Distribution, Sequence Homology, and Homing of Group I Introns among T-even-like Bacteriophages

EVIDENCE FOR RECENT TRANSFER OF OLD INTRONS*

Received for publication, January 28, 2004, and in revised form, March 15, 2004
Published, JBC Papers in Press, March 15, 2004, DOI 10.1074/jbc.M400929200

Linus Sandegren and Britt-Marie Sjöberg‡

From the Department of Molecular Biology & Functional Genomics, Stockholm University, Svante Arrhenius vag 16, F3, SE-10691 Stockholm, Sweden

Self-splicing group I introns are being found in an increasing number of bacteriophages. Most introns contain an open reading frame coding for a homing endonuclease that confers mobility to both the intron and the homing endonuclease gene (HEG). The frequent occurrence of intron/HEG has raised questions whether group I introns are spread via horizontal transfer between phage populations. We have determined complete sequences for the known group I introns among T-even-like bacteriophages together with sequences of the intron-containing genes td, nrdB, and nrdD from phages with and without introns. A previously uncharacterized phage isolate, U5, is shown to contain all three introns, the only phage besides T4 found with a "full set" of these introns. Sequence analysis of td and nrdB genes from intron-containing and intronless phages provides evidence that recent horizontal transmission of introns has occurred among the phages. The fact that several of the HEGs have suffered deletions rendering them non-functional implies that the homing endonucleases are of no selective advantage to the phage and are rapidly degrading and probably dependent upon frequent horizontal transmissions for maintenance within the phage populations. Several of the introns can home to closely related intronless phages during mixed infections. However, the efficiency of homing varies and is dependent on homology in regions flanking the intron insertion site. The occurrence of optional genes flanking the respective intron-containing gene can strongly affect the efficiency of homing. These findings give further insight into the mechanisms of propagation and evolution of group I introns among the T-even-like bacteriophages.

The three self-splicing group I introns in bacteriophage T4 were the first prokaryotic introns to be found. The T4 introns are situated in the genes coding for the small subunit of the aerobic ribonucleotide reductase (nrdB) (1, 2), the anaerobic ribonucleotide reductase (nrdD [formerly called sunY]) (1, 3), and thymidylate synthase (td) (4). The presence of introns in a bacteriophage was unexpected, because phages generally are under strong selective pressure to keep a compact genome to maintain rapid replication and to fit the genome inside the phage capsid. The subsequent finding that the T4 introns are mobile genetic elements, mediated by the presence of a homing endonuclease gene (HEG)1 within each intron, gives an explanation to how they were inserted into the phage genome (5, 6), but the low number of introns among T-even-like phages has hitherto prevented a systematic study of how the introns have propagated within this phage population. Phages of both Gram-negative and Gram-positive bacteria have subsequently been shown to contain group I introns, all belonging to the same subgroup (IA2) of group I introns as the T-even introns indicating a common ancestry of phage introns. Most phage introns also contain homing endonuclease genes (7–16).

Homing endonuclease genes are a diverse group of proteins that generates double strand cuts in DNA at, or near, a generally very long recognition sequence spanning up to 40 bp (for reviews see Refs. 17–19). HEGs are often associated with self-splicing introns or inteins and can confer mobility both to themselves and to their surrounding splicing element via the process of homing. Intron homing occurs when a HEG-containing intron encounters a cognate chromosome copy without an intron (i.e. for phage introns during mixed infection with a cognate or closely related phage without intron). The recognition sites of intron-associated homing endonucleases generally span their respective intron insertion site, and they will therefore only cleave intronless alleles. Intron-containing alleles are immune to cleavage, because the intron interrupts the recognition site. After cleavage the double-strand break is repaired via recombination-dependent replication repair, using an intact, intron-containing chromosome copy as template (20). This results in insertion of the intron and the HEG into the repaired genome. Efficient homing is dependent upon sequence homology between donor and recipient alleles in the sequences flanking the insertion site (5, 21–23). If the requirement of homology in the homing site and surrounding regions is met the efficiency of homing can be close to 100%.

Of the three HEGs in the T4 introns the td and nrdB HEGs (denoted I-TevI and I-TevII, respectively) belong to the GIY-YIG family of endonucleases. They have been shown to specifically cleave intronless alleles of td and nrdD, respectively (6, 24, 25). The T4 nrdB intron contains a non-functional HEG of the H-N-H family, named I-TevIII, which has suffered a large deletion and therefore cannot promote homing (5, 26). Previous studies have shown that T4 is exceptional among its closer relatives in having several introns. Most of the known T-even-like phages completely lack introns, whereas a small number

* This work was supported by the Swedish Science Research Council. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY262125–AY262157.

‡ To whom correspondence should be addressed. Tel.: 46-8-164-150; Fax: 46-8-166-488; E-mail: britt-marie.sjoberg@molbio.su.se.

1 The abbreviations used are: HEG, homing endonuclease gene; IS, intron insertion site.
only contain the td intron (27, 28). The only phage besides T4 that has been shown to have more than one intron is RB3, which has a td intron and an nrdB intron with an intact, functional HEG (26). Given the high efficiency of homing of the two functional T4 introns (6, 24, 25) and the propensity for horizontal gene transfer that appears to exist between closely related phages (29, 30) and perhaps also globally between different phages (31), it is striking how few phages actually contain introns.

Here, we show that the previously uncharacterized T-even-like phage isolate, U5, contains all three known T-even introns. This gives a total of 11 identified introns (6 td, 3 nrdB, and 2 nrdD) within the T-even-like phages and enables a systematic study of how group I introns have propagated within this phage population. We have used a combination of sequence comparisons of the introns and the intron-containing genes from a number of phages belonging to the T-even-like family as well as mixed infections between phages with and without introns/HEGs to determine the history and potential of intron transfer among these phages. Our results provide evidence for a recent horizontal spread of introns mediated by the cognate HEGs among phages with high sequence similarity at the intron insertion sites. We also show that homing can overcome the effect of genetic exclusion that occurs between some phages and that the presence of optional genes flanking the intron-containing genes can strongly affect the efficiency of homing.

EXPERIMENTAL PROCEDURES

**Bacteria and Phages**—Strains of the original T-even, T2L, T2H, T4D, and T6, were kindly provided by Elisabeth Haggd, Dept. of Genetics, Stockholm University. Strains RB2, RB5, RB27, RB52, RB49, RB69, LZ1, LZ7, U5, and TuA were kindly provided by Karin Carlsson, Dept. of Cell and Molecular Biology, Uppsala University. RB14, RB15, RB23, RB26, RB51, and RB61 are from our stocks, originally a gift from Sean Eddy, and phage LZ2 was kindly provided for this study by Sean Eddy, Dept. of Genetics, Washington University of Medicine. The intron-deleted T4 strain T4LIVS was kindly provided by David Shub, Dept. of Biological Sciences and Center for Molecular Genetics, University at Albany, State University of New York. **E. coli** B, and CR63 were used as host strains for all phage work and were from our stocks.

**PCR Amplification and Sequencing of nrdB, nrdD, and td Genes from T-even-like Phages**—Individual plaques were picked and suspended in 40 μl of water per reaction. 20 μl of this suspension was used as template for PCR. Amplifications were performed using 2.5 units of proofreading Pfu DNA polymerase (Stratagene) for 30 cycles according to the manufacturer’s recommendations. Primers for PCR amplification of td, nrdB, and nrdD genes were based on T4 sequences. T4nrdA5'-GCGACTCAAGATTAATTCGC; T2nrdA5'-CCATATGGTCGACACATTAGACC, T4nrdA5'-GCAAATACACCTTCATTACACCGC (compement: C-terminal end of T4 nrdA gene for nrdA amplification), SL10435 5’-CCATTTTTCGCTTTACATTACAG (compement: T4 nrdB-dena A intergenic region for nrdB amplification); T4frd1 5'-AGATGAGAATTACCAACCTTATGCTCC (complement: C-terminal end of T4 frd gene for frd amplification), T4nrdA4 5’-TACCTTATTTTTTCTTTCATCGTC (complement: C-terminal end of T4 nrdA gene for nrdA amplification); PAT23 5’-CATATGCAAATTCAGAAG (start of T4 nrdD gene for nrdD amplification), PAT23 5’-TCTGTCATAATTTGCACGCTTC (complement: start of T4 nrdD gene for nrdD amplification), T4frd1 5’-AGATGAGAATTACCAACCTTATGCTCC (complement: start of T4 nrdD gene for nrdD amplification), T4frd1 5’-CCAGTGACATTGATTATACACCGC (compement: middle of T4 nrdG gene for nrdG amplification). Primers specific for TuA were: Tula1d1 5’-GACATGACACCTTGTTTGACCGA; TuAd1d2 5’-TGCCTTTTGCAATTTGTATCC; TuAd1d3 5’-CATATGCAAATTCAGAAG (start of T4 nrdD gene for nrdD amplification), Primers used for screening these genes were: T2 42,3A, 5’-GGTTGTTAACCGAGGAGCGG; T2 42,3B, 5’-CAATTAACTTAAAGGAAAC; RB49 wacA, 5’-CATGTGACGACCTTTACCA; RB49 wacB, 5’-GGTTAACCACCCACCC; T4 21,5’-CCAGTGACATTGATTATACACCGC (compement: start of T4 nrdD gene for nrdD amplification), T4 21,5’-CCAGTGACATTGATTATACACCGC (compement: start of T4 nrdD gene for nrdD amplification).

**PCR Amplification and Sequencing of nrdB, nrdD, and td Genes from T-even-like Phages**—A total of 21 T-even-like phage strains, including the original T-even phages T2, T4, and T6, were used for amplification of PCR fragments of the nrdB, nrdD, and td genes (Table 1). Strains were chosen to include all previously known intron-containing T-even-like phages together with a number of T-even-like phages lacking introns.
Horizontal Transfer of Phage Group I Introns

Table I
Intron distribution in T-even like phages

| Strain  | td | nrdB | nrdD | Original isolation                  | Intron reference |
|---------|----|------|------|-------------------------------------|------------------|
| T2H, L  | -  | -    | -    | St. Louis?, feces?                  | (26–28)          |
| T4      | +  | +    | -    | Brooklyn, sewage?                   | (26–28), Genome sequenced |
| T6      | +  | -    | +    | Brooklyn, sewage?                   | (26–28)          |
| RB2     | +  | -    | -    | Huntington, NY, sewage              | (27)             |
| RB3     | +  | +    | -    | Oyster Bay, NY, sewage              | (27)             |
| RB14    | -  | -    | -    | Oyster Bay, NY, sewage              | (27)             |
| RB15    | -  | -    | -    | Flushing, NY, sewage                | (27)             |
| RB23    | -  | -    | -    | Flushing, NY, sewage                | (27)             |
| RB26    | -  | -    | ?    | New York, sewage                    |                  |
| RB27    | -  | -    | -    | Long Island, NY, sewage             | Genome sequenced |
| RB32    | -  | -    | -    | Long Beach, NY, sewage              | (27)             |
| RB49    | -  | -    | -    |genome sequenced                     |                  |
| RB51    | -  | -    | -    | Denver, sewage                      | (27)             |
| RB61    | -  | -    | -    | Denver, sewage                      | (27)             |
| RB69    | -  | -    | -    | Denver Zoo, clouded leopard         | (27)             |
| LZ1     | -  | -    | -    | Uppsala, sewage                     |                  |
| LZ2     | +  | -    | -    | Tubingen, sewage?                   | (27)             |
| LZ7     | -  | -    | -    |                                  |                  |
| U5      | +  | +    | +    |                                  |                  |
| Tula    | +  | -    | -    |                                  |                  |

- *T2 strain designations: H, originally from A. Hershey laboratory; L, originally from S. Luria laboratory. Both T2H and T2L were included in the study since there has been an earlier report about incongruence in intron presence between different T2 isolates (28). However, both isolates were identical over all sequenced regions (including intergenic regions) and are therefore referred to as T2 in this study. Strains in *boldface* were chosen for sequencing.
- **a** T2 strain designations: H, originally from A. Hershey laboratory; L, originally from S. Luria laboratory. Both T2H and T2L were included in the study since there has been an earlier report about incongruence in intron presence between different T2 isolates (28). However, both isolates were identical over all sequenced regions (including intergenic regions) and are therefore referred to as T2 in this study. Strains in *boldface* were chosen for sequencing.
- **b** References indicate previous reports about intron presence.
- **c** Lack of Sequence Variation in Intron HEGs Suggests Recent Horizontal Transfer of Phage Group I Introns
- **d** Phylogenetic Analyses Support Horizontal Spread of Introns among the T-even-like Phages—To obtain further understanding of the history of intron gain/loss among the T-even-like phages we performed independent phylogenetic analyses of the three intron-containing genes and compared them to the distribution of introns. Such analyses have been used earlier to differentiate between horizontal and vertical spread of mobile introns (33–36). Phylogenies were constructed for each of the *td*, *nrdB*, and *nrdD* genes using alignments of exon sequences only. The *td* and *nrdB* sequences generate well supported trees (Fig. 3), whereas the *nrdD* sequences are too similar to yield a reliable tree. All intron-containing genes (marked with an “I” in Fig. 3), except the RB3 and LZ2 *td* genes, are more closely related to at least one intronless gene than to another intron-containing gene and separated by branches with 100% bootstrap support. The scattered distribution of genotypes with introns indicates either gain of introns via horizontal transfer, or multiple independent intron losses. If no horizontal intron transfers have occurred the hypothesis predicts that the ancestral phage had introns that were then lost along the lineages leading to taxa with intronless genes. Once an intron is lost by a phage all descendents of that phage would also lack the intron, and, therefore, genes without introns would be clustered on the phylogenies and the number of intron losses would be minimized. If, on the other hand, horizontal transfers have occurred, the intron occurrences would appear scattered on the phylogenetic trees. The minimum number of losses inferred from the trees in Fig. 3 is 5 for *td* and 4 for *nrdD* and not statistically less than if intron presence/absence is randomized on the trees (p = 0.72 and p = 0.18, respectively, n = 100 randomizations). Thus, the scattered distribution of introns together with the very efficient homing properties of the HEG-containing *td* and *nrdD* introns during T4 infection (Refs. 5, 6, 20, 21, and this study, see below) suggest that the T-even-like introns have spread among the phages via horizontal transfer.

Lack of Sequence Variation in Intron HEGs Suggests Recent Transfer of *td* Introns among American Phages—If the introns and the genes they reside in have had different phylogenetic histories, we would expect trees based on the intron sequences...
to differ from the gene trees in Fig. 3. It is, however, not possible to achieve useful phylogenetic intron trees in the current study, because there are only three identified nrdB introns and the six td introns are too similar (the td introns of the American isolates T4, T6, RB3, and LZ2 are nearly identical, see below) to generate a phylogenetic tree with more than three branches. The lack of sequence variation among the American td introns is clearly seen when the differences in the intron-containing td genes are plotted as the number of varying sites in a 20-nucleotide sliding window (Fig. 4). The differences between T4, T6, RB3, and LZ2, are almost exclusively located to the td exons. There is only one nucleotide difference throughout the whole td intron, including the HEG, among these phages. On the contrary, U5 and especially TuIa show substantially more variation both in non-coding intron sequences and in the HEG, compared with T4. The same low degree of variation as for the American td introns is seen in the nrdB and nrdD intron sequences (see Table II).

Because the catalytic parts of group I introns are under high selective pressure to keep their splicing capacity, they are expected to have a low mutation rate, explaining the low number of differences between cognate introns of different phages. The coding regions of the HEGs however can be presumed to accumulate at least synonymous changes over time. Such variations are clearly present in the td exons of all the phages and to a lesser extent in nrdB and nrdD exons. The most plausible explanation for the lack of variation in the HEGs of the American td introns is that there has been a recent invasion and lateral spread of the td intron among the American phages.

Taken together with the phylogenetic data, this further strengthens the view that the td introns have been spread horizontally among T-even-like phages.

Variations at the Intron Insertion Sites—A prerequisite for intron homing is that the recipient phage contains a homing site and enough sequence similarity in flanking regions to promote efficient homologous recombination. One possible explanation for the limited spread of introns among the T-even-like phages would be if the phages without introns lacked homing sites. Fig. 5 shows the sequence variations found around the intron insertion sites (IS) for the td, nrdB, and nrdD genes from intron-containing and intronless phages. Apart from the intron-associated variations in the nrdD gene, there is very little variation in the region around the IS in both the nrdB and the nrdD genes among the T-even-like phages. In contrast, more variation is found around the IS in the td genes.
Horizontal Transfer of Phage Group I Introns

TABLE II

| Gene | ORF | Exon sequences |
|------|-----|----------------|
| Td   | T4-U5 | 2.8 3.5 22.5 |
|      | T4-Am<sup>a</sup> | 0.4 0 7.3/11.3/8.5 |
|      | U5-Am<sup>a</sup> | 3.2 3.5 22.3/21.3/18.9 |
|      | T4-Tula | 2.0 12.1 29.5 |
|      | U5-Tula | 4.4 13.6 22.8 |
|      | Am<sup>a</sup>-Tula | 2.4 12.1 30.9/30.9/29.1 |
| NrdB | T4-U5 | 3.3 0.3 10.8 |
|      | T4-RB3 | 3.3 0.9 11.2 |
|      | U5-RB3 | 0 0.5 11.4 |
| NrdD | T4-U5 | 0 0.4 2.9 |

<sup>a</sup> Because the td introns and HEGs are identical in the American phages T6, RB3, and LZ2, these are treated together as Am. Exon sequence variations are for T6, RB3, and LZ2, respectively.

Mixed Infections between Intron-containing and Intron-less Phages—The efficiency of intron homing within the T-even-like group of phages and between more distantly related T4-type phages was studied in pairwise mixed infections at high multiplicity between intron-containing (donor) phages and intron-less (recipient) phages. We used four different recipient phages: the T-even phages T4ΔIVS (which lacks all three introns) and T2 and the more distantly related T4-type phages RB49 and RB69. The expected frequency of a non-homing gene in the progeny of a mixed infection is 50%, whereas genes capable of homing will increase their occurrence approaching 100% for efficient homing. However, the frequency of genetic markers in the progeny can be strongly affected by the still largely uncharacterized processes of general and localized exclusion (49). To discriminate between exclusion and homing we mapped recombination sites around intron insertion sites, and in mixed infections with T2 or RB49 as recipient phages we also screened for intron-independent markers (Figs. 6B and 7B, inverted gray bars).

With T4ΔIVS as recipient phage, the T4 td and nrdD introns, the U5 td and nrdB introns, and the T6 and Tula td introns were found at frequencies indicative of homing (Fig. 6A). Despite the influence of exclusion when T2 was the recipient phage, the T2 results agree with the T4 results with respect to which introns are found together with the T2 marker indicating homing (Fig. 6B). These results agree well with our sequencing data as to which introns have functional/non-functional HEGs with three exceptions. The first one is the U5 nrdB intron, which was found at high frequency in the mixed infections despite the fact that the U5 td endonuclease is expected to be non-functional according to the DNA sequence. All progeny that had acquired the U5 nrdB intron also contained the U5 td intron and displayed U5 genetic markers throughout the region between these genes, indicating that both introns have been transferred in a single recombination event. It is unlikely that the homing capacity of the td I-TevI can mobilize the nrdB intron 4 kb downstream of td. The region between nrdA and nrdB in U5 contains a previously unknown open reading frame of 544 amino acids without homology to any sequence currently in GenBank™ that replaces the putative freestanding homing endonuclease gene mobE in T4 (see “Discussion”). If this potential U5 gene can promote the transfer of the U5 nrdB intron remains to be investigated (work in progress). The second exception is RB3 in which td and nrdB introns were only found at low frequency or not at all in crosses

even between closely related phages and even among the intron-containing phages. In agreement with the co-conversion of short flanking sequences that accompanies intron homing (21, 24, 37) the American intron-containing phages are very similar around the IS. However, Tula differs markedly from the other five intron-containing phages in the td region between the CS and the IS and is instead more similar to the intronless phage RB69 in concordance with the phylogenetic data. The more distantly related phages RB49 and Aeh1 (classified as pseudo-T-even and schizo-T-even phages, respectively (30, 38, 39)) show extensive sequence divergence around the IS in all three genes, and homing of the T-even introns would most likely not be possible to these phages.

Although the I-TevI endonuclease can tolerate quite a lot of sequence variation within its recognition site (6, 40–42) the differences in this region in some of the intronless phages may limit the efficiency of homing of the td intron. In contrast, the low variability among T-even-like phages in the I-TevI and especially the I-TevII recognition sites together with the fact that homing of the T4 nrdD intron to T2 occurs quite efficiently
with T4ΔIVS or T2. The RB3 td intron is identical to the T4 and T6 introns and should therefore be homing-competent. The RB3 nrdB intron is identical to the T4 and T6 introns and should therefore be homing-competent. The RB3 td-nrdB region of RB3 is strongly excluded by T4 in mixed infections (26, 43). When we mapped the region between frd and nrdA in intron-containing progeny from the RB3/T4/H9004 IVS cross we found that most (8/10) phages had a recombination point within 550 bp downstream of the td intron insertion point and 4/10 had a recombination point within the upstream td exon, clearly indicating that homing has taken place (Fig. 6C). A comparing screen with progeny from the T6/T4ΔIVS cross (high frequency of td intron presence) showed few isolates (2/10) with recombination sites close to the intron, whereas the rest (8/10) had T6 sequences in the mapped region. Our sequence results show that both RB3 and T6 contain a previously unknown open reading frame in-between the frd and td genes (according to T4 nomenclature we call it td.1). To analyze what effect the presence of td.1 has on td homing, we recrossed some of the chimeric progeny isolates from the first RB3 and T6 crosses with T4/H9004 IVS. Progeny from all recrosses showed increased frequencies of td intron presence compared with the original crosses indicating less T4 exclusion of RB3 alleles and a more efficient homing. However, isolates containing td.1 showed lower homing efficiencies of the td intron than isolates without td.1 (Fig. 6, right part), indicating that the presence of larger differences between the flanking sequences of donor and recipient phages may be involved in the exclusion process or interfere substantially with homing. The third exception is the Tula td intron, that differs markedly in sequence from the other td introns, and was found at low frequency in the T2 cross but at high frequency when crossed with T4ΔIVS (Fig. 6, A and B). No recipient sequences were found close to the intron (between the frd and nrdA genes) in the td-containing progeny. We can therefore not know whether the td intron has been transferred through homing or as part of a larger rearrangement. The sequence of the Tula td gene also differs substantially from the T4 and T2 sequences, and even if the I-TevI endonuclease could cleave the intronless alleles recombination may be inhibited by the low degree of sequence similarity. This would result in the survival only of progeny with larger recombination regions, including the td intron and of original Tula phages.

To determine the ability of the T-even introns to home to more distantly related T4-type phages, we also performed
crosses with RB69 and RB49 as recipient phages. Based on sequence homology, RB69 and TuIa form a slightly more distant group of the T-even-like phages (see Fig. 3), whereas RB49 is even more distant in relation, belonging to the pseudo-T-evens (38). In the RB49 crosses there was no co-localization of any of the introns with the RB49 marker (Fig. 7A). RB49 appears to exclude TuIa, and conversely T4 and U5 appear to efficiently exclude RB49 (gray bars in Fig. 7A). In the RB69 crosses, on the other hand, the TuIa td intron was found at high frequency and screening of sequences flanking the td insertion site confirms the presence of recombination points close to the intron indicative of homing (Figs. 7, B and C). In contrast, when T4 or RB3 was used as donor, we found very few intron-containing progeny of which most had all introns and also displayed donor plaque phenotypes and therefore most likely were original donor phages. However, in one of our four T4/ RB69 crosses we found a small number of progeny that only had inherited the nrdD intron. Screening of flanking restriction markers showed recombination points within the nrdD exons in all progeny examined (4/4) clearly indicating homing (Fig. 7D). The nrdD sequences flanking the intron insertion site are very similar between T4 and RB69, whereas td IS sequences differ more. These differences in the I-TevI recognition site together with the lower degree of similarity in flanking regions may decrease the otherwise highly efficient homing of the td IVS in RB69. RB69 has been reported to strongly exclude the original T-even phages and most of the RB-strains (43) explaining the generally low occurrence of T4 and RB3 sequences in the crosses. Our data show that the homing properties of the nrdD I-TevII can overcome part of this exclusion.

**DISCUSSION**

Bacteriophage T4 appears to be an optimal target for homing endonucleases with its highly efficient recombination machinery and multiple copies of the genome present during a large part of the lifecycle. Its response to superinfection (a second phage infecting an already infected cell) is lysis inhibition, a delay in lysis during which replication continues, and an increasing number of new phage are produced (for review see Ref. 44), a perfect environment for transfer of homing endonucleases from one phage to another. In agreement with this the T4 genome contains, apart from the three intron-encoded HEGs, at least 12 genes with homing endonuclease motifs several of which have been shown to possess endonuclease activity (45–50). Many of these free-standing, putative HEGs are not present in other T-even-like phages (49, 51), perhaps for the same reasons as the introns are absent.

Our results indicate that the introns in the T-even-like phages each share a recent common ancestor that has spread horizontally among the phages most likely via mixed infections. The nearly total lack of sequence variation between the td introns of the American phage isolates clearly suggests recent spread of this intron among these phages, further
strengthening the view that horizontal transmission of introns can be very efficient among phages in Nature. The high degree of similarity between the nrdB intron of U5, T4, and RB3 and between the nrdD intron of U5 and T4 indicates recent exchange of these introns between the three phages despite the geographic separation of their isolation sites. The differing td intron allele of U5 and the lack of an nrdD intron in RB3 suggest that the introns have been acquired in separate transfers and not as multiple homing events during one mixed infection.

How recent are the introns in the T-even-like phages? Even if the introns in most of the phages are highly similar, the presence of distinctly different versions of the td intron in geographically separated phage populations show that these introns can persist and evolve among phages over time. Furthermore, the highly T-even specific nature of the introns with respect to HEG promoters and codon usage (32, 52, 53) indicates that the introns and phages have a history of coexistence. We therefore favor the idea that the introns have resided among the T-even-like phages for a long time but that their distribution goes through major fluctuations. A similar conclusion was implicit from studies of the ω intron within saccharomyces yeast populations (34). Goddard and Burt (34) proposed that recurring horizontal transfers of the ω intron and cognate HEG into new host populations might be the only way for the HEG to stay viable over larger tracts of evolutionary time. Once at fixation in a population, selection for endonuclease function is relaxed, because there are no additional homing sites to transfer to. The HEG will then degenerate and eventually be lost, unless it can spread horizontally to a HEG-less population or evolve to home to a different recognition sequence in a second gene and start a new cycle of spreading into HEG-less individuals of the same population. Our results on the transmission of group I introns within the T-even-like bacteriophages are consistent with the steps in the cycle proposed by Goddard and Burt (outlined in Fig. 8) as follows: (i) The initial stage with an intronless gene containing a recognition site for cleavage by the respective intron homing endonuclease is the most abundant among T-even-like bacteriophages. There is little sequence variation around the intron insertion sites for the nrdB and nrdD introns of the T-even-like phages sequenced here and in many of the phages also for the td intron, indicating that functional homing endonuclease recognition sites most likely are present in many intronless phages. Our mixed infections show that intron homing is possible to sites with enough sequence similarity but with varying efficiency mainly due to the process of local marker exclusion. The phylogenetic data also support that a rapid spread of the td intron has occurred between closely related American T-even-like phages. (ii) The second step in the cycle, degeneration of the HEG, is seen in three of the T-even-like introns, U5 nrdB and nrdD and T4 nrdB. All three HEGs have suffered frameshift deletions that render the proteins non-functional. The low number of additional nucleotide differences between functional and non-functional HEGs suggests that these deletions have occurred recently and shortly after intron acquisition by the phage. This frequently observed loss of HEG function indicates that the HEGs present no selective advantage to the phage and that the dependence upon frequent transfer of the introns to intronless hosts is vital for their survival in phages. Whether the loss of homing function is particularly rapid for certain HEGs in some phages (i.e. I-TevIII in T4 and U5 and I-TevII in U5) due to negative effects on phage viability remains to be investigated. No T-even-like phages have been found with an intron that totally lacks an endonuclease gene, but such introns have been reported in other phages (9, 13, 15). (iii) No
Horizontal Transfer of Phage Group I Introns

Fig. 8. Cycle of intron gain and loss. Adapted from Ref. 34.

Possible limitations to horizontal transfer between phages in Nature, together with our observation that the intron HEGs appear to be rapidly degenerating once inserted into the phage, may explain why introns are not a more general feature of phages.

Acknowledgments—We are grateful to Karin Carlsson, Sean Eddy, David Shub, and Elisabeth Haggard for providing phage strains, to Torsten Eriksson for statistical help on phylogenetics, and to Anthony Poole for discussion and critical comments on the manuscript.

REFERENCES

1. Gott, J. M., Shub, D. A., and Belfort, M. (1986) Cell 47, 81–87
2. Sjoberg, B.-M., Hahne, S., Mathews, C. Z., Mathews, C. K., Rand, K. N., and Gati, M. J. (1986) EMBO J. 5, 2031–2036
3. Yang, F., Ohman, M., Xu, M., Shub, D. A., and Sjoberg, B. M. (1994) J. Biol. Chem 269, 20229–20232
4. Chu, F. K., Maley, G. F., Maley, and Belfort, M. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 3049–3053
5. Quirk, S. M., Bell-Pedersen, D., and Belfort, M. (1989) Cell 56, 455–465
6. Bell-Pedersen, D., Quirk, S., Clyman, J., and Belfort, M. (1990) Nucleic Acids Res. 18, 3761–3770
7. Goodrich-Blair, H., Scarlato, V., Gott, J. M., Xu, M. Q., and Shub, D. A. (1990) Cell 63, 417–424
8. Goodrich-Blair, H., and Shub, D. A. (1994) Nucleic Acids Res. 22, 3715–3721
9. Belfort, D. H., Hove, K. K., and Shub, D. A. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 11669–11673
10. Mikkonen, M., and Alatosavva, T. (1995) Microbiology (UK) 141, 2183–2190
11. van Sinderen, D., Karsens, H., Kok, J., Terpstra, P., Rutters, M. H., Venema, G., and Nauta, A. (1996) Mol. Microbiol. 19, 1345–1355
12. Lazarevic, V., Soldo, B., Bisterhoff, A., Hilbert, H., Mauel, C., and Karamata, D. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 1692–1697
13. Landthaler, M., and Shub, D. A. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 7005–7010
14. Foley, S., Bruttin, A., andBrusson, H. (2000) J. Virol. 74, 611–618
15. Landthaler, M., Begley, U., Lau, N. C., and Shub, D. A. (2002) Nucleic Acids Res. 30, 1933–1943
16. Landthaler, M., and Shub, D. A. (2003) Nucleic Acids Res. 31, 3071–3077
17. Belfort, M., and Roberts, R. J. (1997) Nucleic Acids Res. 25, 3379–3388
18. Collier, B. S., and Stoddard, B. L. (2001) Nucleic Acids Res. 29, 3770–3774
19. Jurica, M. S., and Stoddard, B. L. (1999) Cell Mol. Life Sci. 55, 1304–1326
20. Mueller, J. E., Clyman, J., Huang, Y. J., Parker, M. M., and Belfort, M. (1996) Genes Dev. 10, 351–364
21. Clyman, J., and Belfort, M. (1992) Genes Dev. 6, 1269–1279
22. Parker, M. M., Court, D. A., Preiter, K., and Belfort, M. (1996) Genetics 143, 1057–1068
23. Parker, M. M., Belisle, M., and Belfort, M. (1999) Genetics 153, 1513–1523
24. Bell-Pedersen, D., Quirk, S. M., Aubrey, M., and Belfort, M. (1989) Gene (Amst.) 82, 119–126
25. Chu, F. K., Maley, G., Pedersen-Lane, J., Wang, A. M., and Maley, F. (1990) Nucleic Acids Res. 18, 7005–7010
26. Eddy, S. R., and Gold, L. (1991) J. Mol. Evol. 33, 119–126
27. Eddy, S. (1992) Introns in the T-even Bacteriophages. Thesis, University of Colorado, Boulder
28. Quirk, S. M., Bell-Pedersen, D., Tomaszewski, J., Ruger, W., and Belfort, M. (1989) Nucleic Acids Res. 17, 315–321
29. Tettar, F., Desplate, C., and Krisch, H. M. (1998) J. Mol. Biol. 282, 543–556
30. Tettar, F., Desplata, C., Kutateladze, M., Monod, C., Ackermann, H. W., and Krisch, H. M. (2001) J. Bacteriol. 183, 355–366
31. Hendrix, R. W., Smith, M. C., Biggin, M. R., Ford, M. E., and Hatfull, G. F. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 2192–2197
32. Gott, J. M., Zeeb, A., Bell-Pedersen, D., Ehrenman, K., Belfort, M., and Shub, D. A. (1988) Genes Dev. 2, 1791–1799
33. Holst-Jensen, A., Vaage, M., Schumacher, T., and Johansen, S. (1999) Mol. Biol. Evol. 16, 114–126
34. Godlørd, M. R., and Burt, A. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 7768–7768
35. Rud, S., Korsal, T., and Jakobsen, K. S. (2002) J. Bacteriol. 184, 666–671
36. Busse, I., and Frenzel, A. (2003) Prost 154, 57–69
37. Mueller, J. E., Smith, D., and Belfort, M. (1996) Genes Dev. 10, 2158–2166
38. Monod, C., Repoila, F., Kutateladze, M., Tettar, F., and Krisch, H. M. (1997) J. Mol. Biol. 267, 257–249
39. Desplata, C., Despale, C., Tettar, F., Eleaume, H., and Krisch, H. M. (2002) J. Bacteriol. 184, 2789–2804
40. Bell-Pedersen, D., Quirk, S. M., Bly, M., and Belfort, M. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 7719–7723
41. Chu, F. K., Maley, F., Wang, A. M., Pedersen-Lane, J., and Maley, G. (1991) Nucleic Acids Res. 19, 6863–6869
42. Bryk, M., Quirk, S. M., Muller, J. E., Lozios, N., Lawrence, C., and Belfort, M. (1993) EMBO J. 12, 2141–2149
43. Russell, L. R., and Huxley, R. J. (1974) Genetics 78, 989–1014
44. Abedon, S. T. (1994) in Molecular Biology of Bacteriophage T4 (Karam, J. D., ed) 1st Ed., pp. 397–407, American Society for Microbiology Press, Washington, D. C.
45. Sharma, M., Ellis, R. L., and Hinton, D. (1992) Proc. Natl. Acad. Sci. U. S. A. 89, 6658–6662
46. Sharma, M., and Hinton, D. M. (1994) J. Bacteriol. 176, 6439–6448
47. Kadyrov, F. A., Shlyapnikov, M. G., and Kryukov, V. M. (1997) FEBS Lett. 415, 75–80

Postscript

* L. Sandegren, D. Nord, and B.-M. Sjöberg, unpublished data.
Horizontal Transfer of Phage Group I Introns

48. Belle, A., Landthaler, M., and Shub, D. A. (2002) Genes Dev. 16, 351–362
49. Miller, E. S., Kutter, E., Masig, G., Arisaka, F., Kanisawa, T., and Ruger, W. (2003) Microbiol. Mol. Biol. Rev. 67, 86–156
50. Liu, Q., Belle, A., Shub, D. A., Belfort, M., and Edgell, D. R. (2003) J. Mol. Biol. 334, 13–23
51. Kutter, E., Gachechiladze, K., Poglazov, A., Marusich, E., Shneider, M., Aronsson, P., Napuli, A., Porter, D., and Mesyanzhinov, V. (1995) Virus Genes 11, 283–297
52. Shub, D. A., Gott, J. M., Xu, M. Q., Lang, B. F., Michel, F., Tomaschewski, J., Pedersen-Lane, J., and Belfort, M. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 1151–1155
53. Edgell, D. R., Belfort, M., and Shub, D. A. (2000) J. Bacteriol. 182, 5281–5289
54. Michel, F., and Westhof, E. (1990) J. Mol. Biol. 216, 585–610
55. Bryk, M., Belisle, M., Mueller, J. E., and Belfort, M. (1995) J. Mol. Biol. 247, 197–210
56. Leuza, N., Silva, G. H., and Belfort, M. (1996) J. Mol. Biol. 255, 412–424