COMAP: A *comigrating analysis* program for estimating the relationship of adenoviruses on the genome level

Thomas Adrian and Walter Heinrich¹

Abteilung für Virologie, Institut für Medizinische Mikrobiologie und Hygiene, Universitätskliniken, Haus 47, D-6650 Homburg (Saar), FRG, and ¹Rechenzentrum der Universität des Saarlandes

Received 17 July 1985

ABSTRACT

The computer program COMAP is described that enables one to evaluate the relationship among adenovirus 6 strains by comigration analysis with regard to the genome. The required data were obtained by restriction analysis with several enzymes. The program is also important for identifying further adenovirus strains by DNA restriction analysis and to compare new restriction fragment patterns with stored data of known adenovirus strains.

We believe that in this manner the DNA restriction analysis will become a valuable diagnostic procedure.

INTRODUCTION

Since the discovery of adenoviruses by Rowe (1) many strains have been isolated, characterized and identified. In the meantime 41 prototypes could be identified by serological means, i.e. neutralization and haemagglutination inhibition (2,3). In addition many strains with no clearcut antigenic relationship were isolated (4,5). Other strains however were indistinguishable by immunological tests from the corresponding prototypes. These strains were named genome types (6,7). The main interest of our research was the molecular base of the genetic variability of adenoviruses; we therefore analysed about 600 different adenovirus strains by at least 7 restriction endonucleases. The molecular weights of the resulting fragments were determined and stored in a computer (a 8-Bit Z80 based microcomputer). We used the known database system *dBase-II*, which allows to store, edit and manipulate large sets

The program is available in printed form or on a CM/P-8^{II}-floppy disk upon receipt of a self-addressed mailing label and a blank diskette. A small charge will be requested to cover mailing and processing.

of data in an easy and straightforward manner. Since fragments with the same molecular weight are usually located at the same position on the genome of related strains, as shown by hybridization experiments, the comigration analysis has proved to express the relationship between different adenovirus strains (4, 8).

MATERIALS AND METHODS

DNA restriction analysis

The DNA was isolated from adenovirus-infected HeLa cells as reported before (4, 8). The DNA fragments were made visible by ethidium bromide staining and photographed using 303 nm transillumination to stimulate the fluorescence. The migration distances of the DNA fragments, visible on the Polaroid 665 film were scanned densitometrically, and from these data the molecular weights were determined (9).

COMPUTATIONAL ALGORITHM

In this section we will briefly sketch the idea behind the computational procedure we have implemented on our computer: Let V_1 and V_2 be two different fragment patterns with the molecular weights $w_{11}, w_{12}, \dots, w_{1n_1}$ and $w_{21}, w_{22}, \dots, w_{2n_2}$ (where n_1 and n_2 are the number of fragments of V_1 and V_2 resp.). Then one can say that a fragment w_{1i} comigrates with fragment w_{2j} , if the molecular weight w_{1i} is "in the neighbourhood" of w_{2j} , i.e. if w_{1i} lies in the interval $[(1-f)*w_{2j}, (1+f)*w_{2j}]$, (where f is some positive constant <1), and vice versa. We define

$$\text{com}(w_{1i}, w_{2j}) = \begin{cases} 1, & \text{if } (1-f)*w_{2j} \leq w_{1i} \leq (1+f)*w_{2j} \text{ or} \\ & (1-f)*w_{1i} \leq w_{2j} \leq (1+f)*w_{1i} \\ & (\text{where } f \text{ in our case equals } 0,02, \\ & \text{i.e. } 2\% \text{ representing the accuracy,} \\ & \text{with which a strain can be measured}) \\ 0, & \text{otherwise.} \end{cases}$$

Then one can see immediately that the number of comigrating fragments of these two patterns can be obtained by summing

up, namely:

$$\sum_{\text{com}} (w1i, w2j)$$

$$1 \leq j \leq n1$$

$$1 \leq j \leq n2$$

The choice of allowing 2% deviation in the molecular weights for the definition of comigration was made empirically. The apparent molecular weights of the reference strain (AV2) deviated in all cases less than 2% from the absolute values determined by sequencing the whole genome of AV2 (10).

The number of comigrating fragments as a measure for the relationship between adenovirus strains is expressed as the ratio of the comigrating fragments to the absolute number of existing fragments. If this number is different for the two underlying strains, the mean value is chosen.

These formulas can be reformulated algorithmically and programmed in one of the usual programming languages.

We programmed this algorithm in *MSBASIC* as the nucleus of a program that determines the total number of comigrating fragments and the appropriate percentages, provided all fragments of the variants of an adenovirus for the 7 enzymes (BamHI, BglII, BstEII, EcoRI, HindIII, KpnI and SmaI) are known. It is running on a MODUCOMP dt Z80 based microcomputer system under CP/M Vers. 3.0.

RESULTS

As an example for the DNA restriction analysis with a great number of adenovirus strains we present the genome type analysis of 24 AV6 field strains, representing 13 genome types differing from the prototype.

Figure 1a shows the original gel photo after analysis with the enzyme BstEII. Four fragment patterns, deviating from the prototype 1 and named 2 to 5, are visible. A schematic drawing is given in fig. 1b: the migration distances are shown on a logarithmic scale.

Table 1 shows the comigration analysis of AV6 prototype with the 4 variants after the analysis with one enzyme (BstEII). The total number of fragments, the number of hits,

Fig. 1. (a) Gel photo after analysis with BstEII. **(b)** schematic drawing

and the comigration expressed in percentage after comparing the fragments are recorded. A pairwise comigration of all BstEII variants is shown in the same table.

Table 2 contains the summarized number of comigrating fragments of AV6 and its variants after analysis with 7 enzymes. In the second column the total numbers of fragments of each variant are listed. In the lower left corner of the matrix the total number of comigrating fragments (for all 7 enzymes) between the *i*-th and the *j*-th variant, in the upper right corner the corresponding percentages are recorded.

Table 1
Comigration analysis of BstEII-fragments

a) Prototyp with variants						
	AV6		Variants			
	P	2	3	4	5	
fragments	11	13	12	14	14	
hits (total)	11	9	10	10	9	
hits (perc.)	100	69	83	71	64	
b) Comparison among the variants						
	AV6		Variants			
	P	2	3	4	5	
AV6/P	11	9	10	10	9	total hits
VAR 2		13	11	9	8	
VAR 3			12	9	8	
VAR 4				14	11	
VAR 5					14	
AV6/P	100	75	87	80	72	percentages
VAR 2		100	88	67	59	
VAR 3			100	69	62	
VAR 4				100	79	
VAR 5					100	

DISCUSSION

We have introduced the computer program COMAP for several reasons. First, if a largenumber of adenovirus strains must be compared, like those of subgenus D comprising 23 prototypes (2), it is technically not feasible to run more than 20 DNA samples on the same gel. Second, if new strains are to be compared, besides the prototypes also the greater number of already known variants and genome types (4,6,7,8,11) must be taken into account. Third, COMAP is apt to facilitate a comparison of DNA

Table 2
Comigration of the genome types*)

		1\$	2	3	4	5	6	7	8	9	10	11	12	13	14
1\$	63	..	93	96	98	85	91	90	70	65	93	98	96	95	98
2	64	59	..	89	90	85	98	98	71	64	95	91	89	98	92
3	62	60	56	..	94	83	87	86	74	62	89	94	92	91	94
4	62	61	57	58	..	83	88	88	67	62	90	95	94	93	95
5	66	55	55	53	53	..	84	82	67	74	82	83	81	87	84
6	65	58	63	55	56	55	..	95	69	64	98	88	87	95	90
7	63	57	62	54	55	53	61	..	72	66	93	88	86	95	90
8	60	43	44	45	41	42	43	44	..	71	68	68	72	73	70
9	64	41	41	39	39	48	41	42	44	..	61	63	67	66	65
10	64	59	61	56	57	53	63	59	42	39	..	91	89	93	90
11	64	62	58	59	60	54	57	56	42	40	58	..	94	93	95
12	62	60	56	57	58	52	55	54	44	42	56	59	..	91	95
13	63	60	62	57	58	56	61	60	45	42	59	59	57	..	94
14	62	61	58	58	59	54	57	56	43	41	57	60	59	59	..

*) 2 - 14 : genome types

\$) 1 : prototype

restriction results obtained from different laboratories.

The DNA restriction analysis is a powerful tool in epidemiological studies (11); it may also be used as a type-specific adenovirus proof from patients' material (12). As a general advantage of COMAP, for comparing new strains fresh DNA of reference strains must not be prepared.

Obviously an absolute comparison of adenovirus genomes is possible only by determining their whole nucleotide sequences. Although this has been done for AV2 (10), it would be unrealistic to anticipate that this task will be performed for dozens of adenovirus strains in near future. Hence COMAP is expected to be helpful to study the genetic variability of adenoviruses for years to come.

ACKNOWLEDGEMENTS

We gratefully acknowledge the technical assistance by Barbara Best and the typing the manuscript by Monika Reber. The work was aided by a grant from the Deutsche Forschungsgemeinschaft (wi 3/21-3).

REFERENCES

1. Rowe, W.P., Huebner, R.J., Gilmore, L.K., Parrott, R.H. and Ward, T.G. (1953). *Proc.Soc.Exp.Biol.Med.* 84, 570-573.
2. Wigand, R., Bartha, A., Dreizin, R.S., Esche, H., Ginsberg, H.S., Green, M., Hierholzer, J.C., Kalter, S.S., McFerran, J.B., Pettersson, U., Russell, W.C. and Wadell, G. (1982). *Intervirology* 18, 169-176.
3. de Jong, J.C., Wigand, R., Kidd, A.H., Wadell, G., Kapsenberg, J., Muzerie, C.J., Wermenbol, A.G. and Firtzlaff, R.-G. (1983). *J.Med.Virology* 11, 215-231.
4. Adrian, Th., Bastian, B., Benoist, W., Hierholzer, J.C. and Wigand, R. (1985), *Intervirology* 23, 15-22.
5. Wigand, R., Sehn, N., Hierholzer, J.C., de Jong, J.C. and Adrian, Th. (1985), *Arch.Virol.* 84, 63-78.
6. Wadell, G., Varsanyi, T., Lord, A. and Sutton, R.N.P. (1980), *Am.J.Epid.* 112, 619-628.
7. Wigand, R. and Adrian, Th. (1985) in *Adenovirus DNA: The Viral Genome and its Expression*. Doerfler, W. ed., Nijhoff, M. Publishing, Boston.
8. Adrian, Th., Wigand, R. and Hierholzer, J.C. (1985), *Arch. Virol.* 84, 79-89.
9. Duggleby, R.G., Kinns, H. and Rood, J.I. (1981), *Analyt. Biochem.* 110, 49-55.
10. EMBL Nucleotide Sequence Data Library, European Molecular Biology Laboratory, Postfach 10 22 09, D-6900 Heidelberg, FRG.
11. Wadell, G. (1984), *Current Topics Microbiol.* 110, 191-220.
12. Buitenwerf, J., Louwerens, J.J. and de Jong, J.C. (1985), *J.Virol.Meth.* 10, 39-44.