Arch Virol (1988) 103: 83–98

Archives of Virology

Antigenic relationships between avian paramyxoviruses

III. A mathematical model of antigenic drift and a computer-assisted approach for construction of a phylogenetic tree

M. Lipkind[1] and **N. Rishe**[2]

[1] Kimron Veterinary Institute, Beit Dagan, Israel
[2] School of Computer Science, Florida International University, College of Arts and Sciences, Miami, Florida, U.S.A.

Accepted August 3, 1988

Summary. The suggested model of antigenic kinship between related paramyxoviruses is based on another concept of anitgenic determinant, as compared to the previously suggested combinatorial mathematical model by the authors. According to it, antigenic changes of any determinant do not proceed by "leaps" but can be changed *gradually*. Such changed determinant can induce a correspondingly changed type of antibodies which still preserve a certain kinship to the original type of the determinant (before its changing) revealed by cross reaction serological tests. Accordingly, there can be "families" of the determinants differing by degree of relatedness to (or, reversely, by antigenic distance from) the "original" ("ancestor") determinant.

In addition to another interpretation of the antigenic kinship, the new mathematical model was used as an approach for revealing phylogenetic relationships between antigenically related viruses.

Introduction

The group of avian PMV numbers to date 9 antigenic serotypes [3, 4]. In the previous communication [16] diverse multidirectional antigenic interconnections displayed by both hemagglutination (HA) and neuraminidase (Nase) inhibition tests have been found. This has given a basis for some suggestions concerning: (a) a distinct topology of the HA and Nase functional antigenic sites on the HA-Nase (HN) molecule and their independent antigenic evolution; (b) a tentative subdivision of the group of avian PMV into two subgroups; (c) the conception about the presence of the "common" and "serotype-specific" and "conserved" and "variable" portions in the HN gene.

A combinatorial mathematical model describing the found antigenic relationships between the avian PMV serotypes was presented [28]. According to

the model, the HN of each PMV antigenic serotype contains a set of particular (serotype-specific) sort(s) of the determinant(s) which is (are) prevailing as well as some determinants of the other sorts which are shared by the most of the other serotypes and, hence, cause cross reactivity with the other respective serotypes. The model has given an interpretation of the experimental serological results including, in particular, the phenomenon of asymmetric cross reactivity.

The combinatorial character of this model was based on the "all-or-none" principle. According to it, each determinant is an absolutely specific antigenic entity with no relationship to any other sort. Thus, the specificity of the combination of the "serotype-specific" and "common" determinants ascertains the pattern of the relationships of the given serotype wih the others. Accordingly, any possible alteration of any determinant, no matter how minimal but detectable, means a qualitative "leap" change of the determinant antigenicity, so that such determinant is converted into quite another one (the "all-or-none" kind of change). This should be displayed by inability of the antibody against the determinant before changing to bind to the determinant after changing and vice versa. The meaning of this is that any changing of the PMV serotype which may occur (both natural evolutionary and experimentally mutational) and which are expressed by serological tests—hemagglutination inhibition (HI) or neuraminidase inhibition (NI), in the considered case—are associated with the changes in the combination of different sorts of the determinants which will include this newly converted one.

In the present communication an alternative hypothesis is suggested which includes some different postulates as compared to the initial [28] hypothesis. The essence of the new hypothesis is based on a new concept of the antigenic determinant. According to it, the determinant is not an invariable entity but can be changed (drifted) *gradually*. Such a changed determinant can induce a correspondingly changed type of antibodies but these "changed" antibodies still preserve a certain kinship to the original type of the determinant before its changing. Accordingly, there can be a "familiy" of the determinants with different minor changes of different degree each of which reacts with antibodies induced by either member of the family while the "compatibility" between the "heterologous members" of the family is detectably less than that between the "homologous" ones.

In addition to another interpretation of the antigenic kinship, the new hypothesis served as a basis for the mathematical model of phylogenetic relationships between the viruses.

Materials and methods

The same bulk of the experimental data [16, 17] used for the previously developed mathematical models [28] was employed in the present studies. The data concern comparative cross-reactive serological studies which revealed multiple interrelationships between various serotypes belonging to a group of avian paramyxoviruses. These serotypes are listed in Table 1.

Thus, all the serotypes of the avian PMVs, with the only exception of PMV-5/budgerigar/Japan/Kunitachi/74 which was not available, were used in the studies.

Table 1. Reference strains of avian paramyxoviruses

Serotype	Prototype strain	
	Full designation	Abbreviation
1	PMV-1/NDV/LaSota/51	NDV
2	PMV-2/chicken/Yucaipa/California/56	Yucaipa
(a)	PMV-3/turkey/Wisconsin/67	Ty/Wisc
3		
(b)	PMV-3/parakeet/Netherlands/449/75	Pk/Neth
4	PMV-4/duck/Hong Kong/D 3/75	D 3/HK
6	PMV-6/duck/Hong Kong/D 199/77	D 199/HK
7	PMV-7/dove/Tennessee/75/4/75	Dove/Tn
8	PMV-8/goose/Delaware/1053/76	Goose/Del
9	PMV-9/duck/New York/22/80	Duck/NY
?	PMV-?/pigeon/Otaru/76	Pigeon/Ot

Results

Postulates of the general hypothesis

The postulates (1)–(3) are similar to those used in the combinatorial model of the antigenic kinship [28], namely:

(1) Each PMV virion contains a number C of identical HN molecules.

(2) Each HN molecule contains two distinct antigenic domains—HA and Nase—spatially arranged around HA and Nase functionally active sites.

(3) Each HA as well as Nase domain consists of number D_h and D_n, respectively, of antigenic determinants.

The next two postulates are the core of the present hypothesis:

(4) The antigenic determinants are grouped into classes, each class including both identical and similar ("related") determinants.

(5) Each determinant of a certain class induces a sort of antibodies which, besides the identical (homological) determinants, can also inhibit the related determinants of the same class with the effectiveness according to the degree of the relatedness.

Definitions of antigenic kinship and antigenic distance

Let us suggest that a virus **x** evolves into a virus **y** as a result of such mutational changes in the determinant(s) belonging to a class **c** that the changed determinant(s) still preserves relatedness to the initial determinant before mutation, i.e. remains belonging to the class **c** ("similar" or "related" determinants). Then, the antigenic kinship between the viruses **x** and **y**, *with respect to the determinants of the class* **c** determined by functional—either HI or NI—tests is defined as the ratio between: (a) the inhibition of the class-**c** determinants of **x** by the

antibodies against the class-**c** determinants of **y** and (b) the homologous inhibition of the class-**c** determinants of the **x** by the homologous antibodies.

This kinship is denoted as $\mathbf{k}(\mathbf{x}, \mathbf{y}, \mathbf{c})$. Note that $\mathbf{k}(\mathbf{x}, \mathbf{x}, \mathbf{c}) = 1$. The important point is that it is not at all necessary that $\mathbf{k}(\mathbf{x}, \mathbf{y}, \mathbf{c})$ must be equal to $\mathbf{k}(\mathbf{y}, \mathbf{x}, \mathbf{c})$ [the possibility for the asymmetric cross reaction] [16, 17, 28].

The general antigenic kinship between the viruses **x** and **y** *with respect to all the classes of antigenic determinants* detected by the functional inhibition tests (HI and NI) is defined as the ratio between: (a) inhibition of all the determinants of the virus **x** by the antibodies against the determinants of the virus **y** and (b) inhibition of the determinants of **x** by the homologous antibodies.

This general kinship is denoted as $\mathbf{K}(\mathbf{x}, \mathbf{y})$. Note that $\mathbf{K}(\mathbf{x}, \mathbf{x}) = 1$. It is not necessary that $\mathbf{K}(\mathbf{x}, \mathbf{y})$ must be equal to $\mathbf{K}(\mathbf{y}, \mathbf{x})$.

It follows from the experimental data [16] that sometimes $\mathbf{K}(\mathbf{x}, \mathbf{y})$ is several orders lower than $\mathbf{K}(\mathbf{x}, \mathbf{x}) = 1$, (according to the degree of the relatedness) or even is equal to zero (no relatedness).

In the first approximation, we shall postulate a correlation between $\mathbf{K}(\mathbf{x}, \mathbf{y})$ and $\mathbf{k}(\mathbf{x}, \mathbf{y}, \mathbf{c})$. Together with the above postulates and definitions, this correlation is based on the following speculation:

Let us suggest that a concentration of antibodies to the homologous virus **y** is sufficient to inhibit a functional activity of a certain amount of the virus in a certain quantitative test (either HI, or NI). If we perform the same test with a heterologous virus **x** (cross reactivity experiment), only a certain portion of the determinants of the class $\mathbf{c}_1$, namely, that equal to $\mathbf{k}(\mathbf{x}, \mathbf{y}, \mathbf{c}_1)$, a certain portion of the determinants of the class $\mathbf{c}_2$, namely, that equal to the $\mathbf{k}(\mathbf{x}, \mathbf{y}, \mathbf{c}_2)$ and so forth, would have a chance to be inhibited. Approximately, for a given virus there are chances equal to $\mathbf{k}(\mathbf{x}, \mathbf{y}, \mathbf{c}_1)$, $\mathbf{k}(\mathbf{x}, \mathbf{y}, \mathbf{c}_2) \ldots \mathbf{k}(\mathbf{x}, \mathbf{y}, \mathbf{c}_n)$ that a certain amount of the virus's determinants $\mathbf{c}_1, \mathbf{c}_2, \ldots \mathbf{c}_n$, respectively, would be inhibited. Assuming that those "chances" are independent, the chance to inhibit all the classes of the determinants can be expressed as the multiplication of $\mathbf{k}(\mathbf{x}, \mathbf{y}, \mathbf{c}_1)$, $\mathbf{k}(\mathbf{x}, \mathbf{y}, \mathbf{c}_2) \ldots \mathbf{k}(\mathbf{x}, \mathbf{y}, \mathbf{c}_n)$. Such multiplicative correlation, being a common case for processes occurring in living systems, appears to be the only simple correlation which is consistent with our experimental data on cross reactivity inhibition tests [16]. Certainly, some complex correlations are possible to be consistent with the experimental data, but in the first approximation those complex correlations can be reduced to multiplication. The first criterion of the consistency is the degree of antigenic kinship expressed by numerical experimental data [16]. However, this is rather a weak criterion since the degree of the kinship is not connected directly with the problem of ancestry. Another criterion is based on the pairwise comparison of the experimental data on the kinship related to triples of the viruses **x**, **y**, **z**, i.e. $\mathbf{K}(\mathbf{x}, \mathbf{y})$, $\mathbf{K}(\mathbf{y}, \mathbf{z})$ and $\mathbf{K}(\mathbf{x}, \mathbf{z})$ as will be discussed below.

Therefore, on the basis of the above postulates, definitions and considerations, we postulate that, in the first approximation, the primary hypothesis of

antigenic kinship is expressed by the following equation:

$$\mathbf{K(x, y) = k(x, yc, c_1) \cdot k(x, y, c_2) \cdot \ldots \cdot k(x, y, c_n).}$$

A notion of "antigenic distance", which is reverse to the notion of the antigenic kinship and which will be used in the further computations in order to simplify formulas, is expressed as

$$\mathbf{D(x, y) = -\log_2 [K(x, y)].}$$

An auxilliary theorem

Let us assume that:

(a) A virus **y** is a descendant of a virus **x**;

(b) A virus **z** is a descendant of the virus **y**;

(c) The determinant pattern of the virus **z** differs from the determinant pattern of the virus **x** in what the virus **z** differs from the virus **y**, plus in what **y** differs from **x**. This means that if the drift from **x** to **y** caused changes in some classes of the determinants, the drift from **y** to **z** caused changes in *other* classes of determinants, rather than changes in the same classes again.

Then:

(i) The antigenic distance from **x** to **z** is equal to the antigenic distance from **x** to **y** plus the antigenic distance from **y** to **z**.

(ii) The antigenic distance from **z** to **x** is equal to the antigenic distance from **z** to **y** plus the antigenic distance from **y** to **x**.

The proof of the theorem:

Let us assume that the change from **x** to **y** was in the classes $\mathbf{c_1}$ and $\mathbf{c_2}$, and the change from **y** to **z** was in the class $\mathbf{c_3}$. Then:

$$\mathbf{K(x, y) = k(x, y, c_1) \cdot k(x, y, c_2) \cdot \ldots \cdot k(x, y, c_n) =}$$

$$\mathbf{= k(x, y, c_1) \cdot k(x, y, c_2) \cdot 1 \cdot \ldots \cdot 1 = k(x, y, c_1) \cdot k(x, y, c_2).}$$

Similarly:

$$\mathbf{K(y, z) = k(y, z, c_3)}$$

and, similarly, since **x** and **z** differ in three classes,

$$\mathbf{K(x, z) = k(x, z, c_1) \cdot k(x, z, c_2) \cdot k\,(x, z, c_3).}$$

But $\mathbf{c_1}$ and $\mathbf{c_2}$ have not been changed during the drift from **y** to **z**, and $\mathbf{c_3}$ has not been changed during the drift from **x** to **y**.

Thus, for example, $\mathbf{k(x, z, c_1) = k(x, y, c_1)}$ and, hence,

$$\mathbf{K(x, z) = k(x, y, c_1) \cdot k(x, y. c_2) \cdot k(y, z, c_3).}$$

Thus,

$$\mathbf{K}(\mathbf{x},\mathbf{z}) = \mathbf{K}(\mathbf{x},\mathbf{y}) \cdot \mathbf{K}(\mathbf{y},\mathbf{z})$$

and, hence,

$$\mathbf{D}(\mathbf{x},\mathbf{z}) = -\log[\mathbf{K}(\mathbf{x},\mathbf{z})] = -\log[\mathbf{K}(\mathbf{x},\mathbf{y}) \cdot \mathbf{K}(\mathbf{y},\mathbf{z})] = \{-\log[\mathbf{K}(\mathbf{x},\mathbf{y})]\} + \\ + \{-\log[\mathbf{K}(\mathbf{y},\mathbf{z})]\} = \mathbf{D}(\mathbf{x},\mathbf{y}) + \mathbf{D}(\mathbf{y},\mathbf{z}).$$

Similarly,

$$\mathbf{D}(\mathbf{z},\mathbf{x}) = \mathbf{D}(\mathbf{z},\mathbf{y}) + \mathbf{D}(\mathbf{y},\mathbf{x})$$

which is the end of the proof.

The approach to the solution of the problem of ancestry

On the basis of the above theorem, let us designate the viruses **x**, **y**, and **z** as a parent, a child and a grandchild, respectively. According to the assumption (**c**) of the theorem, not every triple of **x-y-z** has to satisfy the corollaries (i) and (ii) of the theorem. That means that when the triple of the viruses **x-y-z** does satisfy both (i) and (ii), then there is a good chance that they are related according to either of the schemes: parent-child-grandchild or grandchild-child-parent. The fact that the antigenic distances (**D**) between different pairs of related viruses (i.e. those responding to the parent-child scheme) are quantitatively different [16, 28], is compatible with the view that the evolutionary changes between the related viruses are not due to one drifting step but rather are the result of several such steps.

Now let us assume that several triples of viruses, for example, **x-y-z**$_1$, **x-y-z**$_2$, and **x-y-z**$_3$ satisfy both the conditions (i) and (ii) of the theorem. From each triple taken separately one cannot affirm whether **x** is the parent of **y** or **y** is the parent of **x**, but several triples considered together provide strong evidence that the **x** is the parent of the **y** and $\mathbf{z}_1$, $\mathbf{z}_2$, and $\mathbf{z}_3$ are the children of the **y**. If this were not true, then each of the $\mathbf{z}_1$, $\mathbf{z}_2$, and $\mathbf{z}_3$ would have a strong probability to be the parent of **y** that is quite unlikely.

Further support for the decision about the parentship relation between **x** and **y** can be given by a triple also satisfying the conditions (i) and (ii) of the theorem but involving another virus **w** on the left side, instead of the virus **z** on the right side, namely, **w-x-y**. This in itself does not determine absolutely who is a parent of whom, but strongly supports the choice between the two possibilities: whether **x** is the parent of **y**, or the **y** is the parent of the **x**.

From the triple **x-y-z** we derived a conclusion about the possible parentship relation between **x** and **y** in one of the directions. From the triple **w-x-y** we also derive a conclusion about the possible parentship between **x** and **y**, thus reinforcing the conclusion from the **x-y-z**.

The data about the antigenic distances between every pair of the viruses have been obtained experimentally. A computer program has been written to

analyze all triples of viruses and extract those triples where $\mathbf{D(x, z) = D(x, y) + D(y, z)}$ approximately (up to the precision of the experimental method) and/ means that these 12 triples could be combined to 6 pairs, each pair of the triples whether it displays a parent-child-grandchild direction or an opposite grandchild-child-parent direction. However, the totality of all our data treated by the analysis of all the triples as described in the following section and the exhaustive computerized combinatorial search for all possible global relationships capable of forming a phylogenetic tree or a part of it based on the above principles have provided for only one possible solution, i.e. only one possible phylogenetic tree.

The experimental data on both the HI and NI tests [16] were computed, according to the above approach and the especially constructed computer program, and the whole total of 720 possible triple combinations of the viruses, separately for HI and NI tests, have been treated.

Phylogenetic relationships between the viruses according to the data of HI test

The treatment of the experimental data related to the HI test [16] has revealed 12 triple combinations (from the whole total of 720 possible ones) of the viruses which could be arranged by the above expression of **x-y-z** triples satisfying (with a minor exclusion) both the conditions (i) and (ii) of the theorem. That means that these 12 triples could be combined to 6 pairs, each pair of the triples corresponding to the direct and reverse directions of the ancestry [conditions (i) and (ii)]. The results obtained (Tables 2 and 3) appeared to be a basis for the construction of the phylogenetic tree.

The example below demonstrates the approach:

The following two triples were found to satisfy the conditions (i) and (ii):

1. Dove/Tn — Pigeon/Ot — Pk/Neth
2. Dove/Tn — Pigeon/Ot — NDV

In each of these cases taken separately, there is a good reason to suspect phylogenic (ancestor-descendant) relatedness, but the ancestry could exist with equal chance in two opposite directions, namely:

Dove/Tn → Pigeon/Ot → Pk/Neth or Pk/Neth → Pigeon/Ot → Dove/Tn and

Dove/Tn → Pigeon/Ot → NDV or NDV → Pigeon/Ot → Dove/Tn.

However, if to consider both the triples together, a chance for the ancestry in the direction from the Dove/Tn significantly increases, namely:

Dove/Tn → Pigeon/Ot → Pk/Neth and Dove/Tn → Pigeon/Ot → NDV

The chance for such direction of the ancestry further increases while considering the other triples satisfying the conditions (i) and (ii) and containing

Table 2. The 6 triples satisfying the condition (i) of the auxilliary theorem

Triples	X-Y-Z			
	D(x, y)	**D(y, z)**	**D(x, y) + D(y, z)**	**D(x, z)**
D 3/HK-Ty/Wisc-Pk/Neth	2.18 ± 0.47	1.75 ± 0.64	3.93 ± 0.86	4.07 ± 1.13
NDV-Pigeon/Ot-Dove/Tn	3.61 ± 0.43	0.05 ± 0.36	3.66 ± 0.56	3.29 ± 0.22
Pk/Neth-Pigeon/Ot-Dove/Tn	3.00 ± 0.34	0.05 ± 0.36	3.05 ± 0.50	3.35 ± 0.20
Ty/Wisc-Pk/Neth-Pigeon/Ot	1.75 ± 0.64	3.00 ± 0.34	4.75 ± 0.72	4.19 ± 0.69
D 3/HK-Dove/Tn-Pigeon/Ot	3.14 ± 0.70	2.40 ± 0.22	5.54 ± 0.073	5.00 ± 0.93
D 199/HK-Dove/Tn-Pigeon/Ot	3.50 ± 0.50	2.40 ± 0.22	5.90 ± 0.55	6.01 ± 0.22

The figures are the mean values with standard errors
The 4th and 5th columns demonstrate $\mathbf{D(x, y) + D(y, z) \approx D(x, z)}$ [absence of the significant difference columns]

Table 3. The 6 triples satisfying the condition (ii) of the auxilliary theorem (combined in the reverse order, as compared to the triples displayed in the Table 2)

Triples	Z-Y-X			
	D(z, y)	**D(y, x)**	**D(z, y) + D(y, x)**	**D(z, x)**
Pk/Neth-Ty/Wisc-D 3/HK	2.95 ± 0.66	1.37 ± 0.21	4.32 ± 0.69	5.52 ± 0.60
Dove/Tn-Pigeon/Ot-NDV	2.40 ± 0.22	3.21 ± 0.43	5.61 ± 0.48	5.37 ± 0.43
Dove/Tn-Pigeon/Ot-Pk/Neth	2.40 ± 0.22	2.05 ± 0.52	4.45 ± 0.56	3.81 ± 0.60
Pigeon/Ot-Pk/Neth-Ty/Wisc[a]	2.05 ± 0.52	1.37 ± 0.21	3.42 ± 0.56	1.75 ± 0.31
Pigeon/Ot-Dove/Tn-D 3/HK	0.05 ± 0.36	5.58 ± 0.72	5.63 ± 0.85	4.06 ± 0.78
Pigeon/Ot-Dove/Tn-D 199/HK[b]	0.05 ± 0.36	∞	∞	5.04 ± 0.96

The same designations as those to the Table 2
[a] The difference between the 4th and 5th columns is on the verge of statistical significance
[b] The only case not satisfying the condition (ii) of the auxilliary theorem

other combinations of the members of the above two initial triples, for example:

Pigeon/Ot — Pk/Neth — Ty/Wisc and Pk/Neth — Ty/Wisc — D 3/HK

Such kind of the analysis involving all the triples (Tables 2 and 3) has led to the following conclusions:

All the triples, with the only exception of D 199 — Dove/Tn — Pigeon/Ot, satisfy condition (ii) whenever they satisfy the condition (i), and vice versa. The triples satisfying (i) and (ii) conditions span a clear phylogenetic tree (Fig. 1) with multiple evicence (from several triples) for each segment of the tree. As to the triple D 199 — Dove/Tn — Pigeon/Ot, it satisfies the condition (i) but

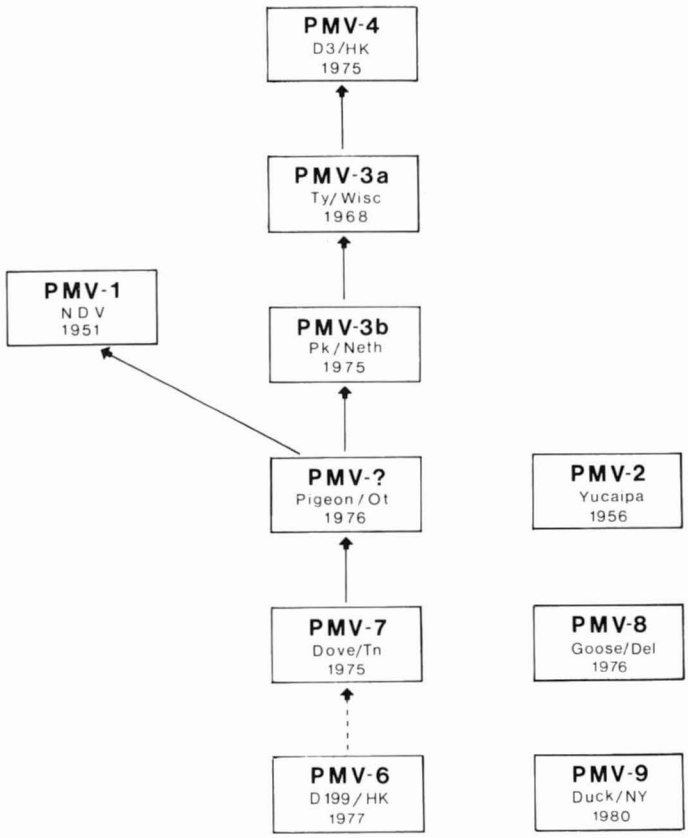

Fig. 1. Phylogenetic relationships between seven serotypes of avian paramyxoviruses. ———> Interconnections satisfying conditions (i) and (ii) of the theorem, – – –> interconnections satisfying condition (i) of the theorem

does not satisfy the condition (ii): Pigeon/Ot — Dove/Tn — D 199/HK. The failure to satisfy the condition (ii) is not in itself a contra-indication to the existence of a phylogenetic relationship, namely, D 199 → Dove/Tn → Pigeon/Ot. Therefore, in the case of the D 199/HK, the conclusion about its location upon the phylogenetic tree is supported only by one piece of evidence [the corresponding branch of the tree is indicated by the arrow with dashes (Fig. 1) while the rest of the tree segments are supported independently by several pieces of evidence].

However, a peculiar situation is connected with the position of D 3/HK on the phylogenetic tree: on one hand, the D 3/HK is located at the top of the tree, being a descendant of the Ty/Wisc, while, on the other hand, the D 3/HK is one of the parents of the Dove/Tn (together with the D 199) being located in the bottom of the tree. Taking into account that a phylogenetic tree cannot be circular and the situation when one descendant is derived from two fathers is impossible, we have had to accept, as the most reasonable, the possibility that the D 3/HK is a child of the Ty/Wisc (Fig. 1).

The results of the computational treatment of the NI test data

Among the 720 triples treated by the computational program, only a very few triples satisfying (i) and (ii) were found but there was no mutual support between them as it was in the case of the HI test. There were some triples which provided a weak support for the results of the HI test. Besides, there were some triples with the viruses Goose/Del and Duck/NY (no phylogenetic relationship was found in the case of the HI test data) but those triples were not compatible with the tree constructed from the HI test data (Fig. 1).

Discussion

The presented mathematical model of antigenic kinship is based on different definition of the concept of antigenic determinant as compared to the previously suggested one [28]. The latter is based on the postulate that any changes occurring in any antigenic determinant proceed according to the "all-or-none" law, i.e. the antibodies against the changed determinant must not react at all with the original determinant (before the mutation). According to the present definition, the determinants may undergo *gradual* changes, this leading to the appearance of the "families" (classes) of antigenically related (cross-reacting) determinants of different degree of the relatedness to each other.

Since the two suggested mathematical models are alternative ones, a natural question is which of them is "correct". At present, however, the choice between them is based on the respective estimation of their "explaining capacity". Each of the models, based on quite different hypotheses, "explains" the described phenomena of the complicated maize-like network of the cross-reaction relationships between avian PMV serotypes, including the phenomenon of asymmetric cross-reactivity. However, the second model presents an original approach for the establishment of the phylogenetic interconnections between the antigenically related viruses.

Generally, the problem of phylogenetic relationships as applied to the viruses has some specific peculiarities. First of all, there is a problem of quantitative estimation of the degree of relatedness between any related viruses. Till the recent time, the relatedness between the viruses was determined by different phenotypic features associated mainly with morphology, chemical composition, antigenicity, genome strategy etc. which were the basis for taxonomy. Recent development of sequencing techniques revealed new possibilities for quantitative estimation of the relatedness on the genomic level. Indeed, the sequencing of viral genome seems to permit an absolute determination of the degree of the relatedness: in this respect, the anti-relatedness (distance) between the related viruses maybe expressed just as a number of nucleotide bases or, correspondingly, of amino acids, by which these viruses differ. However, the differences in some phenotypic viral properties (for example, antigenic) do not always go in parallel to the *quantitative* differences on the genomic level. In the genealogical dendrograms described recently for the case of influenza virus HA gene [2, 7,

15, 36] the quantitative estimation of the antigenic drift was done either by means of calculation of percentage of amino acid sequence difference made by either especial computer program [1, 2, 36], or by calculation of a number of mutations separating different strains [7, 15, 33]. It may seem that the modern approaches based on computations of the sequence data may give a more precise quantitative estimation of the antigenic drift than classic serological tests. However, evidence has been obtained that a single base change in influenza virus HA gene may cause drastic antigenic change [24, 33] which may be pleiotropically associated with changes in biological functions [6, 13]. On the other hand, complete sequence analysis has shown that the hemagglutinins of H 1 (HO) and H 2 influenza subtypes are closely related [10] while serologically they are quite distinct. Besides, the problem is complicated by the fact that new antigenic variants are generated by consequential change of key amino acids within certain antigenic regions and such changes may represent evolutionary endpoints [6]. It was also shown by RNA-RNA hybridization technique that base homology percentage does not always correlated with intraserotype antigenic differences [31a]. Thus, the changes in antigenicity measured serologically and being, in fact, real expressions of a current antigenic drift, do not proceed in parallel with monotonous quantitative accumulation of point mutations. Therefore, together with differences on the genomic level (primary polypeptide structure) the differences in some measurable phenotypic indications should be taken into account.

However, the main difficulty in the elucidation of the phylogenetic relationships lies in revealing the *direction* of the evolutionary changes, i.e. a recognition of ancestry within a considered family of the related viruses. Even in the case when the relatedness between the compared viruses is established and measured quantitatively, it remains obscure what strain must be considered as an "ancestor" to which all the other somehow differing "descendant" viruses could be related. As soon as the direction is determined, the members of the family can be arranged in a "phylogenetic tree" according to the quantitatively determined differences. The absolute criterion of the ancestry—a precedence in time—is not suitable in the case of viruses because of the absence of viral paleontology. In this respect, the only "temporal" criterion is the chronological time (the date) of the viruses' isolation which, however, can often be just an occasional incident not necessarily connected with the time of emergence of the isolated strain in nature. Only in some particular cases such a criterion can really be the objective one, namely, in the cases of those virsues which caused "new" pandemics. The classic example concerns pandemic influenza viruses when new antigenic serotypes appeared "for the first time" ("Asian" and "Hong-Kong" influenza in 1957 and 1968, respectively). Indeed, taking into account the initial appearance of a new influenza strain and a wide global surveillance of the influenza viruses, especially during and after a pandemic period, it is possible to register any evolutionary changes of the initial strain in the real temporal succession [2, 7, 33, 36]. Another demonstrable example is a new

contagious conjunctivitis disease caused by a novel enterovirus E 70 which, being registered for the first time in 1969 [8, 14], during the next decade was isolated all over the world [37] presenting a perfectly surveyed and well-documented pandemic case. On the basis of these data, a phylogenetic tree was constructed [20, 34] using comparative oligonucleotide mapping genetic analysis combined with pairwise comparison of the common oligonucleotide spots [1] and Unweighted Pair Group Method [25].

However, even in such a seemingly clear case as influenza viruses, the constructed phylogenetic relationships [2, 7, 33, 38] reflect only sequential evolution on a single branch since the origin of a strain initiating a pandemic turned out to be obscure. The point is that, according to the recent views on "sero-archeology" (retrospective seroepidemiology) those "new" pandemic strains were circulated long ago [22]. Hence, the "modern" pandemic strains were causative agents of some more "old" pandemics [22]: the "Asian-like", "Hong-Kong"-like, and swine-like influenza virus strains are believed to have caused the influenza pandemic in 1889–1890, 1900, and 1918, respectively [9, 19, 21]. Therefore, even in this case, the date of the first isolation of the pandemic influenza virus does not mean the time of its emergence in nature.

In the case of the animal, especially avian, viruses, like avian influenza or avian paramyxoviruses (PMV), such temporal criterion seems to be fully unsuitable since it does not reflect at all the real chronology of the virus origin in nature. Some data on the monoclonal antibody-mediated analysis of the avian influenza viruses may be an illustration of it [11a, 12, 18]. For example, some of the H 7-containing strains isolated recently in Israel had the same antigenicity by the monoclonal antibody HI test as the prototype FPV/Rostock/34 virus (the time interval between the isolations is 45–46 years) while some other avian influenza virus strains isolated during the same season differed significantly [18].

The above consideration can be applied to the evidence that the viruses Goose/Del and Duck/NY have shown no phylogenetic relationship to the other avian PMVs (Fig. 1), in spite of the multiple antigenic cross reactivity between them [16, 17]. We can suggest two possible explanations:

1. According to the condition (**c**) of the theorem, the phylogenetic relationship performed by the parent—child—grandchild (**x-y-z**) triple analysis can be established only if the changes from **x** to **y** and **y** to **z** affect *different* classes of antigenic determinants (sequential mutation). If, however, the drift is due only to **x-y** changes, the phylogenetic relationship estimated by the criteria accepted here (several independent pieces of evidence obtained from the analysis of different triples) will be absent while antigenic relatedness would be quite well expressed.

2. The phylogenetic connections between the Goose/Del and Duck/NY and the other studied viruses are realized through "hypothetical" segments of the phylogenetic tree represented by the viruses which either "died out" (do not exist) or do exist (i.e. are circulating in the nature) but have not yet been isolated.

For example, it can be easily imagined that the shape of the phylogenetic tree presented here would be quite different if the present analysis had been performed before 1976 when the Pigeon/Ot, occupying the crucial point upon the tree (Fig. 1), had not yet been isolated.

The same considerations can be applied for the explanation of the absence of the phylogenetic relationships found in the case of the NI test. However, the discrepancy between the data related to HI and NI tests remains to be a perplexing phenomenon. On the one hand, it does not compromise the approach as it is. Contrary to it, the failure to find phylogenetic relationships in the case of the NI test is so contrasting with the amazing consistency of the HI test data revealing the phylogenetic relationships, that it serves as a demonstration of a certain real regularity but not just a game of figures. On the other hand, the discrepancy is astonishing if to take into account that both HA and Nase activities belong to the same HN glycoprotein molecule [31, 32, 35]. Such contrasting results on the phylogenetic relationships, together with the difference in the cross reactivity [16, 17] between avian PMVs revealed by HI and NI tests, seem to demonstrate the polar location of the HA and Nase functionally active sites upon the three-dimensional body of the HN molecule as well as the independent antigenic drift of the HA and Nase functionally active sites

As to the point about correlation between genetic relatedness measured by sequencing techniques and the antigenic kinship measured by serological tests, it could be thought that the genetic relatedness is an intrinsic indication of evolution to which any other (phenotypic) indications are related as secondary. In this respect, the phylogenetic tree describing the Enterovirus 70 evolution [20] and constructed on the genetic basis, seems to be more genuine than that based on functional tests which is presented here. However, the main postulate underlying the genetical model of the phylogenetic tree is the assumption that the nucleotide base substitution occurs at a constant rate [34], although some oligonucleotide spots were found to be surprisingly conserved [20] which is not compatible with randomity character of the assumption. Besides, this assumption, seeming suitable for the case of Enterovirus 70, does not take into account such phenomena as persistent and latent viral infections which accelerate mutational changes [11], and a combination of the mystifying "freezing" or conservation effect [11a, 12, 18, 23, 27, 38] with the phenomenon of microheterogeneity [26] which fashions the influenza virus enigma. Therefore, the molecular evolution proceeding at the genomic level and expressed by the base substitution, and the antigenic drift proceeding at the protein tertiary structure level and expressed by serological cross reactivity, may not go in parallel. As it was shown on rhinoviruses, three-dimensional protein structures are generally conserved far longer than primary protein sequences [29]. Thus, there may be differences in the genomic sequence and, hence, in the primary protein structures while the essential polypeptide folding motif is maintained. In accordance with this, the phylogenetic tree including various picornaviruses, some plant spherical RNA viruses and their common hypothetical ancestor—

a primordial cell attachment (ancient receptor) protein [5]—was suggested being based on both primary and tertiary structures of capsid proteins of the compared viruses [30]. This means that the construction of the phylogenetic trees on the only genetic basis of evolutionary divergences may lead to inevitable assumptions simplifying the reality and eliminating important factors.

The suggested presence of "conserved" and "variable" as well as serotype-specific and common portions in the HN gene of avian PMVs [16], the occurrence of high intra-serotype antigenic variability, like that in PMV-2 (Yukaipa-like) case [3, 4] together with minimal inter-serotype relationships [16], and the vice versa situation in the PMV-1 (NDV-like) case, give evidence that the evolutionary changes in HN gene do not seem to be due to monotonous randomic substitutions as it was suggested for the molecular evolution of Enterovirus 70 [20]. Therefore, the suggested mathematical model of elucidation of phlyogenetic relationships on the basis of the antigenic relatedness can be considered as an alternative approach to that based on the genetic relatedness and can be suitable for phylogenetic studies in a large variety of virus groups.

Acknowledgement

The research was supported by the grant No. I-194-80 from the U.S.-Israel Agricultural Research and Development Fund (BARD).

References

1. Aaronson RP, Young JF, Palese P (1982) Oligonucleotide mapping: Evaluation of its sensitivity by computer simulation. Nucleic Acids Res 1: 237–246
2. Air GM, Hall RM (1982) Conservation and variation in influenza genome sequences. In: Nayak D, Fox CA (eds) Genetic variation among influenza viruses, vol 22. Academic Press, New York, pp 45–54
3. Alexander DJ (1980) Avian paramyxoviruses. Vet Bull 50: 737–752
4. Alexander DJ (1982) Avian paramyxoviruses—other than Newcastle disease virus. Worlds Poultry Sci 38: 97–104
5. Argos P, Tsukihara T, Rossmann MG (1980) A structural comparison of concanavalin A and tomato bushy stunt virus protein. J Mol Evol 15: 169–179
6. Both GM, Cheng Hua Shi, Kilbourne ED (1983) Hemagglutinin of swine influenza virus: a single amino acid change pleiotropically affects viral antigenicity and replication. Proc Natl Acad Sci USA 80: 6996–7000
7. Both GM, Sleigh MJ, Cox NJ, Kendal AP (1983) Antigenic drift in influenza virus H 3 hemagglutinin from 1968 to 1980: multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. J Virol 48: 52–60
8. Chatterjee S, Quarcoopome CO, Apenteng A (1970) An epidemic of acute conjunctivitis in Ghana. Ghana Med J 9: 9–11
9. Davenport FM, Hennessy AV, Drescher J, Webster RG (1964) Analytical, serologic, and clinical experiences with the hemagglutinating subunits of influenza A virus. In: Wolstenholme GE (ed) Ciba symposium on cellular aspects of myxovirus infection. CIBA, Basle, pp 272–287
10. Hiti AL, Davis AR, Nayak DP (1981) Complete sequence analysis shows that the hemagglutinin of the H 0 and H 2 subtypes of human influenza virus are closely related. Virology 111: 113–124

11. Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, Van de Pol S (1982) Rapid evolution of RNA genomes. Science 215: 1577–1585
11a. Kida H, Brown LE, Webster RG (1982) Biological activity of monoclonal antibodies to operationally defined antigenic regions on the hemagglutinin molecule of A/Seal/Massachusetts/1/80 (H7N7) influenza virus. Virology 122: 38–47
12. Kida H, Kawaoka Y, Naeve K, Webster RG (1987) Antigenic and genetic conservation of H3 influenza virus in wild ducks. Virology 159: 109–119
13. Kilbourne ED, Gerhard W, Whitaker CW (1983) Monoclonal antibodies to the hemagglutinin Sa antigenic site of A/PR 8/34 influenza virus distinguish biological mutants of swine in infuenza virus. Proc Natl Acad Sci 80: 6399–6402
14. Kono R, Sasagawa K, Ishii K, Sugiura S, Ochi M, Matsumiya H, Uchida Y, Kameyama K, Kaneko M, Sakurai N (1972) Pandemic of new type of conjunctivitis. Lancet i: 1191–1194
15. Krystall M, Young JF, Palese P, Wilson IA, Skehel JJ, Wiley DC (1983) Sequential mutations in hemagglutinins of influenza virus isolates: definition of antigenic domains. Proc Natl Acad Sci USA 80: 4527–4531
16. Lipkind M, Shihmanter E (1986) Antigenic relationships between avian paramyxoviruses. I. Quantitative characteristics based on hemagglutination and neuraminidase inhibition tests. Arch Virol 89: 89–111
17. Lipkind M, Shoham D, Shihmanter E (1986) Isolation of a paramyxovirus from pigs in Israel and its antigenic relationships with avian paramyxoviruses. J Gen Virol 67: 427–439
18. Lipkind M, Weisman Y, Shihmanter E (1987) The isolation of influenza virus from chickens in Israel and studies on its antigenic relationships with other native strains of the same antigenicity by means of monoclonal antibodies. Comp Immunol Microbiol Infect Dis 10: 133–139
19. Masurel N (1968) Antibody response obtained by vaccination with the influenza A/Equi 2 virus in man. Nature 218: 100–101
20. Miyamura K, Tanimura M, Takeda N, Kono R, Yamazaki S (1986) Evolution of Enterovirus 70 in nature: all the isolates were recently derived from a common ancestor. Arch Virol 89: 1–14
21. Mulder J, Masurel N (1958) Pre-epidemic antibody against 1957 strain of Asiatic influenza in serum of older people living in the Netherlands. Lancet 1: 810–814
22. Murphy BR, Webster RG (1985) Influenza viruses. In: Fields BN et al (eds) Virology. Raven Press, New York, pp 1179–1239
23. Nakajima K, Desselberger U, Palese P (1978) Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. Nature 274: 334–339
24. Nakajima S, Nakajima K, Kendal AP (1983) Identification of the binding sites to monoclonal antibodies on A/USSR/90/77 (H1N1) hemagglutinin and their involvement in antigenic drift in H1N1 influenza viruses. Virology 131: 116–127
25. Nei M (1975) Molecular population genetics and evolution. North Holland, Amsterdam, pp 197–202
26. Oxford JS, Abbo H, Corcoran T, Webster RG, Smith AJ, Grilli EA, Schild GC (1983) Antigenic and biochemical analysis of field isolates of influenza B virus: evidence for intra- and interepidemic variation. J Gen Virol 64: 2367–2377
27. Raymond FL, Caton AJ, Cox NJ, Kendal AP, Brownlee GG (1986) The antigenicity and evolution of influenza H1 hemagglutinin, from 1950–1957 and 1977–1983: two pathways from one gene. Virology 148: 275–287
28. Rishe N, Lipkind M (1987) Antigenic relationships between avian paramyxoviruses. II. A combinatorial mathematical model of antigenic kinship. Arch Virol 92: 243–253
29. Rossmann MG, Moras D, Olsen KW (1974) Chemical and biological evolution of a nucleotide-binding protein. Nature 250: 194–199

30. Rossmann MG, Arnold E, Griffith JP, Kamer G, Luo M, Smith TJ, Vriend G, Rueckert RR, Sherry B, McKinlay MA, Diana G, Otto M (1987) Common cold viruses. Trends Biochem Sci 12: 313–318
31. Scheid A, Caliguiri LA, Compans RW, Choppin PW (1972) Isolation of paramyxovirus glycoproteins. Association of both hemagglutinating and neuraminidase acivities with the larger SV 5 glycoprotein. Virology 50: 640–651
31a. Scholtissek C, Rohde W, Von Hoyningen, Rott R (1978) On the origin of the human influenza virus subtypes H2N2 and H3N2. Virology 87: 13–20
32. Shimizu K, Shimizu YK, Kohama T, Ishida N (1974) Isolation and characterization of two distinct types of HVJ (Sendai virus) spikes. Virology 62: 354–360
33. Sleigh, MJ, Both PA, Underwood PA, Bender VJ (1981) Antigenic drift in the hemagglutinin of the Hong Kong influenza subtype: correlation of amino acid changes with alteration in viral antigenicity. J Virol 37: 845–853
34. Tanimura M, Miyamura K, Takeda N (1985) Construction of a phylogenetic tree of Enterovirus 70 Japan. J Genet 60: 142–147
35. Tozawa H, Watanabe M, Ishida N (1973) Structural components of Sendai virus: serological and physicochemical characterization of hemagglutinin subunit associated with neuraminidase activity. Virology 55: 242–253
36. Webster RG, Laver WG, Air GM, Schild GC (1982) Molecular mechanisms of variation in influenza viruses. Nature 296: 115–121
37. Yin-Murphy MM (1984) Acute hemorrhagic conjunctivitis. Prog Med Virol 29: 23–44
38. Young JF, Desselberger U, Palese P (1979) Evolution of human influenza A viruses in nature: sequential mutation in the genome of new H1N1 isolates. Cell 18: 73–83

Authors' address: Prof. M. Lipkind, Kimron Veterinary Institute, Beit Dagan P.O. Box 12, 50250 Israel.

Received December 10, 1987

Printed in Austria