Homologies in Nucleotide Sequences of RNA-Phages Qβ and R 17

Erich Köhler and Hartwig Rohloff

Institut für landwirtschaftliche Virusforschung, Biologische Bundesanstalt für Land- und Forstwirtschaft, Braunschweig

(Z. Naturforsch. **29 c**, 433–436 [1974]; received February 6/March 28, 1974)

Sequence Homology, RNA-phages Qβ and R17, Translational Repression

Statistical calculations on sequence homologies provide evidence that the two phages Qβ and R17 have derived from a common precursor, although they are quite different in several properties. Moreover, duplications within the genom of each phage have occurred. These results are discussed with regard to the lack of translational repression between Qβ and MS2 phage group.

Introduction

Homologies within and between the nucleotide sequences of the phages Qβ and R17 have been mentioned in the last few years by several authors. Steitz [21] compared the initial region of the A-protein-, the C-protein-, and the replicase-cistron of R17 and indicated some similarities in the precistronic regions. In a comparison between Qβ and R17 Adams and Cory [1], and Cory *et al.* [6] pointed out that these phages possibly may have derived from a common ancestor, although they are quite dissimilar in several properties, *e. g.* there is no serological relationship [15, 17]. Recently Weissmann *et al.* [25] drew attention again to homologies between and within the RNA-sequences of Qβ and R17 phages. We concerned ourselves with some specific virus relationships that might be important in explaining the problem of cross protection between closely related virus strains [12, 13]. Cross protection means that a host acquiring one strain of a virus is unable to be infected by a second strain of the same virus. This is a well known but unexplained phenomenon in plant virology [13]. Ward *et al.* [23] have proposed a model at the translational level to explain this phenomenon of bacteriophages. The closely related phages MS2, f2, and fr cross protect each other by means of a translational repression mechanism and coat protein serves as repressor. Qβ neither shows cross protection to this phage group nor a repression activity of its coat protein.

Requests for reprints should be sent to Dr. E. Köhler, Institut für landwirtschaftliche Virusforschung der BBA, *D-3300 Braunschweig*, Messeweg 11/12.

Calculations of sequence homologies

Fig. 1 shows the six sequences we were working on in such an arrangement, that the initial codons (AUG) are in the same position. The comparisons are limited to these initial regions, because the replicase cistron of Qβ is not known over a longer region. The nucleotide sequences on the left side of AUG are supposed to code not for an amino acid sequence but to be necessary for attachment to ribosomes.

For quantitative estimation of sequence homologies it is necessary to postulate some assumptions on evolutionary events. In reference to the stochastic models of Holmquist [8] and Kimura and Ohta [11] we generally assume that point mutations (mutant substitutions) occur spatially at random and in uniform probability over the variable part of the sequence, and that at each site a given nucleotide mutates with equal probability to any one of the remaining three.

Calculation of random similarities

By means of binomial distribution the probability $P_0(n_0, k_0)$ can be calculated that two sequences have in k out of n sites the same nucleotide by chance. If the random probability of an equal or better similarity between two sequences is less than 0.001 we exclude the possibility, that the two sequences simply happen to be similar. Table I shows these random probabilities, $P_0(n_0, k_0)$, which are calculated by the formula of the binomial distribution:

$$P = p^k (1-p)^{n-k} \frac{n!}{k!\,(n-k)!}$$

with $p = 1/4$. The comparison between Qβ-replicase and R17-replicase leads to a probability so small

Table I. n_0 = Number of nucleotides compared, from which k_0 are identical at the same site; $P_0(n_0, k_0)$ = probability that k_0 out of n_0 nucleotides are identical by random events; g = number of gaps; r = number of residues in a gap; N_1 = number of nucleotides in the longer chain; N_2 = number of nucleotides in the shorter chain; f_g/f_0 = ratio of possible comparisons of the two chains after and before introduction of the gap (*c.f.* ref. Cantor); n_g = number of nucleotides compared after the introduction of the gap, from which k_g are identical at the same site; $P_g(n_g, k_g)$ = probability that k_g out of n_g nucleotides are identical by random events; value of C should be smaller than 1 to justify a gap.

Compared Sequences	n_0	k_0	$P_0(n_0, k_0)$	g	r	N_1	N_2	f_g/f_0	n_g	k_g	$P_g(n_g, k_g)$	$C = \frac{P_g \cdot f_g}{P_0 \cdot f_0}$
Qβ-A/R17-A	37	10	1.41×10^{-1}									
Qβ-C/R17-C	38	12	9.11×10^{-2}									
Qβ-r/R17-r	32	18	1.22×10^{-4}	1	2	36	32	98/5	30	21	2.44×10^{-7}	0.04
Qβ-A/Qβ-C	38	17	3.98×10^{-3}									
Qβ-A/Qβ-r	32	12	4.27×10^{-2}	1	2	45	32	422/14	30	16	6.03×10^{-4}	0.43
Qβ-C/Qβ-r	37	10	1.41×10^{-1}									
R17-A/R17-C	37	15	1.56×10^{-2}	1	3	43	37	151/7	34	18	3.21×10^{-4}	0.45
R17-A/R17-r	34	9	1.51×10^{-1}									
R17-C/R17-r	36	15	1.23×10^{-2}									

that we can suppose an ancestral nucleotide sequence from which these two sequences have derived.

Introduction of gaps in the sequence of nucleotides

Any two sequences could be arranged to obtain maximum homology by inserting gaps. Each gap, however, increases the number of possible alignments, and therefore the unreliability of homology as quantitatively explained by Cantor [5]. To introduce gaps in one or both of the sequences at various points in order to extend the apparent homology a procedure has been developed by Cantor to test whether a gap does in fact confirm the homology. The extra homology introduced by a gap must more than compensate for the increased number of comparisons. The equations to calculate the number of possible comparisons of two sequences of the length N_1 and N_2 with gaps of r residues are given by Cantor [5]. Table I shows the ratio of possible comparisons, f_g/f_0, after and before the inserting of a gap. The last column of Table I shows the value of C, which should be smaller than 1 in order to justify a gap. In three cases we succeeded in introducing a gap. Fig. 2 shows the alignment of those sequences which are involved in these comparisons. The decreased random probabilities, $P_g(n_g, k_g)$, in Table I indicate homology also within the genoms of Qβ and R17. We can suppose, that the one or other of these cistrons, or parts of them, have derived from duplications in the course of evolution.

Estimation of mutational distance between sequences

Kimura and Ohta [11] have established a stochastic model of evolution for the estimation of mutational distances between homologous sequences. In a good approximation the average number of nucleotide substitutions (inclusive back mutations and multiple substitutions) per site were estimated by using the formula $K = -3/4 \ln(1 - 4/3\, d)$, where d is the fraction of sites at which two holomogous sequences differ. This value has a variance of $s^2 = d(1-d)/L(1-4/3\, d)^2$, where L is the number of nucleotides in the compared sequences. The values of K and their variances for the three homologous sequences are given in Table II. The value of K gives a quantitative estimation of the evolutionary distance between two sequences and is suitable for quantitative comparisons between different homologies. In the discussion the possible conclusions of different values of K will be deliberated.

Discussion

Quantitative comparisons of nucleotide sequences give the most sensitive evidence of relationships between two organisms. It is possible to suppose distinct mutational events, such as duplications, insertions or deletions, and nucleotide substitutions, that have occurred in the evolution between two structures. Fig. 2 gives an illustration of these supposed events.

R17- and MS2-phages show small differences; their C-protein-cistrons *e. g.* show a difference of 8 substitutions out of 238 nucleotides [16]; this results in a value of $K = 0.0307 \pm 0.0001$. None of these base changes leads to an amino acid difference of coat protein [16, 25]. On the other hand R17 and Qβ, although serology fails to detect any convincing

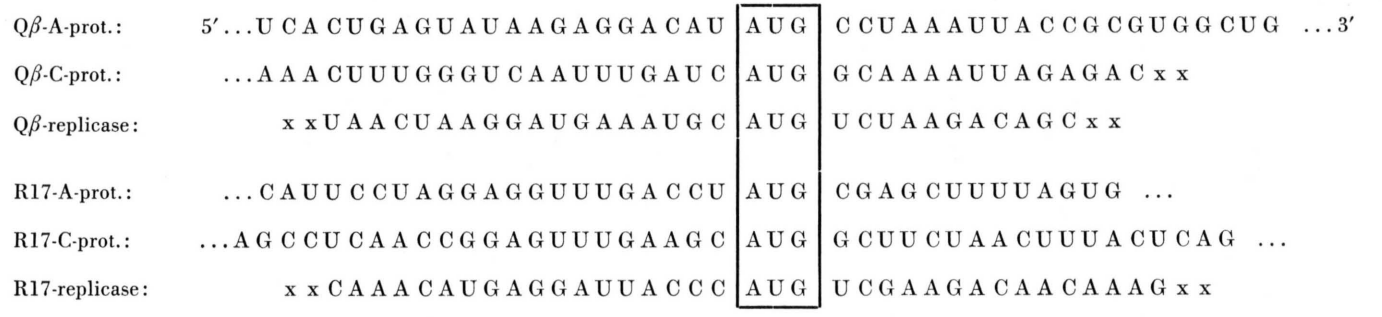

Fig. 1. Nucleotide sequences of the initial regions of the three cistrons of the RNA phages Qβ and R17: the A-protein-, the C-protein-, and the replicase-cistron. They are arranged so that the initiator codons (AUG) are at the same position. Sources of information of Qβ: [2, 3, 7, 9, 10, 14, 18, 21]; sources of information of R17: [4, 14, 19, 20, 22, 24]. Some of the sequences are known over a longer region, this is indicated by a dotted line; unknown bases are indicated by x.

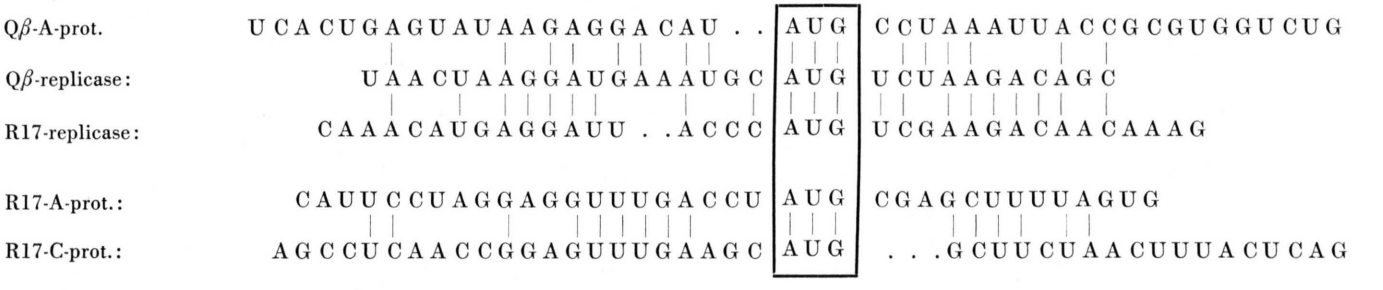

Fig. 2. Alignment of those RNA sequences which show sufficient homology at least in one comparison after introduction of the gaps.

Table II. d = Fraction of sites by which two homologous sequences differ; K = average number of nucleotide substitutions per site; s^2 = variance of K.

Compared Sequences	d	K	s^2
Qβ-r/R17-r	0.30	0.383	±0.019
Qβ-A/Qβ-r	0.467	0.730	±0.058
R17-A/R17-C	0.471	0.741	±0.053

relationship, show at least in the short initial region of replicase cistron a clear homology.

It is necessary to note, that the differences between R17 and Qβ are not equally distributed in the three compared regions: Homology could be made evident only between the initial region of the replicase cistron, but not in the other two compared regions. This would be a contradiction of our general assumption previously mentioned, if we neglect the process of selection. If a function, *e. g.* of replicase enzyme, is closely bound to a special structure, any mutation that would disturb the unity of function will not pass selection processes, while mutations in the nucleotide sequence of an other cistron, *e. g.* of coat protein, would have been established in the evolution. This consideration also resolves a discrepancy: A smaller value of K for two compared sequences than for two others could lead to the assumption that the first duplication would have occurred later in the timespan of evolution than the other; this conclusion is only possible, if we postulate the same pressure of selection on all genetic structures we involved in our considerations. Therefore we cannot deduce a temporal arrangement of duplication events.

In our considerations on the possibilities of evolution we ought not to overlook the possibility that replicase enzyme of different and unrelated phages might be similar resulting from convergent evolution. And it is a problem of considerable interest, in how far the phenomenon of convergent evolution may be expected in genetic structures that code for such an essential function as replication.

Because the sequence comparisons refer only to the initial region of the cistrons, only slight conclusions are possible to the model of translational repression of Ward *et al.*[23]. These authors showed, that coat protein of MS2-, fr-, and f2-phages serve as a type of regulator or repressor of virusspecific replicase synthesis. It is of particular interest that the serologically unrelated virus Qβ showed no cross reaction; its coat showed no repressor activity for the RNA of MS2-phage group. R17 is almost identical to MS2, so that we can extend these results also to R17. We have some evidence for a relationship between Qβ and R17, but there is no homology in the initial region of coat protein cistron. This is in good agreement with the lack of translational repression between Qβ and the MS2 phage group.

[1] J. M. Adams and S. Cory, Nature **227**, 570–574 [1970].
[2] J. M. Adams, S. Cory, and P. F. Spahr, Eur. J. Biochem. **29**, 469–479 [1972].
[3] J. M. Adams, P. F. Spahr, and S. Cory, Biochem. **11**, 976–980 [1972].
[4] M. A. Billeter, J. E. Dahlberg, H. M. Goodman, J. Hindley, and C. Weissmann, Nature **224**, 1083–1086 [1969].
[5] C. R. Cantor, Biochem. Biophys. Res. Commun. **31**, 410–416 [1968].
[6] S. Cory, P. F. Spahr, and J. M. Adams, Cold Spring Harbor Symp. Quant. Biol. **35**, 1–12 [1970].
[7] S. Cory, J. M. Adams, P. F. Spahr, and U. F. E. Rensing, J. Mol. Biol. **68**, 41–56 [1972].
[8] R. Holmquist, J. molec. Evolution **1**, 211–222 [1972].
[9] P. G. N. Jeppesen, Biochem. J. **124**, 357–366 [1971].
[10] P. G. N. Jeppesen, J. A. Steitz, R. F. Gesteland, and P. F. Spahr, Nature **226**, 230–237 [1970].
[11] M. Kimura and T. Ohta, J. molec. Evolution **2**, 87–90 [1972].
[12] E. Köhler, Phytopath. Z. **7**, 1–30 [1934].
[13] E. Köhler, Phytopath. Z. **69**, 175–197 [1969].
[14] M. Kozak and D. Nathans, Bacteriol. Rev. **36**, 109–134 [1972].
[15] R. C. Krueger, J. Virology **4**, 567–573 [1969].
[16] W. Min Jou, G. Haegeman, M. Ysebaert, and W. Fiers, Nature **237**, 82–86 [1972].
[17] L. R. Overby, G. H. Barlow, R. H. Doi, M. Jacob, and S. Spiegelman, J. Bacteriol. **91**, 442–448 [1966].
[18] U. F. E. Rensing and J. G. G. Schoenmakers, Eur. J. Biochem. **33**, 8–18 [1973].
[19] H. D. Robertson and P. G. N. Jeppesen, J. Mol. Biol. **68**, 417–428 [1972].
[20] D. H. Staples, J. Hindley, M. A. Billeter, and C. Weissmann, Nature New Biol. **234**, 202–204 [1971].
[21] J. A. Steitz, Nature **224**, 957–964 [1969].
[22] J. A. Steitz, Nature New Biol. **236**, 71–75 [1972].
[23] R. Ward, M. Strand, and R. C. Valentine, Biochem. Biophys. Res. Commun. **30**, 310–317 [1968].
[24] A. M. Weiner and K. Weber, Nature New Biol. **234**, 206–209 [1971].
[25] C. Weissmann, M. A. Billeter, H. M. Goodman, J. Hindley, and H. Weber, Ann. Rev. Biochem. **42**, 303–328 [1973].