

Supplementary File: Expression of *MAT1* locus elements in *Teratosphaeria destructans*

RNA-Seq reads derived from both the *MAT1-1* (CMW45661) and *MAT1-2* (CMW44962) isolates of *Teratosphaeria destructans* were used to determine expression of the genes at the *MAT1* locus. The reads of both isolates were mapped to the *MAT1-1* genome of CMW45661 as well as to its *MAT1-1* idiomorph (GenBank accession: MN531144) and the *MAT1-2* genome of CMW44962 and its *MAT1-2* idiomorph (MN531146). Table 6-1 shows the count data for these mappings.

Table 7-1 The number^a of RNA-Seq reads of the *MAT1-1* and *MAT1-2* *Teratosphaeria destructans* isolates that mapped^b to the *MAT1* idiomorphs and flanking genes

| | <u>MAT1-1 reads</u> | <u>MAT1-2 reads</u> | <u>MAT1-1 reads</u> | <u>MAT1-2 reads</u> |
|--|---------------------------|---------------------|-------------------------------|---------------------|
| | MAT1-1 Genome (i) | | MAT1-1 idiomorph (iii) | |
| ^cGenes in and around the <i>MAT1-1</i> idiomorph | | | | |
| <i>APC5</i> | 2770 | 1248 | 5033 | 2197 |
| <i>APN2</i> | 13 | 8 | 1123 | 808 |
| <i>MAT1-1-10</i> | 668 | 1 | 545 | 13 |
| <i>MAT1-1-1</i> | 302 | 0 | 1121 | 731 |
| hypothetical protein | 6860 | 6571 | 7019 | 6837 |
| F-box domain-containing protein | 1748 | 1553 | 2557 | 3001 |
| <i>COX6a</i> | 4604 | 4912 | 4656 | 4991 |
| | MAT1-2 Genome (ii) | | MAT1-2 idiomorph (iv) | |
| ^cGenes in and around the <i>MAT1-2</i> idiomorph | | | | |
| <i>APC5</i> | 3512 | 1103 | 6300 | 4131 |
| <i>APN2</i> | 831 | 708 | 1065 | 802 |
| <i>MAT1-2-1</i> | 264 ^d | 2190 | 1239 | 3039 |
| <i>MAT1-2-12</i> | 0 | 280 | 20 | 326 |
| hypothetical protein | 6898 | 6650 | 6956 | 6790 |
| F-box domain-containing protein | 1718 | 1568 | 2138 | 2195 |
| <i>COX6a</i> | 5261 | 5689 | 4669 | 4983 |

^a Numbers represent the average RNA reads counted in the three replicates of each isolate.

^b Reads were mapped against four templates, described in the main text. These were the two *T. destructans* genomes (i and ii) and the *MAT1-1* (iii) and *MAT1-2* (iv) idiomorphs.

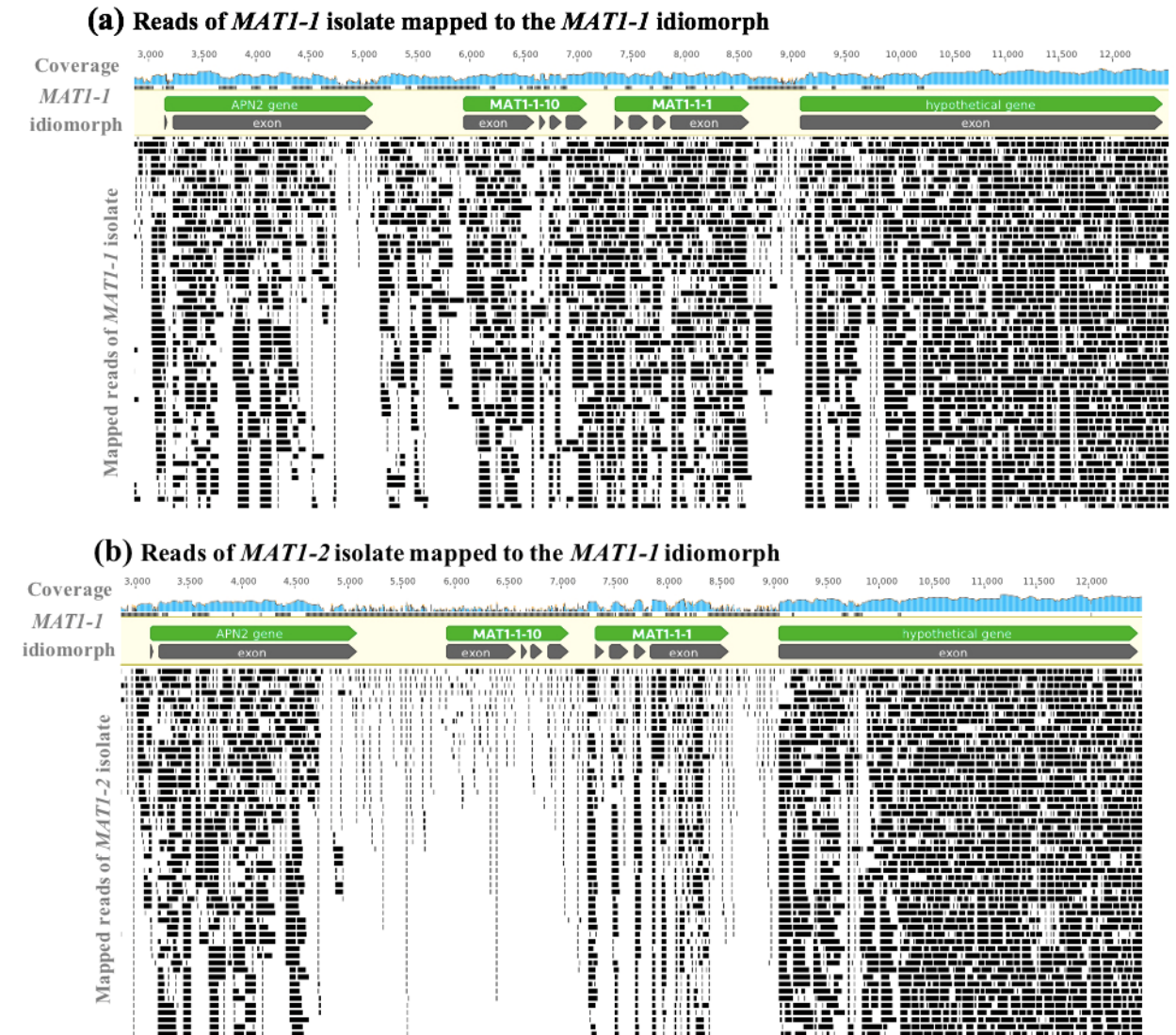
^c Genes are listed in the order they occur in the genome. The *MAT1* genes are in bold.

^d All these *MAT1-1* reads mapped to a portion of the first exon (194 bp) of *MAT1-2-1*, but their sequences correspond to one of the identified *MAT* fragment sequences and, therefore, do not suggest expression of the *MAT1-2-1* gene in a *MAT1-1* isolate.

The counts of the RNA-Seq reads that mapped to the two *T. destructans* genomes (Table 7-1), indicated expression of all flanking genes in the *MAT1-2* isolate and all flanking genes, except *APN2*, in the *MAT1-1* isolate. The *MAT1-1-1* and *MAT1-1-10* genes were expressed in the *MAT1-1* isolate and the *MAT1-2-1* and *MAT1-2-12* genes in the *MAT1-2* isolate. However, when only considering the *MAT1* idiomorphs as templates, a considerable number of RNA-

Aylward *et al.* Novel mating type-associated genes and gene fragments in the genomes of Mycosphaerellaceae and Teratosphaeriaceae fungi

Seq reads of the *MATI-2* isolate mapped to the *MATI-1-1* gene and, similarly, many reads of the *MATI-1* isolate mapped to the *MATI-2-1* gene. These mappings were less than the number of reads identified in the isolate containing the specific mating type gene, but nonetheless suggested expression of these genes in the “incorrect” isolate. A graphical representation of these mappings is shown in Fig. 7-1.



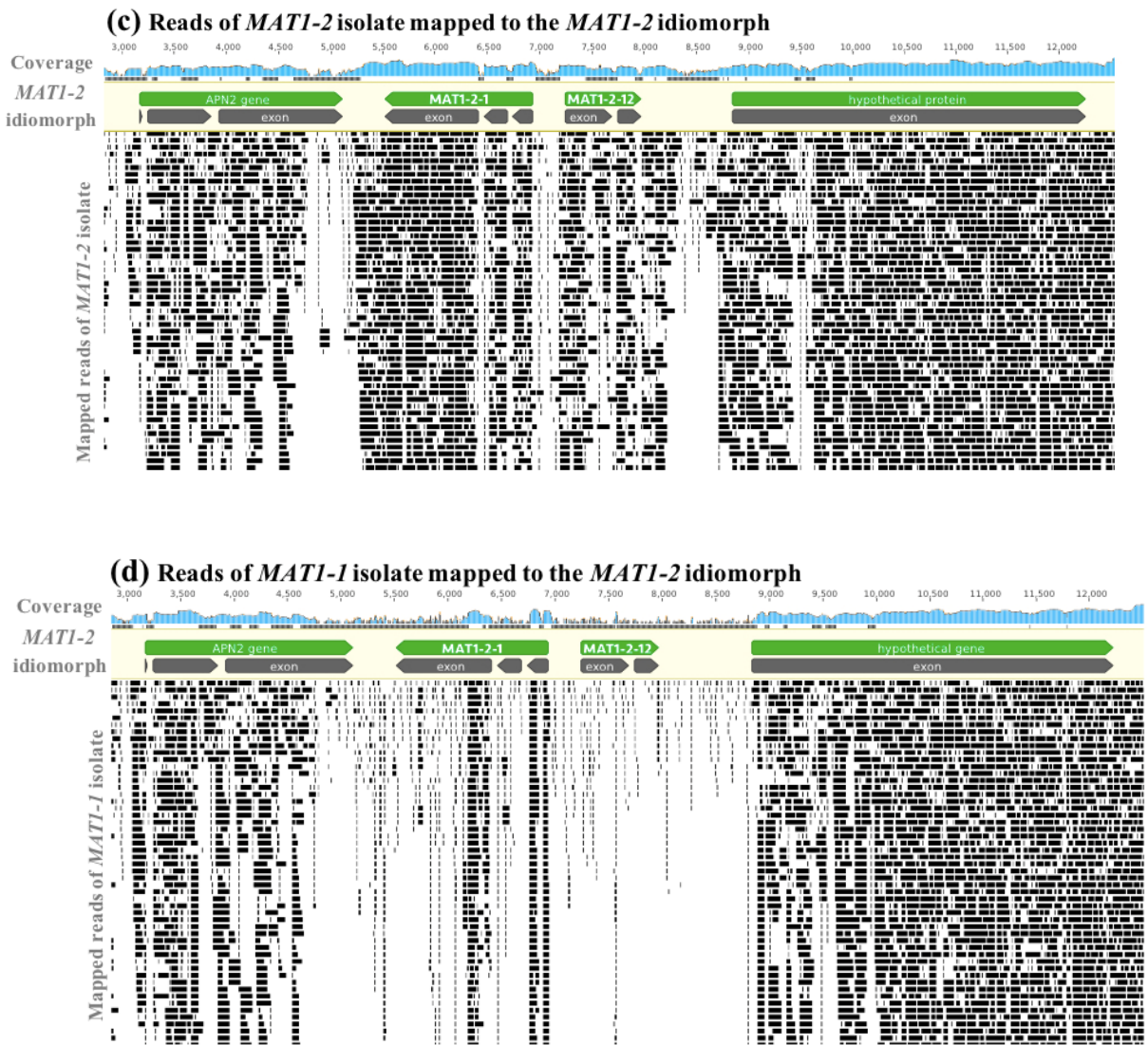


Figure 7-1 Mapping of RNA-Seq reads to the *Teratosphaeria destructans* (a-b) *MATI-1* and (c-d) *MATI-2* idiormorphs. Reads of each isolate were mapped to their own *MATI* idiormorph (a and c) as well as to the opposite idiormorph (b and d). Green and grey arrows represent genes and exons, respectively. Black bars are mapped reads and the relative number of mapped reads is indicated by the coverage plot at the top of the figure. Grey bars below the coverage plot indicate areas with <30 mapped reads.

Aylward *et al.* Novel mating type-associated genes and gene fragments in the genomes of Mycosphaerellaceae and Teratosphaeriaceae fungi

When comparing the mapped reads in Fig. 7-1, it is clear that mapping the reads of an isolate of opposite mating type to a *MAT* idiomorph indicates some level of expression in certain parts of the *MATI-1-1* and *MATI-2-1* genes. However, the entire gene is not expressed in isolates with the opposite mating type. Upon closer inspection, reads that mapped in these scenarios (Fig. 7-1b and 7-1d) aligned primarily in areas homologous to the *MAT* fragment sequences (Fig. 7-2).

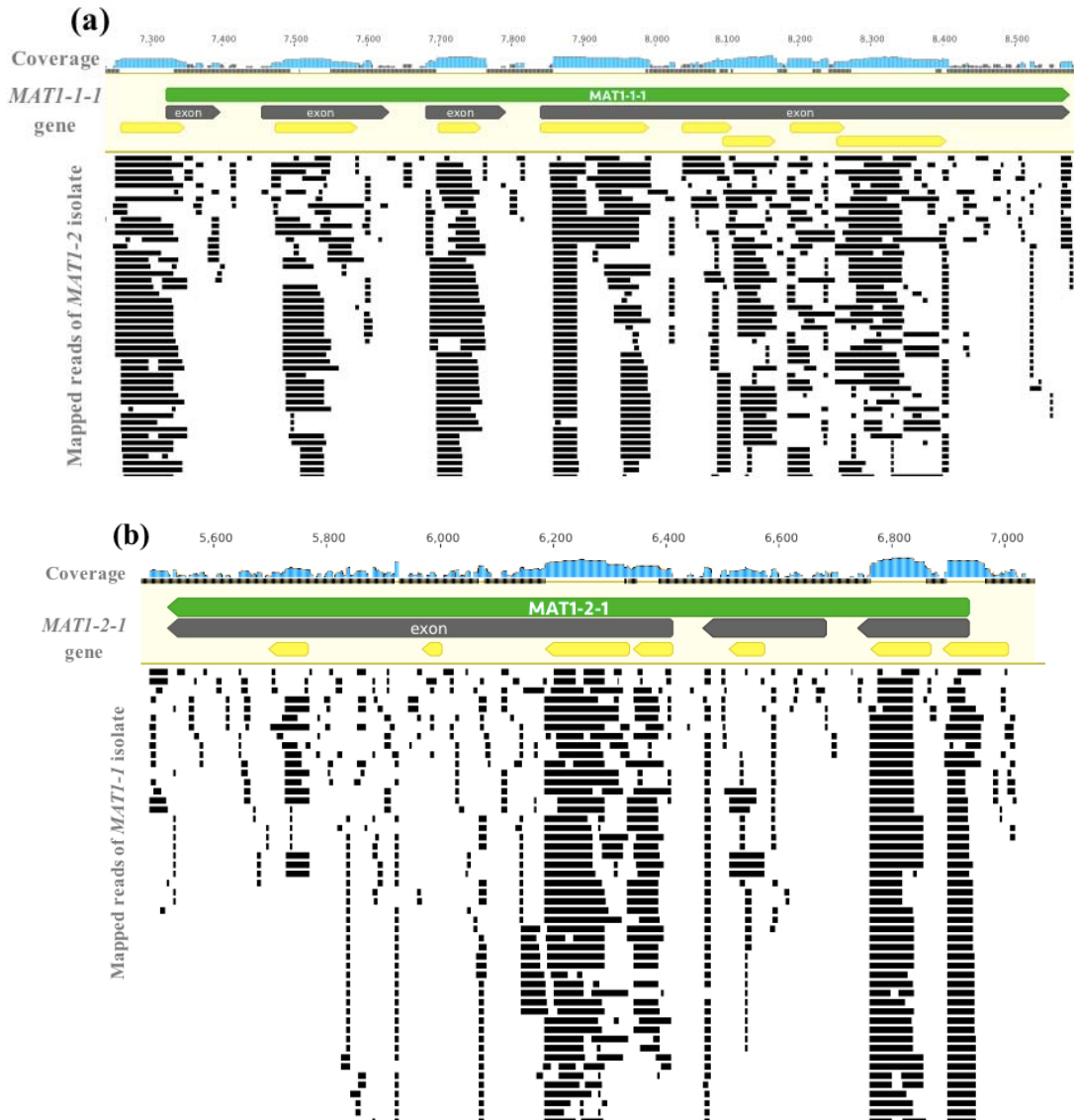


Figure 7-2 RNA-Seq reads of the *Teratosphaeria destructans* (a) *MATI-2* isolate mapped to the *MATI-1-1* gene and (b) *MATI-1* isolate mapped to the *MATI-2-1* gene. Yellow arrows represent regions within the genes (green arrows) that are homologous to the *MAT* fragment sequences identified in the genome.

Two lines of evidence indicated that the *MAT1-1* RNA reads that mapped to the *MAT1-2-1* gene and the *MAT1-2* reads that mapped to the *MAT1-1-1* gene were derived from *MAT* fragment sequences. Firstly, these reads only mapped when the *MAT1* idiomorph, and not the entire genome, was used as a template (Table 7-1). Secondly, the *MAT* fragment sequences and *MAT1* genes had a significant number of single nucleotide polymorphisms (SNPs) and indels (Table 7-2). The RNA reads that mapped to “incorrect” *MAT1* genes had the same SNPs present in the fragment sequences (Fig. 7-3). These mappings, therefore, represent expression of the *MAT* fragment sequences.

Table 7-2 Polymorphisms identified between the *MAT1* genes and their homologous *MAT* fragment sequences in the CMW44962 genome.

| | Length | SNPs | Indels | % Polymorphism |
|----------------------------------|--------|------|--------|----------------|
| <i>MAT1-1-1</i> fragments | | | | |
| A1 | 89 | 7 | 0 | 7.9% |
| B1 | 112 | 13 | 1 | 12.5% |
| C1 | 59 | 5 | 0 | 8.5% |
| D1 | 148 | 11 | 1 | 8.1% |
| E1 | 68 | 4 | 2 | 8.8% |
| F1 | 73 | 7 | 0 | 9.6% |
| G1 | 74 | 11 | 0 | 14.9% |
| H1 | 149 | 17 | 2 | 12.8% |
| <i>MAT1-2-1</i> fragments | | | | |
| A2 | 76 | 10 | 1 | 14.5% |
| B2 | 105 | 10 | 1 | 10.5% |
| C2 | 62 | 4 | 0 | 6.5% |
| D2 | 69 | 6 | 0 | 8.7% |
| E2 | 149 | 11 | 0 | 7.4% |
| F2 | 36 | 1 | 1 | 5.6% |
| G2 | 69 | 6 | 1 | 10.1% |

Aylward *et al.* Novel mating type-associated genes and gene fragments in the genomes of Mycosphaerellaceae and Teratosphaeriaceae fungi

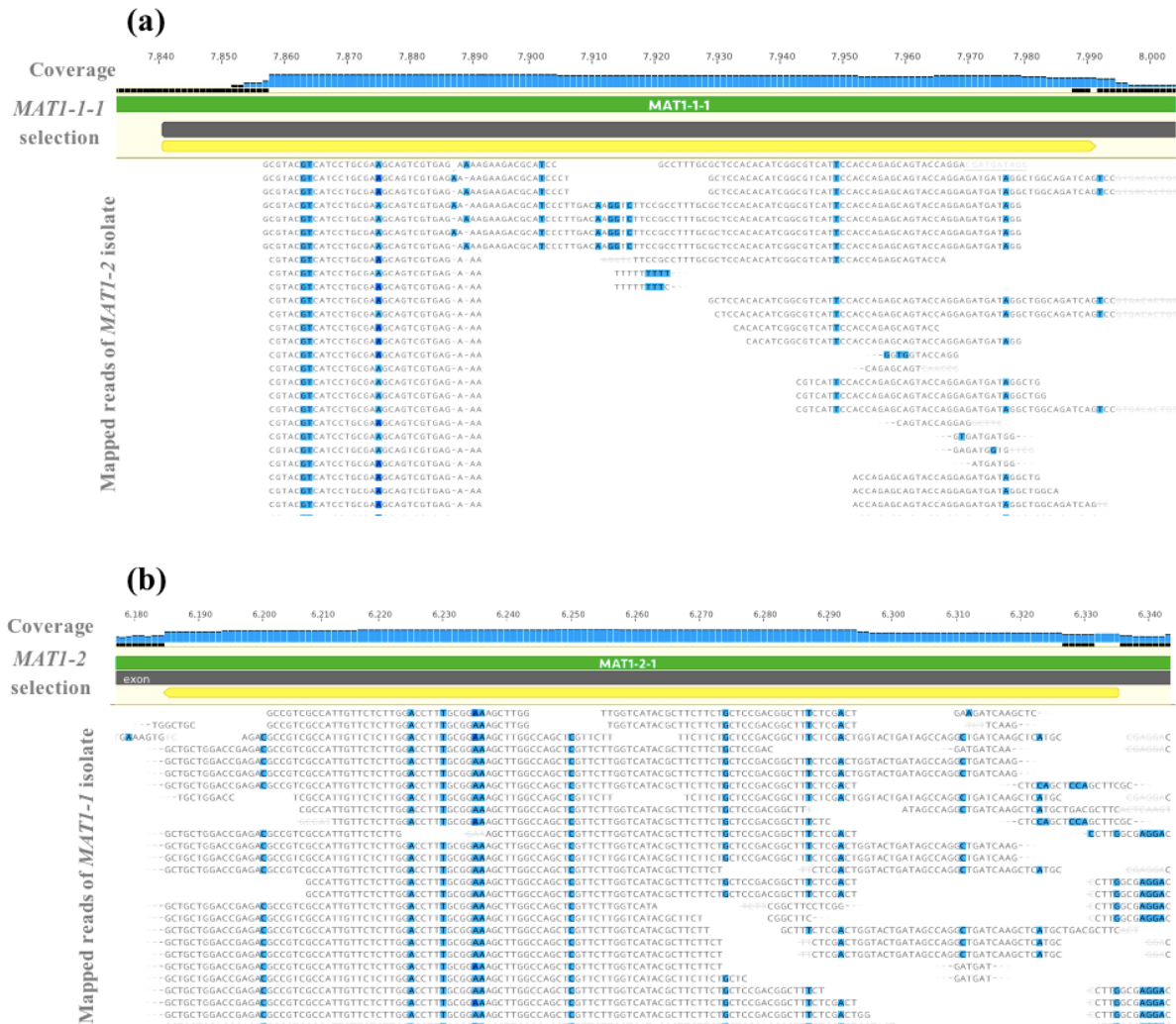


Figure 7-3 An example of a region of *MAT* fragments that mapped to the (a) *MATI-1-1* and (b) *MATI-2-1* genes. The yellow arrow represents the largest region on the *MATI-1-1* or *MATI-2-1* gene (green arrow) homologous to a *MAT* fragment. Bases highlighted in blue are SNPs that correspond to the sequence of the *MAT* fragments.

Aylward *et al.* Novel mating type-associated genes and gene fragments in the genomes of Mycosphaerellaceae and Teratosphaeriaceae fungi

To further confirm expression of the *MAT* fragment sequences, the RNA-Seq reads were also mapped to the genomic locations where these fragments occur (template v in the main text). All of the identified *MAT* fragment sequences, except perhaps fragment F2 on contig 147, had a sufficient number of mapped reads to conclude that these sequences are being expressed in both the *MAT1-1* and *MAT1-2* isolates (Table 7-2). The visual mappings of the reads to the genome, however, showed that expression is not limited to the *MAT* fragment sequences, but that there are also regions of expression around these sequences. The two contigs on which the most *MAT* fragments were identified is shown as an example of this (Fig. 7-4).

Table 7-3 The number^a of RNA-Seq reads of the *MAT1-1* and *MAT1-2* *Teratosphaeria destructans* isolates that mapped to the regions in the CMW44962 genome in which *MAT* fragment sequences were identified

| Gene / <i>MAT</i> fragment ^b | <i>MAT1-1</i> reads | <i>MAT1-2</i> reads |
|---|---------------------|---------------------|
| Contig 28 | | |
| Tdes44962_MAKER_02653 | 403 | 1251 |
| C1_Contig28.1(1) | 694 | 604 |
| B1_Contig28.2(1) | 532 | 617 |
| Tdes44962_MAKER_02654 | 3486 | 3268 |
| Tdes44962_MAKER_02655 | 1126 | 1910 |
| Contig 82 | | |
| Tdes44962_MAKER_05754 | 3108 | 2607 |
| Tdes44962_MAKER_05755 | 2660 | 1687 |
| B2_Contig82.1(2) | 1156 | 463 |
| A2_Contig82.2(2) | 334 | 454 |
| H1_Contig82.3(1) | 735 | 862 |
| A1_Contig82.4(1) | 760 | 558 |
| Tdes44962_MAKER_05756 | 458 | 572 |
| Contig 147 | | |
| Tdes44962_MAKER_07832 | 4992 | 4781 |
| Tdes44962_MAKER_07833 | 1445 | 1228 |
| Tdes44962_MAKER_07834* | 64 | 47 |
| F2_Contig147.1(2) | 21 | 20 |
| Tdes44962_MAKER_07835 | 2170 | 1289 |
| Tdes44962_MAKER_07836 | 1616 | 1587 |
| Tdes44962_MAKER_07837 | 660 | 1089 |
| Contig 300 | | |
| Tdes44962_MAKER_09673 | 9883 | 3587 |
| Tdes44962_MAKER_09674 | 162 | 191 |
| Tdes44962_MAKER_09675 | 228 | 345 |
| Tdes44962_MAKER_09676* | 393 | 456 |
| E2_Contig300.1(2) | 947 | 1312 |
| D1_Contig300.2(1) | 655 | 750 |
| D2_Contig300.3(2) | 195 | 213 |
| G1_Contig300.4(1) | 175 | 206 |

| | | |
|-----------------------|------|------|
| C2_Contig300.5(2) | 54 | 91 |
| G2_Contig300.6(2) | 130 | 198 |
| F1_Contig300.7(1) | 393 | 580 |
| E1_Contig300.8(1) | 47 | 66 |
| Tdes44962_MAKER_09677 | 896 | 1309 |
| Tdes44962_MAKER_09678 | 7266 | 9034 |

^a Counts represent the average of three replicates

^b Genes and *MAT* fragments are listed in the order they occur in the genome. *MAT* fragments are highlighted in yellow and asterisks indicate gene predictions with a *MAT* fragment in the coding sequence.

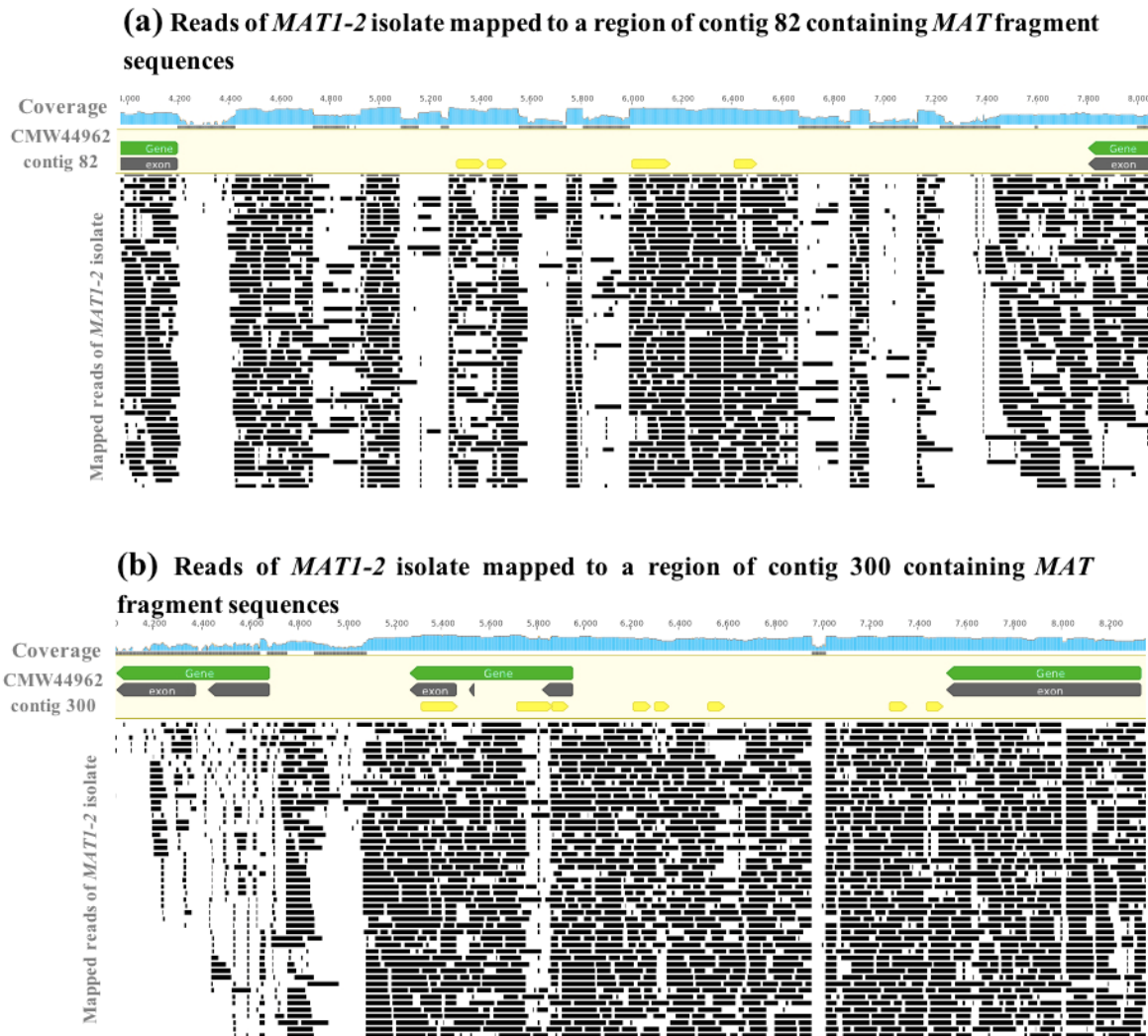


Figure 7-4 RNA-Seq reads of the *MATI-2* isolate mapped to the two regions of the CMW44962 genome in which the most *MAT* fragment sequences (yellow arrows) were identified.