3

Results

3.1 Chloroplast and Mitochondrial Comparison

3.1.1 Chloroplast homologies in the rice nuclear genome.

A substantial amount of homologies were found between the chloroplast and nuclear genomes. Table A-1 (Appendix, pp 101-193) shows the alignment data, including all the alignments found during the initial BLAST search. In order to eliminate small sequence repeats and non-significant alignments only fragments 100 bp and greater with 95% or higher homology were used. This 95% threshold was also used to limit the inclusion of sequences of mitochondrial origin in the chloroplast analysis and *vice versa*.

Using the above mentioned parameters, the rice chloroplast genome aligned to a total of 778678 bp in the nuclear genome. Figure 3.1 shows the distribution of these homologous regions over the rice karyotype. A total of 1070 alignments that made up the 778678 bp were used to draw the map. However fewer positions are visible on the map as insertions tend to group together in certain areas rather than being randomly distributed over the whole genome.

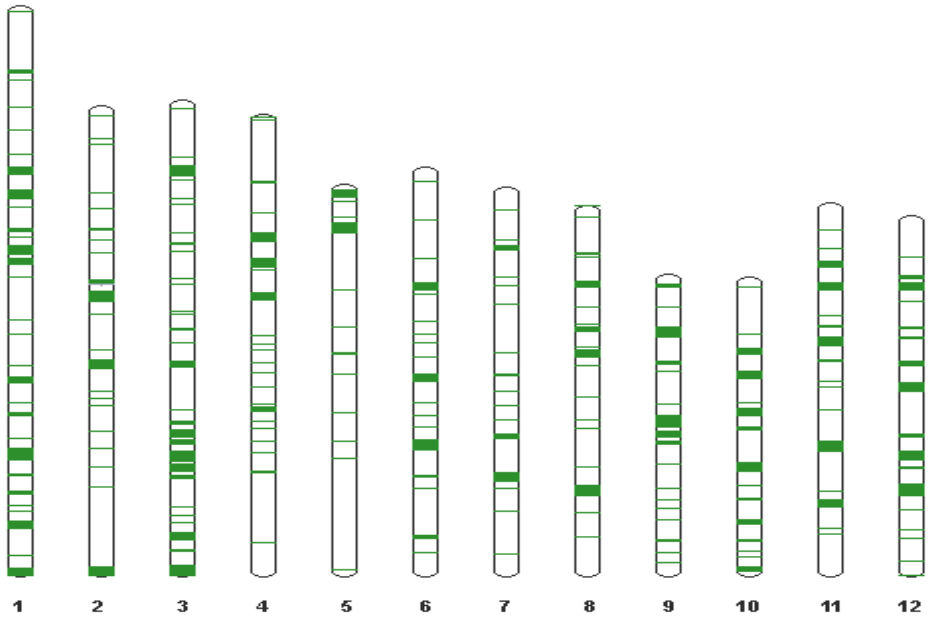

Figure 3.1: Distribution of chloroplast homologies in the rice nuclear genome.

3.1.2 Representation of the chloroplast genome in the rice nuclear genome

Figure 3.2 shows the frequency in which each region of the chloroplast genome is represented in the nuclear genome of rice. It is clear from this figure that while most of the chloroplast genome is present to some level in the nucleus there are some areas like 16S rRNA and the NADH dehydrogenase subunits that is present at a much higher frequency than the rest of the chloroplast genome.

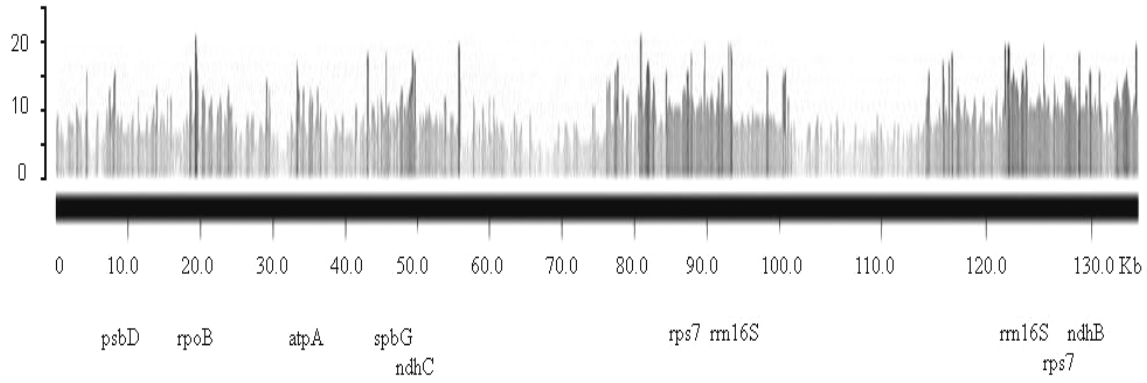

Figure 3.2: Representation of the chloroplast genome in the nuclear genome of rice. Vertical bars indicate the frequency (number of insertions) in which each area is present in the nuclear genome. Below are some of the genes that present in the areas that are highly represented. These are: *psbD* – PSII D2-protein; *rpoB* – RNA polymerase subunit B; *atpA* - H+-ATPase subunit CF1 α; *spbG* – PSII G protein; *ndhC* - NADH dehydrogenase ND3 subunit; *rps3* - 30S r-protein CS3; *rpl2* - 50S r-protein CL2; *rps7* - 30S r-protein CS7; *rrn16S* – 16 S rRNA; *ndhB* - NADH dehydrogenase ND2 subunit.

3.1.3 Chloroplast rRNA genes in the rice nuclear genome

As mentioned one of the chloroplast genome regions that is present at a high frequency in the nuclear genome is that of the rRNA genes. Thirty-nine fragments of the 16S rRNA larger than 100 bp make up 29035 bp (3.7%) of the total chloroplast homology with the nuclear genome. For the 23S rRNA gene 46 fragments and 38012 bp make up 4.8% of the total chloroplast insersions. The 5S rRNA gene is only present 26 times in fragments larger than 100 bp that

accounts for 9071 bp (1.2%). Together these account for 9.8% of the chloroplast insertions in the rice nuclear genome. Figure 3.3 shows the distribution and frequency of these rRNA subunits in the rice nuclear genome.

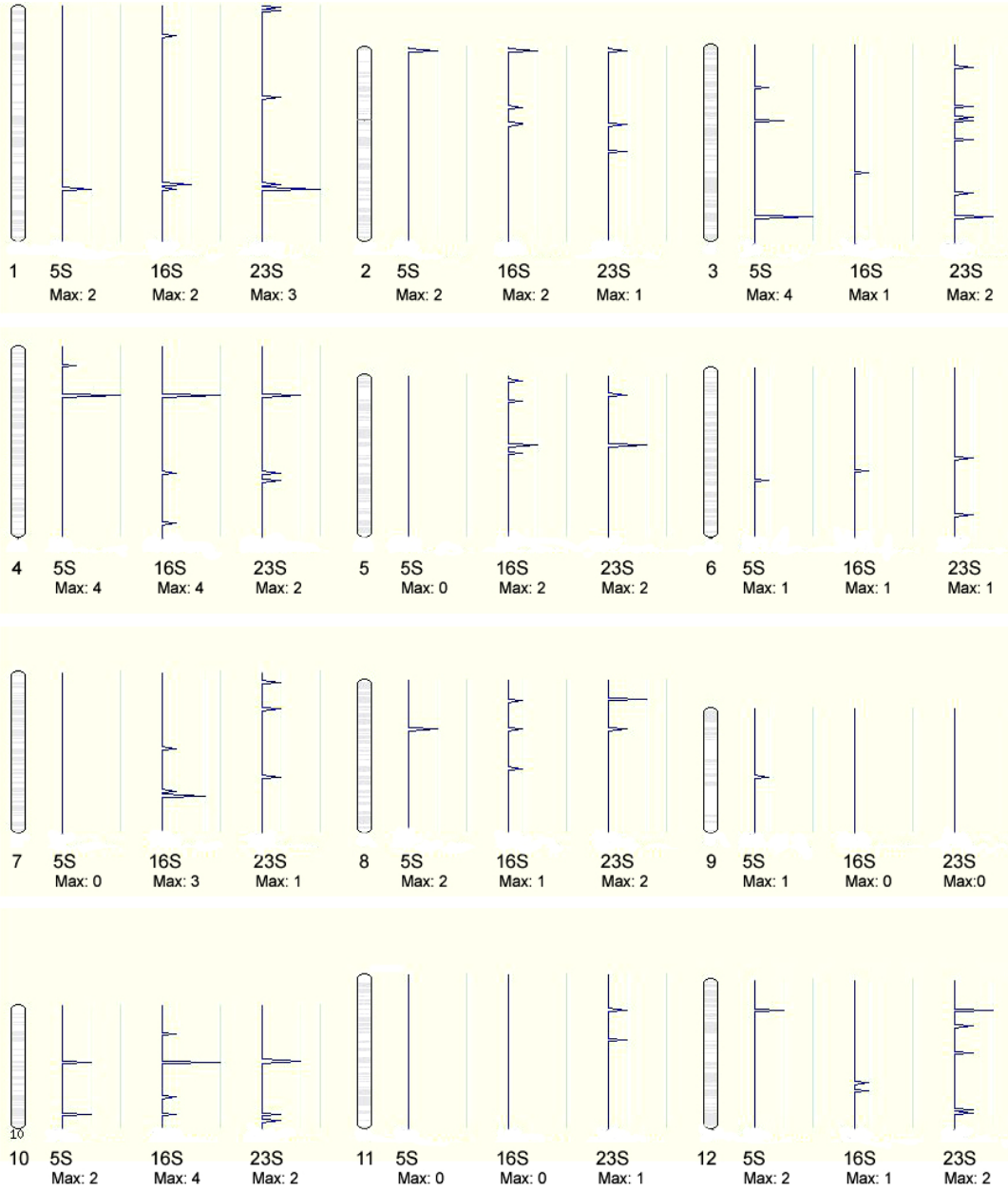

Figure 3.3: Chloroplast rRNA insertions in the nuclear genome showing the position and the density of the insertions. Max refers to the density or number of insertions in a particular region.

3.1.4 Nature of the chloroplast insertions

Though some chloroplast insertions represent large or whole chloroplast genome insertions with subsequent deletions, there are areas that seem to have arisen from the insertion of multiple fragments in the same area. Figure 3.4 shows the composition of such an area on chromosome 12 of rice. The numbers on top shows the position in relation to the chloroplast genome, measured in base pairs from the origin of replication, while the bottom numbers represent the start and end positions of the insertions on chromosome 12. The spacer sequences between the insertions were found contain a high number of thiamine (T) repeats.

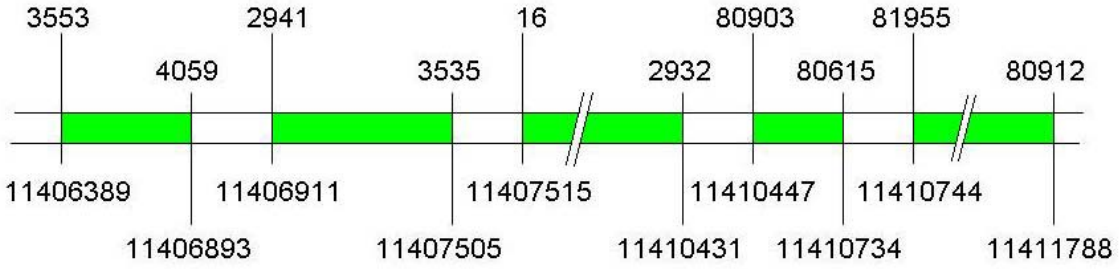

Figure 3.4: Chloroplast insertions on rice chromosome 12.

3.1.5 Mitochondrial homologies in the rice nuclear genome

Homologies between the mitochondrial and nuclear genomes were less than compared with the chloroplast genome. Table A-2 (Appendix, pp 194-264) shows the raw alignment data, including all the alignments found during the initial BLAST search. This includes fragments less than 100 pb with less than 95% homology that were discarded during further analysis.

Using the above mentioned parameters, the rice mitochondrial genome aligned to a total of 614165 bp in the nuclear genome. Figure 3.5 shows the distribution of these homologous regions over the rice karyotype. A total of 1329 alignments were used to draw the map. However as for the chloroplast, individual insertions seem to be grouped together in certain areas rather than being randomly distributed over the whole genome.

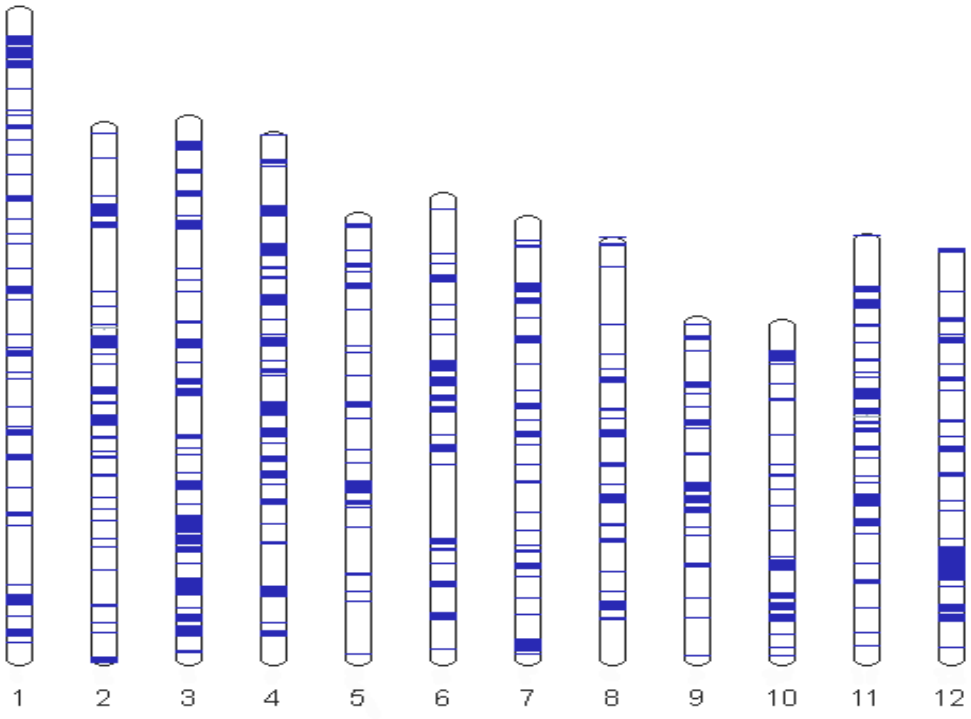

Figure 3.5: Distribution of mitochondrial homologies in the rice nuclear genome.

3.1.6 Representation of the mitochondrial genome in the rice nuclear genome

The representation profile of the mitochondrial genome on the nucleus is shown in Figure 3.6. The distribution pattern shows that only small parts of the genome is represented at high frequencies while most parts are absent from the nuclear genome. Although the NADH dehydrogenase subunits are present at high frequencies in the nuclear genome it is mostly present as fragments rather than complete gene sequences and account for 14841 bp in the nuclear genome. Figure 3.7 and 3.8 shows the frequency and position of these insertions.

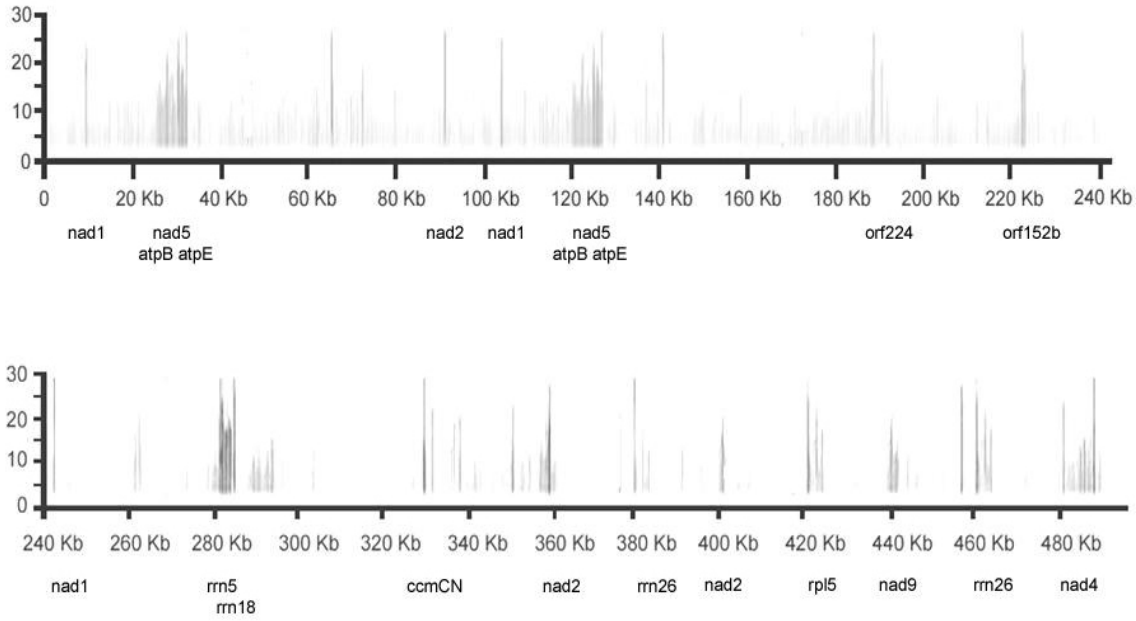

Figure 3.6: Representation of the mitochondrial genome in the nuclear genome of rice. Vertical bars indicate the frequency in which each area is present in the nuclear genome. Below are some of the features in the regions that are highly represented, like the NADH dehyrogenase subunits (nad) the ATPase subunits (atpB and atpE as well as the rRNA genes (rrn5, rrn18 and rrn26).

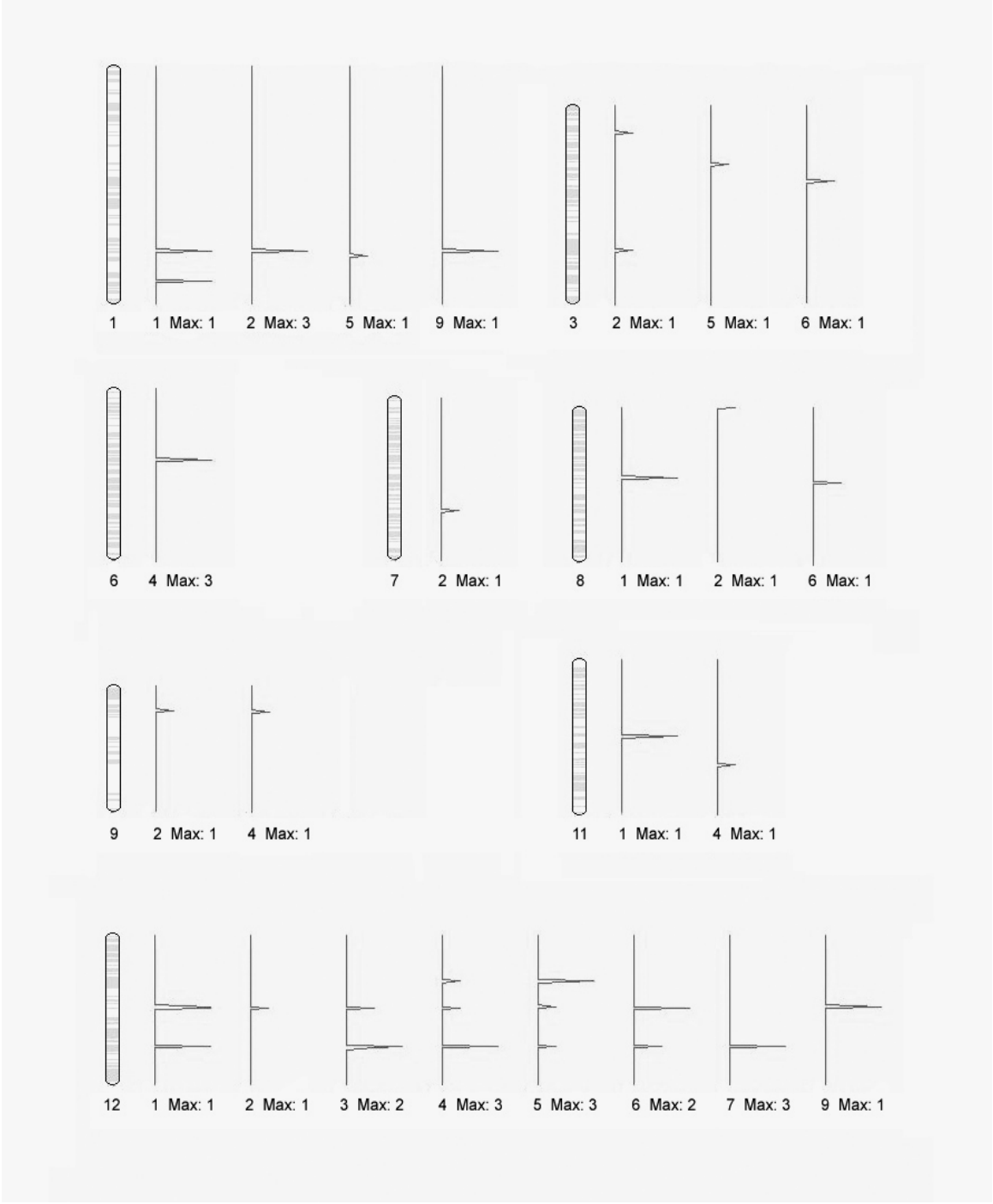

Figure 3.7: Localization of plastid NADH dehydrogenase subunits in the nuclear genome of rice. Max indicates the number of fragments in that area, the height of the peaks is also an indication of the length of the insertions.

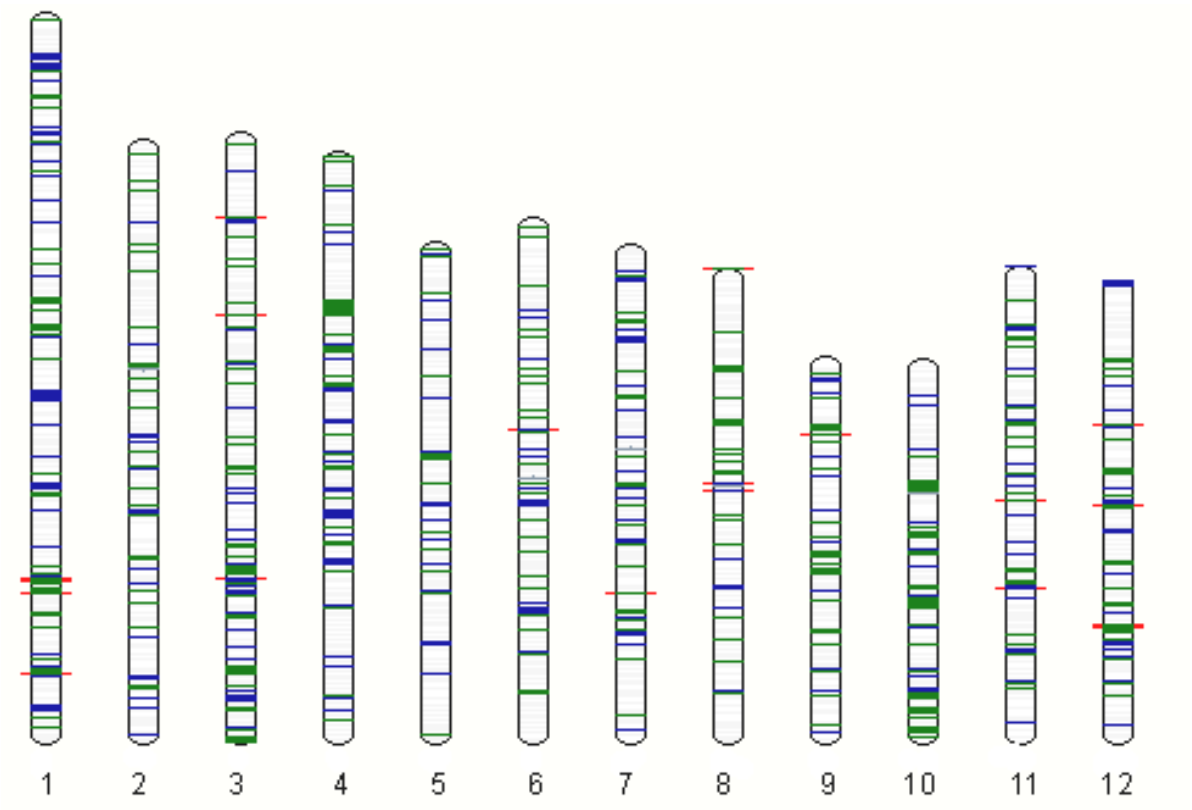

Figure 3.8: Position of the NADH subunits (red wide box) relative to the chloroplast (green) and mitochondrial (blue) insertions.

3.1.7 Mitochondrial ATPase-β genes in the nucleus

Five copies of the chloroplast ATPase CF_1 β-subunit that formed part of the chloroplast insertions in the nuclear genome were found (Chr. 1:14120102-14121600; 5:20783405-20784761; 6:23544648-23546144; 10:10349658-10351154 and 12:5610836-5612332). Two more copies of the ATPase β-subunit were found in the nuclear genome (Chr. 1:28257351-28261681 and 5:27245171-27249646). The first is located between two mitochondrial insertions (1:28097561-28097714 and 1:28361408-28361977) but the second is located well outside of any plastid DNA insertions, (Figure 3.9). One of the distinguishing features of these two copies is eight intron-like sequences that interrupt the coding sequence compared to the chloroplast copies.

A similar copy is found in the genome of *Arabidopsis thaliana* (Chr. 5:2825714-2828663; At5g08690.1). This has been well annotated in terms of intron and exon positions. The intron-like sequences of the two rice copies is located in the same positions as that of the one in *Arabidopsis*. While the exon-sequences were well conserved between the 3 copies the intron-sequences showed no homology between the three copies. Figure 3.9 shows the location of the five chloroplast ATPase CF_1 β-subunit copies and the two intron-containing copies in relation to the chloroplast and mitochondrial insertions in the nuclear genome of rice.

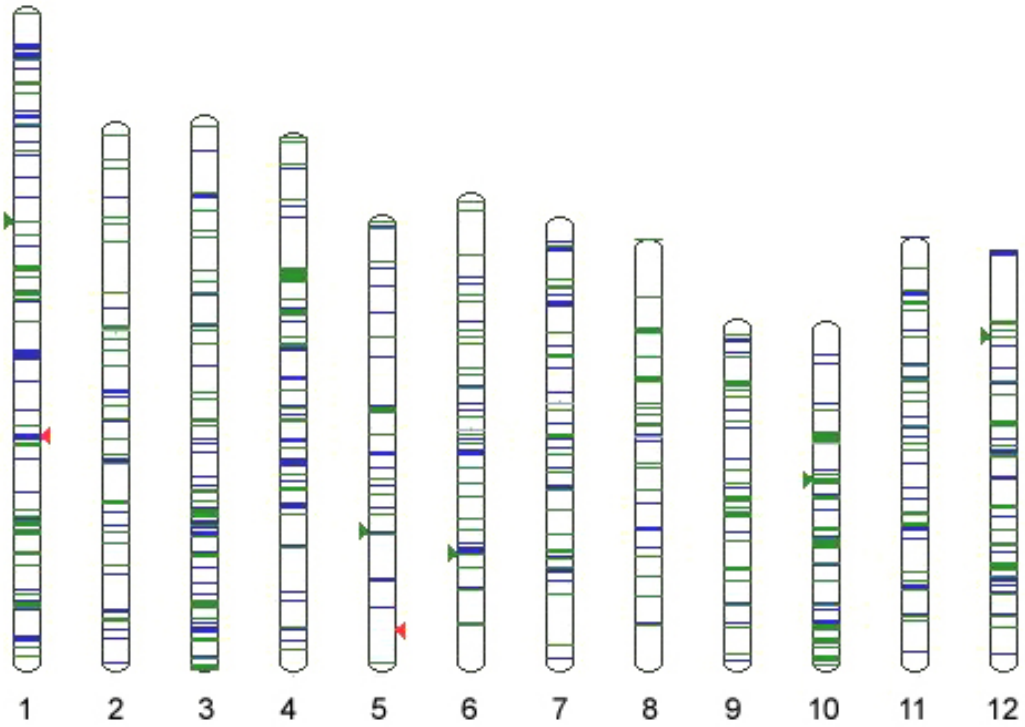

Figure 3.9: Location of the five chloroplast ATPase CF_1 β-subunit copies (green arrow, left) and the two intron-containing copies (red arrows right) in relation to the chloroplast insertions (green boxes) in the nuclear genome of rice.

When the intron sequences are removed from these copies and the translated protein sequences compared to that of the ATPase-β peptide from the chloroplast (NP_039390) as well as the mitochondrial ATPase-β (BAA01372) they show a high degree of homology with the mitochondrial peptide (Figure 3.10), and to a lesser degree with the chloroplast peptide.

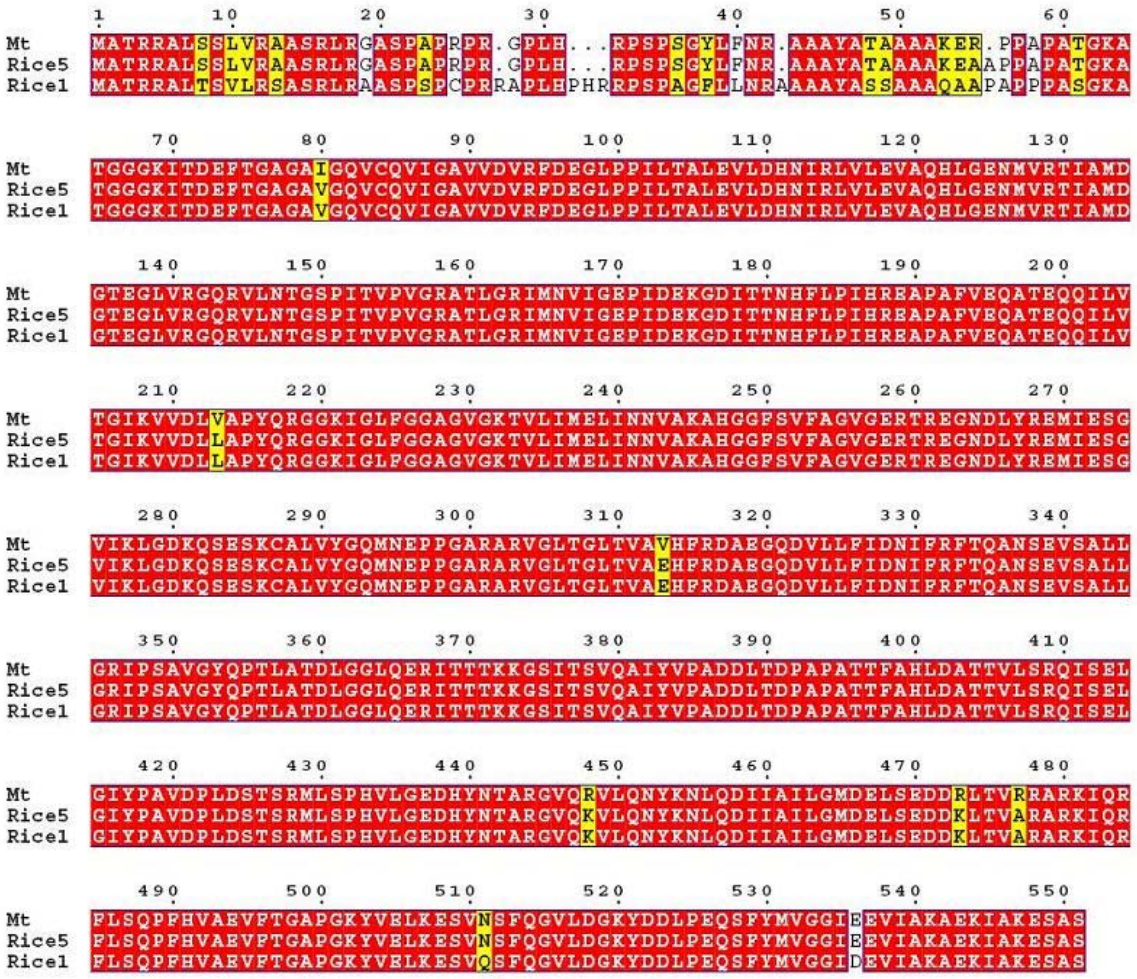

Figure 3.10: Alignment of the translated coding sequences of the two nuclear ATPase-β copies in rice (Rice1 and Rice5) and the rice mitochondrial copy (Mt).

3.1.8 Co-alignments between the chloroplast and mitochondria

Co-alignment of the chloroplast and mitochondrion accounts for some of the homology in the nuclear genome; 8997 bp of the rice chloroplast genome align to 17168 bp in the rice mitochondrial genome, corresponding mainly to the ATP synthase B and E subunits as well as to NADH dehydrodgenase subunits 1, 2 and 5. This accounts for 64453 bp of shared homology in the rice nuclear genome. Table A-3 (Appendix, pp 265-269) shows the exact positions of

these co-alignments in the rice nuclear genome. Figure 3.9 gives a summary of the chloroplast, mitochondrial and shared homology on each chromosome.

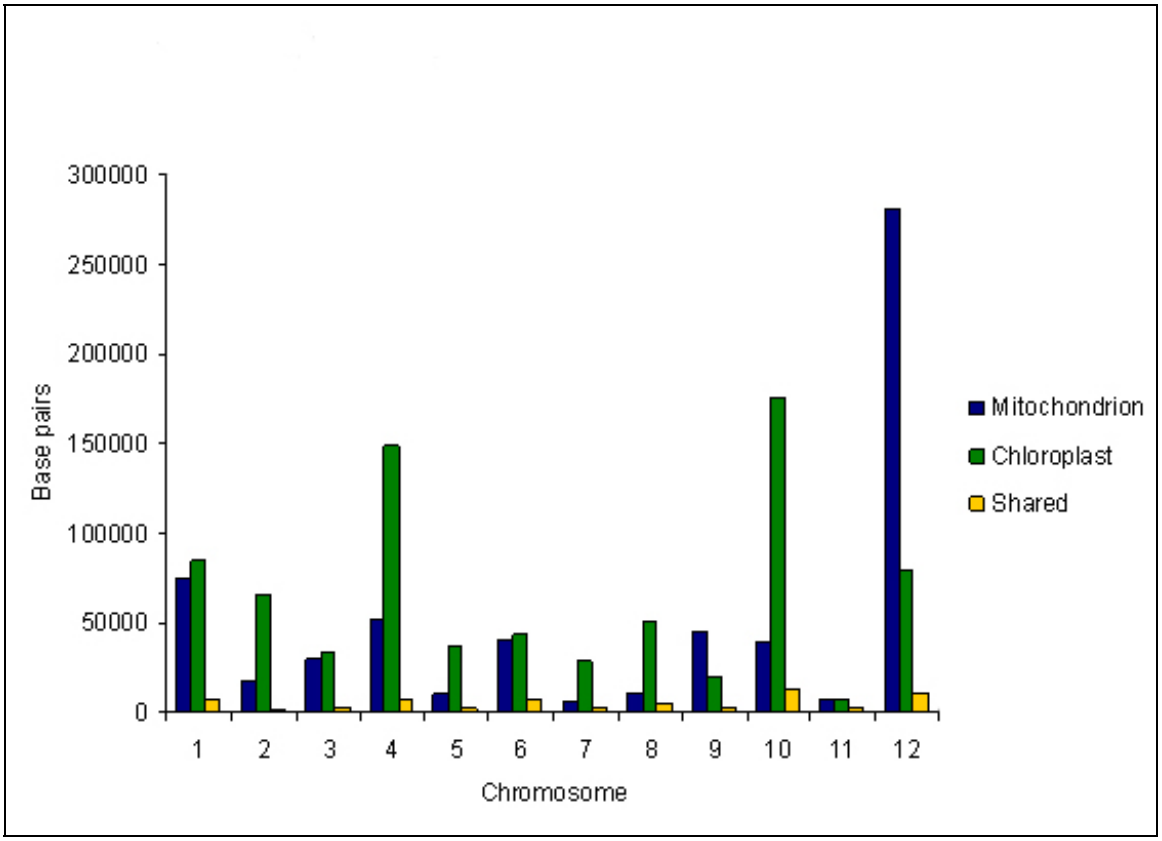

Figure 3.11: Mitochondrion, chloroplast and mitochondrion-chloroplast co-alignments on each of the rice nuclear chromosomes.

3.1.9 Size distribution of insertions

Figure 3.12 shows the size distribution of the chloroplast and mitochondrial insertions. In the filtered data set, using only alignments longer than 100 bp with a homology of 95% and greater, for both the chloroplast as well as the mitochondria the majority of fragments were between 100 - 200 bp in size with a steady decline in the number of fragments with an increase in fragment length.

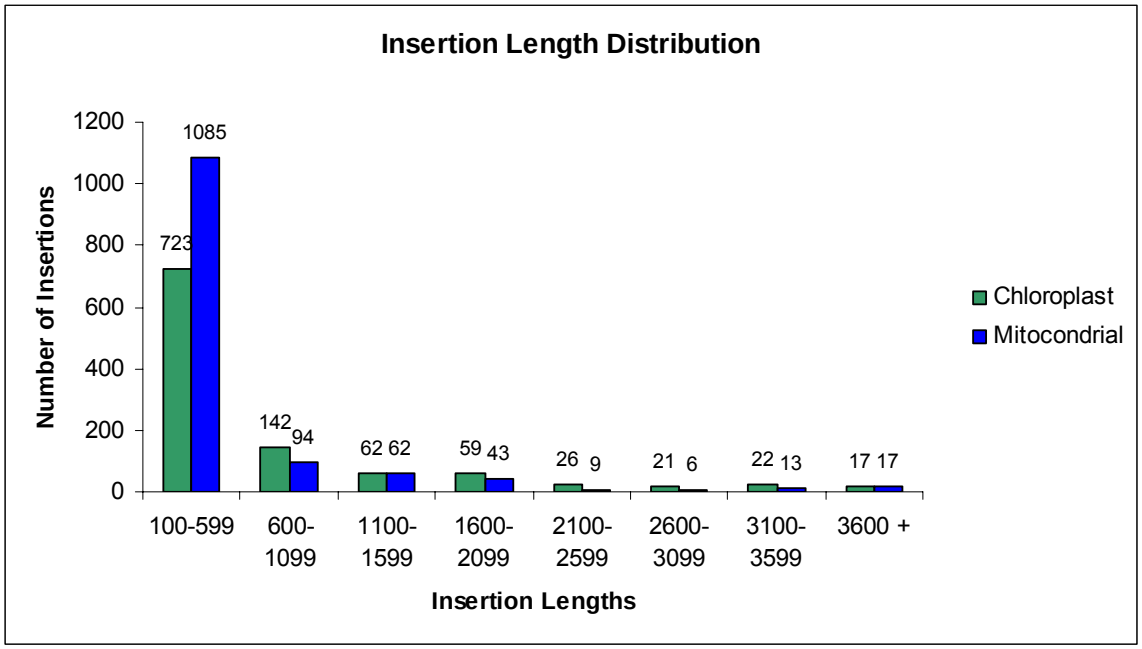

Figure 3.12: Size distribution of plastid insertions in the nuclear genome. The number of insertions in each increment is indicated on top of each bar.

3.2 Viral Comparison

Comparisons between the genomes of several rice- and grass related viruses resulted in the identification of a variety of fragments in the rice genome with distant homology to viral sequences. Figure 3.13 shows the position of the different fragments mapped on the rice karyotype. The largest number of fragments was found in the comparison between the rice genome and the Rice tungro bacilliform virus (RTBV). Previous reports on RTBV-like sequences in the rice genome identified 29 fragments (with a total length of 225 719 bp) (Kunii *et al*., 2004). In this study we identified 409 fragments between 100 and 2443 bps (with a total length of 280 912) with 56 to 71% identity covering most of the RTBV genome (Table B-1, Appendix pp 270-284). Some of the RTBV fragments are represented up to 40 times. Fragments of the RTBV virus in the rice genome include a 844 bp sequence on chromosome 11 (26839202-26839983) corresponding to TIGR Locus LOC_Os11g45180.1 which is a putative disease resistance peptide. There is also a 2436 bp fragment (8: 11221697-11223971) that falls within TIGR Locus LOC_Os08g18820.1 (novel transcript, zinc knuckle family protein). This fragment forms part of the last intron and half of the last exon of this gene in rice. Using the complete gene sequence to do a blastn search identifies the sequence in the *Oryza sativa* Japonica, Indica and

Nipponbare groups as well as in the RTB virus. A translated search (blastx) of this sequence identifies it as a hypothetical protein from RTBV (FAA00012.1) (e=0.0). A third fragment of 2322 bp (8: 11161887-11164077) that falls within TIRG Locus LOC_Os08g18220.1 is annotated as a putative unclassified retrotransposon protein. Using blastn to search the complete gene sequence gives results in *Oryza sativa* Japonica and Indica as well as in the RTB virus and blastx searches identifies the rice gene as a hypothetical protein from RTBV (FAA00009.1) (e=0.0). Using the complete genome of the RTB virus in a blastn search on the NCBI non-redundant database, the top 174 hits include only RTBV (0.0), *Vitis vinifera* (792 bp, 65%, 8e-25), *Oryza sativa Japonica* (6e-14) and Tobacco vein-clearing virus (248 bp, 69%, 3e-10).

Fifty-one additional fragments with sequence similarities between the rice genome and 5 RNA viruses were also found that has not yet been reported in the literature (Table B-2, Appendix pp 285-286). These included the Oryza rufipogon endornavirus (ORE) (NC_007649), Oryza sativa endornavirus (OSE) (NC_007647), Rice black streaked dwarf virus (RBSDV) (segment 4, NC_003735), Rice grassy stunt virus (RGSV) (segment 3, NC_002325 and segment 6, NC_002328) and the Rice stripe virus (RSV) (RNA4, NC_003753) were also found to have similarities to the rice genome. These fragments of similarity ranged between 167 to 1800 bp in length, with an average size of 460 bp and were unevenly spread among the viruses. Fragments from the Rice grassy stunt virus fragment 3 include a 1800 bp fragment on chromosome 6 (6644000-6645675) that does not fall within any expressed portion of the rice genome as well as a 350 bp fragment on chromosome 2 (21462026-21462352) that is also in a 'non-functional' portion of the rice genome and that can only be identified in *Oryza sativa* Japonica using a blast search in the plant database of NCBI. A 550 bp fragment with sequence similarity to Oryza sativa endornavirus is found within the intron of a putative potassium transporter 1 (2:18870105-18870616) in the rice genome. Using this fragment to search the plant database on NCBI only identifies this sequence fragment in *Oryza sativa* Japonica. There is also a 1034 bp fragment with similarity to Oryza rufipogon endornavirus in a non-genic region of chromosome 12 (13977633-13978583) that is also only found within *Oryza sativa* Japonica.

A genome comparison were also done between the rice genome and 8 other monocot-related viral genomes (4 ssDNA and 4 ssRNA) namely the Wheat dwarf virus (NC_003326); Maize streak virus (NC_001346); Sugarcane streak virus (NC_003744); Panicum streak virus (NC_001647); Wheat streak mosaic virus (NC_001886); Wheat eqlid mosaic virus (NC_009805); Sorghum mosaic virus (NC_004035) and the Maize dwarf mosaic virus

(NC_003377). Even with low stringency parameters no significant sequence similarities were found between these genomes and the rice genome

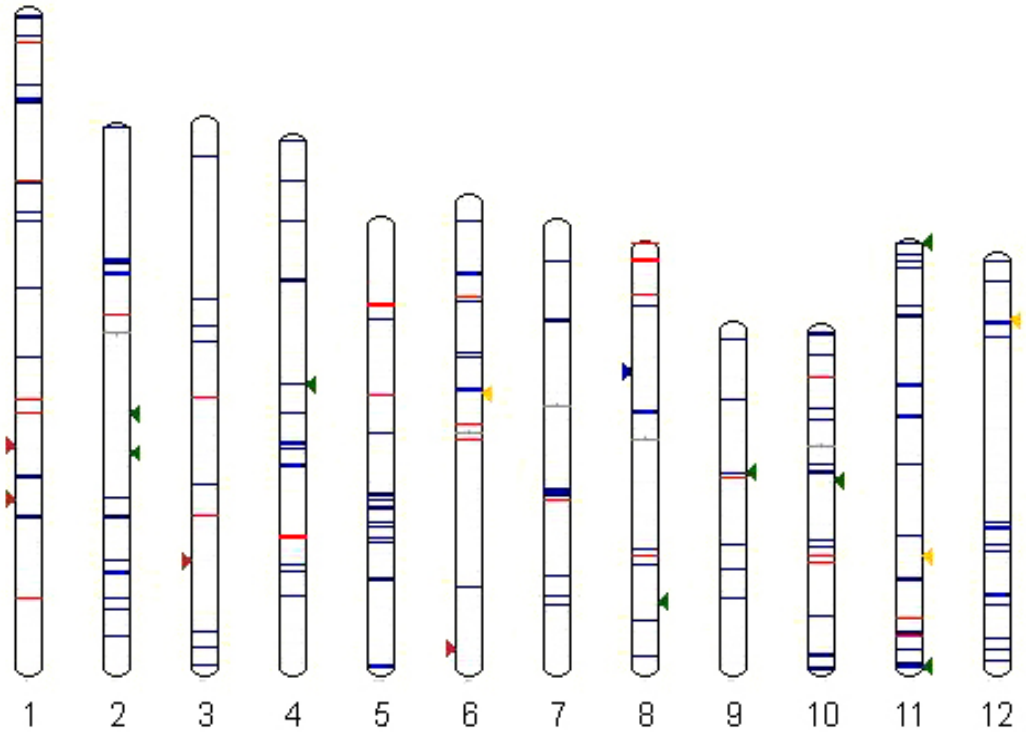

Figure 3.13: Positions of the viral homologies in rice. Oryza rufipogon endornavirus (blue left hand arrow); Oryza sativa endornavirus (green right hand arrows); Rice black streaked dwarf virus (brown left hand arrows); Rice stripe virus (yellow right hand arrows); Rice grassy stunt virus (red filled box) and RTBV (blue filled box).

3.3 Bacterial comparison

3.3.1 *Bacillus* sequence similarities in the rice nuclear genome

Comparisons between the genomes of *Bacillus cereus* and that of rice resulted in the identification of a great number of regions with sequence similarity. Limiting the results only to fragments 100 bp or longer with an e-value equal smaller than 10^{-20} resulted in 74 fragments totaling 12123 bp. The positions of these fragments are shown in figure 3.14.

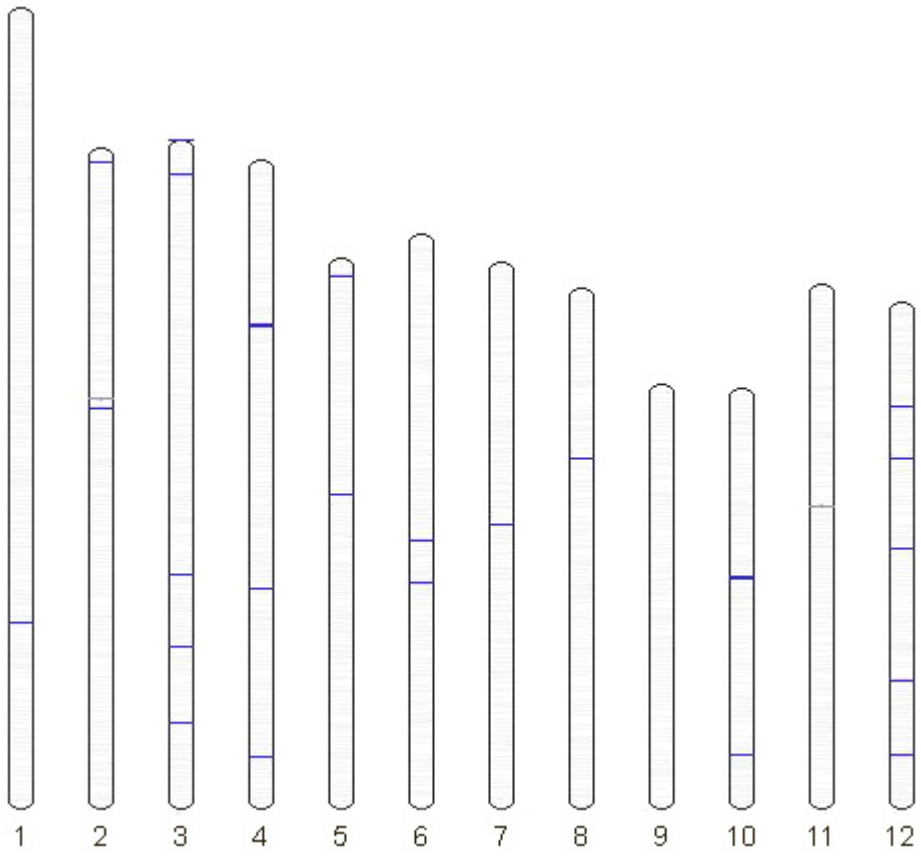

Figure 3.14: Distribution of *Bacillus cereus* alignments on the rice karyotype.

The majority of alignments represent repeats in the *Bacillus* aligning to the same position in the rice genome. The figure of 12123 bp counted regions in rice regardless of the number of alignments in the *Bacillus* genome. Table C-1 (Appendix, pp 287- 306) shows the alignments found in rice.

Taking the results found with *Xanthomonas* and *Pseudomonas*, discussed below into account, and eliminating cross-alignments between these and *Bacillus*, 39 of the 74 fragments totaling 6139 bp were unique between *Bacillus* and rice.

All but one of the alignments between the *Bacillus* genome and the rice genome were within the 23S and 16S rRNA subunits of *Bacillus* and the chloroplast subunits located in the nuclear genome of rice. The cross-homology between the *Bacillus*, *Xanthomonas* and *Pseudomonas* alignments also occurred in this regions. Comparing the 23S subunits of the three bacteria and the chloroplast subunit found a higher degree of homology between the rice chloroplast and *Bacillus* 23S subunits that between *Bacillus* and *Xanthomonas* or *Pseudomonas*. The one fragment found exclusively between *Bacillus* and rice genome (excluding the rRNA fragments), which also had the highest similarity of the fragments identified between rice and *Bacillus* (0.0) was a fragment of 770 bp (3:1817627-1818395) in rice showed 100% (0.0e) sequence similarity with part of the HD domain protein from *Bacillus* (AE017194.1). This region is within the known gene in the rice genome TB2_DP1_HVA22; TB2/DP1, HVA22 family. This family includes members from a wide variety of eukaryotes. It is interesting to note that the similarity with this fragment starts in an intron, includes 2 exons and then terminates in a longer intron. Using Blastn on the NCBI website the two top results (100%) is in rice and *Bacillus* followed by alignments 99% and less in various organisms, including Arabidopsis thaliana. Limiting the results just to plants using the NCBI database the only two plants with significant sequence similarities are rice and Arabidopsis, but with 8 different annotations in Arabidopsis. Using the Gramene database, no results are found with Arabidopsis but 5 fragments varying between 656 to 143 bp are found within the maize genome, not shown in searches using the NCBI database. Using Blastx the translated protein sequence gave significant matches in various bacteria, limiting the search to eukaryotes resulted in only one hit (AAN17394.1, $7e^{-88}$) in *Oryza sativa cv. japonica* annotated as putative *InsB* from *Escherichia coli*.

Various other alignments with different regions in the rice genome not related to the rRNA subunits were also found with the initial alignments but were either less than 100 bp long or had e-values lower than e^{-20}.

3.3.2 *Xanthomonas* sequence similarities in the rice nuclear genome

Comparisons between the genomes of *Xanthomonas oryzae pv. oryzae* and that of rice resulted in the identification of 37 homologous regions 100 bp or longer with an e-value equal smaller than 10^{-20} totaling 5673 bp. The positions of these fragments are shown in figure 3.15. As with the *Bacillus* alignments most alignments represented repeats in the bacterial genome aligning to the same positions in the rice genome, once again the figure of 5673 bp counted regions in rice regardless of the number of alignments in the *Xanthomonas* genome. Table C-2 (Appendix, pp 307-309) shows all the alignments found in rice. Accounting for cross-alignments with the other two bacterial genomes tested, 5 of the 37 fragments totaling 698 bp were unique alignments between *Xanthomonas* and rice. Cross-alignments between *Bacillus* and *Xanthomonas* all occurred in the 16 and 23S sequences, which are discussed above under the *Bacillus* results.

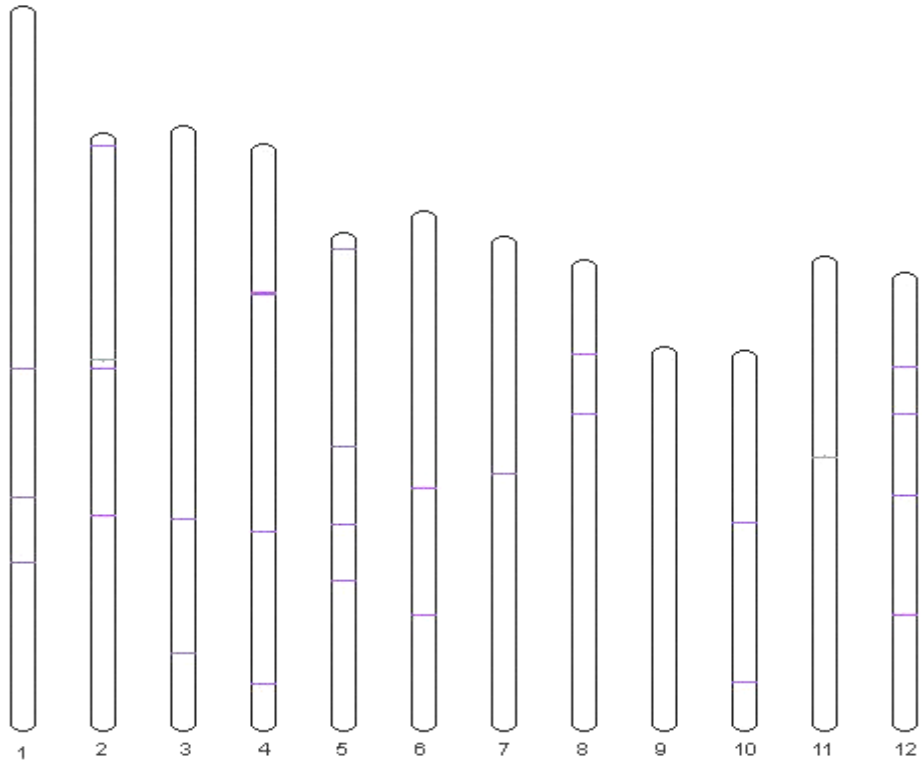

Figure 3.15: Distribution of *Xanthomonas* alignments on the rice karyotype.

Sequences unique to *Xanthomonas* include a 145 bp fragment (AE013598.1) in *Xanthomonas* that aligned to chromosome 1: 21872692-21872833 in rice ($1.4e^{-33}$) that is part of the carbamoyl-phosphate synthase large subunit in *Xanthomonas*; a 147 bp fragment (AE013598.1; elongation factor Tu) aligning to 2: 23057548 – 23057691 in rice ($2.3e^{-32}$); and 3 alignments with the molecular chaperone DnaK (AP008229.1) of 162 bp (5:17472623-17472780; $6.4e^{-22}$), a 101 bp fragment (5:20846967-20847064; $4.5e^{-23}$) and a 143 bp fragment (8:5640788-5640925; $1.6e^{-22}$). Using the 100 bp fragments on each side of the of the 3 alignments resulted in hits only in rice with the first two sequences but using the fragment from chromosome 8 plus 100 bp flanking on each side (8: 5640688 – 5641025), resulted in significant alignments in *Scherffelia dubia* (AJ312020.1|SDU312020; $5.0e^{-34}$) partial mRNA for luminal binding protein, Triticum aestivum (AF005993; $2.0e^{-30}$) 70 kDa heat shock protein, *Lilium longiflorum* (D21824.1|LILLIM18; $8.0e^{-30}$) as well as Sorghum bicolor (AY503363.1; $7.0e^{-27}$). The best alignment with this extended fragment other than rice was with the fungal *Blastocladiella emersonii* heat shock protein (hsp70) gene (L22497.1|BSIAHSP70X; 2e-58).

The only other sequence similarities found between *Xanthomonas* and rice of were 3 fragments of which 2 were also shared with *Pseudomonas* that are annotated as the glycine dehydrogenase gene in the bacterial genomes, aligning to these 3 positions in rice respectively 1:29540592-29540764 (177 bp; $5.7e^{-37}$) (1:29540592-29540767; 181 bp; $3.8e^{-26}$ with *Pseudomonas*); 6:24360517-24360642 (128 bp; $1.3e^{-26}$) (6:24360464-24360643; 185 bp; $5.9e^{-31}$ with *Psuedomonas*) and 8:9239422-9239630 (217 bp; $1.5e^{-52}$).

3.3.3 *Pseudomonas* sequence similarities in the rice nuclear genome

Comparisons between the genomes of *Pseudomonas syringae pv. syringae* and that of rice resulted in the identification of 33 fragments with significant sequence similarity 100 bp or longer with an e-value equal smaller than 10^{-20} totaling 6332 bp. The positions of these fragments are shown in figure 3.16. As discussed in the results above there was substantial cross-homology between the bacterial sequences that aligned with the rice genome. Excluding these the *Pseudomonas* genome had 655 bp of unique similarity with that of the rice genome. This 655 is made up of three fragments of the 23S gene a 156 bp fragment on chromosome 2 in rice (2: 815417 – 815567; $6.2e^{-40}$) and a smaller piece of the same fragment on chromosome 11 (11: 13074159 – 13074260; 9.0 e^{-21}).

A 128 bp fragment (CP000075.1) that is part of the *Pseudomonas* catalase gene with 87% (7e-60) sequence similarity to rice (3:1767897-1768024). The region in the rice genome containing this fragment falls close to the end of the TIGR Locus LOC_Os03g03910 catalase-1 (86 bases past the end of the gene) and prior to the TIGR Locus LOC_Os03g03920 ubiquilin-1 gene. However, when the rice catalase gene was used in Blastx against the *Pseudomonas* genome sequence a region of 374 amino acids (Identities = 201/374, Positives = 260/374) with an e value of e-115 was identified. When the rice catalase gene is used in a Blastn against the nr database, then the first non-plant homology that is identified is the *Pseudomonas* catalase gene fragment. In fact the *Pseudomonas* similarity is greater than that for *Arabisopsis*.

Table C-3 (Appendix, pp 310 - 313) shows the detailed search results.

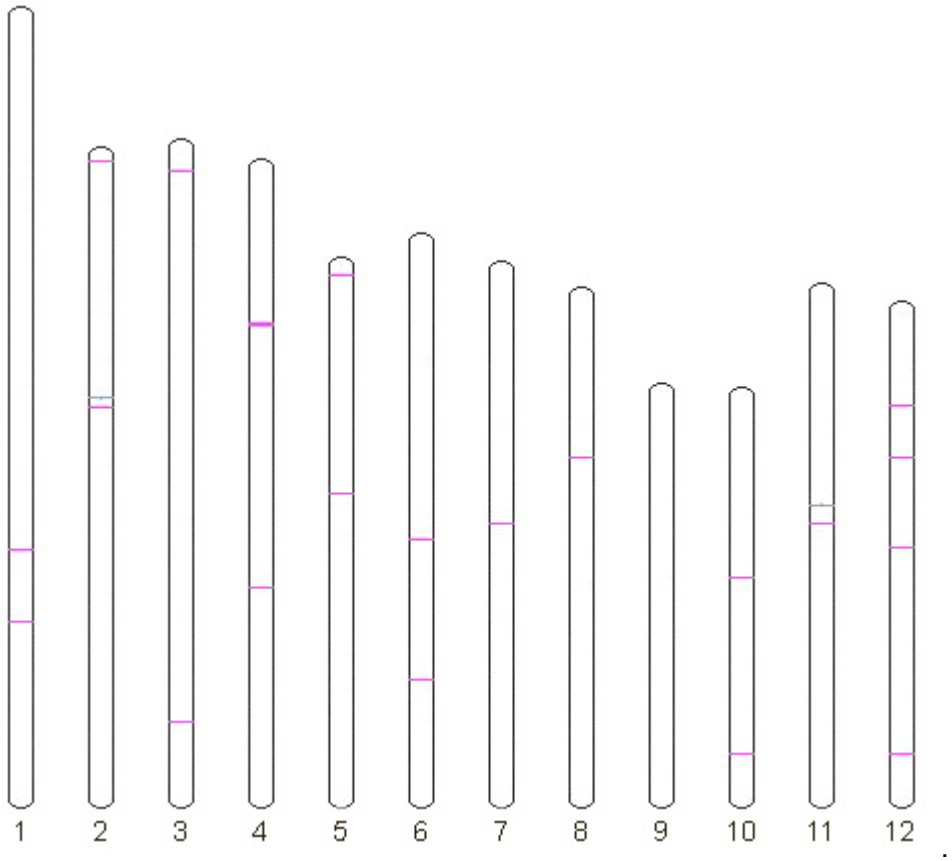

Figure 3.16: Distribution of *Pseudomonas* alignments on the rice karyotype.

3.4 Fungal Comparison

3.4.1 *Magnaporthe* sequence similarities in the rice nuclear genome

Comparisons between the genomes of *Magnaporthe grisea* and that of rice resulted in the identification of 46 fragments (totaling 8274 bp) in the *Magnaporthe* genome and 144 fragments (totaling 25033 bp) of in the rice genome of 100 bp and longer that share significant sequence similarity, with e-values equal smaller than 10^{-20}. Figure 3.17 show the position of these regions on the rice karyotype. Table D-1 (Appendix, pp 314 - 320) shows al the rests of the comparison between rice and *Magnaporthe* that were 100 bp and longer, including repeats that were subsequently removed in further analysis.

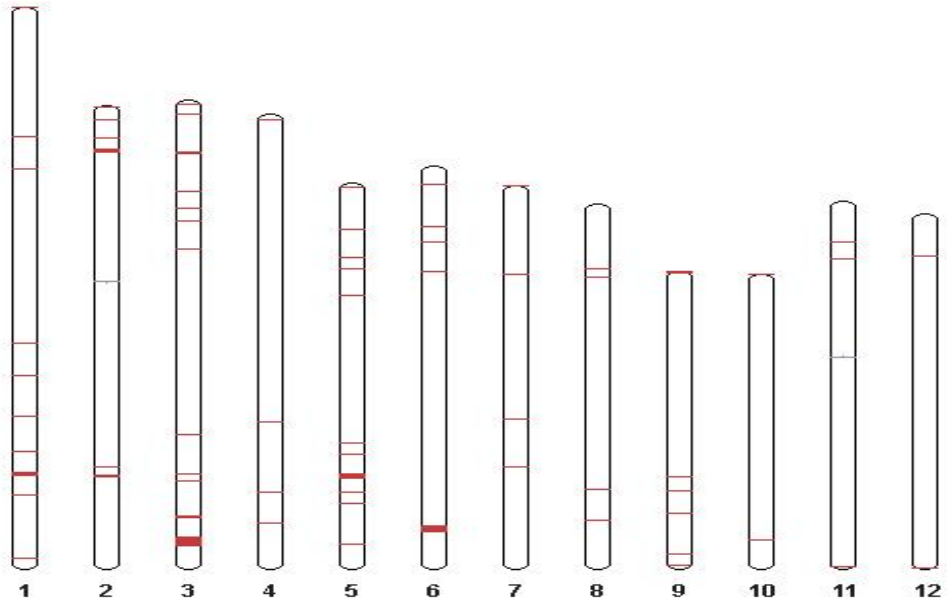

Figure 3.17: Distribution of *Magnaporthe grisea* alignments on the rice karyotype.

There were no co-alignments between the *Magnaporthe* genome and any of the nuclear plastid sequences or with the bacteria and viral sequences. Ninety one percent of the of the alignments were between conserved regions of genic sequences such as beta-tubulin and polyubiquitin while nine percent were either in unknown or *Magnaporthe* related sequences. When the rice sequences identified as similar were used in the reverse blast, these similarities were greatest between rice or monocots and *Magnaporthe*, but there was also wide homology

across different fungi and plants. Some insertions showed significant sequence similarity only to *Magnaporthe grisea* such as the 131 bp insertion on chromosome 5 (5775966 - 5776097) to locus XM_370434.1 of *Magnaporthe grisea* ($9e^{-27}$; 87% identical without gaps), this falls within the TIGR Locus LOC_Os05g10630.1 (putative O-sialoglycoprotein endopeptidase). Another such fragment was a 109 bp sequence on chromosome 4 (23876766-23876857) to locus XM_365128.1 ($9e^{-11}$; 83% identical without any gaps) although the latter of these was eliminated from the final statistics of similarity due to the high e-value. Another fragment on rice chromosome 6: 4668751-4669004 shows significant sequence similarity only in maize (AY104186.1; 8e-63) and the fungus *Coprinopsis cinerea okayama* (XM_001833442.1; 4e-31). In rice this forms part of TIGR Locus LOC_Os06g09290.1 (putative 26S protease regulatory subunit 7).

Multiple sequence alignments using the best fungal and plant related alignments found in Blastn searches on the NCBI database showed three patterns in the sequence dendogram trees as described below.

3.4.2 Dendogram type 1:

This sets of sequences group the fungal sequences together and the plant sequences together. In figures 3.18 and 3.19 *Magnaporthe grisea* groups closest to the plants from the set of fungal sequences used, while in figure 3.20 *Blastocladiella emersonii* is the closest grouping fungi.

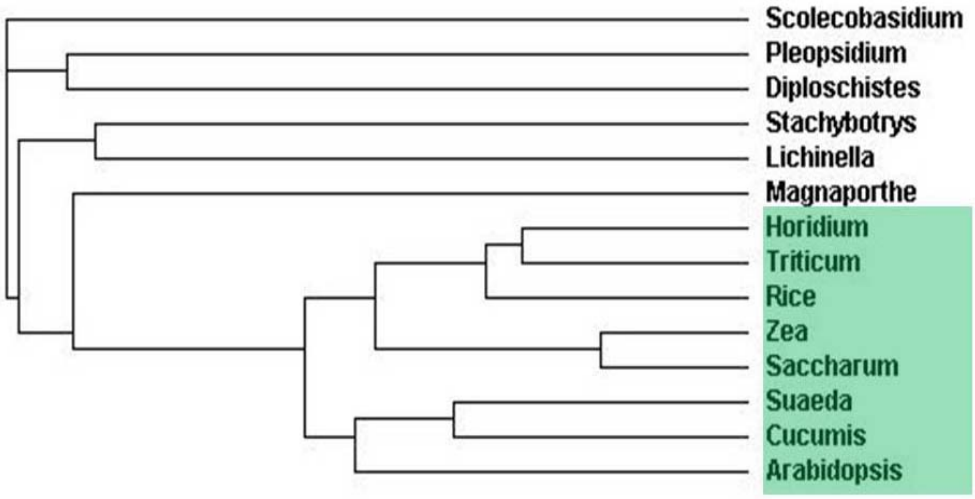

Figure 3.18: Sequence dendogram from a multiple sequence alignment with the rice fragment 3:4074435:4074559 compared to *Hordeum vulgare* (emb|Z50789.1|HVEF1ALFA); *Triticum aestivum* (gb|M90077.1|WHTTEF1X); Zea mays (gb|BT016514.1); *Saccharum officinarum* (gb|AF281361.1); *Suaeda japonica* (dbj|AB073630.1); *Arabidopsis thaliana* (ref|NM_125432.3); *Cucumis sativus* (gb|EF446145.1); *Scolecobasidium humicola* (gb|DQ307355.1); *Magnaporthe grisea* (ref|XM_361098.1); *Stachybotrys chlorohalonata* (gb|AY180269.1); *Lichinella iodopulchra* (gb|DQ832327.1); *Pleopsidium chlorophanum* (gb|DQ782920.1) and *Diploschistes ocellatus* (gb|DQ366251.1). Green blocks indicate the plant sequences.

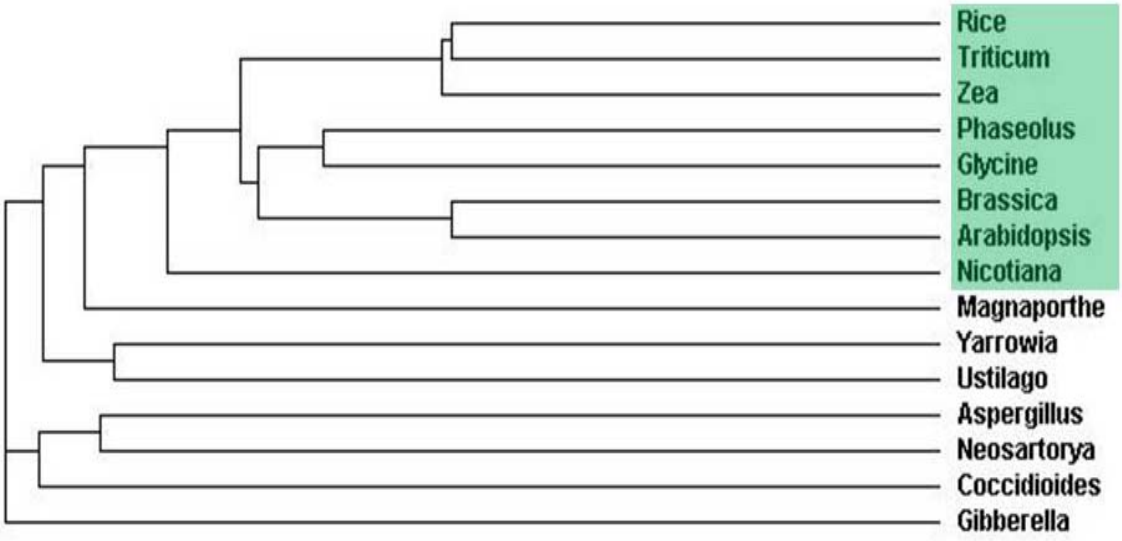

Figure 3.19: Sequence dendogram from a multiple sequence alignment with the rice fragment 10:65974:66122 compared to *Triticum aestivum* (gb|DQ150097.1); Zea mays (gb|AY106100.1); *Phaseolus vulgaris* (gb|AY007525.1); *Glycine max* (gb|DQ139265.1); *Nicotiana tabacum* (gb|AY532656.1); *Brassica rapa* (gb|AC189362.1); *Medicago truncatula* (gb|AC145022.29); *Arabidopsis thaliana* (ref|NM_129380.2); *Magnaporthe grisea* (ref|XM_369484.1); *Aspergillus terreus* (ref|XM_001216938.1); *Gibberella zeae* (ref|XM_385211.1); *Neosartorya fischeri* (ref|XM_001259876.1); *Yarrowia lipolytica* (ref|XM_502217.1); *Ustilago maydis*(ref|XM_754878.1); *Coccidioides immitis* (ref|XM_001242284.1) and *Ashbya gossypii* (gb|AE016815.3). Green blocks indicate the plant sequences.

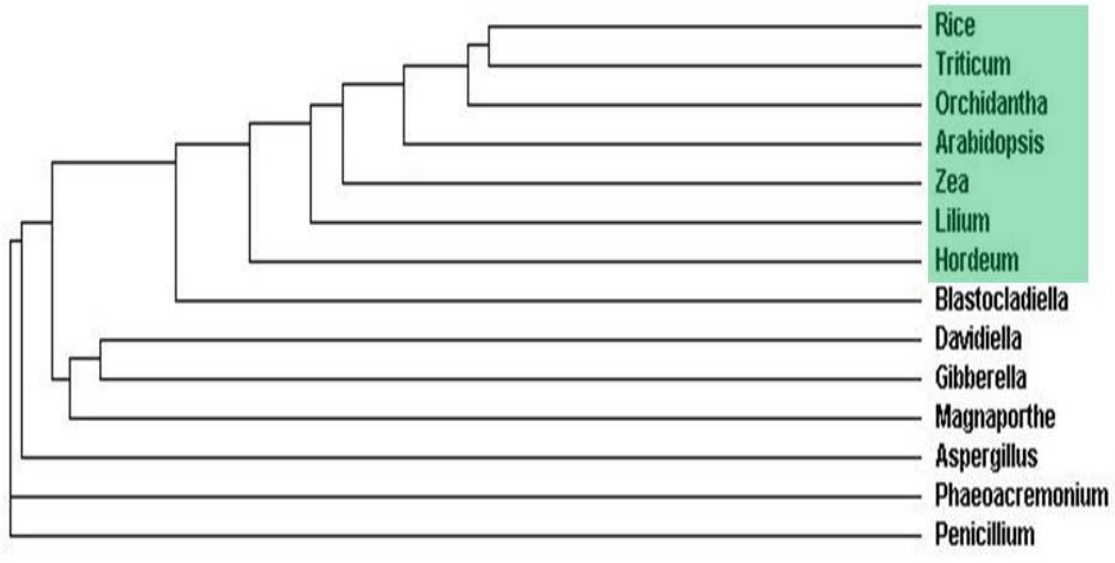

Figure 3.20: Sequence dendogram from a multiple sequence alignment with the rice fragment 3:11502555 - 11502670 compared to *Triticum aestivum* (gb|U49105.1|TAU49105); *Hordeum vulgare* (gb|DQ995512.1); *Zea mays* (gb|BT017909.1); *Orchidantha sp*. (gb|AF430094.1); *Arabidopsis thaliana* (ref|NM_123137.3); *Lilium longiflorum* (emb|Z12839.1|LLCALMOD); *Aspergillus terreus* (ref|XM_001210822.1); *Blastocladiella emersonii* (gb|AF264065.1); *Phaeoacremonium australiense* (gb|AY579280.1); *Davidiella tassiana* (gb|DQ289831.1); *Penicillium roseopurpureum* (gb|AY678541.1); *Gibberella zeae* (ref|XM_382067.1) and *Magnaporthe grisea* (ref|XM_370387.2). Green blocks indicate the plant sequences.

5.3.3 Dendogram type 2:

In this set of dendograms rice sequences group with the fungal sequences. In figures 3.21 and 3.22 the rice sequence group with the fungal sequences and closest to the *Magnaporthe* sequence. In Figure 3.23 Rice as well as maize group with the fungi based on sequence similarity, while the grass species *Setaria italica* group with the other plant sequences used in the alignment. In figure 3.24 the sequence from rice japonica cultivar groups with that of the other plants while the sequence from the indica cultivar group closer to *Magnaporthe* and the other fungi.

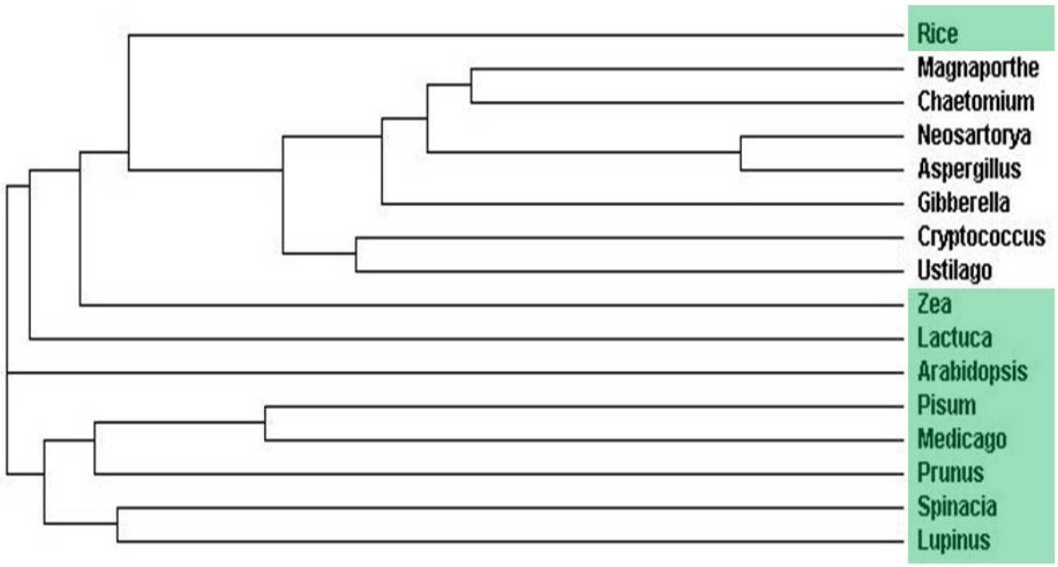

Figure 3.21: Sequence dendogram showing the relationship between a DNA fragment from a multiple sequence alignment with the rice fragment 6:4668751 – 4669004 compared to *Arabidopsis thaliana* (ref|NM_104252.2); *Zea mays* (gb|AY104186.1); *Pisum sativum* (gb|AY623108.1); *Lactuca sativa* (gb|AY243359.1); *Prunus persica* (gb|AF041258.1 AF041258); *Lupinus albus* (gb|DQ118122.1); *Spinacia oleracea* (dbj|D86121.1|SPI26SPAS); *Medicago truncatula* (gb|AC135796.25); *Aspergillus fumigates* (ref|XM_749738.1); *Neosartorya fischeri* (ref|XM_001263659.1); *Cryptococcus neoformans* (gb|AE017347.1); *Gibberella zeae* (ref|XM_380735.1); *Magnaporthe grisea* (ref|XM_363655.1); *Chaetomium globosum* (ref|XM_001220658.1); *Ustilago maydis* (ref|XM_751676). Green blocks indicate the plant sequences.

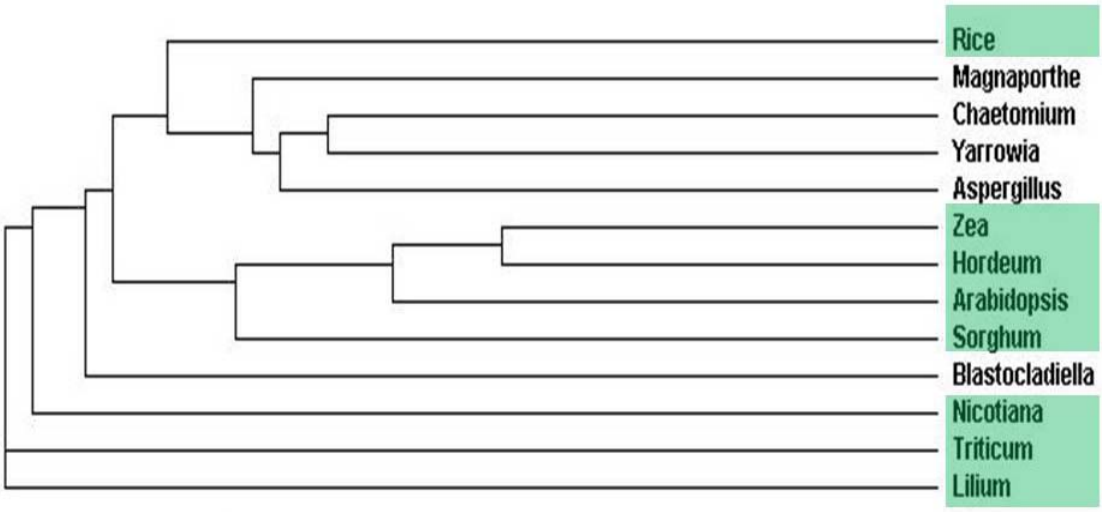

Figure 3.22: Sequence dendogram showing the relationship between DNA fragments from a multiple sequence alignment with the rice fragment 8: 5640822 – 5640946 compared to *Triticum aestivum* (gb|AF005993.1|AF005993); *Zea mays* (gb|U58209.1|ZMU58209); *Hordeum vulgare* (gb|L32165.1|BLYHSPHAA); *Lilium longiflorum* (dbj|D21824.1|LILLIM18); *Sorghum bicolour* (gb|U41653.1|SBU41653); *Nicotiana tabacum* (gb|AY372070.1); *Arabidopsis thaliana* (ref|NM_180788.2); (gb|L22497.1); *Chaetomium globosum* (ref|XM_001227132.1); *Aspergillus kawachii* (gb|AF183893.1); *Yarrowia lipolytica* (ref|XM_503913.1) and *Magnaporthe grisea* (ref|XM_361717.1). Green blocks indicate the plant sequences.

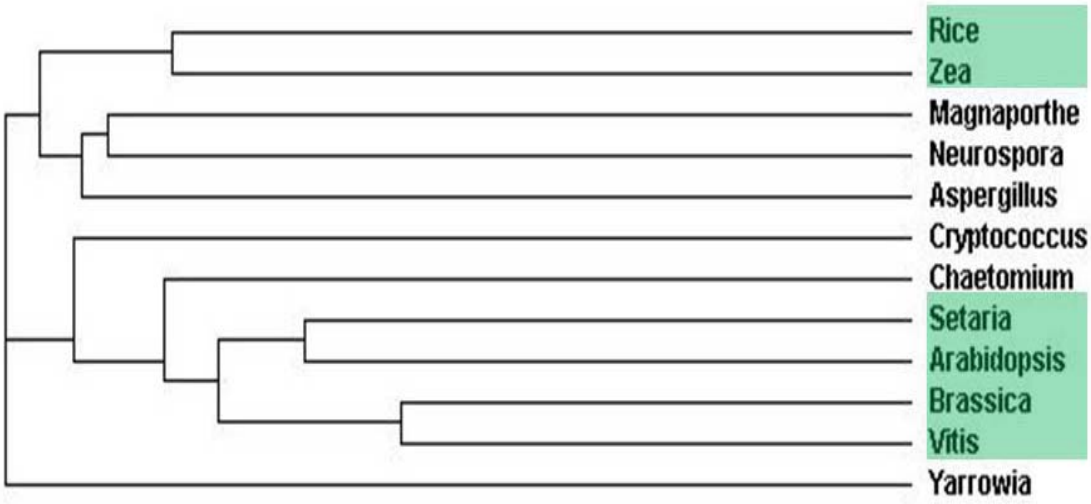

Figure 3.23: Sequence dendogram showing the relationship between DNA fragments from a multiple sequence alignment with the rice fragment 6: 5884317:5884417; *Zea mays* (gb|AY921640.1); *Setaria italica* (gb|DQ393721.1); *Brassica rapa* (gb|AC189331.1); *Vitis vinifera* (emb|AM437355.2); Arabidopsis *thaliana* (emb|BX816218.1|CNS0ABRK); *Magnaporthe grisea* (ref|XM_362751.2); *Aspergillus terreus* (ref|XM_001217694.1); *Cryptococcus neoformans* (ref|XM_572706.1); *Chaetomium globosum* (ref|XM_001223430.1); *Neurospora crassa* (ref|XM_959546.1) and *Yarrowia lipolytica* (emb|CR382127.1). Green blocks indicate the plant sequences.

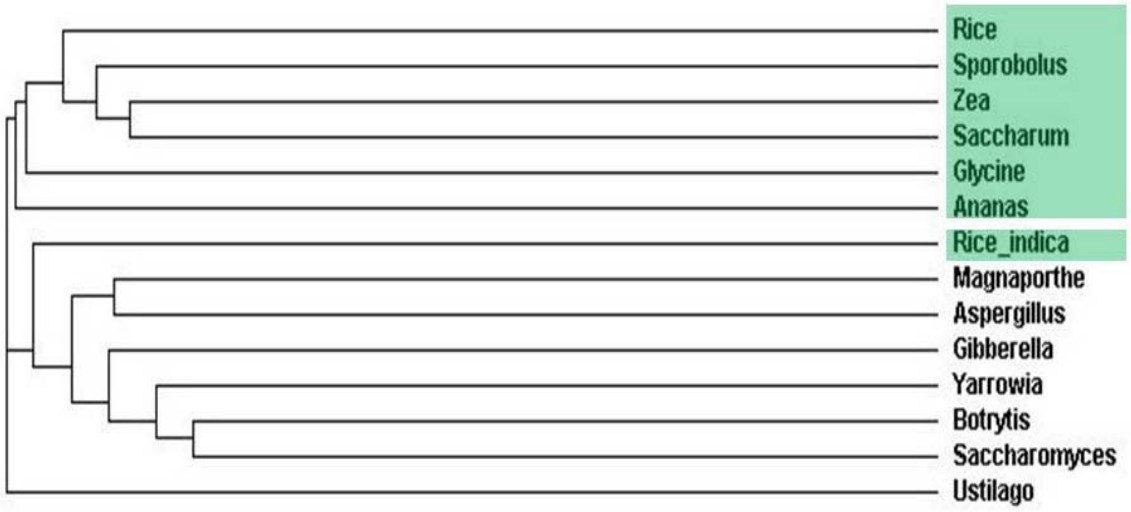

Figure 3.24: Sequence dendogram showing the relationship between DNA fragments from a multiple sequence alignment with the rice fragment 2: 3341370: 3342015; *Oryza sativa* indica cultivar-group (emb|CT837820.1|); *Sporobolus stapfianus* (gb|AF148448.1|AF148448); *Zea mays* (gb|BT016652.1); *Saccharum hybrid* (gb|L41658.1|SCFPOLY); *Ananas comosus* (gb|AY098526.1); *Ustilago maydis* (ref|XM_753127.1); *Glycine max* (dbj|D16248.1|SOYUBI); *Magnaporthe grisea* (ref|XM_363356.1); *Aspergillus niger* (ref|XM_001401978.1); *Gibberella zeae* (ref|XM_388944.1); *Yarrowia lipolytica* (ref|XM_504128.1); *Botrytis cinerea* (emb|AL114489.1|CNS01BOH) and *Saccharomyces cerevisiae* (emb|X05731.1). Green blocks indicate the plant sequences.

5.3.4 Dendogram type 3:

This set of dendograms is characterized by the fact that there seems to be no or little evolutionary order present in the grouping of the species using these sequences. In figure 3.25 rice and *Magnaporthe* group closer than rice and sorghum or maize which groups with the fungi *Chaetomium globosum*. The other cerials wheat and barley group separately while Arabidopsis, grape vine and cotton group with the fungi *Aspergillus terreus* and *Ustilago maydis*. In Figure 3.26 the fungi *Blastocladiella emersonii* interrupts the plant sequences while Nicotiana and Brassica interrupts the fungal sequences.

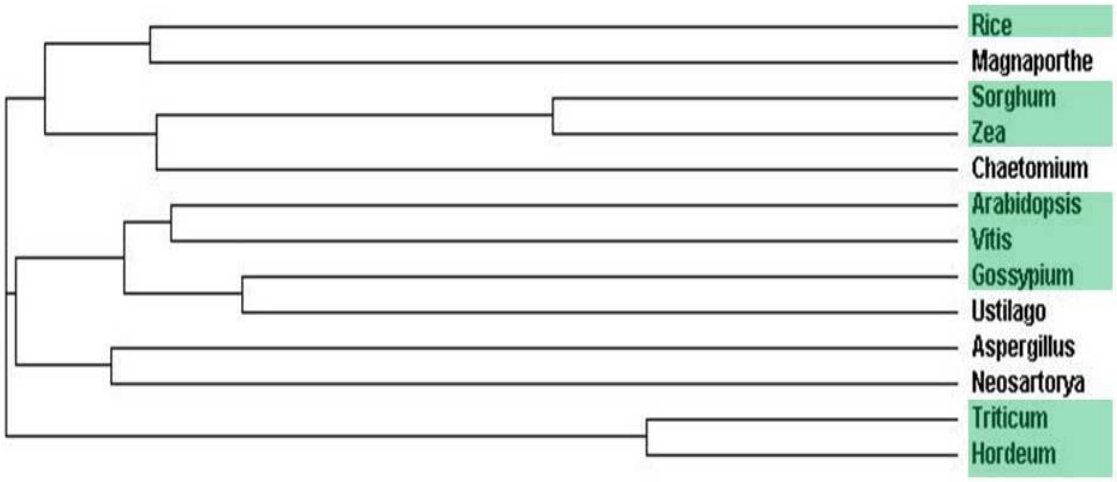

Figure 3.25: Sequence dendogram showing the relationship between DNA fragments from a multiple sequence alignment with the rice fragment 4: 23876766:23876875; *Sorghum bicolor* (gb|AY372819.1); *Zea mays* (gb|AY366085.1); *Arabidopsis thaliana* (ref|NM_116413.1); *Vitis vinifera* (emb|AM472754.2); *Triticum monococcum* (gb|AF459639.1); *Gossypium hirsutum* (gb|AF216497.1); *Hordeum vulgare* (dbj|AB063580.1); *Magnaporthe grisea* (ref|XM_001413121.1); *Chaetomium globosum* (ref|XM_001224928.1); *Aspergillus terreus* (ref|XM_001216110.1) and *Ustilago maydis* (ref| XM_757063.1). Green blocks indicate the plant sequences.

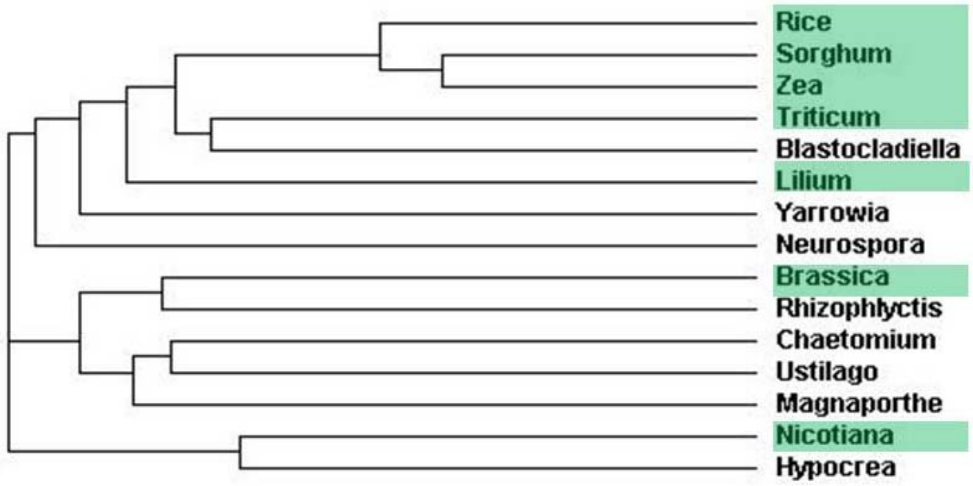

Figure 3.26: Sequence dendogram showing the relationship between DNA fragments from a multiple sequence alignment with the rice fragment 1: 36368256: 36368446; *Sorghum bicolor* (gb|U41653.1|SBU41653); *Zea mays* (gb|BT016692.1); *Triticum aestivum* (gb|AF005993.1); *Lilium longiflorum* (dbj|D21824.1|LILLIM18); *Brassica rapa* subsp. *pekinensis* (gb|AC189429.1); *Nicotiana tabacum* (gb|AY372070.1); *Blastocladiella emersonii* (gb|L22497.1); *Neurospora crassa* (emb|BX284745.1); *Hypocrea jecorina* (gb|AY281746.1); *Yarrowia lipolytica* (ref|XM_503148.1); *Rhizophlyctis rosea* (gb|AY582832.1); *Chaetomium globosum* (ref|XM_001225917.1); *Ustilago maydis* (ref|XM_754845.1) and *Magnaporthe grisea* (ref|XM_370461.2). Green blocks indicate the plant sequences.