

Relationship of Secreted Salivary Protein Variants to Virulence in Hessian Fly (*Mayetiola destructor* (Say))

Yan M. Crane^{1,2}, Charles F. Crane^{1,3}, Brandon J. Schemerhorn^{1,2*}

¹USDA-ARS Crop Production and Pest Control Research Unit, West Lafayette, IN, USA ²Department of Entomology, Purdue University, West Lafayette, IN, USA ³Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN, USA Email: *brandi.schemerhorn@usda.gov

How to cite this paper: Crane, Y.M., Crane, C.F. and Schemerhorn, B.J. (2020) Relationship of Secreted Salivary Protein Variants to Virulence in Hessian Fly (*Mayetiola destructor* (Say)). *Advances in Entomology*, **8**, 15-33.

https://doi.org/10.4236/ae.2020.81002

Received: September 28, 2019 Accepted: November 2, 2019 Published: November 5, 2019

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

http://creativecommons.org/licenses/by/4.0/

Abstract

Salivary proteins are the initial contact between sedentary insect pests and their host plants. It is expected that one or more salivary proteins mediate the interaction between Hessian fly and wheat, in which a feeding site is established to the benefit of the fly. A survey of 52 loci annotated as insect secreted salivary proteins was conducted in 384 individuals evenly distributed among eight biotypes of Hessian fly (B, C, D, E, GP, L, O, and vH9). Amplicons were sequenced with Illumina, and sequence reads were aligned to the reference sequences from which primers had been designed. Positions of consistent base variation (998 in all) were identified and tabulated by biotype. No varying position was associated with biotype-wide virulence to any one of wheat resistance genes H3, H5, H6, H7/H8, H9, H11, H13, and H26. The multiplate pooling strategy utilized in this study is an effective, affordable way to reveal the genotype of hundreds of individuals at tens of genetic loci.

Keywords

Hessian Fly, *Mayetiola destructor*, Virulence, Insect-Plant Interaction, Amplicon Sequencing

1. Introduction

As the third largest producer of wheat in the world, the United States exported 27.2 million tonnes of wheat during the 2018-19 growing season [1]. To fight wheat yield loss due to Hessian fly infestation, it is necessary to be able to identify the molecular mechanisms that operate during wheat-Hessian fly interac-

tions. Secreted salivary proteins are the major point of contact between sucking insects and their plant hosts. There is considerable evidence that some secreted salivary proteins trigger defense responses, while others suppress ongoing defense responses [2]. Virulence requires that the host response is sufficiently suppressed to allow the insect to survive and reproduce. In Hessian fly (*Mayetiola destructor* (Say)), eight biotypes have been defined by their pattern of virulence against a battery of wheat host resistance genes [3] [4] [5]. Here we report an investigation of the distribution of polymorphic nucleotides for 52 secreted salivary proteins in samples of eight biotypes that had been maintained for many years in laboratory culture, since these sites would be potential biomarkers for virulence. We looked for correlation between specific polymorphic sites and biotype, *i.e.*, virulence to a particular set of wheat R genes.

2. Materials and Methods

2.1. Source of Sequence

Translated dipteran nucleotide sequence in GenBank nt was screened for signal peptides with SignalP [6], and 52 such sequences (Supplementary **Table 1**) were chosen for primer construction and amplification. Each 20 μ l PCR was performed by using 10 μ l MyFi Mix from Bioline (Taunton, MA USA), 1 μ l of sample DNA in water, and 1 μ l of each primer for a final primer concentration of 0.5 μ M. Initial denaturation was at 95°C for 2 min followed by 35 cycles of denaturation at 95°C for 30 sec, annealing at 55°C for 30 sec, and extension at 72°C for 1 min. The PCR finished with extension at 72°C for 5 min and holding at 4°C. A multi-plate multiplexing strategy was applied to minimize sequencing cost (**Figure 1**). The four 96-well plates were sequenced as 2x250 base-pair paired-end Illumina reads in samples of 48 individuals from each of Hessian fly biotypes B, C, D, E, GP, L, O and vH9 [7].

2.2. SNP or Indel Discovery

The PCR multiplexing strategy appears as the list of barcodes in Figure 1. Reads

	H3	H5	H6	H7/8	H9	H11	H13	H26
В	+	-	-	+	-	-	-	+
С	_	-	+	+	+	-	-	+
D	+	-	+	+	-	-	-	+
Е	+	-	-	-	-	-	+	+
GP	_	-	-	-	-	-	-	-
L	+	+	+	+	-	MS	-	-
О	+	+	+	-	-	+	-	+
vH9	+	MS	+	+	+	MS	-	MS

Table 1. Hessian fly biotype virulence to wheat R genes.

+virulent, >90% survival on wheat carrying this R gene; –avirulent, <10% survival on wheat carrying this R gene; MS, missing datum, *i.e.*, ambiguous virulence consistent with individual polymorphism.



Figure 1. Amplicon construction for the 52 loci.

were grouped by primer pair, and each group was aligned and displayed in the Integrative Genomics Viewer [8]. Single nucleotide variants (SNPs) were detected by visual inspection. Coordinates of each combination of locus (primer pair) and variable site were saved to a file for each biotype from the Integrative Genomics Viewer.

2.3. Tabulation and Display of Variable Sites

The number of variable sites shared by each possible combination of biotypes was calculated from the coordinates files using the Bioinformatics and Evolutionary Genomics website of the Vib-Ugent Center for Plant Systems Biology of Ghent University [9]. Combinations were enumerated three ways, for all eight biotypes B through vH9 listed above, for the seven biotypes without vH9, and for the six biotypes without vH9 and L, to investigate the sharing of variable sites among biotypes to the exclusion of other biotypes (**Table 1**). The combination counts were presented as partial or complete Venn diagrams of greatly differing appearance, because the problem of drawing such high-order Venn diagrams has not yet been solved completely. The eight-biotype diagram only showed the numbers of variable sites within each biotype and shared in 12 of 28 possible pairs of biotypes (**Figure 2**); it was produced using the Venn Diagram Maker Online [10], and the size of each overlap in the diagram reflected the number of shared variable sites. The complete seven-biotype diagram (**Figure 3**) was drawn in the Adelaide format using R package venn [11]. The complete six-biotype diagram (**Figure 4**) was drawn with tools in the same Vib-Ugent website [9] that counted the combinations.



Figure 2. Partial Venn diagram of the number of variable sites in common in pairs of biotypes.



Figure 3. Numbers of variable sites in common in the biotype combinations that exclude vH9, because the drawing software cannot display a Venn diagram of eight entities with all possible intersections.



Figure 4. Numbers of variable sites in common in all biotype combinations that exclude L and vH9.

3. Results and Discussion

Illumina sequencing produced 27,247,176 merged read pairs that passed quality control and were greater than 150 bases in length. The counts of read pairs ranged from 79,684 for S42 to 1,846,094 for S31 and from 1,148,582 for biotype C to 2,808,266 for biotype D. Counts of read pairs ranged from 261 to 123,110 for individual flies over all loci.

The biotypes display distinct combinations of virulence and avirulence to alleles at eight wheat R genes as shown in **Table 1** [12]. Thus biotypes B, D, E, L, O, and vH9 all survive on H3 wheat, but they share no polymorphic sites in our set of 52 secreted salivary proteins to the exclusion of C and GP (**Table 2**). Likewise biotypes C, D, L, O, and vH9 overcome H6 but share no polymorphic sites to the exclusion of B, E, and GP; B, C, D, L, and vH9 overcome H7/H8 and share no polymorphic sites to the exclusion of E, GP, and O; and C and vH9 overcome H9 but share no polymorphic sites to the exclusion of B, D, E, GP, L, and O (**Table 2**). The relative sizes of circles and overlaps in **Figure 2** show the sharing or lack thereof of 105 positions for single biotypes and 256 positions for pairs of biotypes, but it does not show the 637 polymorphic positions shared among three or more biotypes. The pairs C-GP (148 exclusively shared sites, **Table 2**), E-L (39 sites), E-O (20 sites), and GP-L (16 sites) have the most sharing to the exclusion of other relationships.

Among the tested biotypes, only E overcomes H13 (**Table 2**). It uniquely harbors 22 variable positions, but 13 sites of substitution and one site of deletion reside in one acetylcholinesterase locus, which matches GenBank accession

		Gp	L	0	vH9	GpL	GpO	GpvH9	LO	LvH9	OvH9	GpLO	GpLvH9	GpOvH9	LOvH9	GpLOvH9
		25	15	13	2	16	3	0	2	1	1	1	1	0	0	1
В	3	0	0	0	0	2	1	0	0	0	0	3	0	0	0	0
С	18	148	2	1	0	10	0	0	0	2	1	2	1	1	0	0
D	7	4	1	4	2	1	3	0	1	1	0	0	1	1	1	0
Е	22	4	39	20	1	3	9	0	3	15	0	4	0	0	1	0
BC	0	3	0	0	0	2	4	0	0	0	0	3	0	0	0	0
BD	1	2	3	0	0	0	1	0	1	0	0	5	0	0	0	4
BE	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
CD	1	4	0	0	0	1	1	1	0	0	0	0	2	1	0	1
CE	3	20	5	6	0	43	2	6	0	2	1	21	23	0	0	0
DE	2	4	2	4	0	3	0	0	0	13	1	2	0	2	0	3
BCD	0	0	0	5	0	0	4	1	0	0	0	19	3	2	0	9
BCE	0	0	0	0	0	0	1	0	0	0	0	6	0	0	0	7
BDE	0	0	1	1	0	1	0	0	1	0	7	2	0	0	0	2
CDE	2	1	1	0	0	10	2	1	1	4	0	4	41	0	0	7
BCDE	0	0	0	1	0	1	2	0	3	0	0	15	0	6	0	189

Table 2. Numbers of variable sites in common only among given combinations of eight biotypes.

For example, biotypes C, E, GP, and L have 43 variant sites in common among the 52 amplicons, and variability at these sites is not shared with any other biotype.

ATY49611.1. However, the vH13 gene has previously been mapped and characterized (GenBank accession AEG42082.1) [13], so the distinct variant positions in the acetylcholinesterase of biotype E might be merely coincidental.

Biotypes L and O overcome H5, but the response of vH9 to H5 is not known with certainty. Biotypes L and O uniquely share two variable sites, and the trio of L, O, and vH9 shares no variable sites (**Table 2**). The variants in L and O occur at base 40 in locus 78 (a peroxidase; **Figure 5(a)**) and base 236 in locus 144 (lysosomal acid phosphatase; **Figure 5(b)**), but both occur in only 2% - 5% of individuals and therefore cannot account for the overall virulence of L and O toward H5. Biotypes B, C, D, E, and O overcome H26. These biotypes share polymorphism at base 120 in S34, a sequence similar to Genbank accession XP_020706750.1 (endothelin-converting enzyme; **Figure 5(c)**), again at too low frequency to account for the virulence of these biotypes at large to H26. Only biotype O overcomes H11, but the response of L and vH9 to H11 is not known with certainty; O has 16 apparently unique polymorphic sites.

Figure 3 and **Figure 4** show the distributions of uniquely shared polymorphic sites when vH9 or vH9 and L are not considered. The counts differ somewhat from those in **Table 2**, where all eight biotypes are represented. All three depictions show a large core of 189 to 212 locus-positions that vary in all the represented biotypes and therefore do not appear to be directly involved in virulence. Also, biotype GP uniquely shares more variable positions than any other







Figure 5. Within-biotype genotype frequencies in particular loci that are polymorphic in biotypes virulent to H5 and H26. (a) Genotype frequencies by biotype at nucleotide 40 in locus S78, which matches GenBank ETN66032.1, peroxidase from *Anopheles darlingi*; (b) Genotype frequencies by biotype at nucleotide 236 in locus S144, which matches GenBank KYN33573.1, lysosomal acid phosphatase 1, partial, from *Drosophila nasuta*; (c) Genotype frequencies at nucleotide 120 in locus S34, which matches GenBank XP_020706750.1, endothelin-converting enzyme homolog isoform X1 from *Athalia rosae*.

biotype; L is a distant second. Biotype GP is not virulent against any of the listed resistance genes [14] and thus might be the least modified from ancestral populations existing before the deployment of wheat R genes.

A multiplate pooling strategy was cost-effective and yet allowed reads to be traced to individual fly. This strategy was successfully applied to a panel of 52 genes that encode signal peptides and whose protein products are expectedly secreted. Nevertheless, no frequent nucleotide variant at any particular polymorphic site was limited to biotypes that shared a common virulence toward any of eight wheat resistance genes, and thus none of the 52 queried genes seems to cause virulence.

A more sensitive method to detect gene involvement would be to look at alleles as whole-amplicon variants, but even this is not free of errors. One attempt was to 1) map all the reads to one of the 52 genes, 2) assemble those reads that mapped to any one locus with SPAdes [15], 3) determine individual fly genotype by blastn [16] of the reads against the assembly, and 4) look for high-frequency variant alleles shared by biotypes virulent to a particular resistance gene. This attempt was complicated by the large number of spurious "alleles" arising from assembled sequencing errors, such that the actual alleles comprised less than 5% of the total assembly and accounted for less than 70% of the read counts. Nevertheless, once optimized this method might (or might not) reveal a relationship of a particular multiSNP allele to virulence to a particular resistance gene. A more robust method is linkage mapping with markers traceable to regions in the Hessian fly genome assembly, as performed by Zhao *et al.* [17] with microsatellites to identify vH24.

Acknowledgements

This work was supported by USDA-ARS Research Project Plan 5020-22000-018-00D. We thank Sue Cambron and Jill Nemacheck for helpful advice regarding the virulence profile of biotypes and Lucy Springmeyer for assisting with organizing the data from this work. Mention of a commercial or proprietary product does not constitute an endorsement by the U.S. Federal Government.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Statista (2019) U.S. Imports and Exports of Wheat from 2000/01 to 2018/19 (in Million Metric Tons).
 https://www.statista.com/statistics/237902/us-wheat-imports-and-exports-since-200
- [2] Chaudhary, R., Peng, H.-C., He, J., MacWilliams, J., Teixeira, M., Tsuchiya, T., Chesnais, Q., Mudgett, M.B. and Kaloshian, I. (2018) Aphid Effector Me10 Interacts with Tomato TFT7, a 14-3-3 Isoform Involved in Aphid Resistance. *New Phytolo-*

gist, 221, 1518-1528. https://doi.org/10.1111/nph.15475

- [3] Cartwright, W.B. and Noble, W.B. (1947) Studies on Biological Races of the Hessian Fly. *Journal of Agricultural Research*, **75**, 147-153.
- Kudagamage, C., Foster, J.E., Taylor, P.L. and Chen, B.H. (1990) Biotypes of the Hessian Fly (Diptera: Cecidomyiidae) Identified in the Southeastern United States. *Journal of Entomological Science*, 25, 575-580. https://doi.org/10.18474/0749-8004-25.4.575
- [5] Ratcliffe, R.H., Safranski, G.G., Patterson, F.L., Ohm, H.W. and Taylor, P.L. (1994) Biotype Status of Hessian Fly (Diptera: Cecidomyiidae) Populations from the Eastern United States and Their Response to 14 Hessian Fly Resistance Genes. *Journal* of Economic Entomology, 87, 1113-1121. <u>https://doi.org/10.1093/jee/87.4.1113</u>
- [6] Nielsen, H. (2017) Predicting Secretory Proteins with SignalP. Methods in Molecular Biology, 1611, 59-73. <u>https://doi.org/10.1007/978-1-4939-7015-5_6</u>
- [7] Crane, Y.M., Crane, C.F., SanMiguel, P. and Schemerhorn, B.J. (2019) A Multiplexing Strategy for Mesoscale Targeted Sequencing of Populations. *Academia Journal* of Agricultural Research (Accepted).
- [8] Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P. (2011) Integrative Genomics Viewer. *Nature Biotechnology*, 29, 24-26. <u>https://doi.org/10.1038/nbt.1754</u>
- [9] Vib-Ugent Center for Plant Systems Biology (2019) Calculate and Draw Custom Venn Diagrams. <u>http://bioinformatics.psb.ugent.be/webtools/Venn</u>
- [10] Meta-Chart (2019) Venn Diagram Maker Online. https://www.meta-chart.com/venn
- [11] Dusa, A. (2018) R Package Venn. https://CRAN.R-project.org/package=venn
- [12] Schemerhorn, B.J., Crane, Y.M., Cambron, S.E., Crane, C.F. and Shukle, R.H. (2015) Use of Microsatellite and SNP Markers for Biotype Characterization in Hessian Fly. *Journal of Insect Science*, **15**, 158. <u>https://doi.org/10.1093/jisesa/iev138</u>
- [13] Aggarwal, R., Subramanyam, S., Zhao, C., Chen, M.-S., Harris, M.O. and Stuart, J.J. (2014) Avirulence Effector Discovery in a Plant Galling and Plant Parasitic Arthropod, the Hessian Fly (*Mayetiola destructor*). *PLoS ONE*, 9, e100958. https://doi.org/10.1371/journal.pone.0100958
- [14] Onstad, D.W. and Knolhoff, L. (2014) Arthropod Resistance to Crops. In: Onstad, D.W., Ed., *Insect Resistance Management: Biology, Economics and Prediction*, Academic Press, London, 2nd Edition, 293-326. https://doi.org/10.1016/B978-0-12-396955-2.00009-6
- [15] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A. and Pevzner, P.A. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, **19**, 455-477. https://doi.org/10.1089/cmb.2012.0021
- [16] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: Architecture and Applications. *BMC Bioinformatics*, 10, 421. <u>https://doi.org/10.1186/1471-2105-10-421</u>
- [17] Zhao, C., Shukle, R., Navarro-Escalante, L., Chen, M., Richards, S. and Stuart, J.J. (2016) Avirulence Gene Mapping in the Hessian Fly (*Mayetiola destructor*) Reveals a Protein Phosphatase 2C Effector Gene Family. *Journal of Insect Physiology*, 84, 22-31. <u>https://doi.org/10.1016/j.jinsphys.2015.10.001</u>

Supplementary Table 1 Nucleotide Sequences of the 52 Secreted Salivary Proteins in Fasta Format with Identity, Forward Primer (5'-3'), Reverse Primer (5'-3'), and Genbank Accession in Each Defline

>S10 AAATAATGGCCGCTCATCAC GGCCCAAGGAATGGATTATT FM984359.1 AAATAATGGCCGCTCATCACGATTCACTTTTATACAATCTTCAGCCAAACGTTTCAATGCTTGCGGACAAT CCGAGCGAACTTTTGTCAAATCTGGACGCAAAACACCACGACCAACTTTTAGCAGAATTTCATCTTTGTTA TTTATATGATTATAGGGCAATTGAAGCGTTAACAATTCATACAATACAATTCCGAATGCATAAACATCACT TTGAAATGAATATGGATTTGATTCTCGCATTCGAATTACTTCGGGTGCCATCCAAATTAAACTGCCCGTTG GATATTCCGATTGTTGTTCGCCGGACCATCGTGCTTTTGCCGTTGCAAGACCAAAATCGCCAATTTCACG GACAAATCTTCGTGCAAGAAGATATTATACGATTTGAGATCTCGATGAATAATATTCTTTGCATGTAAATA ATCCATTCCTTGGGCC

>S12 GGTTGCAGACTCCGAAAGAG CCCAAAATCCCCTATCATCC EV465518.1 GGTTGCAGACTCCGAAAGAGGTTAAACCATTTCTTCCGAATGATGTCATGGAAAAAATCCTCCTTAAAAACT GATTTCAAGACATTACAATCGGCCATAAGAGCTTTTGCAGTACCTCCGGATACGATCGTTATGACCCAAA TATTACAAGATCACATAATTGAAATTGTGCCACTTAAAAAGGGTACTGATATCATGAAGTTGAATCCAA CGGCGAATTATTCCTTCAAATCACTAACTATAAACAACTTGAAAATTTCTTCAAAAAATTTGGTGACCGAG TGAAGAAAATTGCCATTGATTACAGACCATTAATTGAGGATGATCAATACAAAAAATTACGAATATT >S122 TCATCGAGTGCTTTCATTGC CAGGCGAACGTAAAGATGGT DQ440437.1 TCATCGAGTGCTTTCATTGCTTGAGCGAAATGATGACCATGATCGATTCGTCCACCTGCGAAATAATTTGA AGTCATCAATATTCAACGATGGTTTAAATGGTTACGAAACGATTATTACATACCTTCGACAAATAAGAA AAAGCCATTTTCATCTTTTTGTAATATTTTGATTGCGGCCTCAGTCATTTCGACGAGATTTGGTTGCTTCAT TCGGTCTGCGTCTAAGTTGAAATCCATGTGACCGGACGAAAAAAGTCCCAACACTTGATCCAATTCTC AAATTTAAATTATCTAATTCTTCTTTGCGTGATATAAATTGCCTTTTGGTTTATCGCCCAACCATTCTTCA

>S144 TCGCAAAAGAAAGCCAGATT CTGGCAACCCATACCAATTC CK810674.1 TCGCAAAAGAAAGCCAGATTTTAGTCGTGACATTTCACGAGTTCCGATCAGTCTTTTTCTCGATTTCGA CGAGCTCTGTGAGCTGATCATAAATATTTGACGCCCATTCTGGAATACTAAACATTTTGGTTAGATTTTA TCTTTATGAAAATGTAACACAAAGCTCAAAACATGTTATTTACGTTCTATTGTGGAGCTGTTGAACCATGA GCGTATCATAAATGCAAATGATATCTCTGAAATTTTCCATTGAAAGAGCACTTTTTTCCAACAAGGAATACG TAAAGCCATTCATATTTTAAACGCCAAGCACTAATTCAGGCTTATTAAAATGTTTTCCAAAAGGAATTCA AATCGATCGCATCGTTTACCGCCAGCCAAAACATATCCAGTTTTAGACGGAATAGTATGAATTGGTATGG GTTGCCAG

>S148 CTGGTGGCTTAGCTTCAAGG ATTCCGATGTTTTTGTTGCC FD93207.1 CTGGTGGCTTAGCTTCAAGGGCGCAACTTGATCGGATGCATTGAAACCAAATGATTCAATGATACTGCG AGCCTTTTCTACTAATTCGTGCCTTAAGTTGGGTGGTGTGTCACGATAATGGAATGTTAATAGTGGACCTT TGTTTTCGACCCAGGCACCGTCACGGCATATCTGTTCAGGAAAACAATGCCATTTTAATAGTGGACCTT CAAATTGAAATCGATGAGGTTGATCAAGTACTTACCGATTCCTGTAGCGATTTAAGCAAATCGCTAACTTT ATCTTCATATTGACTGGGAAGTGGATGAACGAATTTACTACCATCCGAATGTAAAATCTCTAAGCCATGG TTACCGGCATAGGTGATACCTTCAATTCCGACCATTTTAACATTTCGCCGGATATAAC GGCAACAAAACATCGGAAT

>\$15 TGAATGGATAAATGGTGCGA ACCGTGACCGAGTGGATTAG EZ422485.1 TGAATGGATAAATGGTGCGAGTACGTATAACTATTTTTCAGCCAAAAGTCGTGTCGGACCGGTGAGAAAT >\$18 CTTTCTTCAGTGCGTGGACA CCGAATTGGGTTGTTATTCG HS094723.1 CTTTCTTCAGTGCGTGGACACTGAAATATGAACCGTTCTAGAATAGCCCAACTCATTGAAATGACAA CGAAATACCTGCAAAACATCATCCTTCAGCATTTTGAATATGGCATCGCACACTACTGGAATAATTTTCGA TATTGTCGACTGCGGTATACGGAACAGATATTGCAATGACATGAAGCTGTCGCCAGTAGCCAGAAAACGC AATGTCAACGCCAAACGTTCACCTGGTGACACAGCATAGCGCATTGATGTGTCCTCTTTTTTTATGAGCGG CGACACCGAATTCAACAGATTATTAAAATCCACAATATTCATACGAAGGAAATTCTTGTATAAATACGGC ACCTCTTGTTCTAGCTCAACAAGCAATTGCGAATAACAACCCAATTCGG

>S19 GTACATCATGCAGCATTCGG TTCAGGCATGCCAAATGTAA EW780297.1 GTACATCATGCAGCATTCGGCGATGAGCCGGCAGTCATTGAGCAATTGTCGAAGGCTGGCGGTGACTTGA ATGCCCGAAACAAACGACGTCAGACAGCATTGCATATCGCTATCAATAAAGGCCATTTAAACGTTGTAAA AACATTACTCGAACTTAATTGTCACCCAAGTTTACAAGATTCAGAAGGCGATACACCACTCCATGATGCT ATATCCAAAGAACAAGATGAAATGCTTTCGCTGCTACTCGATTTTGGTGCTGATATAACGCTAACAAATA ATAATGGTTTCAATGCTTTACATCATTGCGCCCTTAAGGGAAATCCCAGTGCAATGAAGATTTTGTTATCG AAAACGAATCGGCCCTGGATTGTTGAAGAAAAAAAGATGATGGATATACAGCATTACATTGGCATGCCT GAA

GCATAATGCCAGGACTCCGCTTGACATCTACTGGTAGATATACATTACCGCCATCTAACCCCCCTAACCA ACCATGAATAATAAAGTTTACAGGTTTTTCTAAATTTAGTTTAGAATGGATATCTGGATCATTGAAGTAGA TTTGATGCAGTTCTTTTTTTTGCTACAGAAAATTTAAACAAATATTCGATTCATCGTTAAAATACATAGT TATTACTTCAAAAAGTATTCGCAAGAAAATATTTCACAATGTACTTACGGACGTACACACCAAAAAG

>S28 TGTTTGATTGCGTTTTTCCA TTCCAGAATGCTTACACGGA FM961982.1 TGTTTGATTGCGTTTTTCCAACTCGAACTGCGGTATGAAGCCAAAGTACCGATTGATGAAATTGAAATTCAAT AGAAATCAATACTCAAATCTATTCTTTTAGCGGTTTGGAGTGGCACTTGTACGTGCAGGCCAACTCAAGC TGAATCAATCACACTTCAAGAATGATCTAAGGCCCATTGATGAAACCTGTGATTGCAGGTACATGCAAACA TTATTCACGGGCTTATTTGCACCATATAGTTCGAGTTGAAGCTGTAAGGTGGCAAGTTTGCTAACTGTTCACA ACGTTGCTTATCAGGTGAATTAATATCATTTATTTCGTTAAATTAGTGTACTATTTACCTGAATTTTTGTT

>S3 CTTATCGCGGCACAATTCTT GGATCGCAAACACTTTTGGT AY752844.1 CTTATCGCGGCACAATTCTTCGCCAGCCATTTTTATAAAGGCATTAGATGGTGTAATATCATAAATACCCA GCCAGGATGATATTGTCTATAAAATATGTAAGATTAAATTATTCAACACATATATGCAACATGCAATTTT CCGATCAACATACCCCCAACGCCGAAACGAATGGCGCGATCAAACGAATTGGTGGACTTTCAACGTGCGA CATAAACGCAACTGGAGCGAGTGCATGCATACATGTTGACCTTGGAAAGATATGCAGGTTTTTGAGAGCTC ATCACATAAAAGCGGTGGTTCCTTGTGAATGACCGATGAACTGTAGTTGCGATACACGAGTTTGGTTCA AAATATAATCGATCGTTGCTGGCAAATCGTAATACCAATTGATGCCAGGAAATGACCAAAAGTGTTT GCGATCC >S30 GCAAACCTGTATGCCCAACT GCGAAAGTTTTATGGGCAAA EE571743.1 GCAAACCTGTATGCCCAACTGGACCTGCTAACGGATTACATGACCAAACGACAGCAAATGAGTCGTAATC CGTATCCAAAACATGCATTGTTGTGTCATAATTCACCGGCAATCCAGTATATTTGACCGTAAATGTTGCTT CATCATTTTAGCCGTCACATTCATAACACCCGAAACAATACGCTGAACATTTGTACTATAACAAAATGAA AGGAAAAATGATTTCAAATTCAAAAGCAACAACAATAGTTTTTGGTGTACATTATTGGCAAGGAAAAAAA ATATCTTACAATCGATTGGTCATTTCGTTATTAACATAGACTTTGCCATCTGCACGTCTTTCATATGTTACC ATAACACAACGTGATGCCAATTCAGATACAGCAAAAATAGCGTTCTGCTTCGTACCATTTGCCCATAAAAC TTTCGC

>S31 ATAACCAACGCCAAATCCAG TAGGTAGAATTGGCGGAGCA AY752808.1 ATAACCAACGCCAAATCCAGATGTTGATCCTTTTTGGGCTTTAACAAAATTCGATTCAAATCCTCGTTGAA TATTTTGATAGTTGATATTTATGTATTGATCGCGTTCTTCGCGGTTTTGCTCATGTAAAAATCCGGCTGAA TGCATAAATTCGTGTAATACAGTTCCCAATAATGTTGTACAAGCAGGCGATTGTAAAATTGACTGTTTGTGC TCCGCCAATTCTACCTA

>S35 TAATTTTCTGGGCCTCGTTG TATGCCCAATATGCATCGAA EV466763.1 TAATTTTCTGGGCCTCGTTGCAAAATGGTAGGTCACAGATTTCACATTCACAATATTTGACACCCAACAAT TTGAGCATGTCCATTTCCTTCTGTAATTCCTCTAATTTCTTCCTCACTTCTATCATTTCGTCTCCGGATTGG TTATCGACAACTATAATATTGGTGCTATATAGTTTTCGGATATCAGCGAACACTGCTCGAGCCTTACAAGT TGGACAGCATTGTAAGCGAACTGGGCAAACCGAAAGCCAACTGAGAGAAGAAGAACATGAATTTTATAA AGGTGTCGAACCAATCAAATACATACAACATACTTCCTTACATGGAAACCAGGAATCTCCAAAATATGTCCGCACT

>\$36 ACGGCATGTCACAAATTTCA CACAATCGAAGCGAGTCAAA EV466763.1 ACGGCATGTCACAAATTTCACATTCGCAATACTTGACACCCCAACCTCTTCAGCATATCCATTTCACGTTGT AATTTTTCAAATTTCTCGTTTAGCGTGTGCATTTCTTCTTCGGATTGGCCATCGACAATTAAAATATTGTTA GCATATAGCTTCCGAATTTCACTGGATACTGCACGGGCCTTACAGGTTGGACAGCTTTGTAAGCGGGATG GGCATTCTGACAGCCACCTGAGAGAAAAATATTTTGGATTTTATGATGTGTCAAACCAATCAGTTTGATA CAACATACTTCCTTAAGCAGGAATCACCAAAGAGATGCCCACACTTTGTTGAGACGAGGAAATGGCTACC AGACGATGTACAGGCCTCGTAACAGATTGGACATTTTAATACTTCTAAATCGTCAGTATCTGGTTTGACTC GCTTCGATTGTG

>\$37 CACCGAATTTCGACCAAGAT TGCAGATTCGCAGGTATGAG EV465518.1 CACCGAATTTCGACCAAGATTCATCAATGATATCAACGAAGATCTGCCATTTTTGCCAAATGATTTG ATTGAAATGATTATCAGTAAAACTGATTTCAAAACAATACATGAGGCCGTACAATCACGGATACTACGTA CAACTGACATCGTAGTGAATAGAATTTTACAAGATCACCATAATTATAATTGATAGTCGTAAACGCTTTGAT GTTACTGATATCAATGAATTTCATGCACAAGGCGACTCATACCTGCGAATCTGCA

CAATCCGAGACATACGAAATTGCCAATAACCCAATTTCGTGACGGGCAGCCAAAATTCTGGCTTAACATC ACATCCGCCAATTATCAATTCACCACCAATTTCAGCCGTTTGATTATTCATATGAAAGCAAAAAATCTTTT CATTGATTACACCTTGCTCTTTGAGTCGATCAAGTGCTGTATAATTCCAGGAGGTCCTTGAAATTGTTCCG A

>S48 TCCAAACCGAAGTAAATCGC ATCGGCATTATGTTTGGTGG EE264606.1 TCCAAACCGAAGTAAATCGCATGATGAAATTAATCATCAATTCGTTGTATCGTAACAAAGAAATTTCCTT CGTGAATTGATCTCGAATGCATCGGATGCTTTGGATAAGATTCGTCTGTTGTCATTGACAAAGAAATCGAACGG AAATTGAAACACATCCAGAGATGAAAATTCAAATTAGAGCCGACAAAGATAACAAGGTGTTGCACATTAT CGATACTGGTATTGGTATGACAAGAGAGGATTTGAGCAACAATTTGGGAACCATTGCTAAGTCCGGAACA GCCGATTTTTTGGCAAAACTGCAAGACAGTCCAAAGACAATGAAAATGATATGATTGGTCAATTTGGTG TTGGATTTTATTCGGCATTTTGGTTGCCCGATCGGGTTGTTGTCACCACCACAAACTAATGCCGAT

>S5 AAGGTGAGTTTTGTCAGGCG AGAGCCCAAAACGACAGAGA EV465518.1 AAGGTGAGTTTTGTCAGGCGTGGCATATCAAATGAAAGATTTCCAAATAATTCGTCCATTTGGTC TTCATCCATTTCGTCTTCATCAAACTCAATATAAAAGTGGGTCACTGATGGAAGTTTATATACTGGGTTAT TGTTTTTTCGTAACATAAAGTTTTCACCAGGATTTTCAACTTCAAGTGTTTCTAACTTCGTCAAATATTTTG CAATCGACTGCATTGTCTCTGTCGTTTTGGGCTCT

>S62 AATCAAACATGGCCGGATTA CGCTGCCAGAGAAGAATACC CK810674.1 AATCAAACATGGCCGGATTATTTCCTCCCAAAGATGATCAGATATGGAATAAGAATCTAATGTGGCAAGC GATTCCAATACACACAATACCAGAATATATGGACTATATATTGGCTGCGAAAAGATCGTGCGCCCAATAT GAGCATGCTCTTAAGAGGTATCAACAATCGCCTGAATTTAAAGTTCTATTGGACAAATCTAAGGCCCTTGAT GTTTTATTTGGAGCAGAGTACCGGAAAAGCCATTAGAACACTAACAGATGTAGGAAATCTATACAATACC CTATGGATTGAAAATCTCAAGAACAAAACGTAAATATGTCACATGAAATATATCGAGTGAGGTCTGAGT TTATTCTCTATTTTCTTAGTTTGCCAGAGTGGACAAAAAGGTATTCTTCTCTCTGGCAGCG

>S67 GGCTGATGAATCCCAAATGT TTCTTCGGTTTGCATAGGCT EE264606.1 GGCTGATGAATCCCAAATGTATTGTTCATCATCATGTGTGTTTCGCTGTTACGGTAACTTTGTCGGCAACCA AGAAGGCTGAGTGAAAACCAACATCGAATTGATCAATCATCAATATCAGCCCCCGCTTTCAATGTTTC CATAAATGCAATTGTTCTTGATTTAGCAACTGTACCAAGTTTGATGATCATATCGGTTTTAGTCATACCAA TACCAGTATCAATTATGGTAAGTGTACCAGCAGCTTTGTTGGGAATGATTTTAGTACAATTCTTTGCCG CTGTCGAGTTTTGATGGATCAAAGAGTGATTCATATCGGATTTTGTCCAAAGCCTATGCAAACCGAAGAA

>S7 AATGATTCGGCCATCAAATC GACAAGGGCCCACAATCTTA EV465518.1 AATGATTCGGCCATCAAATCGTTCTCCACTAATGGTGAGCTTTTTCAGGTTTGGCATATCAAATGAAAGTC TTCCGATTGCAAATACATCGTCCATTCTACTACCTAAACTAATATCAAAGTGTATCACTGATGGAAGCTTA TATACTTCATTATTGTTTGCTCGTAACAAAACGTTTTCACCAGGAATTCCAATTACAAGTGTTTCTAACTT TGTTAAATATTTTGCAATCGATTGCATTGTATCCGTCCTGTTTGGCAATGCGAGAGAAATTGTTGTCATGT TCCTATTATTTGCGAAAATATCATTAAGATTGTGGGCCCTTGTC

>\$73 CATTTCGCCATTCGAAAACT TGGCTAAGACCACCATAGGC U11052.1 CATTTCGCCATTCGAAAACTAATAATTTTATTGTTTCATTTCGTTCACAGAATGGTATTCGGTAAGCATT ATCACGGTTACTCAAGTTCAATTCGTCCAGGTGTCTTCAATGAAATTGCAACTAGCGCGTTGCCAGCATTT ATGACCATGTTGCCAAATGGAATGAACAATGAATCAACAACAATTGAAACATTAACGACTTCGGCTATTCGGT TGTTAAAAACCTTAAAACCAAACAATTTGGATTTCAATGAATACTGGACACCAATCGCTCTGGCTATTCAT CGTAGCCGTGACCATGGCATTCCAGCCTATCACAAGCATTGAATCTTTGTGAAGAACGCTTACCAAATA CAAATGGACAAAGACTCACATTTGCCGATCTTCAAGCCTATGGTGGTCTTAGCCA

>S74 AAAGAATGTGCCGTCGAAAC TTCATCTTTGTTTGCACCGA EV466783.1 AAAGAATGTGCCGTCGAAACATCATTTACACCAGAAGATCTGACCAAGCTTCGCACCAATCCAAAAGCTG CCGTTGGCGATGCTAAAGCACAAGTAAGCTATTCGCTCATTTGCAGAAATTCAATTTTGAAAGGAACCAC CAAATGTTGTAGACCTTTCATTTGAAATTTGAAACATTGATGTCGCTGGTGATTTCCAAGAACATGTTGCCACAAAATG AGTAAATGTTTCCTCCACAAATTGGGCTTCATCAATGATGCCGGTGATTTCCAAGAACATGTTGTCATCGA AAAATTGTCAAAAGGTGGTGACCGTGCCAAATTGAAGAAATCGTTAAACAATGCAAATCCGTCATCGGT GCAAACAAAGATGAA

>S75 CGGAGCTTGGAAAAAGTGTC TCCCATTGTCTTTGTGCGTA FD933933.1 CGGAGCTTGGAAAAAGTGTCATAATTTCGCCATTTCCAGTGGCAACCTCTCTCACACTTCTGTTACAAGCA ACAAATGGGACAACTTGCACACAGTTGAAAAATTGTTTGCATTTGCAGGACAATAATATTGCGAAAATCAC ATGTGGCCAATCAGTTTCGCAATTATTTTGCATTAATAAAAAGAGTGCTGGACAATCGGAACTGATGAT TGTCAATCAAATCTATGTACAACAAACATTTCGATTGAATAAAGAGTTTCATGAAAATTGCAGCCAATGAA TTTGATGCCGGTGTTGAATCCGTGAATTTTCCAACACAAACGAAACAACAACAACAATTAATAATCAATAATTCAT ACGCACAAAGACAATGGGA