THE ORIGIN AND STRUCTURAL FEATURES OF A NEWLY-DETECTED PLASMID IN A DEVIAN NEUROSPORA INTERMEDIA STRAIN

by

YAN XU
B.Sc. Anhui University 1986
M.Sc. The University of Science and Technology of China 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

( Department of Botany )

We accept this as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
March 1998

© YAN XU, 1998
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of \textit{Botany}

The University of British Columbia
Vancouver, Canada

Date \textit{April 7, 1998}
Abstract

During a survey of mitochondrial DNA plasmids in natural populations of *Neurospora intermedia* and *Neurospora crassa*, many new mitochondrial plasmids, both linear and circular, were found. Also, new circular and linear plasmids appeared to arise spontaneously from existing plasmids. A new mitochondrial linear DNA plasmid has been detected in *Neurospora intermedia* Harbin strain 3983M-7.0 which is a derivative of *Neurospora intermedia* Harbin strain 3983M. The origin of this new plasmid was studied and its complete nucleotide sequence was determined. This new DNA element is a mitochondrial linear plasmid. It is not derived from the nuclear or mitochondrial genome. It is related to a linear plasmid in *Neurospora intermedia* Harbin strain 3983. This newly-detected plasmid, named Har-7.0, is 7.0 kilobase pairs (kb) in length. It carries perfect terminal inverted repeats (TIR) of 347 base pairs (bp). Extending inward from the terminal repeats are two long open reading frames with similarity to DNA and RNA polymerases. These are separated by a short intergenic region. The plasmid sequence shows remarkable similarity to that of the senescence-inducing plasmid maranhar, originally described in *Neurospora crassa*. Overall the two plasmids have identical genetic organization and are clearly homologous at the sequence level. ORF1 of this newly-detected 7.0 kb plasmid is 2691 bp in size and ORF2 is 3108 bp. In ORF1 there is just mono-nucleotide (1-nt) substitution and mono-nucleotide insertion relative to marDNA. In ORF2 there are several substitutions (up to 4-nt) and small insertions (up to 4-nt). The distribution of maranhar plasmid in different species, *N. crassa* and *N. intermedia*, could be due to horizontal transfer or gene flow between these two.
## Table of contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>List of tables</td>
<td>v</td>
</tr>
<tr>
<td>List of figures</td>
<td>vi</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td><strong>Neurospora</strong> life cycle</td>
<td>1</td>
</tr>
<tr>
<td>DNA plasmids in eukaryotes</td>
<td>2</td>
</tr>
<tr>
<td>Killer plasmids in <em>Kluyveromyces lactis</em></td>
<td>4</td>
</tr>
<tr>
<td>Plasmids S1 and S2 of <em>Zea mays</em></td>
<td>5</td>
</tr>
<tr>
<td>Plasmids in filamentous fungi</td>
<td>6</td>
</tr>
<tr>
<td>Objectives of these studies</td>
<td>13</td>
</tr>
<tr>
<td>Materials and methods</td>
<td>14</td>
</tr>
<tr>
<td>Strains</td>
<td>14</td>
</tr>
<tr>
<td>Media and growth conditions</td>
<td>14</td>
</tr>
<tr>
<td>DNA isolation</td>
<td>15</td>
</tr>
<tr>
<td>Enzyme digestion and gel electrophoresis</td>
<td>16</td>
</tr>
<tr>
<td>Cloning of the mitochondrial plasmid</td>
<td>16</td>
</tr>
<tr>
<td>Labelling nucleic acid</td>
<td>17</td>
</tr>
<tr>
<td>Probes</td>
<td>17</td>
</tr>
<tr>
<td>Southern blot analysis</td>
<td>17</td>
</tr>
<tr>
<td>Dot-blot hybridization</td>
<td>18</td>
</tr>
</tbody>
</table>
PCR reaction 18
DNA sequencing 19
Computer analysis of DNA sequence 20

Chapter one. The origin of the newly-detected 7.0 kb plasmid of deviant *N. Intermedia*
Harbin strain 3983M-7.0 21

Introduction 21
Results 23
Discussion 25

Chapter two. The structural features of the Har-7.0 plasmid. 43

Introduction 43
Results 45

General sequence organization of the Har-7.0 plasmid DNA. 45
Sequence similarity between Har-7.0 and maranhar. 47

ORF1 48
ORF2 48

Terminal inverted repeats 49
The intergenic region 50
The 5' terminal nucleotides 50

Sequence similarity between Har-7.0 and Har-L 50

Discussion 51

Summary 58

Literature cited 82
List of tables

Chapter one

Table 1 - 1. The plasmid content of strain 3983 and its derivatives.  22
Table 1 - 2. The transmission of maternal mitochondrial plasmids to ascospore progeny from crosses.  30

Chapter two

Table 2 - 1. The marDNA-homologous plasmids with size similar to maranhar in a sample of 19 natural isolates of *N. intermedia*.  44
Table 2 - 2. Size (bp) of domains of maranhar and Har-7.0 plasmid.  46
Table 2 - 3. Quantity of changed residues.  47
Table 2 - 4. Quantity of deleted residues.  47
Table 2 - 5. Quantity of added residues.  48
Table 2 - 6. The changed amino acids in ORF1.  48
Table 2 - 7. The changed amino acids in ORF2.  49
# List of figures

## Chapter one

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1 A</td>
<td>The new plasmid Har-7.0 replacing the original Har-L detected in mitochondria of a culture 3983M-7.0 derived from strain 3983M.</td>
<td>31</td>
</tr>
<tr>
<td>1-1 B</td>
<td>Filter of the gel probed with Har-L plasmid.</td>
<td>31</td>
</tr>
<tr>
<td>1-2</td>
<td>Linearity of the Har-7.0 plasmid in strain 3983M-7.0.</td>
<td>33</td>
</tr>
<tr>
<td>1-3</td>
<td>Southern hybridization of the different parts of Har-7.0 to the three Zhisi plasmids.</td>
<td>35</td>
</tr>
<tr>
<td>1-4</td>
<td>Comparison of restriction maps of the plasmids Har-7.0 and maranhar.</td>
<td>38</td>
</tr>
<tr>
<td>1-5</td>
<td>The restriction map of Har-L plasmid.</td>
<td>40</td>
</tr>
<tr>
<td>1-7</td>
<td>The PCR fragment amplified from DNA in mitochondria of strain 3983M-7.0.</td>
<td>41</td>
</tr>
</tbody>
</table>

## Chapter two

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>The nucleotide sequence of the top strand (5' to 3') of Har-7.0.</td>
<td>60</td>
</tr>
<tr>
<td>2-2</td>
<td>The features of the 7050-bp Har-7.0 plasmid.</td>
<td>47</td>
</tr>
<tr>
<td>2-3</td>
<td>The alignment of the sequences of Har-7.0 and maranhar, Har-7.0 and Har-L.</td>
<td>66</td>
</tr>
<tr>
<td>2-4</td>
<td>The sequence alignment of ORF1 of Har-7.0 and maranhar.</td>
<td>74</td>
</tr>
<tr>
<td>2-5</td>
<td>The sequence alignment of ORF2 of Har-7.0 and maranhar.</td>
<td>76</td>
</tr>
<tr>
<td>2-6</td>
<td>The sequence alignment of TIRs of Har-7.0 and maranhar.</td>
<td>78</td>
</tr>
<tr>
<td>2-7</td>
<td>The sequence alignment of the intergenic regions of Har-7.0 and maranhar.</td>
<td>79</td>
</tr>
<tr>
<td>2-8</td>
<td>The comparison of the conserved motifs in plasmid-encoded RNA polymerases of maranhar and Har-7.0.</td>
<td>81</td>
</tr>
</tbody>
</table>
Introduction

**Neurospora** life cycle

The genus *Neurospora* belongs to the class Ascomycetes of the kingdom Fungi. The ascomycetes are the largest class of fungi and provide most of the species that have been widely used in genetics. *Neurospora* shows a mycelial form of cellular organization in which hyphae have partial septa that delineate hyphal compartments. Within each compartment there are several nuclei. The septa have central pores through which the nuclei and cytoplasm can pass thus uniting the various hyphal compartments into a continuous protoplasmic system. This form of cellular organization is referred to as coenocytic. The life cycles of *N. crassa*, *N. sitophila* and *N. intermedia* involve both sexual and asexual propagation (reviewed by Beadle 1945). These species are all heterothallic and consequently require fusion of opposite mating types in order to complete the life cycle. Mating type is determined by a mating type locus which is located on Linkage Group 1 (Perkins *et al.* 1982). The two mating types, designated $\Delta$ and $\delta$, are determined by codominant alleles. On suitable crossing medium, either mating type is capable of producing female sexual structures, protoperithecia. Protoperithecia consist of coiled filaments of specialized hyphae which become surrounded by a thick layer of hyphae. The coiled filaments are destined to become ascogenous hyphae. From each filament a sexual hypha, the trichogyne, is produced, which grows towards and fuses with a fertilizing cell of the opposite mating type. The male cells may be either vegetative hyphae, or asexual spores called macroconidia or the less abundant microconidia. It has been shown that trichogyne growth and localization of the male fertilizing cell is a chemotactic response.
initiated by the presence of a pheromone released by the male fertilizing cell, whose growth is inhibited by female pheromone (Bistis 1983, 1986). After fusion, the nucleus from the male cell is transferred through the trichogyne into the ascogenous hyphae. The paternal and maternal nuclei undergo a number of synchronous mitotic divisions to form a small mass of dikaryotic ascogenous hyphae. At the same time the protoperithecium enlarges and becomes blackened with melanin and eventually forms the mature fruiting body, the perithecium. Karyogamy eventually occurs in the penultimate hyphal compartments of the ascogenous hyphae. Immediately after karyogamy, meiosis occurs and the four products of meiosis undergo a round of mitosis to give a total of eight nuclei which form the ascospores. At maturity the asci elongate and eject their spores through the ostiole of the perithecium. The ascospores germinate under high temperatures, 60°C, and form mycelia. On vegetative medium, aerial mycelium is formed and conidia are produced through mitosis. The conidia can become airborne and give rise to new colonies that continue the life cycle.

**DNA plasmids in eukaryotes**

Plasmids are small extragenomic DNA molecules that can reproduce inside living cells. They replicate separately from the genome, but some can integrate covalently into the genome and replicate as part of genomic DNA. Plasmids were originally discovered in bacteria, but later analogous molecules were found in eukaryotes. The first plasmid detected in a eukaryote was found in baker's yeast, *Saccharomyces cerevisiae* (Sinclair et al. 1967). This plasmid, named '2μ', turned out to be circular and located within the nucleus. Subsequently many plasmids have been recorded in fungi, few in plants, and none in animals. Virtually all the plasmids discovered so far in filamentous fungi are mitochondrial. There are basic two types: circular and linear plasmids, reviewed by Nargang (1985), Esser et al. (1986), Meinhardt et al. (1990), Fecikova (1992), Kempken (1995) and Griffiths (1995).
All circular plasmids with different sizes except Mauriceville and Varkud probably replicate autonomously by a rolling circular mechanism used in the replication of circular elements in prokaryotes, and some of them have concatameric structures of the basic unit (Maleszka 1992). Linear plasmids are found in different sizes ranging from 7 to 9 kb. They possess terminal inverted repeats with terminal 5' bound proteins, with one exception in which the plasmid has a terminal hairpin loop (Miyashita et al. 1990). They exist as monomers: no multimer of their basic unit has been proven, even through such structures have been suggested. These plasmids replicate autonomously, presumably using their 5' terminal proteins as primers, as do adenovirus and bacteriophage φ29 (Salas 1988, Sakaguchi 1990). So far, no particular function or phenotype can be associated with most of these plasmids. Almost all of these plasmids have been found in mitochondria (Meinhardt et al. 1990), some in chloroplasts (Turmel et al. 1986) and some in cytosols (Stam et al. 1986). Some locations are still unknown, but no linear plasmid has been proven to be within the nucleus.

Plasmids residing in the nucleus are transmitted to sexual progeny from either male or female parents. Plasmids in the cytoplasm are passed to sexual progeny when plasmid-containing strains behave as females. However leakage of paternal cytoplasm has been detected (May and Taylor 1989, Erickson et al. 1989, Yang and Griffiths 1993b).

Generally, fungal plasmids are stable through asexual and sexual reproduction. Furthermore, plasmids are transmissible by heterokaryosis and by incompatible cell contact. (Griffiths et al. 1990, Collins and Saville 1990). It has been demonstrated that some laboratory strains of Neurospora carry suppressor alleles that eliminate the kalilo plasmid (Griffiths et al. 1992). In natural isolates of Neurospora there is considerable variation in suppressive ability (Yang and Griffiths 1993c).
Some plasmids are widely dispersed both within and between species (Yang and Griffiths 1993a, Arganoza et al. 1994). Also plasmid distribution is apparently random (Arganoza et al. 1994) suggesting that plasmids are freely mobile in natural fungal populations.

There is only one case of intergeneric distribution of plasmids in fungi, and that is the kalilo plasmid which is found in Neurospora and Gelasinospora (Wei et al. 1996).

Killer plasmids of *Kluyveromyces lactis*

A wild type strain of the yeast *Kluyveromyces lactis* was found to harbor two linear plasmids: K1 (8.9 kb) and K2 (13.4 kb) (Gunge et al. 1981, Wesoloski et al. 1982a, b). Cells lacking these elements are killed by the plasmid-containing strains. The plasmids are thus the genetic basis of the killer phenotype. It has been found by mutagenesis that K1 encodes the killer toxin, while K2 seems to be responsible for maintenance of the plasmids. Loss of K1 resulted in toxin-sensitive non-killers and K1 has never been seen without K2 (Niwa et al. 1981, Wesoloski et al. 1982c). These plasmids were found to reside in the cytoplasm outside mitochondria (Stam et al. 1986). Transformation of these plasmids into *Saccharomyces cerevisiae* confers the killer phenotype (Gunge et al. 1982, Gunge and Sakaguchi 1981). Plasmids are stably inherited through vegetative propagation only in rho° strains (containing no mitochondrial DNA) of *S. cerevisiae*, suggesting genetic incompatibility between the plasmids and the mtDNA (Gunge and Yamane 1984). Transfer of the plasmids and expression of the killer phenotype has also been possible in *S. kluyveri* and *Candida pseudotropicalis* (Sugisaki et al. 1985). The K1 and K2 plasmids share no homology at the DNA level, even in the terminal repeats which are 202 bp and 184 bp respectively (Sor et al. 1983). Each plasmid has 5' bound terminal proteins, which are different between the two plasmids (Kikuchi et al. 1984). Sequence and transcription analysis revealed that the K1 plasmid encodes four ORFs, whereas ten ORFs have been
identified for K2 (Hishinuma et al. 1984, Stark et al. 1984, Sor and Fukuhara 1985, Tommasino et al. 1988). Two of the ORFs of K1 encode for the toxin (Stark and Boyel 1986), one encodes a resistance factor (Tokunaga et al. 1987), and one most probably encodes a DNA polymerase (Jung et al. 1987). One of the ten K2 ORFs probably encodes a DNA polymerase (Tommasino et al. 1988), and a second ORF encodes an RNA polymerase (Wilson and Meacock 1988). It has been shown that the K1 and K2 plasmids replicate like adenovirus (Fujimura et al. 1988).

Plasmids S1 and S2 of Zea mays

Two linear mitochondrial plasmids S1 and S2 have been associated with cytoplasmic male sterile (cms) S lines of Zea mays (Pring et al. 1977). Each S1 and S2 plasmid has 5' bound terminal proteins and terminal inverted repeats (Kemble and Thompson 1982, Kim et al. 1982). They share identical TIRs, and a common sequence of 1462 bp at one terminus (Paillard et al. 1985). Besides plasmids free in the mitochondria, there are integrated copies in the mtDNA of the cms-S line (Lonsdale et al. 1981, Schardl et al. 1984). Schardl et al. (1984) demonstrated that the plasmids can recombine with homologous regions of mtDNA, leading to linearization of the mtDNA genome. Spontaneous reversion to fertility is normally accompanied by the recirculization of the linearized mtDNA and loss of the plasmids, but small deletions can be observed in the integrated copies of the plasmids (Schardl et al. 1985). However, loss of the plasmids during sterility reversion is under nuclear control; it is not an S-cytoplasm trait (Escote et al. 1985). In some cases, free plasmids are present in cytoplasmic revertants (Escote-Carlson et al. 1988). Consequently these plasmids are not the causative agent of cms, which is further evident from the fact that similar plasmids (R plasmids) exist in fertile lines (Weissinger et al. 1982, Levings et al. 1984). R1 and R2 share sequence homology at the DNA level with the S plasmids, which themselves are considered to have come into existence by recombination between R1 and R2.
Sequence analysis of the S1 plasmid revealed an ORF homologous to a DNA polymerase gene, whereas S2 shares homology with an RNA polymerase gene (Kuzmin and Levchenko 1987, Sederoff et al. 1986, Kuzmin et al. 1988, Oeser 1988).

Plasmids in filamentous fungi

Plasmids are a common feature of filamentous fungi, where they show considerable diversity (Esser et al. 1986, Meinhardt et al. 1990, Yang and Griffiths 1993a, Arganoza et al. 1994). The origin and evolution of eukaryotic plasmids generally has been the subject of much research effort. Many authors (Natvig et al. 1984; Taylor et al. 1985; Rohe et al. 1991; Kempken et al. 1992; Wang and Lambowitz 1993, for example) have considered the origins and phylogenetic relationships of fungal plasmids. Others have considered structural rearrangements of plasmids and their interaction with mitochondrial DNA (Bertrand et al. 1985, Akins et al. 1988, Myers et al. 1989, Court et al. 1991, Nargang et al. 1992, Vierula and Bertrand 1992, Oeser et al. 1993, Yang and Griffiths 1993a, Hermanns and Osiewacz 1994, Hermanns et al. 1995). The linear kalDNA and marDNA cause death of the host (Bertrand et al. 1985, Myers et al. 1989, Court et al. 1991). Some circular plasmids interact with the mitochondrial genome at the transcript level, and reverse transcription results in molecular hybrids (Akins et al. 1988). Some plasmids share a homologous region with mtDNA. For example, the Labelle circular plasmid bears a 1.6 kb region that is also found in mtDNA (Nargang et al. 1992). Most Neurospora strains do not carry the LaBelle plasmid (Nargang et al. 1992, Arganoza et al. 1994), but all mtDNAs of all Neurospora species carry the homologous region. The direction of transfer of this DNA historically, whether from mtDNA to plasmid or from plasmid to mtDNA, is still uncertain. In other filamentous fungi virtually all the natural plasmids discovered have been linear (these are listed in Kempken 1995). However, these show many of the transactions described above for the linear plasmids of Neurospora. For example recombination with mtDNA has been found in Agaricus species.
(Robison et al. 1991), **Claviceps purpurea** (Tudzynski and Esser 1986, Oeser et al. 1993) and **Podospora anserina** (Hermanns and Osiewacz 1992, Hermanns et al. 1995). Therefore previous studies on plasmids in filamentous fungi have shown that they can not only spontaneously generate variability in their own genomes, which is undoubtedly important for their own evolutionary flexibility, but that they can recombine with the mitochondrial genome, and that these types of recombinations seem to have been important in the evolutionary histories of several well-studied fungi.

The plasmids of **Neurospora** are among the best studied in fungi. Worldwide sampling of natural isolates has been performed, and these isolates constitute the basis for plasmid studies in the genus. **Neurospora** plasmids may be linear or circular, and they fall into at least six homology groups (Yang and Griffiths 1993a, Arganoza et al. 1994). Most of these homology groups are distributed across more than one species of **Neurospora** (Natvig et al. 1984, Taylor et al. 1985, Yang and Griffiths 1993a, Arganoza et al. 1994, Marcinko-Kuehn et al. 1994) and even across genera (Wei et al. 1996). Within a homology group there is both structural variation and nucleotide substitution. Homologous circular plasmids can be found in different species and distant geographical locations (Natvig et al. 1984; Nargang 1985; Nargang et al. 1993, LaBelle, and Griffiths and Yang 1995, Harbin-1.). One example of homologous linear plasmids is the kalilo plasmids of **N. intermedia** (Chan et al. 1991), **N. tetrasperma** (Marcinko-Kuehn et al. 1994), and **Gelasinospora** species (Wei et al. 1996). Change in plasmid sequence have been observed in culture. Derivatives of the linear kalilo plasmid (kalDNA) of **Neurospora** have been observed to form spontaneously during routine culturing. Similar derivatives have been observed for the maranhar plasmid and the marDNA-homologous plasmids (Yang 1994). Giant derivatives, similar-sized 'sibling' plasmids, nested short forms (Yang and Griffiths 1993a), and hairpins and deletions bearing only the terminal inverted repeats (Vierula and Bertrand 1992) have all been detected as spontaneous variants arising by unknown molecular mechanisms. This type of generation of new plasmids is from...
parental linear types to new linear types. Also, the derivation of new circular plasmids from parental circular types have been observed. A small circular plasmid has been proved to stem from 0.9 kb segment of Harbin-1 in a vegetative culture of _N. intermedia_ strain 3983 (Griffiths and Yang 1995). Modified plasmids of the circular plasmids Varkud and Mauriceville have been observed repeatedly in other studies (Akins _et al._ 1986, 1989). Surprisingly, two new linear plasmids Har-L and Har-L’ in strain 3983M of _N. intermedia_ have been derived from combination of the circular plasmid Harbin-1 and a marDNA-homologous linear plasmid in strain 3983 of _N. intermedia_. Therefore, mitochondrial plasmids evolve constantly over time.

Only the linear kalilo and maranhar plasmids have a readily observable effect on their hosts. The first fungal phenotype shown to be produced by an extragenomic plasmid is senescence in _Neurospora_ (reviewed by Griffiths 1992). The two linear plasmids kalilo and maranhar aggressively insert into mitochondrial DNA, associating with abnormal mitochondrial physiology and ultimately to death of the culture. Also the circular Mauriceville plasmid of _Neurospora_ occasionally recombines with mitochondrial DNA in a variety of ways leading to senescence (Akins _et al._ 1986). None of the natural plasmids examined to date are derived from the _Neurospora_ mitochondrial genome.

Most commonly the natural plasmids encountered in fungi are of the linear type. The natural circular type seems to have been found almost exclusively in _Neurospora_ (Griffiths _et al._ 1995). The circular plasmids that have been characterized range in size from 0.9 to 5.3 kb. They all exist as a series of one or more monomer units joined in a head-to-tail fashion. Based on hybridization studies, the circular plasmids were placed into one of the homology groups, Mauriceville, LaBelle, Fiji, Java, MB1, VS and Harbin-2 (Natvig _et al._ 1984, Nargang 1985, Saville and Collins 1990, Griffiths and Yang 1995).
The complete DNA sequences of five circular (Mauriceville, Varkud, LaBelle, Fiji and VS) (Akins et al. 1988, Pande et al. 1989, Li and Nargang 1993, Saville and Collins 1990) and two linear, kalilo and maranhar, (Chan et al. 1991, Court and Bertrand 1992) mitochondrial plasmids are known. The overall structures of linear plasmids kalilo and maranhar are typical of most other linear plasmids discovered to date. The general type of structure has been termed an 'invertron' (Sakaguchi 1990). The characteristics are as follows. First, there is a terminal inverted repeat, whose size is characteristic of the individual plasmid. Second, at the terminus at each end there is a protein bound to the 5' nucleotide. Third, starting within the terminal repeats there are two large non-overlapping open reading frames running on opposite strands towards the middle of the plasmid. These reading frames are open only if mitochondrial codon usage is assumed. Smaller ORFs are present in other frames, but these are probably insignificant. The presumptive amino acid sequences of the ORFs suggest in one case a viral-type DNA polymerase, and in the other case an RNA polymerase similar to those of bacteriophages and yeast mitochondria (Chan et al. 1991, Court and Bertrand 1992). These ORFs are both transcribed (Vickery and Griffiths 1993, Court and Bertrand 1993). Fourth, there is an intergenic region between the ORFs. No function has been proposed for this region, although it presumably contains transcription termination signals. In all cases except VS (Saville and Collins 1990), sequencing studies have revealed long ORFs that occupy a large portion of the coding capacity of the plasmids. These ORFs encode polymerases that are presumably involved in replication or transcription of the plasmids. The highly similar Mauriceville and Varkud plasmids both give rise to abundant unit length transcripts that carry the information for expression of the long ORF. These plasmids encode reverse transcriptases that function in the replication of the plasmids (Nargang et al. 1984, Akins et al. 1988, Kuiper and Lambowitz 1988, Wang and Lambowitz 1992 & 1993, Kennell and Lambowitz 1994, Lambowitz et al. 1995). The Mauriceville reverse transcriptase is unusual in that it is capable of initiating cDNA synthesis directly opposite the 3'-terminal nucleotide of the template RNA. It has been suggested that the enzyme may be a primitive
reverse transcriptase related to those that first evolved from an RNA-dependent RNA polymerase (Wang et al. 1992). Curiously, a variable ratio of Varkud transcripts contain a 1.2 kb leader sequence that is derived from the 5' end of the mitochondrial small rRNA. A number of possible mechanisms for the synthesis of this unusual transcript have been discussed (Akins et al. 1989).

The LaBelle and Fiji plasmids also give rise to unit length transcripts but these are not abundantly expressed. The LaBelle plasmid ORF was originally thought to be related to reverse transcriptases, but it has now been shown that the plasmid encodes a DNA-dependent DNA polymerase with motifs characteristic of the B family of DNA polymerases (Pande et al. 1989, Schulte and Lambowitz 1991, Li and Nargang 1993). The Fiji plasmid encodes a DNA polymerase with almost 50% amino acid identity to that of LaBelle. The polymerases from these two plasmids are unusual in that they contain the amino acid motif 'thr-thr-asp' in place of 'asp-thr-asp', which is thought to be important for the activity of the B family polymerases (Li and Nargang 1993). The VS plasmid does not encode a polymerase and is apparently dependent on the Mauriceville / Varkud enzyme for its replication (Saville and Collins 1990). VS RNA possesses a self-splicing activity that may be used to generate monomeric VS RNAs from multimeric RNAs transcribed from multimeric versions of VS DNA.

The genetic organization and nucleotide sequence of Neurospora circular plasmids has led to the suggestion that they are related to mitochondrial introns and mobile genetic elements (Nargang et al. 1984, Lambowitz et al. 1985, Lambowitz 1989). Much of the observed behavior of the different plasmids supports this view. The Mauriceville plasmid was first described by Collins et al. (1981), who established the monomer length and demonstrated the presence of circular concatamers of up to six repeats. They also showed that the plasmid produced an approximately full-length transcript. Nargang et al. (1984) sequenced the
plasmid and found structural features reminiscent of introns. First, the ORF showed a codon usage similar to that of fungal mt DNA introns. Second, there were DNA sequences characteristic of the conserved elements E, P, Q, R, E', and S, which interact to promote splicing in group I mitochondrial DNA introns. They were all in correct position and alignment. This suggested that Mauriceville is an intron progenitor, that it is an excised intron that can replicate independently, or that is has recombined with part of an intron. Michel and Lang (1985) also showed that the Mauriceville ORF contained seven conserved blocks of amino acids that are shared with four different group II introns. These observations, together with the existence of a full-length transcript, gave rise to the idea that the plasmid might be a mobile intron, capable of insertion by reverse transcription, a property also shared by retrotransposons (Lambowitz 1989). To test the idea Akin et al. (1986, 1989) did a series of experiments. Their findings suggest insertion through an RNA intermediate. It was also noted that the splice junctions were not those expected for either group I or group II introns, so the plasmid inserts were not acting as introns. The distribution of plasmids may be explained in part by the ability of plasmids to transfer between strains during unstable vegetative fusions (Collins and Saville 1990, Griffiths et al. 1990).

Linear maranhar DNA was originally found in the several field-collected strains of *N. crassa* from India which are prone to precocious senescence and death (Court et al. 1991). In these isolates, senescence is induced by the integration of the 7.0 kb linear mitochondrial plasmids (marDNA) into the mitochondrial genomes (Court et al. 1991). These isolates were named as the Maranhar strains of *N. crassa*. The maranhar is a Sanskrit word meaning moribund. The marDNA is not derived from the mtDNA nor the nuclear genome of Neurospora. The nucleotide sequence of marDNA was determined by Court and Bertrand (1992). The plasmid is 7052-base pairs in length and has perfect terminal inverted repeats of 349 bp. Each DNA strand contains a long open reading frame using the Neurospora mitochondrial genetic code (Heckman et al. 1980). It begins within the TIR and extends toward the center of the plasmid.
ORF-1 codes for a single-subunit RNA polymerase. The ORF-2 product may be a B-type DNA polymerase. A separate coding sequence for the terminal protein could not be identified; however, the DNA polymerase of maranhar has an amino-terminal extension which probably comprises at least the part of the terminal protein of the maranhar plasmid. The maranhar plasmid can integrate into mtDNA by a mechanism that generates the full-length plasmid insertion sequences which invariably are flanked by very long inverted repeats of mtDNA (Court et al. 1991).

The study of fungal plasmids can uncover a wealth of diverse molecular processes that are relevant not only to the plasmids themselves but also to the properties of DNA in general. Plasmids represent easily studied examples of a mysterious class of DNA known as parasitic or selfish DNA, a type of DNA that seems to exist only for the purpose of existing. This class of DNA includes plasmids, introns, transposons, and viruses. Analysis of plasmids has shown possible areas of connection between these different types, leading to new insights into their evolution. Most DNA—even genomic DNA—could be considered selfish, so the study of the plasmids can provide clues about primitive genomes.
Objectives of these studies

Originally, the studies were focused on the origin of a newly-detected linear mitochondrial plasmid Har-7.0 in a deviant strain 3983M-7.0 of *N. intermedia*. Such studies might provide some knowledge about the diversification and evolution of fungal plasmids. The whole sequence analysis of this new plasmid may provide not only functional analysis for itself but also functional comparison with related plasmids.

Incentives for this studies were three-fold.

First, the general structure of this plasmid was of interest. Is it a linear plasmid? Does it have 5’ terminal proteins? Is it a mitochondrial plasmid?

Second, the origin of this plasmid was of interest. Was it derived from the Har-L in strain 3983M or other plasmids? Is it related to any other plasmid in *Neurospora*?

Third, the structural features of this plasmid were of interest, since the restriction map of this plasmid is similar to that of maranhar in Maranhar strain of *N. crassa*. Maranhar is a senescence-causing plasmid: what is the sequence difference between these two plasmids and is it related to its function?
Materials and methods

Strains

*Escherichia coli* strain DH5α was bought from Bethesda Research Laboratories, BRL. The natural isolate Harbin strain 3983 of *N. intermedia* was obtained from the Fungal Genetics Stock Center (FGSC stock number 3983), Department of Microbiology, University of Kansas Medical School, Kansas City, Kansas. It was collected from Harbin, China. Its mating type is a.

Media and growth conditions

A single colony of *Escherichia coli* DH5α strain was inoculated in 3-5 mL LB broth and shaken at 50 rpm for 16 hours. Cultures were harvested by spinning for 20 seconds in 1ml eppendorf tubes.

Vegetative culturing of *Neurospora* was performed exclusively on Vogel's minimal medium containing 2% glucose (Vogel, 1956) at room temperature. Serial subcultures were made in 10 x 75 mm tubes. Subculturing was performed as described by Griffiths and Bertrand (1984).

For the growth of mycelium for nucleic acid isolation, conidia were prepared by suspending in 8-10 ml of sterile distilled water, then this suspension was added to liquid Vogel's medium.
(approximately 10⁶ conidia/ml). The liquid Vogel's medium was inoculated and shaken at 200 rpm for 16-40 hours. Cultures were harvested by suction filtration through Whatman #1 filters in Buchner funnels.

DNA Isolation

Plasmid DNA in *Escherichia coli* strain DH5α was isolated according to the method modified from Greene (1988) and Bimboim (1983). DNA from mitochondria was isolated according to the small scale mtDNA method of Myers (1988), with addition of a proteinase K digestion prior to phenol/chloroform precipitation of protein. All procedures were carried out at 0-4°C, unless otherwise noted. One hundred and fifty microlitres of liquid culture was harvested and stored on ice until needed. The mycelial pellet was ground with a half volume of acid-washed sand (BHL) and suspended in 30 ml DNA isolation buffer (44 mM sucrose, 50 mM Tris-HCl, pH 7.6, 1 mM EDTA). The suspension was centrifuged at 2 krpm for 5 minutes in an SS-34 rotor. The supernatant was transferred to a tube then centrifuged for 15 minutes at 15 krpm in an SS-34 rotor to pellet the mitochondria. The mitochondrial pellet was suspended in 3.5 ml 70% sucrose in T₁₀E₁ (10 mM Tris-HCl, pH 7.6, 1 mM EDTA), and layered with 1 ml 44% sucrose in T₁₀E₁. The flotation gradients were centrifuged at 45 krpm for 1 hour in an SW50.1 rotor. Mitochondria were collected from the interface beneath the 44% sucrose step gradient layer, and diluted to 3 ml with 2 ml T₂₀₀E₁ (200 mM Tris-HCl, pH 7.6, 1 mM EDTA) in microfuge tubes. The tubes were centrifuged at 10 krpm for 10 minutes to pellet the mitochondria. The mitochondrial pellets were pooled and resuspended in 350 ul of a solution of T₂₀₀E₁, and 40 ul of 20% SDS was added to lyse the mitochondria, followed by the addition of 20 ul of a 10 mg/ml solution of proteinase K (BRL). DNA solution was incubated at 37°C overnight. Protein was extracted by the addition of a half volume of Tris-HCl saturated phenol, and a half volume of chloroform/isoamyl alcohol (24:1). Tubes were mixed by inversion and centrifuged for 15 minutes at 10 krpm. The aqueous phase was
collected and this step was repeated once. Nucleic acids were precipitated from the aqueous phase by the addition of 2.5 volumes of ethanol containing 200 mM ammonium acetate, and incubated overnight at -20°C. The nucleic acids were pelleted and washed with 70% ethanol. The mtDNA was resuspended in 100-150 ul T_{10}E_{1} (pH 8.0), with the addition of 5 ul of 20 mg/ml RNase A (SIGMA). RNA was digested at 55°C for an hour. DNA concentration was determined by UV absorption at 260 nm wavelength. Typically, the yield was 10-20 ug of mtDNA per 150 ml of liquid culture.

**Enzyme digestion and gel electrophoresis**

Restriction enzyme (BRL) digestion of DNA was as described by the manufacturer. Restriction digestion of 2 ug DNA was carried out for an hour at 37°C. Digestion of 2 ug DNA with Lambda exonuclease (BRL) was performed by incubating for an hour at 37°C in a solution of 67mM glycine KOH, pH 9.4, 2.5 mM MgCl₂, 50 ug/ml BSA and 10 units of Lambda exonuclease. Digestion of 2 ug DNA with exonuclease III (BRL) was for an hour at 37°C in a solution of 50 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 5 mM DTT, 50 ug/ml BSA and 50 units of exonuclease III. 3.5 ul of 6x loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 40% sucrose in H₂O) was added to the digestion reaction. Samples were loaded into wells of 0.8% agarose gels and separated by size at 40-60 volts for 16 hours. The buffer for gel electrophoresis was 1x TAE (40 mM Tris acetate, pH 7.6, 2 mM EDTA). Gels were stained with ethidium bromide (0.5 ug/ml) for 20-30 minutes and rinsed in water for 5-10 minutes, then photographed (using Polaroid 57 film) under short wave UV illumination.

**Cloning of the mitochondrial plasmid**

Vector pUC18 (BRL) was used for cloning according to the supplier’s instructions. EcoRI, Hind III, Xba I and Pst I were used for digestion of the vector and Harbin strain 3983M-7.0
mtDNA including Har-7.0 plasmid separately. A restriction endonuclease digestion of the vector was ligated to the same restriction endonuclease digestion of mtDNA. The recombinant bacterial plasmids were transformed into Escherichia coli DH5α competent cells (BRL) according to the supplier's instruction.

Labeling nucleic acid

Oligolabelling was carried out according to the protocol from Pharmacia. At total of 25-50 ng of DNA to be labelled was used in 36 ul of dH2O. The DNA solution was denatured by heating for 2-3 minutes in a water bath at 95-100°C, then placing on ice for 2 minutes. 10 ul of reagent mix (dNTP) and 1 ul (10 units) of Klenow fragment was added. 3 ul (30 uCi) of 32P-dCTP (3000 Ci/mmol) was added and mixed. The labelling reaction was carried out at room temperature for 1-16 hours. Usually 1/3 to 1/2 of the 32P was incorporated into the labelled DNA.

Probes

The pUC18 clones of Har-7.0 1.2 kb and 4.5 kb Hind III (BRL) fragments, 2.3 kb and 3.0 kb EcoR I (BRL) fragments and 1.9 kb Xba I (BRL) (see Har-7.0 map in Fig.1-4) were extracted from Escherichia coli DH5α and linearized by Kpn I (BRL) for use as a probe.

Southern blot analysis

Southern blot analysis was performed essentially as described by Southern (1975). DNA separated by gel electrophoresis was denatured for 30-45 minutes in an alkaline solution of 0.5 N NaOH, 1.5 M NaCl, and then neutralized for 45 minutes in 1 M Tris-HCl, 3.0 M NaCl. DNA was transferred to Hybond filter (BRL) with 2x SSC (1x = 0.15 M NaCl, 10 mM...
sodium citrate, pH 7.0) for 24 hours. After transfer, the filters were baked at 80-90°C for 2-3 hours.

DNA fragments were detected by hybridization to 32-P-labelled probes. Filters were pre-hybridized for 2-16 hours at 42°C with a hybridization solution of 40% deionized formamide, 1% SDS, 1x Denhardt's solution (100x = 2% BSA, 2% PVP, 2% ficoll), 1 M NaCl, and 0.4 mg/ml denatured herring sperm DNA (Bertrand 1985). Hybridizations were carried out in the same solution with the addition of labelled probe to 10^6 cpm/ml. Hybridization was for 16-24 hours at 42°C. After hybridization, filters were washed in 2x SSC at room temperature for 5 minutes, twice in 2x SSC, 0.5% SDS at 65°C for an hour. After air drying, blots were wrapped in Saran Wrap and exposed to Kodak X-Omat RP film for the appropriate time.

**Dot-blot hybridization**

A single colony of *Escherichia coli* DH5α strain was inoculated in 1 ml LB broth and was incubated for 16 hours. The culture was harvested and suspended in a solution of 0.2N NaOH and 1% SDS. The suspension was transferred to hybond filter (BRL) through a dot-blot apparatus (BioRad). The filter was saturated with an alkaline solution of 0.5 N NaOH and 1.5 M NaCl for 5 minutes at room temperature to denature DNA, and then neutralized for 5 minutes in a solution of 1M Tris-HCl pH 8.0 and 3.0 NaCl. After air drying, the filter was baked at 80-90°C for two hours. The subsequent procedure was the same as for the Southern blot hybridization.

**PCR reaction**

PCR amplification was carried out using the PerKin-Elmer Cetus DNA Thermal Cycler and PCR Reagent Kit (BRL). DNA from mitochondria of strain 3983M was used as template for
PCR reaction. Within 0.5 ml PCR tube the appropriate amount of template DNA, 10 ul 10X reaction buffer, 5 ul primer 1: 5’ AAGAATTAAGCGGGAAAG 3’ complementing primer binding sites, 3’ TTCTTAATTCGCCCTTTTC 5’, at the 3’ ends of the target sequence on the 3’ to 5’ strand (20 pmol/ul), 5 ul primer 2: 5’ TAAAGCCGAGTCAGGATGA 3’ complementing primer binding sites, 3’ ATTTCGGCTCAGTCCTACT 5’ at the 3’ ends of the target sequence on the 5’ to 3’ strand (20 pmol/ul), 10 ul 2mM mix, 0.5 ul Tag DNA polymerase (5 U/ul) and appropriate amount of ddH$_2$O were added to make total volume of 100 ul. The above mix in the 0.5 ml PCR tube was incubated in the DNA Thermal Cycler using the following parameters:

Thermocycle file:

Segment 1: 94°C 30 sec.

Segment 2: 47.5°C 30 sec.

Segment 3: 72°C 30 sec.

Total 30 cycles.

Soak file: 4°C.

After the PCR amplification the reaction was run in agarose gel (1.5%) for 1.5 hours at 80 V to see if there was any expected PCR fragment present under UV light in agarose gel.

DNA sequencing

DNA sequence was determined using an automated sequencing system (ABI 373A DNA Sequencer). The Taq DyeDeoxy™ terminator Cycle Sequencing Kit (Applied Biosystems) was used for sequencing reactions. Universal primers were used for sequencing the recombinant plasmids containing inserted fragments: the 1.2 kb and 4.5 kb Hind III fragments, 2.3 kb and 3.0 kb EcoR I fragments and 1.9 kb Xba I fragment of Har-7.0 plasmid. The new primers were designed as primers for the remainder of the cloned
fragments. When the sequence of all cloned fragments was determined, the sequences close to the 3' and 5' terminal regions of the whole clone fragments were employed to design primers that could be used to prime polymerization outwards towards the Har-7.0 plasmid ends. Total DNA from mitochondria (mtDNA and plasmid DNA) was purified and used as a template for sequencing both termini. From the first round of such sequencing, secondary primers were designed from the 3' terminal regions of the sequence. This outward walking process was repeated until no more sequence could be obtained, and this was presumed to be the terminus of the plasmid. All segments were sequenced at least twice and conflicts were resolved by further sequencing. Standard DNA fragments were included in every sequencing run and results from these showed that the sequencing procedure was 98-99% accurate.

**Computer analysis**

The DNA sequence was analyzed using Assembly LIGN, MacVector and BLASTN (Altschul et al. 1990) computer programs.
Chapter one

The origin of the newly-detected 7.0 kb plasmid of deviant \textit{N. intermedia} Harbin strain 3983M-7.0

Introduction

The study is based on the senescent strain 3983 of \textit{N. intermedia} from Harbin, China. The plasmid content of the senescent Harbin strain 3983 of \textit{N. intermedia} changed in the process of subculturing. Previous studies on this strain (Yang and Griffiths 1993a, b) showed that it contained a variety of plasmids, listed in Table 1-1. The linear 1.4 kb and 4.0 kb plasmids were not evident in the present study; either they are absent or present at levels that cannot be detected by staining with ethidium bromide. There are five prominent plasmids, three linear and two circular. The three prominent linear plasmids in strain 3983 are plainly visible on a stained gel. These plasmids, of 7, 8, and 9 kb, are referred as "Zhisi" plasmids after a Chinese word for dying. Previous studies showed that all three Zhisi plasmids hybridize to a probe consisting of the senescence plasmid maranhar (marDNA; Yang and Griffiths 1993a). The two circular plasmids in strain 3983 were named Harbin-1 and Harbin-2 (Har-1 and Har-2; Yang and Griffiths 1993a). Generally, circular plasmids are not visible on stained gels and appear as complex ladders in Southern hybridizations. In one subculture, labelled 3983M, the three Zhisi plasmids had apparently disappeared and were replaced by one smaller prominent linear plasmid of 5.5 kb, which was named Har-L, and one less prominent linear type Har-L'
(5.3 kb). In addition, a small new 0.9 kb circular plasmid with a different ladder pattern was sometimes found in strain 3983M. Hybridization studies showed the new circular 0.9 kb plasmid comprised part or parts of Har-L plasmid alone and suggested that the new plasmid Har-L in strain 3983M is formed from a combination of linear Zhisi plasmids and the circular Har-1 plasmid present in the mitochondria of strain 3983. Low levels of the Har-1 plasmid are still present in strain 3983M (Griffiths and Yang 1995). The plasmid content of strain 3983M is listed in Table 1-1.

Table 1-1: The plasmid contents of 3983 strain and its derivatives. The linear plasmids are listed by size in kb except the Har-7.0. Har-1, Har-2 and Har-0.9 are circular plasmids.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Plasmids</th>
</tr>
</thead>
<tbody>
<tr>
<td>3983</td>
<td>9.0, 8.0, 7.0, 4.0, 1.4, Har-1, Har-2</td>
</tr>
<tr>
<td>3983M</td>
<td>5.5, 5.3 (low copy#), 4.0, 1.4, Har-2, Har-0.9, Har-1 at low level</td>
</tr>
<tr>
<td>3983M-7.0</td>
<td>Har-7.0, 4.0, 1.4, Har-2, Har-0.9, Har-1 at low level</td>
</tr>
</tbody>
</table>

In one out of 20 conidial isolates of 3983M, labelled 3983M-7.0 the 5.5 kb Har-L and 5.3 kb Har-L' had disappeared and was replaced by one prominent linear plasmid of 7.0 kb, which was named Har-7.0 (Fig.1-1A). The plasmid content of strain 3983M-7.0 is listed in Table 1-1. The origin of this new 7.0 kb linear plasmid was of interest. The derivation of this plasmid might be relevant to the mechanism of plasmid diversification and evolution in general.

This chapter describes the study on the origin and structure of this newly-detected mitochondrial plasmid (Har-7.0) and shows that the Har-7.0 plasmid in this aberrant 3983M-7.0 strain is probably similar to the original 7.0 kb Zhisi plasmid of strain 3983. The copy number of the original 7.0 kb Zhisi plasmid or its derivative fell to undetectable value in the
aberrant 3983M. Then the copy number of this plasmid rose to the normal detectable value in the aberrant strain 3983M-7.0.

Results

Mitochondrial linear plasmids appear as distinct bands in electrophoretic gels. Therefore this new 7.0 kb band was likely to be a linear plasmid. To confirm its linearity and 5'-bound protein, DNA from the mitochondria of strain 3983M-7.0 digested with 5' exonuclease or 3' exonuclease after proteinase K treatment and DNA from the mitochondria of strain 3983M-7.0 extracted without proteinase K treatment or with proteinase K treatment were run in agarose gel (Fig.1-2). A complete digestion of 7.0 kb DNA was observed with 3' exonuclease and the 7.0 kb DNA was still present after 5' exonuclease digestion. Without the proteinase K treatment of the mitochondrial DNA extract there was no clear 7.0 kb DNA band present in the gel and after proteinase K treatment there was clear 7.0 kb DNA band present in the gel. Results indicated that this 7.0 kb DNA element was sensitive to 3' degradation, but resistant to 5' degradation and present after proteinase K treatment. So this new 7.0 kb DNA element is a linear plasmid and has a 5'-bound protein (Sakaguchi 1990).

The Har-7.0 plasmid did not hybridize to nuclear DNA or mitochondrial DNA (data not shown). It hybridized to the Har-L plasmid of strain 3983M very weakly (Fig.1-1B). But it strongly hybridized to the three 7.0 kb, 8.0 kb and 9.0 kb Zhisi plasmids of strain 3983 (Fig.1-3). Based on the results of digestion and hybridizations, the restriction map of this plasmid was constructed (Fig.1-4). Comparing the restriction maps of Har-L (Fig.1-5) and Har-7.0, it is obvious that they do not share common restriction sites.
But the restriction map of this plasmid is very similar to that of maranhar plasmid in Maranhar strain of *N. crassa* (Fig.1-4). Since the restriction maps of three Zhisi plasmids are not known the direct comparison of restriction maps of Har-7.0 and three Zhisi plasmids can not be made. But these three Zhisi plasmids are marDNA-homologous. Therefore, this new plasmid might be related to the three Zhisi plasmids, somehow. An attempt was made to find out if there is any relationship of the Har-7.0 plasmid with the three Zhisi plasmids, which might be possibly present at very low copy number in strain 3983M. To define the regions of three Zhisi plasmids homologous to Har-7.0 plasmid of strain 3983M-7.0, a series of hybridizations was performed. The 1.2 kb and 4.5 kb *Hind* III fragments, 3.0 kb and 2.3 kb *EcoR* I fragments and 2.0 kb *Pst* I fragment of Har-7.0 were used as probes. The Southern hybridizations of the intact DNA from mitochondria of strain 3983 were performed using these probes. The 1.2 kb and 4.5 kb *Hind* III fragments, 3.0 kb and 2.3 kb *EcoR* I fragments and 2.0 kb *Pst* I fragment of Har-7.0 all strongly hybridize to the three Zhisi plasmids of strain 3983 (Fig.1-3). Based on these hybridizations it is known that the different regions of Har-7.0 plasmid are homologous to all three Zhisi plasmids. Therefore, it is not clear whether Har-7.0 is derived in whole or in part from the three Zhisi plasmids or part (s) of those. But this new Har-7.0 plasmid of strain 3983M-7.0 may be the same plasmid as the 7.0 kb Zhisi plasmid of strain 3983. The idea is proposed that the copy number of the 7.0 kb plasmid in strain 3983M might have reduced to undetectable level and then it rose again to the normal copy numbers in strain 3983M-7.0. It is already known that there is no detectable 7.0 kb marDNA-homologous plasmid in mitochondria from strain 3983M using Southern hybridization (Yang and Griffiths 1993a). To confirm the above idea a PCR reaction has to be performed using DNA from mitochondria of strain 3983M.

After the complete sequence of Har-7.0 plasmid was obtained the PCR primer pair only for the Har-7.0 plasmid was designed according to the OLIGO™4.0 (a software system developed for primer selection by Cetus Co.). It was attempted to design the primer pair to amplify the
7.0 kb plasmid probably present in strain 3983M but not any regions of the Har-L plasmid in strain 3983M or any regions of Zhisi plasmids 8.0 kb, 9.0 kb probably present in strain 3983M. The sequences of the 8.0 kb, 9.0 kb plasmids are not known. But these two plasmids contain marDNA-homologous regions (Yang and Griffiths 1993a). The proposed marDNA-homologous regions of the Zhisi 8.0 kb and 9.0 kb plasmids are shown in Fig.1-6 (Griffiths and Yang 1995). Therefore, this primer pair was designed to amplify the Har-7.0 excluding primer pairs for Har-L plasmid and marDNA using OLIGO™4.0 since the sequences of Har-L and marDNA are known. The positive control and negative control for PCR reaction were also carried out. The template DNA for the positive control was DNA from mitochondria of strain 3983M-7.0. Template DNA was not added in the negative control in order to indicate any possible contaminations during the procedure of the PCR reaction. The expected 0.5 kb PCR fragment was found in the PCR reaction using the primer pair (Fig.1-7). This probably indicated that the 7.0 kb plasmid of strain 3983 still exists in strain 3983M at very low copy number undetectable using Southern hybridization. According to this study it might be suggested that this 7.0 kb plasmid probably recovered to normal copy number in strain 3983M-7.0. This turns out that the Har-7.0 plasmid of strain 3983M is probably the same plasmid as the 7.0 kb Zhisi plasmid of strain 3983 or the mutated Zhisi 7.0 kb plasmid since it is not sure if there is any mutation in the 7.0 kb plasmid during the process.

**Discussion**

In this study a 7.0 kb linear plasmid of the *N. intermedia* Harbin strain 3983M-7.0 was detected. This plasmid is not derived from the nuclear or the mitochondrial genome. It is not derived from other plasmids as is the Har-L plasmid of strain 3983M (Griffiths and Yang 1995). It might be similar to the 7.0 kb Zhisi plasmid of strain 3983. The copy number of the 7.0 kb Zhisi plasmid or its derivative was most likely reduced to an undetectable level in
strain 3983M and then increased its copy number to normal levels in strain 3983M-7.0. The derivative strain 3983M-7.0 was detected from one out of 20 conidial isolates of strain 3983M. When the 7.0 kb plasmid in strain 3983M-7.0 was detected, the Har-L plasmid of strain 3983M had disappeared. But the reason that the Har-L plasmid of strain 3983M was lost in strain 3983M-7.0 is unknown. These two plasmids do not show strong cross hybridization in Southern hybridization. Therefore, they do not share high degree of homology at sequence level. Since the plasmid is small, 5.5 kb, and has limited coding sequence, it almost certainly relies upon the host genome for essential functions concerning replication and expression. The loss of Har-L plasmid may be an example of nuclear function affecting the replication and expression of a plasmid.

Mitochondrial plasmid suppressors have been demonstrated for the first time in natural isolates of *N. intermedia* (Yang and Griffiths 1993c). The suppressors are plasmid specific, located in nuclear genomes, and normally show Mendelian segregation during meiosis. The plasmid suppressor could either eliminate plasmids or reduce copy number to barely detectable levels. Nuclear heterogeneity in natural isolate strain 3983 of *N. intermedia* has been observed (Yang 1994). Strain 3983, 3983M and its derivative 3983M-7.0 might have different nuclear genotypes regarding their support of mitochondrial plasmids. The nuclei with suppressor mutations, Har-L plasmid suppressor, might be present together with normal nuclei in 3983M heterokaryotic mycelia. Somatic segregation of these two types of nuclei results in different support of mitochondrial plasmids, without the Har-L plasmid or with the Har-L plasmid. These nuclei in mycelia could result from spontaneous mutations during their vegetative propagation. This kind of somatic segregation was also proposed to explain the rejuvenated growth of a senescent *Neurospora* strain (Yang 1994). Preliminary genetic studies of the new senescent strains suggest that senescence is probably due to abnormal nuclei with lethal mutations present together with normal nuclei in heterokaryotic mycelia (Yang 1994). Somatic segregation of these two types of nuclei results in both normal growth
and vegetative death. Successful rescue of the very sick isolates of the senescent *Neurospora*
strain by repeated plating and selecting of conidia suggests nuclear heterogeneity in these
senescent strains (Yang 1994). The loss or low copy number of the three Zhisi plasmids in
strain 3983M could be due to the superior replication of the new plasmid Har-L over the
previous three Zhisi plasmids (Griffiths and Yang 1995). Therefore, the loss of Har-L
plasmid in strain 3983M-7.0 could be due to the Har-L plasmid suppressor in strain 3983M-7.0. Meanwhile, an assumption must be made that the Har-L plasmid suppressor could also
suppress the replication, transcription or translation of the Zhisi 8.0 kb and 9.0 kb plasmids
but not 7.0 kb plasmid. It turns out that the copy number of the Zhisi 7.0 kb plasmid or its
derivative in strain 3983M-7.0 rose to normal copy number without Har-L present. Another
possibility is that the Har-L suppressor might only suppress the expression of the Har-L
plasmid but not the expression of the 8.0 kb and 9.0 kb plasmids, but the expression of Zhisi
8.0 kb and 9.0 kb plasmids in strain 3983 of *N. intermedia* might depend on a function (e.g.
plasmid-encoded DNA polymerase) of the 7.0 kb Zhisi plasmid. The Har-7.0, somehow,
might be mutated in strain 3983M-7.0 relative to the 7.0 kb Zhisi plasmid in strain 3983 and
can not support the expression of the 8.0 kb and 9.0 kb Zhisi plasmids any more. The
dependence of one plasmid on another has been seen in the VS plasmid. The VS plasmid
does not encode a polymerase and is apparently dependent on the Mauriceville/Varkud
enzyme for its replication (Saville and Collins 1990). The possibility can’t be ruled out that
the reappearance of 7.0 kb plasmid in strain 3983M-7.0 could be due to the cytoplasmic
segregation (Yang 1991). The mitochondria in 3983M strain might be heterogeneous, some
containing the Har-L, some containing the 7.0 kb plasmid at very low copy number. When
conidia are formed with the mitochondria containing the 7.0 kb plasmid from their parent
cytoplasm mostly containing Har-L, the derivative strain with mitochondria containing the
7.0 kb plasmid without Har-L plasmid forms.
In routine propagation of plasmid-containing strains generally all subcultures show the plasmid in more or less constant amounts. (Senescence plasmids are an exception to this rule.) In the few cases where asexual cells have been examined by isolating single conidial isolates the majority of these isolates shows the plasmids, but occasionally a plasmid-free isolate will be found that might represent some kind of suppressor mutation or cytoplasmic segregation (Yang 1991). The plasmid loss could occur during either sexual reproduction leading to ascospore formation, or asexual reproduction leading to conidiation. For example, in culturing Kalilo strains of Neurospora occasional nonsenescent conidia are produced (Griffiths and Bertrand 1984). Also occasional ascospores lack plasmids, but tests on these have shown no evidence of suppressors (Yang and Griffiths unpublished results), so some type of cytoplasmic segregation is inferred. No studies on regulation of plasmid copy number have been attempted. But the variation of the three Zhisi plasmid copy numbers between different isolates of N. intermedia Harbin strain 3983 has been observed (Yang 1994). The copy number of the 7.0 kb marDNA-homologous plasmid in ascospores from the cross N. intermedia 1940 a x 1766 A appeared to be reduced relative to the copy number of this plasmid in female parent. And also a new 7.0 kb linear plasmid, homologous to the original 9.0 kb plasmid of female strain N. intermedia 3977, arose spontaneously in ascospore progeny from cross 3977 x 3709 (Yang 1994).

During vegetative growth the plasmids in 3983M-7.0 are relatively stable. The transmission of the Har-7.0 plasmid in strain 3983M-7.0 during the sexual cycle is shown in Table 1-2 (Yang and Griffiths, unpublished data). But the stability of the Har-7.0 plasmid in 3983M-7.0 during the sexual cycle is different from that of 7.0 kb plasmid in strain 3983 (Yang 1994, see table 1-2). These different natural isolates of N. intermedia used as paternal parents were from a variety of geographical locations. During the sexual cycle the three Zhisi plasmids either all existed in the progeny or were all lost together. The loss of the three Zhisi plasmids of 3983 strain during the sexual cycle could be caused by plasmid suppression (Yang and
Griffiths 1993c). The loss of Har-7.0 plasmid of 3983M-7.0 in almost all crosses suggests Har-7.0 might be incomplete or defective in some way that makes its passage through the cross impossible. Or it is possible that the transmission of Har-7.0 plasmid through the sexual cycle might rely on the existence of other two Zhisi plasmids. But the marDNA in Maranhar strain of *N. crassa* is maternally inherited (Court 1991). The difference between the inheritance of the Har-7.0 plasmid in the 3983M-7.0 strain of *N. intermedia* and that of marDNA in Maranhar strain of *N. crassa* could be due to the different genetic background of the different parental strains or different structural features between these two plasmids.
Table 1-2. Transmission of maternal mitochondrial plasmids to ascospore

progeny from crosses (Yang and Griffiths, unpublished result).

Inside the bracket are the plasmids which the strain contains.
* label the plasmid homologous to marDNA.
N/A: no cross performed.

<table>
<thead>
<tr>
<th>male parent</th>
<th>female 3983 (9.0*, 8.0*, 7.0*, 4.0, 1.4, Har-1, Har-2)</th>
<th>female 3983M-7.0 (Har-7.0*, 4.0, 1.4, Har-2, Har-0.9, Har-1 at low level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P 605</td>
<td>normal</td>
<td>lost all</td>
</tr>
<tr>
<td>1766</td>
<td>normal</td>
<td>lost all</td>
</tr>
<tr>
<td>1804</td>
<td>normal</td>
<td>lost all</td>
</tr>
<tr>
<td>1808</td>
<td>lost all</td>
<td>N/A</td>
</tr>
<tr>
<td>3350 (7.2*, C)</td>
<td>lost all</td>
<td>N/A</td>
</tr>
<tr>
<td>3697 (Har-1)</td>
<td>normal</td>
<td>lost all but Har-1</td>
</tr>
<tr>
<td>3704 (Han-2, Har-1, C)</td>
<td>lost 1.4</td>
<td>lost all</td>
</tr>
<tr>
<td>3709* (7.5*, Har-1, Han-2, C)</td>
<td>lost 1.4</td>
<td>lost all but Har-1 and Har-7.0 (1/2-1/4 level of normal plasmid)</td>
</tr>
<tr>
<td>3991</td>
<td>lost all but Har-2</td>
<td>lost all</td>
</tr>
</tbody>
</table>
Figure 1-1A The new plasmid Har-7.0 replaced the original Har-L detected in mitochondria of a culture 3983M-7.0 derived from strain 3983M in one out of 20 conidial isolates

A, DNA in the EtBr-stained gel.  
Lane 0, 1 kb ladder.  
Lane 1, 2, 4-20, DNA of strain 3983M.  
Lane 3, DNA of strain 3983M-7.0  

B, filter of the gel probed with the Har-L plasmid
Figure 1-2 Linearity of the Har-7.0 plasmid in strain 3983M-7.0 DNA from strain 3983M-7.0 was extracted without proteinase K treatment or with proteinase K treatment and was digested with 3' or 5' exonuclease after proteinase K treatment.

Lane 1, DNA without proteinase K treatment
Lane 2, undigested DNA with proteinase K treatment
Lane 3, DNA digested with 3' exonuclease after proteinase K treatment
Lane 4, DNA digested with 5' exonuclease after proteinase K treatment
mtDNA -

7.0 kb -
Figure 1-3 The Southern hybridization of the different parts of Har-7.0 to the three Zhisi plasmids DNA from mitochondria of strain 3983

Figure 1-3 (1st):
Lane 3983, DNA from mitochondria of strain 3983
Lane 1.2 kb Hind III insert, the linearized recombinant DNA containing pUC 19 vector and 1.2 kb Hind III fragment of Har-7.0
Lane 4.5 kb Hind III insert, the linearized recombinant DNA containing pUC 19 vector and 4.5 kb Hind III fragment of Har-7.0
Lane 2.3 kb EcoR I insert, the linearized recombinant DNA containing pUC 19 vector and 2.3 kb EcoR I fragment of Har-7.0
Lane 3.0 kb EcoR I insert, the linearized recombinant DNA containing pUC 19 vector and 3.0 kb EcoR I fragment of Har-7.0
Lane 2.0 kb Pst I insert, the linearized recombinant DNA containing pUC 19 vector and 2.0 kb Pst I fragment of Har-7.0

Figure 1-3 (2nd):
A, filter of the gel containing lane 1.2 kb Hind III insert, 3983 and 3983 probed with the cloned 1.2 kb Hind III fragment of Har-7.0
B, filter of the gel containing lane 4.5 kb Hind III insert, 3983 and 3983 probed with the cloned 4.5 kb Hind III fragment of Har-7.0
C, filter of the gel containing lane 2.3 kb EcoR I insert, 3983 and 3983 probed with the cloned 2.3 kb EcoR I fragment of Har-7.0
D, filter of the gel containing lane 3.0 kb EcoR I insert, 3983 and 3983 probed with the cloned 3.0 kb EcoR I fragment of Har-7.0
E, filter of the gel containing lane 2.0 kb Pst I insert, 3983 and 3983 probed with the cloned 2.0 kb Pst I fragment of Har-7.0
Figure 1-4 Comparison of restriction maps of maranhar plasmid (Court et al. 1991) from N. crassa and Har-7.0 plasmid from N. intermedia.
Figure 1-5 Restriction map of Har-L (Griffiths and Yang 1995).

Harbin-L

<table>
<thead>
<tr>
<th>0.15</th>
<th>1.7</th>
<th>2.3</th>
<th>1.2</th>
<th>0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.0</td>
<td></td>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td>1.4</td>
<td>4.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>4.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td>5.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td>5.1</td>
<td></td>
<td>0.2</td>
</tr>
</tbody>
</table>

1 kb

Bgl II
EcoR I
EcoR V
Hind III
Kpn I
Pst I
Xba I
Figure 1-6 The PCR fragment amplified from DNA in mitochondria of strain 3983M.

Lane 0, 1 kb ladder
Lane 1, 2, the PCR fragments amplified from DNA in mitochondria of strain 3983M-7.0 as a positive control
Lane 3, 4, the PCR fragments amplified from DNA in mitochondria of strain 3983M
Lane 5, negative control
Chapter two

The structural features of Har-7.0 plasmid

Introduction

The sequences of the Har-L and marDNA are already known (Griffiths and Yang 1995, Court et al. 1992). The interest in the structure of the newly-detected Har-7.0 plasmid stems from the finding that it weakly hybridized to the Har-L plasmid in the Harbin strain 3983M of N. intermedia in Southern hybridizations and the restriction map of Har-7.0 is similar to that of the marDNA in the Maranhar strain of N. crassa (Chapter 1 & Court et al. 1992). The Harbin strain 3983M-7.0 of N. intermedia and the Maranhar strain of N. crassa are both senescent strains and contain similar linear plasmids. Furthermore, marDNA in Maranhar strains of N. crassa integrates covalently into the mitochondrial DNA as the form IS-marDNA (Court et al. 1991). The inserted form of mtDNA rapidly predominates over the wild type mtDNA, resulting in abnormal mitochondrial function, senescence and death (reviewed by Griffiths 1992). But preliminary study showed that Har-7.0 plasmid seemed not to integrate into mtDNA (X. Yang, unpublished data). So the structure of the Har-7.0 plasmid was of interest. The sequence of this plasmid may provide some clue about the origin of this plasmid. And the sequence comparison between this plasmid and maranhar may have some explanation on the different functions of these two plasmids.
MarDNA hybridized with eleven linear plasmids with sizes similar to marDNA in a sample of 19 natural isolates of *N. intermedia* (Yang and Griffiths 1993a). The marDNA was originally found in *N. crassa*, but these marDNA-homologous plasmids are all in *N. intermedia*. These plasmid species are as follows (Yang and Griffiths 1993a):

Table 2-1. The marDNA-homologous plasmids with size similar to maranhar in a sample of 19 natural isolates of *N. intermedia*

<table>
<thead>
<tr>
<th>Strain#</th>
<th>Mar-homologous plasmid</th>
<th>N/S</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1940</td>
<td>7.0 kb</td>
<td>S</td>
<td>LaBelle, USA</td>
</tr>
<tr>
<td>3336</td>
<td>7.0 kb</td>
<td>S</td>
<td>Monte Alegre, Brazil</td>
</tr>
<tr>
<td>3983</td>
<td>7.0 kb, 8 kb and 9 kb</td>
<td>S</td>
<td>Harbin, China</td>
</tr>
<tr>
<td>3983M</td>
<td>Har-L, Har-L2</td>
<td>S</td>
<td>3983 Derivative</td>
</tr>
<tr>
<td>3983M-7.0</td>
<td>Har-7.0</td>
<td>S</td>
<td>3983M Derivative</td>
</tr>
<tr>
<td>3986</td>
<td>7.2 kb and 7.0 kb</td>
<td>N</td>
<td>Shengyang, China</td>
</tr>
<tr>
<td>4853</td>
<td>7.0 kb</td>
<td>S</td>
<td>Wau, New Guinea</td>
</tr>
<tr>
<td>P27</td>
<td>8.6 kb</td>
<td>N</td>
<td>Manila, Philippines</td>
</tr>
<tr>
<td>3350</td>
<td>7.2 kb</td>
<td>N</td>
<td>Piracunuga, Brazil</td>
</tr>
<tr>
<td>3709</td>
<td>7.5 kb</td>
<td>N</td>
<td>Salinas, Puerto</td>
</tr>
</tbody>
</table>

N= nonsenescent; S= senescent

The marDNA or the marDNA-homologous plasmid or part of it are found in both *N. crassa* and *N. intermedia* and in diverse geographical locations. These hybridizations do not show whether the marDNA-homologous plasmids are substantially identical or share only some homologous regions. Based on previous studies (Yang 1994), these marDNA-homologous...
plasmids do not possess the insertion activities of the marDNA, because no marDNA-homologous fragment was found in the mitochondrial genome of strains carrying these plasmids, and there was no substantial change of DNA of the strains. Therefore the senescence of these strains is probably not associated with the insertion of these marDNA-homologous plasmids like the senescence of the Maranhar strains.

The geographical distribution of the mar-homologous plasmids suggests a relatively ancient origin. The interspecific distribution indicates either the plasmid could migrate through gene flow since *N. crassa* and *N. intermedia* will cross or the plasmid is in some respect infectious. Horizontal transfer of fungal DNA is important in both process and study of fungal evolution.

Two questions have arisen. The first is the degree of homology between the Har-L and the Har-7.0 and between the Har-7.0 and the marDNA. The second is the structural difference between the Har-7.0 and the marDNA. Since the *N. intermedia* and *N. crassa* both belong to the genus *Neurospora*, the full DNA sequence of the Har-7.0 plasmid would provide an interesting evolutionary comparison with the marDNA sequence. Therefore the plasmid was fully sequenced.

**Results**

**General sequence organization of Har-7.0 plasmid DNA**

The complete nucleotide sequence of the 5' to 3' strand of the Har-7.0 plasmid is presented in Fig. 2-1. Only this strand was sequenced. The Har-7.0 plasmid sequence shows remarkable similarity to that of the marDNA in the Maranhar strain of *N. crassa* except some substitutions (up to 5-nt), small insertions (up to 6-nt) and small deletion (up to 5-nt) relative to marDNA. The sizes of the various regions of the plasmid are shown in Table 2-2, together
with comparison with marDNA. The plasmid is 7050 bp in length, 2 bp shorter than marDNA (7052 bp), explaining the same size of Har-7.0 and maranhar in agarose gel. The Har-7.0 DNA sequence shows perfect terminal inverted repeats of 347 bp. The repeats run from nucleotide position 1 to 347 and 6704 to 7050 (This numbering uses the same plasmid orientation as that used by Court et al. 1992).

Table 2-2 Size (bp) of domains of maranhar and Har-7.0 plasmids

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Length</th>
<th>TIR&lt;sup&gt;a&lt;/sup&gt;</th>
<th>ORF1&lt;sup&gt;b&lt;/sup&gt;</th>
<th>IGS&lt;sup&gt;c&lt;/sup&gt;</th>
<th>ORF2&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Har-7.0</td>
<td>7050</td>
<td>347</td>
<td>2691</td>
<td>669</td>
<td>3108</td>
</tr>
<tr>
<td>marDNA</td>
<td>7052</td>
<td>349</td>
<td>2691</td>
<td>716</td>
<td>3069</td>
</tr>
</tbody>
</table>

<sup>a</sup>TIR, terminal inverted repeat  
<sup>b</sup>ORF, open reading frame  
<sup>c</sup>IGS, intergenic sequence

There are two large non-overlapping open reading frames, on opposite strands and running from just inside the terminal repeats towards the center. ORF1 runs from position 292 to 2982, and ORF2 from 3652 to 6759. The ORFs overlap with the terminal repeats in both plasmids, as shown in Fig. 2-1. Their sizes and orientation are similar to those of marDNA. The general sequence organization of the Har-7.0 plasmid is typical of fungal linear plasmid 'invertrons' (Sakaguchi 1990). The features of the plasmid are presented in Fig. 2-2 below.
Fig. 2-2. The features of the 7050-bp Har-7.0 plasmid. The plasmid and its TIRs are represented by thick lines and hatched regions, respectively. Open circles indicate the terminal proteins, which are bound to the 5' nucleotides of the linear plasmid DNA. The locations and orientations of the two large ORFs are indicated by open arrows.

**Sequence similarity between Har-7.0 and maranhar**

Overall, the Har-7.0 plasmid has an AT content of 72.5%, which is similar to that of marDNA (72.3%, Court and Bertrand 1992) and kalDNA (70%, Chan *et al.* 1991), through the TIRs are less AT-rich (60.1%) like that of marDNA (60.5%). An alignment is shown in Fig. 2-3. A high level of nucleotide similarity is seen. The quantity of changed, deleted and added residues relative to marDNA are shown in Table 2-3, 2-4 and 2-5.

**Table 2-3. Quantity of changed residues.**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>17</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>16</td>
<td>11</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2-4. Quantity of deleted residues.**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
Table 2-5. Quantity of added residues.

<table>
<thead>
<tr>
<th>A</th>
<th>T</th>
<th>C</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

ORF1

ORF1 has 2691 bp, the same length as the equivalent ORF in marDNA. ORF1 encodes a presumptive protein of 896 amino acids, which should encode RNA polymerase by homology to marDNA (Court and Bertrand 1993). The ORF1 sequences of the two plasmids are 98.96% similar at the DNA level and 98.44% similar at the amino acid level. The sequence alignment is shown in Fig. 2-4. All of the amino acid differences are substitutions and no insertions or deletions. The changed amino acids are shown in Table 2-6.

Table 2-6. The changed amino acids in ORF1.

| Ala, A | Leu, L | Trp |
| Arg, R | Lys, K | Asn |
| Asn, N | 3 Lys | Met, M |
| Asp, D | Phe, F | Leu |
| Cys, C | Pro, P | |
| Gln, Q | Ser, S | Thr |
| Glu, E | Thr, T | Ile |
| Gly, G | Trp, W | |
| His, H | Tyr, Y | Ser |
| Ile, I | Thr, Leu | Val, V | Ala |

ORF2

ORF2 has 3108 bp, 42 bp larger than the equivalent ORF of marDNA. The difference is accounted for by the 36 bp of extra coding sequence at the end of the ORF2 and in addition there is a two-amino acid insertion in Har-7.0 DNA. The ORF2 DNA sequence shows a
98.2% similarity to marDNA. The Har-7.0 ORF2 codes for a presumptive DNA polymerase protein of 1035 amino acids, which by comparison with the equivalent marDNA ORF codes for a DNA polymerase. The alignment with the marDNA ORF2 is shown in Fig. 2-5. There is 96.69% amino acid similarity. The changed amino acids are shown in Table 2-7.

Table 2-7. The changed amino acids in ORF2.

<table>
<thead>
<tr>
<th>Ala, A</th>
<th>Thr, Pro</th>
<th>Leu, L</th>
<th>Ile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg, R</td>
<td>Lys, K</td>
<td>3 Asn, Arg</td>
<td></td>
</tr>
<tr>
<td>Asn, N</td>
<td>Thr, Ile, Lys</td>
<td>Met, M</td>
<td>Thr</td>
</tr>
<tr>
<td>Asp, D</td>
<td>Phe, F</td>
<td>Leu, Ile</td>
<td></td>
</tr>
<tr>
<td>Cys, C</td>
<td>Ser</td>
<td>Pro, P</td>
<td></td>
</tr>
<tr>
<td>Gin, Q</td>
<td>Ser, S</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>Glu, E</td>
<td>Gln</td>
<td>Thr, T</td>
<td>Arg</td>
</tr>
<tr>
<td>Gly, G</td>
<td>Trp, W</td>
<td></td>
<td></td>
</tr>
<tr>
<td>His, H</td>
<td>Tyr, Y</td>
<td>His, Asn</td>
<td></td>
</tr>
<tr>
<td>Ile, I</td>
<td>Leu</td>
<td>Val, V</td>
<td>Glu, Leu</td>
</tr>
</tbody>
</table>

Terminal inverted repeats

Sequencing primers based on the sequences of the two outermost edges of the unique region produced identical sequences for the proximal several hundred bases of the termini, showing that most likely the plasmid has terminal inverted repeats. From this point on, sequencing primers could not have distinguished between identical terminal structures. The end cloning of the terminal fragments was not tried due to the previous unsuccess of similar linear plasmid (Wei et al. 1996). Hence it was not possible conclusively to demonstrate the identity of the repeats distal to this point even through the sequences obtained for both ends were the same. However, most likely the termini are identical inverted repeats. The terminal inverted repeats are 347 bp long, 2 bp shorter than those of marDNA, which are 349 bp. The alignment is presented in Fig. 2-6. According to the alignment these two terminal inverted repeats show great similarity except four 1-nt substitutions in the inner and middle parts and
one 1-nt addition and one 5-nt substitution and four 1-nt substitution at the very end (terminal). There is a 96.26% similarity at DNA level.

**The intergenic region**

The intergenic region is 669 bp, 47 bp shorter than the equivalent marDNA region which is 716 bp. An alignment is shown in Fig. 2-7. A high level of nucleotide similarity is seen, with the exception of a small block of nucleotide deletion (42nt) relative to marDNA at one end of the intergenic region next to the ORF2 and some substitutions and small length deletions (up to 3-nt) and 1-nt insertions.

**The 5’ terminal nucleotides**

The 5’ nucleotides of Har-7.0 plasmid are T. This is uncommon to the linear mitochondrial plasmids although the 5’ nucleotides of Gel-KalDNA are also T. The 5’ nucleotides of all other characterized mitochondrial linear plasmids are purines (G: pCIK1, Oeser and Tudzynski 1989; pFSC2 Fusarium solani, Samac and Leong 1989; kalilo, Chan et al. 1991; maranhar, Court 1992. A: pFSC1 of F. solani, Samac and Leong 1989; S-1, Paillard et al. 1985; S-2, Levings and Sederoff 1983; 11.3-kb plasmid of Brassica, Turpen et al. 1987). The 5’ nucleotides of numerous bacteriophages are also purines (A: φ29, φ15, M2Y, GACp-1, Escarmis et al. 1984. G: PRD1, Savilahti and Bamford 1986; Gerendasy and Ito 1987).

**Sequence similarity between Har-7.0 and Har-L**

The alignment of Har-7.0 and Har-L nucleotide sequences has shown that there is a 170 bp common nucleotide sequence region shared by both plasmids. This is shown in Fig. 2-3. This common region starts from the 5’ end to number 170 nucleotide in the central part of terminal
inverted repeats of Har-7.0 plasmid. This result is consistent with the idea that the Har-L plasmid contains marDNA-homologous region (Griffiths and Yang 1995).

Discussion

In the Harbin 3983M-7.0 derivative strain of *N. intermedia*, a 7.0 kb linear plasmid, named Har-7.0, was detected that has same size as marDNA. The studies on full sequencing showed that this plasmid is almost identical to the marDNA originally isolated from several field-collected strains of *N. crassa* (Court 1992). General structural features are very similar to those of marDNA. There are terminal inverted repeats, and two major ORFs starting just inside the inverted repeats and running towards the center of the plasmid. There is a short intergenic region. Both 5' ends are resistant to exonuclease digestion, presumably due to the presence of a covalently bound protein (Vierula *et al.* 1990). The nucleotide similarity is 98.55% over all the region, over 98% in the ORFs. The sequence accuracy is 98%-99% (see materials and methods). So the sequence error is 1%-2%. Therefore, these two plasmids are still pretty similar. But these two are not identical due to the different restriction maps. These features make it virtually certain that the two plasmids are related by descent from a common ancestral plasmid. This kind of similarity is also seen in Gel-kalDNA; kalDNA (Wei *et al.* 1996) and LA-kalDNA (Marceko-Kuehn *et al.* 1994). The similarity of kalDNA and LA-kalDNA was judged by restriction endonuclease site mapping and the sequences of the terminal inverted repeats. But the Har-7.0 plasmid is much more similar to marDNA than Gel-kalDNA is to kalDNA according to the nucleotide sequence alignments. Such a situation contrasts with plasmids that show similar organization but are only functionally related. For example the kalDNA, marDNA, pAls and many other linear plasmids are organized along very similar lines, and show conserved blocks of amino acids in the open reading frames (Meinhardt *et al.* 1990, Hermanns and Osiewacz 1992, Kempken *et al.* 1992), but do not
cross-hybridize at the DNA level. In such plasmids the conserved amino acids suggest relatedness to viral polymerases, but if this is true the relatedness is no longer manifested at the DNA level. This kind of similarity might reflect functional constraints on nucleic acid processing enzymes (analogy) rather than relatedness (homology).

The Har-7.0 and marDNA plasmids from *N. intermedia* and *N. crassa* are almost identical: the variation was due to nucleotide substitutions, small length mutations and sequence errors beyond the scope of the current instrumentation's accuracy (see Materials and Methods). These two virtually identical plasmids were found in different species from different locations. The distribution of the homologous plasmids in nature and the presence of these identical plasmids in different species supported the hypothesis that these plasmids could be transmitted between isolates independently of their host mitochondrial genomes (Taylor et al. 1985). This was also shown for a circular plasmid in *N. crassa* 516 (Roanoke, LA), *N. intermedia* 435 (Fiji) and *N. tetrasperma* 2510 (Hanalei, HA) (Taylor et al. 1985). The similarity of this circular plasmid was judged by DNA-DNA hybridization and restriction endonuclease site mapping. Horizontal transmission can be inferred for the plasmids Har-7.0 and marDNA in different *Neurospora* species. Laboratory experiments have shown that horizontal transmission can occur within a species and between species. Horizontal transmission of kalilo and other plasmids has been demonstrated between compatible and incompatible strains of the same species of *Neurospora* (Griffiths et al. 1990, Collins and Saville 1990, Debets et al. 1994). Transmission of the kalilo and Hanalei-2 plasmids from *N. intermedia* to *N. crassa* has also been shown (Griffiths et al. 1990). Even transmission between fungal genera has been demonstrated. Kempken (1995) has transferred the *Ascobolus immersus* linear plasmid pAI2 to *Podospora anserina*, although the plasmid is unstable and gradually disappears in its new host. There is only one case of intergeneric distribution of plasmids in fungi, and that is the kalilo plasmid, which is found in *Neurospora* and *Gelasinospora* species (Wei et al. 1996). It seems that interspecific and intergeneric
incompatibility can be overcome. The presence of these identical plasmids in different species also could be explained by gene flow since \textit{N. intermedia} and \textit{N. crassa} often cross readily in the laboratory (Perkins \textit{et al.} 1976). But the horizontal transfer mechanism is preferred since some sexually-isolated \textit{Neurospora} still contain many plasmids. Curiously the Har-7.0 and marDNA plasmids were not found at the same location. The Har-7.0 plasmid of \textit{N. intermedia} was found in Harbin, China and marDNA plasmid of \textit{N. crassa} was found in India. There seems to be no sufficient proximity to potentially allow contact and transfer. The distribution of the marDNA plasmid family is disjunct shown in table 2-1. This distribution pattern could reflect incomplete sampling. Alternatively it could be a significant biological pattern reflecting sites of evolutionary origin, selection, host suppressor mutation (Griffiths \textit{et al.} 1992, Yang and Griffiths 1993c) or dispersal. Another possible factor might be historical patterns of trade and settlement that could have influenced distribution patterns by transporting spores on produce such as sugar cane. The mar-homologous plasmids are more commonly encountered than the plasmids hybridizing to kalilo in natural population samples (Yang and Griffiths 1993a). The sample size might be too small to come to any conclusions. The comparison of the virulences of the marDNA-homologous plasmids and kalDNA-homologous plasmids might provide some explanation for this kind distribution. If the marDNA-homologous plasmids are not so virulent as the kalDNA-homologous plasmids both the host and the plasmid might not be destroyed and the general incidence of the plasmids in population could be maintained. Alternatively, horizontal transmission of marDNA-homologous plasmids in \textit{N. intermedia} might be easier transferred than that of kalilo-homologous plasmids. But the marDNA-homologous plasmids are less commonly encountered than the Harbin-1/ LaBelle plasmid in natural population samples (Yang and Griffiths 1993a). The marDNA-homologous plasmids are linear and the Harbin-1/LaBelle plasmid is circular. Hanalei-2 circular mitochondrial plasmids can transfer more efficiently to other strains through compatible and incompatible heterokaryon formation (Griffiths \textit{et al.} 1990, Collin and Saville 1990 and Debets \textit{et al.} 1994). It might be conceivable that the
Harbin-1/LaBelle plasmid can transfer more efficiently than the linear marDNA-homologous plasmids between natural isolates.

Both strands of Har-7.0 plasmid were translated by computer using MacVector programme, in all reading frames, using the Neurospora mitochondrial genetic code (Heckman et al. 1980) in which TGA is not a stop codon but encodes tryptophan. The ORFs of marDNA are known to be transcribed (Court and Bertrand 1993) and the same is true for the ORFs of kalDNA (Vickery and Griffiths 1993). Candidate proteins were noted in Maranhar strains, but no translational products have been demonstrated in either case. The comparison of the conserved motifs in plasmid-encoded RNA polymerases of marDNA and the Har-7.0 is shown in Fig. 2-8. Eight of the nine motifs of the putative RNA polymerase (ORF1) of the Har-7.0 are exactly the same as that of marDNA (Court et al. 1992). There are three amino-acid substitutions within motif IV.

In addition to the overall homology to known DNA polymerase, the ORF2 product of Har-7.0 contains the exactly same highly conserved motifs as that of marDNA which are characteristic of the proof-reading (Bernad et al. 1989) and polymerization (Bernad et al. 1987, Jung et al. 1987) domains of B-type DNA polymerases, although there are 22 changed amino acids in ORF2 (Table 2-7). B-type DNA polymerases; possessing the three consensus segments, 1, 2, and 3, which are characteristic of this family of polymerases; are encoded by linear bacteriophages (Escarmis and Salas 1982, Savilahti and Bamford 1987, Salas 1991, Yoshikawa and Ito 1982) and other linear mitochondrial plasmids (Chan et al. 1991, Kempken et al. 1989, Oeser and Tudzynski 1989, Paillard et al. 1985, Robison et al. 1991, Roch et al. 1991, Court et al. 1993). Therefore the putative DNA polymerase of Har-7.0 plasmid is also likely to be a terminal-protein-primed B-type DNA polymerase. All the changed amino acids in ORF2 are outside the proof-reading and polymerization domains. This means the fundamental functions of the putative DNA polymerase are very well
conserved. Furthermore the consistently spaced serine, tyrosine, lysine and asparagine (SKYN) residues thought to represent part of a terminal protein sequence (Oeser and Tudzynski 1989, Chan et al. 1991) is also present in Har-7.0 plasmid ORF 2. There are two 1-aa substitution between these spaced amino acids. There are 12 extra amino-acids at the carboxy-terminus relative to that of marDNA ORF 2 because of the TAT codon which encode lysine in Har-7.0 instead of the TAG stop codon in marDNA. This is curious in the view of conservation of the rest of the ORFs. Similar situation is seen in Gel-kalDNA ORF2 relative to kalDNA ORF2 (Wei et al. 1996).

The DNA polymerase and RNA polymerase ORFs in Har-7.0, Gel-kalilo (Wei et al. 1996), maranhar (Court et al. 1992), kalilo (Chan et al. 1991) and pCLK1 (Oeser and Tudzynski 1989) start in the TIRs of their respective plasmid, suggesting that a single promoter sequence, which is identical for both genes, may be found in the TIRs of each of these genetic elements. The 5' start site of both ORFs of maranhar map very close to the ends of the plasmid within the repeats, at around position 50 (Court and Bertrand 1993). Upstream of the start site, a 23-bp “promoter” region showed 10 matches to the comparable region of the Claviceps plasmid pCLK1 (Gessner-Uhlrich 1994). The kalilo ORFs also have a common start site close to the end of the plasmid, at around position 101 (Vickery and Griffiths 1993). However, the kalilo presumptive promoter region did not resemble that of maranhar. The Har-7.0 and marDNA have almost identical imperfect dyad symmetry except one substitution in the terminal sequences of the TIR, which is thought to be the downstream of the promoter region. The promoter regions of these two plasmids are almost identical. The 3' ends of both kalilo and maranhar map to the intergenic region. Like the intergenic region of marDNA there are several repeats in both direct and inverted orientation within the intergenic region of Har-7.0 plasmid. The arrangement whereby transcripts and ORFs start in the terminal inverted repeat has also been found in several linear plasmids in other fungi [for example, in Claviceps (Gessner-Uhlrich 1994) and Podospora (Hermanns 1994) species].
The terminal nucleotides of the plasmids are likely to be the origin of replication and, in the case of the *Neurospora* kalilo and maranhar plasmids, sites of integration into mtDNA. But a previous study (Court *et al.* 1992) suggest that there is no evolutionary conservation in the nucleotide sequences and palindromic regions between the terminal nucleotides of maranhar plasmid and those of kalilo plasmid even though these two plasmids show the similar insertion behavior. But the mode of kalilo and maranhar insertion is quite novel, having never previously been observed in any other system, eukaryotic or prokaryotic. Both IS-kalDNA (the inserted form of kalilo) and IS-marDNA (the inserted form of maranhar) are found flanked by long inverted repeats of the mtDNA, formed from the DNA to one side of the insertion point. The reciprocal of this product (containing the other flanking sequence) has not been demonstrated. The insertion mechanism must be different in the two plasmids. AR-kalDNA (the free form of kalilo) inserts by matching 5 bp from anywhere within the last 20 bp or so at its terminus with an identical quintet in the mtDNA (Chan *et al.* 1991). Presumably, some kind of crossover event then forms a recombinant molecule. Crossing-over leads to an almost full-length kalilo molecule flanked by the mtDNA to one side of the recombination point (Court *et al.* 1991). Maranhar integrates as a full-length copy, the insertion process has no need for a 5 bp match of terminus and the target sites in the mtDNA (Court *et al.* 1991). Homologous recombination is not responsible for the integration of maranhar plasmid. Previous studies on the kalilo plasmids of *N. intermedia* (Chan *et al.* 1991), *N. tetrasperma* (He 1995 & Marcinka-Kuehn 1994) have shown that the *N. tetrasperma* form (LA-kalDNA) is almost identical to the *N. intermedia* form (kalDNA), differing only by a 60-bp deletion and a 13-bp insertion of unknown origin at the same position in the center of the inverted repeats. This modification might account for the lack of insertion of the LA-kalDNA plasmid into the mtDNA in *N. tetrasperma*. The Gelasinospora plasmid (Gel-kalDNA) is clearly homologous to the kalDNA and LA-kalDNA. Its ORFs are virtually the same as those in the *Neurospora* plasmids; although there are small mutations,
the continuity of the reading frames is preserved. However, the intergenic region and the terminal repeats show numerous small and large mutations, which might prevent its insertion. Similarly the small mutations, including deletions, insertions and substitutions in the intergenic region and the terminal repeats of Har-7.0 plasmid, relative to those of marDNA, might explain the lack of insertion of Har-7.0 plasmid into mtDNA in *N. intermedia*. But the possibility can not be ruled out that the nuclear suppressors might be involved. The suppressors may inhibit plasmid insertion (Griffiths *et al.* 1992).

At one end of Har-L, beginning at position 207 right after the TIR, there is a section of 170 bp with 100% similarity to the termini of the Har-7.0 plasmid. This segment is not present at the other end of Har-L. This segment is the only marDNA-homologous region of the Har-L probably derived from a Zhisi plasmid. This further supports that the 5.5 kb Har-L is partly from a Zhisi plasmid which is marDNA-homologous (Griffiths and Yang 1995).

Without the sequences of the Zhisi plasmids the comparisons of the sequences of the Har-7.0 and the Zhisi plasmids can not be carried out. Therefore, it is not sure if the Har-7.0 is identical to the 7.0 kb Zhisi plasmid or not. And the source of the extra material of the 8.0 kb and 9.0 kb Zhisi plasmid can not be determined.
Summary

After the investigation of this newly-detected plasmid it is found that this new plasmid in a deviant Harbin strain of *N. intermedia* is a linear mitochondrial plasmid with structural features almost identical to maranhar plasmid originally found in Maranhar strain of *N. crassa*. This new plasmid Har-7.0 in *N. intermedia* Harbin strain 3983M-7.0 is related to Zhisi 7.0 plasmid in *N. intermedia* Harbin strain 3983. It is probably the same as 7.0 kb Zhisi plasmid. The mechanism of the evolvement of this new plasmid might provide some clue about plasmid evolution in nature.

Many similarities existed between maranhar and Har-7.0. Plasmid size was identical, the restriction enzyme map was pretty similar on the basis of fragment patterns from 5 enzymes, and both plasmids were resistant to 5' exonuclease digestion.

The identity with maranhar was remarkable at DNA sequence. The nucleotide similarity is 98.55% over all the region. It was the first time a maranhar plasmid had been confirmed in a fungal strain *N. intermedia* other than *N. crassa*.

The distribution of identical plasmids in these two species could be explained by horizontal transfer or gene flow. The horizontal transfer of plasmids has been demonstrated. *N. intermedia* and *N. crassa* may cross. This will allow plasmid transfer between these two species. But the horizontal transfer mechanism is preferred.
More study is needed on the distribution of marDNA-homologous plasmids. The more extensive survey of related plasmids might find more marDNA-homologous plasmids in other species of *Neurospora* or another genera like kalilo plasmid. Plasmids from different species or genera could become usual markers for studying host evolution. Further studies on the replication, transcription and translation of Har-7.0 plasmid could be carried out. Are the terminal nucleotides of the plasmid the origin of replication? Do small deletion and nucleotide substitutions the terminal repeats effect the transmission of this plasmid during the sexual cycle? Are the ORFs transcribed like the ORFs of marDNA and kalDNA? Are there any candidate proteins for translational products like in Maranhar strain? Do the nucleotide substitutions at the terminal repeats effect the integration of this linear plasmid making it avirulent?
Figure 2-1. The nucleotide sequence of the 5' to 3' strand of Har-7.0 plasmid.

The terminal inverted repeats are underlined. The amino-acid sequences of the ORF1 and ORF2 are presented below the nucleotide sequence. This strand of the nucleotide sequence is the sense-strand for ORF1 and the antisense-strand for ORF2. The primary sequences start at the first methionine in each sequence. Potential translation initiation codon (ATG) and termination codons (TAA and TAG) are bold.

1 TGGGGGAGT ACATAAATCC CTACTTTATA AATGAACAT CCCCCCTATA GAGGAAATAG
61 CCTATATAAC ACTATGACTG ATCATAGAT CAGTCTTAG TGTGTACTA TTCTCCTCAAC
121 CTATAACAGT CGTCTTACCT TCAGCGCTTA TGCAGCTCTT CACAGAATGG GGCCAGATCC
181 AAAGACTACA GTAGGGCAGT CTCCAGAACG ATCCGGAGGC GAAGCTTCCT ATTACCTGTCC
241 AAGAGGAGAG CTGAAAAATA GTAGGAATCG GGGTCAGAGA ATATATTTTT AATGATTTAAA
      ORF1-M I K
301 TTTTATTTTA CCACATTTCA CATGAAATTT CCTGCACGTT TATTTTTCAAC TTTCACTACA
      FYFTFTFH MKF PARLFLST SFT
361 TTGAATATAA CGAATATAAT AAATGAAAAA GATGGTTAAAT ATATTACTAA AAACCTTTTA
      LNNTNINNEKDVK YITKKNFL
421 TTGAAAAATT CAAATGTTTT TAATTTAATT AAAATATAA TTAAATCTGA TGAAACATCT
      LENSNGNFLN KNIINSDETS
481 GAAGTTAAAC AAAAGAAAAA TGAAGTTGAA TTTAAAAATA TATGCGCATC TGAAATAACT
      EVKQKKIKEVEELNNIWHТЕTT
541 GACATTTTTA AGAAAAAAG AAGCCTTAGG TTGGATGCAA TAGGAACTTC TATTTTTAGCT
      DILQKKRLGLDAILIGTSIL
601 AAAGATTTTTC AATAATTTAA AGGAGATAAT GAGAATTTTTA TTAATAAAGG AGAACAATAAT
      KDFHNLIGDIENTKFGRTN
661 AAGTTACCAG GCCTTGAAATA CTTCGAACCA TCATTAATAA TTCTTATTGT CTTAGGAAAA
      KLPGVEYLKPSLIIISIVLKG
721 GTTATTCCCTT TGAATTCTTAG ACATTCGATG ATTTTTAACC AACCTACTCA TTCTTTTTAT
      VIPFSLRHSDDLNPHTSLF
781 GCAGAAAAATG GAAAAACATT AAAATAATCA TCTATTTTTG AATTATATCA TAGAATTGGT
      AEIGKTLKYQSSIFELHHRIG
841 CTATTCCAAG ATCTGTTGGA AGAAATTTAA AATCTACTG AAAAAAATT AGTATCTGAA
      LIQNRVEIEKNNTEKKSLLVLVE
901 TTTAATAAAT TAACCAAGGTC ACTAGATGAA TATAAAAACA CTCTAAAAAA CTCTAGAGAA
      FNKLTKVDLEYKNTLKNLEEE
<table>
<thead>
<tr>
<th>Line</th>
<th>Sequence</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>961</td>
<td>TTAAGCGGGG AAAGTATTAT AAAAGTAGGA</td>
<td>TTAGGATTTA CAGCACCTTTT AAGTGAAACT LSGESIKVGLGF TALLS ETT</td>
</tr>
<tr>
<td>1021</td>
<td>AGTGAATTITTT ACTCTCTCGA AGAACA AMATT ATAGCTAAAA ATAAAATCAAT AAGATATATA SEFYSLEEQIIIAKKNKSI RYI</td>
<td></td>
</tr>
<tr>
<td>1081</td>
<td>TTACCTAAAA ATAAATTAAA TACATTTAAT AATAATAAA CATATATGGA TACAATTGAA LPKLNKLNTLINNITLMDTIE</td>
<td></td>
</tr>
<tr>
<td>1141</td>
<td>CTCCCAATGA TTATACCACC ACTTGAA TGA AAAATTGATG ATATAGGAA AATTATTGAA LPMIIPPLEW KDNEKIE</td>
<td></td>
</tr>
<tr>
<td>1201</td>
<td>TATGTTGGAAG ATGATCTTAA TAATATACAC AGGATAGAC CATTAAGAC AAAAACTAGTA YGTTILNKKHRIR PLRTKSV</td>
<td></td>
</tr>
<tr>
<td>1261</td>
<td>GAGAATTCAG ATGCAAATGA TATGACATAT AATAAAGAGT TAGTGATGC CGTAACTTT ENSDANDMTYNKELVD AVPNF</td>
<td></td>
</tr>
<tr>
<td>1321</td>
<td>AAATCAAAAA TACCATATAT TATATAATTTA AAAACTTTTAG ATTTTATTAC TAGAGTGA AKSKPYIKNKLKLDFITTRED</td>
<td></td>
</tr>
<tr>
<td>1381</td>
<td>TTATTTATAC GTGTAAAAA AGATAATGTT ATTATTTATA AACATATTCA TCCTGACTCG FINRDKKDNVIIYKHIHPDS</td>
<td></td>
</tr>
<tr>
<td>1441</td>
<td>GTTATTTGGG GTGAGTATAT GAAGGATCTG AAAAAACCAA AGATTAGTG AATAACGACT AALLGEYMKDKNPKISEITT</td>
<td></td>
</tr>
<tr>
<td>1501</td>
<td>CATAATAGTA AATTCTTTTA TCATTCTCTC ATAAATTTCA TAGCAAAATT AATGAAGAT HNSKFLY HSSISIAIKLMMKD</td>
<td></td>
</tr>
<tr>
<td>1561</td>
<td>GTTAAGGAGT TCATATAGC GTATTTTATA GATTGAAAGGA AAGATTTTTA TACTCAGTT VKEFYMTVFIDWRGRFTYSTS</td>
<td></td>
</tr>
<tr>
<td>1621</td>
<td>TGTGCGCTTA ACATAACAAGG TGGGAATTT GCTAGAAGTT TACTTCTATT TAAAGAAGGA CAILN IQGELARSLLLFKEG</td>
<td></td>
</tr>
<tr>
<td>1681</td>
<td>CAGAAATTTAA ATGATATTGG ATTAAAGCT TTTAAATAT ATACTGCAAAGT GCTTCTCGGT QKLNDIG LKALKIYTANAFG</td>
<td></td>
</tr>
<tr>
<td>1741</td>
<td>CTGATAAAA GATCTAAAGA AGAAAGATTT GATTGAGTCG AACAAGATTT GCATAAATTT LDKRSEKRLBDWVEQNHLKI</td>
<td></td>
</tr>
<tr>
<td>1801</td>
<td>ATGATATAC ACAATATTGA AATATGAAGA GAAGCTGATG AACCATTACT ATTTTTAGCC IDIDTNYEYWRLOEDEPLFLA</td>
<td></td>
</tr>
<tr>
<td>1861</td>
<td>TGTGCCTTTG AGITGAAAGG GTATAAAGAG GACCCCAATT TTATTTCAAG TTTAACCAATT CABLELKG YKD PNFISHLPI</td>
<td></td>
</tr>
<tr>
<td>1921</td>
<td>CTATGGATG CGACATGTAA TGGAATACAT TATATATAATGCA AATTTCTAGG TATTTTTTT GATLMDATCNGLQHLSAMVNDFV</td>
<td></td>
</tr>
<tr>
<td>1981</td>
<td>TTAGCAGAAA AAGTTAACAT ACTTAAGACT ACTGAAAAAG ATATATCTTAG GATTTTATAT LAAEKVNLKSLKSTENDNPRDLY</td>
<td></td>
</tr>
<tr>
<td>2041</td>
<td>TCTGAAGTAA TTTGCTTAAT AAAACAGAA ATTTTGAAGG CATCAAATTC GTAGGAACAC SENVIPHIKQIEILEASKSYEH</td>
<td></td>
</tr>
<tr>
<td>2101</td>
<td>ACAACCTTGG AGCGTATAAA TGTAGAGAGA TGTCTGTTA AAAAAAGACT TATGACTATT TNLERINVERCLVKRGMLMTI</td>
<td></td>
</tr>
</tbody>
</table>
ACATATGGAG CAACCTGAAAG AGGAATTTAT GATCAAAATTG TATCTAAATT CTTTCAAAAA TYGATERGIYDQIVSKFFQK
GATGAATGAA ACAAAACTGC AGGGCTTCAT TTTGTATGTA TAGATTCGGA TATAAGCTCCA DEW NKTAGLHFPVCIDSIDIA P
AAAGATGTTG TATTTACTCA AAAAGAATATA CTCTTTTGGAA GTAAAATTAT TTATAATTCT KDV VFQ KNI LWSKII YNS
TTATTTAAAA TTATCCTCAA TTAAATACT CTAATGGTAT ATTTTAATAG TATAGTTAAA LFK IHPNLNTLMYFNSIVK
GTATTATGTG AATTTAGTTT ACCTGTAAAT TGAGTAACTC CTATGGTTTT AGTAATACAA VCL ELDLPVNWPTYPGLVIQ
CAAAAAATATA ATAAATTTAC TAAATATAAT GAAACAACCT ATGGTGACAT CAAACGATAT QKYNKFKETTYVASKRY
AAATTAGTAT TAAGAAAAAC AGATACTACT TCTATTTCTA AGCAGAAAAA CAACTAAACG KLVLRKLDDTSISKQKQIQA
TTATACCCA ATTTTTGTTCA CTCAATGGAT GGAAGTAAATA TTGTTTTATT AATAAACAFA FIPNFVHSDMGNSIVLILKT
ATAAGAGATG AAGGACCCAA AATTAAACCTT GAAAGTAACT ATGGATGTTTT TGCAACTCAT IRDEGRKINFASIHDFCATH
GCAAATGATA CTGATGTACT ATCATGATAT GTAAGCAAAA GTTTTATACG TATTTATTCT ANDTAWLSTYVKQSPIRYS
GATTCTAGTT TTTTAAAGAG ATTTCTAATA TATATCAAAAT TAAAGATCCA AGCTAAATAT DSSFLRRFHNYIQLRIOAKY
GGAGAAGATG CTCAATTTTCA TAAAGATGTT GTATTAAACCC ATGTAACTAT TGATATAGGT GEDAHFDKDGVTHTVIKG N
ACAGATATTG CTGATGTTAT TGATATTAAC TAAACTCTTG TAGTAATAAAG AGTAATAAAA TDI PDRIKPLVVKNDN
ATATATAAAG AAATCTTCTCA TTCTGAGTAT TTTTTATTTT AATATTTATA TATAGTATTG IYKEILSEYFIN *
AAATAATATA AATATAGTGG GATTAAACCA AAAAGGGACA ACACAAATAG TGTATACCT TTTTTAAAAG GATATTTCTA ATTATGTTAT TTTTAATTAT TATAGTTAAA
AGTAAATACA CATGTAATAA CAAAGTAAATG TTTGTATGTA ATAGGTAGTA CGTGATTGC
ACCAGGATC TACCTCTCTT CAAATCTTTA TATACACTG AAGGTTAAAT CATAGAAAAA
TACCTATATT TAATACTGTT AAGTATATAA TTTATAATTA TAATATTTA TAATATTATA
ACAAAAATTA CACCTGTCCA GCTGTTGACA AGTACTGTTA CGAGTAGGAT AAAACCTCG
AAGGGGAAAA AATRAAAACAA AAAAAAGGAAATG CAAAACCTGCA ATTCCCCATAA AAAATAGGGA
TACACAAATAC TTTTGTTGTT TCTTTTTTTA AAGGTATACAA CAAATACGT TCTTTTTTTA AGGTATACAA CAAATACGT
CTTAAAGGT AACAACATAC TGCTATATATC TTACCCGAA AAAGTATTTA
TCAGGGTAA CCAACAATAC ATTTATAATA TATTTATTAA AATAACATAC ATACCTCTC
Figure 2-3. The alignment of the sequences of Har-7.0 and maranhar and Har-L (Court et al. 1992, Yang and Griffiths)

The stars represent the consensus sequence relative to Har-7.0 plasmid. The horizontal lines represent the deleted residues.

```
7Kb 1 TGCGGGAGTCACATAATCCCTACCTTTATAAAATGAACATCCCCCTTATA GAGGAAATAG
Mar ******************** TC GGG**G** **T****T** T******** *********** ***********
Har-L 206 *************** **************** *********** *********** ***********

7Kb 61 CCTATATAGACTATGACTGATCATTAGATCAGTCTATAGTGTCTTTACTATTTCCTCAAC
Mar ******************** ******************** ********************T
Har-L 375

7Kb 121 GTATAACATGCGTTTTACCTTTCCCGCTTA TGCGGCTCTGACAGAGTTGGGCGCGATCC
Mar G********* **************** ******************** ******************** 375
Har-L

7Kb 181 AAAGACTACAGTACGCACATCTCCAGACAGATCCGGGAGC GAAGCCTACTATTACCTGTC
Mar ******************** ******************** ******************** ********************

7Kb 241 AAGAGGAGAGCTGATAATAGTATGAAATCGGGTCACAGAATATATTTTAATGAATTATA
Mar ******************** ******************** ******************** ********************

7Kb 301 TTTTTTTTTACACCATTTGACATGAAATTTTCTGCACGTTTTATTTTCACA
Mar ******************** ******************** ******************** ********************

7Kb 361 TTGATAATACGAAATATAATAAATGAAAAAGATGTAAATATATTACCTAAACACTCTTTA
Mar ******************** ******************** ******************** ********************

7Kb 421 TTGAAAAATTCAAATGTTTTTAAATTTAATTTAATAAAATATAAAATTTATTTAAATTTTCTGA
Mar ******************** ******************** ******************** ********************

7Kb 481 GAAATTTACAAAGAATAATGAAATACTAATATGAGCTTGACGTGAATTATATATAATATGATTAATCT
Mar ******************** ******************** ******************** ********************

7Kb 541 GACATTACAGAAAAAAGAAGCTTACGGTTGGATGCAAATAGGAACTTCGAGGAGGATAG
Mar ******************** ******************** ******************** ********************

7Kb 601 AAAGATTTTCTATAAAAAATAGGAGATATAAGAATTTTTAAAATATAAAGGAAACAAATT
Mar ******************** ******************** ******************** ********************
```
<table>
<thead>
<tr>
<th>7Kb</th>
<th>661</th>
<th>AAGTTACCAG GCCCTGAATA CTTGAACCA TCATTAAATTA TTCTTATTGT CTTAGGAAAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ******* ********** ******* **</td>
</tr>
<tr>
<td>7Kb</td>
<td>721</td>
<td>GYATATCTCTT TCAGTCTTAG ACATTAGAT ATTTTAACCC AACCCTACTCA TTCTTTATTT</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>781</td>
<td>GCAGAATAG GAAAACATT AAAATATCAA TCTATTTTTG AATTACATCA TAGAATTGGT</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>841</td>
<td>CTATATCCAA ATCGTGTGGA AGAAATTTAA AATTCTACTG AAAAATTATT AGTATCTGAA</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ******* ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>901</td>
<td>TTAAATGAAAT TACAAAAAGT ACTAGATGAA TATAAAAACA CTAAAAAACA CTTAAAGGAA</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>961</td>
<td>TTAAGCGGGG AAAGTATTAT AAAAGTAGGA TTAGGATTTA CAGCACTTTT AAGTGAAGCT</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>1021</td>
<td>AGTGAATTTT ACTCTCTGGA AGAAACAACT ATTTTCTACT CATTAAGAAC AAGGAAAAAT</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>1081</td>
<td>TTACCTAAAA ATAAATTAAA TACATTAATT AATAATATAA CATTAATGGA TCAAATGGA</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>1141</td>
<td>CTCCCAATGA TTATACCACC ACTGAAATGA AAAATTGATG ATTTGAGAA AAAATTGAA</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>1201</td>
<td>TATGGTGGAA CTATACCTAA TAATACACAC AGGATTACG CATGAAAGG AAAATCGTA</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>1261</td>
<td>GAGAATTCAG ATGCACCAAT CAGTACTATAT AAAAAAGAGT TAGTTGATGC CTTAAGCTTT</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>1321</td>
<td>AAAATCAAAA TACCATATAT TATTTATTTA AAACTTTTAT ATTTATTAT TAGGATGAA</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>1381</td>
<td>TTATTAATC GTGATAAAAA AGATAAGGTT ATTATTATA AACATATTCA TCCTGACTCG</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>1441</td>
<td>GCTTTATTGG GTGATGATAT GAAGGATCGT AAAAAACCAA AGATAAGGTG AAATACGACT</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>1501</td>
<td>CATATACTGA ATTTTTCTTA TCATTCTTCT ATTAATTTCA TAGCAAATT AAGGAAAAAT</td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td>** ********** ******* ********** ********** ********** ********** **</td>
</tr>
<tr>
<td>7Kb</td>
<td>1561</td>
<td>GTTAAAGGAGT TCTATATGAC GGTATTTATA GATGGAAGAG GAAGATTTTTA TACTTCAAGT</td>
</tr>
</tbody>
</table>
Mar

7Kb 1621 TGTGCGCTTA ACATACAAGG TGGGAATTG GCTAGAAGTT TACTTCTATT TAAAGAAGGA

Mar

7Kb 1681 CAGAAATTAA ATGATATTGG ATTAAAAGCT TTAAAAATAT ATACTGCAAAG TGCTTTCGGT

Mar

7Kb 1741 CTGTGATAAA GATCTAAAGA AGAAAGATTG GATTGAGTCG AACAAAATTT GCATAAAATT

Mar

7Kb 1801 ATTGATATAG ACAATTATGA AATATGAAGA GAAGCTGATG AACCATTACT ATTTTTAGCC

Mar

7Kb 1861 TGTGCTTTGG AGTGAAGAGG GTATAAAGAG GACCCCAATT TTATTTCACA TTTACCAATT

Mar

7Kb 1921 CT TAATGGATG CAACATGTAA TGGATTACAA CATTTAAGTG CAATGGTAAA TGATTTTGTT

Mar

7Kb 1981 TTAGCAGAAA AAGTTAATTT ACTTAAGAGT ACTGAAAATG ATAATCCTAG GGATTATAT

Mar

7Kb 2041 TCTGAAGTAA TTCCTCATAT AAAACAAGAA ATTTTAGAAG CATCAAAATC GTACGAACAC

Mar

7Kb 2101 ACAAAACCTTG AGCGTATAAA TGTAGAGAGA TGTCTAGTTA AAAGGGGACT TATGACTATT

Mar

7Kb 2161 ACATATGGAG CAACTGAAAG AGGAATTTAT GATCAAATTG TATCTAAATT CTTTCAAAAA

Mar

7Kb 2221 GATGAATGAA ACAAAACTGC AGGGCTTCAT TTTGTATGTA TAGATTCGGA TATAGCTCCA

Mar

7Kb 2281 AAAGATTTG TATTTACTCA AAAGAAATATA CTTCTTTGGA GTAAAATTAT TTATAATTCT

Mar

7Kb 2341 TTATTTAAAA TTATCCTCAA TTTAAAATACT CTAATGGTAT ATTTTTAATAG TATAGTTAAA

Mar

7Kb 2401 GTATTATGGG AATTAGATTT AACTGGTTAT TGAGTAACCT CTTATGGTTT AGTAATACCA

Mar

7Kb 2461 CAAAAATATA AATAATTTAC TAAATATAAT GAAACAATTT ATGTGGCATC TAAACGATAT

Mar
Mar

7Kb 3474 ATACTGTGTA TACC--TTTT TTTAAAGTATA CACAAATCT GTGTATATCT TATACCGAAA
Mar

7Kb 3532 AAGTATTAAT CAAGGTTAAC CCAATAACA TTTATATAT ATTAACCTAA ATAACAATCA
Mar

7Kb 3592 TTACCTCATA TTGTAATATA TGGTGTGAGG AAAATAAAAT GTATATTAG TTGTATATCT
Mar

7Kb 3652 TTGAATACT TTGAATAAT CATATATA--A TAT-----GA TTATATTAT TAAATGTAT
Mar

7Kb 3766 ATTTTGTATT ATGAAATATG GTGAAGTATT AACTCATTTA CCTAAAGAGT ATAGTTTTC
Mar

7Kb 3826 AGTTTATAC TAACCCTGATA TAAGATATAA TTTTGATAT TAAACTGAA CTGTTCTAA
Mar

7Kb 3886 TTCCAGCTCA CGTTCTCACG TTTCTGAAA TAGTACCCAA CTGTCCCAT TATAAGTGC
Mar

7Kb 3946 TTTGAAATCA TTAATATCTA GATTATGTTGT GGTATTATCT TTGTTTTTAG TTATCCCTTT
Mar

7Kb 4006 ACATTATAAT TCTAATTGAC CACAAAATTC TAAATACATAT AATTTTCTCTG AATGAAAAT
Mar

7Kb 4066 TGCACGTATG ATTAATACGA CTTATATATC AGCTTTAAAT GGTCTGATA CGTAAATGTT
Mar

7Kb 4126 AAAAGCCTAC TCTAATTGTT TTTCCACAAAT TATGAAATCA GTATCAGTT ATGCTGAATT
Mar

7Kb 4186 AATTTAGTAG TTATACATCA GAATTCTAGA TCATGATGCT GTTGCTGCAG CAATGGATGT
Mar

7Kb 4246 AGAGTTTTATG ATAAATCCAT CAGTTTTCAT TCCATCTAGG TAAGATAAAA GTGCTGAAAT
Mar

7Kb 4306 TTTTCTGTAT TGTCACACAA ATACAGGACA TGTTTTTTTA CGTATCTCA CGTAATGTCT
7Kb  4366  ATCATCATCT ACTTCAAATT CGTGAATAAC GTTATTAGTT AATTGAATAT TATCTAATTC
Mar          **************************** **************************** ****************************
7Kb  4426  ATTAGTAGTT AGCATTTTAA TTTCATCATTAC ATACCTGTTG TACATATAAA
Mar          **************************** **************************** **************************** G***********
7Kb  4486  AGTATGCTAAA AGCAATTCTAC CCATCTGTATTT CTATTGAACTTT TAGATTATAT CTTTAATTGA
Mar          **************************** **************************** ****************************
7Kb  4546  GGAATAATGCG TCAATATTCTCT ATCTCTCTCT TTTTTCGAAAA CGTAACCTCTC
Mar          **************************** **************************** ****************************
7Kb  4606  AAGCTACTCTCT ATTTTATATTCT CGTACCTCAAC TGCTAAATTCTT ATTCTTACAG AAAATATCA
Mar          **************************** **************************** ****************************
7Kb  4666  CCCCTGTCAT TCACCAATGG GAAAATCTTTA TTTTTGTACTT CCATTGACTT TAACCTGTC
Mar          **************************** **************************** ****************************
7Kb  4726  AGGTAAATCTT GGAATATTATA CTGCTTGGATG AAATTATTTTT GCTTTTCTTAA ATCCAAATAT
Mar          **************************** **************************** ****************************
7Kb  4786  TTCAATTATAG TTTTTAGAAA AATATATGAC TGGTATACCA ACAGGCATTTG GCTACATCAT
Mar          **************************** **************************** ****************************
7Kb  4846  TGCAGTAGGG TAAAGACTTT TAAAATTCAGT TAAAATCAAA TGGAAAAATA TTTATTATTTT TGGAAATAAA
Mar          G*********** **************************** **************************** ****************************
7Kb  4906  AATTTCATTCA CGTCACCAGA AATAACGAGC TCTAACAGGCT TTTTCAAGCT TTCTTTTCTAG
Mar          **************************** **************************** ****************************
7Kb  4966  TTTAGGTTAAA AAATATTTTAG GAGTTAATTCT TTTCTCTCTCT TTTTAGTTCAA ACAATTGATAA
Mar          **************************** **************************** ****************************
7Kb  5026  AAGTTATTTAT GGTGAAGTAG TTTTATCTCT TTCTCTTCTCT TCTACTTCAT TGGGAAATTTAT
Mar          **************************** **************************** ****************************
7Kb  5086  TAACGGTCTTA TAAACAGATA AGGCCAATGC AGATGCAGTT TTAACCTCTAG TAAATATTTAT
Mar          **T************ **************************** ****************************
7Kb  5146  TCTAAAATGT GAAATGTAT TATTACTCAT TCTGATCAT AATGATATA AAGCCCTTTAT
Mar          **************************** **************************** ****************************
7Kb  5206  ATCTTTTTCT AAGTAGATAAA TGCTTTTCTTT TCTCGTGAT CATTTATTAG TATACATAGC
Mar          **************************** **************************** ****************************
7Kb  5266  AGCTCACTCA AATAGTGTAAT TCATTACCCT TTTTGGGA TCAATATAT ACTCATAATC
Mar          **************************** **************************** ****************************
Mar

7Kb 5326 AGGTATTAAAT CCAACATATT CTAAATTATC TTTATTAACA AATTHTGTAAG GAAATTTC
Mar

7Kb 5386 TTTTTTAGTA ATGATATTAT GATCTTTAGC TAATTTATCT AGTGATCCTG GTAGGAGTCT
Mar

7Kb 5446 ACATGAATCG GCAATGGTAA TAGATGCTGA CTCAGGCTTT CCTCCTTTTT TTTTAGGTTC
Mar

7Kb 5506 AAATTTATAT GAAATTTTTA TTGAAAGAAT ATCAAGATCT TTAGATATTA TTTTCTAC
Mar

7Kb 5566 AACAAATTCT TGAACAAATA TTTTGATATAT AAAATTAATT TCAAATTATG AAAAATATG
Mar

7Kb 5626 GCAATATCTG TATGTTTGT CATATTTTTA CATATCCTTT ATACACGCTA GCAAATTTTC
Mar

7Kb 5686 AGTGTGAGAA ATAAAATCAG AAATGTAATA AGTAAGAGAT TTATTTCCAT CATAAAATCC
Mar

7Kb 5746 ACAAGCATAA GCAATCATCG TTGAATCACC ATTCCCAGTA GGAACTTGAA ATGTTTCAAT
Mar

7Kb 5806 ACATATGTCT AATTTTTTTA CATATTTTTA CATATCTTTT ATACATGCTA GCAAATTTTC
Mar

7Kb 5866 ACATTTTTGGAGTTTTCAA CATTGACTAA AGTATTATCA TAGTAATAT AATACATATT
Mar

7Kb 5926 ACATATTTTTA CTAAGCCTAA ACCATTTCTA ACATCTTTA ACCGCTTTAA ACTGATCAA
Mar

7Kb 5986 TAATCTATTA TCAGTGAAA TAAACATCTAT AAGATGATT AACCACATTAAT TGATAATTCT
Mar

7Kb 6046 CACCTTAAGA ATACATTTCT CATTATCAAA AATACGAATG TTATTTTCTT TGATAATTCT
Mar

7Kb 6106 TCTTTACAACA CCAAAATTTA TAAAGCTTTA TGTGTGCTGG ACAAATTTCTGT GTAAGAAC
Mar

7Kb 6166 TTGAGACCTTT CCTTTAGGAA TAGTAGCTGT TTAAGAAGTT TTTTTCTTTT TGATAAGATG
Mar
Figure 2-4. The sequence alignment of ORF1 of Har-7.0 and maranhar.

The stars represent the consensus sequence.

ORF1>
7Kb MIKYFTTFH MKPPARLFST SFTLANTNII NEKDVKYIK NFLLENSNGF NLKNIINSD
Mar ********** ********** ********** ********** ********** **********
7Kb ETSEVKQKRI EVELNNIWH EITDIQKXKR SLOLDAIQTS ILAKDFHNLI GDIEFINKG
Mar **********N** ********** ********** ********** ********** **********
7Kb RTNKLPGVEY LKPSLIIISIV LGKVIFSLR HSDILNPQHT SLPAEIKTL KYQSIHELH
Mar ********** ********** ********** ********** ********** **********
7Kb RIGLQNRVE EIKNSTEKL VSEFNKLIKV LDEYKNTKN LEELSGSII KVGGLTALL
Mar **********Y******** ********** ********** ********** ********** **********
7Kb SETSEFYSLE EQIAKNKSI RYILPKNKLN TLINNITLM TIELPMIIPP LEWKIDNEK
Mar ********** ********** ********** ********** ********** **********
7Kb IIEYGZTILN NKHRIRPLRT KSVENSDAND MTYNKELVDA VNPFSKIPYI INLKLIDFIT
Mar ********** ********** ********** ********** ********** **********T********
7Kb RDEFINRDKK DNVIYKIHII PDSALLGEGYM KDRKNNPKISE ITTHNSKFLY HSSIISIAKL
Mar ********** ********** ********** ********** ********** **********
7Kb MKDVKEFYMT VFIDWRGRFY TSSCALZIGQ GELARSLLL LKGQKLNDIG LKALKIYTN
Mar ********** ********** ********** ********** ********** **********
7Kb AFGLDKRSKE ERLDQVQNL HKIDIDNIE IWEAEPPLFL ACALIEKGO YKEDPNFISH
Mar ********** ********** ********** ********** ********** **N**********
7Kb LPILMDATCN GLQHLSAMVN DFVLAEKVIN LKSTENDNPR DLYSEVIPHI KQIEILEASKS
Mar ********** ********** ********** ********** ********** **********
7Kb YEHTNLERRIN VERCLVRGK MTTIGGTER GIYDQIVSKF FQKDEWNKTE GLHFVCIDSD
Mar ********** ********** ********** ********** ********** **********V********
7Kb IAPKDVVFQ KIILWSKII YNSLFKIHNP LNTLMVYFSNS IVKVLCESDL PVNWTIPYGL
Mar **********L***** ********** ********** ********** ********** **********
7Kb VIQKYNKFT KYNNETTYVAS KRYKLVLRLKA DTTSISKQKQ IQAFIPNPHV SMDGSNIVLL
Mar ********** ********** ********** ********** ********** **********
7Kb IKTIRDEGRK INFASIHDCF AHTANDTAWL SWYVKQSFIR IYSDSSFLRR PHNYIQLRIQ


Figure 2-5. The sequence alignment of ORF2 of Har-7.0 and maranhar.

The stars represent the consensus sequence. The horizontal lines represent the deleted amino-acids.

<table>
<thead>
<tr>
<th>7Kb</th>
<th>FVKLYDYIIH NYKNYHIRKR YWTEGVDSVL A KSIVEFNEN ICLPSTNVWK GLSYLKERC KD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar</td>
<td>---------- --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>YGSILNYKY KVTVGLELS RGWREQFLVR SEGNYLAED SIDLNHTND KNKTIGCKI</td>
</tr>
<tr>
<td>Mar</td>
<td><strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>ELKGGFDLLY LRGSI FIARK ILQGNYBAKFL CGGEGFAS DLPEVFISD TDYASNI IH</td>
</tr>
<tr>
<td>Mar</td>
<td><strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>KYMLIRSWSA TAAAISTSNI IFGDDEGDLY YSLLEYNKES QACLVPCPKK DRYVYH KDDD</td>
</tr>
<tr>
<td>Mar</td>
<td><strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>VEFEHIVNNT LQINDLENT LMKIEAASDN MGTRGYLNL LLKAMQKS NG KTNDKISA PH</td>
</tr>
<tr>
<td>Mar</td>
<td><strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>EIYEKFDDR KEFVYSELVE I KYGYEVALK LEEFYWG TW EGIPFI KQV GNVVRCPLV</td>
</tr>
<tr>
<td>Mar</td>
<td><strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>PINAPIIIK ARVFPI E NL NKSPTHVPIG VPM PMMATP YLSNFDPS FI NNI IFI EN</td>
</tr>
<tr>
<td>Mar</td>
<td><strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>RGFPYAARVA RELRGLKPL FYKPTL KEE KDFLSSLINN PSTTNEEEE EVENPLFLTR</td>
</tr>
<tr>
<td>Mar</td>
<td><strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>YVLFA LASAT KVRTINIRFT S YTNNSMEMM LQYLAKIDKE LYIITEKRTS W KNTYMAAWE</td>
</tr>
<tr>
<td>Mar</td>
<td><strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>F LTIMEGKKP D IYYE YDPIL GYELN DKNV FKYPFKKKT IINHD KALKD LSGP LRC SD</td>
</tr>
<tr>
<td>Mar</td>
<td><strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>AITITH REAK GKK KPEF KY SIKSLLD LS KIIKEVVPE QV LIKI IFN I DF KSFPNH CV</td>
</tr>
<tr>
<td>Mar</td>
<td><strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>THKD YKLMDK IC ALMER QS IFDSIYYTLS KNGDYFGCAY AIMTSDGNT PVQFTEIDPA</td>
</tr>
<tr>
<td>Mar</td>
<td><strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong> --------- <strong>S</strong></td>
</tr>
<tr>
<td>7Kb</td>
<td>LIKKD QKRDKR KTPEINGVKQ P TENVNL TN YYYYYMN DNI WHRK LGDG HK V DEFQY LLRN</td>
</tr>
</tbody>
</table>
Figure 2-6. The sequence alignment of TIR of Har-7.0 and maranhar. The stars represent the consensus sequence. The horizontal lines represent the deleted residues.

| 7Kb   | 1   TGGGGGAGT ACATAATCC CTACTTTATA AAATGAACAT CCCCCTTATA GAGGAAATAG |
|-------|-----|---------------------------------------------------------------------|
| Mar   | -****TC GGG****G** **T****T** *T ******** ********* ********* ********* |
| 7Kb   | 61  CCTATATAAC ACTATGACTG ATCATTAGAT CAGTCTATAG TGGTTTACTA TTTCCCTCAAC |
| Mar   | ********* ********G** ********* ********* ********* ********* *********T |
| 7Kb   | 121 CTATAACATG CTTTTTACCT TTCCGCTTTA TGGGCTCCTT GACAGATTG GCCGAGATCC |
| Mar   | G******** ********* ********* ********* ********* ********* ********* |
| 7Kb   | 181 AAAGACTACA GTAGCACAT CTCCAGACAG ATCCGGGAGC GAAGCTACT ATACCTGTC |
| Mar   | ********* ********* ********* ********* ********* ********* ********* |
| 7Kb   | 241 AAGAGGAGAG CTGAAAATA GTAGGAATCG GGGTCACAGA ATATTATTTT AATGATTTAAA |
| Mar   | ********* ********* ********* ********* ********* ********* ********* |
| 7Kb   | 301 TTATTTTTA CCACATTTCA CATGAAATTT CCTGACGTT TATTTTC--- |
| Mar   | ********* ********* ********* ********* ********* ********* CAC |
Figure 2-7. The sequence alignment of the intergenic regions of Har-7.0 and maranhar.

The stars represent the consensus sequence. The horizontal lines represent the deleted residues. The numbers at the left indicate the positions of the first residues in that line in the whole nucleotide sequence of the Har-7.0 plasmid.

```
7Kb  ATGATAAA TAATGATTGT
    *********** ***********
Mar
7Kb  3001 AAATAATATA AATATAGTGG GATTTAACAA AAAAGGACA ACACAATTAG TGTATACCT
    *********** *********** *********** *********** *********** *********** ***********
Mar
7Kb  3061 TTTTTTAAAG GGAATACACA AATTTGTGTA TTTCCCTTTTT TTTT---AAT TATTATACGG
    *****... ----------- *********** *******TTA*** ***********
Mar
7Kb  3118 TTAAGTAATAT ACACATGTA AAACAAAGTA ATGTGGTTTT ATGATAGTGA GTACGTCGAT
    ********** ********** ********** ********** ********** ********** **********
Mar
7Kb  317 8 TTTTTTAAAG GGAATACACA AATTTGTGTA TTTCCCTTTTT TTTT AAT TATTATACGG
    *****... ----------- *********** *******TTA*** ***********
Mar
7Kb  3238 AAATACCTAT ATTTAAATCT GTTAAGTATA TAATTTTATA TTATAATATT TAATATTAAA
    ********** ********** ********** ********** ********** ********** **********
Mar
7Kb  3298 TAAACAAAAA TTACACCTGT CCAGCTCGTG ACAAGTACTG TAAC GAGATA GGTAAAACCC
    ********** ********** ********** ********** ********** ********** **********
Mar
7Kb  3358 TCGAAGGGAG AAAAATAAAA AAAAAGGGA ATGCAAAACC TGCATTCCCA TAAAAA
    ********** ********** ********** ********** ********** ********** **********
Mar
7Kb  3415 TAGGGATACAA CAAATCTTTT GTGTATCCCT TTTTTTTTCT TTTTTAAAG GTATACCGAA
    ********** ********** ********** ********** ********** ********** **********
Mar
7Kb  347 4 ATACTGTGTA TACC---TTTT TTAAAGTATA CACAAACTG TGTATATCT TATACCGAAA
    ********** ********** ********** ********** ********** ********** **********
Mar
7Kb  3532 AAGTGTTAAC ACAGTTAAAC CAAATAACA TTTATAATAT ATTAAACTAA ATAAACATCA
    ********** ********** ********** ********** ********** ********** **********
Mar
7Kb  3592 TTACCTCATA TTGTAATATA TGGTGTTAGG AAAATAAAT GTAAATATAG TTGTATACCT
    ********** ********** ********** ********** ********** ********** **********
Mar
7Kb  3652 ------------ ------------ ------------ ------------
```
Mar TTAGAATACT TTAGATAAT CATATAGACG CCTGGGGGGG CT
Figure 2-8. The comparison of the conserved motifs in plasmid-encoded RNA polymerases of maranhar and Har-7.0.

The motifs are numbered according to Oeser and Tuzdyski (1989). Because motif I and II could not be identified readily in all of the sequences (Chan et al. 1991) they are not shown. The consensus sequences are joined by vertical lines. The number of residues between the end of one motif and the beginning of the next, or between motif III and the putative amino-terminus of the RNA polymerase, are indicated to the left of the motifs. The sequence presented are derived from the nucleotide sequences of the maranhar (Court et al. 1992).

<table>
<thead>
<tr>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
</table>
| 7kb 280aa TIEIPPEWKIDNEKI 35aa LVDAVFNKSKKPIYINLKLLDF 56aa ISIATMKQVFYMTIFIDWRGRFYTTSMCIALNQGELARSLLL | I
| Mar 280aa TIEIPPEWKIDNEKI 35aa LVDAVFNKSKKPIYINLKLLDF 56aa ISIATMKQVFYMTIFIDWRGRFYTTSMCIALNQGELARSLLL |
| VI | VII | VIII | IX |
| 7kb 51aa IRREADPLLFLACALE13aaLPIIMDATCNGLQHLSAMNDFVLAEVKVNNLKTSTENIPIDLYSEV124aaERCLVRQKLMTTGY85aa VNWTFYGLVIOQKYN |
| Mar 51aa IRREADPLLFLACALE13aaLPIIMDATCNGLQHLSAMNDFVLAEVKVNNLKTSTENIPIDLYSEV124aaERCLVRQKLMTTGY85aa VNWTFYGLVIOQKYN |
| X | XI | XIII |
| 7Kb 28aa SKQKIQAFIPFNFHSMGDSN 12aa RKNFASIHDCPATHANDTAWLWSWVQKQRIYSD 35aa IDNIDIPRIDIPKPVNKNKIQKHEILHSEFIN |
| Mar 28aa SKQKIQAFIPFNFHSMGDSN 12aa RKNFASIHDCPATHANDTAWLWSWVQKQRIYSD 35aa IDNIDIPRIDIPKPVNKNKIQKHEILHSEFIN |


Bertrand H and Griffiths A J F 1989. Linear plasmids that integrate into mitochondrial DNA in *Neurospora*. Genome 31: 155-159


Broda P 1979. Plasmids. WH Freeman, New York


Dujon A and Belcour L 1989. Mitochondrial DNA instabilities and rearrangements in yeasts


Kemble R J and Thompson R D 1982. S1 and S2, the linear mitochondrial DNAs present in a male-sterile line of maize, possess terminal attached proteins. Nucleic Acid Res. 10: 8181-8190.


Kuzmin E V and Levchenko I V 1987. S1 plasmid from cms-S maize mitochondria encodes a viral type DNA polymerase. Nucleic Acid Res. 15: 6758


Lonsdale D M, Thompson R D and Hodge T P 1981. The integrated forms of the S1 and S2 DNA elements of maize sterile mitochondrial DNA are flanked by a large repeated sequence. Nucleic Acid Res. 9: 3657-3668


Nargang F E, Pande S, Kennell J C, Akins R A and Lambowitz A M 1992. Evidence that a 1.6 kilobase region of Neurospora mtDNA was derived by insertion of part of the LaBelle mitochondrial plasmid. Nucleic Acid Res. 20: 1101-1108


Paillard M, Sederoff and Levings III C S L 1985. Nucleotide sequence of the S-1 mitochondrial DNA from the S cytoplasm of maize. EMBO J. 4: 1125-1128


Sor F and Fukuhara H 1985. Structure of a linear plasmids of the yeast Kluyveromyces lactis: compact organization of the killer genome. Nucleic Acid Res. 9: 147-155

Sor F, Wesolowski M and Fukuara H 1983. Inverted terminal repetitions of the two linear DNA associated with the killer character of the yeast Kluyveromyces lactis. Nucleic Acid Res.11: 5037-5044


Turmel M, Bellemare G, Lee R W and Lemieux C 1986. A linear DNA molecule of 5.9 kilobase-pairs is highly homologous to the chloroplast DNA in the green alga Chlamydomonas moewasii. Plant Mol. Biol. 6: 313-319


Wang H and Lambowitz A M 1993. The Mauriceville plasmid reverse transcriptase can initiate cDNA synthesis de novo and may be related to reverse transcriptase and DNA polymerase progenitor. Cell 75:1071-1081


