VIRULENCE OF *MAYETIOLA DESTRUCTOR* (SAY) FIELD POPULATIONS IN THE GREAT PLAINS AND LEVANASE/INULASE-LIKE GENES IN THE HESSIAN FLY GENOME

by

SANDRA GARCÉS CARRERA

B. S., Central University of Ecuador, Ecuador, 1988
M. S., The Ohio State University, 2004

AN ABSTRACT OF A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Entomology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Abstract

The Hessian fly, *Mayetiola destructor* (Say), is a major pest of wheat, and is controlled mainly through deploying fly-resistant wheat cultivars. This study investigated five *M. destructor* populations collected from Texas, Louisiana, and Oklahoma, where infestation by Hessian fly has been high in recent years. Eight resistance genes including *H12, H13, H17, H18, H22, H25, H26*, and *Hdic*, were found to be highly effective against all tested *M. destructor* populations in this region, conferring resistance to 80% or more of plants containing one of these resistant genes. The frequency of biotypes virulent to resistant genes ranged from 0 to 45%. A logistic regression model was established to predict biotype frequencies based on the correlation between the percentages of susceptible plants obtained in a virulence test. In addition to the virulence test, the log-odds of virulent biotype frequencies were determined by a traditional approach to predict the logistic regression model.

Characterization of a bacterial artificial chromosome (BAC) clone identified a gene encoding a protein with sequence similarity to bacterial levanases. Blast searching with the levanase-like protein identified 14 levanase/inulase-like genes or gene fragments. In this study, we determined the expression levels of these genes in different developmental stages and different tissues of 3-d old larvae of *M. destructor*. Sequence analysis revealed that six genes encode full length proteins, three were truncated at the 5' end, and five truncated at the 3' end. Sequences of putative proteins showed approximately 42% similarities to bacterial levanases or inulases, and 36% similarity to fungal levanases or inulases. No sequence similarities were found with any known animal or plant proteins. Comparative analysis of sequences among 14 levanase/inulase-like genes revealed that positions for intron/exon boundaries are conserved

among different genes even though the length of each intron and exon varied among different genes. The expression patterns of the levanase/inulase-like genes were different among developmental stages and larval tissues of *M. destructor*. Interestingly, three genes presented alternative splicing bands in different developmental stages, and two genes exhibited splicing bands in different tissues of 3 d old *M. destructor*. This study would be useful for future studies of the characterization and function of levanase/inulase-like genes of these enzymes in plant-insect interactions.

VIRULENCE OF *MAYETIOLA DESTRUCTOR* (SAY) FIELD POPULATIONS IN THE GREAT PLAINS AND LEVANASE/INULASE-LIKE GENES IN THE HESSIAN FLY GENOME

by

SANDRA GARCÉS CARRERA

B. S., Central University of Ecuador, Ecuador, 1988
M. S., The Ohio State University, 2004

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Entomology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Approved by:

Co-Major Professor
Ming-Shun Chen

Approved by:

Co-Major Professor
C. Michael Smith

Copyright

Abstract

The Hessian fly, *Mayetiola destructor* (Say), is a major pest of wheat, and is controlled mainly through deploying fly-resistant wheat cultivars. This study investigated five *M. destructor* populations collected from Texas, Louisiana, and Oklahoma, where infestation by Hessian fly has been high in recent years. Eight resistance genes including *H12, H13, H17, H18, H22, H25, H26*, and *Hdic*, were found to be highly effective against all tested *M. destructor* populations in this region, conferring resistance to 80% or more of plants containing one of these resistant genes. The frequency of biotypes virulent to resistant genes ranged from 0 to 45%. A logistic regression model was established to predict biotype frequencies based on the correlation between the percentages of susceptible plants obtained in a virulence test. In addition to the virulence test, the log-odds of virulent biotype frequencies were determined by a traditional approach to predict the logistic regression model.

Characterization of a bacterial artificial chromosome (BAC) clone identified a gene encoding a protein with sequence similarity to bacterial levanases. Blast searching with the levanase-like protein identified 14 levanase/inulase-like genes or gene fragments. In this study, we determined the expression levels of these genes in different developmental stages and different tissues of 3-d old larvae of *M. destructor*. Sequence analysis revealed that six genes encode full length proteins, three were truncated at the 5' end, and five truncated at the 3' end. Sequences of putative proteins showed approximately 42% similarities to bacterial levanases or inulases, and 36% similarity to fungal levanases or inulases. No sequence similarities were found with any known animal or plant proteins. Comparative analysis of sequences among 14 levanase/inulase-like genes revealed that positions for intron/exon boundaries are conserved

among different genes even though the length of each intron and exon varied among different genes. The expression patterns of the levanase/inulase-like genes were different among developmental stages and larval tissues of *M. destructor*. Interestingly, three genes presented alternative splicing bands in different developmental stages, and two genes exhibited splicing bands in different tissues of 3 d old *M. destructor*. This study would be useful for future studies of the characterization and function of levanase/inulase-like genes of these enzymes in plant-insect interactions.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my deepest gratitude to my co-major advisors, Drs. Ming-Shun Chen and Charles Michael Smith for their valuable and constructive suggestions during the planning and development of this research work. Their willingness to give their time so generously has been appreciated.

I would like to offer my special thanks to Dr. Yoonseong Park for his continuous support and guidance throughout my research work. His suggestions and moral support helped me to intensify and improve my research efforts.

I am thankful to Dr. Bob Bowden, who provided assistance for my research work during these three years. I am grateful to Dr. John Fellers, Dr. Myron Bruce and Kerri Neugebauer, of the Department of Plant Pathology, for their advice, suggestions and assistance for Real-Time PCR analysis.

I would also like to extend my thanks to Shauna Dendy for her assistance during my research work in the greenhouse. I could not have completed this work without the support from my colleagues in Dr. Chen's, Dr. Park's, and Dr. Zhu's laboratories.

My special thanks to Drs. Jeff Whitworth and James Shroyer for their support as Committee Members and to Dr. Michael Kanost for serving as the chairperson for my examination committee.

I would also like to extend thanks to all faculty, staff members and graduate students in the Department of Entomology, Kansas State University for their support during my Ph. D. studies.

I am immensely grateful to the Government of Ecuador for provide me a scholarship for my Ph. D. studies at the Department of Entomology. I am grateful to The National Research Agriculture Institution (INIAP) for supporting my Ph. D. program.

Special thanks should be given to all my friends: Phillip Defoe, Andrés Salcedo, Chammi Attanayake, Miguel Arango, Priscilla Mfombep, Moustapha Soumaila, Krissana-Ruang-Rit, Don Hung Kim, and Maricela Robles-Murguia for offering help and support in one way or another.

Finally, I wish to thanks my family for their support and encouragement throughout my studies.

Dedication

I dedicate my dissertation work to my family. I have a special feeling of gratitude to my parents, Olmedo and Rocío for their encouragement and support throughout my entire life; to my son, Juan Esteban, for his unwavering support throughout the doctorate program; to my sisters, Janeth and Guisela who have never left my side and are very special to me; to my brother in law, Tarquino, my nephew Santiago Andrés, and my niece Ivonne Irina for supporting me all the time.

Chapter 1 - Literature Review

M. destructor biology and feeding mechanism.

The Hessian fly, *Mayetiola destrutor* (Say), is a member of the gall midge family (Diptera: Cecidomiiydae) (Mamaev 1975; Gagne 1994 cited by Stuart et al. 2008) which is suitable for genetic analysis due to its short life cycle (28 days) and a small genome (male 120.0 MB and female 157.9 Mb) (Johnston et al. 2004). The adult *M. destructor* has a life span of 2 to 3 days; long enough for mating and laying eggs (Stuart et al. 2008). Adult females fly from plant to plant, examine leaf surfaces, and generally oviposit on the adaxial leaf surface of the youngest leaves in a short time (approximately 3 h) (Kanno and Harris 2002). However, adult females avoid wheat plants that carry conspecific larvae, but not plants with conspecific eggs. This oviposition preference could be related with larvae feeding causes. Wheat seedlings attacked by *M. destructor* larvae stop growing. The growth of the third leaf is stunted, the leaf blades occupied by *M. destructor* larvae are blue-green instead of bright green, and no other leaves are generated (Byers et al. 1971; Anderson et al. 2006).

Although adult females have only a few living days and lay between 50-400 eggs (Harris et al. 2003), the ovipositing females must optimize the oviposition-site to allow neonate larvae to manipulate the plant to become susceptible by reprogramming the epidermal cells or reactive cells to become nutritive cells. However, if an adult female made a bad oviposition choice, neonate larvae can make choices by leaving that location and move to a better location within the plant to find reactive cells. For instance, offspring of females that placed their eggs on the abaxial surface of the third leaf shift to the adjacent abaxial surface of the fourth leaf, and also, when a *M. destructor* female laid eggs on the oldest leaf, neonate larvae can move away from that location and find a younger leaf (Ganehiarachchi et al. 2013). It is possible that *M.*

destructor larvae can determine where the epidermal cells can be reprogrammed to become nutritive cells. Based on environmental conditions, the eggs hatch after a period of 3-4 days at 20 ˚C (Stuart et al. 2008). *M. destructor* larva has three instars, but only the first and second larval instars feed and cause damage to wheat plants (until 10-14 days). The first instar larvae, after hatching from the egg, migrate down along the leaf blade, by moving between the leaf sheaths to the base of the plant, where a permanent feeding site is established (Harris et al. 2003). The third instar larva is a non-feeding stage, and forms a puparium after 6-7 days at 20˚C. The non-feeding larvae can be maintained in diapause at 4˚C for more than one year (Stuart et al. 2008). The pupae stage also occurs within the dark brown puparium. The length of the pupal stage ranges from 7-10 days at 20˚C before adult flies emerge to mate and lay eggs.

The feeding mechanism of *M. destructor* was first studied in the 1930's. It was believed that larval secretions softened cell walls of wheat seedlings, allowing larvae to obtain nutrition from plant cells and causing a reduction of plant growth (Refai 1956). Subsequent studies have shown that first instar larvae inject salivary fluids into the plants via highly specialized mandibles. Hatchett et al (1990) reported that these salivary fluids modify plant cell components and provide an opportunity for larvae to ingest nutrients. During this process, first instar larvae inject effector proteins through its salivary glands into the epidermal cells, and these effectors bind with their respective resistant cognate proteins in the plant so there is an interspecific molecular recognition (Stuart et al. 2012). Inside the epidermal cells, *M. destructor* larval feeding causes nuclear breakdown, cytoplasmic organelle degradation, and an increase in the number and size of vacuoles (Harris et al. 2006a). After two to three days of the initial larval attack, epidermal and mesophyll cells become nutritive or nurse cells because of an increase in cytoplasmic staining, number of cellular organelles, and numerous small vacuoles (Harris et al.

2006a). The nutritive cells in wheat seedlings probably act as sink tissues to support the growth of the larvae by importing photoassimilates (Harris et al. 2006a). As a result of these changes in the plant, the first virulent instar larvae have elevated consumption of carbohydrates, and there is an elevated synthesis of amino acids which are required nutrients for the larvae (Zhu et al. 2008; Saltzmann et al. 2010). Additionally, the carbon-nitrogen ratio is reduced at the feeding site. These nutrients allow the first virulent instar larvae to develop into the second instar larvae (Stuart et al. 2012). The physiological response of susceptible plants to *M. destructor* feeding is visible through permanent plant stunting and a characteristically dark green leaf color that results from an increase in the concentration of leaf chloroplasts (Robinson et al. 1960). The effectors induce susceptibility suppressing plant basal defense.

M. destructor control measures.

In the United States' Great Plains and upper Midwest areas, *M. destructor* control is based on prevention rather than remedial control because once an infestation occurs; there is no strategy to control larval feeding damage. During the wheat seedling stage, larval feeding causes irreversible stunting and will eventually kill the plants. During the wheat adult stage, larval feeding can prevent spike development; reduce grain fill; weaken stems, resulting in lodging; and decrease yield (Castle del Conte et al. 2005). As the research on the biology of *M. destructor* has advanced, strategies and tactics of dealing with *M. destructor* have changed due to their ineffectiveness. Currently, four major tactics are recommended for wheat growers to control *M. destructor*: (1) Planting resistant cultivars, (2) applying insecticide seed treatment, (3) late planting (fly-free date) to avoid the *M. destructor* fall generation damage, and (4) cultural practices such as crop rotation, destruction of volunteer wheat, and destruction of infested stubble (Buntin et al. 1992). These tactics can alleviate *M. destructor* damage. However, their

effectiveness depends on the specific areas of production, wheat cultivar, weather, wheat growers, and *M. destructor* infestation levels. Late planting for winter wheat is applied to avoid oviposition by the last autumn generation of *M. destructor*. However, this tactic is not effective in all regions of *M. destructor* distribution. For instance, in the southeastern region of the United States, the emergence and oviposition of *M. destructor* population occurs throughout the winter season (Buntin et al. 1990). *M. destructor* problems can be prevented by applying crop rotation to reduce the presence of wheat stubble and volunteer wheat which facilitate new *M. destructor* infestations. The lack of co-operation and adequate co-ordination among wheat growers has been a limiting factor. Some of the growers prefer to maintain the wheat stubble to reduce the loss of soil moisture (Weisz 2012). Seed treatment applications can help to manage *M. destructor* infestations in specific situations, such as when wheat growers plant susceptible cultivars in wheat growing regions that historically have *M. destructor* problems, and early planting. The use of systemic insecticide seeds treatment should be applied at the correct rate and on susceptible cultivars. The decision to use insecticide seed treatments should be taken after careful considerations due to expensive treatment and it only provides control of *M. destructor* larvae for a short period of time (up to 30 days) (Whitworth 2005).

The greatest success in controlling *M. destructor* has come from the development of *M. destructor*-resistant wheat cultivars. The early benefits of *M. destructor*-resistant cultivars are well documented, and include the value of "Pawnee" wheat in Nebraska and Kansas during 1942 and 1943 (Painter 1968; Webster and Kenkel 1999). In Georgia, wheat growers reported that *M. destructor*-resistant cultivars were unaffected by significant *M. destructor* populations, and growers obtained large economic benefits when compared with susceptible cultivars. Consequently, growers discontinued application of systemic insecticides which reduced

environmental contamination (Buntin et al. 1992). Due to these successes, wheat breeders have continued deploying *M. destructor*-resistant wheat cultivars and have been adapted to most U.S. wheat production areas.

Wheat resistance to *M. destructor* larvae

Numerous sources of resistance have been identified and are being used for developing *M. destructor*-resistant wheat cultivars, especially when old wheat cultivars lose effectiveness (Liu et al. 2005). Currently, 34 resistant genes from common wheat, durum wheat, rye, and goatgrass have been identified and incorporated into wheat varieties (Li et al., 2013). Most of these are single genes inherited as dominant traits, except *h4* which is recessive and *H7/H8* that are effective only when they are together (Stuart et al. 2012). The majority of the *M. destructor*-resistant genes were derived from *Triticum turgidum* ssp. *durum*. Three genes, H3, H5 and H12 were derived from wheat (*Triticum aestivum* L.). *Hdic* gene was identified from *Triticum turgidum* ssp. *dicoccum* (Liu et al. 2005). Finally, the newly identified gene *H34* was from wheat cultivar "Clark" (*Triticum aestivum* L.) (Li et al. 2013).

The category of resistance related to these genes is antibiosis, which is manifested as the death of first instar larvae within 5 to 6 days after establishment (Stuart et al. 2012). Even though all resistant genes have the same mechanism of resistance, their individual resistant genes are different. Currently, there is a hypothesis that gall midges use an effector-based strategy which is similar to that used by plant-pathogenic organisms (Stuart et al. 2012). The wheat-*M. destructor* interaction is characterized by a gene-for-gene recognition where resistant genes provide *M. destructor* resistance. Furthermore, the presence of some *M. destructor* genes encoding putative effector proteins suggest that wheat-*M. destructor* interactions use the same strategy as plants with their parasitic fungi, oomycetes and nematodes. Most resistant genes cloned to date are

encoded with the nucleotide binding-leucine rich repeat (NB-LRR) proteins, which mediate recognition of diverse effectors from all different classes of plant pathogens. Among 34 *M. destructor* resistant genes in wheat, only the gene *Hdic* was cloned, and it contains the NB-LRR proteins that recognize the effector proteins of plant pathogens (Stuart et al. 2012). The *Hdic* confers resistance to GP and L biotypes, and the strains vH9 and vH13 of the *M. destructor*. The reaction of the gene *Hdic* is different to the pattern reactions obtained with the other *M. destructor* genes that are in the chromosomes 1A (Liu et al. 2005). In addition, results obtained in virulence tests reported that *Hdic* confers resistance to 80% or more to *M. destructor* populations collected from Grayson and Fannin Counties, (Texas); Kay County in Oklahoma, and Scott County in Kansas (Chen et al. 2009). These results make the gene *Hdic* very useful in wheat breeding programs in these states. In addition, only one insect resistant gene (*Mi-1.2*) has been cloned due to its resistance to root-knot nematodes (Milligan et al. 1998) and its ability to confer resistance to both potato aphids and white flies (Rossi et al. 1998; Goggin et al. 2001; Nombela et al. 2003).

Due to *M. destructor* population changes, resistant genes *H3, H5, H6* and *H7/H8* have lost effectiveness in most wheat growing areas (Ractliffe et al. 2000). Some other resistant genes, such as gene *H18,* are a very temperature sensitive gene that loses effectiveness above 20°C (Cambron et al. 2010). Although these genes confer important levels of resistance to *M. destructor* populations, resistant genes derived from goatgrass and ryes have problems in cultivar development programs when they are associated with alien linkage drag. In addition, many *M. destructor* genes have been defeated 5 to 8 years after being deployed; probably through the selection of virulent *M. destructor* genotypes in agriculture fields (Harris et al. 2010). Therefore, gene deployment strategies need to be applied for reducing biotype development, which will

allow us to use the existing resistant genes efficiently. Two strategies have been proposed to improve the durability of the resistance: sequential gene deployment (Cox and Hatchett 1986), and pyramiding gene (Gould 1986). Sequential releases of wheat cultivars containing the resistant genes may reduce *M. destructor* outbreaks. Combining several genes in one cultivar may increase the duration of the resistance. However, the effectiveness of these two models depends on the genotypic interaction between resistance genes and *M. destructor* virulence genes (Bouhssini et al. 2001). The selection of the most effective breeding method will depend on the results of further studies into the field movement and genetic changes that occurs in *M. destructor* field populations.

M. destructor biotypes.

The term "biotype" is primarily used by biologists to differentiate insect populations (as well as other organisms) due to a highly variable range of underlying causes (Diehl and Bush 1984). Some "insect biotypes" are referred to variants in color, shape, size, insecticide resistance, migration, dispersal tendencies, seasonal activities, pheromones differences, and diseases vector capacities (Russell 1978). Some of these biological differences have been used to describe populations as biotypes in the literature. In the case of *M. destructor*, biotype was defined according to the insect virulence to different resistant genes (Painter 1951).

The deployment of wheat cultivars with high levels of antibiosis has closely paralleled the development of *M. destructor* biotypes. High levels of antibiosis to first instar *M. destructor* larvae exert a strong selection pressure on *M. destructor* populations that favors biotypes to reproduce and survive on resistant cultivars (Ratcliffe et al. 1994). Virulence depends on the presence of the homozygous recessive condition in the *M. destructor* virulent gene at a locus corresponding to a specific dominant plant resistance gene (Clement and Quisenberry 1999).

Hence, *M. destructor* biotype composition in wheat growing areas changes with the exposure of wheat cultivars carrying a specific *M. destructor*-resistant gene(s). Analysis of allozyme variation among 18 *M. destructor* populations distributed thorough the United States suggests considerable variation among populations (Black et al. 1990). The explanation for this local variation among populations is considered to be caused by genetic drift within the natural populations (Johnson et al. 2004). This local differentiation in conjunction with mating behavior indicates how the combination of virulent genes becomes established in the field (Black et al. 1990). Factors affecting gene flow in local populations could be related to the short life span of adults *M. destructor* and therefore the limited dispersal of the *M. destructor* (Johnson et al. 2004). Norton and Schemerhorn (2013) suggested that the *M. destructor* population structure of North America could be closely associated with a local adaptation of a particular wheat class. The changes in wheat varieties planted over time and space, and the gene-for-gene interaction between *M. destructor* and wheat, indicate that balancing selection is common in *M. destructor* populations. This would be because of the survival of heterozygous as well as homozygous population in the same host plant (Gallun 1977; Baluch et al. 2012). However, further research is needed in order to explore whether the *M. destructor* population structure in North America is the result of the association between *M. destructor* populations and wheat class or if other factors related to the biology of the *M. destructor* may be responsible (Norton and Schemerhorn 2013).

M. destructor biotype composition was recognized for the first time in 1930 by Painter (Clement and Quisenberry 1999). Since then, 16 biotypes (identified GP and A to O) have been identified on the basis of their differential response to the resistant genes *H3, H5, H6*, and a gene combination *H7/H8* in wheat (Ratcliffe et al. 1994). Biotype Great Plains (GP) has been considered as predominant in *M. destructor* field populations, and the biotypes “A” through “O”

differ in the number of resistance genes to which these biotypes express virulence. The most virulent biotype is "L" which affects wheat varieties that contain resistant genes, *H3, H5, H6,* and a gene combination *H7/H8*. All biotypes have been identified in the field, at various frequencies in soft winter wheat areas of the eastern United States. Major shifts in biotype composition and virulence to resistant genes in wheat occurred throughout the Eastern United States from the mid-1980's to the late 1990's that reduced the effectiveness of all deployed genes (Ratcliffe et al. 1994). However, the effectiveness of the four resistant genes differs according to the wheat-growing region. Actually, this biotype system is useful only for the studies of biotype genetics and wheat-*M. destructor* interactions. It no longer provides meaningful information for effective protection against *M. destructor* populations. However, this biotype system provides only information on the virulence/avirulence of the *M. destructor* to three resistant genes and a gene combination (Chen et al. 2009). These three resistant genes and the one gene combination are no longer effective in most wheat growing regions. Additionally, the number of the biotypes defined with the system is 2^n (where n is the number of resistant genes used in the analysis). Following the same system, the number of the resistant genes identified to date, will be around 233 possible *M. destructor* biotypes (genotypes), if each biotype differs in at least one gene specially matching one of the host's 33 resistance genes (Cambron et al. 2010). As the number of resistant genes increase, the potential number of biotypes will be unmanageable, and the designation of biotype will be impossible and questionable (Diehl and Bush 1984). Another problem with this system is that it is very labor intensive and impractical for analyzing large number of *M. destructor* field samples. Moreover, there is insufficient information of all known resistant genes at the present moment. Instead, virulence tests in flats are being used to evaluate the percentage of virulence of *M. destructor* females which correlate with the percentage of the

susceptible plants obtained in the virulence test (Chen et al 2009; Cambron et al. 2010). Thus, as *M. destructor* biotypes continue to evolve as resistant wheat cultivars are deployed, knowledge of the frequency of *M. destructor* virulent gene(s) is important for predicting field durability of *M. destructor* resistance genes and the elaboration of new resistant gene deployment strategies.

M. destructor-associated bacteria

Insects are a group of multicellular organism that live together with many different microorganisms either inside or outside of their body in symbiotic relationship as mutually beneficial, neutral or as parasitism (Buchner 1965). The microorganisms inside the insect body include endoparasites, gut microbe, extracellular symbiosis and intracellular symbiosis or endosymbionts. Endosymbionts is the most important association between insect and microorganisms. This association is maintained through generations where insects and microorganism equally benefit from this association. There are several examples of the insect pests that live in association with microorganism. Symbiotic bacteria perform different functions in insects. For instance, the variable susceptibility of Pea aphids (*Acyrthosiphon pisum*) to *Aphidius ervi* predation is linked to the presence or absence of the aphid's endosymbionts (Oliver et al. 2003). Other studies revealed that phage-borne toxin genes and several toxins which provide defense to the aphid host depend on the genome of the pea aphid symbiont, *Hamiltonella defensa* (Moran et al. 2005). Several studies revealed that food plant use of herbivorous insects can be directly increased by facultative endosymbionts (Su et al. 2013). Another example of the influence of symbionts on insect-plant interactions is the case of stinkbug, *Megacopta punctatissima* performed well on crop legumes while the non-pest species, Megacopta cribari, reduced the rate of eggs hatching. However, it is not clear how their symbionts interact with insects (Hosokawa et al. 2007). Insect symbionts have also been reported to benefit their hosts

through physiological changes in insect host through a complex signaling transduction response to insect attacks. The best example is ambrosia beetles and their mutualistic symbiotic fungi of bark. Fungi make wood digestible for their host larvae and assist beetles in overcoming tree resistant mechanisms (Paine et al. 1997). In addition, symbionts induced alteration of insect host behavior where symbionts and hosts both benefit from the behavior.

In the case of *M. destructor*, different types of bacteria have been found in all developmental stages and they are maternally transmitted (Bansal et al. 2010). The most abundant bacteria found in *M. destructor* belong to genera *Enterobacter*, *Pantoea*, *Klebsiella*, *Pseudomonas*, *Sthenotropomonas*, *Staphylococcus* and *Achromobacter* (Bansal et al. 2010). The function of the bacteria in *M. destructor* is not clear yet; however, it is possible that gut bacteria in *M. destructor* synthesize nutrients for larvae or digest nutrients that are inaccessible for the insect. Since *M. destructor* larvae have the ability to manipulate host plants through secretions from salivary glands, it is possible that bacteria could be released through salivary glands and thus participates in specific metabolic pathways for the interaction with the host plant.

References

Baluch, S., Ohm, H., Shukle, J and C. Williams. 2012. Obviation of Wheat Resistance to the Hessian Fly Through Systemic Induced Susceptibility. Source: Journal of Economic Entomology. 105:642-650.

Black, C., Hatchett, J and L. Krchma. 1990. Allozyme Variation among Populations of the Hessian Fly (Mayetiola destructor) in the United-States. Journal of Heredity. 81:331-337.

Bouhssini, M., Hatchett, J., Cox, T and G. Wilde. 2001. Genotypic interaction between resistance genes in wheat and virulence genes in the Hessian fly Mayetiola destructor (Diptera: Cecidomyiidae). Bulletin of Entomological Research. 91:327-331.

Buntin, G., Ott, S and J. Johnson. 1992. Integration of plant resistance, inseticides, and planting date for management of the Hessian fly (Diptera: Cecidomyiidae) in winter wheat. J. Econ. Entomol. 85:530-538.

Butin, G., Brunker, P and J. Johnson. 1990. Management of Hessian fly (Diptera: Ccidomyiidae) in Georgia by delayed planting of winter wheat. Ann. Entomol. Soc. Amer. 83:1025-1033.

Buchner P. 1965. Endosymbiosis of Animals with Plant Microorganisms. Interscience Publishers, New York. Pg. 909.

Brown-Guedira G., Hatchett, J., Liu, X., Fritz, A., Owuoche, J., Gill, B., Sears, R., Cox, T and M. S. Chen. 2005. Registration of KS99WGRC42 Hessian fly-resistant hard red winter wheat germplasm. Crop Sci. 45:804-805.

Cambron, S., Buntin, G., Weisz, R., Holland, J., Flanders, K., Schemerhorn, B and R. Shukle. 2010. Virulence in Hessian Fly (Diptera: Cecidomyiidae) field collections from the Southeastern United States to 21 Resistance Genes in wheat. J. Econ. Entomol. 103: 2229-2235.

Castle Del Conte, S., Bosque-Pérez, N., Schotzk, J and S. Guyo. 2005. Impact of Tillage Practices on Hessian Fly-Susceptible and Resistant Spring Wheat Cultivars. Journal of Economic Entomology. 98:805-813.

Clement, S., and S. Quisenberry. 1999. Global Plant Genetic Resources for Insect-Resistant Crops, pp. 25-49. Boca Ratón, Florida.

Cox, T and J. Hatchett. 1986. Genetic model for wheat Hessian fly (Diptera: Cecidomyiidae) interaction: strategies for deployment of resistance genes in wheat cultivars. Environmental Entomology. 15:24-31.

Chen, M. S., Echegaray, E., Whitworth, J., Wang, H., Sloderbeck, P., Knutson, A., Giles, K and T. Royer. 2009. Virulence analysis of Hessian fly populations from Texas, Oklahoma, and Kansas. J. Econ. Entomol. 102:774-780.

Chen, M., Fellers, J., Zhu, Y., Stuart, J., Hulbert, S., El-Bouhssini, M and X. Liu. 2006. A super-family of genes coding for secreted salivary gland proteins from the Hessian fly, Mayetiola destructor. J. Insect Sci. 6:1536-2442.

Diehl, S and G. Bush. 1984. An Evolutionary and Applied Perspective of InsectBiotypes. Annual Review of Entomology 29:471-504.

Elmer, J. 1939. Grasses of the tribe Hordeae a hosts of the Hessian fly. J. Econ. Entomol. 32:505-510.

Elmore, J., Lin, Z and G. Coaker. 2011. Plant NB-LRR signaling: upstreams and downstreams. Current Opinion in Plant Biology.14:365-371.

El Boushssini, M., Amri, M and J. H. Hatchett. 1988. Wheat genes conditioning resistance to the Hessian fly (Diptera:Cecidoymiidae) in Morocco. J. Econ. Entomol. 81:709-712.

Gallun, R. L. 1977. Genetic basis of Hessian fly epidemics. Ann N Y Acad Sci 287: 223-229.

Ganehiarachchi, G., Anderson, K., Harmon, J and M. Harris. 2013. Why Oviposit There? Fitness Consequences of a Gall Midge Choosing the Plant's Youngest Leaf. Environ. Entomol. 42:123-130.

Goggin, F., Williamson, V and D. Ullman. 2001. Variability in the response of Macrosiphum euphorbiae and Myzus persicae (Hemiptera: Aphididae) to the tomato resistance gene Mi. Environ. Entomol. 30:101-106.

Gould, F. 1986. Simulation models for predicting durability of insect-resistant germ plasm: Hessian fly (Diptera: Cecidomyiidae)-resistant winter wheat. Environmental Entomology. 15:11-23.

Harris, M., Stuart, J., Mohan, M., Nair, S., Lamb, R and O. Rohfritsch. 2003. Grasses and gall midges: Plant defense and insect adaptation. Annu. Rev. Entomol. 48:549-77.

Harris, M., Freeman, T., Rohfritsch, O., Anderson, K., Payne, S and J. Moore. 2006a. Virulent Hessian Fly (Diptera: Cecidomyiidae) Larvae induce a nutritive tissue during compatible interactions with wheat. Ann. Entomol. Soc. Am. Am. 99: 305-316.

Harris, M., Anderson, K., Anderson, K and H. Kanno. 2006b. Proximate Cues for Reduced Oviposition by Hessian fly on Wheat Plants Attacked by Conspecific larvae. Environ. Entomol. 35:83-93.

Harris, M., Freeman, T., Moore, J., Anderson, K., Payne, S., Anderson, K and O. Rohfritsch. 2010. H-gene-mediated resistance to Hessian fly exhibits features of penetration resistance to fungi. Phytopathology. 100:279-289.

Hatchett, J., Kreitner, G and R. J. Elzinga. 1990. Larval mouthparts and feeding mechanism of the Hessian fly (Diptera: Cecidomyiidae). Ann. Entomol. Soc. Am. Am. 83:1137-1147.

Hosokawa, T., Kikuchi, Y and T. Fukatsu. 2007. How many symbionts are provided by mothers, acquired by offspring, and needed for successful vertical transmission in an obligate insect-bacterium mutualism? Mol Ecol. 16:5316-25.

Johnston, J., Ross, L., Beani, L., Hughes, D and J. Kathirithamby J. 2004. Tiny genomes and endoreduplication in Strepsiptera. Insect Mol Biol. 13:581-585.

Johnson, J., Schemerhorn, B and R. Shukle. 2004. A first assessment of mitochondrial DNA variation and geographic distribution of haplotypes in Hessian fly (Diptera : Cecidomyiidae). Annals of the Entomological Society of America. 97:940-948.

Kanno, H and M. Harris. 2002. Avoidance of occupied hosts by the Hessian fly: oviposition behaviour and consequences for larval survival. Ecological Entomology. 27:177-188.

Li, Ch., Chen, M., Chao, S., Yu, J and G. Bai. 2013. Identification of a novel gene, *H34*, in wheat using recombinant inbred lines and single nucleotide polymorphism markers. Theor Appl Genet. 126:2065-2071. DOI 10.1007/s00122-013-2118-5

Liu, X., Brown-Guedira, G., Hatchett, J., Owuoche, J and M. -S. Chen. 2005. Genetic characterization and molecular mapping of a Hessian fly-resistance gene transferred from T. turgidum ssp. Dicoccum to common wheat. Theor Appl Genet. 111:1308-1315

Liu, X., Gill, B and M. S. Chen. 2005. Hessian fly resistance gene H13 mapped to a distal cluster of R genes in chromosome 6DS of wheat. Theor. Appl. Genet. 111:243-249.

Liu, X., Fritz, A., Reese, J., Wilde, J., Gill, B and M. S. Chen. 2005. H9, H10, and H11 compose a cluster of Hessian fly-resistance genes in the distal gene-rich region of wheat chromosome 1AS. Theor. Appl. Genet. 110:1473-1480.

Liu, X., Smith, C., Friebe, B and B. Gill. 2005. Molecular mapping and allelic relationships of Russian wheat aphid resistance genes. Crop Sci. 45:2273-2280.

Maddur, A., Liu, X., Zhu, Y., Fellers, J., Oppert, B., Park, Y., Bai, Y., Wilde, G and M.-S. Chen. 2006. Cloning and characterization of protease inhibitor-like cDNAs from the Hessian fly, Mayetiola destructor (Say). Insect Mol. Biol.. 15:485-496.

Milligan, S., Bordeau, J., Yaghoobi, J., Kaloshian, I and M. Williamson. 1998. The root-knot nematode resistance gene Mi from tomato is a member of leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. Plant Cell. 10:1307-1319.

Moran, N., Russell, J., Koga, R and T. Fukatsu. 2005. Evolutionary relationships of three new species of Enterobacteriaceae living as symbionts of aphids and other insects. Appl Environ Microbiol. 71:3302-10.

Morgan, G., Sansone, C and A. Knutson. 2005. Hessian fly in Texas Wheat. Texas Cooperative Extension. The Texas A&M University System. http://varietytesting.tamu.edu/wheat/otherpublications/HessianFlyinTexasWheatE-350.pdf

Nombela, G., Williamson, V and M. Muniz. 2003. The root-knot nematode resistance gene Mi-1.2 of tomato is responsible for resistance against the whitefly Benisia tabaci. Mol. Plant Micro. Interact. 16:645-649.

O'brien, D. 2011. World Wheat Market-Demand Trends. K-State research and extension. http://www.agmanager.info/marketing/outlook/newletters/archives/GRAIN-OUTLOOK_03-04-11.pdf

Oliver, K., Russell, J., Moran, N and M. Hunter. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proc Natl Acad Sci USA. 100:1803-7

Painter, R. H. 1968. Crops that resist insects provide away to increase world food supply. Kans. Agric. Exp. Stn. Bull. 520.

Painter, R. H. 1951. Insect resistance in crops plants. Macmillan, New York.

Paine, T., Raffa, K and T. Harrington. 1997. Interactions among Scolytid bark beetles, their associated fungi, and live host conifers. Annu Rev Entomol. 42:179-206.

Ratcliffe, R., Cambron, S., Flanders, K., Bosque-Perez, N., Clement, S and H. Ohm. 2000. Biotype composition of Hessian fly (Diptera: Cecidomyiidae) populations from the southeastern, midwestern, and northwestern United States and virulence to resistance genes in wheat. J. Econ. Entomol. 93:1319-1328.

Ratcliffe, R., Safranski, G., Patterson, F., Ohm, H and P. Taylor. 1994. Biotype status of Hessian Fly (Diptera: Cecidomyiidae) populations from the eastern United States and their response to 14 Hessian fly resistance genes. J. Econ. Entomol. 87:1113-1121.

Ratcliffe, R and J. Hatchett. 1997. Biology and genetics of the Hessian fly andresistance in wheat, pp 47-56. In K. Bondari [ed.], New developments inentomology. Research Signpost, Scientific Information Guild, Trivandum, India.

Refai, F. Y. 1956. The feeding mechanism of Hessian fly larvae. J. Econ. Entomol. 49:182- 184.

Rossi, M., Goggin, L., Milligan, S., Kaloshian, I., Ullman, D and V. Williamson. 1998. The nematode resistance gene Mi of tomato confers resistance against the potato aphid. Proc. Natl. Acad. Sci. U.S.A. 95:9750-9754.

Royer, T. A. 2005. Hessian fly reported in Oklahoma wheat, http://www.ento.okstate.edu/Pddl/2005/PDIA4-1.pdf.

Russell, G. 1978. Plant Breeding for Pest and Disease Resistance. Boston: Butterworths.

Weisz, R. 2012. Small Grain production Guide. http://www.smallgrains.ncsu.edu/_Pubs/PG/_PG2012Rev.pdf

Roberts, J., Foster, J and F. Patterson. 1988. The Purdue-USDA small grain improvement program-a model of research productivity. J. Prod. Agric. 1:239-241.

Robinson, R., Miller, B., Mitchell, H., Mussman, H., Johnson, J and Jones. E. T. 1960. Chloroplast number in leaves of normal wheat plants and those infested with Hessian fly or treated with maleic hydrazide. J. Econ. Entomol.53:560-562.

Shukle, R.H. 2000. Molecular and cytological characterization of an actin gene from Hessian fly (Diptera: Cecidomyiidae). Ann. Entomol. Amer. 93:1164-1172.

Saltzmann, K., Giovanini, M., Zheng, C and C. Williams. 2008. Virulent Hessian fly Larvae Manipulate the Free Amino Acid Content of Host Wheat Plants. J. Chem. Ecol. 34:1401-1410.

Smith, R. 2007. Hessian Fly Takes Big Bite Out of NE Texas Wheat Crop. Southwest Farm Press. http://southwestfarmpress.com/hessian-fly-takes-big-bite-out-ne-texas-wheat-crop

Stuart, J and J. Hatchett. 1987. Morphogenesis and cytology of the salivary gland of the Hessian fly, Mayetiola destructor (Diptera: Cecidomyiidae). Ann. Entomol. Soc. Am. 80:475-82

Stuart, J., Chen, M and M. Harris. 2008. Hessian Fly. USDA Agricultural Research Service-Lincoln, Nebraska. USDA/UNL Faculty.

Stuart, J., Chen, M., Shukle, R and M. Harris. 2012. Gall Midges (Hessian flies) as Plant Pathogen. Ann. Rev. Phytopathol. 50:339-357

Su, Q., Zhou, X and Y. Zhang. 2013. Symbiont-mediated functions in insect hosts. Communicative & Integrative Biology. 6:3, e23804. www.landesbioscience.com

Tan, Ch., Carver, B., Chen, M., Gu, Y and L. Yan. 2013. Genetic association of OPR genes with resistance to Hessian fly in hexaploid wheat. BMC Genomics, 14:369. http://www.biomedcentral.com/1471-2164/14/369

USDA. 2011. Small Grains 2011. Summary (September 2011). USDA, National Agriculture Statistics Service. http://usda01.library.cornell.edu/usda/current/SmalGraiSu/SmalGraiSu-09-30-2011.pdf

Webster, J. A., and P. Kenkel. 1999. Benefits of Managing Small-Grain Pests with Plant Resistance. In: Economic, Environmental, and Social Benefits of Resistance in Field Crops. Entomological Society of America. Lanham, Maryland. 87p.

Williams, C., Nemacheck, J., Shukle, J., Subramanyam, S., Saltzmann, K and R. Shukle. 2011.Induced epidermal permeability modulates resistance and susceptibility of wheat seedlings to herbivoy by Hessian fly larvae. Journal of Experimental Botany. 62:452-4531.

Withworth, J. 2005. Wheat Growers Urged to Consider Hessian Fly Threat Before Planting. http://www.ksre.k-state.edu/news/

Zantoko, L. and R. Shukle. 1997. Genetics of virulence in the Hessian fly to resistance gene H13 in wheat. J. Heredity. 88:120-123.

Zeiss, M., Brandenburg, R and J. van Duyn. 1993. Suitability of seven grass weeds as Hessian fly (Diptera: Cecidomyiidae) hosts. J. Agric. Entomol. 10:107-119.

Zhao, H., Liu, X and M.-S. Chen. 2006. H22, a major resistance gene to the Hessian fly (Mayetiola destructor), is mapped to the distal region of wheat chromosome 1DS. Theor. Appl. Genet. 113:1491-1496.

Chapter 2 - Virulence and Biotype Analyses of Hessian Fly, *Mayetiola destructor* (Say), Populations from Texas, Louisiana, and Oklahoma

Abstract

The Hessian fly, *Mayetiola destructor* Say, is a major pest of wheat, and is controlled mainly through deploying fly-resistant wheat cultivars. The challenge for the plant resistance approach is that virulence of *M. destructor* populations in the field is dynamic and wheat cultivars may lose resistance within 6-8 years. To ensure continuous success of host plant resistance, *M. destructor* populations in the field need to be constantly monitored to determine which resistance genes remain effective in different geographic regions. This study investigated five *M. destructor* populations collected from Texas, Louisiana, and Oklahoma, where infestation by *M. destructor* has been high in recent years. Eight resistance genes including *H12*, *H13*, *H17*, *H18*, *H22*, *H25*, *H26*, and *Hdic*, were found to be highly effective against all tested *M. destructor* populations in this region, conferring resistance to 80% or more of plants containing one of these resistance genes. The frequencies of biotypes virulent to resistance genes *H13* (biotype *vH13*), *H18* (vH18), *H21* (*vH21*), *H25* (*vH25*), *H26* (*vH26*), and *Hdic* (*vHdic*) were determined, and were found to vary from population to population, ranging from 0 to 45%. A logistic regression model was established to predict biotype frequencies based on the correlation between the percentages of susceptible plants obtained in a virulence test and the log-odds of virulent biotype frequencies determined by a traditional approach.

Introduction

The Hessian fly (*Mayetiola destructor* Say) has been a major pest of wheat since it was brought to the U.S. around 1779 during the American Revolutionary War (Buntin 1999, Pauly

2002, Stuart et al. 2012). The *M. destructor* can damage wheat plants from seedling to reproductive stages. In seedling plants, *M. destructor* larvae feed between leaf-sheaths, resulting in growth inhibition and death of the infested plant, unless a new tiller develops (Byers et al. 1971). In reproductive plants, larvae feed on the stem, resulting in plant lodging and reduced grain filling. In recent years, *M. destructor* outbreaks have occurred more frequently, especially in the Southwestern USA, probably due to no-till cultivation and climate change (Royer 2005; Watson 2005; Comis 2007; Knutson and Swart 2007; Smith 2007; Huang et al. 2011).

Currently, the *M. destructor* can be suppressed by seed treatments with systemic insecticides, late planting to avoid infestation in the fall (so called fly-free date), and deployment of resistant wheat cultivars (Buntin and Bruchner 1990, Buntin and Hudson 1991, Buntin 1992, Buntin et al. 1992, Morgan et al. 2005, Giles and Royer 2011). Among these control measures, deployment of resistant cultivars is the most economic and effective means to reduce *M. destructor* infestations. Seed treatment is only effective for about 14 days, and late planting can only be adopted in the northern U.S. and the fly-free date varies from year to year. Currently, 34 resistance genes have been identified and many of them have been deployed to the field (Li et al. 2013). All resistance genes except *h4* are inherited as dominant traits with antibiosis effects on fly larvae. A reliable host plant resistance strategy remains a challenge because *M. destructor* field populations are dynamically changing, and the effectiveness of deployed resistance genes is relatively short-lived (Ratcliffe and Hatchett 1997, Gould 1998, Ratcliffe et al. 2000). Rare and uncommon biotypes virulent to specific resistance genes can become prevalent due to selection pressure after resistance genes have been deployed. Even without selection pressure from deployed resistance genes, *M. destructor* populations are constantly evolving in the field due to genetic adaptation mechanisms formed in the long course of wheat – *M. destructor* co-evolution

(Chen et al. 2010). To safeguard the effectiveness of the plant resistance strategy, the virulence of *M. destructor* populations in different regions must be constantly monitored so that breeders and producers know which resistance genes remain effective in their regions, and which biotypes are currently prevalent.

In the southern U.S., including the states of Georgia, Oklahoma, and Texas, *M. destructor* has historically been a major problem of wheat production. In recent years, heavy infestations have become more frequent and have occurred in larger areas in Oklahoma and Texas (Royer 2005, Watson 2005, Comis 2007, Knutson and Swart 2007, Smith 2007, Alvey 2009). *M. destructor* damage had not been observed until 1989 in Louisiana (Colyer et al. 1989), where the pest has recently become a serious problem of wheat production (Huang et al. 2011). *M. destructor* field virulence in Texas and Oklahoma was first reported in 2009 (Chen et al. 2009) and field virulence in Georgia and several other southern states was reported in 2010 (Cambron et al. 2010).

Objectives

The objectives of the present study were to provide current information on *M. destructor* virulence and biotype compositions in field populations collected from heavily infested areas in Texas, Louisiana, and Oklahoma; and to establish a method to estimate biotype frequency based on results of virulence assays so that the time consuming process for direct biotype analysis of individual females can be avoided in the future.

Materials and Methods

M. destructor Samples

M. destructor populations were sampled by collecting wheat plants infested with *M. destructor* in pupal (so called flaxseeds) stage. Infested plants were packed in boxes, sent to the USDA-ARS *M. destructor* Research Laboratory in Manhattan, KS, and stored in a cold room for 180 d to break diapause before assaying for virulence and biotype. During 2010-2011, *M. destructor* populations from Grayson, McClennan and Hill Counties in Texas were collected and evaluated. A population collected in 2008 from Grayson County, Texas, had been previously analyzed (Chen et al., 2009). *M. destructor* populations from McClennan and Hill Counties, Texas, had not been evaluated since 1987 (Hoelscher et al. 1987). These samples were designated as Grayson-TX-FD-11, McClennon-TX-FD-10, and Hill-TX-FD-11, respectively. FD represents 'field' samples analyzed directly without increasing the population in greenhouse. A *M. destructor* population from Franklin Parish, Louisiana, was collected in 2011, and a population from Okeene, Oklahoma collected in 2012. Prior to the current study, the status of *M. destructor* virulence in these two locations was unknown. Due to their small size, the Oklahoma and Louisiana populations were increased in the greenhouse for one generation before analyses, and were designated as Franklin-LA-GH-11 and Okeene-OK-GH-12, respectively (GH - 'greenhouse' increased).

Greenhouse Increase of M. destructor Populations

M. destructor samples containing less than 8,000 pupae were increased in the greenhouse before virulence and biotype assays. Wheat stubble collected from fields was placed in a mesh tent (243.8 x 61 x 91.4 cm) in the greenhouse to facilitate *M. destructor* adult emergence. When adult flies started to emerge, approximately 3,500 seedlings of 'Karl 92, a *M. destructor*-susceptible

cultivar, at the 1.5 leaf stage were placed into the cage to collect eggs. Adult flies were allowed to lay eggs on the plants for three to four days, depending on egg densities. When egg density reached an average of ~8 eggs per plant, seedlings were transported to a different greenhouse to allow eggs to hatch and larvae to develop to pupation. Newly obtained pupae were collected along with wheat seedlings and stored in a cool room for at least three months before assays (Chen et al., 2009).

Virulence Analysis

A set of 22 wheat cultivars, each carrying a different *M. destructor*-resistance gene or gene combination, was assembled. This set of wheat cultivars contained resistant genes *H3, H5, H6, H7/H8, H9, H10, H11, H12, H13, H14, H16, H17, H18, H19, H21, H22, H23, H24, H25, H26, H31*, and *Hdic*. Twenty to 25 seeds of each testing line were planted in a row in a randomized design in a 54 x 36 x 8 cm flat, which included two rows of Karl 92 in the middle of each flat. To maintain *M. destructor* populations, additional flats of Karl 92 seedlings were planted and used to collect eggs. Wheat stems containing *M. destructor* pupae were placed into a 243.8 x 61 x 91.4 cm greenhouse tent three days before the virulence test to facilitate *M. destructor* adult emergence. Water was sprayed daily onto the tent to maintain moisture. When *M. destructor* adults started to emerge, seedlings at the 1.5 leaf-stage were placed inside the same tent for infestation.

Female flies oviposit on the adaxial surface of plants in a free-choice manner. To reduce variations in the test results, infestation was terminated when the number of eggs reached an average of 8 eggs per plant (Chen et al., 2009). Plants were categorized as susceptible or resistant 21 days after infestation. Plants were recorded as resistant if they contained dead (slim and reddish) first-instar larvae and were growing normally. Plants were recorded as susceptible if

they were stunted and contained live (fat and white) larvae. Plants with no dead or alive larvae were categorized as escapes and were excluded from analysis.

As described previously (Chen et al., 2009), a gene was defined as highly resistant to a *M. destructor* population if ≥ 80% plants of the wheat cultivar containing the gene were identified to be resistant to the *M. destructor* population in a virulence assay. A gene was considered as moderately resistant to a fly population if 50-80% plants of the variety containing the gene were resistant to the insect population, while a gene was judged as susceptible if <50% plants of the variety containing the gene were resistant to the population.

Analysis of Biotype Composition

Six resistant genes (*H13*, *H18*, *H21*, *H25*, *H26*, and *Hdic*) shown previously to confer ≥80% of plants resistant to a Texas field population (Chen et al. 2009) were selected to determine *M. destructor* biotype frequencies. Biotype composition was determined using a procedure similar to that described by Ratcliffe et al. (1994).

Relationship Between Percentages of Susceptible Plants and Virulence

Chi-square tests were performed using percent resistant plants in virulence tests to determine if *M. destructor* populations were dependent on or independent of cultivars carrying different resistant genes. A P-value of <0.0005 suggests that resistant genes are not independent of *M. destructor* populations. Comparison of the percentages of resistant plants obtained with different fly populations was conducted by ANOVA of virulence test data for each resistance gene. Twenty-two resistant genes tested and a multiple comparison adjustment based on Bonferroni correction (Bonferroni 1935) was made to declare significance for each test. A virulence test was considered significant at $\alpha = 0.05$ level when the P-value was less or equal to $0.05/22 = 0.0022$. For a test yielding a significant P-value, pairwise comparisons were conducted

using the Tukey's Honestly Significant Difference (HSD) method to identify populations differing from others on percentages of resistant plants. The Tukey's HSD adjusted P-values were obtained and compared to 0.05 to identify populations that differ significantly in the percentages of resistant plants for each gene.

To model the relationship between biotype composition and the percentages of susceptible plants, a logistic regression model was developed, using the counts of virulent and avirulent *M. destructor* females as the response variable. Significance predictors included in the model were the percentages of susceptible plants in virulence tests, the type of resistant gene, and their interaction. All predictors were highly significant at $P \leq 0.01$ (the *P*-values based on deviance test from the logistic regression model for these factor/variables are < 2.2x, 2.133x, and 0.0011, respectively). Pearson correlation coefficients between the observed percent of virulent biotypes and their predicted value from the logistic model were calculated. All the statistical analyses were conducted with R software version 3.0.1 (http://cran.r-project.org/).

Results

Virulence of Three Texas M. destructor Populations

Eight genes, *H12*, *H13*, *H17*, *H18*, *H21*, *H22*, *H25*, *H26*, and *Hdic*, were highly resistant to all three Texas *M. destructor* populations, conferring resistance in 80% or more of plants containing one of these genes to the three fly populations (Table 2. 1), a level considered highly effective based on historic observations (Ratcliffe et al. 2000, Chen et al. 2009, Cambron et al. 2010). In addition, *H3* and *H11* conferred resistance in 80% or more of plants containing either of these genes to the Grayson-TX-FD-11 and Hill-TX-FD-11 populations. *H5* conferred resistance in 82% of plants to the Grayson-TX-FD-11 population. *H9*, *H16*, *H19*, and *H23*, exhibited moderate resistance, conferring resistance in 50% or more of plants containing one of

these genes to at least one of the three Texas populations. The *H6*, *H7H8*, *H10*, *H14*, *H24*, and *H31* genes or gene combination conferred resistance in less than 50% of plants containing one of the genes to any one of the three Texas populations.

Virulence of a Louisiana M. destructor Population

Fourteen genes, *H3*, *H9*, *H12*, *H13*, *H16*, *H17*, *H18*, *H19*, *H21*, *H22*, *H24*, *H25*, *H26*, and *Hdic*, were highly effective against the Franklin-LS-GH-12 population, conferring resistance in 80% or more of plants containing one of these genes (Table 2. 2). *H10*, *H14*, and *H23* exhibited moderate resistance, conferring resistance in 50 to 80% of plants containing one of the genes. *H5*, *H6*, *H7H8*, *H31* were less effective, conferring resistance in only 50% or less of plants containing one of these genes.

Virulence of an Oklahoma M. destructor Population

H3, *H5*, *H6*, *H12*, *H13*, *H17*, *H18*, *H19*, *H21*, *H22*, *H25*, *H26*, and *Hdic* were highly effective against the Okeene-OK-GH-12 population, conferring resistance in 80% or more of plants containing one of these genes (Table 2. 3). *H9*, *H10*, *H11*, and *H16* exhibited moderate resistance, conferring resistance in 50 to 80% of plants containing one of the genes, and *H7H8*, *H14*, *H23*, *H24*, and *H31* were less effective, conferring resistance in only 50% or less of plants containing one of these genes.

Biotype Composition Analyses

Although 100 females were analyzed from each of the three Texas populations, success rates were only 21 to 83% as some females failed to oviposit (Table 2. 4). The percentages of biotypes virulent to the wheat resistant genes *H13* (biotype vH13), *H18* (vH18), *H21* (vH21), *H25* (vH25), *H26* (vH26), and *Hdic* (vHdic) were 4%, 5%, 16%, 10%, 13%, and 0%, respectively, for the Grayson-TX-FD-11 population; 1%, 21%, 45%, 25%, 10%, and 4%, respectively, for the

McClennan-TX-FD-10 population; and 5%, 19%, 22%, 22%, 13%, and 2%, respectively, for the Hill-TX-FD-11 population.

In the Louisiana and Oklahoma populations, fewer females were analyzed due to limited availability of mated females when plants were at the appropriate stage for evaluation. The virulence frequencies of the Oklahoma and Louisiana populations were much lower. The frequency of vH13, vH18, vH21, vH25, vH26, and vHdic was 0%, 0%, 14%, 10%, 11%, and 0%, respectively, for the Franklin-LS-GH-12 population, and 0%, 0%, 27%, 18%, 9%, and 0%, respectively, for the Okeene-OK-GH-12 population.

Significant differences were observed in the percentages of resistant plants carrying *H11*, *H16*, or *H24* to the three Texas populations, with *P*-values 2.94x, 1.59 x, and 5.92 x, respectively (Table 2. 5). When all five populations are compared, the percentages of resistant plants showed differences among some populations for the following genes (*P*-value in parenthesis): *H5* (3.93x), *H6* (1.27x), *H9* (1.16x), *H11* (4.11x), *H14* (5.63x), *H19* (1.35x), *H24* (4.39x), and *H31* (1.79x).

Relationship Between Percentages of Susceptible Plants and Virulence

Virulence tests and biotype frequency analyses were conducted simultaneously on the five *M. destructor* populations (Table 2. 4). The percentages of susceptible plants obtained in virulence tests conveyed strong information about the frequencies of virulent biotypes obtained in biotype analyses. A logistic regression that predicted biotype frequencies (Figure 2. 1), yielded predicted values quite similar to and highly correlated with actual observed values (r = 0.93, 95% confidence interval of the correlation is (0.86, 0.97), *P*-value < 8.4x). We further tested the model with data reported by Chen et al. (2009), and the predicted frequencies of biotypes virulent to *H13* (vH13) and *H18* (vH18) are fairly consistent with the observed values (Figure 2. 1 F). However, the predicted frequencies for biotypes vH21 and vH26 are 9.5% and 9.7%,

respectively, which differ from the corresponding observed frequencies of 4% and 22%, respectively.

Discussion

Variation in virulence among *M. destructor* populations from different regions has been observed previously (Chen et al., 2009, Cambron et al., 2010). In this study, we also observed variation in virulence among the five regional populations (Table 2. 5). Specifically, wheat cultivars carrying *H11, H16*, or *H24* showed significant difference in percentages of resistant plants among the three Texas populations. Wheat cultivars carrying *H5*, *H6*, *H9*, *H11*, *H14*, *H19*, *H24*, or *H31* showed significant difference in percentages of resistant plants when all five populations were compared. For example, 94% of the plants with the *H5* gene were susceptible to the Franklin-LA-GH-11 population, whereas only 18% of plants with this gene were susceptible to the Grayson-TX-FD-11. Variation in sample collection and testing conditions might have partially affected results, but the standardized virulence detection procedure yields fairly consistent results as described previously (Chen et al. 2009). Thus, the variation in population virulence was likely due to differences in biotype composition of the five *M. destructor* populations. Among possible factors that might have caused difference in population virulence, regional selection pressure from deployment of cultivars containing specific resistance genes is likely playing a role. This can be seen from the detection of the most virulent Hessian fly population to date in Kay County in Oklahoma, the nursery site where new cultivars containing various *M. destructor* resistance genes are being tested (Chen et al. 2009). The virulence of the Kay county population is very different from that of populations collected from neighboring regions including the Okeene-OK-GH-12 population. Another possible reason for virulence variation among regional *M. destructor* populations is the intrinsic mechanism that allows *M. destructor* to constantly generate genetic diversification in genes likely involved in virulence, which include large numbers of genes encoding putative effector proteins (Chen et al. 2010).

One of the main objectives of this study was to provide information on which resistance genes are still effective to *M. destructor* populations in the field in three southern states. Despite the variation in virulence described above, some resistance genes conferred resistance against all of the *M. destructor* populations tested. As shown in Table 2. 5, 80% or more of plants containing *H12*, *H13*, *H17*, *H18*, *H22*, *H25*, *H26*, or *Hdic*, were resistant to all five tested populations, indicating that these genes remain highly effective in protecting wheat in these regions. *Hdic*, a newly identified resistance gene from *Triticum turgidum* ssp. *dicoccum* (Liu et al. 2005), exhibited the highest effectiveness, conferring resistance in > 98% of plants containing *Hdic* to all five *M. destructor* populations. In addition, *H3* and *H21* also conferred resistance in 80% or more plants to four of the five populations, and in nearly 70% plants to the McClennan-TX-FD-10 population. However, *H3* was reported to be ineffective to other populations collected from southern U. S. (Cambron et al. 2010). Accordingly, we recommend that *H12*, *H13*, *H17*, *H18*, *H22*, *H25*, *H26*, and *Hdic* to be used in breeding programs against *M. destructor* infestations in Texas, Louisiana, and Oklahoma.

The determination of biotype frequencies in field populations is time-consuming, and becomes increasingly un-achievable as the numbers of the identified resistance genes increase. In the wheat–*M. destructor* interaction, a previous study indicated the frequency of a virulent biotype is proportional to the percentage of susceptible plants that carry a resistance gene in a virulence test (Chen et al. 2009). In the current study, a logistic regression model predicted frequencies of biotypes virulent to some genes, but was less reliable for other genes. This discrepancy indicates that further data accumulation and improvement of model parameters are needed for more accurate biotype predictions. Despite this need for improvement, the strong relationship between virulence frequency and the percent of susceptible plants in virulence tests suggests that it may be practical to predict biotype frequency based on virulence test results.

References

Alvey, D. 2009**.** Evaluation of the impact of Hessian fly (*Mayetiola destructor*) on Oklahoma winter wheat systems. MS Thesis, Oklahoma State University.

Bonferroni, C. 1935. "Il calcolo delle assicurazioni su gruppi di teste." In Studi in Onore del Professore Salvatore Ortu Carboni. Rome: Italy, pp. 13-60.

Buntin, G. 1992. Assessment of a microtube injection system for applying systemic insecticides at planting for Hessian fly control in winter wheat. Crop Protection. 11: 366-370.

Buntin, G. 1999. Hessian fly (Diptera: Cecidomyiidae) injury and loss of winter wheat grain yield and quality. J. Econ. Entomol. 92: 1190-1197.

Buntin, G and P. Bruckner. 1990. Effect of planting date on Hessian fly infestation and production of triticale. Appl. Agric. Res. 5: 82-88.

Buntin, G and R. Hudson. 1991. Spring control of the Hessian fly (Diptera: Cecidomyiidae) in winter wheat using insecticides. J. Econ. Entomol. 84: 1913-1919.

Buntin, G., Ott, S and J. Johnson. 1992. Integration of plant resistance, insecticides, and planting date for management of the Hessian fly (Diptera: Cecidomyiidae) in winter wheat. J. Econ. Entomol. 85: 530-538.

Byers, R and R. Gallun. 1971. Ability of the Hessian fly to stunt winter wheat. I. Effect of larval feeding on elongation of leaves. J. Econ. Entomol. 65: 955-958.

Cambron, S., Buntin, G., Weisz, R., Holland, J., Flanders, K., Schemerhorn, B and R. Shukle. 2010. Virulence in Hessian Fly (Diptera: Cecidomyiidae) field collections from the Southeastern United States to 21 Resistance Genes in wheat. J. Econ. Entomol. 103: 2229-2235.

Chen, M. S., Echegaray, E., Whitworth, J., Wang, H., Sloderbeck, P., Knutson, A., Giles, K and T. Royer. 2009. Virulence analysis of Hessian fly populations from Texas, Oklahoma, and Kansas. J. Econ. Entomol. 102: 774-780.

Chen, M.S., Liu, X., Yang, Z., Zhao, H., Shukle, R., J. Stuart, J and S. Hulbert. 2010. Unconventional conservation among genes encoding small secreted salivary gland proteins from a gall midge. BMC Evol. Biol. 10: 296.

Clement, S., Elberson, L., Young, F., Ratcliffe, R., Alldredge, J and C. Hennings. 2003. Changing crop rotation and tillage practices increase Hessian fly populations in wheat. J. Kan. Entomol. Soc. 76: 567-577.

Comis, D. 2007. Virulent Hessian flies renew attack on U.S. wheat. USDA Agricultural Research – News and Events. January 24, 2007. http://www.ars.usda.gov/is/pr/2007/070124.htm.

Colyer, P., Vernon, P and T. Riley. 1989. First report of Hessian fly (Diptera: Cecidomyiidae) on winter wheat in Louisiana. J. Agric. Entomol. 7: 147-148.

Giles, K and T. Royer. 2011. Characterization of insect populations and resistance in wheat. Partners in Progress: Wheat Research at OSU 2010. Oklahoma State University. DASNR-OAES-OCES. Stillwater, OK.

Gould, F. 1998. Sustainability of transgenic insecticidal cultivars: Integrating pest genetics and ecology. Annu. Rev. Entomol. 43: 701-726.

Hoelscher. C., Turney, H and M. Lidell. 1987. Hessian Oy in Texas wheat. Texas Agricultural Extension Service. (Publication 8-1518) College Stalion, 8 pp.

Huang, F., Harrison, S., Leonard, B., Ghimire, M and P. Price. 2011. Hessian fly, an emerging wheat pest in Louisiana. Louisiana Agriculture Magazine http://text.lsuagcenter.com/en/communications/publications/agmag/Archive/2011/Winter/Hessian-Fly.htm

Knutson, A and J. SwartJ. 2007. Hessian fly outbreaks showing up in many north Texas wheat fields. Southwest Farm Report. (http://southwestfarmpress.com/grains/041807-hessian-fly/).

Li, C., Chen, M., Chao, S., Yu, J and G. Bai. 2013. Identification of a novel gene, *H34*, in wheat using recombinant inbred lines and single nucleotide polymorphism markers. Theor. Appl. Genet. In press.

Liu, X., Brown-Guedira, G., Hatchett, J., Owuoche, J and M. S. Chen. 2005. Genetic characterization and molecular mapping of a Hessian fly resistance gene *Hdic* transferred from *T. turgidum* ssp. *dicoccum* to common wheat. Theor. Appl. Genet. 111: 1308-1315.

Morgan, G., Sansone, C and A. Knutson. 2005. The Hessian Fly in Texas wheat. E-350. Texas AgriLife Extension Bulletin. http://agrilifebookstore.org.

Pauly, P. J. 2002. Fighting the Hessian fly. Environ Hist 7: 385-507.

Ratcliffe, R., Cambron, S., Flanders, K., Bosque-Perez, N., Clement, S and H. W. Ohm. 2000. Biotype composition of Hessian fly (Diptera: Cecidomyiidae) populations from the southeastern, Midwestern, and northwestern United States and virulence to resistance genes in wheat. J. Econ. Entomol. 93: 1319-1328.

Ratcliffe, R and J. Hatchett. 1997**.** Biology and genetics of the Hessian fly and resistance in wheat. In: Bondari K (ed) New Developments in Entomology. Research Signpost, Trivandrum, India, pp 47-56

Ratcliffe, R., Safranski, G., Patterson, F., Ohm, H and P. Taylor. 1994. Biotype status of Hessian fly (Diptera: Cecidomyiidae) populations from the eastern United States and their response to 14 Hessian fly resistance genes. J. Econ. Entomol. 87: 1113-1121.

Royer, T. A. 2005. Hessian Fly Reported in Oklahoma Wheat. (http://www.ento.okstate.edu/Pddl/2005/PDIA4-1.pdf).

Smith, R. 2007. Widespread outbreaks of Hessian fly reported in wheat. Southwest Farm Report. http://southwestfarmpress.com/grains/051807-hessian-fly/

Watson, S. 2005. Hessian fly problems have been increasing in recent years in the Central Plains. The Wheat Farmer/Row Crop Farmer. 9 (3): 4-5.

Figure and Tables

Table 2.1 Percent ± SD *M. destructor* resistance in plants with different wheat *M. destructor* resistance genes to the three *M. destructor* populations Grayson-TX-FD-11, McClennan-TX-FD-10, and Hill-TX-FD-11.

R Gene	Wheat Cultivar	Hessian Fly Population Collection Site											
		Grayson-TX-FD-11				McClennan-TX-FD-10				Hill-TX-FD-11			
		R	S	%R	SD	R	S	%R	SD	R	S	%R	SD
H3	Ike	69	7	**91**	1.3	43	18	70	3.3	61	15	**80**	2.8
H5	Magnum	56	12	**82**	4.7	31	30	51	3.6	41	24	63	5
H6	Cadwell	5	59	8	1	20	37	35	1.8	33	39	46	1
H7H8	Seneca	0	78	0	0	13	46	22	1.7	0	75	0	0
H9	Iris	23	50	32	3.1	46	12	79	1.3	38	30	56	3.9
H10	Joy	26	29	47	3.4	17	38	31	1	28	42	40	3.6
H11	Karen	59	0	**100**	1.7	24	37	39	1.4	60	8	**88**	2.6
H12	**Lola**	**66**	**3**	**96**	**1.9**	**36**	**9**	**80**	**3.5**	**51**	**10**	**84**	**1.3**
H13	**Molly**	**55**	**4**	**93**	**1.7**	**47**	**2**	**96**	**1**	**54**	**2**	**96**	**1.3**
H14	D6647 H14	12	58	17	1.2	8	51	14	1.4	0	57	0	0
H16	D6647 H16	52	17	75	1.4	20	42	32	1.4	57	16	78	3.3
H17	**D6647 H17**	**45**	**5**	**90**	**2.6**	**45**	**9**	**83**	**2.9**	**64**	**9**	**88**	**2**
H18	**Redland**	**64**	**2**	**97**	**0.8**	**45**	**6**	**88**	**1.7**	**61**	**5**	**92**	**4.3**
H19	84702B14	53	14	79	5.7	7	35	17	1.3	17	45	27	1.7
H21	**Hamlet**	**51**	**13**	**80**	**1.5**	**20**	**9**	**69**	**0.8**	**68**	**10**	**87**	**1.4**
H22	**KS85WGRC01**	**75**	**0**	**100**	**1.3**	**64**	**1**	**98**	**0.8**	**70**	**3**	**96**	**2.1**
H23	KS89WGRC03	40	34	54	2.4	20	31	39	1.2	25	52	32	2.6
H24	KS89WGRC06	23	50	32	1	1	66	1	0.5	21	56	27	1.9
H25	**KS92WGRC20**	**70**	**5**	**93**	**2.4**	**57**	**3**	**95**	**1**	**67**	**6**	**92**	**3.4**
H26	**KS93WGRC26**	**60**	**6**	**91**	**2.6**	**54**	**5**	**92**	**0.6**	**69**	**5**	**93**	**2.5**
H31	921696-H31	12	57	17	1.8	8	45	15	0.8	8	63	11	1.8
Hdic	**KS99WGRC42**	**64**	**0**	**100**	**3.2**	**63**	**1**	**98**	**1.5**	**76**	**1**	**99**	**1.2**

Bold font indicates plants exhibiting 80% or more resistance.
Tan highlight indicates plants exhibiting 80% or more resistance to all three populations.

Table 2.2 Percent ± SD *M. destructor* resistance in plants with different wheat *M. destructor* resistance genes to the Hessian fly population Franklin-LS-GH-11.

R		Number of plants		Percent ± SD
Gene	Wheat Cultivar	Resistant	Susceptible	Resistant plants
H3	**Ike**	**67**	**4**	**94±1.3**
H5	Magnum	4	59	6±1.2
H6	Cadwell	16	55	23±1.8
H7H8	Seneca	12	64	16±2.9
H9	**Iris**	**67**	**0**	**100±1.5**
H10	Joy	50	17	75±2.5
H11	Karen	19	40	32±1.3
H12	**Lola**	**66**	**0**	**100±0.6**
H13	**Molly**	**60**	**0**	**100±0.8**
H14	D6647 H14	37	17	69±5.4
H16	**D6647 H16**	**57**	**0**	**100±1**
H17	**D6647 H17**	**56**	**5**	**92±4.3**
H18	**Redland**	**62**	**0**	**100±1.3**
H19	**84702B14**	**55**	**0**	**100±1.3**
H21	**Hamlet**	**63**	**7**	**90±1.5**
H22	**KS85WGRC01**	**67**	**0**	**100±1.5**
H23	KS89WGRC03	34	30	53±1.7
H24	**KS89WGRC06**	**68**	**2**	**97±1.8**
H25	**KS92WGRC20**	**72**	**5**	**94±0.8**
H26	**KS93WGRC26**	**70**	**4**	**95±0.6**
H31	921696-H31	27	34	44±3.1
Hdic	**KS99WGRC42**	**65**	**0**	**100±1.5**

Bold font indicates plants exhibiting 80% or more resistance.

Table 2.3 Percent ± SD *M. destructor* resistance in plants with different wheat *M. destructor* resistance genes to the Hessian fly population Okeene-OK-GH-12.

R		Number of plants		Percent ± SD
Gene	Wheat Cultivar	Resistant	Susceptible	Resistant plants
H3	**Ike**	**60**	**2**	**97±2.2**
H5	**Magnum**	**50**	**12**	**81±1**
H6	**Cadwell**	**46**	**6**	**88±5**
H7H8	Seneca	0	61	0±**0**
H9	Iris	28	24	54±2.2
H10	Joy	38	17	69±3.1
H11	Karen	29	11	73±2.2
H12	**Lola**	**60**	**1**	**98±4.8**
H13	**Molly**	**62**	**0**	**100±2.1**
H14	D6647 H14	8	37	18±1.6
H16	D6647 H16	41	15	73±5.6
H17	**D6647 H17**	**56**	**2**	**97±3.5**
H18	**Redland**	**54**	**0**	**100±4.7**
H19	**84702B14**	**42**	**7**	**86±3.9**
H21	**Hamlet**	**46**	**10**	**82±6.8**
H22	**KS85WGRC01**	**70**	**0**	**100±0.6**
H23	KS89WGRC03	18	43	30±5.3
H24	KS89WGRC06	23	43	35±3.5
H25	**KS92WGRC20**	**68**	**2**	**97±2.7**
H26	**KS93WGRC26**	**66**	**1**	**99±1**
H31	921696-H31	20	38	34±0
Hdic	**KS99WGRC42**	**56**	**0**	**100±4.7**

Bold font indicates plants exhibiting 80% or more resistance.

Table 2.4 Biotype composition of *M. destructor* populations from Louisiana, Oklahoma and Texas

		Hessian fly resistance gene					
Hessian fly population	**Virulence**	***H13***	***H18***	***H21***	***H25***	***H26***	***Hdic***
Grayson-TX-FD-11	# Avirulent	80	79	70	75	72	83
	# Virulent	3	4	13	8	11	0
	%Virulent	4	5	16	10	13	0
McClennan-TX-FD-10	# Avirulent	68	65	38	52	74	79
	# Virulent	1	17	31	17	8	3
	%Virulent	1	21	45	25	10	4
Hill-TX-FD-11	# Avirulent	61	44	50	50	47	53
	# Virulent	3	10	14	14	7	1
	%Virulent	5	19	22	22	13	2
Franklin-LS-GH-11	# Avirulent	21	54	18	19	48	54
	# Virulent	0	0	3	2	6	0
	%Virulent	0	0	14	10	11	0
Okeene-OK-GH-12	# Avirulent	22	54	16	18	49	54
	# Virulent	0	0	6	4	5	0
	%Virulent	0	0	27	18	9	0

% Virulent = Percentage of *M. destructor* female offspring virulent to the corresponding wheat resistance gene.

Table 2.5 Percent ± SD *M. destructor* resistance in plants with different *M. destructor* resistance genes resistant to Hessian fly populations from Texas, Louisiana, and Oklahoma.

		Percent ± SD Resistant plants				
R Gene	**Wheat Cultivar**	**Grayson-TX-FD-11**	**McLennan-TX-FD-10**	**Hill-TX-FD-11**	**Franklin-LA-GH-11**	**Okeene-OK-GH-12**
H3	Ike	**91±1.3**	70±3.3	**80±2.8**	**94±1.3**	**97±2.2**
H5	Magnun	**82±4.7**	51±3.6	63±5	6±1.2	**81±1.0**
H6	Cadwell	8±1	35±1.8	46±1	23±1.8	**88±5.0**
H7H8	Seneca	ND	22±1.7	ND	16±2.9	ND
H9	Iris	32±3.1	79±1.3	56±3.9	**100±1.5**	54±2.2
H10	Joy	47±3.4	31±1.0	40±3.6	75±2.5	69±3.1
H11	Karen	**100±1.7**	39±1.4	**88±2.6**	32±1.3	73±2.2
H12	Lola	**96±1.9**	**80±3.5**	**84±1.3**	**100±0.6**	**98±4.8**
H13	Molly	**93±1.7**	**96±1.0**	**96±1.3**	**100±0.8**	**100±2.1**
H14	D6647-H14	17±1.2	14±1.4	ND	69±5.4	18±1.6
H16	D6647-H16	75±1.4	32±/1.4	78/3.3	**100±1.0**	73±5.6
H17	D6647-H17	**90±2.6**	**83±2.9**	**88±2**	**92±4.3**	**97±3.5**
H18	Redland	**97±0.8**	**88±1.7**	**92±4.3**	**100±1.3**	**100±4.7**
H19	84702B14	79±5.7	17±1.3	27±1.7	**100±1.3**	**86±3.9**
H21	Hamlet	**80±1.5**	69±0.8	**87±1.4**	**90±1.5**	**82±6.8**
H22	KSWRCG 01	**100±1.3**	**98±0.8**	**96±2.1**	**100±1.5**	**100±0.6**
H23	KSWRCG 06	54±2.4	39±1.2	32±2.6	53±1.7	30±5.3
H24	KSWRCG 03	32±1	1±0.5	27±1.9	**97±1.8**	35±3.5
H25	KSWRCG 20	**93±2.4**	**95±1.0**	**92±3.4**	**94±0.8**	**97±2.7**
H26	KSWRCG 26	**91±2.6**	**92±0.6**	**93±2.5**	**95±0.6**	**99±1.0**
H31	P921696A1	17±1.8	15±0.8	11±1.8	44±3.1	34±0.0
Hdic	KSWRCG 42	**100±3.2**	**98±1.5**	**99±1.2**	**100±1.5**	**100±4.7**

Bold font indicates plants exhibiting 80% or more resistance.
Tan highlight indicates plants exhibiting 80% or more resistance to all five populations.
ND – not determined.

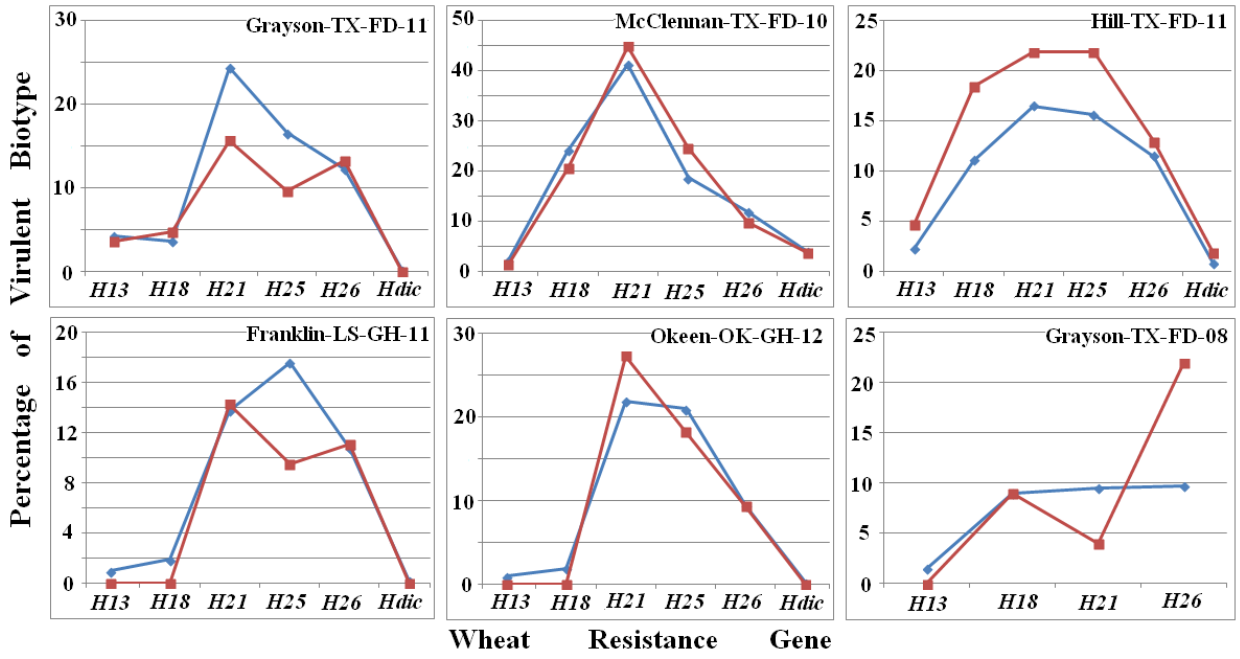


Figure 2.1 Biotype frequency prediction based on virulence test results.

The curves with squares represent the actual biotype frequencies obtained following the procedure described by Ratcliffe et al. (1994) (Table 4). The curves with diamonds represent predicted biotype frequencies using the logistic regression model $y=100(1+e^{-f(x)})$, where y is the percentages of virulent biotypes and f(x) is a function of the percentages of susceptible plants (percSP) in a virulence assay with cultivars carrying a specific resistance gene. Mathematically, f(x) = -4.629 + 0.668 *I(Gene = 'H*18*') + 2.086*I(Gene = '*H21*') + 3.521* I(Gene = '*H25*') + 2.326*I(Gene = '*H26*') -1.869*I(Gene = '*Hdic*') + 0.217*percSP + 0.0178*I(Gene = '*H18*')*percSP – 0.147*I(Gene = '*H21*')*percSP – 0.290*I(Gene = '*H25*')*percSP – 0.180*I(Gene = '*H26*')*percSP + 1.44*I(Gene = '*Hdic*')*percSP, and I(condition) is an indicator function that takes value 1 if the condition is true and takes value 0 if the condition is false. The formula can be simplified to $y=100(1+e^{-(B+b+K_x+k_x)})^{-1}$ for a specific calculation, where B is the Intercept, b is the modification factor to the intercept associated with a specific resistance gene, K is the slope, and k is the modification factor to the slope associated with a specific resistance gene, and x is percSP for a specific resistance gene. For example, to calculate the predicted biotype vH18 based on the Grayson-TX-FD-11, the calculation is: $y=100(1+e^{-(B+b+K_x+k_x)})^{-1} =100(1+ e^{-(-4.629+0.668+0.217x+0.0178x)})^{-1} =100/ (1+ e^{-(-4.629+0.668+0.217(3)+0.0178(3))})^{-1}$ = 3.7%. The data shown in the last graph with the Grayson-TX-FD-08 was based on results of a previous report (Chen et al., 2009).

Chapter 3 - Levanase/Inulase-like genes with possible bacterial origin in the Hessian Fly genome, *Mayetiola destructor* (Say).

Abstract

Previous characterization of a bacterial artificial chromosome (BAC) clone fromidentified a gene encoding a protein with sequence similarity to bacterial levanases (Chen et al., unpublished observation). Blast searching with the levanase-like protein identified 14 levanse/inulase-like genes or gene fragments in the *M. destructor* genome sequence. The 14 genes were named *MDL-1* to *MDL-14*. In this study, the expression level of these genes in different developmental stages and different tissues of 3 d old *M. destructor* larvae were determined. Sequence analysis revealed that six genes encode full length proteins, that three are truncated at the 5' end, and that five are truncated at the 3' end. Sequences of putative proteins showed approximately 42% similarity to bacterial levanases or inulases, and 36% similarity to fungal levanases or inulases. No sequence similarities were found with any known animal or plant proteins. This result suggests that levanase/inulase-like genes in the *M. destructor* genome were originated from bacteria through horizontal gene transfer. Of the 14 levanse/inulase-like genes, six genes, *MDL-3, MDL-5, MDL-9, MDL-11, MDL-12,* and *MDL-13*, are full length, and contain a promoter region, three exons, two introns, and a 3' untranslated region. Three genes, *MDL-1, MDL-8* and *MDL-14* are truncated at the 5' end, and five genes, *MDL-2, MDL-4, MDL-6, MDL-7,* and *MDL-13* are truncated at the 3' end. Comparative analysis of sequences among the 14 MDL genes revealed that positions for intron/exon boundaries are conserved among different genes even though the length of each intron and exon varied among different genes. Reverse-transcription PCR (RT-PCR) and quantitative real-time PCR (qPCR) of different *M. destructor* developmental stages and tissues yielded differential expression patterns of the

levanase/inulase-like genes. Interestingly, *MDL-2, MDL-9* and *MDL-10* presented alternative splicing bands in different developmental stages, and *MDL-3* and *MDL-9* exhibited splicing bands in different tissues of 3 d old *M. destructor* larvae. Results of this study will be useful for future characterization of MDL genes and their function in plant- *M. destructor* interactions.

Introduction

Polymers of fructose molecules are called fructans (Vijn and Smeekens 1999). Based on the linkage position of the fructose residues, two types of fructans exist in nature, inulin and levan. In inulins, the fructosyl residues are linked by β-2,1-linkages, whereas in levan, the fructosyl residues are linked by β-2, 6-linkages (Martel et al.2011). Inulins are produced by many types of plants and are typically used as a means of storing energy in roots and rhizomes (Scotta et al. 2011). Inulins are degraded by the enzyme inulinase, which is also called inulase or 2,1-β-D-fructan fructanohydrolase (Brevnova et al. 1998). Inulinases are divided into endo-inulinases and exo-inulinases (Scott et al. 2011). Endo-inulinases degrade inulins within the polysaccharide chains into oligofructans, whereas exo-inulinases degrade inulins or oligofructans at the end of the chain, yielding simple sugars fructose and glucose. Inulin is increasingly used in processed foods because it has unusually adaptable characteristics and can enhance the growth of bifidobacteria colon (Guibellini et al. 2009).

Levan can be produced by either plants or microorganisms. In plants, levan has been found as a reserve carbohydrate in many monocotyledons, including grasses (*Dactilis glomerata*, *Poa secunda*, and *Agropyrum cristatum*), wheat (*Triticum* spp.) and barley (*Hordeum vulgare* L.) (Vijn and Smeekens, 1999). Levans are widely distributed in different bacteria as storage of carbon and energy and have been reported in *Eschericha coli* (Romberg 2001), *Bacillus subtilis* (Matrin et al. 1990), and various halophilic archaea, including *Haloarcula marismortui*, and

Haloferax volcanii (Pickl et al. 2012). Bacterial levan is produced extracellularly by a single enzyme, levansucrase (EC. 2. 4. 1. 10), which produces levan directly from sucroses (Han 1990). For the degradation of levan, bacteria produce specific enzymes called levanases (EC 3.2.1.65), which hydrolyze the β-2, 6-linked main chain of levan to oligofructans and/ or fructoses. Like inulinases, levanases are also divided into endo- and exo-levanases. Endo-levanases hydrolyze within the chain of levan, and the product is usually various sizes of oligofructans. Exo-levanases hydrolyze at the end of the levan chain, yielding only a single size product, usually levanbiose (Vijn and Smeekens, 1999). The degraded products are used by microorganisms as nutrients since they possess necessary enzymes required for the degradation and oxidation of these sugars.

Several soil-borne bacteria such as *Pseudomonas putida*, *Streptococcus* and *Bacillus subtilis* produce levan-degrading enzymes (Murakami et al. 1990, Burne et al. 1987, Wanker et al. 1995). Levanase activities have also been found in yeasts *Kluyveromyces fagilis* (Grootwassink and Hewitt, 1983) and filamentous fungi (Vandamme and Derycke 1983; Barthomeuf et al. 1991). Several levanases have been purified and characterized from different bacteria such as *Bacillus* sp., *Rhodotorula* sp., *Streptomyces* sp., and *Pseudomonas* sp. (Wanker et al 1995, Miasnikov 2006, Kang et al., 1999, Chaudhary et al., 1996, Lim et al. 1998). Genes encoding levanases have been cloned from *Bacillus* sp. L7 (Miasnikov 1997), *Bacillus subtilis* (Martin et al. 1987), endophytic bacterium *Gluconacetobacter diazotrophicus* SRT4 (Menéndez et al. 2012), and *Streptococcus* sp. (Burne et al. 1987; Burne and Penders, 1992). *Bacillus polymyxa* CF43, which is a growth promoting rhizobacteria in the wheat rhizosphere was detected as expressing sucrose hydrolase activity (Bezzate et al. 1994).

Numerous genes found in microorganisms, particularly genes encoding metabolic enzymes, have also been found in the genomes of the eukaryotic parasites, likely through horizontal gene transfer (Whitaker et al. 2009; Jain et al. 2003). For example, a bacterial gene encoding a mannase was transferred to the genome of *Hypothenemus hampei*, a devastating pest of coffee (Acuña et al. 2012). A bacterial gene encoding a cyanase was transferred to the genomes of the spider mites *Tetranychus urticae* (Grbic et al. 2011), *Tetranychus evansi*, and *Panonicus citri* (Wybouw et al. 2012), and to the genomes of parasitic nematodes (Haegeman et al. 2011). Most of the documented horizontal transfers involve endosymbionts, where the association between the bacteria and host cells are closely related and genetic exchanges can occur in the germ line relatively easily (Dunning Hotopp et al. 2007; Nikoh, N and A. Nakabachi. 2009).

Genes encoding metabolic enzymes are more likely to be involved in horizontal gene transfer from bacteria to eukaryotes than other genes because of their metabolic processes are more similar than other processes (Tokumasa et al. 2001; Lake et al. 1999). Metabolic enzymes are often conserved among different kingdoms (Whitaker et al. 2009). For instance, phylogenetic analyses of 2,257 metabolic enzymes revealed a preference for enzymes encoded by genes gained through horizontal and endosymbiotic transfer (Whitaker et al. 2009). There are lines of evidence to suggest that high levels of endosymbiotic gene transfer have allowed further acquisition of bacterial genes in host genomes. Several cases of endosymbiotic partnership have been described between a eukaryotic host with prokaryotic or eukaryotic endosymbionts (Hoffmeister and Martin, 2003; Wernegreen 2004).

The Hessian fly, *Mayetiola destructor*, is a gall midge (Diptera: Cecidomyiidae) and a destructive pest of wheat (Hatchett et al. 1987; Butin et al. 1999; Pauly 2002). The insect has the

ability to manipulate plant metabolism, and a single larva can induce the formation of nutritive cells at the feeding site (Harris et al. 2006), inhibit growth of wheat plants (Byers and Gallun 1971), and altering the physiological pathways of infested plants (Liu et al. 2007). Insects inject proteins into host plants that act as effectors to suppress plant defenses and/or to reprogram host plant physiological pathways (Tjallingii 2006; Mutti et al. 2008), and it is hypothesized that gall midges use effector-based strategies similar to those used by plant pathogenic bacteria, fungi, oomycetes and nematodes (Grant et al. 2006; De Wit et al. 2009; Kamoun 2006; Pate et al. 2010; Stuart et al.2012). The identification of different bacterial genera throughout the *M. destructor* life cycle and their ovarial transmission suggests their intimate relationship with *M. destructor*. In addition, the elimination of bacteria from *M. destructor* with antibiotics causes high mortality of larvae infesting wheat seedlings, indicating that symbiotic bacteria are essential for survival (Bansal et al. 2011). Thus, it is possible that bacteria associated with *M. destructor* play a role in plant manipulation and facilitate larval feeding. It is also possible that genes encoding proteins and enzymes in these bacteria are transferred to the *M. destructor* genome through horizontal gene transfer and that these genes may have been an essential evolutionary step in *M. destructor* adaptation to a parasitic life style.

Objectives

A systematic search of the *M. destructor* genome (http://www.k-state.edu/hessianfly/genomics.html) revealed several types of genes that were likely derived from bacteria. Among these genes are a group of genes encoding proteins that share significant sequence similarity with bacterial levanases and inulases. The objective of this research was to investigate the expression profiles of these putative levanase/inulase genes in different

developmental stages of the insect and in different tissues of *M. destructor* larvae, the only feeding stage.

Materials and Methods

M. destructor

Hessian flies used in this study were derived from a field collection from Scott County, Kansas in 2005 (Chen et al. 2009). Since then, insects have been maintained on seedlings of the susceptible wheat cultivar 'Karl 92' in the greenhouse. The majority of the flies were the GP (Great Plains) biotype, which survives on and stunts (expresses virulence to) plants of wheat cultivars carrying the *H3, H5, H6 M. destructor* resistance genes and the *H7H8* gene combination (Ratcliffe et al. 1994; 1997; 2000).

M. destructor life stages, larvae tissue and RNA isolation

Whole body samples from, 1-, 3-, 5- and 10 d old larvae, and pupae were collected by dissecting the crowns of infested wheat plants. Female and male adults were collected after emergence. Approximately 200 samples of salivary gland, gut, malpighian tubule, and fat body tissues were extracted from 3 d old (first instar) larvae. Salivary glands were dissected in phosphate buffered saline (PBS) by grasping the posterior end of a larva with a pair of forceps while with another pair of forceps pulling away the anterior tip of larvae with the salivary glands and mouth parts. Then salivary glands were removed from mouthpart tissues. Gut, malpighian tubule and fat body tissues were removed from larvae and collected in PBS. There were three biological and three technical replications per treatment.

Total RNA was extracted from whole *M. destructor* life stages and specific larval tissues using TRI REAGENT following the protocol provided by the manufacture (Molecular Research Center, Inc. Cincinnati, OH). RNA samples were treated first with DNase I Amplification Grade

(Invitrogen, Carlsbad, CA, USA) to eliminate potential DNA contamination and analyzed immediately, using the Agilent 2100 Bioanalyzer. RNA was reverse-transcribed into cDNA using (oligo-dT)20 primers with the SuperSript® III First-Strand cDNA Synthesis System (Invitrogen, Carlsbad, CA, USA).

Sequence Analysis and Data Searching

The identity of putative cDNAs was farther confirmed by Blastx search in GeneBank (National Center of Biotechnology Information, Bethesda, MD). Signal peptide cleavage sites were predicted using the SignalP 4.1 Server (Petersen et al. 2011). Multiple alignments of gene data matrices and protein sequences were generated using similarity calculated with ClustalW (Larkin et al. 2007).

PCR Analyses

Primers were designed using Beacon Designer 7.0 Software (Biosoft, Palo Alto, CA), and IDT Integrated DNA Technology, http://www.idtdna.com/Primerquest/Home/Index.

Semi-quantitative PCR amplification was carried out for 30 to 45 cycles, depending on the intensity of the PCR product, as follow: 240 s at 94 °C; 60 s at 94 °C; 60 s at 55 °C; 120 s at 72 °C and 60 s at 72 °. DNA fragments from PCR reactions were separated on 1.5 – 3% agarose gel, depending on the size of the product. Gels were stained with (0.5 mg/ml) ethidium bromide. Actin was used as reference gene for normalization. DNA bands were photographed with a BIODoc-ItTM system (UVP, Uplan, CA).

Melting curve analyses of quantitative real time (qPCR) primers were conducted on an iCycler real time detection system (Bio-Rad, Herculese, CA). Each reaction was performed with 6 µl of cDNA mix, 6.5 µl of primer mix, and 12.5 µl of iQ SYBR Green super mix in a 25-µl total volume. Each reaction was done in duplicate wells per primer using PCR tube strips with

optical flat caps (Bio-Rab Laboratories). Melting curve analyses were performed to ensure the absence of primer-dimer formation and amplicon specificity. The qPCR amplification was carried out under the following conditions: one cycle at 95 °C (3 min), followed by 39 cycles of denaturation at 95°C (10 s), annealing and extension at 62°C for 30 s. Melt curve analyses were done by heating the PCR mixtures from 65 to 95 °C (0.5°C per cycle of 5 s) with measurements of the SYBR Green signal intensities at the same time. There were 79 cycles until reaching 95°C.

Relative expression analyses were performed using elongation factor-1α (EF1a) as an internal reference gene after validation tests of EF1α, actin, ribosomal protein S4, ribosomal protein S30, and ubiquitin. To calculate the relative expression values of each gene, Ct, threshold cycle were obtained from each target genes per biological replicate and technical replicate. Averages from the biological and technical replicates per treatment were calculated, transformed to $\log_2$, and used as input data for statistical analyses. The log-transformed arbitrary expression values were analyzed by ANOVA using MINITAB 16 (2013). Tukey's pairwise comparison and 95% confidence intervals were used to separate data into groups with significant differences.

Results

Levanase- and inulase- like genes in the M. destructor genome

Blast searching *M. destructor* genome sequences with the putative levanase protein identified 14 levanase- and inulase-like genes or gene fragments which were named *Mayetiola destructor levanase gene one* (*MDL-1*) to *MDL-14*. Further sequence analysis revealed that six of the genes encode full-length proteins, three were truncated at the 5'-end, and five were truncated at the 3'end (Table 3. 1). Blast searches of the Genbank database with these putative proteins revealed that the first hit targets were from six different bacteria, namely, *Flexithrix dorotheae* , *Bacillus mojavensis*, *B. subtilis*, *Emticicia oligotrophica* , *Segetibacter koreensis*,

and *Clostridium acetobutylicum*. The E-values for these hits ranged from e-12 to e-110. Of the 14 levanase/ inulase genes, 10 were identified with a putative levanase function, and four with a putative exo-inulase/levanase function (Table 3. 1).

Gene structure of levanase- and inulase-like genes

The structural organization of the 14 putative levanase/ inulase-ike genes (Figure 3.1) includes a promoter region, three exons, two introns, and a 3' untranslated region (UTR). *MDL-3, MDL-5, MDL-9, MDL-11, MDL-12,* and *MDL-13* are full length. Genes *MDL-1, MDL-8* and *MDL-14* are truncated at the 5' end, with *MDL-1* and *MDL-14* truncated at exon 2 and *MDL-8* truncated at exon 3. *MDL-2, MDL-4, MDL-6, MDL-7,* and *MDL-13* are truncated at the 3' end, with *MDL-2, MDL-4* and *MDL-7* truncated at exon 3, *MDL-6* truncated at exon 2, and *MDL-13* truncated at intron 1. Although the numbers and overall arrangement of introns/exons are conserved among different genes, the lengths of each exon and intron varied among different genes.

Sequence conservation and variation of the M. destructor levanase/inulase genes and proteins.

Comparative analysis of sequences among different levanase/inulase-like genes revealed sequence conservation and diversification in different regions (Fig. 3. 2A). Conserved regions are exons with average scores of 74 to 77%. Diversified regions include introns with average scores of 52 to 57%, promoter regions with an average score 49%, and 3' UTRs with an average score of 47%.

Comparative analysis of the putative full-length proteins revealed conservation and diversification among members as well. Overall, the N-terminal region, which constitutes a putative secretion signal peptide, is highly diversified (Fig. 3. 2B). The C-terminal region with

around 150 amino acids is more diversified than the N-terminal mature protein region. Even in the N-terminal mature protein region, there are patches of diversified residues among different proteins. The intron/exon boundaries are conserved in term of protein encoding.

Reverse transcription PCR analysis

Expression levels of 11 of the 14 levanase/ inulase-like genes varied among all developmental stages (Figure 3. 3A) (*MDL-6, MDL-7,* and *MDL-8* did not amplify by RT-PCR). *MDL-1* and *MDL-4* exhibited similar expression patterns, and were expressed in all developmental stages with higher levels in the larval stage. *MDL-5, MDL-11* and *MDL-13* exhibited significantly greater expression levels in larval tissue, but very little expression in pupae and adults. On the other hand, *MDL-3, MDL-12* and *MDL-14* were preferentially expressed only in larvae and undetectable in other developmental stages. The alternative splicing bands in *MDL-2, MDL-9* and *MDL-10* were differentially expressed, with the bands in *MDL-2* and *MDL-10* expressed in pupae and adults, and the bands in *MD-l9* expressed in 10 d old larvae, pupas and adults. The actin reference control was expressed at all developmental stages.

The expression pattern of the levanase/inulase genes was also examined with samples from different tissues including salivary glands, malpighian tubules and fat bodies of 3 d old *M. destructor* larvae (Figure 3. 3B). The expression level of the 11 levanase/inulase-like genes was abundant in tissues except *MDL-9* and *MDL-14*. *MDL-3* exhibited splicing bands in all larvae tissues. *MDL-9* was expressed a low level in salivary glands, gut, and malpighian tubules, and exhibited splicing bands in gut, malpighian tubules and fat bodies samples. MDL-14 was weakly expressed in all tissues, especially in the gut and malpighian tubules.

Real-time PCR analysis

MDL-2, MDL-3, MDL-4, MDL-5, MDL-9, MDL-10, MDAL-11 and *MDL-14*, were successfully amplified in qPCR, whereas *MDL-1, MDL-6, MDL-7, MDL-8, MDL-12* and *MDL-13* did not. The expression patterns were different from those revealed by RT-PCR for some genes. *MDL-2* exhibited higher expression in pupae compared with other developmental stages, whereas similar expression levels were detected in other stages. *MDL-3* and *MDL-4* exhibited less expression in larvae compared with other genes. *MDL-3* was highly expressed in pupae and adults, whereas *MDL-4* was highly expressed in pupae, but less in adults compared to *MDL-3*. *MDL-5, MDL-9, MDL-10,* and *MDL-14* were almost equally expressed in all developmental stages. Different expression was observed in *MDL-11* compared with other genes. This gene was highly expressed in 1 d, 3 d, pupa, and adult (female and male) *M. destructor* stages, and less expressed in 5 d and 10 d old *M. destructor* larvae, and in pupa and adult (female and male) stages, but expressed at lower levels in 3 d and 5 d old larvae.

The tissue expression patterns of the eight levanase/inulase-like genes were given Figure 3. 3B. Overall, more genes exhibited higher expression levels in malpighian tubules and fat body tissues, and lower expression levels in the gut. Considering the big error bars, *MDL-2 MDL-5, MDL-9, MDL-10,* and *MDL-14* were roughly equally distributed among different tissues. *MDL-3* and *MDL-4* were expressed at low level in the gut, whereas *MDL-11* was expressed at lower level in Malpighian tubules.

Discussion

Origin of levanase/inulase-like genes in the M. destructor genome

Blast search of the Genebank database with the respective putative proteins showed approximately 42% similarity to bacterial sequences with E values smaller than e^{-12} (Table 3. 1).

Sequence analysis also showed an average 36% similarity to fungal levanase or inulase proteins (data not shown). No apparent sequence similarity was found to any know vertebrate proteins, indicating that levanase/inulase genes in the *M. destructor* genome originated from bacteria, likely through horizontal gene transfer (Acuña et al. 2012).

Diverse bacteria are associated with different developmental stages of *M. destructor* and many different types of bacteria are maternally transferred from generation to generation (Bansal et al. 2011). Levanase genes were likely transferred from a symbiotic bacterium to a *M. destructor* embryo, providing a selective advantage to *M. destructor.* Genes encoding metabolic enzymes are known to have been transferred from bacteria to eukaryotes through horizontal- or lateral gene transfer (Whitaker et al. 2009), and it has been suggested that invertebrate bacterial endosymbionts transfer genes more frequently because of the close and constant proximity of cells in both organisms (Dunning et al. 2007). However, the similarity of the putative levanases or inulases to fungal proteins does not exclude that these *M. destructor* genes originated from fungi. Fungal genes, for example, encoding a carotenoid cyclase and a carotenoid desaturase were found to have been transferred from fungi to the genome of the two-spotted spider mite, *Tetranychus urticae* (Altincicek et al. 2012), suggesting that arthropods acquire genes from fungal symbionts.

There are at least two possible origins of the *M. destructor* levanase/inulase-like multi-gene family. One is from a single bacterial gene transfer, which then duplicated and diversified into different genes. Alternatively, multiple transfer events may have occurred, each resulting in a different *M. destructor* gene. Given that overall gene structure is conserved, namely all full-length gene having three exons and two introns and the conservation of intron/exon boundaries, it is highly likely that these multi-gene families were derived from a single gene transfer. This

postulation is consistent with the fact that bacterial genes usually have no introns. Therefore, the two introns must have been derived after the bacterial gene was transferred to the *M. destructor* genome. If multiple transfer events occurred, the numbers of introns and the intron locations likely differ for each gene. The expansion of a single bacterium-derived levanase gene into a mulit-member gene family suggests that the presence of these horizontally transferred genes play some critical roles in the *M. destructor* genome.

Sequence Conservation and Diversification of Levanase/Inulase Genes

Sequence comparisons revealed conservation and diversification of levanase/inulase –like genes in different gene regions. Overall, exons were relatively highly conserved, with an average 74-77% identity (Figure 3. 2), but introns are more diversified, with an average of only 52-57% identity. The higher rates of conservation in exons than in introns suggest that the genes are functional and under selection constraint. Assuming that multiple levanase/inulase genes were derived from a single ancestor, the highly diversified introns suggest that horizontal transfer has occurred for a long time, and that these levanase/inulase-like genes have played important roles in shaping *M. destructor* evolution. Therefore, the ancestral bacterial gene gained introns after transfer to the host genome without interrupting its original open reading frame given a potential advantage to levanase/inulase gene activities (Wybouw etal. 2012).

The promoter region and the 3'-UTRs are even more diverse, with sequence similarity of only 47 to 49% (Figure S2). In the promoter region, no TATA box occurred within the first 300 bp upstream sequence in one of the six full-length genes (Figure S2). A TATA box does exist in the other five genes within 300 bp upstream sequences, but their location varied greatly from gene to gene (Figure S2). The high diversification in the 5' and 3' regulatory regions of the levanase/inulase genes suggests that different genes have different regulatory mechanisms for

expression. This is consistent with different expression patterns of these genes in different developmental stages and different tissues (Figure 3. 3).

Protein Diversification and Alternative Splicing

Despite the relative high rates of conservation in exons in comparison to introns, proteins encoded by different genes were diversified, especially in the C-terminal region (Figure 3. 2B). The diversification in protein sequence indicates that it is likely that different proteins perform different functions, or have different specificities. In addition to sequence diversification, alternative splicing forms apparently exist for some genes. For example, alternative splicing forms were detected in *MDL-2, MDL-9* and *MDL-10* in 10-day old *M. destructor* larvae, pupae and adult females and males (Figure 3.3A). Likewise, alternative splicing forms were observed for *MDL-3* and *MDL-9* in different tissues of 3-day old larvae (Fig. 3. 3A and B). Sequence diversification and alternative splicing forms in different levanase or inulase genes gained different functions or specificity in *M. destructor*. Further research is needed to reveal the exact functions of proteins with bacterial origin in *M. destructor* biology.

Signal peptides

Among the 14 levanase/inulase putative proteins, four were truncated at the N-terminal, making it impossible to determine whether they have a secretion signal peptide. For the remaining 10 putative proteins, nine had a typical secretion signal peptide, and the remaining one lacked a signal peptide. Proteins with a signal peptide usually are produced in one location, and secreted (transported) to another location for function (Zanen et al. 2005). Since the majority of levanase/inulase proteins have a signal peptide, they are likely secreted out of the cell for function. Proteins produced in the salivary glands may be secreted into wheat tissue for pre-oral

digestion of fructans (levans or inulins) that function as reserve carbohydrate in wheat and other monocotyledons (Vijn and Smeekens 1999). Proteins produced in gut cells may be secreted into the gut and function as digestive enzymes.

Expression patterns in different M. destructor developmental stages and tissues

The expression pattern of the levanase/inulase-like genes differed among developmental stages and larval tissues of *M. destructor*. The difference could be related to the specific function of each larval tissue, and metabolic processes in different developmental stages. Expression in larval gut and salivary gland tissues may be to aid digestion enzymes in pre-oral or gut digestion. However, expression of these genes in pupae and adults as alternative splicing forms may be linked to an abundance of symbiotic bacteria (Bansal, 2010). Levanases or inulases may play a regulatory role in modulating or regulating the bacterial population by removing bacterial levan or other fructons, or providing components for synthesizing levan-like molecules via reversed enzymatic functions.

The identification of a gene with similarity to levanase bacterial gene in the *M. destructor* genome and the determination of the expression level of these genes in different developmental stages and different larval tissue of the insect provide scientific support for further characterization of the levanase/inulase-like genes and the specific role of these enzymes in plant insect interaction.

References

Acuña, R., Padilla, B., Flórez-Ramosa, C., Rubioa, D., Herrera, J., Benavides, P., Lee, S., Yeats, T., Egan, A., Doyle, J and J. Rose. 2012. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. Proceedings of the National Academy of Sciences. 109:4197-4202.

Altincicek, B., Kovacs, J and N. Gerardo. 2012. Horizontally transferred fungal carotenoid genes in the two-spotted spider mite *Tetranychus urticae* . Biol. Lett. 8:253-25 doi:10.1098/rsbl.2011.0704.

Aminov, R. 2011. Horizontal gene exchange in environmental microbiota. Frontiers in Microbiology. Review article. 2:1-19. doi: 10.3389/fmicb.2011.00158

Andersen, C., Jensen, J and J. Orntoft TF. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Research. 64:5245-5250.

Bezzate, S., Steinmetz, M and S. Aymerich. 1994. Cloning, Sequencing, and Disruption of a Levanase Gene of *Bacillus polymyxa* CF43. Journal of Bacteriology. 176:2177-2183

Barthomeuf, C., Regerat, F and H. Pourrat. 1991. Production of inulinase by a new mold of *Penicillium rugulosum*. Journal of Fermentation and Bioengineering. 17:491-494.

Burne, R., Schilling, K., Bowen, W and R, Yasbin. 1987. Expression, purification, and characterization of an Exo-β-D-fructosidase of *Streptococcus* mutants. Journal of Bacteriology. 169:4507-4517.

Burne, R and E. Penders. 1992. Characterization of the *Streptococcus* mutans GS-5 *fruA* Gene Encoding Exo-,1-D-Fructosidase. American Society for Microbiology. 60:4621-4632

Bansal, R., Hulbert, S., Schemerhorn, B., Reese, J., Whitworth, R., Stuart, J and M. Chen. 2011. Hessian Fly-Associated Bacteria: Transmission, Essentiality, and Composition. PLoS ONE. 6:23170. doi:10.1371/journal.pone.0023170

Buntin, G., 1999. Hessian fly (Diptera: Cecidomyiidae) injury and loss of winter wheat grain yield and quality. J. Econ. Entomol. 92:1190-1197.

Byers, R and R. Gallun. 1971. Ability of the Hessian Fly to stunt winter wheat. 1. Effect of larval feeding on elongation of leaves. J. Econ Entomol. 65:955–958.

Brevnova, E., Kozlov, D., Efremov, B and S. Benevolensky. 1998. Inulase-Secreting Strain of Saccharomyces cerevisiae Produces Fructose. Biotechnology and Bioengineering**.** 60:492-497.

Cote, G and J. Ahlgren. 1993. Metabolism in microorganisms, Part I. Levan and levansucrase. *In* M Suzuki, NJ Chatterton, eds, Science and Technology of Fructans. CRC Press, Boca Raton, FL, pp 141–168.

Chen, I., Christie, P and D, Dubnau. 2005. The ins and outs of DNA transfer in bacteria. Science. 310:1456-1460. DOI:10.1126/science.1114021.

Chen, M. S., Echegaray, E., Whitworth, J., Wang, H., Sloderbeck, P., Knutson, A., Giles, K and T. Royer. 2009. Virulence analysis of Hessian fly populations from Texas, Oklahoma, and Kansas. J. Econ. Entomol. 102:774-780.

Dawes, E and D. Ribbons. 1966. Sucrose utilization by *Zymomonas mobilis*: formation of a levan. Biochemical Journal. 98:804-812.

Dahech, I., Ben Ayed, H., Srih Belghith, K., Belghith, H and H. Mejdoub. 2013. Microbial production of levanase for specific hydrolysis of levan. International Journal of Biological Macromolecules. 60:128-133.

Dawes, E and D. Ribbons. 1966. Sucrose utilization by *Zymomonas mobilis*: formation of a levan. Biochem. J. 98:804–812.

Dunning, J., Clark, M., Oliveira, D., Foster, J., Fischer, P., Muñoz Torres, M., Giebel, J., Kumar, N., Ishmael, N., Wang, S., Ingram, J., Nene, R., Shepard, J., Tomkins, J., Richards, S., Spiro, D., Ghedin, E., Slatko, B., Tettelin, H and J. Werren. 2007. Multicellular Eukaryotes Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes. Science. 317:1753-1756.

Dunning, J. 2011. Horizontal gene transfer between bacteria and animals. Trends in Genetics. 27:157-163. doi:10.1016/j.tig.2011.01.005.

De Wit, P., Mehrab, I., Van den Burg, H and S, Stergiopoulo. 2009. Fungal effector proteins: past, present and future. Mol Plant Pathol. 10:735-747 DOI: 10.1111/J.1364-3703.2009.00591.X

Fondi, M and R, Fani. 2010. The horizontal flow of the plasmid resistome: clues from inter-generic similarity networks. Environ Microbiol. 12:3228-3242.

Gupta, A and N. Kaur. 2000. Carbohydrtae reserves in plants – Synthesis and regulation. Elsvier Science B. V. p.223

Grootwassink, W and M. Hewitt. 1983. Inducible and constitutive formation of P-Fructofuranosidase (inulase) in batch and continuous cultures of the yeast *Kluyveromyces fragilis*. Jounal of General Microbiology. 129:31-41.

Grant, S., Fisher, E., Chang, J., Mole, B and J. Dangl. 2006. Subterfuge and manipulation: Type III effector proteins of phytopathogenic bacteria. Ann Rev Microbiol. 60:425-449. doi: 10.1146/annurev.micro.60.080805.142251

Giubellini, A., Ribbera, P., Merusi, P., Turroni, F., Foroni, E., Pizzetti, P., Cagnasso, V., Bizzarri, B., Luigi de'Angelis, G., Fergus, S., Douwe van, S and M. Ventura. 2009. Exploring the Diversity of the Bifdobacterial Population in the Human Intestinal Tract. Appl. Environ. Microbiol. 75:1534-1545.

Han, Y. W. 1990. Microbial levan. Adv. Appl. Microbiol. 35:171–194.

Hoffmeister, M. and W. Martin. 2003. Interspecific evolution: microbial symbiosis, endosymbiosis and gene transfer. Environ. Microbiol. 5:641–649.

Haegeman, A., J. Jones and E. Danchin. 2011. Horizontal gene transfer in nematodes: a catalyst for plant parasitism? Mol. Plant-Microbe Interact. 24:879-887. doi:10.1094/MPMI -03-11-0055.

Hatchett, J., Starks, K., Webster, J., 1987. Insect and mite pests of wheat. In: Wheat and Wheat improvement. Agronomy Monograph No. 13:625–675.

Harris, M., Freeman, T., Rohfritsch, O., Anderson, K and S. Payne. 2006. Virulent Hessian Fly (Diptera: Cecidomyiidae) Larvae Induce a Nutritive Tissue During Compatible Interactions with Wheat. Annals of the Entomological society of America. 99:305-316.

Jain,R., Rivera, M., Moore, M and J. Lake. 2003. Horizontal Gene Transfer Accelerates Genome Innovation and Evolution. Mol. Biol. Evol. 20:1598–1602.

Keeling, P and J. Palmer. 2008. Horizontal gene transfer in eukaryotic evolution. Nature Review Genetics. 9:605-618. doi:10.1038/nrg2386

Kamoun, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. Ann Rev Phytopathol. 44:41-60. doi: 10.1146/ annurev.phyto.44.070505.143436

Komberg, H. 2001. Routes for fructose utilization by *Escherichia coli*. J. Mol. Microbiol Biotechnol. 3:355-359.

Lake, J., Jain, R and M. Rivera. 1999. Mix and Match in the Tree of Life. Science. 283: 20227-2028. DOI: 10.1126/science.283.5410.2027

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T. and D. Higgins. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23:2947-2948.

Liu, X., Bai, J., Huang, L., Zhu, L., Weng, N., Reese, J., Harris, M., Stuart, J and M.-S. Chen. 2007. Gene expression of different wheat genotypes during attack by virulent and avirulent HF (*Mayetiola destructor*) larvae. J Chem Ecol. 33:2171–2194.

Martel, C., Parker, J., Jackson, C., Warrilow, A., Rolley, N., Greig, C., Morris, S., Donnison, I., Kelly, D and Kelly, S. 2011. Expression of bacterial levanase in yeast enables simultaneous saccharification and fermentation of grass juice to bioethanol. Bioresource Technology. 102:1503–1508.

Martin-Verstraete, I., Débarbouillé, M., Klier, A and G. Rapoport. 1990. Levanase operon of *Bacillus subtilis* includes a fructose-specific phosphotransferase system regulating the expression of the operon. Journal of Molecular Biology. 214:657-671.

Murakami, H., Muroi, H., Kuramoto, T., Tamura,Y., Mizutani, K., Nakano, H and S. Kitahata. 1990. Purification and some properties of a levanase from *Streptomyces* sp. No. 7-3. Agric. Biol. Chem. 54:2247–2255.

Martin, I., Debarbouillé, M., Ferrari, E., Klier, A., and Rapoport, G. 1987. Characterization of the levanase gene of *Bacillus subtilis* which shows homology to yeast invertase. Molecular Genetics and Genomics. 208:177-184.

Miasnikov, A. 1997. Characterization of a novel endo-levanase and its gene from *Bacillus* sp. *L7*. FEMS Microbiology Letters. 154:23–28.

Menéndez, C., Hernández, L., Selman, G., Mendoza, M., Hevia, P., Sotolongo, M., Arrieta, J. 2002. Molecular cloning and expression in *Escherichia coli* of an exolevanase gene from the endophytic bacterium *Gluconacetobacter diazotrophicus SRT4*. Current Microbiology. 45:5-12.

Mutti, N., Pappan, L., Begum, K., Pappan, K., Chen, M., Park, Y., Reese, J and G, Reeck. 2008. A protein from the salivary glands of the pea aphid, *Acyrthosiphon pisum*, is essential in feeding on a host plant. Proceedings of the National Academy of Sciences. 105:9965-9969. www.pnas.org_cgi_doi_10.1073_pnas.0708958105

Nikoh, N and A. Nakabachi. 2009. Aphids acquired symbiotic genes via lateral gene transfer. BMC Bio. 7:12.

Pauly, P. 2002. Fighting the Hessian fly. Environ. Hist. 7:385-507.

Petersen, T., Brunak, S., von Heijne, G and H. Nielsen, H. 2011. SignalP 4.0: iscriminating signal peptides from transmembrane regions. Nat. Methods. 8:785-786. http://www.cbs.dtu.dk/services/SignalP/.

Pickl, A., Johnsen, U and P. Schönheit. 2012. Fructose Degradation in the Haloarchaeon *Haloferax volcanii* Involves a Bacterial Type Phosphoenolpyruvate-Dependent Phosphotransferase System, Fructose-1-Phosphate Kinase, and Class II Fructose-1,6-bisphosphate aldolase. J. Bacteriol.194:3088-97. doi: 10.1128/JB.00200-12.

Pinney, J., Shirley, M., McConkey, G and D, Westhead. 2005. metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. Nucleic Acids Res. 33:1399-1409. doi:10.1093/nar/gki285

Popa, O and T. Dagan. Trends and barriers to lateral gene transfer in prokaryotes. Current Opinion in Microbiology. 14:1–9. DOI 10.1016/j.mib.2011.07.027.

Pate, N., Hamamouch, N., Li, C., Hewezi, T., Hussey, R., Baum, T., Mitcchum, M and M, Davis. 2010. A nematode effector protein similar to annexins in host plants. J Exp Bot. 61:235-248. doi:10.1093/jxb/erp293

Pfaffl, M., Tichopad, A., Prgomet, C and T, Neuvians. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnology letters. 26:509-515.

Ratcliffe, R., Safranski, G., Patterson, F., Ohm, H and P. Taylor. 1994. Biotype status of Hessian fly (Diptera: Cecidomyiidae) populations from the eastern United States and their response to 14 Hessian fly resistance genes. J. Econ. Entomol. 87:1113-1121.

Ratcliffe, R and J. Hatchett. 1997. Biology and genetics of the Hessian fly and resistance in wheat, pp 47-56. *In* K. Bondari [ed.], New Developments in Entomology. Research Signpost, Trivandrum, India.

Ratcliffe, R., Cambron, S., Flanders, F., Bosque-Perez, N., Clement, S and H. Ohm. 2000. Biotype composition of Hessian fly (Diptera: Cecidomyiidae) populations from the southeastern, Midwestern, and northwestern United States and virulence to resistance genes in wheat. J. Econ. Entomol. 93:319-1328.

Silver, N., Best, S., Jiang., J and S. Thein. 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC molecular biology. 7:33.

Scotta, K., Martina, J., Chassarda, C., Clergeta, M., Potrykusa, J., Campbella, G., Mayerb, G., Younga, P., Rucklidgea, G., Ramsaya, A and H. Flint. 2011. Substrate-driven gene expression in *Roseburia inulinivorans*: Importance of inducible enzymes in the utilization of inulin and starch. PNAS. 108:4672-4679

Thomas, C and K, Nielsen. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat. Rev. Microbiol. 3:711-721. doi:10.1038/nrmicro1234 http://dx.doi.org/10.1016/j.ibmb.2012.08.002

Tokumasa, T., Hamada, K., Kanaya, S and T. Shinozawa. 2001. Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. Natural Cell Biology. 3:210-214.

Tjallingii, W. 2006. Salivary secretions by aphids interacting with proteins of phloem wound responses. J Exp Bot. 57:739-745. doi:10.1093/jxb/erj088

Vandamme, E and Derycke, D. 1983. Microbial Inulinases: Fermentation process, properties, and Aplications. Advances in Applied Microbiology. 29:139-176.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A and F. Speleman. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology. 3:research0034.1–0034.11

Vijn, I and S. Smeekens 1999. Fructan: More than a reserve carbohydrate?. Plant Physiol. 120:351–359.

Whitaker, J., McConkey, G and D. Westhead. 2009. The transferome of metabolic genes explored: analysis of the horizontal transfer of enzyme encoding genes in unicellular eukaryotes. Genome Biology. 10:R36. doi:10.1186/gb-2009-10-4-r36

Wernegreen, J. 2004. Endosymbiosis: lessons in conflict resolution. PLoS Biol. 2:0307-0311. DOI: 10.1371/journal.pbio.0020068

Wybouw, N., Balabanidou, V., Ballhorn, D., Dermauw, W., Grbi, M., Vontas, J and T. Van Leeuwen. 2012. A horizontally transferred cyanase gene in the spider mite *Tetranychus urticae* is involved in cyanate metabolism and is differentially expressed upon host plant change. Insect Biochemistry and Molecular Biology. 42:881-889.

Weigel, l., Clewell, D., Gill, S., Clark, N., McDougal, L., Flannagan, S., Kolonay, J., Shetty, I., Killgore, G and F. Tenover. 2003. Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. Science. 302:1569-1571.

Wanker, E., Huber, A and H. Schwab. 1995. Purification and Characterization of the *Bacillus subtilis* Levanase produced in *Escherichia coli*. Applied and Environmental Microbiology. 61:1953-1958.

Zanen, G., Houben, E., Neima, R., Tjalsma, H., Jongbloed, J., Westers, H., Oudega, B., Luirink, J., van Dijl1, J and W. Quax. 2005. Signal peptide hydrophobicity is critical for early stages in protein export by *Bacillus subtilis*. FEBS Journal 272:4617-4630.

Figures and Tables

Table 3.1 Levanase/ inulase like-genes in the *M. destructor* genome*.

Gene name	Accession	Organism	Putative function	E-value	Gene length
MDL-1	WP_020531995	*Flexithrix dorotheae*	Exo-inulase/levanase	$2e^{-22}$	5'-truncate
MDL-2	WP_010335070	*Bacillus mojavensis*	Levanase	4e-^{90}	3'-truncate
MDL-3	WP_020531995	*F. dorotheae*	Levanase	$2e^{-106}$	Full length
MDL-4	WP_010335070	*B. mojavensis*	Levanase	$1e^{-87}$	3'-truncate
MDL-5	YP_003866957	*B.subtilis*	Levanase	$1e^{-100}$	Full length
MDL-6	YP_006875301	*Emticicia oligotrophica*	Levanase	$2e^{-36}$	3'-truncate
MDL-7	WP_020531995	*F. dorotheae*	Exo-inulase/levanase	$3e^{-94}$	3'-truncate
MDL-8	WP_020531995	*F. dorotheae*	Levanase	$3e^{-42}$	5'-truncate
MDL-9	WP_020531995	*F. dorotheae*	Exo-inulase/levanase	$4e^{-97}$	Full length
MDL-10	WP_018617377	*Segetibacter koreensis*	Levanase	$4e^{-110}$	Full length
MDL-11	WP_020531995	*F.dorotheae*	Levanase	$1e^{-99}$	Full length
MDL-12	WP_020531995	*F. dorotheae*	Levanase	$4e^{-106}$	Full length
MDL-13	NP_348090	*Clostridium acetobutylicum*	Levanase	$8e^{-15}$	3'-truncate
MDL-14	WP_020531995	*F. dorotheae*	Exo-inulase/levanase	$6e^{-12}$	5'-truncate

*Putative protein sequences were used to blast Genbank. The accession number, organism names, putative functions, and E-values were given according to the first hit. Six genes encode full length protein while others were truncate at either 5’ or 3’ ends.

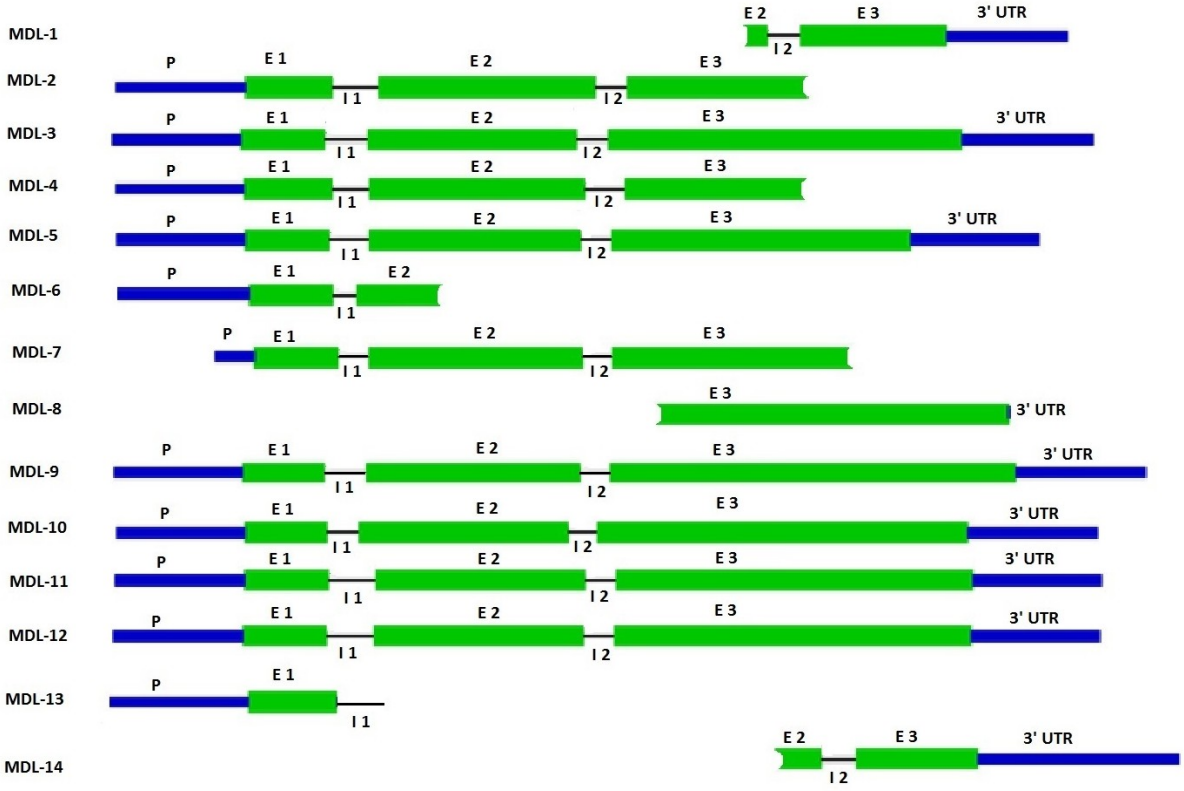
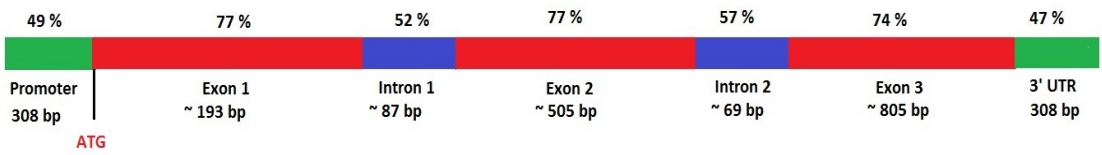


Figure 3.1 Structural organization of 14 putative levanase/inulase genes in the *M. destructor* genome.
Blue rectangles indicate 5' or 3' regulatory regions. Green rectangles indicate exons, lines between exons indicate introns. P - promoter region, E1 - exon 1, I1 - intron 1, E2 - exon 2, I2 - intron 2, E3 - exon 3, 3'UTR - 3'-untraslated region.

A



B

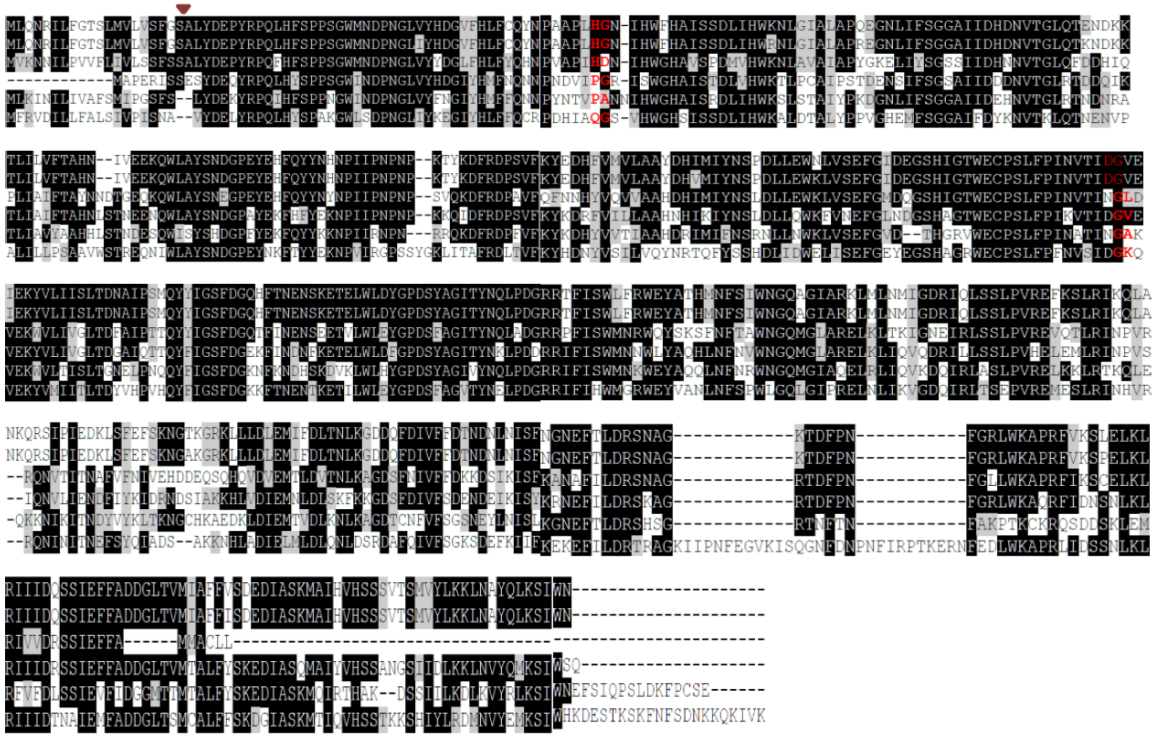
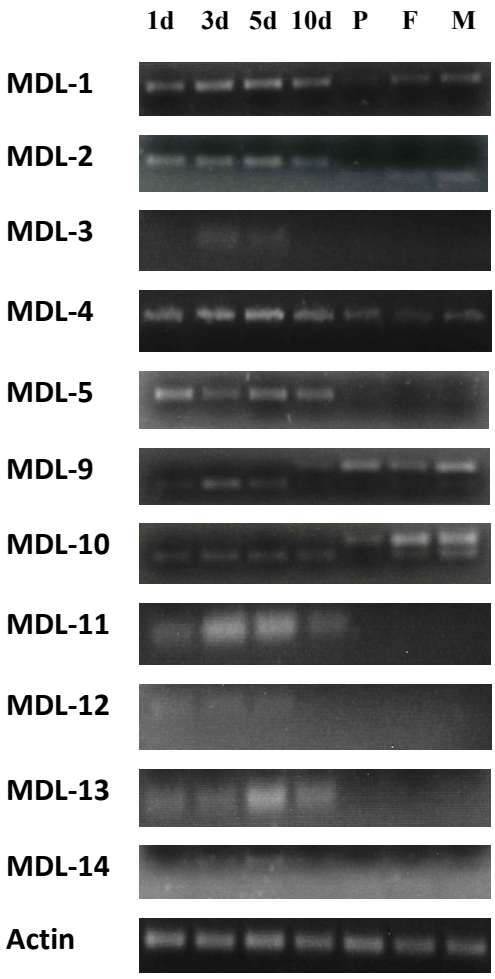


Figure 3.2 Conservation and diversification of putative levanase/inulase genes and proteins in the *M. destructor*.

A. Comparison of promoter, exon, intron and 3' untranslated regions in different genes. Average scores (above rectangles) from pairwise comparisons of respective regions of all 14 genes in all possible combinations using ClusterW. Green - unconserved regions, red - conserved regions, blue - diverse regions. Average lengths are below rectangles. **B**.
Sequence alignments of putative proteins. The red triangle indicates the putative signal peptide cleavage site (19 to 22) in the amino acid sequences. Red residues in the sequence alignment are sites of exon/intron splicing. The alignment was produced using ClusterW and printed with Boxshade.

A


1d
3d
5d
10d
P
F
M
MDL-1
MDL-2
MDL-3
MDL-4
MDL-5
MDL-9
MDL-10
MDL-11
MDL-12
MDL-13
MDL-14
Actin

B

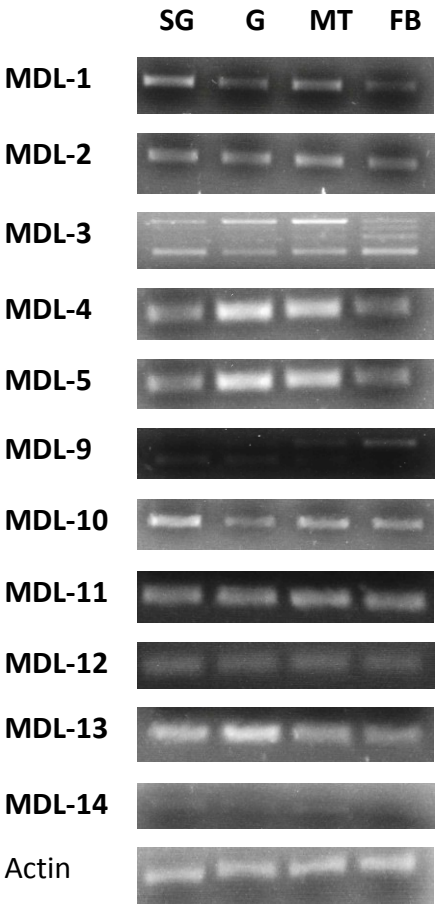
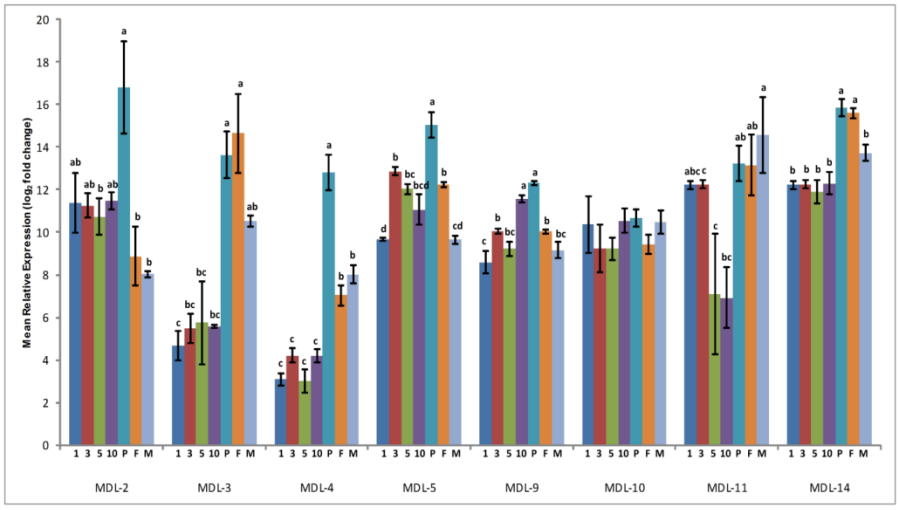


Figure 3.3 Reverse-transcription PCR analysis of levanase/inulase genes expressed in *M. destructor*.
A. 1d, 3d, 5d, 10d - Tissues from whole body of 1-, 3-, 5-, and 10-d old larvae. P - whole body of pupa, F and M - female and male adults, respectively. **B.** Tissues from salivary glands (SG), gut (G), Malpighian tubules (MT) and fat bodies (FB) of 3 d old larvae fed Hessian fly-susceptible "Karl 92" wheat. Relative quantification of Levanase/ inulase genes by RT-PCR

A



B

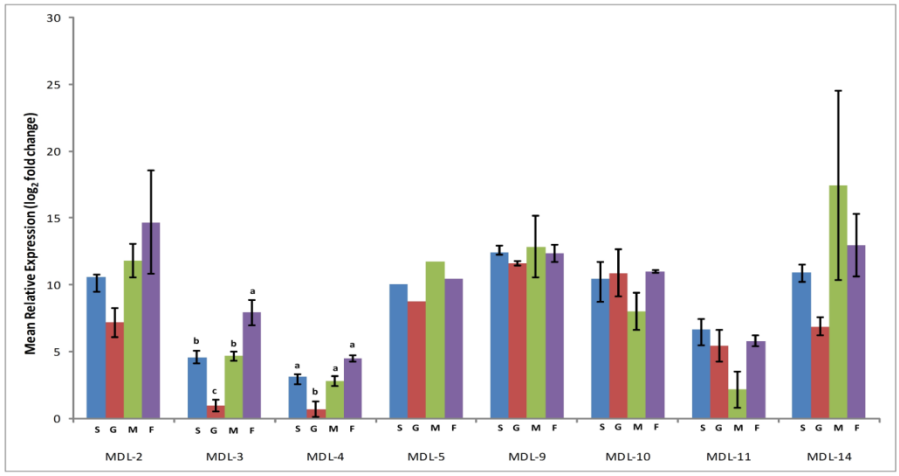


Figure 3.4 Mean ± S. E. expression of 8 levanase/inulase genes in tissues of different *M. destructor* developmental stages.
A. 1-, 3-, 5-, and 10 day old larvae, pupae, and adult females and males. **B.** Salivary glands (SG), gut (G), malpighian tubules (MT) and fat bodies from 3d old larvae. n = 3 biological replicates, n = technical replicates per gene.

Supplementary data Figure S1.

>MDL-1

ACTTTTTACCGCTCATAATAGGGCTACATATGAGCAGGAACAATGGTGAGCATACA**GT**AA
F L P L I I G L H M S R N N G E H T
TGATAGCCCTGAATTCAAACACTCCAACTACAATCTATACGAAATTTTTACGTCTGTAAT

TTATAGAGTGAA**AG**AAATTTACAAATGAAAATTCTCCGGAAACCACAATATGGGTGGACT
K F T N E N S P E T T I W V D
ATGGCCCGGATAGATATGCTGGAATCACATACAATCAATTACCCGATGGTCGTCGTGTTT
Y G P D R Y A G I T Y N Q L P D G R R V
TTGTCAGTTGGATGAATAAATGGGAATACGCAGAGCAATTGAATTTCAATGTTTGGAACG
F V S W M N K W E Y A E Q L N F N V W N
GACAAGTGGGACTGATGCGAGAACTTAAATTGAAACAAGTTGGAGATCAAATTCGATTCG
G Q V G L M R E L K L K Q V G D Q I R F
TTTCCTTCCCTGTTCGTGAAGTAGAAAAATTGAGAACTTATCTGGTTCGCAAAGAGAATA
V S F P V R E V E K L R T Y L V R K E N

TTAAGGATCAATTTATTTTTATTAAGTACAAAATTAATTTTTTGAGGACCATTTATCCGA
I K D Q F I F I K Y K I N F L R T I Y P
TT**TAA**AGACCAGTTTATTTTTTTACGGATTTTTTTTATTTGAAAAAAGGACCCAAAAATC
I
TGATTCAAAGACCATTTATACGAAAAAAGGACCAATTTCACGCGACCTTTTCTGAATCTT
TCGAATGTTGCAAACTTCAAGGCATAAGCCCCACCCCACACTGAGTTAGAGCTGGTACAA
ATAAATAAATAAATAAATAAATATATATAAAACGATTTTATAACTTAATGTTTCGCATCA
AGCAAAACAGAAAAAAGAAAGAGAAGCATTTTCTATGTAGGAATTAAAATTTACAATGAA
ATGCCTATAGAGATAAAGAACGAAACGAATACAGATAAATTTAAAAACTATGTGCTAATT
ATGCTATAAACCAAATTGATACATTCAGGTTCAGAGAAACAAATTAAATTAATAATTATT
AATGTTTTTAAAATAAGGAAAAAGAAGCCTAAGGCGTGCTAAATAAATAAATAAAAAAAT
GAAAATTAAACAATAAAACTTCAAAAGTTGTTGTCTTAAGCGTTGTATTGCTGTCTTATT
CCCCAATGTTTGCATTTTTATGATTGATAATATGTAATAGTAATCGACGAAAATCAGATT
CGGTCAACAAATAAAAATTCATATCAATCGACACTTTGAGCATAACATTTTACAAATAGA
ATTTGAATAACATTATTAAATTAAGTCCAAAGGTTTGGTGTTCTGAATGAGATCCCAACA
GGTTTTATTTTTTTCAGCAGTATTTAGCTGATAACACTTATAGCATTTAATGAGATATTT
TCGCGTTTGTAAACATGCGCAAAAAGAAGTTTCAGTAGTGTTCATCATAAAAAATAACTA
TTAACTGTTAGAAGGCCATTTCATTTCAGTGATAAGAATCAAATGCTCAAATCGAATGGT
AATATAGAAAAAAAAT**TATAA**AAGGCTGGATGGTTTCGAATTGGGAAATAGTAATCAACA
AATATTGGATAGAGTAGAAGGGTCCTGTGTAGAAAGATCCTGACTATTCTCAAT

>MDL-2
AATGAGATCCCAACAGGTTTTATTTTTTTCAGCAGTATTTAGCTGATAACACTTATAGCA
TTTAATGAGATATTTTCGCGTTTGTAAACATGCGCAAAAAGAAGTTTCAGTAGTGTTCAT
CATAAAAAATAACTATTAACTGTTAGAAGGCCATTTCATTTCAGTGATAAGAATCAAATG
CTCAAATCGAATGGTAATATAGAAAAAAAAT**TATAA**AAGGCTGGATGGTTTCGAATTGGG
AAATAGTAATCAACAAATATTGGATAGAGTAGAAGGGTCCTGTGTAGAAAGATCCTGACT

ATTCTCAAT**ATG**TCTCGAGCAAATATTTTACTCTGTACAACTTTGTTAGTTTTAATTTCA
M S R A N I L L C T T L L V L I S
TTAACAAGTGCGGTTTACGATGAACTTTATCGGCCACAAATACATCTTAGTCCACCATCA
L T S A V Y D E L Y R P Q I H L S P P S
GGATGGATGAGCGATCCGAATGGTTTGGTTTATCATGATGGAGTATATCACGTTTTCTAT
G W M S D P N G L V Y H D G V Y H V F Y
CAACACAACCCAAATGACACCAAACCCAACGACTTCC**GT**AAGTGAAATTCACAAAAAATG
Q H N P N D T K P N D F
GATCAAACGATGTTTGATGTTTGCCCAGAAAAAAAAAACAAAACCAACACTCAAATTAAT

AATTAAATTTTTTTTTTTCATAAATTT**AG**CACAAATGTATTGGGGTCATGCAATTTCAAC
P Q M Y W G H A I S T
TGATTTGGTTCATTGGAAAAACCTCGATATTGCGCTTTCTCCACCCACTGACAAATCATT
D L V H W K N L D I A L S P P T D K S F
TTTCTCTGGCAGCGGAATCATTGATCAACATAATGTAACTGGCTTCCAAATTGATGACGA
F S G S G I I D Q H N V T G F Q I D D D
TAAAAAACCAATAATTTTGTTATTCACCTCCAATAATATGTCCACCGCAGAACAAGAACA
K K P I I L L F T S N N M S T A E Q E Q
ATGGATAGCATACAGCAACGACGCCCCTGAGTACAAGACATTTGAATATTATAAAAATAA
W I A Y S N D A P E Y K T F E Y Y K N N
TCCAGTTATCAAAAAGTTGAATTTGGAAGGGGAAAAGTCAATCAATTTTAGAGATCCACA
P V I K K L N L E G E K S I N F R D P Q
GATTATTGAATGGGAAACTGGTCAATATGTTACATTAATAACCCAAGACAATAAATCAAT
I I E W E T G Q Y V T L I T Q D N K S M
GATTTATAATTCACGTGATATGATAAATTGGGAATTGGTTAGCCAATTTGGAGAATATGA
I Y N S R D M I N W E L V S Q F G E Y E
AGGAAAACATGGCGGAACATGGGAATGTCCTTCACTCTTTACCCTTAACGTTACCATCAA
G K H G G T W E C P S L F T L N V T I N
TGG**GT**AAATACAAATAAAAGAAAAAATCAATTAAAAGAATCAGTTAAATGAATTTCTTTT
G
CTGTTGGATATTT**AG**AACGATGGTCGAGAAGCAGGCTCTGTTGATCACTTTAACTGACGC
T M V E K Q A L L I T L T D A
TGTAATACCGGCCATTCAATATTTCATTGGATCTTTTGATGGTAAAACTTTCAAAAATGA
V I P A I Q Y F I G S F D G K T F K N E
AAATCCACCAGAAACAATTTTGTGGTACGATTATGGACCGGACAGCTTTGCTGGTAGTAC
N P P E T I L W Y D Y G P D S F A G S T
ATTTAATCATGTGCCGAGTAATCGACGGATTTTCCTCAGCTGGATGTCACGCTGGGAATA
F N H V P S N R R I F L S W M S R W E Y
TGCGAAGAGAGTAGGTCCTGGTCCCTGGAGTGGAAATATGGGTCTTCCTAGAGAATTGAA
A K R V G P G P W S G N M G L P R E L N
TTTGAAACAAGTTGGAAATGATATTAGATTTGCATCTTTACCCGTTTCTGAATTGAAAGT
L K Q V G N D I R F A S L P V S E L K V
CTTGAGGATGAGTCAGTTTCGTGAGAAAAACGTAACGATCAGTTCAAAGTACGTTTCGAA
L R M S Q F R E K N V T I S S K Y V S K
AATGGTTGAAGATGACGGTGAAGACA
M V E D D G E D

>MDL-3
TGATGAAACTATTAAGAACTGATGTTCTTTAGGAAAATTCGATAGATAAAAAGATTAGGA
ACATTTGATATCGTCATTTTTAAATCGAGAGTTGAAGGGTTAATTATGTTTTGAATCAAT
TCAGCGACAAAAATAACAATTCACAATTAATTGAA**TATAA**ACGCTCAACGAACGTATAAA
AGGATGGTAAATTTTGTATTTGATGGTAGTCATCAGTTGGTAATCCCCCAACAGAGTTCT
AAAAAATAGACTGAAGACGAAGAGTGAATTCAACTTGCTAAACTTTGAATTGAAACAAAG
GAACAAGA**ATG**TTACAAAATAGAATTTTGTTCGGTACATCGTTGATGGTTTTAGTTTCAT
M L Q N R I L F G T S L M V L V S
TTGGAAGTGCATTGTATGATGAACCATATCGACCACAATTACATTTCAGTCCACCATCAG

```
F  G  S  A  L  Y  D  E  P  Y  R  P  Q  L  H  F  S  P  P  S
GATGGATGAACGACCCAAATGGTCTAGTCTATCATGATGGTGTTTTTCATCTATTCTGCC
G  W  M  N  D  P  N  G  L  V  Y  H  D  G  V  F  H  L  F  C
AATATAATCCAGCTGCACCTCTTCATGGTAAATAGAAGTGTCCAACTAAGAACCAATTTC
Q  Y  N  P  A  A  P  L  H
TGATCCAATTAACCGATTTACTTTAAATAAAAAAAATAAAATAAAGATTTCACATCAAAT

ATTTTAGGTAACATTCATTGGTTCCATGCAATTTCTTCGGATTTGATTCATTGGAAAAAT
      G  N  I  H  W  F  H  A  I  S  S  D  L  I  H  W  K  N
CTTGGCATTGCACTTGCTCCACAGGAAGGAAATTTAATTTTTTCAGGCGGCGCAATCATT
 L  G  I  A  L  A  P  Q  E  G  N  L  I  F  S  G  G  A  I  I
GATCATGACAACGTAACTGGTCTACAAACAGAAAACGATAAAAAAACATTAATTTTGGTG
 D  H  D  N  V  T  G  L  Q  T  E  N  D  K  K  T  L  I  L  V
TTCACCGCTCATAATATCGTTGAGGAAAAACAATGGTTGGCATACAGCAACGATGGACCC
 F  T  A  H  N  I  V  E  E  K  Q  W  L  A  Y  S  N  D  G  P
GAATACGAACACTTTCAATACTATAACCATAATCCAATCATACCAAATCCAAATCCGAAA
 E  Y  E  H  F  Q  Y  Y  N  H  N  P  I  I  P  N  P  N  P  K
ACGTATAAAGATTTCAGAGATCCATCAGTTTTCAAGTATGAAGACCATTTTGTGATGGTA
 T  Y  K  D  F  R  D  P  S  V  F  K  Y  E  D  H  F  V  M  V
TTAGCCGCATACGATCATATCATGATTTATAATTCACCGGATTTATTGGAATGGAATCTG
 L  A  A  Y  D  H  I  M  I  Y  N  S  P  D  L  L  E  W  N  L
GTGAGTGAATTTGGAATCGACGAAGGATCCCATATCGGGACATGGGAATGCCCGTCCCTC
 V  S  E  F  G  I  D  E  G  S  H  I  G  T  W  E  C  P  S  L
TTTCCTATTAACGTAACAATTGATGGGTATGTTTTCGCGAATATTTTTTTTATTTAATCA
 F  P  I  N  V  T  I  D  G
TCAATTTGTGAGTGAAAAATTTATTTTTGACAATTTTCAGCGTCGAAATTGAAAAATACG
                                        V  E  I  E  K  Y
TTTTGATCATCAGCCTAACGGACAATGCAATACCAAGCATGCAGTACTATATAGGATCTT
V  L  I  I  S  L  T  D  N  A  I  P  S  M  Q  Y  Y  I  G  S
TTGATGGACAACATTTCACGAATGAAAACTCCAAAGAGACAGAACTATGGCTCGATTATG
F  D  G  Q  H  F  T  N  E  N  S  K  E  T  E  L  W  L  D  Y
GTCCGGATAGCTATGCTGGCATTACATATAATCAATTGCCGGACGGTCGACGTACTTTCA
G  P  D  S  Y  A  G  I  T  Y  N  Q  L  P  D  G  R  R  T  F
TAAGTTGGTTATTTCGATGGGAATACGCTACACATATGAATTTCAGTATTTGGAATGGAC
I  S  W  L  F  R  W  E  Y  A  T  H  M  N  F  S  I  W  N  G
AAGCCGGAATTGCTAGAAAATTGATGTTGAACATGATTGGTGATCGCATTCAACTTTCCT
Q  A  G  I  A  R  K  L  M  L  N  M  I  G  D  R  I  Q  L  S
CTTTACCTGTTCGTGAATTTAAATCTTTGCGAATCAAACAACTCGCCAATAAACAAAGAA
S  L  P  V  R  E  F  K  S  L  R  I  K  Q  L  A  N  K  Q  R
GTATACCGATTGAGGACAAACTTAGTTTTGAATTTTCTAAAAATGGCACTAAAGGAAGGA
S  I  P  I  E  D  K  L  S  F  E  F  S  K  N  G  T  K  G  R
AACTTTTGCTAGACTTAGAAATGATTTTCGATTTAACGAATCTAAAAGGAGACGACCAAT
K  L  L  L  D  L  E  M  I  F  D  L  T  N  L  K  G  D  D  Q
TTGATATTGTATTCTTTGACACTAATGACAATTTGAACATTTCATTCAATGGAAATGAAT
F  T  L  F  D  I  V  F  F  D  T  N  D  N  L  N  I  S  F  N
TCACTTTAGATCGATCGAATGCAGGAAAAACCGATTTCCCGAATTTCGGGCGACTTTGGA
G  N  E  D  R  S  N  A  G  K  T  D  F  P  N  F  G  R  L  W
AAGCTCCACGTTTTGTTAAAAGTCTAGAATTAAAATTAAGAATAATTATTGATCAATCTT
K  A  P  R  F  V  K  S  L  E  L  K  L  R  I  I  I  D  Q  S
CAATTGAATTTTTTGCCGATGATGGACTTACAGTTATGATAGCGTTTTTCGTTTCCGATG
S  I  E  F  F  A  D  D  G  L  T  V  M  I  A  F  F  V  S  D
AAGATATTGCTTCAAAAATGGCCATTCACGTTCATTCATCTTCAGTGACGTCCATGGTTT
E  D  I  A  S  K  M  A  I  H  V  H  S  S  S  V  T  S  M  V
ATTTAAAAAAATTAAATGCTTATCAATTGAAAAGTATTTGGAATTAATGCATTTTGATAA
Y  L  K  K  L  N  A  Y  Q  L  K  S  I  W  N
TAATAACATGTACAGCATTACAGTAAAATCTCTCTAAAAAAAGTCATCTCCCATTTTTGA
ACCAATTAGCCGGCAACCAAAATTCTCTTTATATTACATAAACATTATACACATACTTAT
```

ATTTTAATAATTTTTTCCACATCCTTTGTTCATCGTATTCATTTATTAATAAGATCAGCT
GTAAAACACCAAATCGTACGCAACAATAGAATTCCAAATTTGAACCTTATTTCATAATAC
TCTTATGAGCTTTATGTGTATGGCGAATTTTTTTCCATTCTCTTTCTACTAGTTA

>MDL-4

TTATAGCATTTATTGGAATATTTTCGCGTTCGTAATCATGCGCAAATAGAAATTTCAATA
GTGTTCATCATAAATTAACTATTAACTGTTAGAAGGTCATTCCACTTCAGTGATAAGAAT
CAAATGCCTAAATTCAACACTGATATGGCAAAAATA**TATAA**AAGGCTGGATGGTTTCGTA
TTGGAAAACAGTAATCAATAAATATTGGATCGAGTAGAAGGAGCCTGTGTGGTACAAAAC
TAAAAAAAAGATTTAACATTGGTCAATTTTTAGGAAATCATTACTAAATAGGAAGACTAT
TTTTCAAT**ATG**TCTCGAACAAATATTTTACTCTGTACAACTTTGTTAGTTTTAATTTCAT
M S R T N I L L C T T L L V L I S
TAACAAGTGCAGTGTACGATGAACTTTATCGTCCACAAATACATCTTAGTCCTCCAAATG
L T S A V Y D E L Y R P Q I H L S P P N
GATGGATGAGCGATCCAAATGGTTTGGTCTATCATGATGGAGTATATCACGTCTTCTATC
G W M S D P N G L V Y H D G V Y H V F Y
AACACAATCCAGTTGATACCGAACCCGACAAATTCC**GT**AAGTTGTTTGCTCAAAAAAAAA
Q H N P V D T E P D K F
AGAAAAAAACAAAACCAACACTCAAATTAATAATTAAATTGGTTTTTTTCTTCATACATT

T**AG**CACAAATGTATTGGGGCCATGCAATTTCAACCGATTTGGTTCATTGGAAAAACCTCG
P Q M Y W G H A I S T D L V H W K N L
GTATTGCGCTTTCTCCACCCACTGGCAGATCATTTTTCTCTGGCGGCGGAATCATTGATC
G I A L S P P T G R S F F S G G G I I D
ATCATAATGTAACTGGATTCCAAATCGATGACGATAAAAAACCATTAATTTTGTTATTCA
H H N V T G F Q I D D D K K P L I L L F
CATCCTATAATATATCCACCGAAGAACAAGAACAATGGATAGCATACAGCAATGATGCGC
T S Y N I S T E E Q E Q W I A Y S N D A
CCGAGTACAATACATTTGAATATTATAAAAATAATCCAGTTATCAATCAGTTGAATTTGG
P E Y N T F E Y Y K N N P V I N Q L N L
AAGGGGAAAAGTCAATCAATTTTAGAGATCCACAAATTTATGAATGGGAAACTGGTCAAT
E G E K S I N F R D P Q I Y E W E T G Q
ATGTTACATTTATATCCCAAGACAATAAATCAATGATTTATAATTCACGTGATATGATAA
Y V T F I S Q D N K S M I Y N S R D M I
ATTGGGAATTGGTTAGCCAATTTGGAGAATATGAAGGGAAACATGGTGGAATATGGGAAT
N W E L V S Q F G E Y E G K H G G I W E
GTCCTTCACTCTTCACCCTTAACGTCACCATCAATGG**GT**AAATACAATTAAAAGAAAGCT
C P S L F T L N V T I N G
TTCTTCATTTTCTTTACTTTCAATTTACATGAATCAATTAAATGAATATTGCCATTTCTG

TTGGATATTT**AG**AACGATGGTCGAGAAGGAGGCTCTGTTGATCACTTTAACTGACTTCGT
T M V E K E A L L I T L T D F V
GATACCGGCCATTCAATATTTCATTGGATCTTTTGATGGTAAAACTTTCAAAAATGAAAA
I P A I Q Y F I G S F D G K T F K N E N
TCCACCAGAAACAATTTTGTGGTACGATTATGGACCGGACAGCTTTGCTGGTAGTACATT
P P E T I L W Y D Y G P D S F A G S T F
TAATCATGTGCCGAGTAATCGACGAATTTTCCTCAGCTGGATGTCACGCTGGGAATATGC
N H V P S N R R I F L S W M S R W E Y A
GATGAGTGTAGGTCTTGGTCCCTGGAATGGAAATCTGGGTCTACCCAGAGAATTGCATTT
M S V G L G P W N G N L G L P R E L H L
GAAACAAGTTGGAAATGATATTCGTCTTGCATCTTTACCCGTTTATGAATTGAAAGACTT
K Q V G N D I R L A S L P V Y E L K D L
GAGGACGAGTCAGTTTCGTAAGAAAAATGTGAGGATCAGTTCAAAGTATGTTTCGAAAAT
R T S Q F R K K N V R I S S K Y V S K I
TGTTGAAGATGACGGTGAAGACC
V E D D G E D

>MDL-5

AAATCCTTATCGATTTCTTTTATCTACAGACAAAAAGTTGAACTTACAATTTTTCAAATC
AAGAAGCTTTATTCGCAGTTAAATAAAAACGTTTATTGATACCTAATATGTATAGTGCTT
GAACGCAATTTTTGGAATGTAAAACATCCAGTGTAATGTTGATTATGAACGAAGAATTGA
ATCAG**TATAA**AAGGCTGGATATTCTACGAAGAAAAACATTGTACGAATATTTGTTGAAAG
AAACATAGCAATTTTAAAGAGAAATCCAAAAAGTAAAACTATAATTTAATTCAAAAATTC
ATTTGAAA**ATG**GTGAAAAACAATATCCTGCCTGTTGTATTTCTGATAGTTTTAAGTTCAT
M V K N N I L P V V F L I V L S S
TTTCAAGTGCACTGTACGATGAACCATATCGACCACAATTTCATTTTAGTCCACCTTCAG
F S S A L Y D E P Y R P Q F H F S P P S
GATGGATGAACGATCCAAATGGTCTAGTTTATTATGATGGTTTATTTCATTTATTCTATC
G W M N D P N G L V Y Y D G L F H L F Y
AACATAATCCAGTTGCACCCATTCATG**GT**AAGTTTTAACGACCCGAAAACTAAATGAAAA
Q H N P V A P I H
TCAGCGACGAAAAATAATGCTAAATAAAATAAATTACAATATTCATTCGCAATGTAATT**A**

GATAATATCCATTGGGGTCATGCAGTATCACCGGATATGGTACATTGGAAAAATCTTGCC
D N I H W G H A V S P D M V H W K N L A
GTTGCAATTGCTCCATATGGGAAAGAATTAATCTATTCGGGTAGTTCGATCATTGATCAT
V A I A P Y G K E L I Y S G S S I I D H
AACAATGTAACTGGTCTTCAATTTGATGATCATATTCAGCCATTGATTGCCATATTCACC
N N V T G L Q F D D H I Q P L I A I F T
GCTTACAATAACGACACCGGAGAACAAAAACAATGGTTGGCCTATAGCAATGAGGGTCCT
A Y N N D T G E Q K Q W L A Y S N E G P
GAATACGAACACTTCCAATACTATAACTATAATCCAATCATTCCGAATCCAAATCCAAGT
E Y E H F Q Y Y N Y N P I I P N P N P S
GTACAAAAGGATTTCAGAGATCCGGCCGTTTTCCAATTCAACAATCATTATGTTCAGGTT
V Q K D F R D P A V F Q F N N H Y V Q V
GTAGCTGCTCACGATCATATCATGATTTATAATTCACTGGATTTATTGGAATGGAAACTT
V A A H D H I M I Y N S L D L L E W K L
GTGAGTGAATTTGGAATGGATCAAGGATCACATATTGGAACATGGGAATGTCCGTCACTA
V S E F G M D Q G S H I G T W E C P S L
TTTCCCATCAATGTCACAATTAATGG**GT**AAAACGACAAAATTTAATTTTTTCCGAATTCA
F P I N V T I N G
TTGTGATTATAAATTTTGTTTTGTTCTATTTTTC**AG**ATTGGATGTTGAAAAATGGGTTTT
L D V E K W V L
GATTGTCGGTTTGACCGATTTTGCGATACCAACCACACAATACTATATTGGCTCCTTTGA
I V G L T D F A I P T T Q Y Y I G S F D
TGGACAAACCTTTATCAATGAGAATTCCGAAGAAACAGTTCTATGGCTGGAATATGGACC
G Q T F I N E N S E E T V L W L E Y G P
AGATAGCTTCGCTGGTATTACTTACAATCAATTAGCGGATGGTCGTCGCCCTTTCATCAG
D S F A G I T Y N Q L A D G R R P F I S
CTGGATGAATCGATGGCAATATAGTAAGAGTTTCAATTTTACCGCTTGGAATGGACAAAT
W M N R W Q Y S K S F N F T A W N G Q M
GGGCCTGGCAAGAGAATTAAAATTGACCAAAATTGGAAATGAAATTCGACTTTCGTCATT
G L A R E L K L T K I G N E I R L S S L
ACCTGTTCGTGAAGTACAAACCTTGAGAATTAATCCGGTTCGGAGACAAAATGTTACGAT
P V R E V Q T L R I N P V R R Q N V T I
CACCAATGCATTCGTATTCAATATTGTCGAACATGATGATGAACAATCCCAACATCAAGT
T N A F V F N I V E H D D E Q S Q H Q V
GGATGTTGAAATGACACTCGATGTGACGAATCTTAAAGCAGGCGACTCATTTAATATCGT
D V E M T L D V T N L K A G D S F N I V
ATTTTTTGACAAAAAAGATTCGATAAAAATTTCCTTCAAAGCAAACGCGTTTATTTTGGA
F F D K K D S I K I S F K A N A F I L D
CCGATCGAATGCAGGAAGAACTGATTTTCCCAATTTCGGGCTACTTTGGAAGGCGCCAAG
R S N A G R T D F P N F G L L W K A P R

GTTCATAAAAAGTTGTGAATTAAAATTGAGAATTGTTGTCGATAGATCATCAATTGAATT
  F I K S C E L K L R I V V D R S S I E F
CTTCGCGATGATGGCCTGTCTGTTA**TGA**CTGCACTTTTCTTTTCAATGAAGACATCGCAT
  F A M M A C L L
CGAAGATGGCCATTAATGTTTATTCATCCTCAACCAATTCGAGTGTTCATTTAAATCAGC
TGAATGCTTACCAAATGAAAAGTATTTGGAATTAATTTTTAAGTAAAATATTTGAATGTT
CTCAATTATATTTAAGAAACAATAAATCTATGTTATTCCAATGAAAAAAAAAAAGTTCTA
CACGTTTCAATAATATCTTATCTAAGTTATTATTGACATTTTTCCCATGCGTTCATTGAA
ATGTTACTGGTTACACGCTAATAAATCAAAACAGAC

>MDL-6

ATCATTCTTCGAAAACAATCAAAAGAGTTCGGAAGAAAATACTTAATTAAAAGCATAAAT
TTAAGAAATTCTTTTCAAAATGGTGAAAAACGATATACTACTTATTGCATCTCTGATTGC
TTTAAGTTCTTTTTCATATGATATTTCCGATGATTTTCGTTTGCTGATTTTCAGTTCCTG
ATTATTCAATTTTGAACTTACTTCAGCAAATTGAAATTCTTATCATGTTTCACAGAAGAA
TTCTGTTTATTTGAAATTAATGTGTTCGTGACGAATGGAACATGAAACTAGAACTATGTT
AATTGTCATTTGTCAATTGATAGTAGCCGAGTGTTTGAAGGTAATTGATTACCCAAATGC
AATGATAAAAATCGAATATACAATGAATTGTATTTAGCGATAAAATTATTATA**TATAA**AT
GGCTGGATAATTTGCATTAGAACCCTAGTTTTCAATCATTTTTCAACGGATCCACATGAA
AATCCGTCAAAACCAGAAAAACACAAATTTGAAGAGAATTATATTTCGTGGATTGAAGT
AAGAAGTAGAAAAGTGCTTTAGA**ATG**TTACGAATAAATATCCAGCTCATTATATCAATAG
                        M L R I N I Q L I I S I
TTGTTTTGGATTCATATTCGAGTGAATTATATGACGAACTATATCGGCCTCAATTAACTT
V V L D S Y S S E L Y D E L Y R P Q L T
TTAGTCCACCATCAGGATGGATAAATGATCCGAATGGTCTAGTCTATTATGATGGTGTTT
F S P P S G W I N D P N G L V Y Y D G V
TTCACCTTTTTTGTCAATATAATCCAAATAGCACACTTCATG**GT**ATGTAAAATGGTTCGA
F H L F C Q Y N P N S T L H
TTAACTGGAAATAATGATAATTATAATTTTTTTGACCT**AG**GTAATCTTCATTGGTATCAT
                                          G N L H W Y H
GCGATCTCTCCGGATTTGGTTCATTGGGAAAATCTTGGTATTGCACTTGCTCCAACGAAT
 A I S P D L V H W E N L G I A L A P T N
GGAAATTTAATTTTTTCGGGCAGCGCAATCATTGATCATGGAAATGTAACTGGTCTTCAA
 G N L I F S G S A I I D H G N V T G L Q
GCTAATGACGATAAGAAAACATTAATAGCCATTTTTACGGCTCATGATTTGTCAACCAAT
 A N D D K K T L I A I F T A H D L S T N

>MDL-7
TTTCTGAAAAATAGAAACTTAATAGTAATCATTTGACAGTACATAGTAGATAGTGTATTA
TTCTTGTTGAGAAAATTGTGAGATTATCCTCAAAAC**ATG**GTGCACATTAATATTTTACTT
                                        M V H I N I L L
ATTGCATTAATAATGGTTTTAAATTCATTTTCAAGTGGATTATATGACGAACCATATCGG
 I A L I M V L N S F S S G L Y D E P Y R
CCACAATTACATTTTAGTCCTCCAGTAGGATGGATGAATGACCCAAATGGTCTAGTCTAT
 P Q L H F S P P V G W M N D P N G L V Y
CATGATGGTATTTTTCATTTATTTTATCAACATGATCCAAAGACAACAATTCAAG**GT**AAC
 H D G I F H L F Y Q H D P K T T I Q
TTCAAACATATATGTGACATTGTTTGCGTCGATTCAAAATTAAAATCAATTTTTTCCATC

AATTT**AG**GTCAAATGCATTGGGGCCATGCAATTTCCACCGATATGATTTATTGGACGAAC
       G Q M H W G H A I S T D M I Y W T N
CTTGATATAGCGCTCACTCCACCTAAAGGAATTGCATATTTTTCTGGTGGTGCAATCATC
 L D I A L T P P K G I A Y F S G G A I I
GATTATAACAATGTAACTGGCTTCCAAATGTCTGCCAATGTAAAACCTTTGATTGCAATT
 D Y N N V T G F Q M S A N V K P L I A I

```
TTTACCGCTCATAATAGGGCTACATATGAGCAGGAACAATGGTTAGCCTACAGTAATGAT
 F  T  A  H  N  R  A  T  Y  E  Q  E  Q  W  L  A  Y  S  N  D
GGTCCTGAATACAAACACTTTGAACTTTATGAAAATAATCCAATAATTGGTAAATTGAAT
 G  P  E  Y  K  H  F  E  L  Y  E  N  N  P  I  I  G  K  L  N
ACGGATCCAAGTAAACCAATCGACTTCAGAGATCCAGCAATTACCAAATGGGGTGATCAT
 T  D  P  S  K  P  I  D  F  R  D  P  A  I  T  K  W  G  D  H
TTTGTTTTATTCTTGGCTCAATATAACAAATCCTTGATTTATAATTCACTGGATCTGAAG
 F  V  L  F  L  A  Q  Y  N  K  S  L  I  Y  N  S  L  D  L  K
AATTGGGAATTGGTCGGCGATTTTGGCGAATATGATGGATCACATTCAGAGGTTTGGGAA
 N  W  E  L  V  G  D  F  G  E  Y  D  G  S  H  S  E  V  W  E
TGTCCTTCACTCTTTCCACTTAATGTTACAATTAATGGGTAAAATGCAATGAATTTTCTA
 C  P  S  L  F  P  L  N  V  T  I  N  G
GCTCTAACTACAATTTATATGAAATTTTCCGATCTGTTTTTTACAGAGTGAATGTTGAAA
                                          V  N  V  E
AGCATGTTTTGATCGTTGGCTTAACTGGCAGCGTACTTCCGACAACACAATATTTTATTG
K  H  V  L  I  V  G  L  T  G  S  V  L  P  T  T  Q  Y  F  I
GATCTTTCGATGGAAAAAAATTTACAAATGAAAATTCTCCGGAAACCACATTATGGGTGG
G  S  F  D  G  K  K  F  T  N  E  N  S  P  E  T  T  L  W  V
ACTATGGCCCGGATAGCTATGCTGGAATCACATACAATCAATTACCCGATGGTCGTCGTG
D  Y  G  P  D  S  Y  A  G  I  T  Y  N  Q  L  P  D  G  R  R
TTTTTGTCAGTTGGATGAATAAATGGGAATACGCAGAACAATTGAATTTCAATGTATGGA
V  F  V  S  W  M  N  K  W  E  Y  A  E  Q  L  N  F  N  V  W
ATGGACAAATGGGACTGATGCGAGAACTTAAATTGAAACAAGTTGGAGATCAAACTCGAC
N  G  Q  M  G  L  M  R  E  L  K  L  K  Q  V  G  D  Q  T  R
TCGTTTCCTACCCTGTTCGTGAAGAAGAAAAATTACGAACATATGTGGTTCGTAAAGAGA
L  V  S  Y  P  V  R  E  E  E  K  L  R  T  Y  V  V  R  K  E
ATATTAAAATATCCACGAATAGATGTGTATATAAAATCACACCTGACGGAAAAACTAAAC
N  I  K  I  S  T  N  R  C  V  Y  K  I  T  P  D  G  K  T  K
ATACTGTCGATATGGAAATTACTCTCGACGTAACAGATCTCAAGAAAGGCGATTCGATCG
H  T  V  D  M  E  I  T  L  D  V  T  D  L  K  K  G  D  S  I
AATTTACGTTTTTCGATAAAAATGACAATTTAACTATTTCGCTCACGGAAAATGAATTCA
E  F  T  F  F  D  K  N  D  N  L  T  I  S  L  T  E  N  E  F
CTTTGGAGCGAAACAATACA
T  L  E  R  N  N  T
```

>MDL-8

```
AATTTGCAGAGTGCAAGTGGAAAAATATGTTTTGATAGTTGGCCTAACTGACAATTCAAT
  I  C  R  V  Q  V  E  K  Y  V  L  I  V  G  L  T  D  N  S  I
ACCGACGACACAATACTATGTCGGTTCTTTCGATGGTCAAACTTTTACGAATGAAAACTC
  P  T  T  Q  Y  Y  V  G  S  F  D  G  Q  T  F  T  N  E  N  S
AAAAGAAACAATACTATGGTTAGACTACGGTCCGGATAGCTATGCTGGCATTACATATAA
  K  E  T  I  L  W  L  D  Y  G  P  D  S  Y  A  G  I  T  Y  N
TCAATTACCGGATGGCCGACGAATATTCATCAGTTGGGAGAATCGTTGGCAATATGCTCA
  Q  L  P  D  G  R  R  I  F  I  S  W  E  N  R  W  Q  Y  A  Q
ACAATTGAATTTCAATGTTTGGAATGGACAAATGGGCTTAGCTCGAGAACTGACATTGAA
  Q  L  N  F  N  V  W  N  G  Q  M  G  L  A  R  E  L  T  L  N
TAAAATTGATAATCGAATTTTAATTTCGTCTTTACCTGTTCGGGAAACGAAAATGTTACG
  K  I  D  N  R  I  L  I  S  S  L  P  V  R  E  T  K  M  L  R
TATCGATCATGTTCGCAAGCAAAATATACCGATCGAAAATTACTTGTCTTTTGAAATTGC
  I  D  H  V  R  K  Q  N  I  P  I  E  N  Y  L  S  F  E  I  A
GAAATTTGACGACGAAAATAGAAAACAATCGGTAGATATAGAAATGATGCTTAACATAGC
  K  F  D  D  E  N  R  K  Q  S  V  D  I  E  M  M  L  N  I  A
GAATCTCAAAGCAGGCGATTCATTTAATATTGTATTTTTTGGAGTTAATGATGCATTGAA
  N  L  K  A  G  D  S  F  N  I  V  F  F  G  V  N  D  A  L  N
CATTTCATTCAATGGAAATGAATTCATTTTAGATCGATCAAAAGCAGGAAGAATCGATTT
  I  S  F  N  G  N  E  F  I  L  D  R  S  K  A  G  R  I  D  F
```

```
TCCGAATTTTGGGCGGCTATGGAATGCTCCGAGAATGATCGAGAGTTCTATTTTGAAATT
  P  N  F  G  R  L  W  N  A  P  R  M  I  E  S  S  I  L  K  L
GAGAATAATTATCGATCGATCTTCAATTGAATTTTTTGCGGACGATGGATTAACAGTTAT
  R  I  I  I  D  R  S  S  I  E  F  F  A  D  D  G  L  T  V  M
GACAGCACTTTTTTATTCCGAAGAAGATATTGCGTCGAAAATGGTCATTCAAGTTCATTC
  T  A  L  F  Y  S  E  E  D  I  A  S  K  M  V  I  Q  V  H  S
AGCTTCCCTGGATTCAATGATTGCTGTGAAAGAATTTAATGTTTATAAGATGAAAAGTAT
  A  S  L  D  S  M  I  A  V  K  E  F  N  V  Y  K  M  K  S  I
TTGGAAATAAT
  W  K
```

MDL-9

```
GGAATAATCATCCGAGAATACATACAACATTTTTCATGATATAAGATCTGTTCGTTCAAA
AACCAAACTCTAATACAATAAAAATTTCAGACGAATTCATTTATTTGGGACACGTTGGAT
GTTGAAAAACATCTGTAGTTAGTTGAATGACGTGACAAATAATTTTGTATTTCTAAAAGT
AGAAAACATTTGCATTGAATGTATCAATGATAAAAGTGATCGTTTATGAGAATTAAAGGT
GAAAGTAAAATAAAGTTTTGATAGAAAATTCGGCTATACAAAATTGAACTCAGAAATCAT
CGATCATAATGTTTCGAGTTGATATTTTATTGTTTGCGCTTTCGATAGTTCCTATTTCGA
         M  F  R  V  D  I  L  L  F  A  L  S  I  V  P  I  S
ATGCAGTTTATGACGAACTTTATCGGCCACAATTACATTATAGCCCAGCAAAAGGATGGT
N  A  V  Y  D  E  L  Y  R  P  Q  L  H  Y  S  P  A  K  G  W
TAAGCGATCCGAATGGTTTGATTTACAAAGAGGGCATATATCATTTATTTTTCCAATGTC
L  S  D  P  N  G  L  I  Y  K  E  G  I  Y  H  L  F  F  Q  C
GTCCGGATCATATAGCGCAAGGTAAGCAGCAAATTTAACGGAATTTATTGAAAAAATATG
R  P  D  H  I  A  Q
TTCCTATTTCATGCAAAGTCGAATCGATGATAAATAAATTCCAAATTTTAAATTATTAGG
                                                            G
AAGTGTGCATTGGGGGCATTCAATATCTTCTGATTTGATTCATTGGAAAGCTCTCGATAC
  S  V  H  W  G  H  S  I  S  S  D  L  I  H  W  K  A  L  D  T
TGCGCTTTATCCACCTGTAGGACATGAAATGTTTTCGGGTGGCGCAATTTTCGATTACAA
  A  L  Y  P  P  V  G  H  E  M  F  S  G  G  A  I  F  D  Y  K
AAATGTTACTAAACTTCAAACCAACGAGAATGTGCCGGCATTAATTTTGTTACCATCTGC
  N  V  T  K  L  Q  T  N  E  N  V  P  A  L  I  L  L  P  S  A
AGCTGTTTGGTCGACACGTGAGCAAAACATATGGTTAGCGTACAGTAATGATGGTCCTGA
  A  V  W  S  T  R  E  Q  N  I  W  L  A  Y  S  N  D  G  P  E
GTACAATAAATTTACATATTATGAAAAAAATCCAGTAATTCGTGGACCATCAAGTTATGG
  Y  N  K  F  T  Y  Y  E  K  N  P  V  I  R  G  P  S  S  Y  G
CAAACTAATAACAGCATTCCGAGACCTAACAGTTTTCAAGTATCACGATAATTATGTTTC
  K  L  I  T  A  F  R  D  L  T  V  F  K  Y  H  D  N  Y  V  S
AATATTGGTTCAATATAATCGCACCCAGTTCTATAGTTCGCACGATTTAATAGACTGGGA
  I  L  V  Q  Y  N  R  T  Q  F  Y  S  S  H  D  L  I  D  W  E
GCTGATCAGTGAGTTTGGAGAATACGAAGGCTCACACGCAGGAAGATGGGAATGCCCTTC
  L  I  S  E  F  G  E  Y  E  G  S  H  A  G  R  W  E  C  P  S
ACTATTTCCCTTTAATGTTTCGATTGATGGGTAAATCGGCTTACAATGAAATTTTCGCCT
  L  F  P  F  N  V  S  I  D  G
AATTTTTACGTCAGTAACATTTCGTTTCTCCTTTTTCAGAAAACAAGTTGAAAAATACGT
                                        K  Q  V  E  K  Y  V
CATGATAATTACTTTAACCGATTATGTACATCCGGTCCATCAATACTTCATAGGCTCTTT
  M  I  I  T  L  T  D  Y  V  H  P  V  H  Q  Y  F  I  G  S  F
TGACGGTAAAAAATTCACAAATGAAAACACCAAAGAAACAATATTATGGCTAGAGTACGG
  D  G  K  K  F  T  N  E  N  T  K  E  T  I  L  W  L  E  Y  G
ACCAGACAGTTTTGCTGGCGTTACATACAATGAATTACCGGATGGTCGACGTATTTTTAT
  P  D  S  F  A  G  V  T  Y  N  E  L  P  D  G  R  R  I  F  I
CCATTGGATGGGTCGATGGGAATATGTTGCAAATTTGAATTTCAGCCCTTGGCTTGGCCA
```

```
  H  W  M  G  R  W  E  Y  V  A  N  L  N  F  S  P  W  L  G  Q
ATTGGGCATTCCCAGGGAATTAAATTTGATAAAAGTAGGTGATCAAATTCGGCTCACATC
  L  G  I  P  R  E  L  N  L  I  K  V  G  D  Q  I  R  L  T  S
TGAACCTGTTCGTGAAATGGAATCGTTAAGAATTAATCATGTTCGCAGGCAAAACATTAA
  E  P  V  R  E  M  E  S  L  R  I  N  H  V  R  R  Q  N  I  N
TATTACGAATGAATTTAGTTATCAAATCGCTGATAGTGCAAAAAAAAATCATTTAGCGGA
  I  T  N  E  F  S  Y  Q  I  A  D  S  A  K  K  N  H  L  A  D
TATTGAACTGATGCTGGATTTGCAAAATCTTGATTCACGTGACGCATTTCAGATTGTGTT
  I  E  L  M  L  D  L  Q  N  L  D  S  R  D  A  F  Q  I  V  F
TTCAGGCAAAAGTGATGAATTTAAAATTATCTTTAAAGAAAAGGAATTTATTTTGGATCG
  S  G  K  S  D  E  F  K  I  I  F  K  E  K  E  F  I  L  D  R
TACACGCGCTGGTAAAATAATTCCAAATTTCGAGGGTGTGAAAATTTCACAAGGCAATTT
  T  R  A  G  K  I  I  P  N  F  E  G  V  K  I  S  Q  G  N  F
TGATAACCCAAATTTCATAAGGCCAACAAAAGAAAGAAATTTCGAAGATCTGTGGAAAGC
  D  N  P  N  F  I  R  P  T  K  E  R  N  F  E  D  L  W  K  A
TCCACGATTGATTGATAGCTCTAATTTGAAATTAAGAATTATTATCGATACAAATGCAAT
  P  R  L  I  D  S  S  N  L  K  L  R  I  I  I  D  T  N  A  I
TGAAATGTTTGCCGATGATGGATTAACAAGTATGTGCGCCCTTTTCTTCTCCAAAGATGG
  E  M  F  A  D  D  G  L  T  S  M  C  A  L  F  F  S  K  D  G
CATTGCATCAAAAATGACCATTCAAGTCCATTCGTCCACCAAAAAATCGCACATTTATTT
  I  A  S  K  M  T  I  Q  V  H  S  S  T  K  K  S  H  I  Y  L
GAGGGACATGAATGTTTATGAAATGAAAAGTATTTGGCACAAGGATGAAAGTACTAAAAG
  R  D  M  N  V  Y  E  M  K  S  I  W  H  K  D  E  S  T  K  S
CAAATTCAATTTTTCGGATAATAAAAAACAAAAAATTGTAAAATAAAAGAAATCAAGAA
  K  F  N  F  S  D  N  K  K  Q  K  I  V  K
AATGAACTGCTTGCGATGCCCAGATTTGATGTGTGATGGGCTCCTAGTTTTCCATTCAAC
GAATAAAAGTCTTCGAAATCATTTTTTGTTTGTGTTTCATTTGAATAGCTCCCGATTCTT
ATTACTGTAGTTTCATTTACCAGAAATATAGCATCATATTTTCTACAATATTTTCTAACC
CATGGATTTAATTCAGCTGAAAAAATATTGAATAAGTACGAAATAAATGATTTAGCTACT
TCGATGTACTTTGGCTTCGATTCGGATTCGCTATAGCAATTCGAGTCACAATAT
```

>MDL-10

```
CGCACCCTCGTCTCAAACTGATAACTGAACTATCATTTTTCTATTCATTCAGTCTTCAGT
GTGAATGTCAAACGATACAGACGTACAAATCACATGTGTAATAAAATCGTTAAGTGTAAT
GGAATTTAGTTTCGATTTTTGAATCAAATTGAATTGAATTATATTCAAAAATTTCTAATC
ATTGTGAATTTAGCAAAATTTTATCAAAATTACACGATAATTCATAAAATTTCGAACCAG
TGACGATCGTAAACAATTTTTTCAAAATTAGAAAATTAGTGTGAATCACTTATCAGTTCA
TATGATCA**ATG**GCGCCAGAACGAATATCAAGTGAATCATATGACGAACAATATCGCCCAC
        M  A  P  E  R  I  S  S  E  S  Y  D  E  Q  Y  R  P
AATTACATTACAGCCCACCATCAGGATGGATAAATGATCCAAATGGTTTAGTTTATCACG
Q  L  H  Y  S  P  P  S  G  W  I  N  D  P  N  G  L  V  Y  H
ATGGTATTTATCATATGTTTAATCAAAACAATCCGAATGATGTAATTCCCG**GT**GAGCAGG
D  G  I  Y  H  M  F  N  Q  N  N  P  N  D  V  I  P
AACATAGCAAAACAAAACAGTAAACCTTTGTCTGATGAAACGCAATTGTGCTTTTGT**AG**G
                                                        G
TCGAATTTCATGGGGACATGCGATTTCCACTGATTTAGTTCATTGGAAAACACTTCCATG
  R  I  S  W  G  H  A  I  S  T  D  L  V  H  W  K  T  L  P  C
TGCTATTCCTTCAACGGATGAGAATTCAATTTTTTCGGGCAGTGCAATTATTGATGATGA
  A  I  P  S  T  D  E  N  S  I  F  S  G  S  A  I  I  D  D  D
TAATGTAACCGGTCTTCGAACAGATGACCAAATAAAAACATTAATAGCCATTTTTACGGC
  N  V  T  G  L  R  T  D  D  Q  I  K  T  L  I  A  I  F  T  A
TCATAATTTGTCAACCAATGAGGAAAACCAATGGTTGGCATACAGTAACGATGGCCCTGC
  H  N  L  S  T  N  E  E  N  Q  W  L  A  Y  S  N  D  G  P  A
ATACGAGAAATTCCATTTTTATGAAAAAAATCCAATTATTCCGAATCCAAATCCGAAGAA
  Y  E  K  F  H  F  Y  E  K  N  P  I  I  P  N  P  N  P  K  K
ACAAATCGATTTTAGAGATCCATCGGTTTTCAAATACAAAGATCGATTTGTGATATTATT
  Q  I  D  F  R  D  P  S  V  F  K  Y  K  D  R  F  V  I  L  L
```

GGCTGCACACAATCACATCAAAATTTATAATTCACTGGATTTGTTGCAATGGAAATTTGT
  A A H N H I K I Y N S L D L L Q W K F V
GAATGAATTTGGACTAAACGATGGATCGCATGCAGGAACCTGGGAATGTCCTTCACTATT
  N E F G L N D G S H A G T W E C P S L F
TCCTATAAAAGTCACAATTGATGG**GT**GAGAAAATTGTGAATAGAATATGTCACTCAATTT
  P I K V T I D G
TTGTATGCATTAATTTTTTGCCATTTAATTCGT**AG**CGTGGAAGTTGAGAAGTATGTTTTG
                                       V E V E K Y V L
ATAGTTGGCCTAACCGACGGCGCGATACAGACCACACAATATTTTATTGGATCCTTCGAT
 I V G L T D G A I Q T T Q Y F I G S F D
GGCGAAAAATTCATCAATGACAACTTCAAGGAGACCGAATTATGGCTAGACTTTGGACCA
 G E K F I N D N F K E T E L W L D F G P
GATAGTTATGCTGGTATTACTTATAACAAATTACCAGATGATCGGCGTATTTTCATCAGT
 D S Y A G I T Y N K L P D D R R I F I S
TGGATGAATAATTGGCTATACGCTCAACATTTGAATTTTAACGTTTGGAACGGACAAATG
 W M N N W L Y A Q H L N F N V W N G Q M
GGTCTCGCTCGAGAATTAAAATTGATTCAAGTGCAAGATCGAATTCTGCTCTCATCTTTG
 G L A R E L K L I Q V Q D R I L L S S L
CCTGTTCATGAATTGGAAATGTTGCGAATCAATCCAGTTAGCATTCAAAATGTATTGATC
 P V H E L E M L R I N P V S I Q N V L I
GAAAATGACTTTATTTATAAAATTGATAGGAATGATAGCATCGCAAAGAAACATTTAGTA
 E N D F I Y K I D R N D S I A K K H L V
GACATAGAAATGAACTTAGATCTATCGAAGTTCAAGAAAGGCGACTCATTTGATATTGTA
 D I E M N L D L S K F K K G D S F D I V
TTTTCTGATGAAAATGATGAAATAAAGATTTCCTACAAACGAAACGAATTTATATTGGAT
 F S D E N D E I K I S Y K R N E F I L D
CGATCAAAAGCAGGAAGAACAGATTTTCCAAATTTTGGACGGCTATGGAAAGCCCAGAGA
 R S K A G R T D F P N F G R L W K A Q R
TTCATCGATAATTCCAATTTAAAATTGAGAATAATTATCGATCGATCTTCCATTGAATTT
 F I D N S N L K L R I I I D R S S I E F
TTTGCTGATGATGGCTTAACAGTTATGACTGCACTTTTTTATTCCAAGGAGGATATTGCT
 F A D D G L T V M T A L F Y S K E D I A
TCGCAAATGGCCATTTATGTTCATTCATCGGCAAACGGCTCAATCATTGATTTGAAGAAA
 S Q M A I Y V H S S A N G S I I D L K K
TTAAATGTTTATCAAATGAAAAGTATCTGGAGTCAA**TAA**TGAAACATGTATGCAAATATT
 L N V Y Q M K S I W S Q
TATCAAAAATCAAATAATTATTCTAAAACTCAATGTAATAAATTGTGCTCGCAATAAATT
TTTCAAATGTTTTATGATTTGAGATCTTGCCGTATCTATTGTTTGAATATTTCTTTTATA
GCGCTAAAAAAATTTATCATTGCAGATTGCATAACAATATCATCAATAAGCCGGCGCGTG
ATTGTCAAAAACGAAAGCCTAAAATATTTTCTATTTGAAATTGTAAACAGTAGATGATTA
CGCGATCCTTATATTATGCATGGGTCGGAAGCGGGCACTGAGTTGTC

MDL-11

AATTGTAAACAGTAGATGATTACGCGATCCTTATATTATGCATGGGTCGGAAGCGGGCAC
TGAGTTGTCATGTGAGAATAAATCGCTTGCAATAAAATGTAGAATAAATATTCAAGATTA
AATATTTTTTTAATATTCAGATAATGCATATAATAAATTTCATTTGAAAATAACAAAAAA
AAAACCTGTCTTTTGAACCGAATTGAATATTG**TATAA**AAGACTAGGAGGATTTATTTTAG
CAGCCAGTTAACAATTGTTCCTCAAATGAGGCTCAATGAAGTAAGTGATCATCGGAGTCT
TAGGAAAA**ATG**TTAAAAATCAATATTTTAATCGTTGCATTTTCAATGATTCCGGGCTCAT
        M L K I N I L I V A F S M I P G S
TTTCATTATATGATGAAAAATATCGACCACAAATACATTTTAGTCCACCAAACGGATGGA
 F S L Y D E K Y R P Q I H F S P P N G W
TAAACGATCCAAATGGATTAGTTTATTTTAATGGCATTTATCACATGTTCTTCCAAAATA
I N D P N G L V Y F N G I Y H M F F Q N
ATCCATATAATACCGTTCCTG**GT**GAGTTAAAGATTCAAAATTATTCAATTTAGGAAAAAT
N P Y N T V P
TAATAAATCAGTTTAATCGAATTTCTGAATGACAC**AG**CTAACAACATTCATTGGGGACAT

```
                                        A  N  N  I  H  W  G  H
GCCATTTCACGGGATTTGATTCATTGGAAATCCCTTTCTACTGCAATTTATCCAAAAGAT
 A  I  S  R  D  L  I  H  W  K  S  L  S  T  A  I  Y  P  K  D
GGAAATCTAATATTTTCGGGCGGTGCAATCATTGATGAACACAATGTGACTGGTCTTCGA
 G  N  L  I  F  S  G  G  A  I  I  D  E  H  N  V  T  G  L  R
ACCAATGACAACAGAGCAACATTAATAGCAGTGTATGCGGCTCATCATTTGTCAACCAAT
 T  N  D  N  R  A  T  L  I  A  V  Y  A  A  H  H  L  S  T  N
GATGAAAGTCAATGGATTTCATACAGTCATGATGGCCCCTTTTACGAGAAATTCCAATAC
 D  E  S  Q  W  I  S  Y  S  H  D  G  P  F  Y  E  K  F  Q  Y
TATAAAAAGAATCCAATCATTCGGAATCCAAATCGTCGACAAAAAGATTTTCGCGATCCA
 Y  K  K  N  P  I  I  R  N  P  N  R  R  Q  K  D  F  R  D  P
TTCGTTTTCAAGTACAAAGATCACTATGTTGTCACAATAGCTGCACATGATCGCATCATG
 F  V  F  K  Y  K  D  H  Y  V  V  T  I  A  A  H  D  R  I  M
ATTTTTAATTCACGCAATTTATTGAATTGGAAGCTGGTGAGCGAATTCGGTGTAGATACA
 I  F  N  S  R  N  L  L  N  W  K  L  V  S  E  F  G  V  D  T
CATGGAAGAGTTTGGGAATGTCCTTCACTTTTTCCAATCAACGCAACAATCAATGGGTAA
 H  G  R  V  W  E  C  P  S  L  F  P  I  N  A  T  I  N  G
ATTGAAATTTGAATGACAAAGTAGTTTTCCGATTATTATGATTTTTTTTTTCTACAATTT

TCAGAGCCAAAGTTGAAAAATGGGTTTTGACAATTAGCCTAACTGGCAACGAATTACCAA
      A  K  V  E  K  W  V  L  T  I  S  L  T  G  N  E  L  P
ATCAACAATATTTCATAGGATCTTTTGATGGCAAAAACTTTAAGAATGATCACTCTAAAG
N  Q  Q  Y  F  I  G  S  F  D  G  K  N  F  K  N  D  H  S  K
ATGTGAAATTATGGCTGCACTATGGACCCGATAGCTATGCTGGAATCGTTTACAATCAAT
D  V  K  L  W  L  H  Y  G  P  D  S  Y  A  G  I  V  Y  N  Q
TGCCAGATGGTCGTCGAATTTTTATCAGTTGGATGAATAAATGGGAATATGCACAGCAAT
L  P  D  G  R  R  I  F  I  S  W  M  N  K  W  E  Y  A  Q  Q
TGAATTTCAACAGATGGAATGGACAAATGGGCATCGCTCAAGAATTGAGATTGATACAGG
L  N  F  N  R  W  N  G  Q  M  G  I  A  Q  E  L  R  L  I  Q
TAAAAGATCAAATTCGATTGGCATCATTGCCTGTTCGTGAGCTGAAAAAATTAAGAACCA
V  K  D  Q  I  R  L  A  S  L  P  V  R  E  L  K  K  L  R  T
AGCAACTTGAGCAAAAGAAAAATATTAAAATCACGAATGACTACGTTTATAAATTAACTA
K  Q  L  E  Q  K  K  N  I  K  I  T  N  D  Y  V  Y  K  L  T
AGAATGGATGTCATAAAGCAGAAGATAAATTGGACATTGAAATGACGGTCGACCTAAAGA
K  N  G  C  H  K  A  E  D  K  L  D  I  E  M  T  V  D  L  K
ATCTCAAGGCCGGCGACACATGCAATTTTGTGTTTTCTGGTTCAAATGAATATTTGAACA
N  L  K  A  G  D  T  C  N  F  V  F  S  G  S  N  E  Y  L  N
TTTCTCTGAAAGGAAATGAATTCACTTTGGATCGATCGCATTCGGGAAGAACCAATTTTA
I  S  L  K  G  N  E  F  T  L  D  R  S  H  S  G  R  T  N  F
CAAATTTTGCAAAGCCAACGAAATGCAAAAGGCAAAGTGATGACTCAAAGTTAGAAATGA
T  N  F  A  K  P  T  K  C  K  R  Q  S  D  D  S  K  L  E  M
GATTTGTTTTCGATCTATCATCGATTGAAGTCTTCATTGACGGTGGAATGACAACTATGA
R  F  V  F  D  L  S  S  I  E  V  F  I  D  G  G  M  T  T  M
CAGCACTTTTCTATTCCAAAGAAGACATCGCATCAAAGATGCAGATTCGTACACATGCAA
T  A  L  F  Y  S  K  E  D  I  A  S  K  M  Q  I  R  T  H  A
AGGATTCATCGATTATTTTGAAAGATTTGAAAGTTTATAGATTGAAAAGTATTTGGAATG
K  D  S  S  I  I  L  K  D  L  K  V  Y  R  L  K  S  I  W  N
AATTCTCCATTCAGCCATCGCTCGATAAGTTTCCGTGTAGCGAGTGAAGAGGAAATATTA
E  F  S  I  Q  P  S  L  D  K  F  P  C  S  E
ATCAAAAACTAATTTATCGGCATAGAAATTTCCCATTTTGAATTGGTAGGAAATGCTTTG
ACGCTTAGTTCTTTACAATCCTTTATACTTGATGTTTCATGCTATTTGGTTCGCCAAAAA
TAAATATGAAATAATAATAATAATATCGTAATTTATTGCGGGTTAGTCTTTATTCATAAA
CATATCATTTTAATTACATCGTTCAATGAACTGTACGGAGAATATTACGCTTGACATGGC
ATTAAAATGGTTTCCAGCACAATGATTCGATGTAGCGGGAATTTAAATAGCGTTC
```

>MDL-12

```
TTCTTGTCCTGAAGCGATTTATTCTCAAATTATGAACATTCGATAGATAAAAAGATGGGG
```

AAGATTTGATATCGTCATTTTTAAATCGAGAGTTGAAGGGTTAATTATGTTTTGAATCGA
TTCAGCGACAAAAATAACAATTCACAATTAATTGAATATAAACGCTCAACGAACG**TATAA**
AAGGATGGTAAATTTTGTATTTGATGGTAGTCATCAATTGGTAATCTCCCAACAAAGTTT
TAAAAAATAGACTGAAGACGAAGAGTAAATTCAACTTGCTAAACTTTGAATTAAAACAAA
GGAACAAGA**ATG**TTACAAAATAGAATTTTGTTCGGTACATCGTTGATGGTTTTAGTTTCA
M L Q N R I L F G T S L M V L V S
TTTGGAAGTGCATTGTATGATGAACCATATCGACCACAATTACATTTCAGTCCACCATCA
F G S A L Y D E P Y R P Q L H F S P P S
GGATGGATGAACGATCCAAATGGTTTGATCTATCATGATGGTGTTTTTCATCTATTCTGC
G W M N D P N G L I Y H D G V F H L F C
CAATATAATCCAGCTGCACCTCTTCATG**GT**AAATAGAAGTGTCCAACTAAGAACCAATTT
Q Y N P A A P L H
CAAACTGTGCATTTTAGACCCATTTAACCGATTTACTTTAAATAAAGAAATAATAATAAT

GATTTCACATCTAATATTTT**AG**GTAACATTCATTGGTTTCATGCAATTTCTTCGGATTTG
G N I H W F H A I S S D L
ATTCATTGGAGAAATCTTGGCATTGCACTTGCTCCACGGGAAGGAAATTTAATTTTTTCA
I H W R N L G I A L A P R E G N L I F S
GGCGGCGCAATCATTGATCATGACAACGTAACTGGTCTACAAACAAAGAACGATAAAAAA
G G A I I D H D N V T G L Q T K N D K K
ACATTAATTTTGGTGTTCACCGCTCATAATATCGTTGAGGAAAAACAATGGTTGGCATAC
T L I L V F T A H N I V E E K Q W L A Y
AGCAACGATGGGCCCGAATACGAACACTTCCAATACTATAACCATAATCCAATCATACCA
S N D G P E Y E H F Q Y Y N H N P I I P
AATCCAAATCCGAAAACGTATAAAGATTTCAGAGATCCATCAGTTTTTAAGTATGAAGAC
N P N P K T Y K D F R D P S V F K Y E D
CATTTTGTGATGGTATTAGCTGCATACGATCATGTCATGATTTATAATTCACCGGATTTA
H F V M V L A A Y D H V M I Y N S P D L
TTGGAATGGAAACTGGTGAGTGAATTTGGAATCGACGAAGGATCCCATATCGGGACATGG
L E W K L V S E F G I D E G S H I G T W
GAATGCCCGTCCCTCTTTCCTATTAACGTAACAATTGATGG**GT**ATGTTTTCGCGAATATT
E C P S L F P I N V T I D G
TTTTTTATTTAATCATCAATTTGTGAGTGTAAAATTTATTTTTGACAATTTTC**AG**TGTCG
V
AAATTGAAAAATACGTTTTGATCATCAGCCTAACGGACAATGCAATACCAAGCATGCAGT
E I E K Y V L I I S L T D N A I P S M Q
ACTACATAGGATCTTTTGATGGACAACATTTCACGAATGAAAACTCCAAAGAGACGGAAC
Y Y I G S F D G Q H F T N E N S K E T E
TATGGCTCGATTATGGTCCGGATAGCTATGCTGGCATTACATATAATCAATTGCCGGACG
L W L D Y G P D S Y A G I T Y N Q L P D
GTCGACGTACTTTCATAAGTTGGTTATTTCGATGGGAATACGCTACACATATGAATTTCA
G R R T F I S W L F R W E Y A T H M N F
GTATTTGGAATGGACAAGCCGGAATTGCTAGAAAATTGATGTTGAACATGATTGGTGATC
S I W N G Q A G I A R K L M L N M I G D
GCATTCAACTTTCCTCTTTACCTGTTCGTGAATTTAAATCTTTGCGAATCAAACAACTCG
R I Q L S S L P V R E F K S L R I K Q L
CCAATAAACAAAGAAGTATACCGATTGAGGACAAGCTCAGTTTTGAATTTTCTAAAAATG
A N K Q R S I P I E D K L S F E F S K N
GCGCTAAAGGAAGGAAACTTTTGTTGGACTTGGAAATGATTTTCGATTTAACGAATTTAA
G A K G R K L L L D L E M I F D L T N L
AAGGAGACGACCAATTCGACATTGTATTCTTTGACACTAATGACAATTTGAACATTTCAT
K G D D Q F D I V F F D T N D N L N I S
TCAATGGAAATGAATTCACTTTAGATCGATCGAATGCAGGAAAAACCGATTTCCCGAATT
F N G N E F T L D R S N A G K T D F P N

TCGGGCGACTTTGGAAAGCTCCACGTTTTGTTAAAAGTCCAGAATTAAAATTAAGAATAA
F G R L W K A P R F V K S P E L K L R I

TTATTGATCAATCTTCAATTGAATTTTTTGCCGATGATGGACTTACAGTTATGATAGCGT
I I D Q S S I E F F A D D G L T V M I A
TTTTCATTTCCGATGAAGATATTGCTTCAAAAATGGCCATTCACGTTCATTCATCTTCAG
F F I S D E D I A S K M A I H V H S S S
TGACGTCCATGGTTTATTTAAAAAAATTAAATGCTTATCAATTGAAAAGTATTTGGAAT**T**
V T S M V Y L K K L N A Y Q L K S I W N
AATGCATTTTGATAATAATAACATGTACAGCATTACAGTAAAGTCTCTCTAAAAAGACCA
TCTCTCATTTTTGAACCAATTAACCGGCAACCAAAGTGATCCTCATAGTGGAACTATTCT
CTTTATATTACATAAACATTATACACATACTTATATTTCATTAATTTTATCCGCATCCTT
TGTTCATCGTATTCATTTATTAATAAGATCAGCTGTAAAACACCAAATCGTAGGCAACAA
CAGAATTCCAAATTTGAACCTTATTTCATAATTCTTTTATTATTTGTATTATTTGTATGG
CGAATTTTTT
>MDL-13

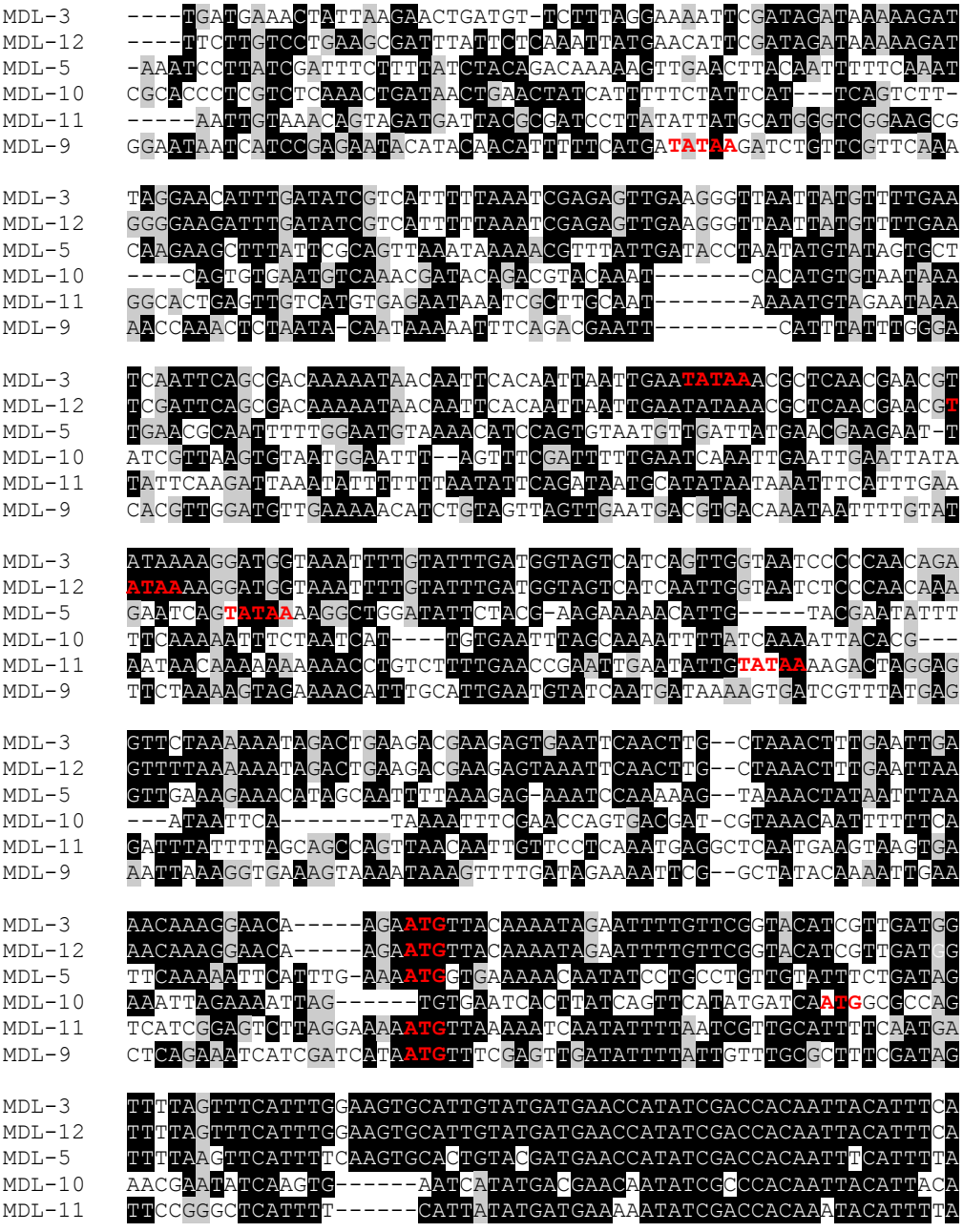
GATTCGGTCAACAAATAAAAATTCATATCAGTCGACACTTCGAGCATGACATTTGACAAA
TAGAATTTGAATGAAATTGTTAATTAAAGTCCAAAGGTTTGGTCTTCTGAATGAGATCGC
AATATGTTTTATTTTTTTTTAGCAGTATTTAGATAATAACACTTATAGCATTTATTGAAA
TATTTTTGCGTTCGTAATCATGCGCAAAAATATGTTTCAGTAGTGTTCATCATAAATTAA
CTATTAACTGTTAGAAGGTCATTCCACTTCAGTGATAAGAATCAAATGTTTAAATTCAAC
ACTGATATAGCAAAAATA**TATAA**AAGGCTGGATGATTTCGTATTGGAAAACAGTAATCAA
TAAATATTGGATAGAGTAAAAGGATCTGGTGTGGTACAAAACTAAAAAAAAGATTTAACA
TTGGTCAATTTTTAAGGAATCATCACTAAATTGAAAGACTATTTTTCAAT**ATG**TCTCGAA
M S R
CAAATATTTTACTCTGTACAATTTTGTTAGTTTTAATTTCATTAACAAGTGCAGTGTACG
T N I L L C T I L L V L I S L T S A V Y
ATGAACTTTATCGTCCACAAATACATCTTAGTCCTCCAAATGGATGGATGAGCGATCCAA
D E L Y R P Q I H L S P P N G W M S D P
ATGGTTTAGTCTATCATGATGGAGTATATCACGTCTTCTATCAACACAATCCAGTTGATA
N G L V Y H D G V Y H V F Y Q H N P V D
CCGAACCCGACAAATTCC**GT**AAGTTGTTTGCTCAAAAAAAAAAAAACAAAACCAACATTC
T E P D K F
AAATTAATAATTAACTTTGTTTTTTTTTCTTC

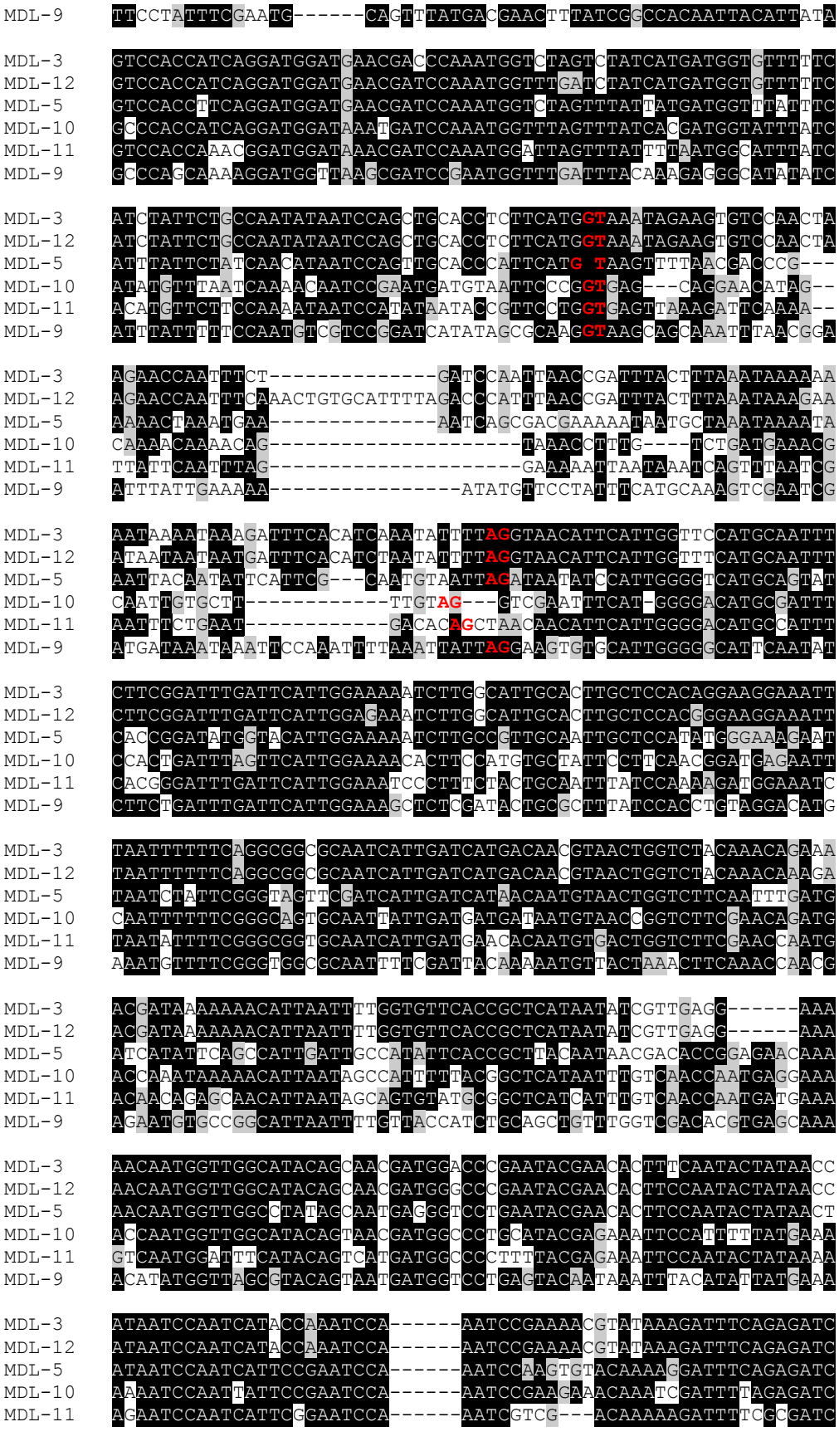
>MDL-14
AATTTTTACCGCTCATAATAGGGCTACATATGAGCAGGAACAATGGTTAGCATACAGTAA
I F T A H N R A T Y E Q E Q W L A Y S N
TGATAGCCCTGAATACAAACACTCCAACTACAATCTATACGAAATTTTTAC**GT**CTGTAAT
D S P E Y K H S N Y N L Y E I F T
TTATAGAGTGAAAGAAGTTTACAAATGAAAATTCCCCGGAAACCACATTATGGGTGGACT

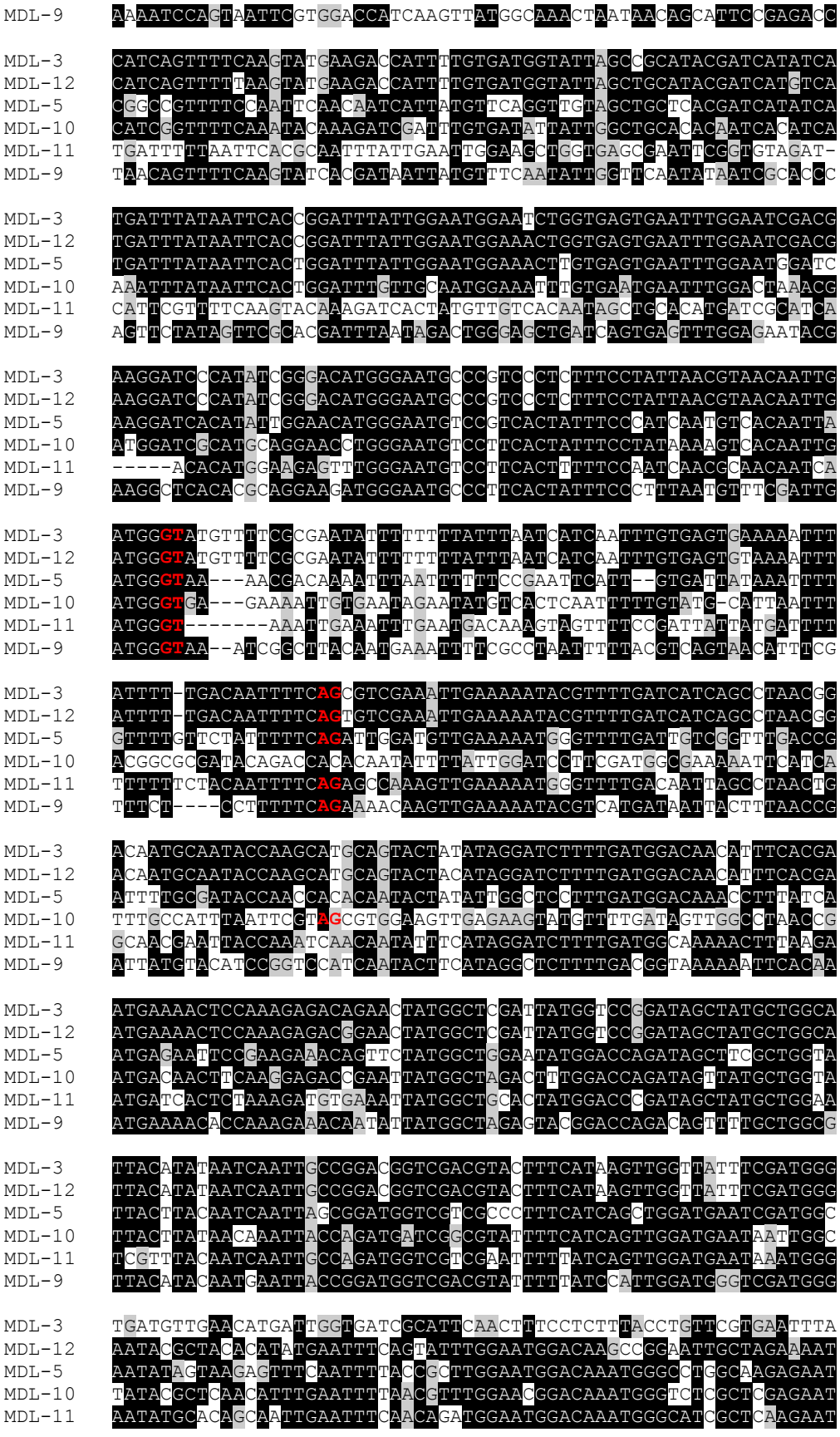
ATTGCCCGGAT**AG**ATATGCCGGCATCACATACAATCAATTGCCCGATGGTCGTCGTGTTT
Y A G I T Y N Q L P D G R R V
TTGTCAGTTGGATGAATAAATGGGAATCTGCAGAGCAATTGAATTTCAATGTTTGGAACG
F V S W M N K W E S A E Q L N F N V W N
GACAAGTTGGACTGATGCGAGAACTTAAATTGAAACAAGTTGGAGATCAAACTCGACTCG
G Q V G L M R E L K L K Q V G D Q T R L
TTTCCTACCCTGTTCGTGAAGTAGAAAAATTGAGAACTTATCTGGTTCGCAAAGAGAATA
V S Y P V R E V E K L R T Y L V R K E N
TTAAGGATCAATTTATTTTTATTAAGTACAAAATTAATTTTTTGAGGACCATTTATCCGA
I K D Q F I F I K Y K I N F L R T I Y P
TT**TAA**AGACCAGTTTATTTTTTTACGGATTTTTTTTAATTGAAAAAAGGACGCAAAAAAC
I
TGATTCAAAGACCATTTATATGAAAAAAGGACCAATTTCACGCGACCTTTTCTGAATCTT
TCGAATGTTGCAAACTTCAAGGCATAAGCTCCACCCCACACCGAGTTAGAGCTGGTACAA
ATG**AATAAA**TGAATAAATATATCGGGTGTGGAAATGAATAATGTCGATTTTTTGGCTTTA
AGAAAGCATTTATTTCAATAACAAGATTAATGAACAGATTATCCAATATATTGCCCATCG

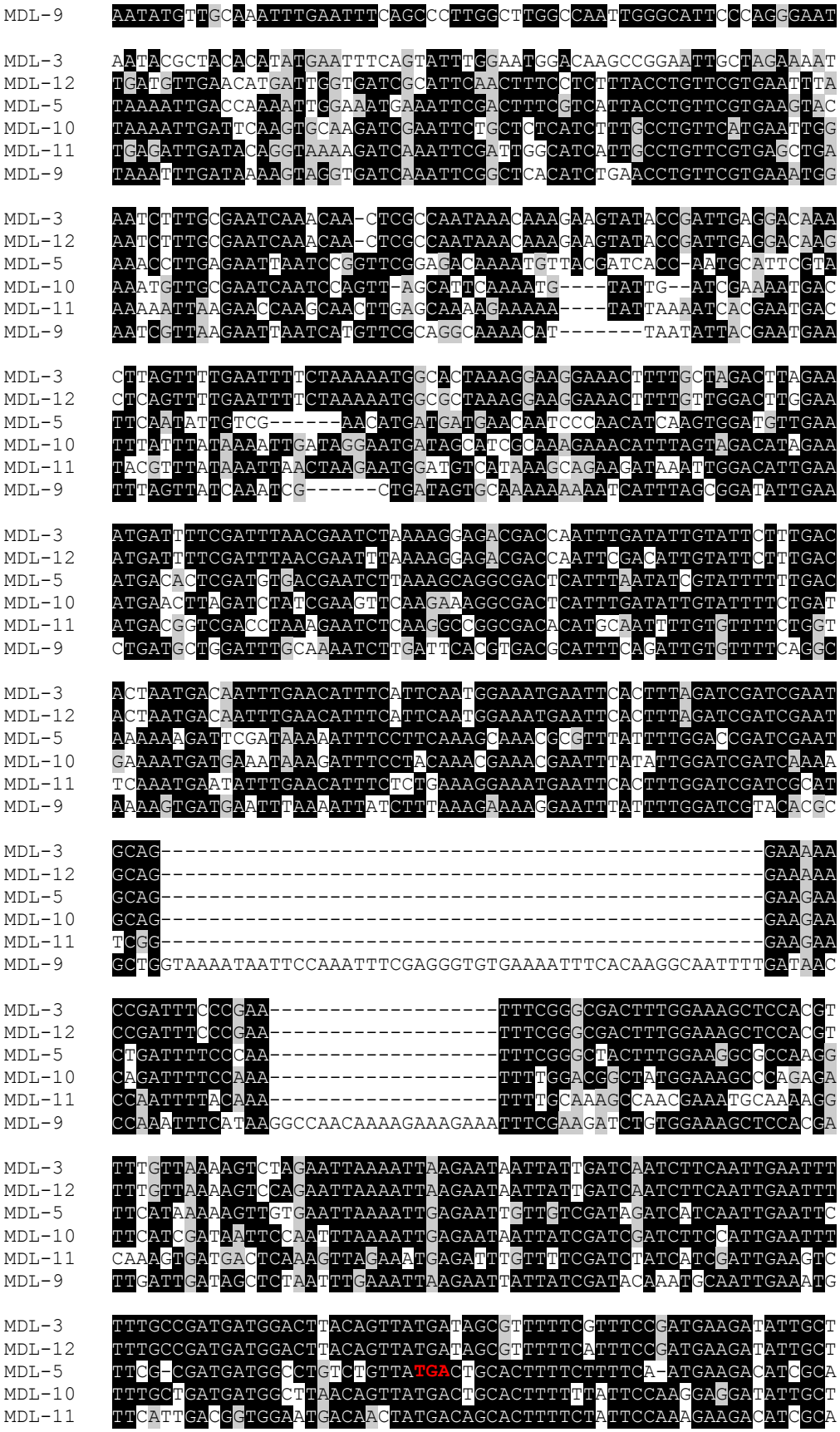
TTAGCTATGACTTCCTCCCATCTTCTCGGAAGTTTATTGATTCCTCGCTTAAAGAAGTCC
TCATCTTTTGAGGCAATCCAGTCGTCGACCCATTTTCGCACTTGTCCGACATTTTCAAAT
TGTATATCGGCCAGGCTAATTTGCATCGATCGGAACAAATGATAATCAGAAGGAGCAAGG
T

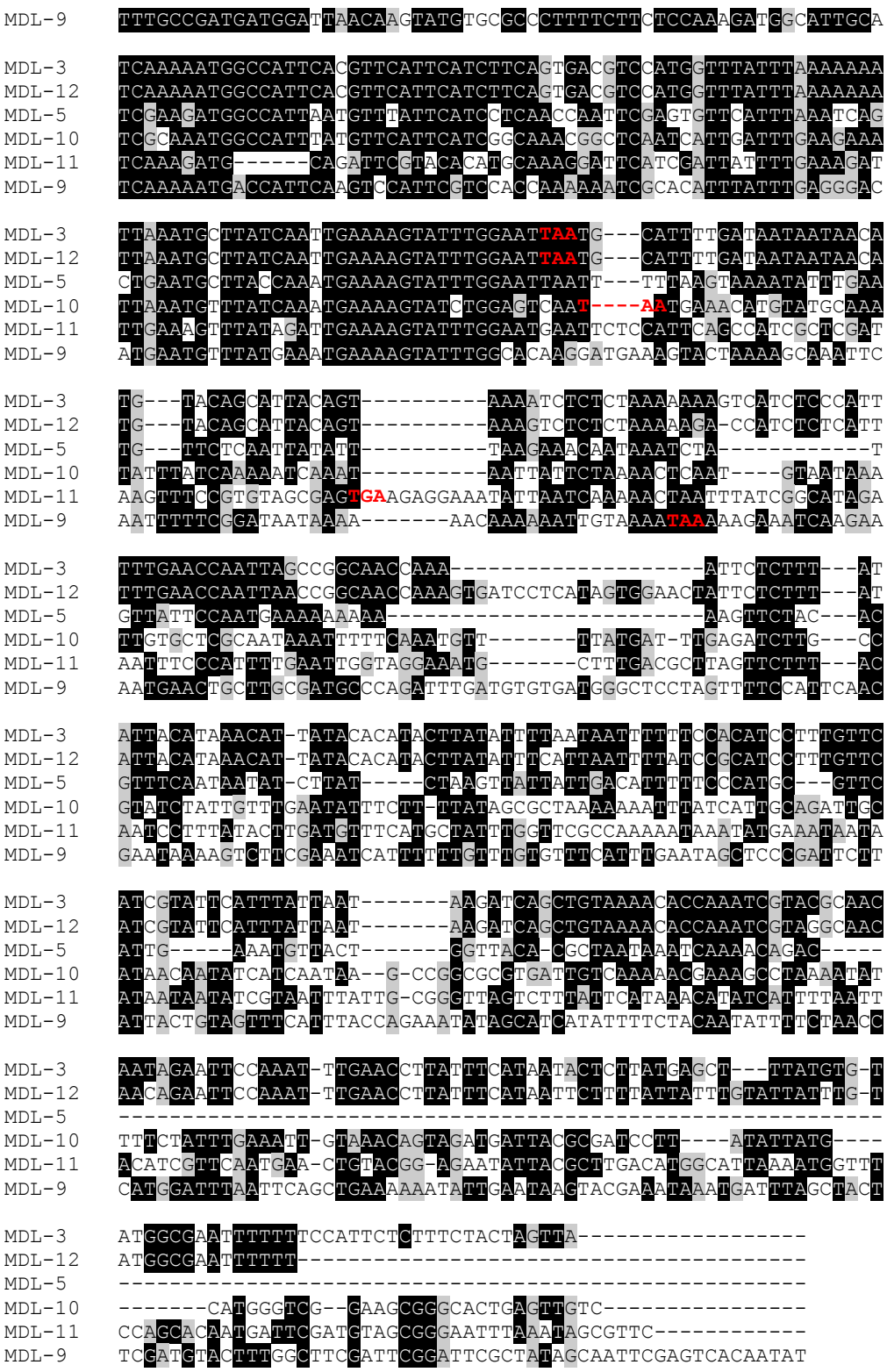
Supplementary data Figure S2. Nucleotide alignments of the 14 levanse/inulase –like genes. Identical residues are highlighted in black. Gaps in the alignment are indicated by dashes**.**




MDL-9 TTCCTATTTCGAAATG------CAGTTTATGACGAACTTTATCGGCCACAATTACATTATA
MDL-3 GTCCACCATCAGGATGGATGAACGACCCAAATGGTCTAGTCTATCATGATGGTGTTTTTC
MDL-12 GTCCACCATCAGGATGGATGAACGATCCAAATGGTTTGATCTATCATGATGGTGTTTTTC
MDL-5 GTCCACCTTCAGGATGGATGAACGATCCAAATGGTCTAGTTTATTATGATGGTTTATTTC
MDL-10 GCCCACCATCAGGATGGATAAATGATCCAAATGGTTTAGTTTATCACGATGGTATTTATC
MDL-11 GTCCACCAAACGGATGGATAAACGATCCAAATGGATTAGTTTATTTTAATGGCATTTATC
MDL-9 GCCCAGCAAAAGGATGGTTAAGCGATCCGAATGGTTTGATTTACAAAGAGGGCATATATC
MDL-3 ATCTATTCTGCCAATATAATCCAGCTGCACCTCTTCATGGTAAATAGAAGTGTCCAACTA
MDL-12 ATCTATTCTGCCAATATAATCCAGCTGCACCTCTTCATGGTAAATAGAAGTGTCCAACTA
MDL-5 ATTTATTCTATCAACATAATCCAGTTGCACCCATTCATG TAAGTTTTAACGACCCG---
MDL-10 ATATGTTTAATCAAAACAATCCGAATGATGTAATTCCCGGTGAG---CAGGAACATAG--
MDL-11 ACATGTTCTTCCAAAATAATCCATATAATACCGTTCCTGGTGAGTTAAAGATTCAAAA--
MDL-9 ATTTATTTTTCCAATGTCGTCCGGATCATATAGCGCAAGGTAAGCAGCAAATTTAACGGA
MDL-3 AGAACCAATTTCT--------------GATCCAATTAACCGATTTACTTTAAATAAAAAA
MDL-12 AGAACCAATTTCAAACTGTGCATTTTAGACCCATTTAACCGATTTACTTTAAATAAAGAA
MDL-5 AAAACTAAATGAA--------------AATCAGCGACGAAAAATAATGCTAAATAAAATA
MDL-10 CAAAACAAAACAG----------------------TAAACCTTTG----TCTGATGAAACG
MDL-11 TTATTCAATTTAG----------------------GAAAAATTAATAAATCAGTTTAATCG
MDL-9 ATTTATTGAAAAA----------------ATATGTTCCTATTTCATGCAAAGTCGAATCG
MDL-3 AATAAAATAAAGATTTCACATCAAATATTTTAGGTAACATTCATTGGTTCCATGCAATTT
MDL-12 ATAATAATAATGATTTCACATCTAATATTTTAGGTAACATTCATTGGTTTCATGCAATTT
MDL-5 AATTACAATATTCATTCG---CAATGTAATTAGATAATATCCATTGGGGTCATGCAGTAT
MDL-10 CAATTGTGCTT------------TTGTAG---GTCGAATTTCAT-GGGGACATGCGATTT
MDL-11 AATTTCTGAAT------------GACACAGCTAACAACATTCATTGGGGACATGCCATTT
MDL-9 ATGATAAATAAATTCCAAATTTTAAATTATTAGGAAGTGTGCATTGGGGGCATTCAATAT
MDL-3 CTTCGGATTTGATTCATTGGAAAAATCTTGGCATTGCACTTGCTCCACAGGAAGGAAATT
MDL-12 CTTCGGATTTGATTCATTGGAGAAATCTTGGCATTGCACTTGCTCCACGGGAAGGAAATT
MDL-5 CACCGGATATGGTACATTGGAAAAATCTTGCCGTTGCAATTGCTCCATATGGGAAAGAAT
MDL-10 CCACTGATTTAGTTCATTGGAAAACACTTCCATGTGCTATTCCTTCAACGGATGAGAATT
MDL-11 CACGGGATTTGATTCATTGGAAATCCCTTTCTACTGCAATTTATCCAAAAGATGGAAATC
MDL-9 CTTCTGATTTGATTCATTGGAAAGCTCTCGATACTGCGCTTTATCCACCTGTAGGACATG
MDL-3 TAATTTTTTCAGGCGGCGCAATCATTGATCATGACAACGTAACTGGTCTACAAACAGAAA
MDL-12 TAATTTTTTCAGGCGGCGCAATCATTGATCATGACAACGTAACTGGTCTACAAACAAAGA
MDL-5 TAATCTATTCGGGTAGTTCGATCATTGATCATAACAATGTAACTGGTCTTCAATTTGATG
MDL-10 CAATTTTTTCGGGCAGTGCAATTATTGATGATGATAATGTAACCGGTCTTCGAACAGATG
MDL-11 TAATATTTTCGGGCGGTGCAATCATTGATGAACACAATGTGACTGGTCTTCGAACCAATG
MDL-9 AAATGTTTTCGGGTGGCGCAATTTTCGATTACAAAAATGTTACTAAACTTCAAACCAACG
MDL-3 ACGATAAAAAAACATTAATTTTGGTGTTCACCGCTCATAATATCGTTGAGG------AAA
MDL-12 ACGATAAAAAAACATTAATTTTGGTGTTCACCGCTCATAATATCGTTGAGG------AAA
MDL-5 ATCATATTCAGCCATTGATTGCCATATTCACCGCTTACAATAACGACACCGGAGAACAAA
MDL-10 ACCAAATAAAAACATTAATAGCCATTTTTACGGCTCATAATTTGTCAACCAATGAGGAAA
MDL-11 ACAACAGAGCAACATTAATAGCAGTGTATGCGGCTCATCATTTGTCAACCAATGATGAAA
MDL-9 AGAATGTGCCGGCATTAATTTTGTTACCATCTGCAGCTGTTTGGTCGACACGTGAGCAAA
MDL-3 AACAATGGTTGGCATACAGCAACGATGGACCCGAATACGAACACTTTCAATACTATAACC
MDL-12 AACAATGGTTGGCATACAGCAACGATGGGCCCGAATACGAACACTTCCAATACTATAACC
MDL-5 AACAATGGTTGGCCTATAGCAATGAGGGTCCTGAATACGAACACTTCCAATACTATAACT
MDL-10 ACCAATGGTTGGCATACAGTAACGATGGCCCTGCATACGAGAAATTCCATTTTTATGAAA
MDL-11 GTCAATGGATTTCATACAGTCATGATGGCCCCTTTTACGAGAAATTCCAATACTATAAAA
MDL-9 ACATATGGTTAGCGTACAGTAATGATGGTCCTGAGTACAATAAATTTACATATTATGAAA
MDL-3 ATAATCCAATCATACCAAATCCA------AATCCGAAAACGTATAAAGATTTCAGAGATC
MDL-12 ATAATCCAATCATACCAAATCCA------AATCCGAAAACGTATAAAGATTTCAGAGATC
MDL-5 ATAATCCAATCATTCCGAATCCA------AATCCAAGTGTACAAAAGGATTTCAGAGATC
MDL-10 AAAATCCAATTATTCCGAATCCA------AATCCGAAGAAACAAATCGATTTTAGAGATC
MDL-11 AGAATCCAATCATTCGGAATCCA------AATCGTCG---ACAAAAAGATTTTCGCGATC

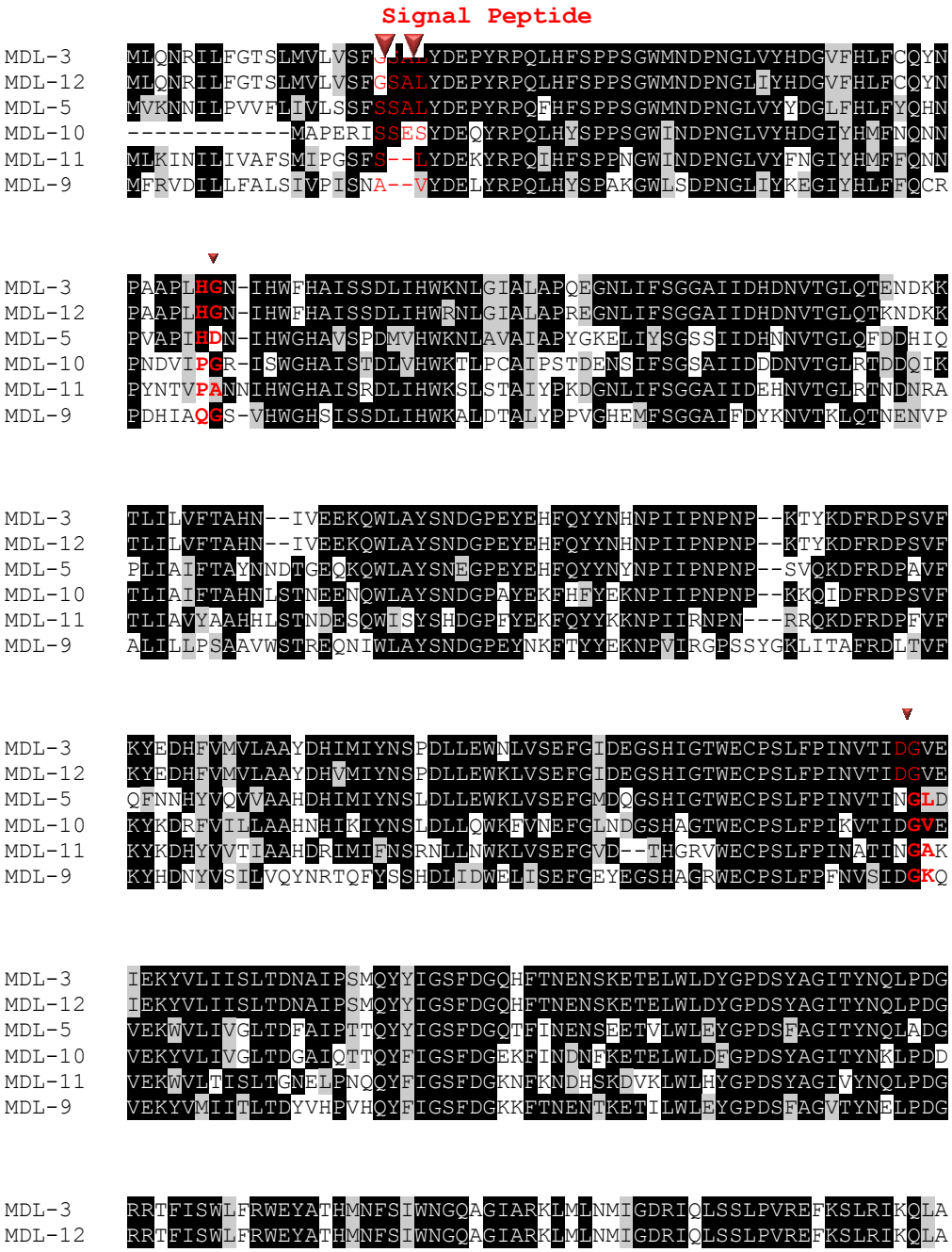

MDL-9 AAAATCCAGTAATTCGTGGACCATCAAGTTATGGCAAACTAATAACAGCATTCCGAGACC
MDL-3 CATCAGTTTTCAAGTATGAAGACCATTTTGTGATGGTATTAGCCGCATACGATCATATCA
MDL-12 CATCAGTTTTTAAGTATGAAGACCATTTTGTGATGGTATTAGCTGCATACGATCATGTCA
MDL-5 CGGCCGTTTTCCAATTCAACAATCATTATGTTCAGGTTGTAGCTGCTCACGATCATATCA
MDL-10 CATCGGTTTTCAAATACAAAGATCGATTTGTGATATTATTGGCTGCACACAATCACATCA
MDL-11 TGATTTTTAATTCACGCAATTTATTGAATTGGAAGCTGGTGAGCGAATTCGGTGTAGAT-
MDL-9 TAACAGTTTTCAAGTATCACGATAATTATGTTTCAATATTGGTTCAATATAATCGCACCC
MDL-3 TGATTTATAATTCACCGGATTTATTGGAATGGAATCTGGTGAGTGAATTTGGAATCGACG
MDL-12 TGATTTATAATTCACCGGATTTATTGGAATGGAAACTGGTGAGTGAATTTGGAATCGACG
MDL-5 TGATTTATAATTCACTGGATTTATTGGAATGGAAACTTGTGAGTGAATTTGGAATGGATC
MDL-10 AAATTTATAATTCACTGGATTTGTTGCAATGGAAATTTGTGAATGAATTTGGACTAAACG
MDL-11 CATTCGTTTTCAAGTACAAAGATCACTATGTTGTCACAATAGCTGCACATGATCGCATCA
MDL-9 AGTTCTATAGTTCGCACGATTTAATAGACTGGCAGCTGATCAGTGAGTTTGGAGAATACG
MDL-3 AAGGATCCCATATCGGGACATGGGAATGCCCGTCCCTCTTTCCTATTAACGTAACAATTG
MDL-12 AAGGATCCCATATCGGGACATGGGAATGCCCGTCCCTCTTTCCTATTAACGTAACAATTG
MDL-5 AAGGATCACATATTGGAACATGGGAATGTCCGTCACTATTTCCCATCAATGTCACAATTA
MDL-10 ATGGATCGCATGCAGGAACCTGGGAATGTCCTTCACTATTTCCTATAAAAGTCACAATTG
MDL-11 -----ACACATGGAAGAGTTTGGGAATGTCCTTCACTTTTTCCAATCAACGCAACAATCA
MDL-9 AAGGCTCACACGCAGGAAGATGGGAATGCCCTTCACTATTTCCCTTTAATGTTTCGATTG
MDL-3 ATGGGTATGTTTTTCGCGAATATTTTTTTTATTTAATCATCAATTTGTGAGTGAAAAATTT
MDL-12 ATGGGTATGTTTTTCGCGAATATTTTTTTTATTTAATCATCAATTTGTGAGTGTAAAATTT
MDL-5 ATGGGTAA---AACGACAAAATTTAATTTTTTCCGAATTCATTT--GTGATTATAAATTTT
MDL-10 ATGGGTGA---GAAAAATTGTGAATAGAATATGTCACTCAATTTTTGTATG-CATTAATTT
MDL-11 ATGGGT-------AAATTGAAATTTGAATGACAAAGTAGTTTTCCGATTATTATGATTTT
MDL-9 ATGGGTAA--ATCGGCTTACAATGAAATTTTCGCCTAATTTTTACGTCAGTAACATTTCG
MDL-3 ATTTT-TGACAATTTTCAGCGTCGAAATTGAAAAATACGTTTTGATCATCAGCCTAACGG
MDL-12 ATTTT-TGACAATTTTCAGTGTCGAAATTGAAAAATACGTTTTGATCATCAGCCTAACGG
MDL-5 GTTTTGTTCTATTTTTCAGATTGGATGTTGAAAAATGGGTTTTGATTGTCGGTTTGACCG
MDL-10 ACGGCGCGATACAGACCACACAATATTTTATTGGATCCTTCGATGGCGAAAAATTCATCA
MDL-11 TTTTTTCTACAATTTTCAGAGCCAAAGTTGAAAAATGGGTTTTGACAATTAGCCTAACTG
MDL-9 TTTCT----CCTTTTTCAGAAACAAGTTGAAAAATACGTCATGATAATTACTTTAACCG
MDL-3 ACAATGCAATACCAAGCATGCAGTACTATATAGGATCTTTTGATGGACAACATTTCACGA
MDL-12 ACAATGCAATACCAAGCATGCAGTACTACATAGGATCTTTTGATGGACAACATTTCACGA
MDL-5 ATTTTGCGATACCAACCACACAATACTATATTGGCTCCTTTGATGGACAAACCTTTATCA
MDL-10 TTTGCCATTTAATTCGTAGCGTGGAAGTTGAGAAGTATGTTTTGATAGTTGGCCTAACCG
MDL-11 GCAACGAATTACCAAATCAACAATATTTCATAGGATCTTTTGATGGCAAAAACTTTAAGA
MDL-9 ATTATGTACATCCGGTCCATCAATACTTCATAGGCTCTTTTGACGGTAAAAATTCACAA
MDL-3 ATGAAAACTCCAAAGAGACAGAACTATGGCTCGATTATGGTCCGGATAGCTATGCTGGCA
MDL-12 ATGAAAACTCCAAAGAGACGGAACTATGGCTCGATTATGGTCCGGATAGCTATGCTGGCA
MDL-5 ATGAGAATTCCGAAGAAACAGTTCTATGGCTGGAATATGGACCAGATAGCTTCGCTGGTA
MDL-10 ATGACAACTTCAAGGAGACCGAATTATGGCTAGACTTTGGACCAGATAGTTATGCTGGTA
MDL-11 ATGATCACTCTAAAGATGTGAAATTATGGCTGCACTATGGACCCGATAGCTATGCTGGAA
MDL-9 ATGAAAACACCAAAGAAACAATATTATGGCTAGAGTACGGACCAGACAGTTTGCTGGCG
MDL-3 TTACATATAATCAATTGCCGGACGGTCGACGTACTTTCATAAGTTGGTTATTTCGATGGG
MDL-12 TTACATATAATCAATTGCCGGACGGTCGACGTACTTTCATAAGTTGGTTATTTCGATGGG
MDL-5 TTACTTACAATCAATTAGCGGATGGTCGTCGCCCTTTCATCAGCTGGATGAATCGATGGC
MDL-10 TTACTTATAACAAATTACCAGATGATCGGCGTATTTTCATCAGTTGGATGAATAATTGGC
MDL-11 TCGTTTACAATCAATTGCCAGATGGTCGTCGAATTTTTATCAGTTGGATGAATAAATGGG
MDL-9 TTACATACAATGAATTACCGGATGGTCGACGTATTTTTATCCATTGGATGGGTCGATGGG
MDL-3 TGATGTTGAACATGATTGGTGATCGCATTCAACTTTCCTCTTTACCTGTTCGTGAATTTA
MDL-12 AATACGCTACACATATGAATTTCAGTATTTGGAATGGACAAGCCGGAATTGCTAGAAAAT
MDL-5 AATATAGTAAGAGTTTCAATTTTACCGCTTGGAATGGACAAATGGGCCTGGCAAGAGAAT
MDL-10 TATACGCTCAACATTTGAATTTTAACGTTTGGAACGGACAAATGGGTCTCGCTCGAGAAT
MDL-11 AATATGCACAGCAATTGAATTTCAACAGATGGAATGGACAAATGGGCATCGCTCAAGAAT

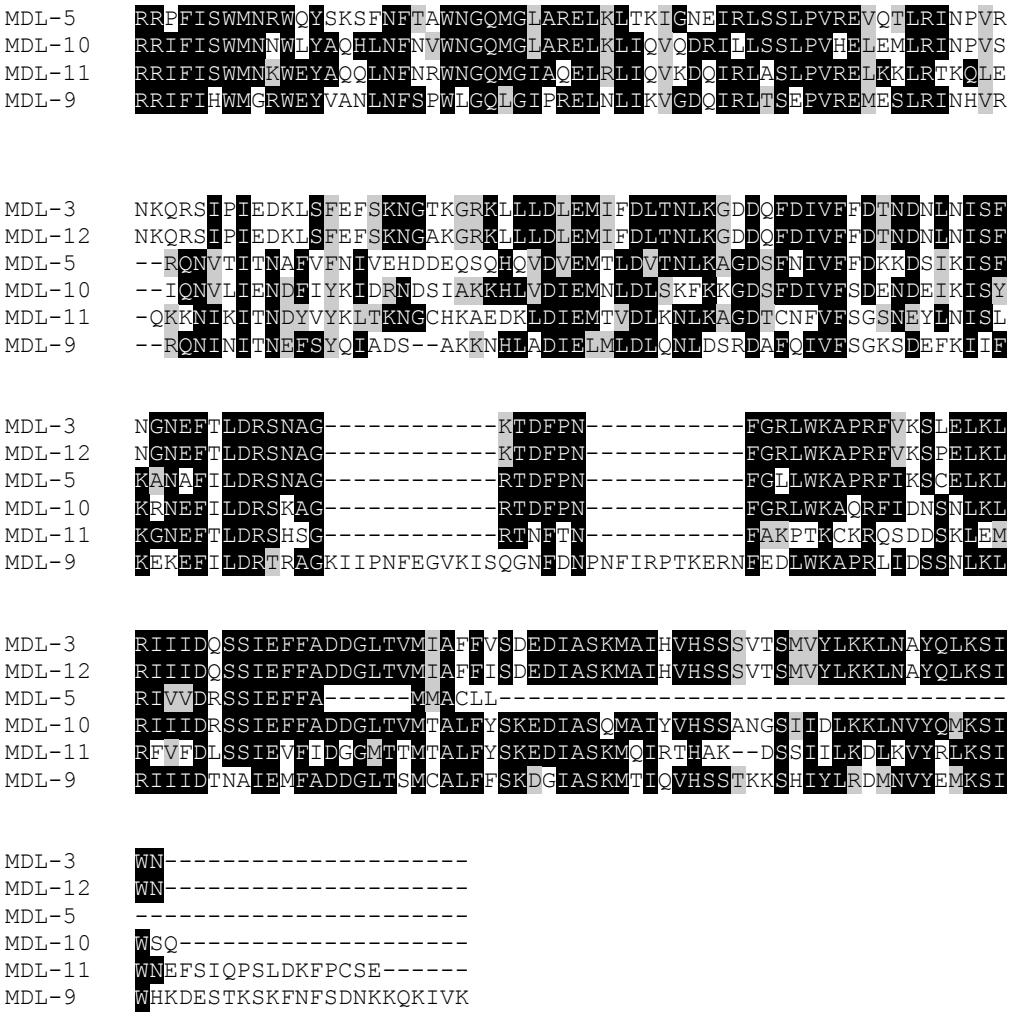

MDL-9 AATATGTTGCAAATTTGAATTTCAGCCCTTGGCTTGGCCAATTGGGCATTCCCAGGGAAT
MDL-3 AATACGCTACACATATTGAATTTCAGTATTTGGAATGGACAAGCCGGAATTGCTAGAAAAT
MDL-12 TGATGTTGAACATGATTGGTGATCGCATTCAACTTTCCTCTTTACCTGTTCGTGAATTTA
MDL-5 TAAAATTGACCAAAATTGGAAATGAAATTCGACTTTCGTCATTACCTGTTCGTGAAGTAC
MDL-10 TAAAATTGATTCAAGTGCAAGATCGAATTCTGCTCTCATCTTTGCCTGTTCATGAATTGG
MDL-11 TGAGATTGATACAGGTAAAAGATCAAATTCGATTGGCATCATTGCCTGTTCGTGAGCTGA
MDL-9 TAAATTTGATAAAAGTAGGTGATCAAATTCGGCTCACATCTGAACCTGTTCGTGAAATGG
MDL-3 AATCTTTGCGAATCAAACAA-CTCGCCAATAAACAAAGAAGTATACCGATTGAGGACAAA
MDL-12 AATCTTTGCGAATCAAACAA-CTCGCCAATAAACAAAGAAGTATACCGATTGAGGACAAG
MDL-5 AAACCTTGAGAATTAATCCGGTTCGGAGACAAAATGTTACGATCACC-AATGCATTCGTA
MDL-10 AAATGTTGCGAATCAATCCAGTT-AGCATTCAAAATG----TATTG--ATCGAAATGAC
MDL-11 AAAAATTAAGAACCAAGCAACTTGAGCAAAAGAAAAA----TATTAAAATCACGAATGAC
MDL-9 AATCGTTAAGAATTAATCATGTTCGCAGGCAAAACAT-------TAATATTACGAATGAA
MDL-3 CTTAGTTTTGAATTTTCTAAAAATGGCACTAAAGGAAGGAAACTTTGCTAGACTTAGAA
MDL-12 CTCAGTTTTGAATTTTCTAAAAATGGCGCTAAAGGAAGGAAACTTTGTTGGACTTGGAA
MDL-5 TTCAATATTGTCG------AACATGATGATGAACAATCCCAACATCAAGTGGATGTTGAA
MDL-10 TTTATTTATAAAATTGATAGGAATGATAGCATCGCAAAGAAACATTTAGTAGACATAGAA
MDL-11 TACGTTTATAAATTAACTAAGAATGGATGTCATAAAGCAGAAGATAAATTGGACATTGAA
MDL-9 TTTAGTTATCAAATCG------CTGATAGTGCAAAAAAAAATCATTTAGCGGATATTGAA
MDL-3 ATGATTTTCGATTTAACGAATCTAAAAGGAGACGACCAATTTGATATTGTATTCTTTGAC
MDL-12 ATGATTTTCGATTTAACGAATTTAAAAGGAGACGACCAATTCGACATTGTATTCTTTGAC
MDL-5 ATGACACTCGATGTGACGAATCTTAAAGCAGGCGACTCATTTAATATCGTATTTTTTGAC
MDL-10 ATGAACTTAGATCTATCGAAGTTCAAGAAAGGCGACTCATTTGATATTGTATTTTCTGAT
MDL-11 ATGACGGTCGACCTAAAGAATCTCAAGGCCGGCGACACATGCAATTTTGTGTTTTCTGGT
MDL-9 CTGATGCTGGATTTGCAAAATCTTGATTCACGTGACGCATTTCAGATTGTGTTTTCAGGC
MDL-3 ACTAATGACAATTTGAACATTTCATTCAATGGAAATGAATTCACTTTAGATCGATCGAAT
MDL-12 ACTAATGACAATTTGAACATTTCATTCAATGGAAATGAATTCACTTTAGATCGATCGAAT
MDL-5 AAAAAAGATTCGATAAAAATTTCCTTCAAAGCAAACGCGTTTATTTTGGACCGATCGAAT
MDL-10 GAAAATGATGAAATAAAGATTTCCTACAAACGAAACGAATTTATATTGGATCGATCAAAA
MDL-11 TCAAATGAATATTTGAACATTTCTCTGAAAGGAAATGAATTCACTTTGGATCGATCGCAT
MDL-9 AAAAGTGATGAATTTAAAATTATCTTTAAAGAAAAGGAATTTATTTGGATCGTACACGC
MDL-3 GCAG----------GAAAAA
MDL-12 GCAG----------GAAAAA
MDL-5 GCAG----------GAAGAA
MDL-10 GCAG----------GAAGAA
MDL-11 TCGG----------GAAGAA
MDL-9 GCTGGTAAAATAATTCCAAATTTCGAGGGTGTGAAAATTTCACAAGGCAATTTTGATAAC
MDL-3 CCGATTTCCCGAA----------TTTCGGCGACTTTGGAAAGCTCCACGT
MDL-12 CCGATTTCCCGAA----------TTTCGGCGACTTTGGAAAGCTCCACGT
MDL-5 CTGATTTTCCCAA----------TTTCGGCCTACTTTGGAAGGCGCCAAGG
MDL-10 CAGATTTTCCAAA----------TTTTGGACGGCTATGGAAAGCCCAGAGA
MDL-11 CCAATTTTACAAA----------TTTTGCAAAGCCAACGAAATGCAAAAGG
MDL-9 CCAAATTTCATAAGGCCAACAAAAGAAAGAAATTTCGAAGATCTGTGGAAAGCTCCACGA
MDL-3 TTTGTTAAAAGTCTAGAATTAAAATTAAGAATAATTATTGATCAATCTTCAATTGAATTT
MDL-12 TTTGTTAAAAGTCCAGAATTAAAATTAAGAATAATTATTGATCAATCTTCAATTGAATTT
MDL-5 TTCATAAAAAGTTGTGAATTAAAATTGAGAATTGTTGTCGATAGATCATCAATTGAATTC
MDL-10 TTCATCGATAATTCCAATTTAAAATTGAGAATAATTATCGATCGATCTTCCATTGAATTT
MDL-11 CAAAGTGATGACTCAAAGTTAGAAATGAGATTTGTTTTCGATCTATCATCGATTGAAGTC
MDL-9 TTGATTGATAGCTCTAATTTGAAATTAAGAATTATTATCGATACAAATGCAATTGAAATG
MDL-3 TTTGCCGATGATGGACTTACAGTTATGATAGCGTTTTTCGTTTCCGATGAAGATATTGCT
MDL-12 TTTGCCGATGATGGACTTACAGTTATGATAGCGTTTTTCATTTCCGATGAAGATATTGCT
MDL-5 TTCG-CGATGATGGCCTGTCTGTTATGACTGCACTTTTCTTTTCA-ATGAAGACATCGCA
MDL-10 TTTGCTGATGATGGCTTAACAGTTATGACTGCACTTTTTTTATTCCAAGGAGGATATTGCT
MDL-11 TTCATTGACGGTGGAATGACAACTATGACAGCACTTTTCTATTCCAAAGAAGACATCGCA


MDL-9 TTTGCCGATGATGGATTAACAAGTATGTGCGCCTTTTCTTCTCCAAAGATGGCATTGCA
MDL-3 TCAAAAATGGCCATTCACGTTCATTCATCTTCAGTGACGTCCATGGTTTATTTAAAAAAA
MDL-12 TCAAAAATGGCCATTCACGTTCATTCATCTTCAGTGACGTCCATGGTTTATTTAAAAAAA
MDL-5 TCGAAGATGGCCATTAATGTTTATTCATCCTCAACCAATTCGAGTGTTCATTTAAATCAG
MDL-10 TCGCAAATGGCCATTTATGTTCATTCATCGGCAAACGGCTCAATCATTGATTTGAAGAAA
MDL-11 TCAAAGATG------CAGATTCGTACACATGCAAAGGATTCATCGATTATTTGAAAGAT
MDL-9 TCAAAAATGACCATTCAAGTCCATTCGTCCACCAAAAAATCGCACATTTATTTGAGGGAC
MDL-3 TTAAATGCTTATCAATTGAAAAGTATTTGGAATTAATG---CATTTTGATAATAATAACA
MDL-12 TTAAATGCTTATCAATTGAAAAGTATTTGGAATTAATG---CATTTTGATAATAATAACA
MDL-5 CTGAATGCTTACCAAATGAAAAGTATTTGGAATTAATT---TTTAAGTAAAATATTTGAA
MDL-10 TTAAATGTTTATCAAATGAAAAGTATCTGGAGTCAAT----AATGAAACATGTATGCAAA
MDL-11 TTGAAAGTTTATAGATTGAAAAGTATTTGGAATGAATTCTCCATTCAGCCATCGCTCGAT
MDL-9 ATGAATGTTTATGAAATGAAAAGTATTTGGCACAAGGATGAAAGTACTAAAAGCAAATTC
MDL-3 TG---TACAGCATTACAGT----------AAAATCTCTCTAAAAAAAGTCATCTCCCATT
MDL-12 TG---TACAGCATTACAGT----------AAAGTCTCTCTAAAAAGA-CCATCTCTCATT
MDL-5 TG---TTCTCAATTATATT----------TAAGAAACAATAAATCTA----------T
MDL-10 TATTTATCAAAAATCAAAT----------AATTATTCTAAAACTCAAT----GTAATAAA
MDL-11 AAGTTTCCGTGTAGCGAGTGAAGAGGAAATATTAATCAAAAACTAATTTATCGGCATAGA
MDL-9 AATTTTTCGGATAATAAA-------AACAAAAAATTGTAAAATAAAAAGAAATCAAGAA
MDL-3 TTTGAACCAATTAGCCGGCAACCAAA----------ATTCTCTTT---AT
MDL-12 TTTGAACCAATTAACCGGCAACCAAAGTGATCCTCATAGTGGAACTATTCTCTTT---AT
MDL-5 GTTATTCCAATGAAAAAAAAA----------AAGTTCTAC---AC
MDL-10 TTGTGCTCGCAATAAATTTTTCAAATGTT-------TTATGAT-TTGAGATCTTG---CC
MDL-11 AATTTCCCATTTTGAATTGGTAGGAAATG-------CTTTGACGCTTAGTTCTTT---AC
MDL-9 AATGAACTGCTTGCGATGCCCAGATTTGATGTGTGATGGGCTCCTAGTTTTCCATTCAAC
MDL-3 ATTACATAAACAT-TATACACATACTTATATTTTAATAATTTTTTCCACATCCTTTGTTC
MDL-12 ATTACATAAACAT-TATACACATACTTATATTTCATTAATTTTATCCGCATCCTTTGTTC
MDL-5 GTTTCAATAATAT-CTTAT-----CTAAGTTATTATTGACATTTTTCCCATGC---GTTC
MDL-10 GTATCTATTGTTTGAATATTTCTT-TTATAGCGCTAAAAAAATTTATCATTGCAGATTGC
MDL-11 AATCCTTTATACTTGATGTTTCATGCTATTTGGTTCGCCAAAAATAAATATGAAATAATA
MDL-9 GAATAAAAGTCTTCGAAATCATTTTTGTTTGTGTTTCATTTGAATAGCTCCCGATTCTT
MDL-3 ATCGTATTCATTTATTAAT-------AAGATCAGCTGTAAAACACCAAATCGTACGCAAC
MDL-12 ATCGTATTCATTTATTAAT-------AAGATCAGCTGTAAAACACCAAATCGTAGGCAAC
MDL-5 ATTG-----AAATGTTACT-------GGTTACA-CGCTAATAAATCAAAACAGAC-----
MDL-10 ATAACAATATCATCAATAA--G-CCGGCGCGTGATTGTCAAAAACGAAAGCCTAAAATAT
MDL-11 ATAATAATATCGTAATTTATTG-CGGGTTAGTCTTTATTCATAAACATATCATTTTAATT
MDL-9 ATTACTGTAGTTTCATTTACCAGAAATATAGCATCATATTTTCTACAATATTTTCTAACC
MDL-3 AATAGAATTCCAAAT-TTGAACCTTATTTCATAATACTCTTATGAGCT---TTATGTG-T
MDL-12 AACAGAATTCCAAAT-TTGAACCTTATTTCATAATTCTTTTATTATTTGTATTATTTG-T
MDL-5 ----------
MDL-10 TTTCTATTTGAAATT-GTAAACAGTAGATGATTACGCGATCCTT----ATATTATG----
MDL-11 ACATCGTTCAATGAA-CTGTACGG-AGAATATTACGCTTGACATGGCATTAAAATGGTTT
MDL-9 CATGGATTTAATTCAGCTGAAAAATATTGAATAAGTACGAAATAAATGATTTAGCTACT
MDL-3 ATGGCGAATTTTTTTCCATTCTCTTTCTACTAGTTA----------
MDL-12 ATGGCGAATTTTTT----------
MDL-5 ----------
MDL-10 -------CATGGGTCG--GAAGCGGGCACTGAGTTGTC----------
MDL-11 CCAGCACAATGATTCGATGTAGCGGGAATTTAAATAGCGTTC----------
MDL-9 TCGATGTACTTTGGCTTCGATTCGGATTCGCTATAGCAATTCGAGTCACAATAT

Supplementary data Figure S2.

Protein alignment




MDL-5
MDL-10
MDL-11
MDL-9
MDL-3
MDL-12
MDL-5
MDL-10
MDL-11
MDL-9
MDL-3
MDL-12
MDL-5
MDL-10
MDL-11
MDL-9
MDL-3
MDL-12
MDL-5
MDL-10
MDL-11
MDL-9

Supplementary data Table S1. Primers used for semi-quantitative PCR analyses

	Sense Primer	Anti-sense Primer	Amplicon length (bp)	Tm (°C)
MDL-1	5’ -TGCTGGAATCACATACAATCA	5’ -TCACGAACAGGGTAGGAAA	183	13.9/-13.6
MDL-2	5’ -CGCTTTCTCCACCCACTGA	5’ -TTGCTGTATGCTATCCATTGTT	166	20.1/40.6
MDL-3	5’ -CGATGGGAATACGCTACACATAT	5’ -CGAACAGGTAAAGAGGAAAG	119	20.4/0.0
MDL-4	5’ -TCCACCGAAGAACAAGAACA	5’ -AACATATTGACCAGTTTCCCATTC	171	-24.1/21.4
MDL-5	5’ -CCGCTTACAATAACGACAC	5’ -GCTACAACCTGAACATAATGAT	187	31.3/21.4
MDL-9	5’ -TTAYCGRCCACAATTWCATT	5’ -TATTSAATGCCCCCRAYGCA	148	-6.8/21.3
MDL-10	5’ -ATCAGGATGGATWAATGATC	5’ -GGAAGTGTTKTCCAATGAAC	150	22.2/21.1
MDL-11	5’ -GGAAGCTGGTGAGCGAAT	5’ -CTCCATTGATTGTTGCGTTG	90	39.0/18.3
MDL-13	5’ -CATCTTAGTCCTCCAAATGG	5’ -AGAAGACGTGATATACTCCAT	76	27.1/19.6
MDL-14	5’ -TGGAACGGACAAGTTGGA	5’ -CTTCACGAACAGGGTAGGA	88	29.9/1.5