about the relation between mobility and MW.

Correction for differential distortion of tracks on the same gel

The raw data obtained from the gels were first preprocessed according to the method of Jackman (1985) to remove distortions that cause discrepancies between the mobilities of homologous oligonucleotides on the same gel. The mobility of any fragment identified as homologous to an AD169 fragment is set to be identical to the mobility of its AD169 homologue on that gel. Secondly, the mobilities of the unidentified fragments are adjusted in such a way as to preserve their relative distance from identified fragments. Chaney *et al.* (1983) used the same technique, but applied it to the MWs themselves rather than their mobilities. Column 4 of Table 2 illustrates how the mobilities are corrected. The corrected mobility value for those bars in Table 2 that have a letter code is made identical to that of like-coded bands in the AD169 of Table 1. The other mobilities are meanwhile shifted in a piecewise-linear fashion.

Mapping from mobilities to molecular weight

For each gel, the transformation of the reciprocal of the fragments' mobilities into their molecular weight was performed with a cubic spline equation. These are composite curves calculated by fitting cubic functions to each set of four adjacent data points. In this context, they are simply a method of calculating a smooth curve which will pass exactly through the mobility/MW data points of the AD169. The MW of the fragments in any test pattern on the gel could then be obtained by solving the cubic spline equation at the value of the reciprocal of its adjusted mobility. The resulting MW values are shown in Table 2, column 4. Because the cubic spline curves are constructed to pass exactly through the mobility/MW points of known fragments, the MWs of identified fragments in the test patterns are made to be the same as the corresponding fragments of AD169, and necessarily the same on every gel.

In this way the program makes allowance for gels running to different lengths, and for between-gel variations arising from distortions which affect all tracks on a gel in the same way.

The log scale

To aid visual examination of the standardized patterns, and of the results of the matching process, an intermediate logarithmic scale (column 5 of Tables 1 and 2) was used. A transformation which gave the most convenient linear spread was $Y = 90 - 85 \log_{10}(\text{MW})$. The matching algorithm has been run with this scale.

The matching algorithm

The matching algorithm is applied to each pair of test patterns. It specifies that two bands are considered to 'match' if, and only if, each is the closest band to the other on the log scale. Additionally, the distance between them must not exceed a certain value, which is made dependent on their average position on the scale. Matching bands were required to have MWs within about 2% of each other for heavier fragments, ranging up to 4·5% at the lighter end, reflecting the decreasing precision in measuring the position of the more mobile fragments. This proximity requirement for matching bands was arrived at by trial and error, though the program's performance was not particularly sensitive to its exact setting.

The algorithm also generates various statistics on the results of the pairwise matching: the number of bands, the maximum that could be matched, the number of unmatched bands, and the distances (on the log scale) between the four worst-matching matches. The program will generate a printout of all this information permitting an easy scan for potentially matching patterns. It can also be set to give a more detailed look at the matching process, including a display of the MWs of both test patterns plotted on the log scale.

Program development

The computer program was written in IBM VS Fortran and made extensive use of the Numerical Algorithm Group subroutine library (NAG Fortran Library, Mark 10, 1983) to calculate and evaluate polynomial regressions and cubic splines. The program was initially developed through a number of versions using a set of nine patterns after *Eco* R1 digestion on four gels. This set did not include any gels used in the computer matching study. A little more adjustment of the proximity requirement for matching was carried out after the first attempts to process the present *Eco* R1 data. No further adjustments were made after the program was applied to the *Bgl* II digests. To a large extent, then, the program was developed on a set of 'training' gels and then applied 'blind' to the gels in this study.

RESULTS

Performance as a screening device

The sets of *Eco* R1 and *Bgl* II digest patterns subjected to computer matching had already been examined by eye. It had been judged that in both sets there were two pairs of matching patterns. Unlike any others, these were from isolates from mother and infant pairs. In addition, duplicates of two cultures had been put onto other gels after *Eco* R1 digestion, so that two further matching pairs should be identified by the program. There were therefore four related pairs out of the 91 possible pairings of the 14 *Eco* R1 patterns and 2 related pairs out of 66 pairings of

https://doi.org/10.1017/S0950268800067200 Published online by Cambridge University Press

Table 3. *Numbers of false positive and false negative matches under different matching criteria*

	Eco R1		*Bgl* II		Both simultaneously	
Number of patterns	14		12		12	
Number of pairings	91		66		66	
Number of related pairs	4		2		2	
Matching criterion (maximum number of unmatched bars)	False +ves	False −ves	False +ves	False −ves	False +ves	False −ves
0	0	3	0	1	0	1
1	0	2	0	1	0	1
2	0	1	0	1	0	1
3	2	0	2	0	0	0
4	5	0	7	0	0	0
5	12	0	13	0	1	0

the 12 *Bgl* II patterns. In 4 of the 6 pairs, the members were on different gels, in 2 they were on the same gel.

For each pair of patterns the number of bands in each pattern that were not matched to a band in the other was used as the measure of pairwise similarity. The fewer the number of unmatched bands, the higher the level of similarity.

To evaluate the program as a device which screens for possible matches, it is necessary to consider how many pairs of patterns it incorrectly matched (false positives) and how many true matches it missed (false negatives), as the matching criterion is made more or less restrictive.

Table 3 shows that the most stringent matching criterion, no unmatched bands at all, would give no false negative matches, but would miss 3 of the 4 pairs of related strains among the *Eco* R1 digests and 1 of the 2 among the *Bgl* II digests. If having 4 or fewer unmatched bars is taken as the criterion of a match, all 4 *Eco* R1 matches and 1 false positive would be picked out, and after *Bgl* II digestion both matches and a further 5 false positive would be selected. For either digestion considered individually, 3 or fewer unmatched bands is the most stringent criterion which still picks out all related pairs.

In most applications failure to detect a true match (a false negative) would be more serious than several false positives, and hence a fairly lax matching criterion is preferable. It is therefore very significant that if a match on *both Eco* R1 and *Bgl* II simultaneously is required, then very lax criteria can be adopted without risking a flood of false positives. If five or fewer unmatched fragments are required on each of the two digests, both related pairs are discovered with only one false positive.

Perfect matches were achieved only when the two tracks were on the same gel, and it is of interest to examine the program's 'errors' at the fragment matching level. In one instance, the unmatched fragment was the result of overlooking a band when measuring the gels by ruler. In two instances, two or three very close bars on one pattern were lined up with an indistinct or broad band on the other. As is to be expected, this tends to happen when one pattern runs much further

https://doi.org/10.1017/S0950268800067200 Published online by Cambridge University Press

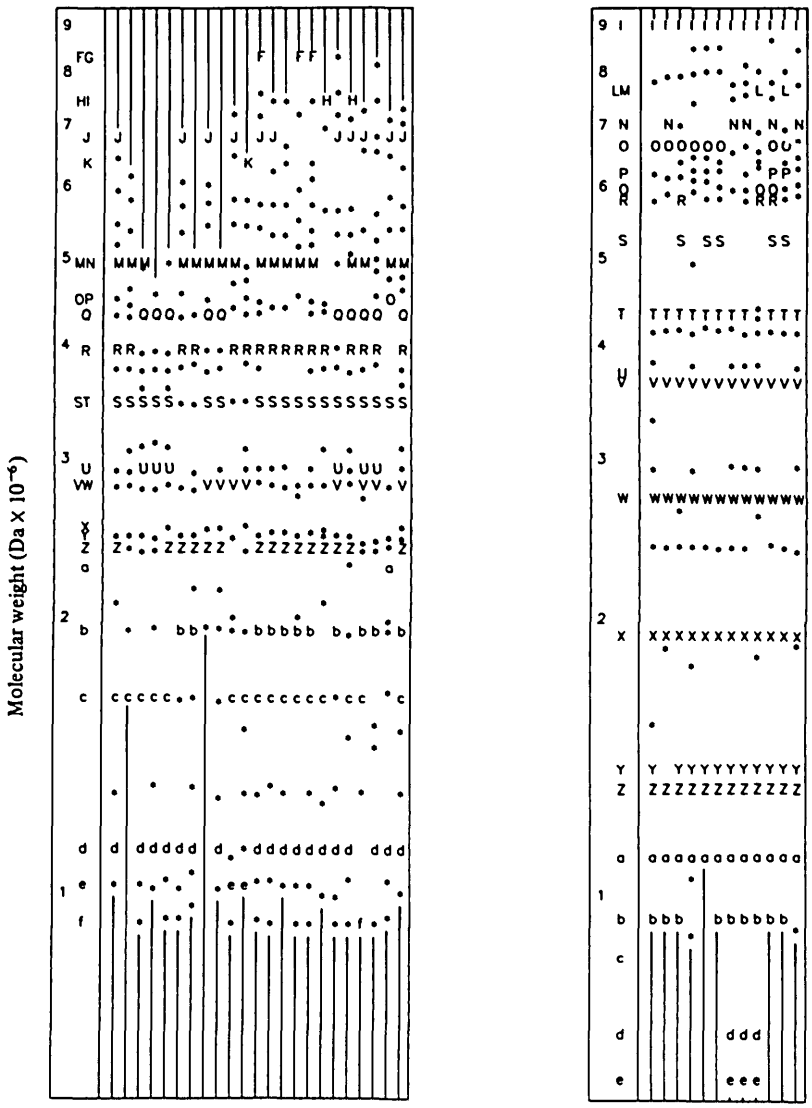

Fig. 2. All 23 *Eco* R1 isolates and 12 *Bgl* II isolates are displayed. Each point represents the MW of a single fragment. The positions of fragments of known MW are indicated by letters, while those of unknown or uncertain MW are indicated by asterisks and the position of AD169 fragments are indicated by the left axis. Solid lines represent portions of the autoradiogram where bands were too indistinct or faint to be resolved by eye.

down the gel than the other, or when one was much clearer than the other. In a further two instances, a very weak band present in one pattern was either invisible or visible only with hindsight in the other. In this respect it is important to emphasize that CMV strains may mutate, and therefore that the search for 'perfect' matching may be inappropriate.

https://doi.org/10.1017/S0950268800067200 Published online by Cambridge University Press

The number of different strains

In Fig. 2 the entire collection of patterns is shown, including the nine original patterns after *Eco* R1 digestion which were used to develop the program. Each column is a plot of the MW of fragments from a single viral isolate. In all there were 23 *Eco* R1 digestions and 12 *Bgl* II digestions. In those cases where the fragments were identified as homologous to AD169 fragments, they are labelled using the terminology of Spector, Hock & Tamashiro (1982). It is evident that many *Eco* R1 and *Bgl* II cleavage points occur in the DNA of all the strains in our study. This implies that the probability of identifying false positive matches from digests by a single restriction enzyme is quite high. However, all the CMV strains in this study were collected within about 2 years from a relatively small area of West London, and their similarity after restriction enzyme digestion and analysis may reflect a real genetic relatedness.

DISCUSSION

The difficulties in comparing two tracks, particularly if they are on different gels, are the same whether it is done by eye or by computer. Firstly, gels are run for different times and under different voltages, and consequently run to different lengths. Secondly, lack of homogenity in the gel, or in the temperature or voltage gradients across it, can cause further between-gel variation in fragment mobility. Thirdly, these distortions can bend the tracks, so that homologous fragments may appear to have different mobilities even on the same gel. A fourth set of problems relates to over- and under-exposure of the autoradiograph, the concentration of DNA in the crude extract, and other factors that can make it difficult to resolve individual bands.

In other contexts a number of methods have been proposed to avoid these problems. For example, given a standard preparation with fragments of known molecular weight, a curve has been fitted so that plots of fragment MW against fragment mobility can be expressed as a parabolic (Duggleby, Kinns & Rood, 1981), logarithmic (Elder & Southern, 1983), or reciprocal relationship (Schaffer & Sederoff, 1981; Elder & Southern 1983; Russell, 1984). An estimated MW may then be ascribed to an unknown fragment by solving the fitted equation at the observed mobility.

Common to all these methods, which may be termed 'global fit', is the aim of characterizing each gel using an equation with only one or two adjustable parameters. This imposes on the data a relationship between MW and mobility that may not reflect the physics of electrophoretic gels in general, and will certainly prove too restrictive to adjust for local distortions unique to each specific gel. It also seems unsatisfactory that the fitted equation does not necessarily pass through the known MW data points of the standard preparation, with the rather curious result that fragments of known molecular weight will not be ascribed the MW that they are known to have.

In applications relating to taxonomy, other investigators have estimated the MW of an unknown fragment by linear interpolation between the MW of the nearest adjacent fragments of known MW (Chaney *et al.* 1983). Alternatively, to

neutralize within gel variation, the mobility pattern can itself be adjusted, by lining up its known fragments with homologous fragments belonging to the standard, and then compressing or stretching the intervening segments in a piecewise-linear fashion (Jackman, 1985). The advantages of using 'local' features for calibration, rather than a single overall equation had earlier been stressed by Elder and Southern (1983).

The present program was developed on a set of training isolates before being applied 'blind' to the study gels. It used local correction to remove within-gel variation, followed by global curve fitting using cubic spline curves. This should correct both for differential distortions of tracks on the same gel and for quite complex gel-specific deviations from a strictly reciprocal relationship between MW and mobility. It may be an advantage of the program that it appears to make no assumptions about the relationship between MW and mobility, nor about the kinds of distortion to which gels are subject.

The second stage of the program, pairwise matching, employed a simple algorithm which counted two fragments of similar MW as 'matching' if each was the closest band to the other. The results suggest that a computer program of this sort can make a useful contribution to epidemiological studies of cytomegalovirus. The matching criteria can be set so as to detect all the true matches without picking out too many false positives. The program could thus be used to alert a researcher to examine the selected pairs more carefully, and put up the paired isolates on the same gel to confirm identity of strain, or to investigate the possibility of transmission of infection in terms of time and place. The method could very probably be applied to other viruses, herpes simplex for example, and possibly to other contexts where banding patterns need to be matched, or compared to known templates.

In principle, this type of program could be used either for calculating MW or for matching, in any situation involving band patterns. In practice, though, the present method may only be usable if the test patterns contain at least two or three bands that are homologous to bands on the standard pattern. This is essential for the preliminary adjustment for differential distortion of tracks on the same gel. About one third of all test isolate bands were considered homologous to AD169 bands, and were thus anchored to fixed points on the MW scale. (This was after *Eco* R1 digestion: over half were fixed on the *Bgl* II patterns). Theoretically, the more within-gel distortion there is, the more fragments of known MW will be needed in the test patterns to correct for it. In the example in Table 2 mobilities are never corrected by more than 0·5 mm, but adjustments of up to 1·5 mm were common, representing up to $0{\cdot}3 \times 10^{-6}$. In cases where tracks were bent, distortions of as much as 7 mm occurred. At the other extreme, of course, if there was no within-gel distortion at all, then the initial adjustment of mobility would be unnecessary and there would no longer be any requirement that the test preparation have some bands homologous to the standard. Only under these circumstances could the subjective element involved in making a judgement of homology be eliminated, and the system made entirely automatic.

In this study, data were keyed in after measurement with a ruler. This is a lengthy and error-prone process. A larger study would require automated data input. In a similar application, digital densitometry has proved to be much more

accurate than analog densitometry or a digitizing tablet (Elder *et al.* 1983) although a median of several readings was used. Densitometer output is in the form of a continuous absorbance trace. When digitized this would be represented by an 8-bit absorbance reading for every 50 microns of gel, or 3000 readings for a 150 mm gel. The computer program described here was designed for a discrete on-off representation of absorbance, but could be adapted to handle digitized continuous absorbance traces.

Adapting the routines that calculate the MW of unknown fragments would be relatively easy, but the matching algorithm and its output would have to be very different if continuous absorbance traces were used, and we suspect that a number of problems remain to be solved relating to the definition of the baseline absorbance, and the densitometer's response to changes in autoradiogram exposure and in DNA concentration. A better approach may therefore be to reduce the continuous densitometer traces to a sequence of 'spikes' each representing a single fragment. The resulting mobilities could then be used as input to the present program or one like it.

Although this procedure disregards much of the information in the original trace, the detail that is lost would probably not assist a robust matching process which can cope with imperfections in electrophoretic gels. Furthermore, use of a simplified 'spike' representation of absorbance must save considerable computing time. The current matching routine will take about 20 seconds central processing time on an Amdahl computer (model 470-V/8) to process 50 patterns (1225 pairs). By contrast, we would expect that use of the full absorbance traces might take 100 times longer or more, rendering it impractical for large numbers of isolates, particularly on microcomputers.

We thank Mrs D. E. Warren for skilled assistance in the preparation of the viral DNA and the electrophoretograms.

REFERENCES

Adler, S. (1985). The molecular epidemology of cytomegalovirus transmission among children attending a day care center. *Journal of Infectious Disease* **152**, 760–768.

Chaney, S. M. J., Warren, K. G., Kettyls, J., Zbitnue, A. & Subak-Sharpe, J. H. (1983). A comparitive analysis of restriction enzyme digests of the DNA of herpes simplex virus isolated from genital and facial lesions. *Journal of General Virology* **64**, 357–371.

Duggleby, R. G., Kinns, H. & Rood, J. I. (1981). A computer program for determining the size of DNA fragments. *Analytical Biochemistry* **110**, 49–55.

Dworsky, M., Lakeman, A. & Stagno, S. (1984). Cytomegalovirus transmission within a family. *Pediatric Infectious Disease* **3**, 236–238.

Elder, J. K., Amos, A., Southern, E. M. & Shippey, G. A. (1983). Measurement of DNA length by gel electrophoresis. 1. Improved accuracy of mobility measurements using a digital microdensitometer and computer processing. *Analytical Biochemistry* **128**, 223–226.

Elder, J. K. & Southern, E. M. (1983). Measurement of DNA length by gel electrophoresis. 2. Comparison of methods for relating mobility to fragment length. *Analytical Biochemistry* **128**, 227–231.

Garrett, A. J. & Warren, D. E. (1985). A simple technique for endonuclease mapping of cytomegaloviruses. *Journal of Virological Methods* **10**, 187–194.

Grillner, M. D. & Strangert, K. (1986). Restriction endonuclease analysis of cytomegalovirus DNA from strains isolated in day care centers. *Pediatric Infectious Disease* **5**, 184–187.

HUANG, E.-S., KILPATRICK, B. A., HUONG, Y. T. & PAGANO, J. S. (1976). Detection of human cytomegalovirus and analysis of strain variation. *Yale Journal of Biological Medicine* **49**, 29–43.

HUANG, E.-S., HUONG, S.-M., TEGTMEIER, G. E. & ALFORD, C. (1980). Cytomegalovirus: genetic variation of viral genomes. *Annals of the New York Academy of Sciences* **354**, 332–346.

JACKMAN, P. (1985). Computer analysis of electrophoretic protein patterns in microbial systematics. *Binary* **5**, 11–13.

NAG FORTRAN LIBRARY MARK 10 (1983). Oxford: Numerical Algorithms Group.

PECKHAM, C. S., GARRETT, A. J., CHIN, K. S., PREECE, P. M., NELSON, D. B. & WARREN, D. E. (1986). Restriction enzyme analysis of cytomegalovirus DNA to study transmission of infection. *Journal of Clinical Pathology* **39**, 318–324.

RUSSELL, G. C. (1984). DNA SIZE – a program to calculate the sizes of DNA fragments in agarose gels from the mobility of standard fragments. *Binary* **4**, 31–32.

SCHAFFER, H. E. & SEDEROFF, R. R. (1981). Improved estimation of DNA fragment lengths from agarose gels. *Analytical Biochemistry* **115**, 113–122.

SPECTOR, D. H., HOCK, L. & TAMASHIRO, J. C. (1982). Cleavage maps for human cytomegalovirus DNA strain AD169 for restriction endonucleases *Eco* R1, *Bgl* II and *Hind* III. *Journal of Virology* **42**, 558–582.

WILFERT, C. M., HUANG, E.-S. & STAGNO, S. (1982). Restriction endonuclease analysis of cytomegalovirus deoxyribonucleic acid as an epidemiological tool. *Pediatrics* **70**, 717–722.

YOW, M. D., LAKEMAN, A. D., STAGNO, S., REYNOLDS, R. B. & PLAVIDAL, F. J. (1982). Use of restriction enzymes to investigate the source of a primary cytomegalovirus infection in a pediatric nurse. *Pediatrics* **70**, 713–716.