J Mol Evol (1992) 34:141–152

Journal of Molecular Evolution

Nucleotide Sequence of the Genome of the Filamentous Bacteriophage I2-2: Module Evolution of the Filamentous Phage Genome

Alphons P.M. Stassen,[1] Eric F.P.M. Schoenmakers,[1] Maoxiao Yu,[2] John G.G. Schoenmakers,[1] and Ruud N.H. Konings[1,3]

[1] Department of Molecular Biology, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands
[2] Institute of Microbiology, Academia Sinica, Beijing, Peoples Republic of China
[3] Department of Biophysical Chemistry, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands

Summary. The nucleotide sequence of the circular single-stranded genome of the filamentous *Escherichia coli* phage I2-2 has been determined and compared with those of the filamentous *E. coli* phages Ff (M13, f1, or fd) and IKe. The I2-2 DNA sequence comprises 6744 nucleotides; 139 nucleotides less than that of the N- and I2-plasmid-specific phage IKe, and 337 (336) nucleotides more than that of the F-plasmid-specific phage Ff. Nucleotide sequence comparisons have indicated that I2-2, IKe, and Ff have a similar genetic organization, and that the genomes of I2-2 and IKe are evolutionarily more closely related than those of I2-2 and Ff. The studies have further demonstrated that the I2-2 genome is a composite replicon, composed of only two-thirds of the ancestral genome of IKe. Only a contiguous I2-2 DNA sequence of 4615 nucleotides encompassing not only the coat protein and phage assembly genes, but also the signal required for efficient phage morphogenesis, was found to be significantly homologous to sequences in the genomes of IKe and Ff. No homology was observed between the consecutive DNA sequence that contains the origins for viral and complementary strand replication and the replication genes. Although other explanations cannot be ruled out, our data strongly suggest that the ancestor filamentous phage genome of phages I2-2 and IKe has exchanged its replication module during evolution with that of another replicon, e.g., a plasmid that also replicates via the so-called rolling circle mechanism.

Offprint requests to: R.N.H. Konings

Key words: Filamentous bacteriophage — Module evolution — Single-stranded genome — Nucleotide sequence comparison

Introduction

A major characteristic of filamentous bacteriophages is that infection of their host cell occurs via attachment of a specific phage end, i.e., the one at which the coat proteins encoded by genes III and VI are located, to the tip of a conjugative pilus present at the surface of the host cell (Caro and Schnös 1966). This pilus is composed of filamentous polymers of identical protein subunits (pilin), generally encoded by conjugative plasmids (Marvin and Folkhard 1986). On the basis of their plasmid (pilus) specificity, filamentous phages have been subdivided into several groups that differ both in phage morphology and protein composition. The best-studied filamentous phages are the IncF-plasmid-specific phage Ff [M13, f1, and fd; for a recent review see (Model and Russel 1988)] and the IncI2- and IncN-plasmid-specific phage IKe (Khatoon et al. 1972; Bradley et al. 1983; Peeters et al. 1985). They both have *Escherichia coli* as a host. Filamentous phages with other pilus- and host-specificity are Pf3, a *Pseudomonas* phage, and C-2, a *Salmonella* phage. They are specific for IncP-1- and IncC-plasmids, respectively (Stanisich 1974; Bradley et al. 1982).

The filamentous *E. coli* phage I2-2, which has been isolated from Pretoria sewage, is specific for pili encoded by conjugative plasmids of the IncI2-, IncN-, and IncP-incompatibility groups (Coetzee et

al. 1982; Bradley et al. 1983). Its serological relationship with phage IKeh, a host range mutant of IKe that also infects *E. coli* cells containing plasmids of the IncP-incompatibility group, is indicative of an evolutionary relationship between these two phages (Grant et al. 1978; Bradley et al. 1983). Based upon this relationship it has been suggested that I2-2 is a naturally occurring host range mutant of IKe (Bradley et al. 1983).

Because we are interested in the evolutionary relationships of filamentous phages and because comparison of sequences of evolutionarily related genes or genomes might provide important clues about the significance of particular nucleotide and/or amino acid sequences with respect to their function, the nucleotide sequence of the circular single-stranded genome of I2-2 has been determined and compared with those of Ff and IKe. From these comparisons it became apparent that I2-2, IKe, and Ff have a similar genetic organization and that the genomes of I2-2 and IKe are evolutionarily more closely related than those of I2-2 and Ff. The homology between the nucleotide sequences of the genomes of I2-2 and IKe or Ff is, however, only partial and restricted to the *cis*-acting nucleotide sequence required for phage assembly and the genes coding for the coat and phage assembly proteins.

Materials and Methods

Bacteria, Bacteriophages, and Plasmids. Bacteriophage I2-2 and its host *E. coli* JE2571[N3] [*leu, thr, str, fla, pil* (N3, *tet, strep, sulf*)] were kindly provided by Dr. D. E. Bradley, St. John's, Newfoundland (Watanabe et al. 1964; Bradley et al. 1982). For propagation of recombinant pKUN19 phagemids (Konings et al. 1987), *E. coli* JM83 [*ara,* del(*lac-proAB*), *rpsL, phi80, lac-Z*delM15] (Yanisch-Perron et al. 1985) was used. For the production of single-stranded phagemid DNA by recombinant pKUN19, phagemid-harboring cells [*E. coli* JM101 (*supE, thi,* del(*lac-proAB*) (F', *traD36, proAB, lacIqZ*delM15))], the helper phages IR1 (Enea and Zinder 1982), Mike (Konings et al. 1986), or Mike-delta (Konings et al. 1987) were used. The single-stranded bacteriophage vectors M13mp18 and M13mp19, which were propagated on *E. coli* JM101, have been described (Yanisch-Perron 1985).

Enzymes and Chemicals. Restriction endonucleases were purchased from Boehringer-Mannheim (Germany), New England Biolabs (Beverly, MA, USA), and Bethesda Research Laboratories (Gaithersburg, MD, USA). T4 DNA ligase, *E. coli* DNA polymerase I (Klenow fragment), and calf intestine alkaline phosphatase were purchased from Boehringer-Mannheim. All enzymes were used according to the manufacturers' instructions.

The 2',3'-dideoxyribonucleosidetriphosphates and 2'-deoxyribonucleosidetriphosphates were purchased from Pharmacia (Uppsala, Sweden) and Boehringer-Mannheim, respectively. α-^{32}P-labeled deoxyadenosine triphosphate and α-^{35}S-labeled thiodeoxyadenosine triphosphate were obtained from Amersham (Buckinghamshire, UK).

The oligonucleotides, used for nucleotide sequence analysis of the viral strand, are listed in Table 1. They were synthesized either by a Cyclone Plus DNA synthesizer (MilliGen/Biosearch) or by an Applied Biosystems DNA synthesizer.

Table 1. Synthetic oligonucleotides and *Hae*III fragments of I2-2 RF that have been used as nucleotide sequence analysis primers for the viral strand of phage I2-2

Primer	Nucleotide positions in the I2-2 genome
5'-dCCCAGCCTAATTTACGGGC-3'	226–244
5'-dGTGAACTTTCGGTAATTG-3'	394–411
5'-dGCGCACAGGAGAATTG-3'	1070–1085
5'-dACTGTTATTGCCTGGGTA-3'	1485–1502
5'-dTTAACGCTTGCAGCC-3'	1895–1909
5'-dGGCAATGATAGCGTG-3'	3020–3033
5'-dGACGATACGATTCTAG-3'	3406–3421
5'-dGGCACAGACAGTTCC-3'	5515–5529
5'-dACATCTTCCTTTGTC-3'	5757–5771
5'-dCGCTATCTGGTAGTC-3'	6558–6572
*Hae*III fragments	4490–4708
	4709–4832
	4833–5229
	5230–91

I2-2 replicative form DNA and single-stranded DNA. Replicative form DNA (RF) was isolated from I2-2-infected *E. coli* JE2571[N3] cells, using the alkaline lysis method of Birnboim and Doly (1979) as described by Konings et al. (1987). The procedures for the preparation of a high-titer I2-2 phage stock and the isolation of viral single-stranded DNA have been described previously (Konings et al. 1987).

Cloning of I2-2 DNA and nucleotide sequence analysis. Basic DNA manipulations, which will not be elaborated here, were mainly performed according to Sambrook et al. (1989).

A library of the I2-2 genome was constructed by random cloning of *Hae*III restriction fragments of I2-2 RF in the *Sma*I site of either pKUN19, M13mp18, or M13mp19. After digestion with *Sau*3AI, the *Hae*III restriction fragments were subcloned in the *Bam*HI restriction site of pKUN19.

The nucleotide sequence of the cloned *Hae*III and *Sau*3AI fragments was determined by the chain termination method developed by Sanger et al. (1977). For the analyses also the single-stranded I2-2 genome was used as a template. As sequencing primers, synthetic oligonucleotides and *Hae*III fragments of I2-2 RF (Table 1) were used. To prevent band compression with dGTP analog 7-deaza-dGTP, instead of dGTP, was used.

The nucleotide sequence was analyzed by computer programs developed by Staden (1980, 1982) and Wilbur and Lipman (1983), and the computer program package IntelliGenetics Suite (release 5.37) of IntelliGenetics, Inc. (Mountain View, California, USA). To search for similar nucleotide sequences, the EMBL (release 23) and Genbank (release 60) Nucleic Acid Databases were scanned. The sequence data are stored in the EMBL, Genbank, and DDBJ Nucleotide Sequence Databases under the accession number X14336.

To estimate the evolutionary relationship between the homologous genes (I, III, IV, VI, VII, VIII, and IX) of I2-2, IKe (Peeters et al. 1985), and Ff (Van Wezenbeek et al. 1980), the extent of synonymous (Ks = synonymous substitutions/synonymous sites) and nonsynonymous amino acid substitutions (Ka = amino acid substitutions/amino acid sites) was calculated according to Miyata and Yasunaga (1980) and corrected for back mutations by the method of Jukes and Cantor (1969). Triplet

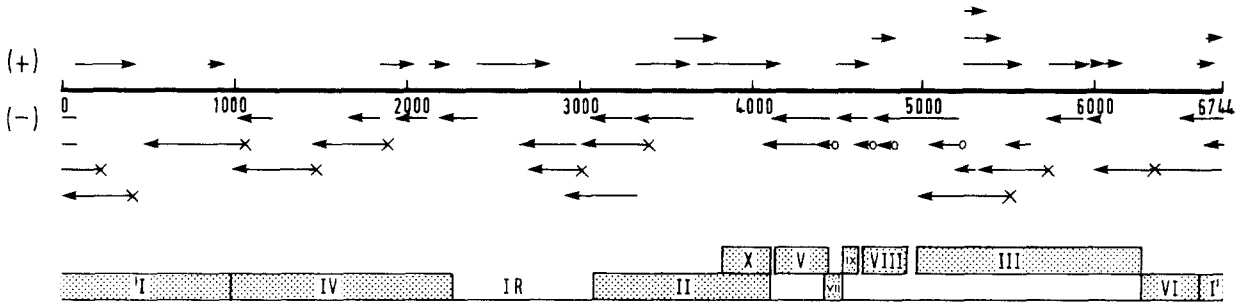

Fig. 1. Positions of individual gel readings in the final nucleotide sequence of the I2-2 genome. The axis represents the nucleotide map of the I2-2 genome, starting at the unique *Hin*dIII recognition site, numbered in the 5′–3′ direction of the viral strand. Arrows pointing to the right and left refer to nucleotide sequences established for the viral and complementary strand, respectively. The length of the sequence derived is represented by the length of the arrow. Nucleotide sequences indicated by → or ← are obtained from cloned DNA, whereas those indicated by ←x and ←o are obtained from nucleotide analysis of viral DNA primed with synthetic oligonucleotides or *Hae*III restriction fragments, respectively. The linear genetic map of the I2-2 genome is given in the lower portion of the diagram. Genes are denoted by Roman numerals; IR refers to the major intergenic region in which the origins for viral and complementary strand synthesis and the signal required for efficient phage assembly are located.

insertions/deletions were excluded from the comparisons. The Ks and Kas calculated are thus overestimates.

Results and Discussion

Nucleotide Sequence of I2-2 DNA

The sequence analysis of the I2-2 genome was initiated by sequencing of I2-2 *Hae*III fragments randomly cloned into phagemid pKUN19. Because three *Hae*III fragments were longer than 1000 bp, and due to subcloning difficulties, several gaps remained. These gaps were filled in by using the viral genome as a template with synthetic oligonucleotides or purified *Hae*III fragments as sequencing primers (vide supra). Furthermore, the nucleotide sequence of several *Sau*3AI subclones of *Hae*III fragments was determined. The sequencing strategy followed is outlined in Fig. 1.

The majority of the I2-2 sequence was established via sequence analysis of both the viral and complementary strand. On the average each nucleotide is supported by at least three separate gel readings. We therefore feel confident that the nucleotide sequence as presented in Fig. 2 is correct. As nucleotide number 1 we have chosen the first nucleotide (A) of the unique *Hin*dIII restriction endonuclease cleavage site. The circular single-stranded genome of I2-2 is 6744 nucleotides long. This is 139 nucleotides less than that of IKe (Peeters et al. 1985), and 337 (336) nucleotides more than that of Ff (Beck et al. 1978; Van Wezenbeek et al. 1980; Beck and Zink 1981; Hill and Petersen 1982).

Gene Identification

The five coat protein genes (i.e., genes III, VI, VII, VIII, and IX), as well as the two phage assembly genes (I and IV) and the phage morphogenetic signal located immediately distal to gene IV, have been identified by alignment of the I2-2 sequence with that of phage IKe. The respective gene products are not only highly homologous but also of similar lengths (Fig. 2). Although no significant homology was observed between the open reading frames II, X, and V of I2-2 and the DNA replication genes II, X, and V of IKe, as demonstrated here, we feel confident in concluding that the proteins encoded by these open reading frames fulfill an identical function in the DNA replication process. This conclusion is based partly on the observation that the biochemical and physical properties (DNA-binding properties, amino acid composition, and molecular weight) of the I2-2 encoded single-stranded DNA-binding protein are in accordance with those of the protein encoded by open reading frame V (data not shown). Furthermore, genetic and biochemical studies have unambiguously demonstrated that the filamentous phage genome codes for three different functions that are indispensable for the phage (DNA) replication process. One of these functions is encoded by gene II and indispensable for initiation and termination of viral strand replication. The second function is encoded by gene X, a gene that completely overlaps with the 3′ terminal region of gene II, and is somehow involved in the regulation of the synthesis of the complementary strand. The third function is the single-stranded DNA-binding protein encoded by gene V that, at a late stage of the infection cycle, is indispensable for the sequestration of the newly synthesized viral strand from the DNA replication cycle. Because the only DNA region that is left in the I2-2 genome that might code for these proteins is located between the major intergenic region and gene VII (Fig. 3), and on the basis of the strong resemblance with the genetic or-

ganization of IKe and Ff, we conclude that the open reading frames indicated code for these indispensable DNA replication functions. As in the phages IKe and Ff, in I2-2 the 3′ terminal part of gene II overlaps gene X completely.

The proposed genetic organization is further substantiated by the observation that all initiation codons of the proposed genes and/or open reading frames are preceded by a sequence that is significantly homologous to the consensus sequence for ribosome binding (Table 2) (Shine and Dalgarno 1974). All I2-2 genes start with an ATG codon, except gene III, which starts with the codon GTG. Similar observations have been made during our studies of the genomes of Ff and IKe (Van Wezenbeek et al. 1980; Peeters et al. 1985).

Genetic Organization

The genetic maps of the genomes of I2-2, IKe, and Ff (M13) are depicted in Fig. 3 (Van Wezenbeek et al. 1980; Peeters et al. 1985). The three genomes consist of an equal number of genes, which are arranged in an identical order (vide supra). Comparison of the genomes of I2-2 and Ff reveals that, with the exception of genes II and VII, all I2-2 genes are larger than their Ff counterparts. The larger length of I2-2 compared with Ff is mainly the result of length differences between their respective major intergenic regions (Fig. 3).

Apart from the major intergenic region that is located between genes IV and II, a much smaller one is located between genes VIII and III. It contains a rho-independent transcription termination signal, identical in sequence to that of IKe and very similar to that of M13 (Fig. 4) (Van Wezenbeek et al. 1980; Peeters et al. 1985). In Ff and IKe the DNA region encompassing genes II through VIII is expressed in a nested set of mRNAs whose synthesis is initiated at different promoters but terminated at the aforementioned rho-independent transcription termination signal (Edens et al. 1978). Because in the DNA region of I2-2 encompassing the major intergenic region and genes II–VIII, multiple nucleotide sequences are found that resemble the consensus sequence of *E. coli* promoters and which have an almost identical map position as the promoters found in IKe and Ff (data not shown), we suggest that this part of the I2-2 genome is expressed via a similar transcription mechanism.

Besides the complete overlap of gene X with the 3′ terminal region of gene II, in I2-2 the reading frames of genes V and VII, VII, and IX, and of III and VI overlap each other with a few nucleotides. A similar compact genetic organization has been found to be present in the genomes of IKe (Peeters et al. 1985) and Ff (Van Wezenbeek et al. 1980), but with the exception that in Ff genes IX and VIII and genes I and IV overlap, rather than genes V and VII and genes III and VI (Figs. 2 and 3).

Intergenic Region

The major intergenic region (IR; 508 nucleotides) of phage Ff is a noncoding sequence that encompasses all *cis*-acting elements required for phage assembly and DNA replication (Zinder and Horiuchi 1985). As one moves from the 5′ to the 3′ end, there are five consecutive regions (designated A–E) of dyad symmetry capable of forming stable stem-loop structures, followed by an AT-rich sequence without self-complementarity and that fulfills an enhancer function in the DNA replication process (Cleary and Ray 1980, 1981; Dotto et al. 1981, 1982). Region A functions both as a rho-dependent transcription termination signal (Moses and Model 1984; Smits et al. 1984) and as a morphogenetic signal for the proper packaging of the viral strand into filamentous particles (Dotto and Zinder 1983). Regions B and C function as origin for the synthesis of the complementary viral strand (Schaller et al. 1976; Gray et al. 1978), whereas the "rabbit ears structure" consisting of regions D and E fulfills an indispensable role in the initiation and termination processes of viral strand replication (Meyer et al. 1979).

The major intergenic region of the genome of IKe is 738 nucleotides long. Similar to Ff this intergenic region consists of five regions with dyad symmetry, which are homologous to those found in the intergenic region of Ff (Peeters et al. 1985). The homology does not hold, however, for the major intergenic region of I2-2, although in this region of 817 nucleotides five stretches with dyad symmetry are also located, which resemble in number, relative position, and structure those found in the intergenic regions of IKe and Ff (Fig. 5). Only stretch A has a homologous counterpart in the genomes of Ff and IKe (Fig. 5B). Because the phage-encoded proteins that cross-talk with this sequence (i.e., the coat proteins and probably also the phage assembly proteins encoded by genes I and IV) are highly conserved, we propose that this sequence functions in I2-2 also as a phage morphogenesis signal. The observation that the secondary structure characteristics of the intergenic regions of I2-2, IKe, and Ff are very much alike highly favors the hypothesis that, despite the lack of homology, in I2-2 the DNA stretches of dyad symmetry have the same function in the phage replication cycle as they do in the phages Ff and IKe.

Evolution of the Filamentous Phage Genome

Based upon the data presented it is clear that the bacteriophages I2-2, IKe, and Ff are evolutionarily

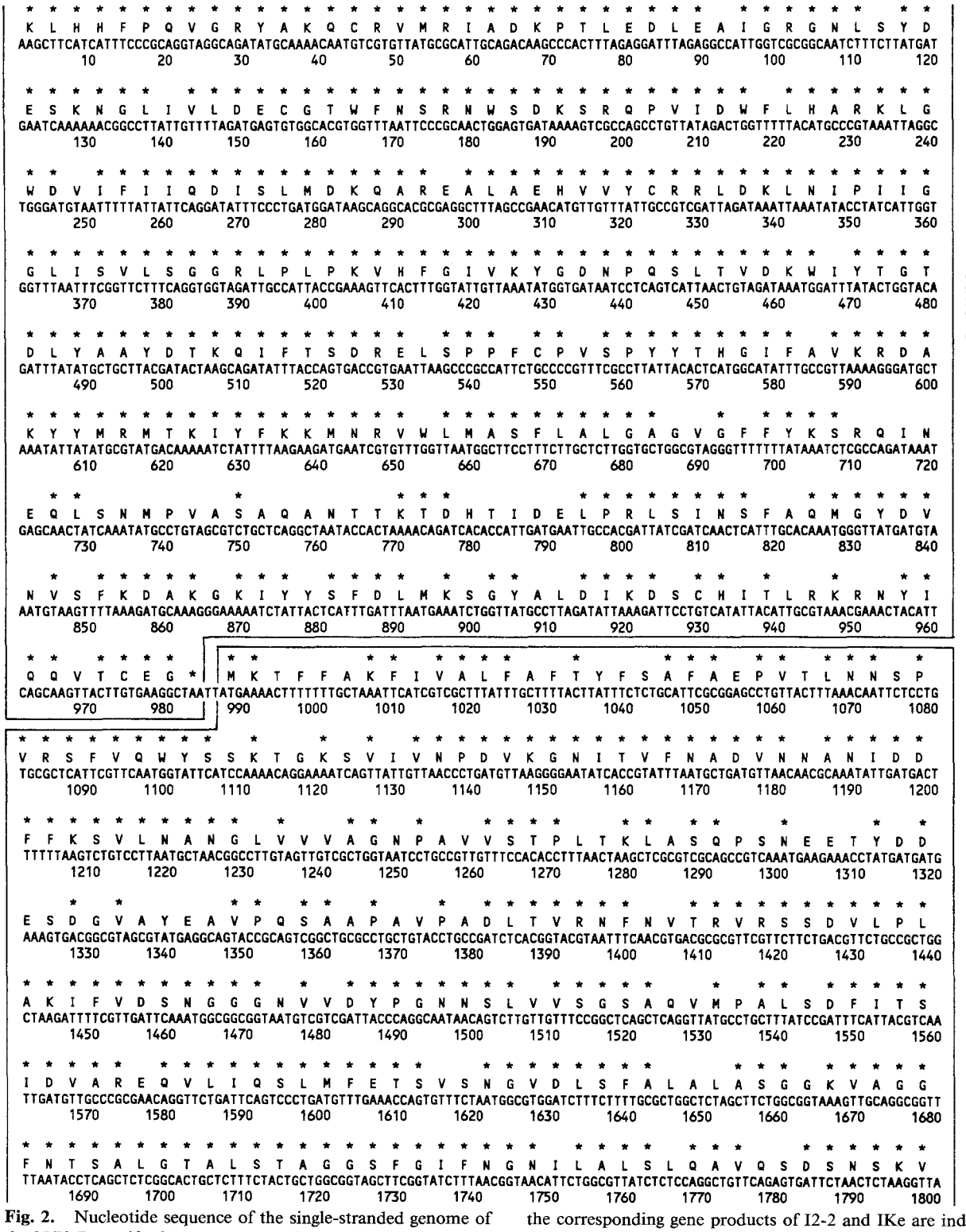

Fig. 2. Nucleotide sequence of the single-stranded genome of the N/I2/P-specific filamentous *E. coli* phage I2-2. Numbering is in the 5′–3′ direction of the viral strand and begins at the first nucleotide of the unique *Hin*dIII recognition site. The genes I–X are boxed and the amino acid sequences of their encoded proteins are shown. For clarity, gene VII is boxed with a broken line. The amino acid residues that have an identical position in the corresponding gene products of I2-2 and IKe are indicated by asterisks. Termination codons are denoted by asterisks above the nucleotide sequence. IR refers to the major intergenic region in which the origins for viral and complementary strand synthesis and the signal required for efficient phage assembly are located. Continued on pp. 146–148.

related. Their genetic organization is almost identical and genes I, III, IV, VI, VII, VIII, and IX of I2-2 and IKe are highly homologous. Furthermore, there are no significant differences in codon usage and at the third codon position there is a strong preference for a T (46%). The evolutionary relationship between filamentous phages can be estimated by comparing the extent of nucleotide se-

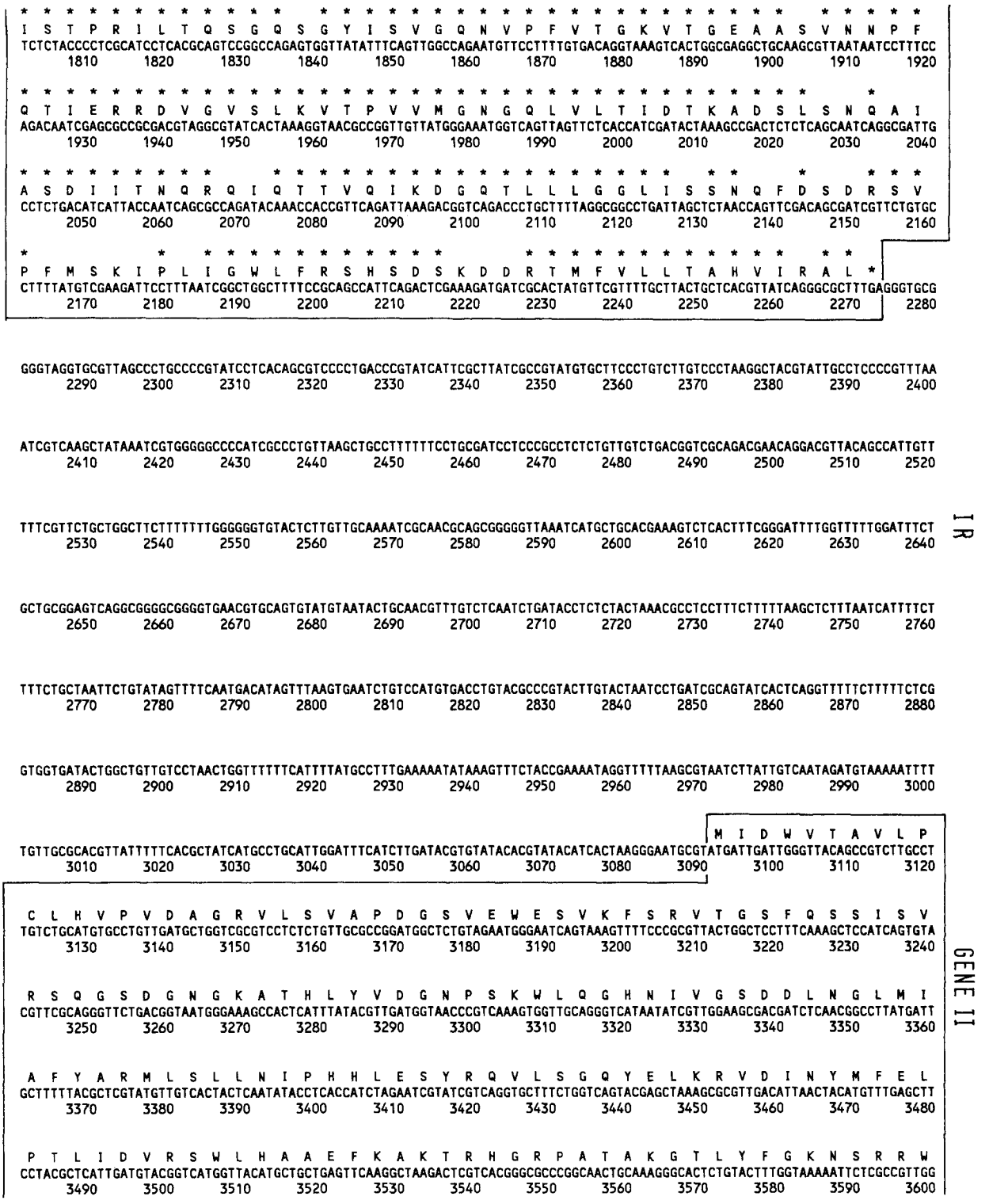

Fig. 2. Continued

quence similarity of genes I, III, IV, VI, VII, VIII, and IX of I2-2, IKe, and Ff(M13), respectively. The extent of synonymous and nonsynonymous amino acid substitutions between the various genes was estimated according to Miyata and Yasunaga (1980) (Table 3). In virtually all comparisons the Ks value is larger than 1, indicating that the genes under consideration did not diverge recently. The reason why the nucleotide sequences of genes III and VIII of I2-2 and IKe are more strongly conserved than all others is unclear. The Ka values (Table 3), i.e., the values for the relative number of amino acid substitutions, clearly indicate that I2-2 is more closely related to IKe than to Ff(M13).

Genetic and cell biologic studies indicate that the protein encoded by gene III of Ff is specifically re-

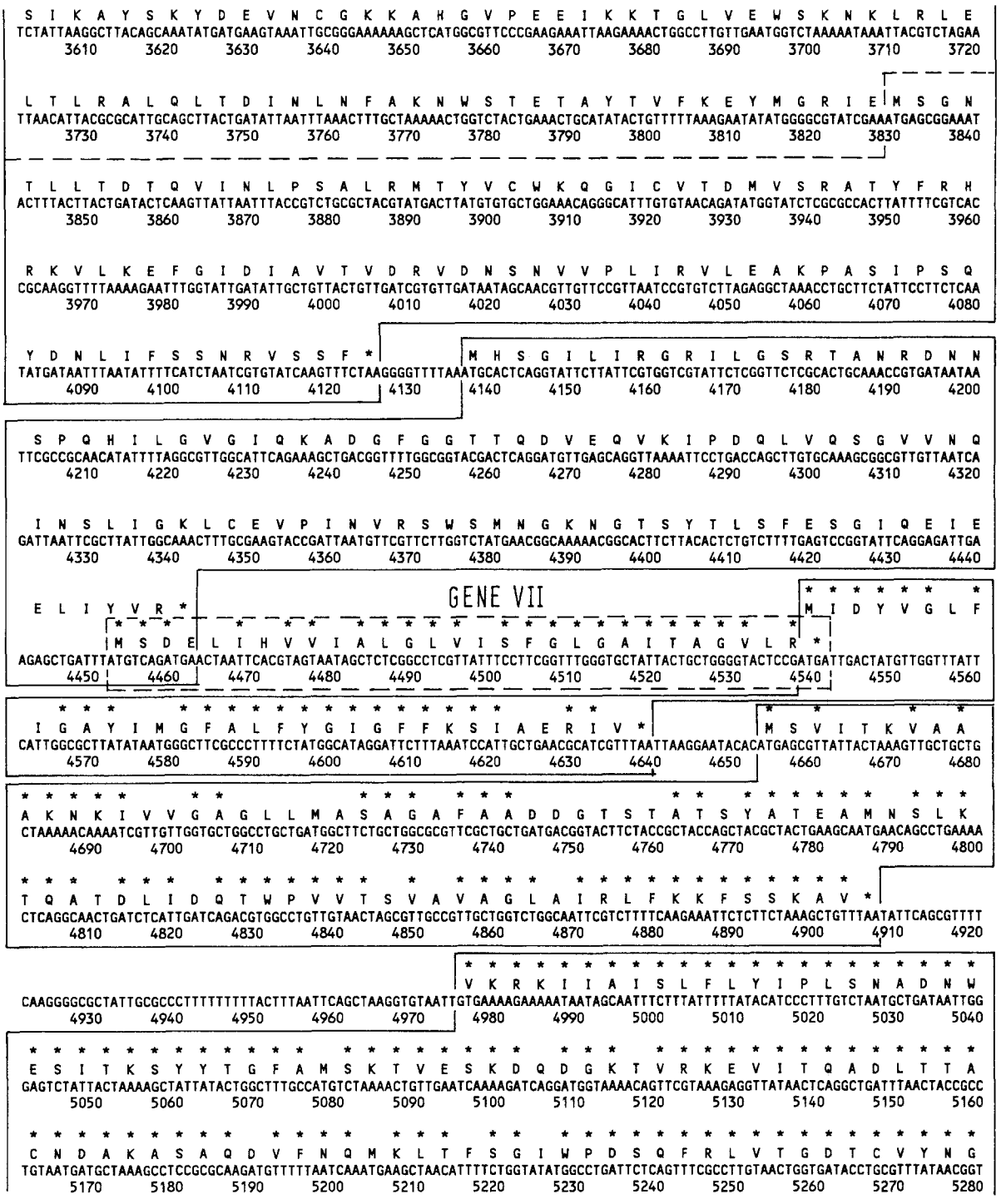

Fig. 2. Continued

quired for attachment of the phage to the conjugative pilus of the host cell. The minor, randomly distributed differences between the gene III proteins of IKe and I2-2 (Fig. 2; 30 out of 434 amino acids) indicate that only a few amino acid substitutions are sufficient to alter/extend the plasmid specificity of the phage (vide supra). The observed serological relationship (Bradley et al. 1983) between IKe, IKeh (a host range mutant of IKe), and I2-2, is in accordance with the high amino acid sequence homology found between the coat proteins encoded by gene VIII of both I2-2 and IKe. However, our sequencing data prove that, contrary to earlier speculations (Bradley et al. 1983), I2-2 is not a host range mutant of IKe. This can also be concluded from dot-matrix comparisons of the genomic sequences of I2-2, IKe, and Ff (Fig. 6). In these analyses the homology between successive stretches of heptanucleotides of the genomes of either I2-2 and IKe (Fig. 6A), I2-2 and Ff(M13) (Fig. 6B), or IKe and Ff(M13) (Fig. 6C) are

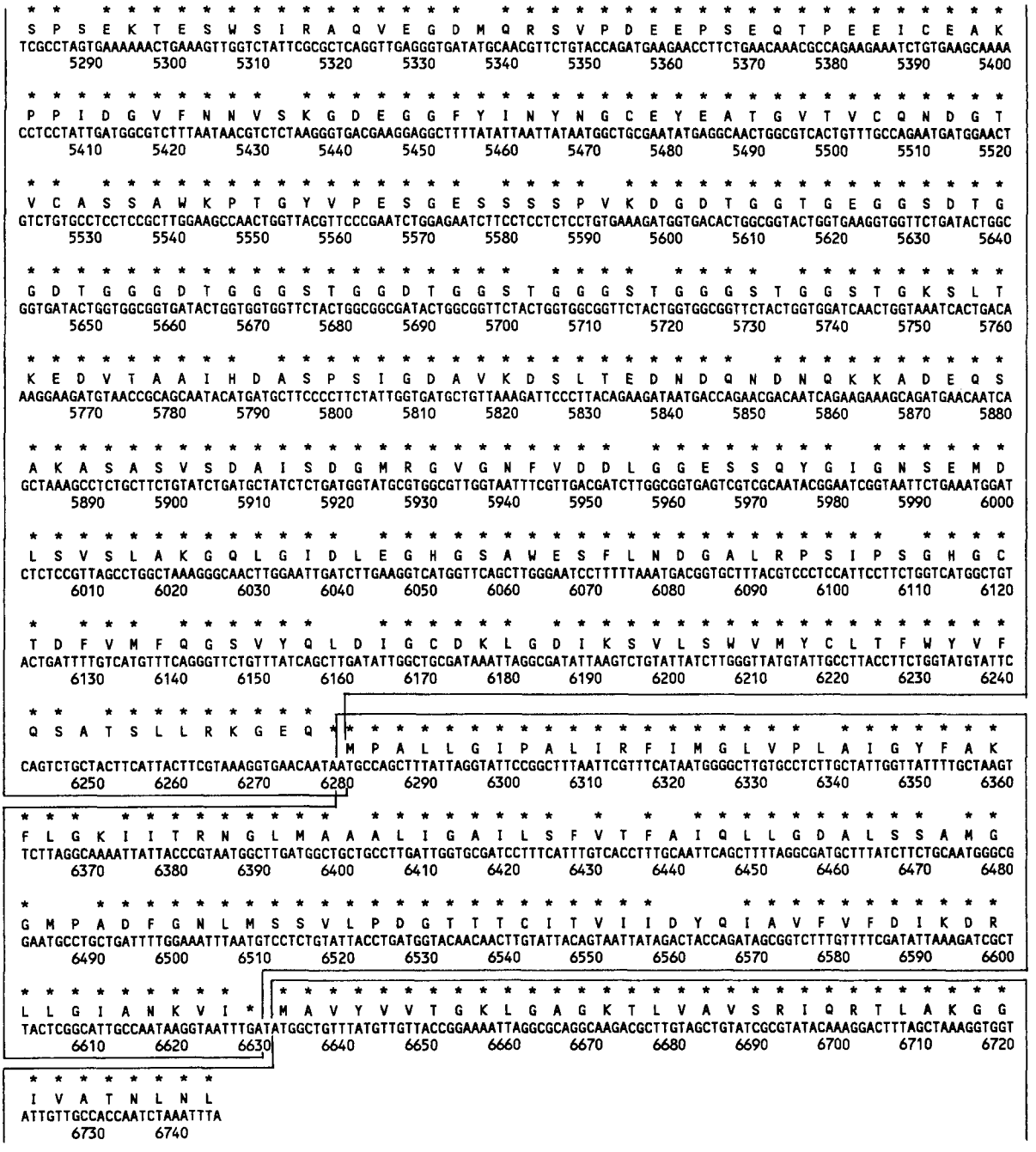

Fig. 2. Continued

compared. The genomes are aligned by using the initiation codon of gene VII as a reference point. From these plots it can be concluded that genes VII, IX, VIII, III, VI, I, and IV and the 5′ end of the intergenic region (hairpin A) of I2-2 and IKe are highly similar. Immediately after the morphogenetic signal the sequence correspondence decreases abruptly. A similar observation has been made when the genomes of I2-2 and Ff(M13) were compared, although in this case the similarity between the DNA sequence located between the 5′ end of gene VII and the 3′ end of hairpin A is less pronounced (Fig. 6B).

The observation that no significant homology exists between the DNA regions of I2-2 and IKe or Ff that encompass the major part of the intergenic region (vide supra), genes II, X, and V suggests that the genome of I2-2 is composed of DNA sequences derived from two different replicons. We propose that the I2-2 genome is the result of a recombination event in which the replication cassette, i.e., the DNA fragment containing the viral and complementary strand origins of DNA replication and the phage-encoded replication genes of its filamentous phage ancestor genome, has been exchanged for the rolling circle replication cassette of another replicon. Whether the latter cassette originated from another

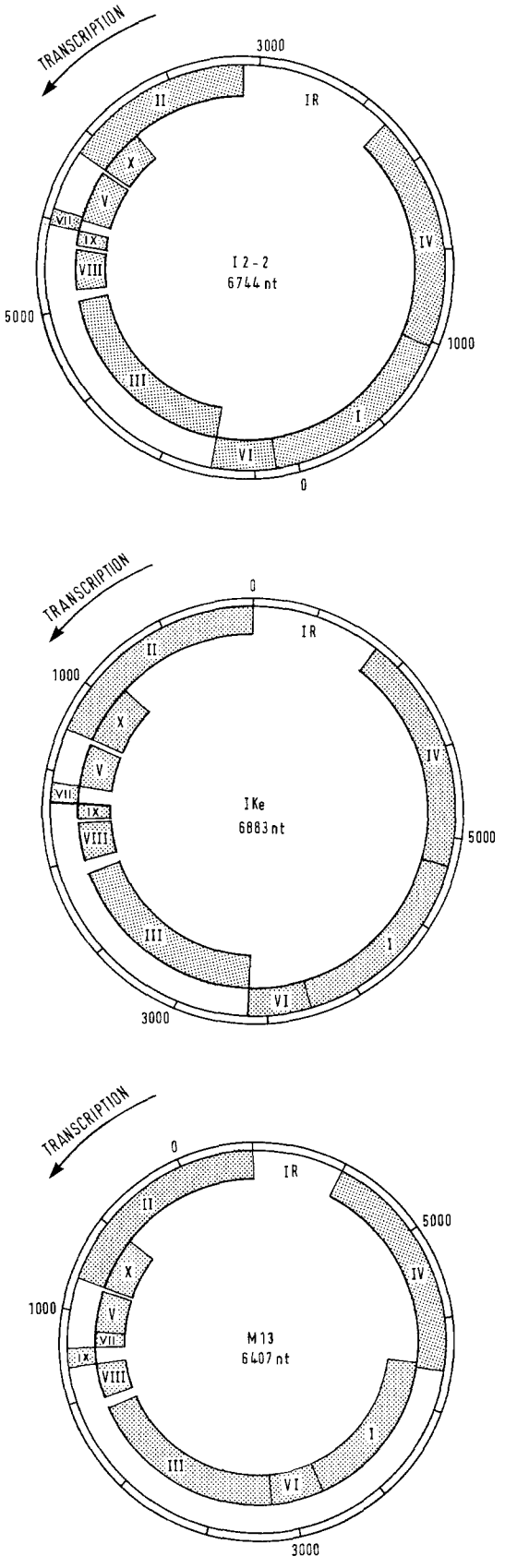

Fig. 3. Comparison of the circular genetic maps of the genomes of the filamentous phages I2-2, IKe, and Ff(M13). Genes are indicated by Roman numerals. IR refers to the major intergenic region located between genes IV and II, in which both the origins for viral and complementary strand synthesis and the *cis*-acting element required for efficient phage assembly are located.

Table 2. Shine–Dalgarno sequences for the assigned and proposed genes of the I2-2 genome

16S RNA	3′ OH-AUUCCUCCACUAG
Gene I	CAATAAGGTAATTTGAT.**ATG**
Gene II	TAAGGGAATGCGT.**ATG**
Gene III	TAAGGTGTAATT.**GTG**
Gene IV	TGTGAAGGCTAATT.**ATG**
Gene V	TAAGGGGTTTTAA.**ATG**
Gene VI	CGTAAAGGTGAACAATA.**ATG**
Gene VII	TGAAGAGCTGATTT.**ATG**
Gene VIII	TAAGGAATACAC.**ATG**
Gene IX	TGCTGGGGTACTCCG.**ATG**
Gene X	TATGGGGCGTATCGAA.**ATG**

Nucleotides that are complementary to the 3′ terminal end of 16S ribosomal RNA are underlined (Shine and Dalgarno 1974). Initiation codons are indicated in bold

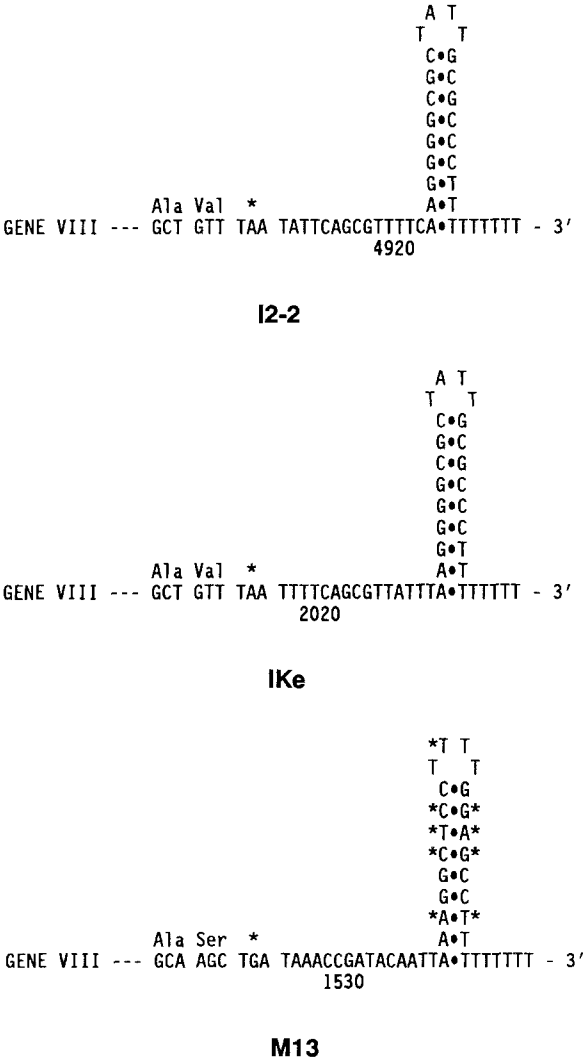

Fig. 4. Nucleotide sequences and presumed secondary structures of the rho-independent transcription termination signals of phages I2-2, IKe, and Ff(M13), which are located in the intergenic region between genes VIII and III. Nucleotides in the Ff sequence, encompassing the region of dyad symmetry, that are different from those in I2-2 and IKe are indicated with asterisks.

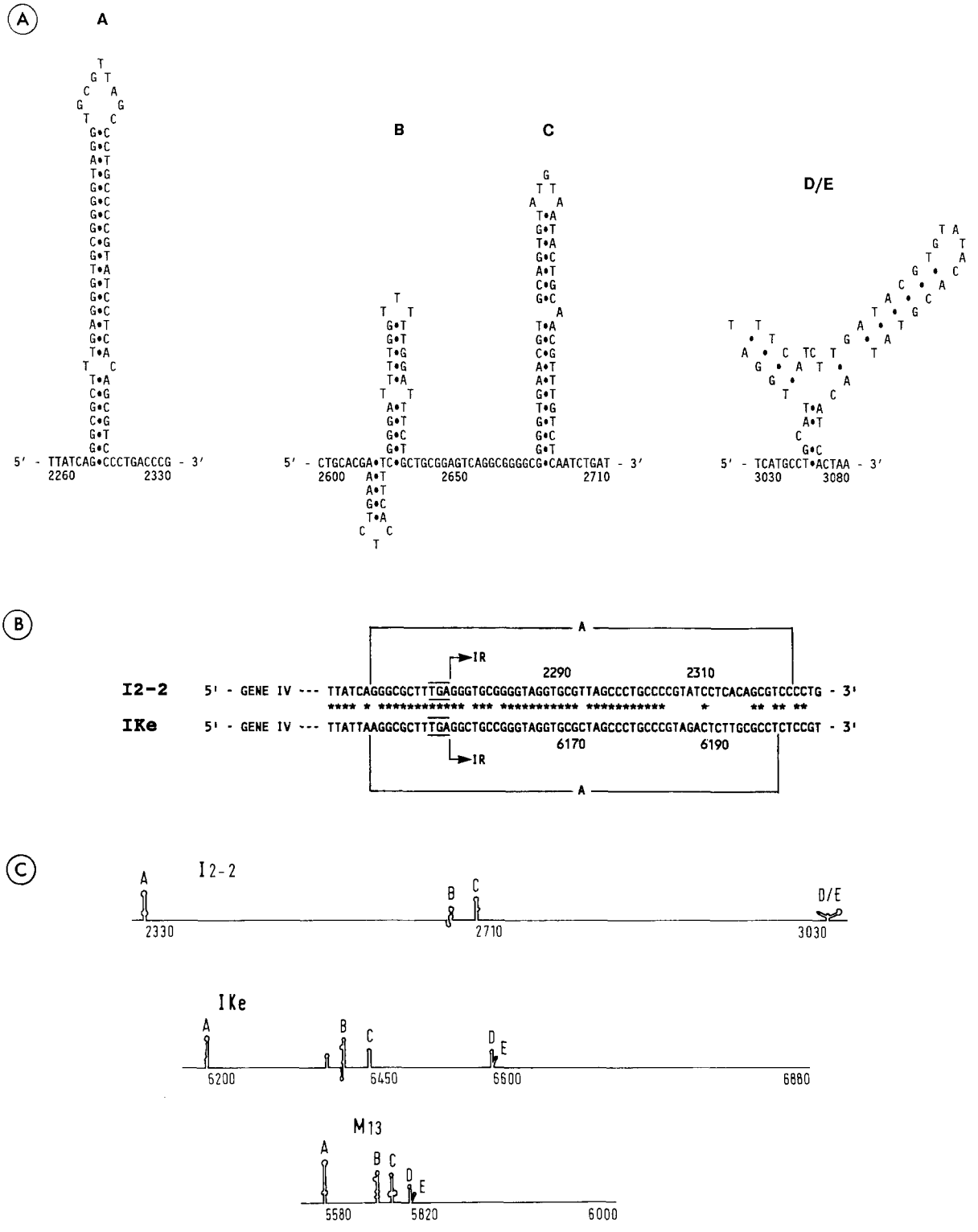

Fig. 5. Comparison of the nucleotide sequence and secondary structural configuration of the regions of dyad symmetry in the major intergenic regions of the phages I2-2, IKe, and Ff(M13). **A** Nucleotide sequence (5'–3') and potential secondary structures in the I2-2 major intergenic region. **B** Comparison of the nucleotide sequence of region A of the intergenic region of the phages I2-2 and IKe. Within this sequence both the phage morphogenetic signal and a rho-dependent transcription termination signal are located. Sequences are aligned in such a way that maximum homology is achieved. Regions of dyad symmetry are indicated by brackets. The stop codons of genes IV of I2-2 and IKe are over- and underlined. **C** Comparison of morphology and relative map position of the potential stem-loop structures (hairpins A through E) in the major intergenic region of the phages I2-2, IKe, and Ff(M13). The configuration are drawn to scale.

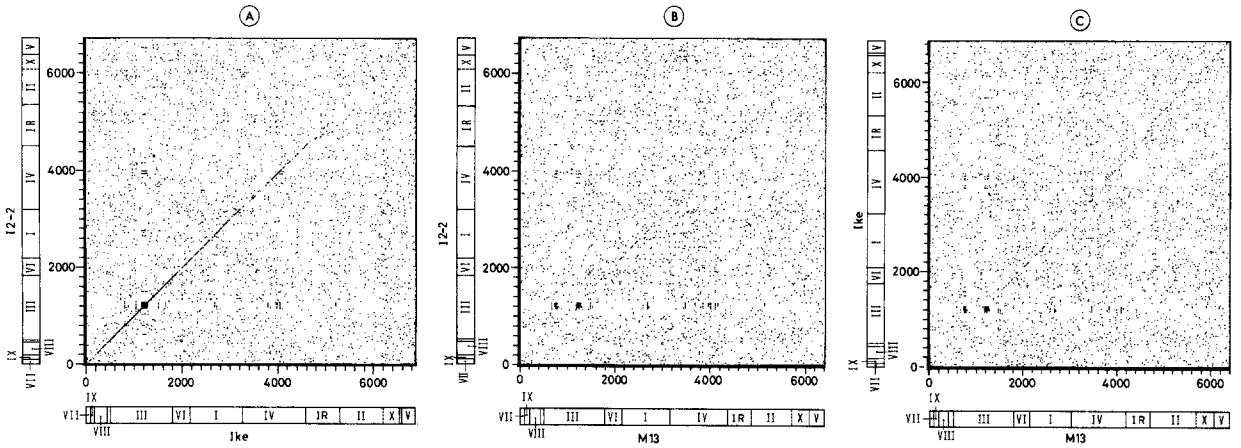

Fig. 6. Dot-matrix analysis of the extent of homology between the genomes of the phages I2-2, IKe, and Ff(M13). Starting from the 5′ terminal nucleotide of gene VII consecutive stretches of seven nucleotides are compared. Each dot represents a stretch of seven identical nucleotides. Parallel to the axes the genetic maps are indicated. **A** Dot-matrix of I2-2 versus IKe. **B** Dot-matrix of I2-2 versus Ff(M13). **C** Dot-matrix of IKe versus Ff(M13).

Table 3. Extent of synonymous (Ks) and nonsynonymous (Ka) substitutions in genes I–IX of the phages I2-2, IKe, and Ff(M13)

	Sites compared	Synonymous difference (Ks)	Non-synonymous difference (Ka)
I2-2 vs IKe			
Gene I	365	>1	0.107
Gene III	432	0.115	0.032
Gene IV	427	>1	0.163
Gene VI	116	>1	0.062
Gene VII	29	>1	0.148
Gene VIII	79	0.563	0.158
Gene IX	33	>1	0.120
I2-2 vs M13			
Gene I	342	>1	0.395
Gene III	338	>1	0.709
Gene IV	408	>1	0.458
Gene VI	109	>1	0.737
Gene VII	28	>1	0.573
Gene VIII	63	>1	0.429
Gene IX	31	>1	0.786
IKe vs M13			
Gene I	342	>1	0.410
Gene III	337	>1	>1
Gene IV	409	>1	0.482
Gene VI	109	>1	0.729
Gene VII	31	>1	0.818
Gene VIII	64	>1	0.571
Gene IX	31	0.874	0.752

The extent of synonymous and nonsynonymous substitutions was calculated as indicated in Materials and Methods

filamentous phage or from a plasmid with a rolling circle replication mode is not yet clear, as computer searches in the EMBL (release 23) and Genbank (release 60) nucleic acids databases have not yet revealed the existence of nucleotide sequences that are evolutionarily related to the replication module of I2-2.

Acknowledgments. We thank our colleagues, Lettie Lubsen and Graham Elliott, for reading the manuscript and making corrections and helpful suggestions. The work has been supported (Maoxiao-Yu) by a grant from the Dutch Ministry of Education and Science.

References

Beck E, Zink B (1981) Nucleotide sequence and genome organisation of filamentous bacteriophages f1 and fd. Gene 16: 35–58

Beck E, Sommer R, Auerswald EA, Kurz C, Zink B, Osterburg G, Schaller H (1978) Nucleotide sequence of bacteriophage fd DNA. Nucleic Acids Res 5:4495–4503

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513–1523

Bradley DE, Sirgel FA, Coetzee JN, Hedges RW, Coetzee WF (1982) Phages C-2 and J: IncC and IncI plasmid-dependent phages, respectively. J Gen Microbiol 128:2485–2498

Bradley DE, Coetzee JN, Hedges RW (1983) IncI2 plasmids specify sensitivity to filamentous bacteriophage IKe. J Bacteriol 154:505–507

Caro LG, Schnös M (1966) The attachment of the male-specific bacteriophage f1 to sensitive strains of *Escherichia coli*. Proc Natl Acad Sci USA 56:126–132

Cleary JM, Ray DS (1980) Replication of the plasmid pBR322 under the control of a cloned replication origin from the single-stranded DNA phage M13. Proc Natl Acad Sci USA 77: 4638–4642

Cleary JM, Ray DS (1981) Deletion analysis of the cloned replication origin region from bacteriophage M13. J Virol 40: 197–203

Coetzee JN, Bradley DE, Hedges RW (1982) Phages Iα and I2-2: IncI plasmid-dependent bacteriophages. J Gen Microbiol 128:2797–2804

Dotto GP, Zinder ND (1983) The morphogenetic signal of bacteriophage f1. Virology 130:252–256

Dotto GP, Enea V, Zinder ND (1981) Functional analysis of bacteriophage f1 intergenic region. Virology 114:463–473

Dotto GP, Horiuchi K, Zinder ND (1982) Initiation and termination of phage f1 plus-strand synthesis. Proc Natl Acad Sci USA 79:7122–7126

Edens L, Konings RNH, Schoenmakers JGG (1978) A cascade

mechanism of transcription in bacteriophage M13 DNA. Virology 86:354–367

Enea V, Zinder ND (1982) Interference resistant mutants of phage f1. Virology 122:222–226

Grant RB, Whiteley MH, Shapley AJ (1978) Plasmids of incompatibility group P code for the capacity to propagate bacteriophage IKe. J Bacteriol 136:808–811

Gray CP, Sommer R, Polke C, Beck E, Schaller H (1978) Structure of the origin of DNA replication of bacteriophage fd. Proc Natl Acad Sci USA 75:50–53

Hill DF, Petersen GB (1982) Nucleotide sequence of bacteriophage f1 DNA. J Virol 44:32–46

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Khatoon H, Iyer RV, Iyer VN (1972) A new filamentous bacteriophage with sex-factor specificity. Virology 48:145–155

Konings RNH, Luiten RGM, Peeters BPH (1986) Mike, a chimeric filamentous phage designed for the separate production of either DNA strand of pKUN vector plasmids by F^+ cells. Gene 46:269–276

Konings, RNH, Verhoeven EJM, Peeters BPH (1987) pKUN, vectors for the separate production of both DNA strands of recombinant plasmids. Methods Enzymol 153:12–34

Marvin DA, Folkhard W (1986) Structure of F-pili: reassessment of the symmetry. J Mol Biol 191:299–300

Meyer TF, Geider K, Kurz C, Schaller H (1979) Cleavage site of bacteriophage fd gene II-protein in the origin of viral strand replication. Nature 278:365–367

Miyata T, Yasunaga T (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Mol Evol 16:23–36

Model P, Russel M (1988) Filamentous bacteriophage. In: Calendar R (ed) The bacteriophages, vol 2. Plenum, New York, pp 375–456

Moses PB, Model P (1984) A rho-dependent transcription termination signal in bacteriophage f1. J Mol Biol 172:1–22

Peeters BPH, Peters RM, Schoenmakers JGG, Konings RNH (1985) Nucleotide sequence and genetic organization of the genome of the N-specific filamentous bacteriophage IKe: comparison with the genome of the F-specific filamentous phages M13, fd and f1. J Mol Biol 181:27–39

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor University Press, Cold Spring Harbor NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467

Schaller H, Uhlmann A, Geider K (1976) A DNA fragment from the origin of single-strand to double-strand DNA replication of bacteriophage fd. Proc Natl Acad Sci USA 73: 49–53

Shine J, Dalgarno L (1974) The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci USA 71:1342–1346

Smits MA, Jansen J, Konings RNH, Schoenmakers JGG (1984) Initiation and termination signals for transcription in bacteriophage M13. Nucleic Acids Res 12:4071–4081

Staden R (1980) A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res 8: 3673–3694

Staden R (1982) An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucleic Acids Res 10:2951–2961

Stanisich VA (1974) The properties and host range of male-specific bacteriophages of *Pseudomonas aeruginosa*. J Gen Microbiol 84:332–342

Van Wezenbeek PMGF, Hulsebos TJM, Schoenmakers JGG (1980) Nucleotide sequence of the filamentous bacteriophage M13 DNA genome: comparison with phage fd. Gene 11:129–148

Watanabe T, Nishida H, Ogata C, Arai T, Sato S (1964) Episome-mediated transfer of drug resistance in Enterobacteriaceae: VII. Two types of naturally occurring R factors. J Bacteriol 88:716–726

Wilbur WJ, Lipman DJ (1983) Rapid similarity searches of nucleic acid and protein data banks. Proc Natl Acad Sci USA 80:726–730

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119

Zinder ND, Horiuchi K (1985) Multiregulatory element of filamentous bacteriophages. Microbiol Rev 49:101–106

Received May 22, 1991/Accepted August 20, 1991