SCIENTIFIC CORRESPONDENCE

HIV/HTLV gene nomenclature

SIR—The complexities of the genomes of human retroviruses (the human T-cell leukaemia viruses, HTLV-I and HTLV-II, and the AIDS-causing human immunodeficiency viruses, HIV-1 and HIV-2) are being unravelled at a rapid pace which is likely to continue and expand. In addition to containing a large ensemble of positive and negative regulatory genes that orchestrate virus expression, these viruses are also remarkable in that they seem to have converged onto parallel regulatory pathways. Two of the regulatory genes of the immunodeficiency viruses are analogous to the two regulatory genes of the leukaemia viruses, although their detailed mechanisms of action may be quite different. Deciphering the modes of action of the regulatory genes of these viruses is crucial to the understanding of their pathogenesis as well as to development of therapeutic agents. Because of the tremendous activity in this field, more than one name has sometimes been given to a single gene and the same name may also apply to more than one gene. In the interest of the many new investigators entering the field for the first time, we feel it is important that we reach a standard nomenclature for all known genes of HIV and HTLV. We propose the scheme outlined in the table.

ROBERT GALLO
FLOSSIE WONG-STAAL
National Cancer Institute, NIH,
Bethesda, Maryland 20892, USA
LUC MONTAGNIER
Department of Virology,
Institut Pasteur,
*75724 Paris Cedex 15, France*
WILLIAM A. HASELTINE
Dana-Farber Cancer Institute,
Boston, Massachusetts 02115, USA
MITSUAKI YOSHIDA
Department of Viral Oncology,
Cancer Institute, Tokyo 170, Japan

| Proposed name (and derivation)	Previous names	Molecular mass ($\times 10^{-3}$)	Known function
HTLV-I and HTLV-II genes:			
tax_1 (transactivator)	*x-lor, p40x, tat*$_1$	41, 41, 42	Transactivator of all viral
tax_2	*tat*$_2$, *TA*	38	proteins
rex_1 (regulator of expression	*pp27x, tel*	27	Regulates expression
rex_2 virion proteins)		25	of virion proteins
HIV genes:			
tat (transactivator)	*tat-3, TA*	14	Transactivator of all viral proteins
rev (regulator of expression of virion proteins)	*art, trs*	19, 20	Regulates expression of virion proteins
vif (virion infectivity factor)	*sor, A, P', Q*	23	Determines virus infectivity
vpr (R)	*R*	?	Unknown
nef (negative factor)	*3'orf, B, E', F*	27	Reduces virus expression, GTP-binding
vpx (X) (only in HIV-2 and SIV)	*X*	16, 14	Unknown
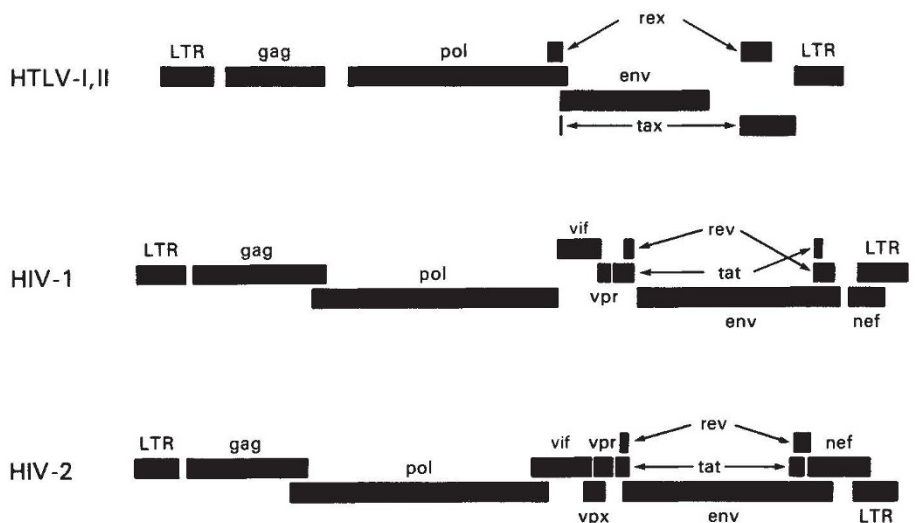
Vpr and *vpx* are temporary names and may be changed when more information about their functions is available. Subscripts 1 and 2 would be used to distinguish genes of HIV-1 and HIV-2 (for example, *rev*$_1$ and *rev*$_2$). It is expected that genes of the simian viruses (STLV-I, SIV) would follow similar nomenclature with the subscripts STLV or SIV as appropriate.

Estimating the incubation period for AIDS patients

SIR—The nonparametric analyses of the data on transfusion-related AIDS considered by Medley *et al.*[1] indicate problems of identifiability. With data obtained by retrospective determination of the time of infection for diagnosed AIDS cases, it is only possible to estimate the early part of the incubation distribution up to a constant of proportionality. The same applies to the total number of infections by blood transfusion before any given time. The transfusion data themselves are unable to discriminate between high infection rates coupled with long incubation times on the one hand, or low infection rates and short incubation times on the other.

As do Medley *et al.*[1], we postulate a function $h(x)$ which specifies the increase over time of the number of HIV-infected individuals who eventually develop AIDS, and a probability density function $f(s)$ for the incubation time of those individuals. The corresponding likelihood function can be maximized jointly with respect to h and f. As the likelihood depends only on the product of h and f, it is not possible to estimate either of these fuctions completely; they may be individually estimated only up to constants of proportionality c and c^{-1}, respectively. Nonparametric estimates of the proportion of eventual AIDS cases that are diagnosed within t years of infection, $F(t) = \int_0^t f(u)\mathrm{d}u$, are given in the figure for the three age groups considered by Medley *et al.*. In this figure we show the estimates of $F(t)$ so that for each group, $c = F(7.5)$. For the children, the levelling of the estimate of $F(t)$ by about 3.5 years suggests that the whole of the distribution of incubation times has been seen; it may then be reasonable to suppose that $c = 1$ but, as also noted by Medley *et al.*, a second wave of incubation times that exceed 7.5 years is not excluded by these data. For the other two age groups, there is nothing in the transfusion data themselves to suggest a value for c. As a consequence, it is impossible to place any upper bound on the median incubation time. To estimate this,
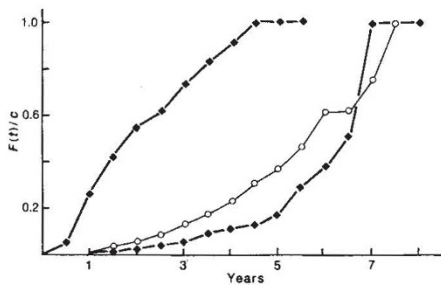
Nonparametric estimates of $F(t)$, the probability that the incubation time is less than t for children (◆, top curve), adults (o) and elderly patients (◆, bottom curve). The unspecified constant c is the probability that the incubation time is less than 7.5 years. Note that c cannot be estimated from the transfusion data alone, and that different values of c could apply to the separate age groups.

it is necessary to obtain an external estimate of c.

Medley *et al.* comment that estimates of incubation time characteristics such as means and medians are based on extrapolation, and this is true. In fact the difficulties are even more severe. Because of the identifiability problems, even low percentiles of the incubation distribution cannot be estimated without additional assumptions. With parametric assumptions, such as those used by Medley *et al.*, identifiability problems still persist. Thus although point estimates of median incubation times can be obtained, the confidence intervals associated with those estimates are very broad. Under a Weibull model for $f(s)$ and an exponential model for $h(x)$, point estimates (with 95 % confidence intervals) of the median incubation times in years are 2.0 (1.4, 4.0), 7.3 (4.6, ∞), and 5.8 (4.3, ∞) for the young, middle, and elderly age groups, respectively. (Here, ∞ indicates a large but finite upper terminus.) So, not much faith can be placed in the point estimates obtained. Similar remarks apply to the estimation of the total number of infected individuals.

J.D. KALBFLEISCH
J.F. LAWLESS
Department of Statistics and Actuarial Science,
University of Waterloo,
Waterloo, Ontario, Canada N2L 3G1

MEDLEY *ET AL.* REPLY — Kalbfleisch and Lawless are correct in their assertion that if the probability distribution of the incubation period of AIDS is truncated at some time point T (all observations beyond the present time, T, are unobserved), then only the conditional distribution up to T can be estimated without parametric assumption concerning the form of the full distribution of incubation periods. For the matters of concern to epidemiologists and public-health planners, namely understanding and prediction of observed trends, the conditional distribution is of rather little interest and hence parametric formulation is inevit-

able. We emphasized the obvious hazards. of this approach by showing explicitly the different estimates arising from two plausible parametric assumptions (the Weibull and gamma distributions)[1]. We agree that the point estimates of median and mean incubation periods have large errors, and that these are best assessed by likelihood ratio arguments.

In our previous papers[1,2] we noted the importance of updating estimates of the distribution of the incubation period as more data become available. The Centers for Disease Control (CDC) in Atlanta provided a recent data set covering reports of AIDS cases received up to 8 February 1988. In these data there are 560 cases with certain dates of transfusion (infection), diagnosis and report, compared with 297 in the previous data. We have so far considered only those cases who were older than 12 years at diagnosis (512 cases).

In this preliminary analysis we have adopted the same approach as before [1,2]. We assume parametric forms for the distributed incubation period and the distributed reporting lag, and an exponential increase in the incidence of infective transfusions up to early 1985 when routine screening of donated blood was introduced in the United States. We assume that the infection rate has been constant since then. CDC estimate that even after the introduction of screening, about 400 infective donations per year are tested during the period before antibodies against HIV are produced, and therefore appear negative (M. Morgan, personal communication). We are able to discriminate between different parametric forms for the incubation period by comparison of the independent assessment of the rate of infective transfusions with that
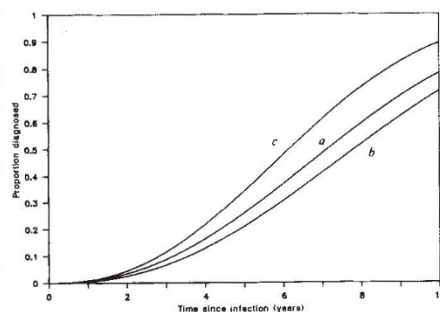
Comparison of the cumulative distributions (Weibull) for the incubation period of AIDS in transfusion-associated patients estimated from: *A*, all adults (>12 yr) notified to CDC by 8 February 1988; *b*, people older than 4 yr and less than 60 yr; and *c*, all cases, regardless of age. Estimates *b* and *c* are from a data set in which people were diagnosed before July 1986 and notified to CDC by January 1987. Exact comparisons are not possible because of the difference in age groupings between the two data sets. The most notable features are the similarity of shape between the distributions, and the fact that the means and medians are similar between *a* and *b* and between *a* and *c* after 2 yr further observation.

estimated from the data.

We have considered the Weibull and gamma distributions as possible candidates for the incubation-period distribution. The estimated means (medians) are 7.59 yr (7.32 yr) and 24.07 yr (20.84 yr), respectively, and, as previously, the log-likelihood values do not allow a distinction to be made between the distributions. The estimated annual rates of infective transfusion since the introduction of screening, however, are 333 and 2,161, respectively, which leads us to reject the gamma distribution in favour of the Weibull as a description of the data.

Also of interest is the observation that the estimate of the overall mean for the adult age classes is similar to that reported in previous publications [1,2]. It is difficult to make direct comparisons as the format of the data has altered. The similarity of the estimates may suggest that the estimate is approaching the true mean of the full distribution, however. Past estimates of this mean have tended to increase with the length of the observation period of transfusion-associated AIDS cases [1-3].

G.F. MEDLEY
R.M. ANDERSON
Department of Pure and Applied Biology,
Imperial College London,
London, SW7 2BB, UK

D.R. COX
Department of Mathematics,
Imperial College London,
London SW7 2BB, UK

L. BILLARD
Department of Statistics,
University of Georgia,
Athens, Georgia 30602, USA

1. Medley, G.F., Anderson, R.M., Cox, D.R. & Billard, L. *Nature* **328**, 718–721 (1987).
2. Medley, G.F., Billard, L. Cox, D.R. & Anderson, R.M. *Proc. R. Soc.* B**233**, 367–377 (1988).
3. Lui, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3051 (1986).

Fish farming and aquaculture

SIR—The article "Fish farming and influenza pandemics" by Christoph Scholtissek and Ernest Naylor (*Nature* **331**, 215; 1988) has been widely reported in Asian and Western newspapers. We feel that it deserves clarification because of adverse effects it might have on the promotion of aquaculture in developing countries.

The authors claim that widespread development of integrated farming systems involving pigs, poultry and fish could lead to the creation of influenza A viruses with new surface antigens against which the human population does not possess specific neutralizing antibodies. The article suggests that pigs may be "mixing vessels" where normally separate avian and human influenza virus reservoirs meet, and where genetic reassortment takes place between them, giving rise to new human pandemic influenza strains.

We share the authors' concern that the