JOURNAL OF VIROLOGY, Dec. 1978, p. 972-991
0022-538X/78/0028-0972$02.00/0

Vol. 28, No. 3
Printed in U.S.A.

Nucleotide Sequence Relationships Between the Genomes of an Endogenous and an Exogenous Avian Tumor Virus

JOHN M. COFFIN,* MARGARET CHAMPION, AND FRANCIS CHABOT

Department of Molecular Biology and Microbiology and Cancer Research Center, Tufts University School of Medicine, Boston, Massachusetts 02111

Received for publication 25 July 1978

We have used mapping of large T_1 oligonucleotides to examine the genome of Rous-associated virus-0 (RAV-0), an endogenous virus of chickens, and to compare it with that of Prague strain Rous sarcoma virus, subgroup B, (Pr-RSV-B), an exogenous sarcoma virus. To extend the sensitivity of such comparisons, we have developed a system of nucleic acid hybridization and hybridization-competition combined with fingerprinting. This method allows us to estimate the relative degree of relatedness of various portions of the viral genomes. From the results of this study, we have concluded that the genomes of Pr-RSV-B and RAV-0 are related in the following way. The 5'-terminal half of the genomes (corresponding to the *gag* and *pol* regions) is virtually identical, with only scattered single nucleotide differences. This region is followed by a region comprising 25 to 30% of the genome (the *env* region) which contains substantial nucleotide sequence differences, most or all of which are due to single base changes. The *env*-coding region can be further subdivided into three regions: a more variable region probably containing sequences coding for subgroup specificity, flanked by relatively common sequences on each side. To the 3' side of the *env* region, the RAV-0 genome contains a very short sequence not found in Pr-RSV-B, whereas the Pr-RSV-B genome contains a much longer unrelated sequence. The central portion of this sequence comprises the *src* gene as defined by transformation-defective mutants. Particularly striking is the absence, in the RAV-0 genome, of any nucleotide sequence related to the "c region" found very near the 3' end of all exogenous tumor viruses. Both the Pr-RSV-B and RAV-0 genomes contain the identical terminally redundant sequence of 21 nucleotides near each end of the genome.

Endogenous oncovirus genomes have been observed in the DNA of uninfected cells of many vertebrates. In chickens, these viruses can be classified into two types: complete and incomplete. Incomplete "virus" genomes lead to the production of proteins (chicken helper factor and group-specifc antigens) closely related to structural proteins of exogenous genomes (5, 16), but not virus particles; complete genomes give rise, in cells from some lines of chickens, to infectious virus (for a review, see H. Robinson, Curr. Top. Microbiol. Immunol., in press). In contrast to the situation with mice, only a few very closely related or identical complete endogenous viruses have been found in chicken cells. The best studied of these viruses is Rous-associated virus-0 (RAV-0) spontaneously produced by cells of certain line 7 or line 100 chickens (12, 45). This virus is closely related to exogenous viruses of chickens as determined by antigenicity and polyacrylamide gel electrophoresis of virion proteins (27, 33, 37) and by cross-hybridization of virion RNA and DNA (23, 29, 39).

In spite of this close relationship, RAV-0 differs from exogenous avian oncoviruses, such as Rous sarcoma virus (RSV) and other strains of RAV, in at least three major properties; (i) it does not induce any disease in chickens made viremic after either spontaneous expression of the virus or exogenous infection of the animals (28); (ii) RAV-0 virion glycoproteins confer the host range and antigenicity of subgroup E to the virus (48)—a subgroup never found in viruses isolated from naturally occurring tumors; and (iii) it is subject to a partial growth restriction in certain types of cells (25, 34). The molecular basis and interrelationship of these differences are poorly understood.

The close chemical relationship of exogenous and endogenous tumor viruses makes it clear that a large portion of the genomes of these viruses must have evolved from a common ancestor. As a first step toward attempting to infer the mechanisms involved in such evolution,

we have used oligonucleotide mapping combined with nucleic acid hybridization to examine and compare the genomes of RAV-0 and a "standard" strain of RSV. Our results, presented below, imply a role for both recombination and point mutation selection events in specific parts of the genome in the processes that give rise to avian tumor virus genomes.

A preliminary report of part of this work has been previously presented (8).

MATERIALS AND METHODS

Cells and viruses. Prague strain RSV subgroup B (Pr-RSV-B) was derived by cloning infected cells in soft agar from a stock previously used by us for physical mapping studies (7). This clone, designated 3H, was identical to the standard virus previously used by us and others (1, 7, 20), as determined from its fingerprint (Fig. 1), and had a very low proportion of transformation-defective (*td*) virus. It was routinely grown in SPAFAS C/E chf$^-$gs$^-$chicken embryo fibroblast cultures. All cultures used to grow the virus for complementary DNA (cDNA) or RNA were derived from the same embryo.

RAV-0 was harvested from the supernatant medium of V$^+$ C/O line 100 chicken cells (12). All cultures were derived by passage of a culture originally provided by G. Cooper. We have also examined virus spontaneously released from closely related V$^+$ C/A line 7 cells, as well as virus induced by bromodeoxyuridine from line 7 cells (35). All of these viruses yielded identical fingerprints.

Viral RNA and DNA. ^{32}P-labeled viral RNA was prepared from the supernatant medium of cells labeled with $^{32}PO_4$ (5 or 10 mCi per 100-mm culture dish) for 2 days and harvested daily for 4 days after the beginning of the labeling period. 70S RNA was prepared from virions as previously described (7). For oligonucleotide mapping, RAV-0-producing cell cultures were labeled with [^{3}H]uridine (20 Ci/mmol, New England Nuclear Corp.) under the same conditions, except that virus was harvested at 12-h intervals. Such virus has a ratio of degraded to intact virion RNA suitable for physical mapping without further treatment (7).

For preparation of cDNA, virions were purified by equilibrium density gradient centrifugation. Typical reactions contained 1.5 mg of virion protein in 2 ml of 50 mM Tris-hydrochloride (pH 8.3), 5 mM $MgCl_2$, 30 mM dithiothreitol, 0.02% (wt/vol) Triton X-100, 2 mM each of dATP, dGTP, dCTP, and dTTP, and 2 mg of DNase-digested salmon sperm DNA per ml (43). The dTTP was adjusted with [^{3}H]dTTP (New England Nuclear) to a specific activity of 2×10^6 cpm/μmol to quantitate the product and monitor its purification. Reactions were incubated at 37°C for 16 to 18 h, and the DNA synthesized was purified by sodium dodecyl sulfate-Pronase digestion, phenol extraction, alkaline hydrolysis, and chromatography on Sephadex G-100 essentially as previously described (31). All purified cDNA preparations were tested for the uniformity of representation of viral sequences by fingerprinting hybrids between the cDNA and 70S [^{32}P]RNA prepared at ratios which gave approximately 50% of the RNA as hybrid.

Strong-stop DNA (17) prepared from Pr-RSV-B was a gift from W. Haseltine.

Nucleic acid hybridization. Small-scale annealing mixtures contained 10 μl of 0.5 M NaCl-10 mM Tris-hydrochloride (pH 7.5) (NT buffer), 20 μg of repurified yeast RNA, and the indicated [^{32}P]RNA and cDNA. The annealing was carried out in small conical polypropylene tubes with 10 μl of paraffin oil added to prevent evaporation of the sample. Incubation was in a 66°C polyethylene glycol bath (to prevent condensed water from leaking into the tubes) for 18 to 24 h. After annealing, 100 μl of NT buffer containing 5 μg of RNase A and 5 U of RNase T_1 was added, and the mixture was incubated at 37°C for 30 min. The sample was adjusted to 250 μl containing 10 μg of yeast RNA and 6% trichloracetic acid and held on ice for 10 min, and the precipitated nucleic acid was recovered by filtration through a filter (Millipore Corp.) and quantitated by liquid scintillation counting. All values are expressed as percentage of input radioactivity on the filters after correction for a background annealing with no added DNA (2 to 5%). Annealings with the amount of salmon sperm primer DNA expected to be present if it copurified completely with the cDNA yielded no detectable hybrid.

Large-scale hybridizations for fingerprints were carried out in 100 μl of NT buffer with 20 μg of yeast RNA and amounts of viral RNA and DNA determined from preliminary small-scale experiments. After annealing (66°C, 18 to 24 h) the samples were digested with either 10 U of RNase T_1 or 5 U of RNase T_1 plus 5 μg of RNase A at 37°C for 30 min. The samples were adjusted to 0.2% (wt/vol) sodium dodecyl sulfate, incubated at 66°C for 10 min to disrupt aggregated digestion products, and applied to a small column (0.6 by 10 cm) of Sephadex G-100 in NT buffer plus 0.2% sodium dodecyl sulfate (NTS buffer) essentially as described previously (31). Fractions containing the void volume (usually between 0.9 and 1.3 ml) were located by Cherenkov counting, pooled, and extracted twice with redistilled phenol after adding 100 μg of yeast RNA. After precipitation with ethanol, the hybrids were dissolved in 100 μl of 20 mM Tris-hydrochloride (pH 7.5)-5 mM EDTA, denatured by heating for 30 s in a boiling-water bath, and precipitated again before RNase T_1 digestion and two-dimensional gel electrophoresis. When tested for background by analyzing annealing mixtures with no added DNA, this procedure yielded 1 to 2% of input radioactivity in the void volume of the Sephadex columns. In no case were any oligonucleotides {except polyadenylic acid [poly(A)]} detected in subsequent fingerprints of such "hybrids", under conditions where 0.1% of the amount of radioactivity in an oligonucleotide in a control fingerprint would have been observable. This combination, therefore, has an extremely low background, and it is possible to analyze sequences comprising 1% of the viral genome with no interference (see Fig. 7).

3'-Terminal fragments. The procedure for preparation of poly(A)-containing fragments of virus RNA has been described in detail (7). Briefly, 70S [^{32}P]RNA was incubated at 50°C for 15 min in 0.5 M Na_2CO_3, neutralized, precipitated with ethanol, and applied to a column of polyuridylic acid [poly(U)]-Sephadex in NTS buffer. The bound [poly(A)-containing] material

was eluted with 90% formamide, 10% (0.5 M Tris-hydrochloride [pH 7.5]–5% sodium dodecyl sulfate), and precipitated twice with ethanol before use.

3′-Terminal poly(A)-linked oligonucleotide (10) was prepared from such fragments by complete digestion with RNase T_1 and subsequent chromatography on poly(U)-Sephadex as described above.

Oligonucleotide analysis. Samples were digested with RNase T_1 and analyzed by two-dimensional gel electrophoresis (13) with the modifications described (2, 7). In general, electrophoresis was until the faster dye (bromophenol blue) had migrated 20 cm in the first dimension and 18 cm in the second. In cases where RNA of substantially lower complexity than the whole genome was analyzed, migration in the second dimension was for 12 cm to retain smaller oligonucleotides in the gel. Large oligonucleotides were located by autoradiography and numbered according to previously established conventions (7, 21). For further analysis, gel pieces containing large oligonucleotides were cut with a cork borer and soaked in 0.5 ml of water containing 60 μg of yeast RNA per ml. After 48 h or more at 4°C, the eluant was removed, and the gel piece was rinsed with 0.3 ml of the same solution. The pooled eluant was dried under vacuum in a polystyrene tube (10 by 75 mm) and digested with 10 μl of a solution containing 150 μg of RNase A per ml of 20 mM Tris-hydrochloride (pH 7.5)–2 mM EDTA for 30 min at 37°C. RNase A digestion products were separated by ionophoresis on DEAE paper (3) and quantitated relative to the G-containing product (2).

Physical mapping of the large T_1 oligonucleotides was performed by the procedure of Coffin and Billeter (7) as modified (9). After denaturation of randomly fragmented RAV-0 70S [^{3}H]RNA and separation of size classes on a sucrose gradient, the poly(A)-containing portions were isolated on poly(U)-Sephadex as described above and fingerprinted after the addition of RAV-0 [^{32}P]RNA as marker for localization and quantitation of the oligonucleotides. The molar yield of each oligonucleotide relative to unfractionated control 70S [^{3}H]RNA was computed from the formula: relative yield = (^{3}H yield/^{32}P yield)$_{sample}$/(^{3}H yield/^{32}P yield)$_{control}$ where the yield is defined as the amount of radioactivity in an oligonucleotide divided by the amount present immediately before RNase T_1 digestion, with all figures corrected for background determined by counting five blank areas from the same gel.

RESULTS

Large T_1 oligonucleotides of Pr-RSV-B and RAV-0. To a first approximation, the extent of relatedness of two RNA molecules can be estimated by comparison of large oligonucleotides obtained by RNase T_1 digestion, because regions of identical nucleotide sequence will yield identical oligonucleotides. Figure 1 shows a comparison of the genomes of Pr-RSV-B and RAV-0. ^{32}P-labeled 70S RNA from each virus was digested with RNase T_1 and fingerprinted by two-dimensional gel electrophoresis. The large oligonucleotides were numbered consistently with previous publications (7, 21), and their identity was confirmed by digestion with RNase A (Table 1). It is readily apparent that these two virus genomes share large regions of identical or very similar sequence. Of the pure, numbered oligonucleotides, 14 are common to the two viruses, whereas 15 are unique to RAV-0 and 19 are unique to the Pr-RSV-B genome.

Most of the common oligonucleotides have previously been mapped in the 5′ half of the Pr-RSV-B genome (7). To confirm this order for RAV-0 and to locate dissimilar oligonucleotides,

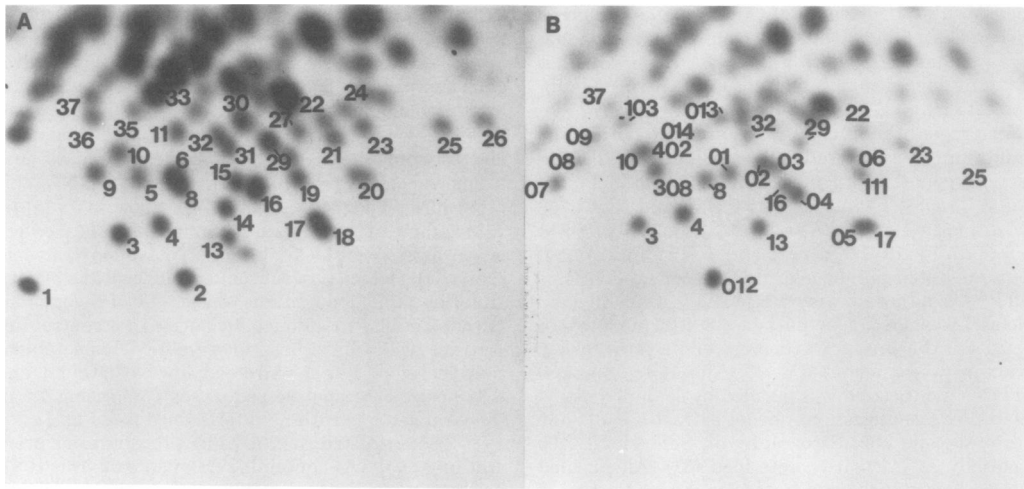

FIG. 1. *Fingerprints of RAV-0 and Pr-RSV-B genomes. Samples of ^{32}P-labeled 70S RNA from either (A) Pr-RSV-B grown on C/E cells or (B) RAV-0 spontaneously released from line 100 cells were digested with RNase T_1, and the digests were analyzed by two-dimensional gel electrophoresis. The numbering is consistent with previous publications. All numbered oligonucleotides were taken for further analysis (Table 1).*

TABLE 1. *Large T_1 oligonucleotides of Pr-RSV-B and RAV-O 70S RNA*

Oligonucleotide[a]	Composition[b]	Presence in:		Oligonucleotide[a]	Composition[b]	Presence in:	
		Pr-RSV-B	RAV-0			Pr-RSV-B	RAV-0
1	$3A_4C$, A_2U, 4AC, C, 2U, G	+	−	26	3AU, 8U, G	+	−
2	$2A_2U$, A_2C, 2AU, AC, 7C, 6U, G	+	−	27	A_2U, AU, 3C, 5U, G	+	−
				29	AU, AC, 4C, 4U, G	+	+
3	A_2C, AU, 2AC, 7C, 6U, A_2G	+	+	30	2AU, AC, 3C, 3U, G	+	+
4	A_3C, AU, AC, 9C, 5U, G	+	+	31	A_3C, AC, 3C, 6U, A_3G	+	−
5	A_3C, AU, AC, 8C, 5U, G	+	−	32	A_2U, AC, 5C, 4U, G	+	+
6A + B	2AU, 5AC, 13C, 8U, AG, G	+	−	33	AU, 2AC, 6C, 5U, G	+	−
8	A_4C, 6C, 6U, A_2G	+	+	35	A_2C, 5C, 3U, G	+	−
9	A_2C, AU, 2AC, 6C, 2U, G	+	−	36	A_2C, AU, 2AC, 2C, U, AG	+	−
10	A_2C, AU, 2AC, 4C, 2U, G	+	+	37	2AC, 4C, U, G	+	+
11	AC, 6C, 5U, A_4G	+	−	103	A_5C, AU, 2AC, C, U, G	−	+
13	m^7GpppGmCCAUUUUACC-AUUCACCACAUUG[c]	+	+	111	A_2U, AU, 4C, 6U, G	−	+
				308	AU, 3AC, 6C, 3U, G	−	+
14	A_2U, A_2C, 2AU, AC, 3C, 4U, G	+	−	01	AU, 9C, 5U, G	−	+
				02	A_2C, 3AC, 2C, 6U, G	−	+
15	AU, 8C, 6U, G	+	−	03	$3A_2U$, 2AU, AC, 5C, 6U, AG	−	+
16A	4AU, 2AC, 4C, 6U, AG	+	+	04	A_2U, A_2C, 5C, 9U, AG	−	+
17	AU, 7C, 9U, G	+	+	05	A_2U, 2AU, 2C, 5U, G	−	+
18	A_2U, 2AU, 3C, 4U, G	+	−	06	A_2U, AU, AC, 3C, 6U, G	−	+
19	AU, 2AC, 3C, 7U, G	+	−	07	A_2U, 2AC, 10C, U, A_2G	−	+
20	A_2U, AU, 3C, 7U, G	+	−	08	A_3U, A_3C, AU, AC, 6C, G	−	+
21	A_2U, AC, 2C, 6U, G	+	−	09	A_2C, AU, 2AC, 4C, 2U, AG	−	+
22	A_2U, 3AU, 2C, 4U, G	+	+	012	A_2U, A_2C, 2AU, 2AC, 7C, 11U, AG	−	+
23	2AU, 3C, 5U, G	+	+				
24	AU, AC, 2C, 5U, G	+	−	013	A_2C, AC, 5C, 3U, G	−	+
25	AU, 2C, 6U, AG	+	+	014	AU, AC, 9C, 4U, G	−	+

[a] The numbers correspond to those in Fig. 1 and are consistent with previous numbering systems.

[b] Determined by complete digestion with RNase A.

[c] The exact sequence of Pr-RSV-B oligonucleotide 13 is known from the sequence of strong-stop DNA (18, 38).

a physical map of large T_1 oligonucleotides in the RAV-0 genome was prepared. ^{3}H-labeled RAV-0 70S RNA was denatured and separated by sedimentation in a sucrose gradient into size classes of 29 to 35S (largely full-length molecules), 22 to 28S, 14 to 21S, and 6 to 13S. We have previously shown that the size classes of RNA less than full length consist of random breakdown products of genome RNA (7). All fractions were chromatographed on poly(U)-Sephadex to isolate the original 3′ ends. The poly(A)-containing fraction of each size class was mixed with ^{32}P-labeled (unfractionated) RAV-0 70S RNA, and the mixtures were fingerprinted along with a control mixture containing the same [^{32}P]RNA and a sample of the unfractionated [^{3}H]RNA. After fingerprinting, the ratio of ^{3}H to ^{32}P in each oligonucleotide was normalized to the same value from the control fingerprint. This procedure allows us to locate the oligonucleotides in their relative order with reasonable precision (7, 19). The results of these experiments are shown in Fig. 2.

Oligonucleotides were scored as common if: (i) they had identical mobilities in the two dimensional gels; (ii) they yielded identical digestion products with RNase A (except for small discrepancies in the number of C and U residues); and (iii) their map positions were nearly identical. Note that the left-hand (5′) half of these two genomes (constituting the *gag* and *pol* genes) contains 10 of the 13 mapped common oligonucleotides. The remaining three common oligonucleotides are all in the *env* region. One of these three (oligonucleotide 4) is found in all strains of avian leukosis virus (subgroups A to E) so far examined; another (oligonucleotide 10) has been observed in all endogenous and many exogenous avian leukosis virus strains regardless of subgroup. Because there is a small uncertainty in the order of large oligonucleotides obtained by our procedure, the RAV-0 and Pr-RSV-B maps were adjusted slightly to put identical oligonucleotides in identical (or analogous) positions. This adjustment was performed by examining the original data for both maps and using the estimate of position considered to be more reliable. Small errors (one to two oligonucleotides) in relative location probably still remain, but those do not affect conclusions from the experiments that follow. The adjusted oligonucleotide maps are shown in Fig. 2E.

Analysis of relationships by hybridizations and fingerprinting. The fingerprinting method used here is very sensitive to small differences in nucleotide sequence. At least four

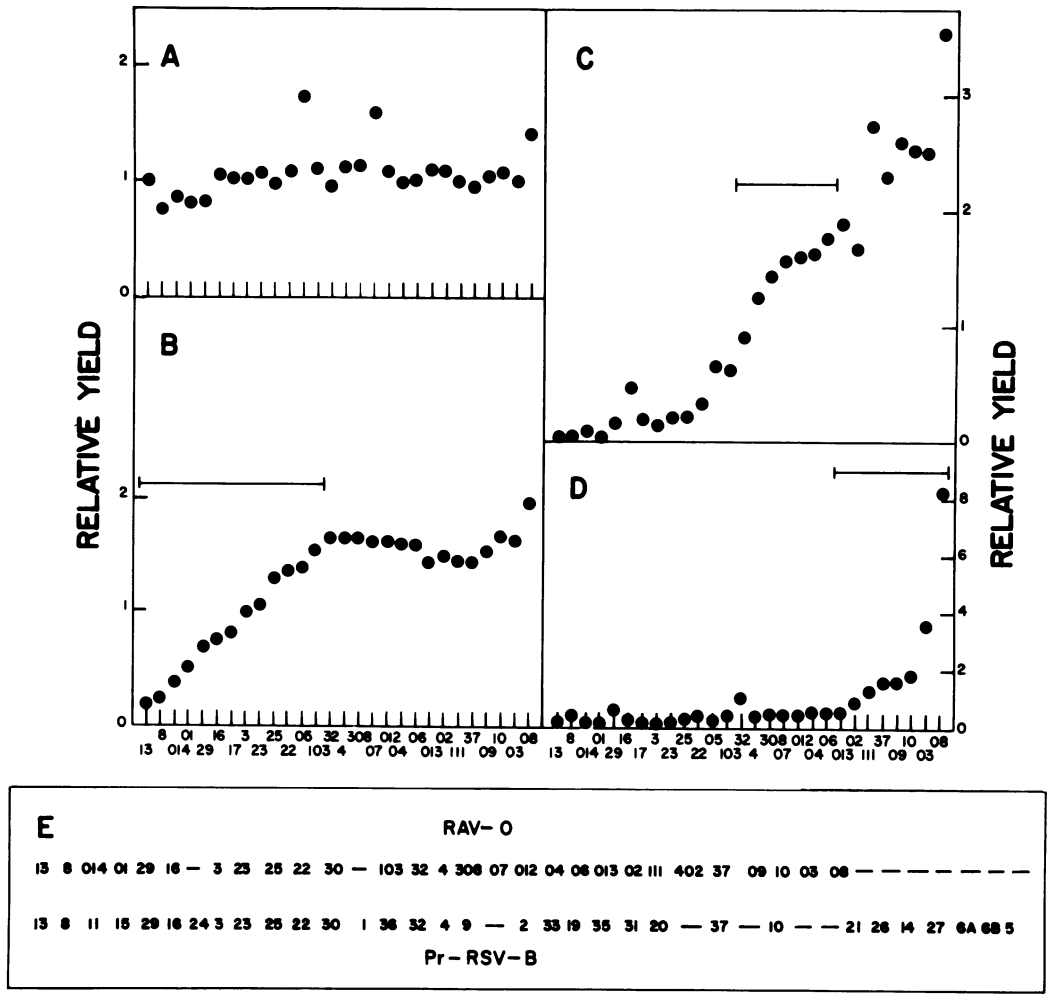

FIG. 2. *Physical mapping of the RAV-0 genome. ^{3}H-labeled RAV-0 70S RNA was denatured and fractionated into size classes by sedimentation in sucrose density gradients. Each size class was chromatographed on poly(U)-Sephadex, the bound [poly(A)-containing] material was mixed with 5.6 × 10^5 cpm of RAV-0 70S [^{32}P]RNA, and the mixture was digested with RNase T_1 and fingerprinted. The relative ^{3}H and ^{32}P yields of each oligonucleotide were normalized relative to a control mixture of the same [^{32}P]RNA and 3.6 × 10^5 cpm of unfractionated [^{3}H]RNA. The oligonucleotides were placed in their most probable order according to the relative molar yields of the oligonucleotides in the regions of varying yields (shown by the horizontal bar). S values were determined relative to 28 and 18S chicken rRNA. The fractions used were: (A) 29 to 35S (input to fingerprint, 1.6 × 10^5 cpm); (B) 22 to 28S (1.2 × 10^5 cpm); (C) 14 to 21S (1.1 × 10^5 cpm); (D) 6 to 13S (3.9 × 10^4 cpm). (E) shows the final working maps of Pr-RSV-B and RAV-0 as described in the text.*

pairs of oligonucleotides (9 and 308, 2 and 012, 15 and 01, and 20 and 111) probably differ by only one or a few nucleotide changes, as judged by their mobilities (Fig. 1), patterns of RNase A digestion products (Table 1), map locations (Fig. 2), and the observation that recombinants between Pr-RSV-B and RAV-0 contain one member of each pair or the other but never both (P. Tsichlis, personal communication). Because of its sensitivity to small nucleotide differences, direct comparison of fingerprints cannot easily be used to assess the extent of sequence relationships in regions of nonidentical sequence. To surmount this problem, we have taken advantage of the relative insensitivity of nucleic acid hybridization to small differences in nucleotide sequence. Our approach is outlined in Fig. 3. [^{32}P]RNA from one virus is annealed with unlabeled cDNA complementary to the genome of another virus. An imaginary example of such a heteroduplex is shown in Fig. 3A. The hybrids are then treated with RNase T_1 or RNase A plus T_1 and separated from relatively low-molecular-weight digestion products by chromatography

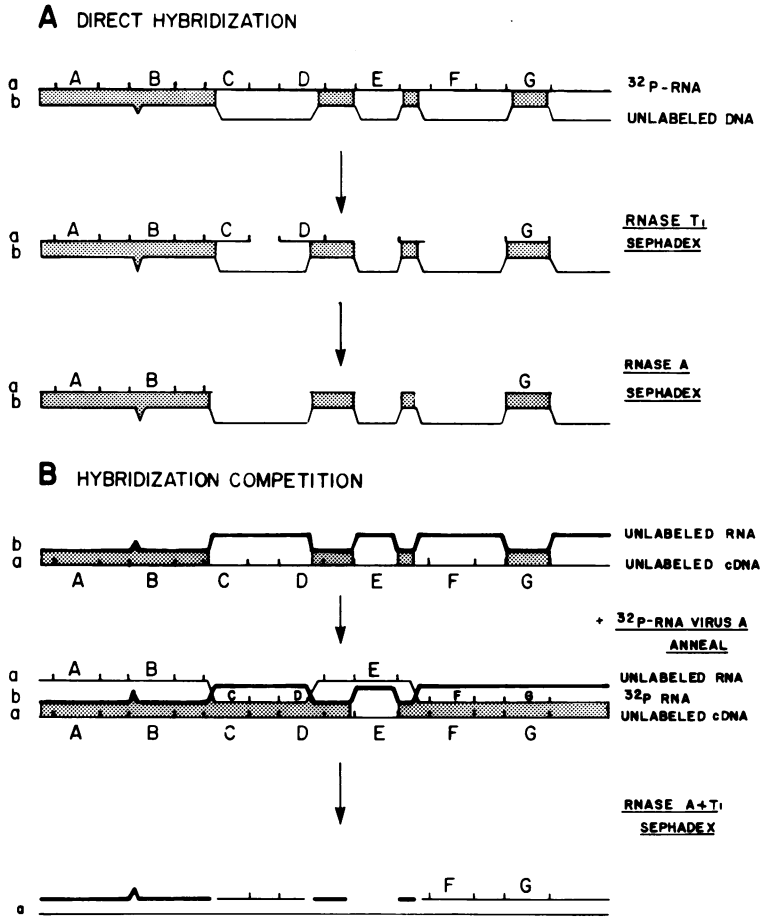

FIG. 3. *Relationship analysis of hybridization and fingerprinting. (A) Direct hybridization. At the top is shown an imaginary heteroduplex between cDNA (thin line) of virus b and [^{32}P]RNA (thick line) of virus a. The positions of six large oligonucleotides are shown (A to G). Shading between the lines indicates regions of base pairing. After treatment with RNase T_1 and chromatography on Sephadex (middle), only regions of essentially complete nonhomology are removed from the hybrid. After treatment with RNase A and T_1 and Sephadex chromatography (bottom), only oligonucleotides with essentially complete homology are preserved intact. (B) Hybridization-competition. The top line shows an imaginary heteroduplex between the same pair of viruses, except with the cDNA from virus a and unlabeled RNA (thick line) from virus b. The complementary nucleotide sequences to the 6 oligonucleotides are shown as A to G. Nonhybridized areas above a certain length are free to anneal to [^{32}P]RNA of virus a in a subsequent step (middle). After RNase digestion and chromatography on Sephadex (bottom), only these areas will yield large labeled oligonucleotides. For simplicity, the RNA and DNA molecules are drawn as continuous species. In practice, both species are in small fragments, but the principle of the experiments is exactly the same.*

on Sephadex (31). The large oligonucleotides within the hybrids are then identified by complete RNase T_1 digestion after denaturation and fingerprinting of the [^{32}P]RNA component. From the patterns of the oligonucleotides and their map positions, the extent of relatedness of various portions of the genome can be inferred. Table 2 shows the classification of oligonucleotides according to the outcome of the experiment. Digestion of the hybrids with RNase T_1 alone will lead to the recovery of all oligonucleotides with sufficient partial homology to remain base paired through the isolation procedure (Fig. 3A: oligonucleotides A, B, C, D, and G). Only oligonucleotides with very little or no homology (as a result of extensive single base changes, deletions, or substitutions) will be lost. In contrast, treatment of such hybrids with RNase A will cleave internally most oligonucleotides which are not fully hydrogen bonded to the cDNA (oligonucleotides C, D, E, and F). Even if cleaved only once by RNase A, such oligonu-

TABLE 2. *Classification of oligonucleotides by extent of relatedness*[a]

Class	Identical?	Protected against:		Competed?	Examples	
		RNase T_1	RNase T_1 + RNase A		Pr-RSV-B	RAV-0
A	+	+	+	+	3, 8, 10, 16, 22, 23, 25, 32, 37	3, 8, 10, 16, 22, 23, 25, 32, 37
B	−	+	+	+	1, 15, 24	01, 014, 103
C	−	+	−	−	2, 9, 19, 20, 21, 26, 36	02, 03, 06, 012, 013, 308, 111
D	−	+	−	+	11	
E	−	−	−	+		
F	−	−	−	−	5, 6, 14, 27, 33, 35	04, 07, 08
G	+	+	+	−	4	4

[a] In the type of experiment shown in Fig. 3.

cleotides will not be found at their usual positions in the fingerprint. Depending on the specific nonhomologous sequence and the conditions of digestion, some oligonucleotides with small differences (oligonucleotide B) will survive the treatment. By combining information from two experiments, it should be possible to determine which oligonucleotides derive from regions of the genome differing only by single base changes, so long as overall homology is preserved to an extent sufficient to allow hybrids to form in the first place.

A related type of experiment should allow us to determine at least qualitatively the extent of relatedness of the region within which an oligonucleotide is located. This can be accomplished by a hybridization-competition experiment of the sort shown in Fig. 3B. By using the same pair of viruses as an example, unlabeled cDNA from virus a is annealed with unlabeled RNA from virus b. The hybrid formed is then tested by subsequently annealing with ^{32}P-labeled RNA from virus a. The hybrid is again isolated after RNase A plus RNase T_1 digestion, and its complement of large oligonucleotides is determined. Oligonucleotides in closely related regions (A, B, and C) will be competed for by the unlabeled RNA and, therefore, will not be found in the fingerprint. Oligonucleotides in unrelated regions (F) will not be competed for and will appear in the fingerprint, as will oligonucleotides in short regions of substantial homology in otherwise unrelated regions (G), because the labeled RNA will displace the competitor from the hybrid. Conversely, substantially unrelated oligonucleotides in a region of high homology (E) would not leave sufficient open regions for hybrid formation with the labeled RNA, and such oligonucleotides will not appear in the fingerprint.

Characterization of Pr-RSV-B and RAV-0 cDNA probes. We have applied these approaches to the analysis of sequence relationships between the genomes of RAV-0 and Pr-RSV-B. cDNA's were prepared by using the endogenous DNA polymerase activity of the two viruses with the addition of oligodeoxynucleotide primers (43). To characterize these DNAs and establish conditions for fingerprinting experiments, saturation hybridization and hybridization-competition curves were obtained by using these DNAs and labeled 70S RNA from the two viruses (Fig. 4). At the apparent saturation point, Pr-RSV-B cDNA protected approximately 90% of the Pr-RSV-B and 80% of the RAV-0 RNA from RNase digestion. The RAV-0 cDNA protected 90% of RAV-0 [^{32}P]RNA but only 40% of the Pr-RSV-B RNA. These results are consistent with those reported by others (23, 29, 39) and suggest that nucleotide sequences homologous to most of the RAV-0 genome are found in the Pr-RSV-B genome, whereas only about 50% of the Pr-RSV-B genome is contained within the RAV-0 genome.

Analysis of the cDNA's and viral RNAs by hybridization-competition is shown in Fig. 4C and D. Two steps were taken to reduce to a minimum the amount of reagents necessary for this experiment. Because preliminary fingerprinting experiments (see below) showed that the viral genomes were uniformly protected from RNase A plus T_1 digestion by homologous cDNA at ratios of DNA:RNA sufficient to yield substantially less than 100% hybridization, the DNA:[^{32}P]RNA ratio in these experiments was adjusted to give approximately 30 to 60% hybridization in the absence of competitor. Additionally, the competition-hybridization was done in two steps: first, the cDNA was preannealed with unlabeled RNA; then, the labeled RNA was added and the mixture was annealed again for the same amount of time. An example of the advantage of the two-step procedure over the more standard method of mixing all reagents

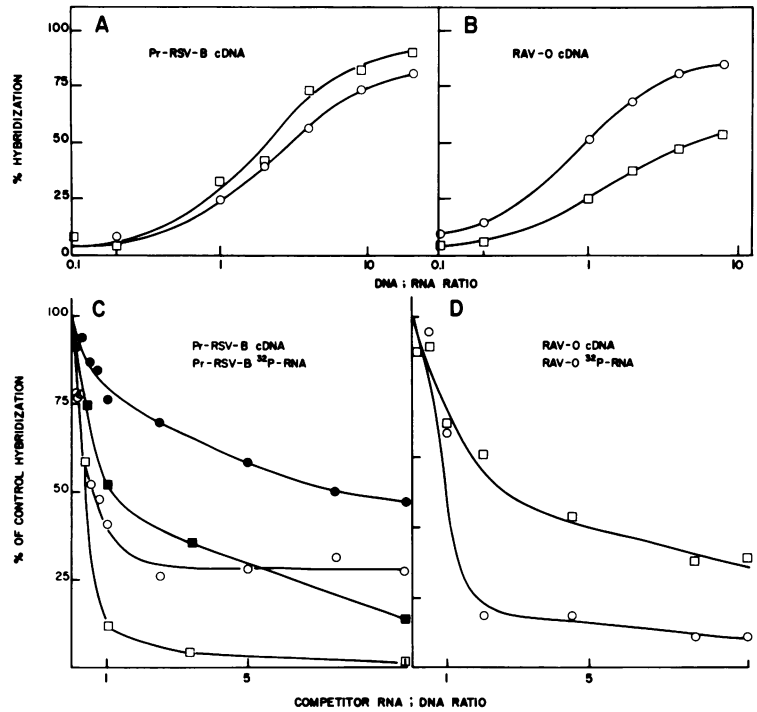

FIG. 4. *Characterization of RAV-0 and Pr-RSV-B cDNA's. (A) Increasing amounts of Pr-RSV-B cDNA were annealed with* 6×10^{-4} *μg of* ^{32}*P-labeled RAV-0 (○; specific activity,* 4×10^{6} *cpm/μg) or* 4×10^{-4} *μg of Pr-RSV-B (□; specific activity,* 7×10^{6} *cpm/μg) 70S RNA. (B) RAV-0 cDNA was annealed with* 6.8×10^{-4} *μg of Pr-RSV-B (□;* 3×10^{6} *cpm/μg) or RAV-0 (○,* 3×10^{6} *cpm/μg) [*^{32}*P]RNA. (C) (Open symbols) Pr-RSV-B cDNA (*1.4×10^{-3} *μg) was annealed with varying amounts of unlabeled Pr-RSV-B (□) or RAV-0 (○) 70S RNA. In a second step,* 1.4×10^{-3} *μg of Pr-RSV-B 70S [*^{32}*P]RNA (3,200 cpm) was added and the mixture was annealed again. (Closed symbols) The same DNAs and RNAs were annealed together in one step. The control annealing with no added competitor yielded 33% of the input radioactivity as hybrid. (D) A two-step annealing, similar to (C), was performed by using* 8.6×10^{-4} *μg of RAV-0 cDNA and* 8.5×10^{-3} *μg of RAV-0 70S[*^{32}*P]RNA (720 cpm). The control annealing yielded 63% of input radioactivity as hybrid. All annealings were for 18 to 24 h in 10 μl of NT buffer. The proportion of labeled RNA in hybrid was determined by RNase treatment and acid precipitation as described in the text.*

together initially is shown in Fig. 4C. In the experiment shown, competition of hybridization of Pr-RSV-B [^{32}P]RNA to Pr-RSV-B cDNA was essentially complete at a ratio of RAV-0 to Pr-RSV-B RNA of 2.5, whereas a ratio of more than 10 was required for a similar extent of competition by using the standard method. The results of these experiments are somewhat different from the direct hybridization analysis shown in Fig. 4A and B, because in both cases the heterologous competitor RNA gave a maximum of about 70% competition. We attribute this apparent difference both to the small number of points taken for this preliminary experiment and to the greater sensitivity of the competition method to differences in the *env* region (see below).

Note that the amounts of RNAs and cDNA's required for these experiments were quite small. For example, an annealing mixture containing 1.4 ng of cDNA, 1.4 ng of Pr-RSV-B [^{32}P]RNA (3,200 cpm), and 3.5 ng of RAV-0 RNA yielded maximal competition of the hybridization of Pr-RSV-B RNA to Pr-RSV-B cDNA.

Fingerprinting of hybrids obtained under various conditions. From the results of the experiments shown in Fig. 4, conditions were established for large-scale hybridization experiments. This was done by selecting the direct hybridization ratios that gave approximately 50% of homologous hybridization, or competition ratios sufficient to ensure that all possible cDNA was hybridized to competitor, and multiplying the amount of each reagent by an amount sufficient to yield enough radioactivity in the hybrid for a visible fingerprint (e.g., 10^{5} cpm for the total genome). Hybrids were isolated by chromatography on Sephadex G-100 essentially as described previously (31) with a few modifications. The most significant modification

was the introduction of a brief heat treatment after RNase digestion to denature aggregated oligonucleotides and thus reduce nonspecific background hybridization significantly. The fingerprints obtained from the hybridization experiments are shown in Fig. 5 and 6 and summarized in Table 2.

Figure 5 shows fingerprints of hybrids obtained with Pr-RSV-B cDNA. From the homologous hybridization (Fig. 5C), it can be seen that even at less than saturating ratios of cDNA to RNA (76% hybridization), all portions of the genome were approximately equally present in the hybrid. Because twofold differences in the molar yields of the oligonucleotides would have been detectable visually, the cDNA must have contained no significant amounts of nucleotide sequence in substantially lower concentration than the average. When RAV-0 [^{32}P]RNA was annealed with the Pr-RSV-B cDNA and the hybrid was isolatd with RNase A plus T_1, only a subset of the RAV-0 oligonucleotides was found in a fingerprint of the hybrid (Fig. 5A). This subset comprised all oligonucleotides common to the two viruses as well as several others (01, 09, 014, 103). These four oligonucleotides must therefore be quite closely related to sequence in the Pr-RSV-B genome. When a similar hybrid was prepared by using RNase T_1 alone (Fig. 5B), all oligonucleotides of the RAV-0 genome were found except for 04, 07, and 08. These results suggest that the Pr-RSV-B genome contains most or all of the RAV-0 genome in a related form.

To assess the extent of relationship in these areas of the genome between oligonucleotides, a

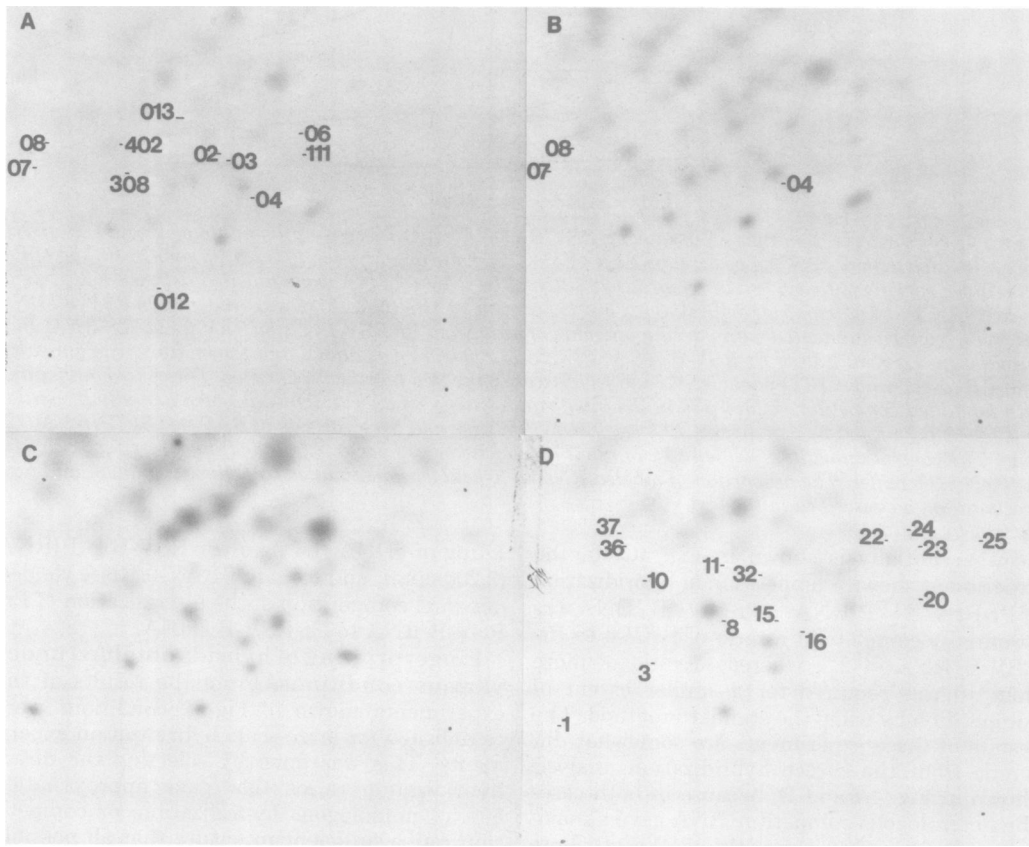

FIG. 5. *Fingerprints of hybrids with Pr-RSV-B cDNA. Pr-RSV-B cDNA (4.7×10^{-1} μg) was annealed with 8.2×10^{-2} μg (4.5×10^5 cpm) of RAV-0 70S [^{32}P]RNA in duplicate. One sample (A) was treated with RNase A and RNase T_1; and the other (B) was treated with RNase T_1 alone. Another preparation of Pr-RSV-B cDNA (9.4×10^{-2} μg) was annealed with 4.7×10^{-2} μg (6×10^5 cpm) of Pr-RSV-B 70S [^{32}P]RNA after (C) no treatment or (D) preannealing with 3.8×10^{-1} μg of RAV-0 70S RNA, and the hybrids were prepared by digestion with RNase A plus T_1. All annealings were as in Fig. 4 except that the volume was 100 μl. Hybrids were isolated by chromatography on Sephadex and fingerprinted as described in the text. Only oligonucleotides missing from the fingerprints are numbered, and their expected locations are shown by a short horizontal line. Identification of any doubtful oligonucleotides was confirmed by complete digestion with RNase A.*

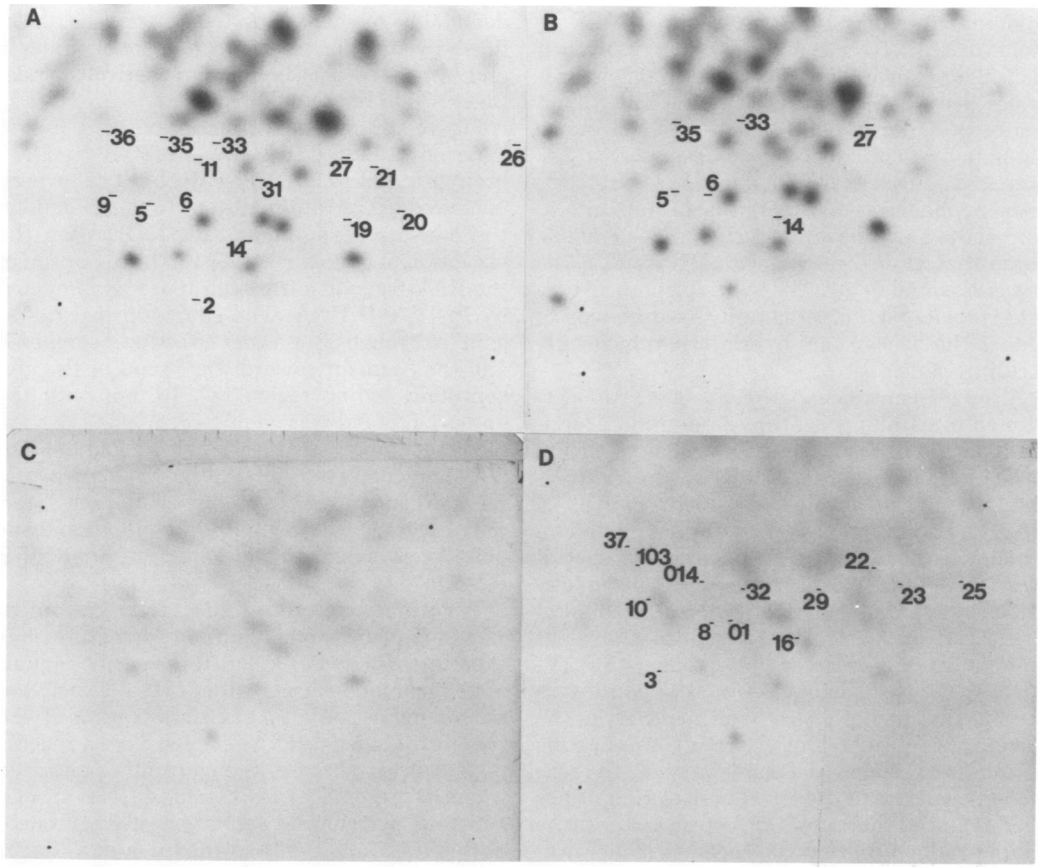

FIG. 6. *Fingerprints of hybrids with RAV-0 cDNA. All conditions were as described in the legend to Fig. 5. (A) RAV-0 cDNA (9×10^{-2} μg) was hybridized to Pr-RSV-B 70S [^{32}P]RNA (4.5×10^{-2} μg; 1.4×10^{6} cpm) and hybrids isolated with RNase A plus T_1. (B) The same annealing mixture, except that hybrid isolation was with RNase T_1 alone. (C) RAV-0 cDNA (1.1×10^{-1} μg) hybridized to RAV-0 RNA (1.1×10^{-1} μg; 5×10^{5} cpm); (D) The same as (C) except the cDNA was preannealed with 8.7×10^{-1} μg of Pr-RSV-B 70S RNA.*

hybridization-competition experiment was performed (Fig. 5D). Pr-RSV-B cDNA was annealed with a fourfold excess of unlabeled RAV-0 RNA and subsequently with Pr-RSV-B [^{32}P]-RNA. From the results of this experiment, the Pr-RSV-B genome can be divided into four segments (see the map in Fig. 2) as follows. (i) With the exception of oligonucleotide 13 (see below), the left-hand half of the Pr-RSV-B genome, through oligonucleotide 32, was absent from the fingerprint and therefore must have had homology to nucleotide sequences in the RAV-0 genome. (Oligonucleotide 29, apparently not competed, is a mixture of two oligonucleotides, one of which is in or near *src*.) (ii) This closely related region is followed by a region competed poorly, if at all, by the RAV-0 RNA. This segment of the Pr-RSV-B genome has been genetically identified as coding for virion subgroup specificity (20). Note that it contains one identical oligonucleotide (oligonucleotide 4) and two (oligonucleotides 9 and 2) which differ from RAV-0 sequences by only a few base changes. (iii) The region defined by oligonucleotides 20 through 10 is not present in the fingerprint and therefore is common or very closely related in these two genomes. This portion of the genome is probably within the region coding for envelope glycoprotein as defined by *env* deletion mutants SRN8 (47) and B-RSV (−) (unpublished data). (iv) The large oligonucleotides nearest the 3′ end (21 through 5) were not detectably competed by the RAV-0 RNA.

In parallel with the experiments shown in Fig. 5, a number of additional control hybridizations were performed (data not shown). To test for nonspecific background, Pr-RSV-B or RAV-0 [^{32}P]RNA was annealed with no added DNA or with an amount of oligodeoxynucleotide primer equivalent to that used to make the cDNA. In all cases, 1 to 5% of the total radioactivity was isolated in the excluded volume of Sephadex

columns after digestion either with RNase T_1 alone or with RNase A plus RNase T_1. When such hybrids were fingerprinted, no oligonucleotides (less than 10% of the amount in hybrids) were found. Thus, none of the oligonucleotides found in fingerprints of hybrids, especially those prepared with RNase T_1 alone, was excluded from Sephadex G-100 in sufficient amounts to contribute a significant background. It should be noted that after digesting with RNase T_1 alone it was essential to heat the hybridization mixture briefly before chromatography (see above) to reduce the background hybridization below detectable values.

A set of control experiments was also performed in parallel with the competition experiment shown in Fig. 5C and D (data not shown). Instead of unlabeled RAV-0 RNA, the Pr-RSV-B cDNA was annealed first with Pr-RSV-B RNA at two different concentrations. When an amount of Pr-RSV-B RNA identical to that of RAV-0 RNA (Fig. 5C) was used, no oligonucleotides were found. However, when an amount of Pr-RSV-B RNA calculated to give an equivalent amount of competition (51%) as seen with RAV-0 RNA was used, oligonucleotides 13 and 1 were present in a higher yield than the others in the fingerprint. We interpret this result as indicating an excess of sequences complementary to these oligonucleotides in the DNA preparation. It has been reported that such DNA species constitute a large majority of Pr-RSV-B cDNA made without primers (4). For this reason, we cannot interpret the failure of oligonucleotide 13 of Pr-RSV-B RNA to be competed as due to nonhomology in nucleotide sequence. As we will show below, oligonucleotide 13 is identical in the two viruses and in a very similar environment.

Figure 6 shows the results of a similar hybridization experiment with RAV-0 cDNA. Again, this cDNA represented, in approximately equal amounts, all portions of the RAV-0 genome detectable as large oligonucleotides (Fig. 6C). The hybrid of Pr-RSV-B [^{32}P]RNA to RAV-0 cDNA prepared with RNase T_1 and RNase A (Fig. 6A) again contained the set of oligonucleotides identical in the two genomes plus oligonucleotide 15 and, in reduced yield, oligonucleotides 1 and 24. In contrast to the results shown in Fig. 5B, a larger set of Pr-RSV-B oligonucleotides was absent from hybrids prepared with RNase T_1 alone (Fig. 6B). As in the case of the hybrid between RAV-0 RNA and Pr-RSV-B DNA, two subgroup-specific oligonucleotides (33 and 35) were not present in the hybrids. In addition to these, however, the five oligonucleotides nearest the 3′ end (14, 27, 5, and the two components of 6) were not present. This result indicates the absence from the RAV-0 genome of sequences detectably related either to the *src* gene as defined by deletion mutants (oligonucleotides 5 and 6; reference 7) or to the genetically undefined region between *src* and *env*.

Figure 6D shows a fingerprint of a hybrid between RAV-0 [^{32}P]RNA and RAV-0 cDNA prehybridized to Pr-RSV-B RNA. This experiment yielded results analogous to those found in the converse experiment (Fig. 5D). With the exception of oligonucleotide 13, the 5′ portion of the RAV-0 genome (through 103) was competed by Pr-RSV-B RNA. The subgroup-specific region (4 through 02) was not effectively competed and the segment toward the 3′ end of the glycoprotein coding region (37, 10, but with the exception of 09) was competed. A small portion of the genome near the 3′ end (03 and 08) was also not competed. Thus, the Pr-RSV-B and RAV-0 genomes seem to have different 3′-terminal regions; the region between *env* and the poly(A) is substantially longer in Pr-RSV-B RNA.

Terminal regions of the virus genomes. The sort of methodology used in the previous experiments can be extended to specific regions of the genome by using either cDNA selected to be complementary to a specific region of the genome or labeled RNA selected from a specific region. In either case, a fingerprint is obtained which represents a less complex RNA species and thus permits the resolution of small oligonucleotides found only in mixtures in a fingerprint of the total genome. Examination of these small, pure oligonucleotides allows a detailed characterization of specific portions of the virus genome.

We have applied this approach to a comparison of nucleotide sequences near the 5′ and 3′ termini of the Pr-RSV-B and RAV-0 genomes. As a specific probe for 5′-terminal sequences strong-stop DNA from Pr-RSV-B was used. This DNA species is a copy of the 5′-terminal 101 nucleotides of the viral genome (10, 18, 38). We have previously characterized five Pr-RSV-B oligonucleotides in this region. These oligonucleotides were precisely located relative to the 5′ end by reference to nucleotide sequence of strong-stop DNA (11). Figure 7 shows the fingerprints of ^{32}P-labeled Pr-RSV-B (Fig. 7A) and RAV-0 (Fig. 7B) RNA in hybrids with Pr-RSV-B strong-stop DNA. The hybrids were prepared with RNase T_1 alone by using the same methodology as before. All oligonucleotides from these fingerprints were identified by their mobility and RNase A digestion products (data not shown). Of the five characteristic oligonucleotides from this region, four (13, a, b, and d) were identical in the two viruses. Oligonucleotide c of Pr-RSV-B was not present in the RAV-0 hybrid

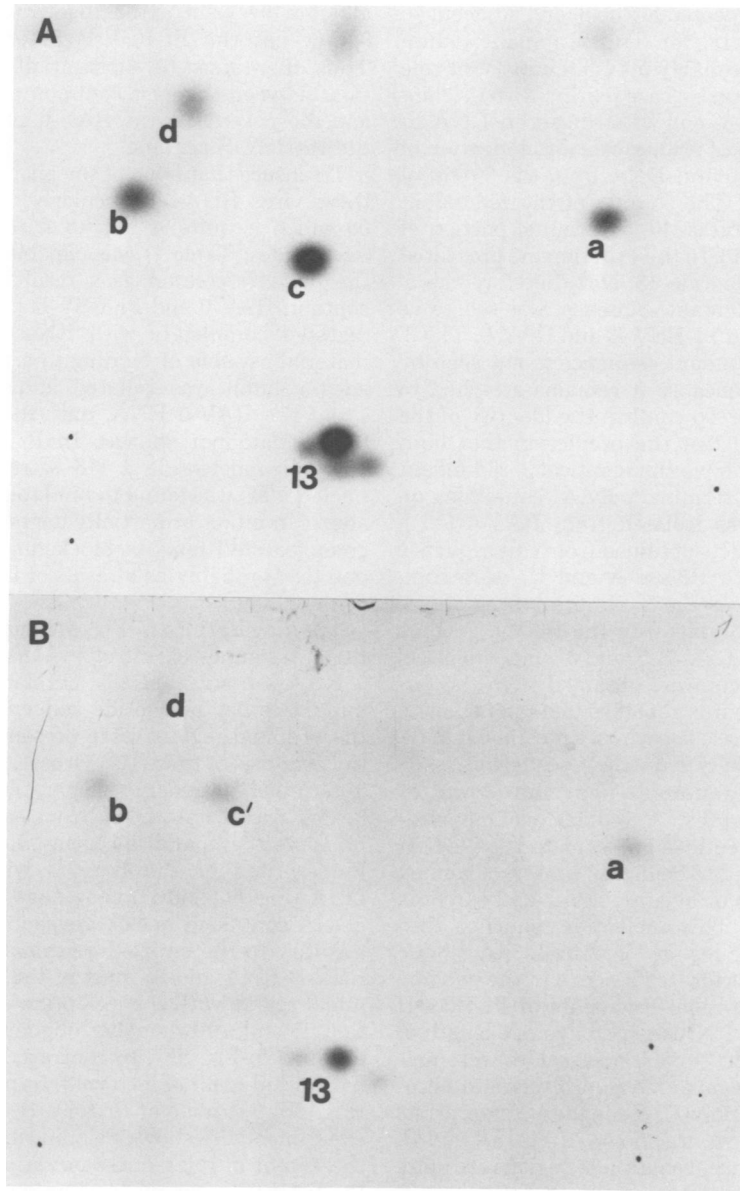

FIG. 7. *Oligonucleotides near the 5' terminus of Pr-RSV-B and RAV-0 RNA. ^{32}P-labeled Pr-RSV-B (A) (0.7 μg; 2.3 × 10^6 cpm) or RAV-0 (B) (0.36 μg; 1.7 × 10^6 cpm) 70S RNA was annealed with 2.2 × 10^{-2} or 1.3 × 10^{-2} μg, respectively, of Pr-RSV-B strong-stop DNA. The hybrids were isolated by chromatography on Sephadex after digestion with RNase T$_1$ alone, and fingerprinted. To retain the smaller oligonucleotides, electrophoresis in the second dimension was until the bromophenol blue dye had migrated 12 cm instead of the usual 18 cm. Numbering of the oligonucleotides is the same as described (11).*

but was replaced by an oligonucleotide (c') of somewhat greater mobility in the second dimension. The RNase A digestion products of oligonucleotide c' were similar to those of c except that an AU present in c at positions 53 to 54 was missing in c' (data not shown). Thus these two oligonucleotides probably differ only by base changes at position 53 and/or 54 from the 5' end. Because oligonucleotides 13, a, b, c, and d together contain 57 of the first 103 nucleotides of the Pr-RSV-B genome, we conclude that this region is very similar in Pr-RSV-B and RAV-0 RNAs.

The 5'-terminal 21 nucleotides of the Pr-RSV-

B genome are repeated at the 3′ end adjacent to the poly(A) (10, 18, 36). This terminally redundant sequence probably plays an important role in RSV replication by base pairing with its complement at the 3′ end of strong-stop DNA to provide a template primer pair for elongation of minus-strand proviral DNA from the 3′ end of the genome (6). The 5′ copy of the redundant sequence constitutes the 5′-terminal portion of oligonucleotide 13. In the experiments presented above, oligonucleotide 13, and therefore the 5′ copy of the redundant sequence, was scored as being identical in Pr-RSV-B and RAV-0. The 3′ copy of the redundant sequence is not seen by this procedure because it remains attached to the poly(A) (10). To confirm the identity of the 5′ sequence and test the prediction that both genomes should have the identical 3′ redundant sequence, the 3′-terminal poly(A)-containing oligonucleotide was isolated from RAV-0 [^{32}P]-RNA. The products obtained from this sequence by digestion with RNase A and U_2 were compared with the products obtained from the 5′ end (oligonucleotide 13) of the RAV-0 genome and those of Pr-RSV-B 5′ and 3′ ends. Identical digestion products were obtained from the termini of the two virus genomes (data not shown), and therefore these two viruses most likely have identical terminally redundant sequences.

Nucleotide sequences near the 3′ end of the viral genomes. A sequence of approximately 900 nucleotides adjacent to the poly(A) has been reported to be highly conserved among different strains of avian leukosis and sarcoma viruses (46, 47). This nucleotide sequence, designated as the "c region," is of unknown genetic function. To test for its presence in the genome of RAV-0, 3′-terminal fragments of Pr-RSV-B and RAV-0 70S RNA with an average length of approximately 300 to 500 nucleotides were prepared by partial alkaline hydrolysis and chromatography on poly(U)-Sephadex. Fingerprints of these fragments are shown in Fig. 8A and B. It is readily apparent that these regions are quite different in nucleotide sequence. Of the approximately 12 large oligonucleotides likely to be unique in an RNA fragment of this size, none of the oligonucleotides of Pr-RSV-B was found in the RAV-0 genome. Similarly, none of the RAV-0 3′ proximal oligonucleotides, with the exception of 10, was found in Pr-RSV-B RNA. Note that this region of the RAV-0 genome contains, in high yield, two oligonucleotides (03 and 08) which are unique in the fingerprint of total 70S RNA. This result is in contrast to that with Pr-RSV-B and all other strains of exogenous avian leukosis virus in which the 3′-terminal 500 or more nucleotides yield no large oligonucleotides. Note also that oligonucleotide 10 is present, albeit in low yield, in the fingerprint of the RAV-0, but not the Pr-RSV-B 3′-terminal region. Thus, there must be substantially less information between this last common oligonucleotide and the poly(A) in the RAV-0 genome than in the Pr-RSV-B genome.

To ensure that none of the oligonucleotides of these virus RNAs, particularly oligonucleotide 08 which is quite A rich (8 A residues per 19 nucleotides; Table 1) was capable of binding to the poly(U) column as a result of its own A content, RAV-0 and Pr-RSV-B [^{32}P]RNA were digested completely with RNase T_1, and the material capable of binding to a poly(U)-Sephadex column was isolated and fingerprinted. From the RAV-0 RNA, only the poly(A) was found (data not shown). In Pr-RSV-B RNA, only oligonucleotide 1 (18 A residues per 27 nucleotides) was found to bind to poly(U)-Sephadex. [In intact or partially degraded RNA, oligonucleotide 1 must be blocked from binding to poly(U), probably as a result of internal hydrogen bonding.]

The fingerprints in Fig. 8A and B show that the 3′-terminal regions of Pr-RSV-B and RAV-0 RNA are substantially unrelated. To determine whether nucleotide sequences related to the terminal regions were present elsewhere in the genomes of these two viruses, slightly longer 3′-terminal fragment preparations were annealed with total cDNA from each virus. Hybrids were prepared by using RNase T_1 alone. Fingerprints of the hybrids with Pr-RSV-B DNA (Fig. 8C and D) show that the Pr-RSV-B cDNA contained nucleotide sequences complementary to the entire 3′-terminal region of Pr-RSV-B RNA and to most of the RAV-0 3′-terminal region with the exception of oligonucleotide 08 and a few smaller oligonucleotides (underlined in Fig. 8F). In contrast, RAV-0 cDNA hybridized to all of its own 3′-terminal sequence (Fig. 8F) but none of Pr-RSV-B (Fig. 8E). This experiment also shows a significant variability in the extent of relatedness of various portions of the RAV-0 3′ proximal sequence to the Pr-RSV-B cDNA, because oligonucleotide 10 was much more prominent relative to 03 in the RAV-0 3′ end hybridized to Pr-RSV-B cDNA (Fig. 8F) than that hybridized to RAV-0 cDNA (Fig. 8D).

From this experiment, we conclude that the RAV-0 genome does not contain nucleotide sequences related to those found near the 3′ end of the Pr-RSV-B genome. The Pr-RSV-B genome does contain nucleotide sequences at least distantly related to most of those near 3′ end of the genome. However, the RAV-0 cDNA did not protect any nucleotide sequences in Pr-RSV-B RNA to the right of oligonucleotide 21. Therefore, the Pr-RSV-B sequence related to the re-

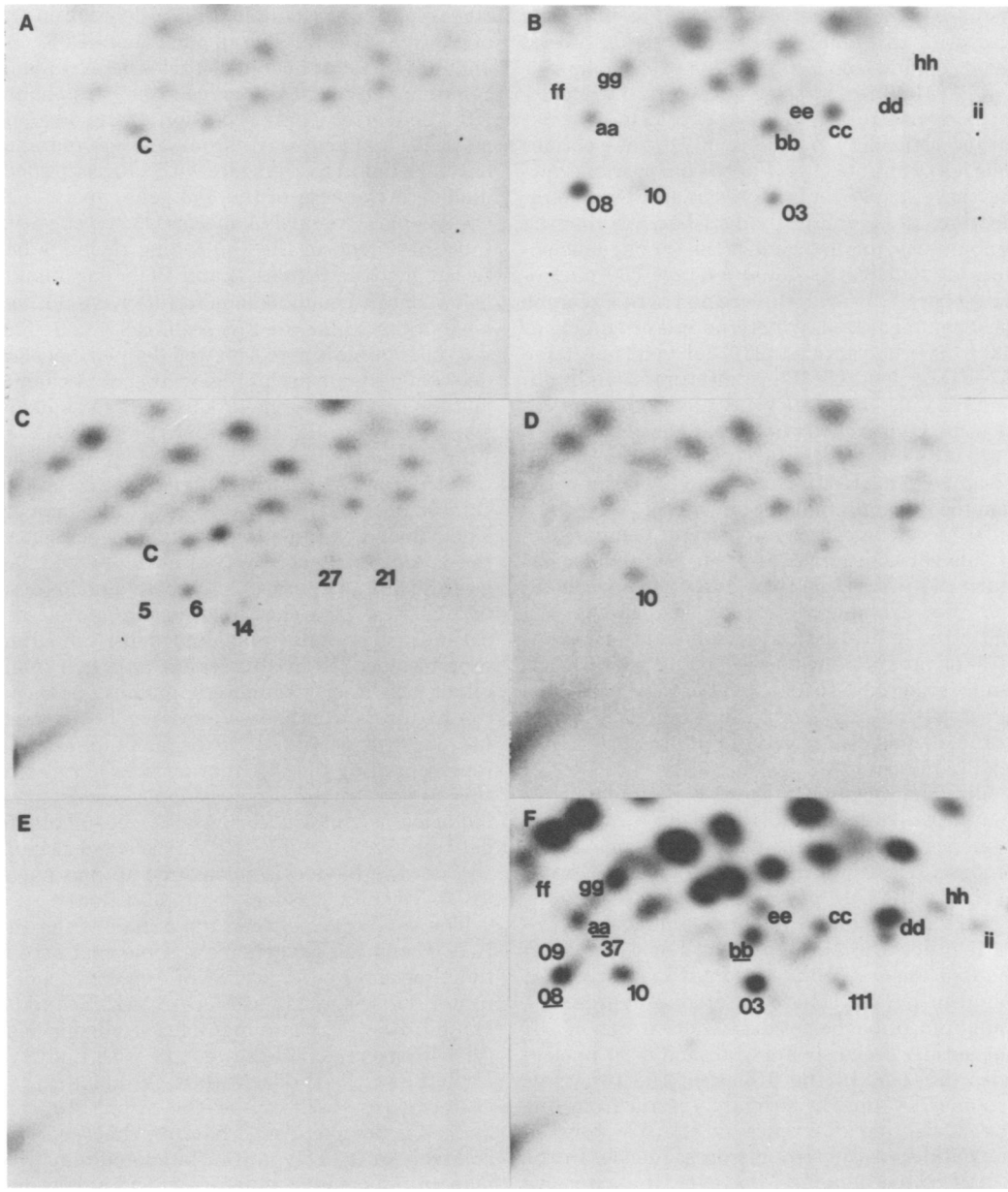

FIG. 8. *3'-Terminal regions of Pr-RSV-B and RAV-0 RNA. Poly(A)-containing fragments (approximately 400 nucleotides in length) were prepared from ^{32}P-labeled 70S RNA of (A) Pr-RSV-B or (B) RAV-0, and fingerprinted directly. In another experiment, somewhat longer 3'-terminal fragments (approximately 30 ng; 3 × 10^5 cpm) were annealed with 90 μg of Pr-RSV-B or RAV-0 cDNA, and the hybrids were isolated by digestion with RNase T$_1$ alone and fingerprinted. Electrophoresis in the second dimension was for the shorter distance as described in the legend to Fig. 7. The identity of numbered oligonucleotides was confirmed by RNase A digestion (data not shown). The hybrids were: (C) Pr-RSV-B RNA to Pr-RSV-B cDNA; (D) RAV-0 RNA to Pr-RSV-B cDNA; (E) Pr-RSV-B RNA to RAV-0 cDNA; and (F) RAV-0 RNA to RAV-0 cDNA. Oligonucleotides found in (F) but not in (D) are underlined.*

gion very near the poly(A) in RAV-0 RNA must be quite distant from the poly(A) in Pr-RSV-B RNA.

DISCUSSION

A relationship map derived from the results

of those experiments is shown in Fig. 9. Before discussing this map, we wish to bring up several general points concerning these experiments.

(i) At the present state of the art, oligonucleotide maps do not yield precise locations, but rather yield only positions of oligonucleotides relative to one another. For the purpose of drawing these maps, we have assumed the oligonucleotides to be uniformly distributed across the genome. By this reasoning, the 30 oligonucleotides of Pr-RSV-B should average 330 nucleotides apart. However, the c region of this genome has been measured by heteroduplex mapping at 800 to 900 nucleotides (22), but yields no large T_1 oligonucleotides. Thus, interpretation of oligonucleotide maps to yield physical distances must be approached with great caution.

(ii) The large oligonucleotides represent only a sample constituting about 6% of the virus genome (2), and small but biologically important differences could escape notice in comparisons of this sort. Nevertheless, this methodology is much more sensitive than other methods available. For example, heteroduplex mapping (22) and cross-hybridization experiments (41) show little or no difference between *env* genes of different subgroup specificity. Also, whenever, a selected probe is available, for example, near the 5′ and 3′ ends, the sensitivity of our approach is readily increased.

(iii) The results of these experiments give qualitative estimates of the relative extents of sequence divergence of various portions of the genome, but are difficult to interpret quantitatively. In the case of closely related sequences, a rough guess can be made, assuming the oligonucleotides are representative. For example, of the 15 5′ oligonucleotides of Pr-RSV-B, 5 are not found in RAV-0. These differences represent a minimum of 5 nucleotides different out of 267 nucleotides, corresponding to about 90 nucleotides different in the 5′ half of the two virus genomes. A similar estimate results from one oligonucleotide difference in the 57 observed nucleotides of the strong-stop sequence. In the 3′ half of the genomes, however, the extensive variation in degree of relatedness between different regions makes such an estimate very difficult.

(iv) The boundaries of specific coding regions in the virus genome are not known, although genetic experiments have allowed the coding regions to be ordered along the genome and have provided a correlation between certain oligonucleotides and gene products. Figure 9 gives these correlations according to the best available evidence, as obtained from analysis of deletion mutants and recombinants. The apparent boundary points obtained by these methods do not always correspond. For example, the *td* deletion in *src* of Pr-RSV-B is missing oligonucleotides 5, 6A, and 6B (1, 7), but oligonucleotides 5, 6A 6B, and 14 are linked to *src* by analysis of recombinants (20). Conversely, only a portion of the oligonucleotides mapped in *env* from deletion mutants has been found to segregate with subgroup specificity. In the case of the *env* gene, analysis of the 28S putative mRNA species (19, 49) suggests that the 5′ end of the *env*-coding region is between oligonucleotides 32 and 4 (26; our unpublished data). Unambiguous results have not yet been obtained for the 21S *src* mRNA (26). The *gag-pol* border is even less well defined, because deletion mutants in this region are not available, and the combined gene products are most likely translated from the same mRNA (30).

From the relationship map (Fig. 9), it is clear that there are substantial differences in relationship of different regions of the viral genomes. The 5′ half, containing the redundant sequence, the complement of strong-stop DNA, and the *gag* and *pol* genes, is very closely related in these two viruses. This close relationship is not surprising in view of the functional identity of virion proteins encoded by this region, although small differences in electrophoretic mobility between the *gag* proteins of these viruses can be detected (33, 37). It is possible that the small differences represented by the different oligonucleotides in this region of the virus genomes have some biological significance, because recombinants between RAV-0 and Pr-RSV-B selected as env^E src^+ always have oligonucleotide 15 and never 01 (P. Tsichlis, personal communication).

The similarity between the 5′ half of the Pr-RSV-B and RAV-0 genome is somewhat exceptional because other strains of exogenous avian tumor viruses, such as RAV-6 and Schmidt-Ruppin RSV-D, share many fewer oligonucleotides in this region of the genome with Pr-RSV-B than Pr-RSV-B does with RAV-0 (3, 47).

The close similarity of the strong-stop sequence is not surprising, because this sequence is largely or entirely untranslated and may provide an important regulatory function as in polymerase or ribosome binding. It is also probably of importance in mRNA processing, because it appears to be "spliced" onto mature, subgenomic mRNA's (26, 49). The presence of the identical terminally redundant sequence in these two viruses is particularly interesting, because this sequence is almost certainly involved in allowing the elongation of nascent DNA chains beyond the 5′ end of the genome (9, 10; W. Haseltine, J. Coffin, and T. Hageman, submitted for publication).

The *env* gene extends rightward from approximately the middle of the genome. Analysis of

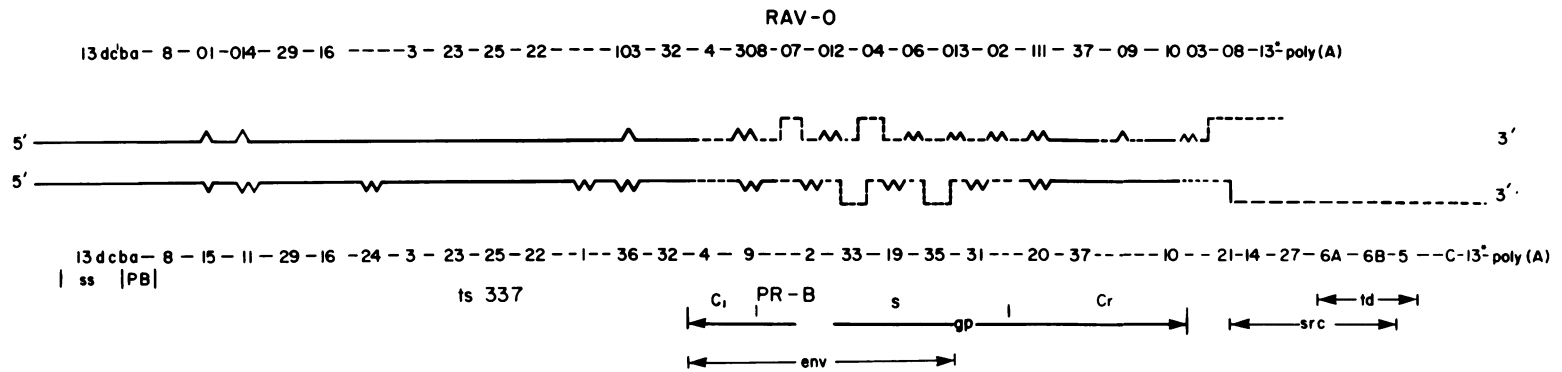

FIG. 9. *Relationship map of the RAV-0 and Pr-RSV-B genomes. The RAV-0 map is shown in the top row of numbers and Pr-RSV-B in the bottom. The large oligonucleotides are drawn as if they were evenly distributed over the genome. The lines between the numbers show the relationship of various oligonucleotides between the two viruses. Symbols: ——, identical oligonucleotide; _/_, different oligonucleotide, found in hybrids prepared with both RNase A and T_1; _/\/_, different oligonucleotide, found only in hybrids prepared with RNase T_1 alone; _┌¬, oligonucleotide not found in any of the hybrids. Solid lines show oligonucleotides not found after competition and therefore present in regions of close homology and dashed lines show oligonucleotides remaining in the fingerprint after a hybridization competition experiment and therefore in regions of little or no homology. Physical and genetic regions are shown at the bottom. SS is the complement of the 101-nucleotide strong-stop DNA sequence (10, 18, 38); PB is the $tRNA^{trp}$ primer binding site (42); ts 337 is a temperature-sensitive mutant defective in polymerase (21, 24); env is the location of subgroup specificity as defined by mapping recombinants (20, 21); gp is the region deleted in mutants lacking envelope glycoprotens (47); C_l, S, and C_r are the three subsections of this region as defined in the text; td is the region deleted in transformation-defective mutants (7, 46); src is the location of sequences necessary for cell transformation as determined by mapping of recombinants (20); C is the oligonucleotide characteristic of the common region (46); and 13* is the 3' copy of the redundant sequence (10, 36).*

recombinants of altered subgroups suggests that subgroup specificity is determined in the relatively short region marked *env* in Fig. 9. Analysis of deletion mutants (47) and glycosylation sites on the gp85 protein (15), however, suggests that the envelope glycoprotein-coding region is much larger, probably extending through oligonucleotide 10. On the basis of these results, the region coding for envelope glycoprotein can be divided into three subregions (labeled C_l, S, and C_r in Fig. 9). The left-hand common (C_l) region contains two oligonucleotides highly conserved among various strains of avian oncoviruses. Oligonucleotide 4 has been found in all strains of subgroups A through E examined to date (at least 12 different isolates; 7, 1, 20, 2, 21, 47, and 46). Oligonucleotide 308 has been found in viruses of subgroups A, C, D, and E, but in subgroup B viruses it is altered to either 9 or another closely related oligonucleotide. This closely related region must be quite short, because homologous hybridization of oligonucleotides 4 and 308 (or 9) is not effectively competed by the heterologous RNA. The subgroup coding (S) region invariably segregates with subgroup specificity in recombinants (20, 21; J. Coffin, M. Champion, and E. Hunter, unpublished data). Of the five RAV-0 or Pr-RSV-B oligonucleotides in this region, three are related to sequences in the other virus, as judged by their presence in hybrids prepared with RNase T_1 alone. In each case two oligonucleotides could not be demonstrated to contain related sequences. This may be a consequence of substitution of unrelated nucleotide sequences, but we consider it more likely to be a result of extensive single base differences. These unrelated segments do not appear to map exactly in the same place, but the apparent difference in location is smaller than the accuracy of the method, and it is possible that oligonucleotides 07 and 04 in RAV-0 and 33 and 35 in Pr-RSV-B represent contiguous opposing regions of nonhomology in these genomes.

The right-hand constant (C_r) region contains two identical oligonucleotides (37 and 10) and one pair of closely related oligonucleotides (111 and 20). Oligonucleotide 111 is found in all strains of avian oncoviruses (subgroups A to E) except Pr-RSV-B and Schmidt-Ruppin RSV-D. This region is long enough and sufficiently closely related that its hybridization is effectively competed by heterologous RNA. It may be the region encoding the C-terminal region of gp85 or the smaller viral glycoprotein (14) or both.

The 3′ regions of these viruses are substantially different following the last common oligonucleotide (10). In the RAV/0 genome this region contains the two large oligonucleotides 03 and 08. Because 08 is very close to the poly(A) sequence (probably within 50 nucleotides; unpublished data) and not far from 10 (Fig. 8), the amount of coding by this region is probably quite limited, and it is possible that no additional product is encoded by the RAV-0 genome. Its map would then be 5′ *gag-pol-env*^{E} 3′. With the exception of a short region including 08, a region partially homologous to 3′ proximal sequences of RAV-0 exists in the Pr-RSV-B genome. We have placed this region between oligonucleotide 10 and 21 in the Pr-RSV-B genome because no homology could be detected between RAV-0 cDNA and any portion of the Pr-RSV-B genome to the right of oligonucleotide 21 (Fig. 8E).

The nucleotide sequence of the Pr-RSV-B genome between oligonucleotide 21 and the 3′ copy of the redundant sequence apparently has no homology at all in the RAV-0 genome. The function of the left-hand part of this region is unknown. It may be a continuation of the C_r portion of the *env* region, or it may be part of the *src* gene, or it may have a role in processing of the *src* mRNA. In the former case, there should be a substantial difference in the structure of the RAV-0 and Pr-RSV-B glycoproteins. In the latter cases, it should be genetically linked to *src*. Preliminary results suggest that this part of the genome is linked to *src* in recombinants between Pr-RSV-B and RAV-0 (P. Tsichlis, personal communication).

The absence from the RAV-0 genome of any nucleotide sequence related to the c region is particularly striking, because this sequence has been reported to be quite highly conserved among exogenous avian oncoviruses (46, 47). Some preliminary reports have suggested that this region may have a gene product (32), but there is no genetic evidence associating any viral function with the c region, and it is not excluded that it forms part of the *src* gene. Because this region of the viral genome has been found to be duplicated at the left side of the integrated DNA provirus, it has also been suggested that it may be important in the regulation of viral RNA synthesis (J. Taylor and H. Varmus, personal communications). Such a hypothesis would be consistent with the lower rate of virus production in RAV-0 as compared with RSV-infected cells (25, 34). The absence in the RAV-0 genome of any sequence detectably related to the c oligonucleotide is particularly interesting, because this oligonucleotide has been found in all exogenous avian leukosis virus stains (1, 47; our unpublished data). Recent evidence from DNA sequencing has shown this oligonucleotide to be 21 to 32 nucleotides from the poly(A) and immediately adjacent to the redundant sequence

(D. Schwartz and W. Gilbert, personal communication). Therefore, the nonidentity in the 3′ ends of these two viral genomes begins at or slightly after the 22nd nucleotide from the poly(A).

These results are consistent with and extend those of Neiman et al. (29), who found that the only differences in the Pr-RSV-B (or Pr-RSV-C) and RAV-0 genomes detectable by stringent hybridization of labeled RNA to infected or uninfected cell DNA resided in the 3′ proximal portions of the two genomes, although their results suggested a somewhat longer region of nonhomology in RAV-0 than we find. This discrepancy between our results and those of Neiman et al. (29) could reflect either an uneven distribution of oligonucleotides near the 3′ end or that the relationship of the region of the RAV-0 genome containing oligonucleotide 03 is not detected in a more stringent hybridization assay.

In addition to its different subgroup, RAV-0 differs from exogenous viruses in at least two significant properties: it does not cause any disease in animals (28), and it is susceptible to an intracellular restriction of its replication not found in exogenous virus (25, 34). How these properties are related to one another and to the sequence differences in the viral genomes is not yet known. Because the major differences between the genomes of Pr-RSV-B (and other exogenous viruses) and RAV-0 (and other complete endogenous viruses; J. Coffin and H. Robinson, unpublished data) are near the 3′ end, it is tempting to postulate a role of these sequences also in leukemogenesis, or intracellular growth restriction, or both.

Implications for the evolution of tumor viruses. Interpretation of our results in terms of how avian tumor viruses might have evolved is complicated by the high recombination rates of these viruses and the uncertain passage history of most laboratory strains of exogenous virus. It is likely that the original strain of Pr-RSV-B was passaged on cells expressing RAV-0 or other endogenous viral genes and acquired sequences from these endogenous viruses. The possibility of a convergent evolution by this sort of mechanism makes it very difficult to infer the actual events giving rise to the diversity of these viruses. Nevertheless, Pr-RSV-B remains a virus capable of causing tumors in animals, and all the RAV-0 used for these experiments was obtained directly from spontaneously producing (line 100) cells. Thus, the viruses examined here must retain the biologically important sequence differences.

From the results presented here, it is clear that a major portion of each virus must have evolved from a common ancestor or ancestors. We do not know which virus, if either, is closer to these ancestral sequences. It is not improbable, however, that tumor-producing viruses are recently evolved from endogenous "viral" genes along the lines originally suggested by Temin (44). We will therefore consider a hypothetical pathway leading from RAV-0 to exogenous viruses.

In such a pathway, the 5′ portion of the genome, including the terminal untranslated region and the *gag* and *pol* genes, must have been highly conserved. It is uncertain whether the small sequence differences in this part of the genome would have profound biological effects on the virus.

The *env* genes also vary largely or entirely by small changes. In this case, however, the changes are much more extensive and reduce the overall homology to a low level, such that cross-hybridization in this region is not observed under conditions of high stringency. It is highly likely that a number of changes in the *env* region must occur together to lead to an altered subgroup, because a number of oligonucleotides are linked to subgroup specificity, and mutations from one subgroup to another have not been found.

As mentioned above, it is likely that the genome of RAV-0 ends very near the 3′ end of the region coding for envelope glycoprotein. All other strains of avian tumor viruses so far examined, however, contain additional information at the 3′ end. In all cases this information contains the c oligonucleotide, and it therefore is likely that both sarcoma and leukemia viruses may have evolved at least partially by the same pathway. One way this could occur is by addition of a sequence containing c at the 3′ end of RAV-0 followed by the insertion of the *src* sequence between c and *env*. A simple mechanism would be to add the *src* gene and c region simultaneously by recombination of a RAV-0-type virus with endogenous sarc sequences (40), and then have a *td* deletion and other changes give rise to a leukemia virus.

ACKNOWLEDGMENTS

We thank C. Barker for expert technical assistance, W. Haseltine for his gift of strong-stop DNA, and G. Cooper for the gift of line 100 cells.

This project was supported by Public Health Service grants CA-17659 and CA-12924 from the National Cancer Institute and by grant NP246A from the American Cancer Society.

LITERATURE CITED

1. **Beemon, K., P. Duesberg, and P. K. Vogt.** 1974. Evidence for crossing-over between avian tumor viruses based on analysis of viral RNAs. Proc. Natl. Acad. Sci. U.S.A. **71**:4254-4258.
2. **Billeter, M. A., J. T. Parsons, and J. M. Coffin.** 1974. The nucleotide sequence complexity of avian tumor

virus RNA. Proc. Natl. Acad. Sci. U.S.A. **71**:3560-3564.
3. **Brownlee, G. G.** 1972. Determination of sequences in RNA. North Holland Publishing Co., Amsterdam.
4. **Cashion, L. M., R. Joho, M. A. Planitz, M. A. Billeter, and C. Weissman.** 1976. Initiation sites of Rous sarcoma virus RNA-directed DNA synthesis in vitro. Nature (London) **262**:186-190.
5. **Chen, J. H., and H. Hanafusa.** 1974. Detection of a protein of avian leukoviruses in uninfected chick cells by radioimmunoassay. J. Virol. **13**:340-346.
6. **Coffin, J. M.** 1976. Genes responsible for transformation by avian RNA tumor viruses. Cancer Res. **36**:4282-4288.
7. **Coffin, J. M., and M. A. Billeter.** 1976. A physical map of the Rous sarcoma virus genome. J. Mol. Biol. **100**:293-318.
8. **Coffin, J. M., M. A. Champion, and F. Chabot.** 1978. Genome structure of avian RNA tumor viruses: relationships between exogenous and endogenous viruses. p. 191-208. *In* S. Barlati and C. deGiuli-Morghen (ed.), Avian RNA tumor viruses. Piccin Medical Books, Padua.
9. **Coffin, J. M., T. C. Hageman, A. M. Maxam, and W. A. Haseltine.** 1978. Structure of the genome of Moloney murine leukemia virus: a terminally redundant sequence. Cell **13**:761-763.
10. **Coffin, J. M., and W. A. Haseltine.** 1977. Terminal redundancy and the origin of replication of Rous sarcoma virus RNA. Proc. Natl. Acad. Sci. U.S.A. **74**:1908-1912.
11. **Coffin, J. M., and W. A. Haseltine.** 1977. Nucleotide sequence of Rous sarcoma virus RNA at the initiation site of DNA synthesis: the 102nd nucleotide is U. J. Mol. Biol. **117**:805-814.
12. **Crittenden, L. B., F. J. Smith, R. A. Weiss, and P. W. Sarma.** 1974. Host gene control of endogenous avian leukosis virus production. Virology **57**:128-138.
13. **DeWachter, R., and W. Fiers.** 1972. Preparative two-dimensional polyacrylamide gel electrophoresis of ^{32}P-labeled RNA. Anal. Biochem. **49**:184-187.
14. **Diggelmann, H., and R. Klemenz.** 1978. Characterization of the common precursor polypeptide to the two envelope glycoproteins of Rous sarcoma virus. p. 191-208. *In* S. Barlati and C. deGiuli-Morghen (ed.), Avian RNA tumor viruses. Piccin Medical Books, Padua.
15. **Galehouse, D. M., and P. H. Duesberg.** 1978. Glycoproteins of avian tumor virus recombinants: evidence for crossing-over. J. Virol. **25**:86-96.
16. **Halpern, M. S., D. P. Bolognesi, R. R. Friis, and W. S. Mason.** 1975. Expression of the major viral glycoprotein of avian tumor virus in cells of chf^{+} chicken embryos. J. Virol. **15**:1131-1140.
17. **Haseltine, W. A., D. G. Kleid, A. Panet, E. Rothenberg, and D. Baltimore.** 1976. Ordered transcription of RNA tumor virus genomes. J. Mol. Biol. **106**:109-132.
18. **Haseltine, W. A., A. Maxam, and W. Gilbert.** 1977. The Rous sarcoma virus genome is terminally redundant: the 5′ sequence. Proc. Natl. Acad. Sci. U.S.A. **74**:988-993.
19. **Hayward, W. S.** 1977. Size and genetic content of viral RNAs in avian oncovirus-infected cells. J. Virol. **24**:47-63.
20. **Joho, R. H., M. A. Billeter, and C. Weissmann.** 1975. Mapping of biological function on RNA of avian tumor viruses: location of regions required for transformation and determination of host range. Proc. Natl. Acad. Sci. U.S.A. **72**:4772-4776.
21. **Joho, P. H., E. Stoll, R. R. Friis, M. A. Billeter, and C. Weissmann.** 1976. A partial genetic map of Rous sarcoma virus RNA: location of polymerase, envelope, and transformation markers. p. 127-145. *In* D. Baltimore, A. Huang, and C. F. Fox, (ed.), Animal virology. Academic Press Inc., New York.
22. **Junghans, R. P., S. Hu, C. A. Knight, and N. Davidson.** 1978. Heteroduplex analysis of avian RNA tumor viruses. Proc. Natl. Acad. Sci. U.S.A. **74**:447-481.
23. **Kang, C. Y., and H. M. Temin.** 1973. Lack of sequence homology among RNAs of avian leukosis-sarcoma viruses, reticuloendotheliosis viruses, and chicken endogenous RNA-directed DNA polymerase activity. J. Virol. **12**:1314-1324.
24. **Linial, M., and W. S. Mason.** 1973. Characterization of two conditional early mutants of Rous sarcoma virus. Virology **53**:258-273.
25. **Linial, M., and P. E. Neiman.** 1976. Infection of chick cells by subgroup E virus. Virology **73**:508-520.
26. **Mellon, P., and P. H. Duesberg.** 1977. Subgenomic, cellular Rous sarcoma virus RNAs contain oligonucleotides from the 3′ half and the 5′ terminus of virus RNA. Nature (London) **270**:631-634.
27. **Mizutani, S., and H. M. Temin.** 173. Lack of serological relationship among DNA polymerases of avian leukosis-sarcoma viruses, reticuloendotheliosis viruses, and chicken cells. J. Virol. **12**:440-448.
28. **Motta, J. V., L. B. Crittenden, H. G. Purchase, H. A. Stone, W. Okazaki, and R. L. Witter.** 1975. Low oncogenic potential of avian endogenous RNA tumor virus infection or expression. J. Natl. Cancer Inst. **55**:685-689.
29. **Neiman, P. E., S. Das, D. MacDonnel, and C. McMillan-Helsel.** 1977. Organization of shared and unshared sequences in the genomes of chicken endogenous and sarcoma viruses. Cell **11**:321-329.
30. **Oppermann, H., J. M. Bishop, H. E. Varmus, and L. Levintow.** 1977. A joint product of the genes *gag* and *pol* of avian sarcoma virus: a possible precursor of reverse transcriptase. Cell **12**:993-1005.
31. **Parsons, J. T., J. M. Coffin, R. K. Haroz, P. A. Bromley, and C. Weissmann.** 1974. Quantitative determination and location of newly synthesized virus-specific ribonucleic acid in cells infected with Rous sarcoma virus. J. Virol. **11**:761-774.
32. **Purchio, A. F., E. Erikson, and R. L. Erikson.** 1977. Translation of 35S and subgenomic regions of avian sarcoma virus RNA. Proc. Natl. Acad. Sci. U.S.A. **74**:4661-4665.
33. **Rettenmier, C. W., and H. Hanafusa.** 1977. Structural protein markers in the avian oncoviruses. J. Virol. **24**:850-864.
34. **Robinson, H. L.** 1976. Intracellular restriction on the growth of induced subgroup E avian type C viruses in chicken cells. J. Virol. **18**:856-866.
35. **Robinson, H. L., C. A. Swanson, J. F. Hruska, and L. B. Crittenden.** 1976. The production of unique C type viruses by chicken cells grown in bromodeoxyuridine. Virology **69**:63-74.
36. **Schwartz, D. E., P. Zamecnik, and L. Weith.** 1977. Rous sarcoma virus genome is terminally redundant: the 3′ sequence. Proc. Natl. Acad. Sci. U.S.A. **74**:994-998.
37. **Shaikh, R., M. Linial, J. Coffin, and R. Eisenmann.** 1978. Recombinant avian oncoviruses. I. Alterations in the precursor to the internal structural proteins. Virology **87**:326-338.
38. **Shine, J., A. P. Czernilofsky, R. Friedrich, H. M. Goodman, and J. M. Bishop.** 1977. Nucleotide sequence at the 5′ terminus of the avian sarcoma genome. Proc. Natl. Acad. Sci. U.S.A. **74**:1473-1477.
39. **Shoyab, M., and M. A. Baluda.** 1976. Ribonucleotide sequence homology among avian oncornaviruses. J. Virol. **17**:106-113.
40. **Stehelin, D., H. E. Varmus, J. M. Bishop, and P. K. Vogt.** 1976. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. Nature (London) **260**:170-172.
41. **Tal, J., D. J. Fugita, S. Kawai, H. E. Varmus, and J. M. Bishop.** 1977. Purification of DNA complementary

to the *env* gene of avian sarcoma virus analysis of relationships among the *env* genes of avian leukosis-sarcoma viruses. J. Virol. **21**:497-505.

42. **Taylor, J. M., and R. Illmensee.** 1975. Site on the RNA of an avian sarcoma virus at which primer is bound. J. Virol. **16**:553-558.
43. **Taylor, J. M., R. Illmensee, and J. Summers.** 1976. Efficient transcription of RNA into DNA by avian tumor virus polymerase. Biochim. Biophys. Acta **42**:324-330.
44. **Temin, H. M.** 1972. The protovirus hypothesis and cancer. J. Natl. Cancer Inst. **46**:III-VII.
45. **Vogt, P. K., and R. R. Friis.** 1971. An avian leukosis virus related to RSV(0): properties and evidence for helper activity. Virology **34**:223-234.
46. **Wang, L.-H., P. H. Duesberg, K. Beemon, and P. K. Vogt.** 1975. Mapping RNase T_1-resistant oligonucleotides of avian tumor virus RNAs: sarcoma-specific oligonucleotides are near the poly(A) end and oligonucleotides common to sarcoma and transformation-defective viruses are at the poly(A) end. J. Virol. **16**:1051-1070.
47. **Wang, L. H., P. H. Duesberg, S. Kawai, and H. Hanafusa.** 1976. Location of envelope-specific and sarcoma-specific oligonucleotides on RNA of Schmidt-Ruppin Rous sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. **73**:447-451.
48. **Weiss, R. A., R. R. Friis, E. Katz, and P. K. Vogt.** 1971. Induction of avian tumor viruses in normal cells by physical and chemical carcinogens. Virology **46**:920-938.
49. **Weiss, S. R., H. E. Varmus, and J. M. Bishop.** 1977. The size and genetic composition of virus-specific RNAs in the cytoplasm of cells producing avian sarcoma-leukosis virus. Cell **12**:983-992.