Supplemental Material

The draft genome assembly of *Dermatophagoides pteronyssinus* supports identification of novel allergen isoforms in Dermatophagoides species

Thomas A. Randall[1], Jim Mullikin[2], NISC Comparative Sequencing Program[2],

Geoffrey A. Mueller[1*]

Short Title: Draft Genome of *Dermatophagoides pteronyssinus*.

1 Intramural Program National Institiute of Environmental Health Sciences

2 NIH Intramural Sequencing Center, National Human Genome Research Institute

* Corresponding author

Geoffrey.Mueller@nih.gov

111 T.W. Alexander Drive, MD-MR01

Research Triangle Park, NC, 27709 USA

Methods

DP was cultured at room temperature on Diet A as described [1]. As food became depleted, mites of all active life stages migrated onto the lids, were collected by aspiration onto a 38 mm stainless steel and were killed by freezing. Once food was completely consumed, the spent culture was also killed by freezing. Both mites and spent culture were lyophilized prior to use. Two aqueous mite extracts were prepared from the mite material described above using methods adapted from [1]. Both extracts were prepared in sterile Dulbecco's Phosphate Buffered Saline. One extract was prepared in the presence of HALT Protease Inhibitor Cocktail and EDTA (Thermo Scientific) according to the manufacturer's instructions. The second extract was prepared without these inhibitors. Mites were collected as indicated above from culture lids and surface sterilized as described for scabies mites [2]. Genomic DNA was isolated from mites of all active stages using the Wizard SV genomic DNA purification system. To enable proper digestion of mite tissues using this kit, mites were ground in digestion buffer using a Dounce homogenizer. RNA was isolated from surface-sterilized mites using the Direct-zol RNA MiniPrep w/ TRI-Reagent The sample prep was similar to that for DNA with ~50 mg of live mites being ground in 500 ul of TriReagent on ice.

Genome Assembly

The genome assembled from 303,594 PacBio ccs reads generated by the NIH Intramural Sequencing Center (NISC) using SPAdes 3.8.1 [3]. SPAdes assembles genomes using a range of kmers and the K127 genome was the most complete, as judged by having the fewest scaffolds and the highest N50 and was used as the starting point for the final assembly. From the initial assembly of 2486 scaffolds, those under 300 bp were removed, resulting in a preliminary assembly of 879 scaffolds. Separately, 216,079,207 Illumina 126 bp paired end reads were assembled with Phusion [4], generating an approximately 70 Mb assembly with an N50 of 8.7 kb. BLAST of both assemblies to an NCBI bacterial genome collection showed that the predominant component of the microbiome was found to be *Serratia marcescens*. The genome of *S. marcescens* WW4 (Genbank accession CP003959.1) was found to be the top hit. 55 contigs with an E value of <= 1-30 were removed as potential bacterial contamination to result in a final assembly of 834 contigs with an N50 of 376 kb and a genome size 52.5 kb with a GC content of 29.03 %. Repeat content was determined using misa for microsatellites (default settings) [5] and Repeatmasker [6].Dotplot alignments were performed with Gepard 1.4 (PMID: **17309896).** All raw genomic data used in this study has been submitted to either SRA as Bioproject PRJNA395246.

Protein prediction and comparison

Whole organism RNA-seq data for DP was obtained as described [7]. 317,387,496 300 bp paired-end reads were assembled using Trinity [8] and from this, Transdecoder predicted a set of 25,446 proteins from which a non-redundant set of 14,409 proteins > 100 aa was obtained. To explore the genome for proteins that may have not been identified in our RNA-seq sample, we used two algorithms, Augustus 2.2.5 [9] and SNAP (snap-2013-11-29) [10], to independently predict proteins sets from our genome assembly. Of the available training sets for each, we empirically determined that the jewel wasp *Nasonia vitripennis* training set was the most optimal available for protein prediction. Augustus predicted 9006 proteins; 3536 not in the RNA-seq dataset while SNAP predicted 17,876 proteins; 6480 not in the RNA-seq dataset. Combined, these two algorithms predicted a total of 4959 proteins > 100 aa not previously predicted. The final set contains 19,368 proteins (Supplemental file 1). We also used SNAP (same training set as above) to make protein predictions for the five species in Table 1 for which there were no publicly available predicted proteomes and other *Sarcoptes scabiei* (SS) genomes. We used BUSCO v1.1b1 (ref; using the arthropod geneset) along with the extended CEGMA dataset of 2748 genes to assess the completeness of our genome. For genome-level orthology prediction of proteomes of the six mites and ticks for which we had publicly available protein predictions, we used Orthovenn, with an E value of 1e-05 and an inflation of 1.5 [11]. For orthology prediction of specific classes of gene sets of interest, we used the script orthoparahomlist.pl with a cutoff E = 1e-05 [2].

Results

Mitochondial DNA

The mitochondrial (mt) genome of DP has previously been published, and is approximately 14.2 kb in size [12]. A full length mt genome of 14,680 bp was extracted from the PacBio data, this was 95.17 % identical with, and co-linear with the published DP mt genome (Supplemental Figure 2) and contains the same complement of 13 CDS and 22 tRNAs. The 12S ribosomal RNA in the our mt genome has a four bp deletion and the 16S ribosomal gene has 7 SNPs and a single base pair deletion, both relative to the published genome. The major difference between the two mt genomes is a largely AT rich region of 495 bp (Supplemental Figure 3). This insertion occurs within the D loop located at 11483-11768 of the published mitochondrial genome. The D loop itself is a variable region itself involved in both replication and the transcription of nearby genes [13].

Microbiome

The composition of some of the microbiome could be assessed because DNA was extracted from whole mites. The Phusion genome assembly was examined using a collection of complete bacterial genomes downloaded from NCBI for better quantification. Any contig with an E value of 1e-30 or lower was considered to be bacterial contamination. Supplemental Table 4 and Supplemental Figure 5 shows that the microbiome of DP is dominated by *Serratia marcescens* but a range of Bartonella species were also present. The PacBio assembly was able to generate a nearly complete 5 Mb genome of *S. marcensens*; this was most closely related to *S. marcescens* WW4. The microbiome has a very distinct GC content from the *D. pteronyssinus* genome. For example, the reconstructed *S. marcescens* genome is 59 % GC, consistent with published *S. marcescens* genomes, in contrast to 29 % for the DP genome. For comparison, the DF genome is 29.5 % GC and the SS genome is 33.26 % GC. A dotplot alignment of these is shown in Supplemental Figure 4. In contrast, the microbiome of DF was predominantly Enterobacter species [14]. However, we note that the Wizard SV genomic DNA purification system used for DP was not optimized for lysis of Gram + bacteria, so these results should be not be considered as a direct analysis of the DP microbiome. For more information, a comprehensive review of the microbiomes of 143 mite species was recently published [15]. Consistent with our data, this review shows that the microbiome of mites that have been examined are dominated by gram negative Enterobacteria.

Repeat Content

The repeat content of the DP genome was examined using several algorithms. Misa was previously used to characterize the simple sequence repeat (SSR) content of SS [2,16,17], which, along with DF was shown to have a much elevated level of SSRs (>2400 per Mb) compared to the other sequenced mite genomes. Our analysis of SSRs within DP with Misa supported this as the total number of identified SSRs is 103,106 and we see a level of just under 2000 SSR per Mb supporting the previous suggestion that the Hypoorder of Astigmata is selecting for an elevated level of SSRs. RepeatMasker was used to assess the general repeat content of *D. pteronyssinus*. The results of both these analyses are in Supplemental Table 5.

Transfer RNA

We used tRNAscan-SE (Supplemental Table 5) to identify the tRNAs within our assembly [18]. We found 68 tRNAs, a similar complement to that of DF (65 tRNAs) which we confirmed by re-running the tRNAscan analysis on the DF genome. However, the genomic organization of the tRNAs may be quite different between the two species. In DF, 26 of the tRNAs are located within 185 kb on a single genomic scaffold (KN266225.1) while the tRNA complement of DP is much more widely dispersed; only one scaffold contained five tRNAs. A caveat to this interpretation is that a different fragmentation of our assembly could, in part, cause this difference; however, our genome N50 of 376 kb compares favorably with that of the DF genome (187 kb). The exon/intron structure of the tRNA genes of these species differ also in that 33 of the DF tRNAs, including all 26 on scaffold KN266225.1, are intronless while only 18 of the 68 tRNAs in DP do not contain an intron.

References

1 Avula-Poola S, Morgan MS, Arlian LG: Diet Influences Growth Rates and Allergen and Endotoxin Contents of Cultured Dermatophagoides farinae and Dermatophagoides pteronyssinus House Dust Mites. Int Arch Allergy Imm 2012;159:226-234.

2 Rider SD, Morgan MS, Arlian LG: Draft genome of the scabies mite. Parasite Vector 2015;8:585.

3 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA: SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 2012;19:455-477.

4 Mullikin JC, Ning ZM: The phusion assembler. Genome Res 2003;13:81-90.

5 Thiel T, Michalek W, Varshney RK, Graner A: Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet 2003;106:411-422.

6 Smit AFA, Hubley R, Green P: RepeatMasker,

7 Ogburn RN, Randall TA, Xu Y, Roberts JH, B. M, Karnuta JM, Rider SD, Kissling GE, London RE, Pomes A, Arlian LG, Fitzgerald MC, Mueller GA: Are dust mite allergens more abundant and/or more stable than other Dermatophagoides pteronyssinus proteins? J Allergy Clin Immun 2017

8 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011;29:644-652.

9 Stanke M, Waack S: Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 2003;19:Ii215-Ii225.

10 Korf I: Gene finding in novel genomes. BMC bioinformatics 2004;5

11 Wang Y, Coleman-Derr D, Chen GP, Gu YQ: OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Research 2015;43:W78-W84.

12 Dermauw W, Van Leeuwen T, Vanholme B, Tirry L: The complete mitochondrial genome of the house dust mite Dermatophagoides pteronyssinus (Trouessart): a novel gene arrangement among arthropods. Bmc Genomics 2009;10

13 Chang DD, Clayton DA: Priming of Human Mitochondrial-DNA Replication Occurs at the Light-Strand Promoter. Proceedings of the National Academy of Sciences of the United States of America 1985;82:351-355.

14 Chan TF, Ji KM, Yim AK, Liu XY, Zhou JW, Li RQ, Yang KY, Li J, Li M, Law PT, Wu YL, Cai ZL, Qin H, Bao Y, Leung RK, Ng PK, Zou J, Zhong XJ, Ran PX, Zhong NS, Liu ZG, Tsui SK: The draft genome, transcriptome, and microbiome of Dermatophagoides farinae reveal a broad spectrum of dust mite allergens. J Allergy Clin Immunol 2015;135:539-548.

15 Chaisiri K, McGarry JW, Morand S, Makepeace BL: Symbiosis in an overlooked microcosm: a systematic review of the bacterial flora of mites. Parasitology 2015;142:1152-1162.

16 Mofiz E, Holt DC, Seemann T, Currie BJ, Fischer K, Papenfuss AT: Genomic resources and draft assemblies of the human and porcine varieties of scabies mites, Sarcoptes scabiei var. hominis and var. suis. Gigascience 2016;5:23.

17 Arlian LG, Morgan MS, Rider SD: Sarcoptes scabiei: genomics to proteomics to biology. Parasite Vector 2016;9:380.

18 Lowe TM, Eddy SR: tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 1997;25:955-964.

Supplemental Figure Legends

Supplemental Figure 1. Kmer analysis of *D. pteronyssinus* Illumina data

Kmer analysis of two of the Illumina datasets used in this study. A) MiSeq, B) HiSeq2500, C) Illumina sequence from *S. scabiei* (Rider genome), D) unpublished sequence data from a haploid fungal genome. Heterozygosity is indicated as blue line is the distribution of heterozygous kmers, green line homozygous kmers

Supplemental Figure 2. Dotplot of mitochondrial genomes

On the horizontal axis is the sequence of the published *D. pteronyssinus* mitochondrial genome and on the vertical axis is the mitochondrial genome assembled from the PacBio sequence used in this study.

Supplemental Figure 3. Multiple sequence alignment of the mitochondrial D loop region

The D loop region of both the published mitochondrial genome of *D. pteronyssinus* and our assembled mitochondrial genome is highlighted.

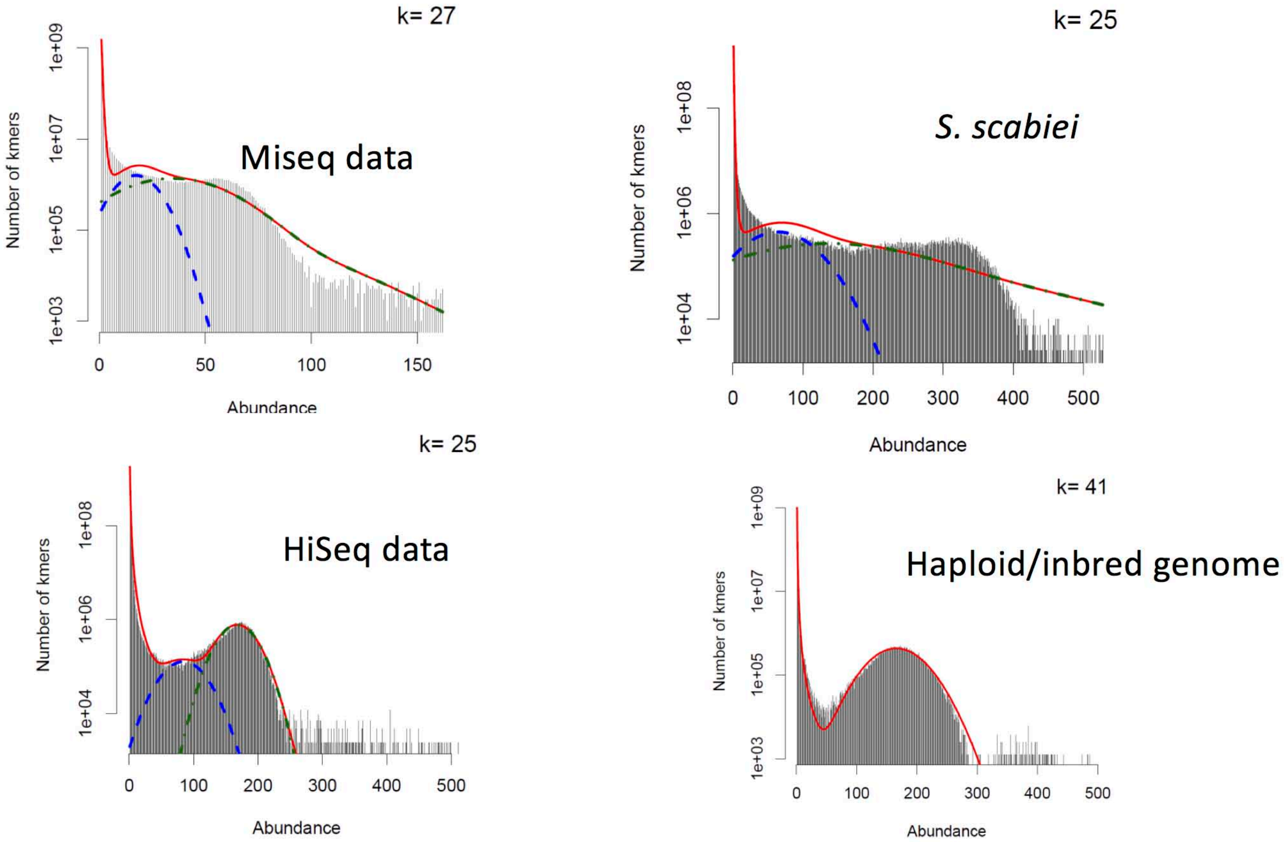
Supplemental Figure 4. Dotplot of *Serratia marcescens* genomes

On the horizontal axis is the sequence from *S. marcescens* WW4 and on the vertical axis is the *S. marcescens* genome assembled from the PacBio sequence used in this study.

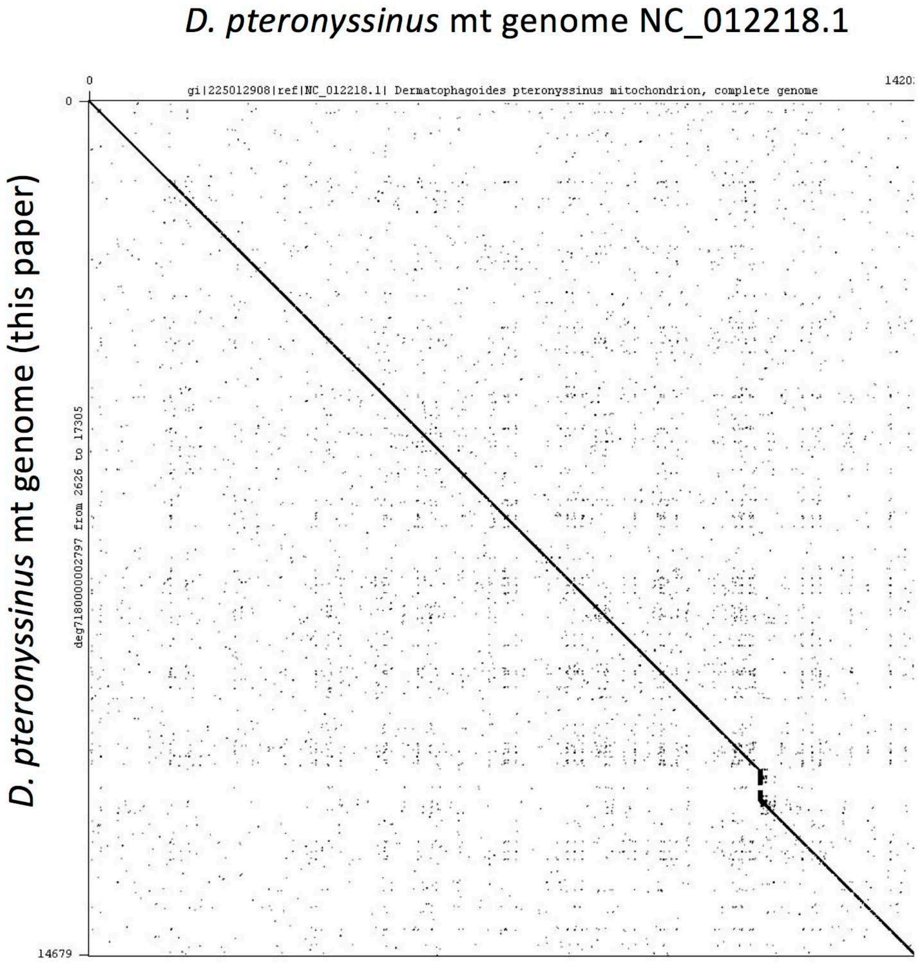
Supplemental Figure 5. Predominant bacterial species in the DP microbiome

The relative percentage of the top ten species found in the DP microbiome is shown as a pie chart.

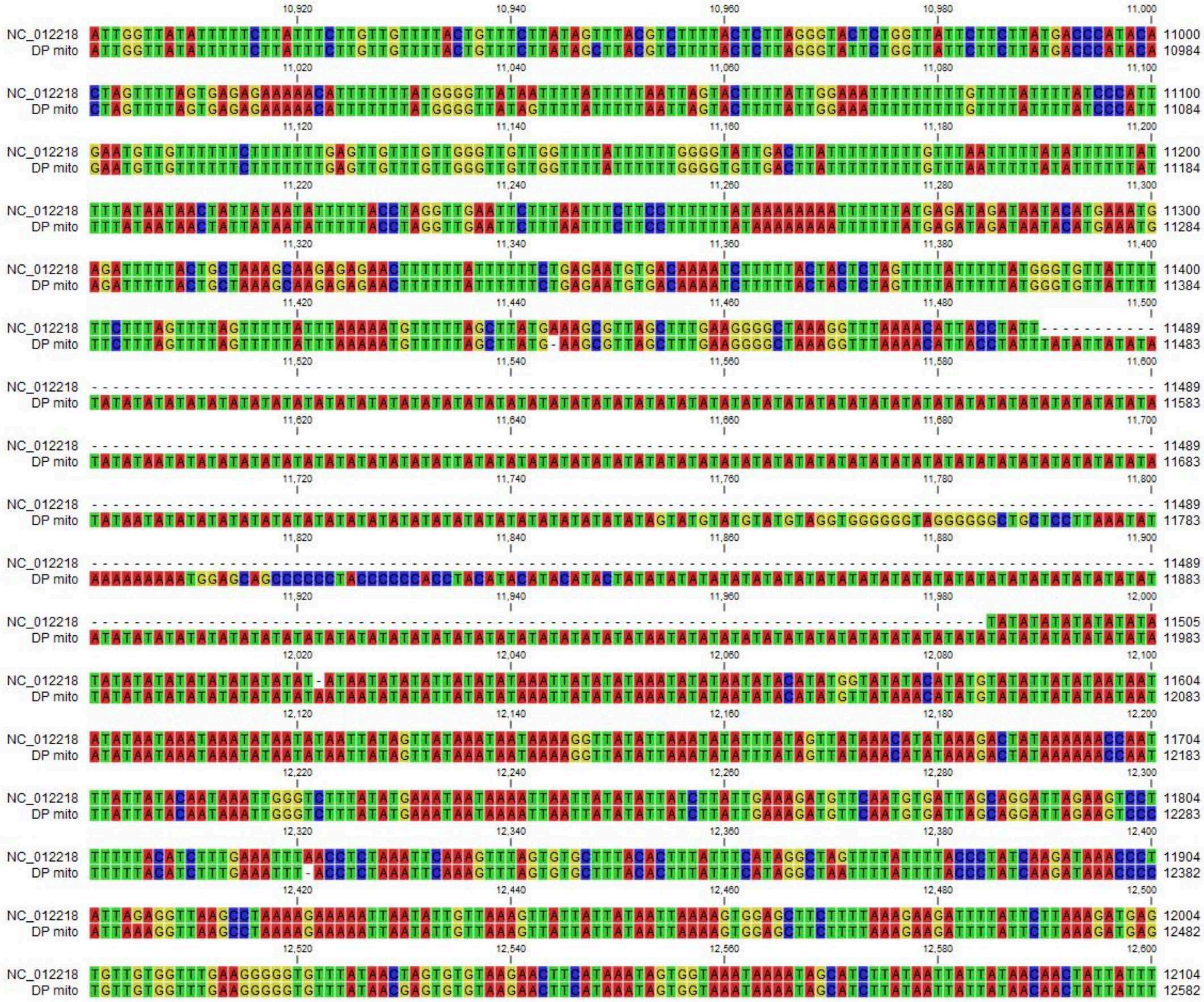
Supplemental Figure 1



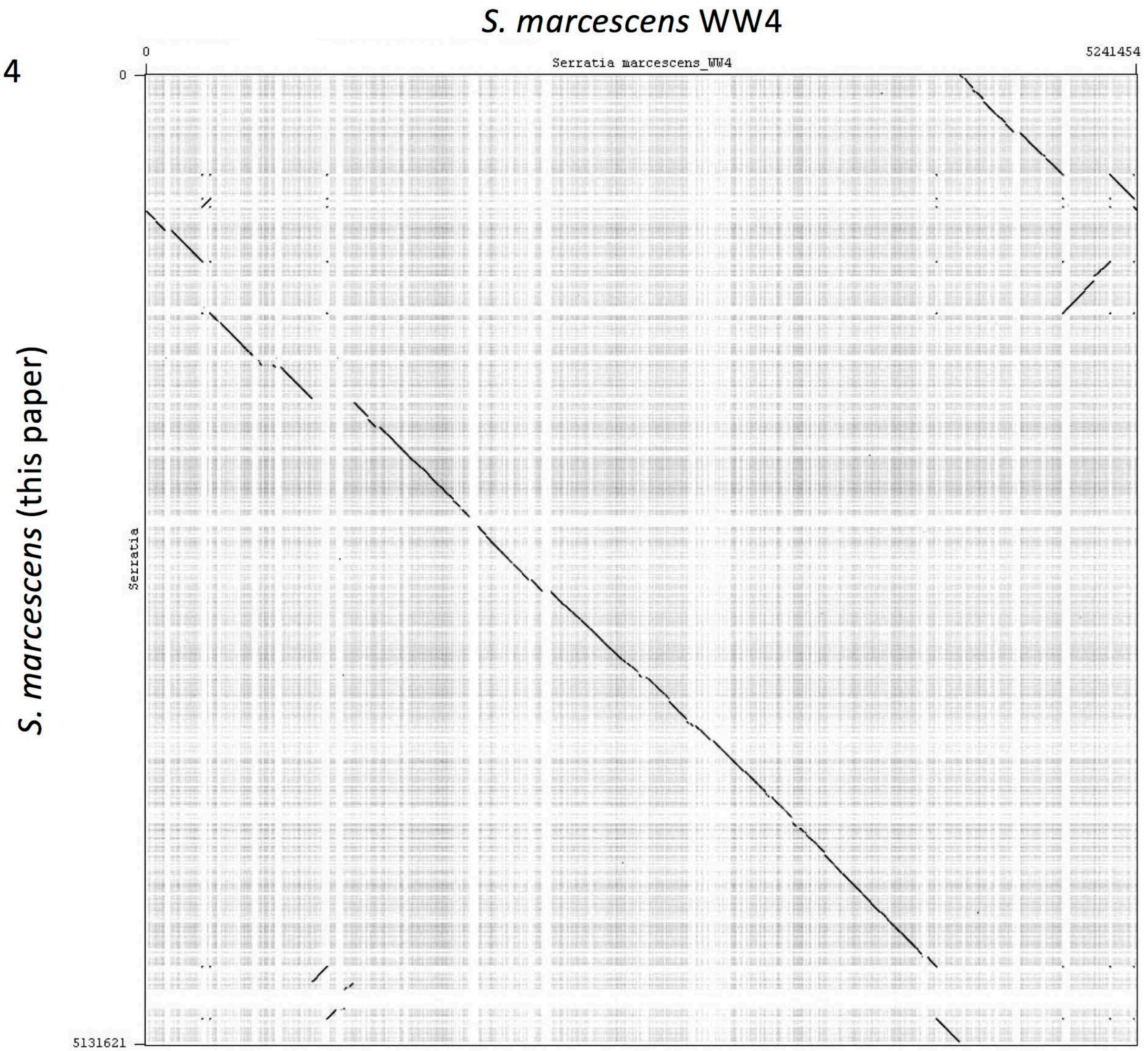
Supplemental Figure 2



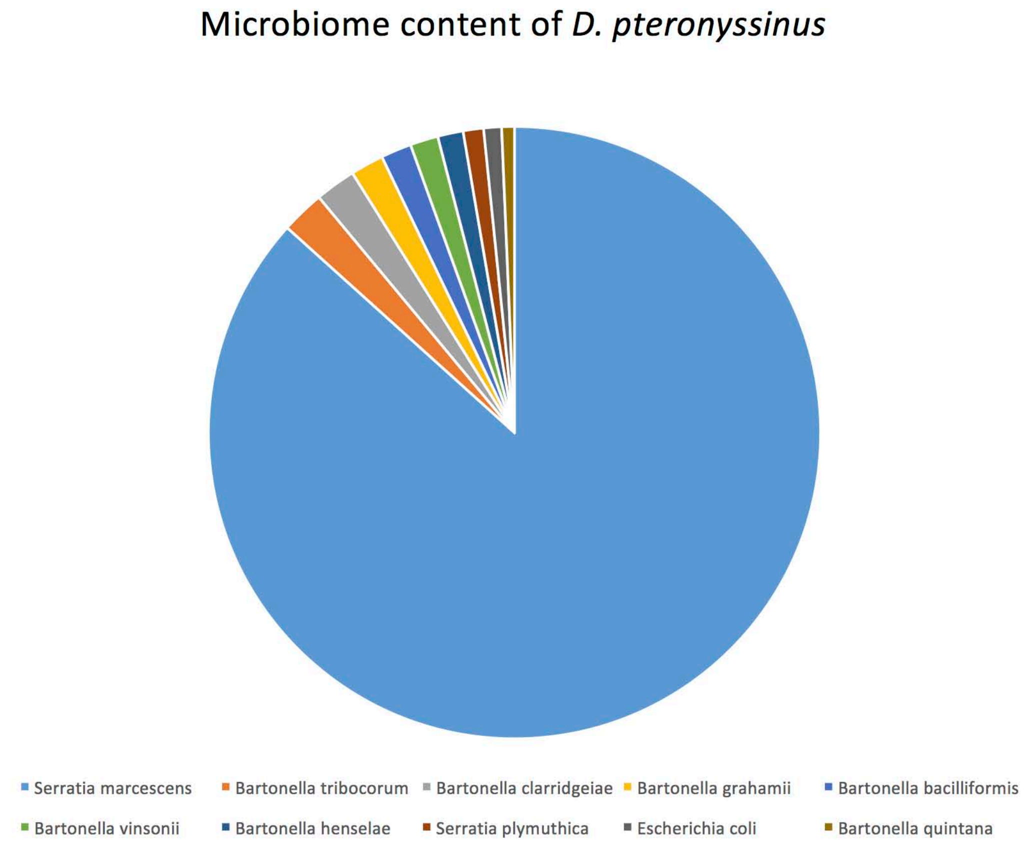
Supplemental Figure 3



Supplemental Figure 4



Supplemental Figure 5



Supplemental Table 1: Conserved Allergen Orthologs

Orthologs		
missing in all outside Dermatophagoides: 1, Derp6		
conserved in all outside Dermatophagoides: 20, Derp1,3,8,10,11,13,14,15,16,20,24,25,26,27,28,29,30,31,32,33		
D farinae	**Gene ID**	**% ID**
Derp1	DEFA_073880	83
Derp2	DEFA_057430	87
Derp3	DEFA_036500	79
Derp4	DEFA_092370	88
Derp5	DEFA_009370	77
Derp6	DEFA_160240	81
Derp7	DEFA_012670	87
Derp8	DEFA_081580	70
Derp9	DEFA_108510	84
Derp10	DEFA_012620	98
Derp11	DEFA_029610	98
Derp13	DEFA_016640	95
Derp14	DEFA_023480	88
Derp15	DEFA_127470	93
Derp16	DEFA_053360	90
Derp18	DEFA_042810	89
Derp20	DEFA_064710	96
Derp21	DEFA_009360	71
Derp22	DEFA_072800	76
Derp23	DEFA_123860	73
Derp24	DEFA_162130	75
Derp25	DEFA_001450	88
Derp26	DEFA_126820	91
Derp27	DEFA_144510	43
Derp28	DEFA_150350	98
Derp29	DEFA_018720	93
Derp30	DEFA_057540	83
Derp31	DEFA_002000	100
Derp32	DEFA_053550	90
Derp33	DEFA_107410	81
S scabiei	**Gene ID**	**% ID**
Derp1	KPM05040.1	45
Derp2	KPM02728.1	29
Derp3	KPM04366.1	43

Derp7	KPM06925.1	37
Derp8	KPM11587.1	65
Derp9	KPM04162.1	35
Derp10	KPM09025.1	98
Derp11	KPM04483.1	94
Derp13	KPM07763.1	90
Derp14	KPM11048.1	59
Derp15	KPM07813.1	44
Derp16	KPM04580.1	76
Derp20	KPM07362.1	90
Derp24	KPM06595.1	93
Derp25	KPM10468.1	81
Derp26	KPM03769.1	99
Derp27	KPM03560.1	99
Derp28	KPM03927.1	80
Derp29	KPM10308.1	85
Derp30	KPM04725.1	83
Derp31	KPM08623.1	97
Derp32	KPM05552.1	68
Derp33	KPM02536.1	99
Tetranychus urticae	**Gene ID**	**% ID**
Derp1	tetur12g01820.1	32
Derp3	tetur22g01230.1	38
Derp4	tetur11g04920.1	53
Derp8	tetur03g09230.1	50
Derp9	tetur01g02990.1	34
Derp10	tetur09g03350.1	71
Derp11	tetur03g04010.1	79
Derp13	tetur04g05930.1	55
Derp14	tetur03g09300.1	23
Derp15	tetur19g03220.1	47
Derp16	tetur05g02070.1	46
Derp18	tetur17g02930.1	29
Derp20	tetur11g04410.1	78
Derp24	tetur07g05320.1	46
Derp25	tetur04g06650.1	69
Derp26	tetur32g00200.1	72
Derp27	tetur20g01980.1	28
Derp28	tetur11g01700.1	86
Derp29	tetur01g12670.1	80

Derp30	tetur19g02170.1	68
Derp31	tetur35g01350.1	75
Derp32	tetur16g01440.1	58
Derp33	tetur03g00230.1	97
Ixodes scapularis	**Gene ID**	**% ID**
Derp1	ISCW017550-PA	30
Derp3	ISCW011961-PA	39
Derp4	ISCW006372-PA	43
Derp7	ISCW010031-PA	25
Derp8	ISCW022157-PA	55
Derp10	ISCW008202-PA	92
Derp11	ISCW016446-PA	77
Derp13	ISCW015316-PA	45
Derp14	ISCW009032-PA	27
Derp15	ISCW012986-PA	39
Derp16	ISCW007199-PA	37
Derp18	ISCW016746-PA	32
Derp20	ISCW021028-PA	67
Derp21	ISCW014231-PA	30
Derp22	ISCW005419-PA	24
Derp23	ISCW021741-PA	36
Derp24	ISCW012532-PA	41
Derp25	ISCW008869-PA	68
Derp26	ISCW001124-PA	37
Derp27	ISCW015204-PA	32
Derp28	ISCW017456-PA	88
Derp29	ISCW008497-PA	80
Derp30	ISCW023334-PA	64
Derp31	ISCW006326-PA	74
Derp32	ISCW000237-PA	61
Derp33	ISCW003527-PA	97
Metaseiulus occidentalis	**Gene ID**	**% ID**
Derp1	XP_003741376.1	31
Derp2	XP_003741573.1	30
Derp3	XP_003741728.1	39
Derp4	XP_003742001.1	33
Derp5	XP_003738455.1	27
Derp7	XP_003744979.1	22
Derp8	XP_003742682.1	49

Derp10	XP_003745222.1	86
Derp11	XP_003744495.1	75
Derp13	XP_003748361.1	38
Derp14	XP_003742265.1	21
Derp15	XP_003739697.1	43
Derp16	XP_003737166.1	25
Derp18	XP_003743171.1	32
Derp20	XP_003743522.1	81
Derp22	XP_003745314.1	22
Derp23	XP_003742368.1	34
Derp24	XP_003737660.1	39
Derp25	XP_003744176.1	62
Derp26	XP_003746758.1	58
Derp27	XP_003748025.1	30
Derp28	XP_003740401.1	87
Derp29	XP_003744457.1	76
Derp30	XP_003740424.1	59
Derp31	XP_003741592.1	67
Derp32	XP_003745528.1	56
Derp33	XP_003741823.1	82
Daphnia pulex	**Gene ID**	**% ID**
Derp1	EFX87400	30
Derp3	EFX80959	43
Derp4	EFX81580	51
Derp8	EFX88365	50
Derp10	EFX89938	82
Derp11	EFX65159	60
Derp13	EFX85386	45
Derp14	EFX88552	27
Derp15	EFX90412	43
Derp16	EFX88163	44
Derp20	EFX89163	73
Derp22	EFX84782	31
Derp24	EFX83282	43
Derp26	EFX86510	48
Derp27	EFX68557	28
Derp28	EFX70674	85
Derp29	EFX80656	77
Derp30	EFX74776	58
Derp31	EFX71219	51

Derp32	EFX69116	58
Derp33	EFX88974	82
Drosophila melanogaster	**Gene ID**	**% ID**
Derp1	FBpp0086763	28
Derp3	FBpp0311472	42
Derp4	FBpp0086263	48
Derp10	FBpp0088903	76
Derp11	FBpp0076321	59
Derp13	FBpp0099726	47
Derp14	FBpp0292877	22
Derp15	FBpp0071463	41
Derp16	FBpp0078607	37
Derp20	FBpp0076270	75
Derp22	FBpp0077492	27
Derp24	FBpp0084648	47
Derp25	FBpp0084949	64
Derp26	FBpp0084566	49
Derp27	FBpp0089193	29
Derp28	FBpp0082519	85
Derp29	FBpp0074017	81
Derp30	FBpp0073576	48
Derp31	FBpp0072097	55
Derp32	FBpp0072250	58
Derp33	FBpp0081153	83
Steganacarus magnus	**Gene ID**	**% ID**
Derp1	LBFN01096833.1-Steganacarus.2	30
Derp2	LBFN01088795.1-Steganacarus.1	35
Derp3	LBFN01086154.1-Steganacarus.1	46
Derp4	LBFN01097791.1-Steganacarus.2	61
Derp7	LBFN01099527.1-Steganacarus.4	32
Derp8	LBFN01048479.1-Steganacarus.1	43
Derp10	LBFN01095513.1-Steganacarus.2	55
Derp11	LBFN01091078.1-Steganacarus.4	84
Derp13	LBFN01095944.1-Steganacarus.1	47
Derp14	LBFN01080809.1-Steganacarus.1	26
Derp15	LBFN01082179.1-Steganacarus.1	47
Derp16	LBFN01089221.1-Steganacarus.1	43
Derp18	LBFN01098015.1-Steganacarus.1	43
Derp20	LBFN01054462.1-Steganacarus.1	85

Derp23	LBFN01078325.1-Steganacarus.1	61
Derp25	LBFN01087895.1-Steganacarus.1	59
Derp26	LBFN01053059.1-Steganacarus.1	30
Derp27	LBFN01082615.1-Steganacarus.2	29
Derp28	LBFN01097326.1-Steganacarus.1	64
Derp29	LBFN01044838.1-Steganacarus.2	66
Derp30	LBFN01071913.1-Steganacarus.1	73
Derp32	LBFN01094543.1-Steganacarus.5	64
Derp33	LBFN01097433.1-Steganacarus.1	83
Varroa	**Gene ID**	**% ID**
Derp1	ADDG01024948.1-Varroa.1	31
Derp3	ADDG01008081.1-Varroa.1	35
Derp8	ADDG01001667.1-Varroa.1	42
Derp10	ADDG01028318.1-Varroa.1	84
Derp11	ADDG01041042.1-Varroa.1	74
Derp13	ADDG01023963.1-Varroa.1	42
Derp15	ADDG01084696.1-Varroa.1	39
Derp16	ADDG01005446.1-Varroa.1	29
Derp20	ADDG01001580.1-Varroa.1	66
Derp26	ADDG01077195.1-Varroa.1	35
Derp27	ADDG01001747.1-Varroa.1	27
Derp28	ADDG01033659.1-Varroa.1	86
Derp29	ADDG01150133.1-Varroa.1	62
Derp30	ADDG01015535.1-Varroa.1	53
Derp31	ADDG01043615.1-Varroa.1	39
Derp32	ADDG01045790.1-Varroa.1	66
Derp33	ADDG01073340.1-Varroa.1	87
Hypochthonius	**Gene ID**	**% ID**
Derp1	LBFL01080588.1-Hypochthonius.1	29
Derp2	LBFL01132356.1-Hypochthonius.1	38
Derp3	LBFL01024592.1-Hypochthonius.3	46
Derp4	LBFL01002821.1-Hypochthonius.2	43
Derp7	LBFL01023569.1-Hypochthonius.1	33
Derp8	LBFL01001952.1-Hypochthonius.2	46
Derp10	LBFL01001323.1-Hypochthonius.3	86
Derp11	LBFL01010485.1-Hypochthonius.8	83
Derp15	LBFL01022083.1-Hypochthonius.5	44
Derp16	LBFL01007633.1-Hypochthonius.7	36
Derp18	LBFL01009890.1-Hypochthonius.1	39

Derp20	LBFL01038475.1-Hypochthonius.2	72
Derp23	LBFL01048614.1-Hypochthonius.1	42
Derp25	LBFL01017303.1-Hypochthonius.1	69
Derp26	LBFL01001088.1-Hypochthonius.8	73
Derp27	LBFL01009618.1-Hypochthonius.1	29
Derp28	LBFL01012381.1-Hypochthonius.1	86
Derp29	LBFL01002554.1-Hypochthonius.1	53
Derp30	LBFL01067974.1-Hypochthonius.1	75
Derp32	LBFL01001011.1-Hypochthonius.1	63
Derp33	LBFL01011273.1-Hypochthonius.5	83
Platynothrus	**Gene ID**	**% ID**
Derp1	LBFO01056876.1-Platynothrus.1	29
Derp2	LBFO01085972.1-Platynothrus.1	39
Derp3	LBFO01036856.1-Platynothrus.1	44
Derp4	LBFO01037278.1-Platynothrus.1	50
Derp7	LBFO01066228.1-Platynothrus.1	23
Derp8	LBFO01104553.1-Platynothrus.3	51
Derp9	LBFO01090592.1-Platynothrus.1	46
Derp11	LBFO01087269.1-Platynothrus.1	87
Derp13	LBFO01090415.1-Platynothrus.1	56
Derp14	LBFO01093439.1-Platynothrus.2	27
Derp15	LBFO01101302.1-Platynothrus.1	42
Derp16	LBFO01097977.1-Platynothrus.1	41
Derp20	LBFO01102491.1-Platynothrus.1	79
Derp21	LBFO01104769.1-Platynothrus.1	32
Derp23	LBFO01092702.1-Platynothrus.1	38
Derp25	LBFO01092861.1-Platynothrus.1	73
Derp26	LBFO01090133.1-Platynothrus.1	34
Derp27	LBFO01105086.1-Platynothrus.2	32
Derp28	LBFO01088396.1-Platynothrus.1	64
Derp29	LBFO01081721.1-Platynothrus.1	86
Derp30	LBFO01104688.1-Platynothrus.1	80
Derp32	LBFO01096309.1-Platynothrus.3	65
Derp33	LBFO01101372.1-Platynothrus.2	80

Supplemental Table 2: GO terms associated with gene clusters

Enriched in DP				**Enriched in DF**			
GO ID	Name	Namespace	p-value	GO ID	Name	Namespace	p-value
GO:0003848	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase activity	molecular_function	0.033	GO:0015020	glucuronosyltransferase activity	molecular_function	0.007
GO:0008240	tripeptidyl-peptidase activity	molecular_function	0.033	GO:0045595	regulation of cell differentiation	biological_process	0.015
GO:0008238	exopeptidase activity	molecular_function	0.033	GO:0042684	cardioblast cell fate commitment	biological_process	0.027
GO:0008976	polyphosphate kinase activity	molecular_function	0.033	GO:0006477	protein sulfation	biological_process	0.027
GO:0004791	thioredoxin-disulfide reductase activity	molecular_function	0.033	GO:0035223	leg disc pattern formation	biological_process	0.027
GO:0004592	pantoate-beta-alanine ligase activity	molecular_function	0.033				
GO:0046656	folic acid biosynthetic process	biological_process	0.033				
GO:0004612	phosphoenolpyruvate carboxykinase (ATP) activity	molecular_function	0.033				
GO:0004156	dihydropteroate synthase activity	molecular_function	0.033				
GO:0004150	dihydroneopterin aldolase activity	molecular_function	0.033				
GO:0006799	polyphosphate biosynthetic process	biological_process	0.033				
GO:0006797	polyphosphate metabolic process	biological_process	0.033				
GO:0009358	polyphosphate kinase complex	cellular_component	0.033				

Top 20 Clusters Common to all 6 mites

Number of proteins	Swiss-Prot Hit	GO Annotation
93	Probable cytochrome P450 28c1	GO:0005783; C:endoplasmic reticulum; GO:0016020; C:membrane; GO:0020037; F:heme binding;
70	Esterase-5A	GO:0005576; C:extracellular region; GO:0052689; F:carboxylic ester hydrolase activity;
34	Lipase member K, putative {ECO:0000313|EMBL:EEC11220.1}	GO:0004806; F:triglyceride lipase activity; GO:0006629; P:lipid metabolic process;
29	Glutathione S-transferase 1	GO:0004364; F:glutathione transferase activity;
26	Cadherin-3	GO:0030054; C:cell junction; GO:0005737; C:cytoplasm; GO:0016021; C:integral component of membrane;
26	Deoxyribonuclease-2	GO:0005622; C:intracellular; GO:0005764; C:lysosome; GO:0005634; C:nucleus; IMP:WormBase
25	Histone H2B	GO:0005811; C:lipid particle; ISS:; GO:0000786; C:nucleosome; GO:0005634; C:nucleus;

24	Phospholipase A1 2	GO:0005576; C:extracellular region; GO:0052740; F:1-acyl-2-lysophosphatidylserine acylhydrolase activity; GO:0008970; F:phosphatidylcholine 1-acylhydrolase activity;
24	Chorion class B protein M2807	GO:0042600; C:chorion; GO:0005213; F:structural constituent of chorion; GO:0007304; P:chorion-containing eggshell formation;
24	N/A	N/A
20	N/A	N/A
20	Late histone H2A.2.1	GO:0000786; C:nucleosome; GO:0005634; C:nucleus; GO:0003677; F:DNA binding;
19	Transcription factor Ken {ECO:0000250|UniProtKB:O77459}	GO:0005634; C:nucleus; GO:0003677; F:DNA binding; GO:0046872; F:metal ion binding;
19	Protein doublesex	GO:0005634; C:nucleus; GO:0003677; F:DNA binding; GO:0046872; F:metal ion binding;
18	Peroxisomal multifunctional enzyme type 2 {ECO:0000303|PubMed:21320074}	GO:0005777; C:peroxisome; GO:0080023; F:3R-hydroxyacyl-CoA dehydratase activity; GO:0004300; F:enoyl-CoA hydratase activity;
18	Acetylcholinesterase	GO:0030054; C:cell junction; GO:0005576; C:extracellular region; GO:0005886; C:plasma membrane;
18	Cuticle protein 16.8	GO:0008061; F:chitin binding; GO:0042302; F:structural constituent of cuticle;
17	Putative cytochrome P450 CYP13A2	GO:0020037; F:heme binding; ; GO:0005506; F:iron ion binding; ; GO:0004497; F:monooxygenase activity;
17	N/A	N/A

Supplemental Table 3: Conservation of genes in insect sexual development

D. melanogaster	***D. pteronyssinus***	**FPKM**	***D. farinae***	**FPKM**	***S. scabiei***	***T. urticae***	***I. scapularis***	***V. destructor***	***M. occidentalis***	***D. pulex***
Sex lethal (AAO41638.2)	DEPT_13667 (5e-44)	22.2	DEFA_088490 (4e-27)	7.5	KPM06079.1 (1e-41)	tetur05g00810.1 (2e-43)	ISCW021980-PA (4e-36)	ADDG01104496.1-Varroa.1 (1e-28)	Mocc_002464 (2e-44)	EFX75395 (1e-62)
transformer (AAF49441.1)	*Not Detected*	N/A	*Not Detected*	N/A	*Not Detected*	*Not Detected*	*Not Detected*	*Not Detected*	*Not Detected*	*Not Detected*
transformer-2 (AAF58232.2)	DEPT_14379 (6e-36)	124.9	DEFA_077550 (1e-35)	14.9	KPM11452.1 (1e-36)	tetur12g02510.1 (3e-10)	ISCW010309-PA (1e-37)	ADDG01000316.1-Varroa.1 (6e-07)	Mocc_000614 (8e-34)	EFX90042 (2e-35)
dmrt11E (AAF48261.2)	DEPT_22290 (2e-20)	34.4	DEFA_059960 (8e-24)	0.0	KPM06218.1 (6e-23)	tetur10g00950.1 (4e-23)	ISCW002258-PA (1e-19)	ADDG01002179.1-Varroa.1 (4e-14)	Mocc_00398f (1e-23)	EFX74782 (2e-13)
doublesex (AAF54169.1)	DEPT_02151 (2e-12)	1.0	DEFA_123420 (3e-12)	0.0	KPM06218.1 (1e-12)	tetur09g04250.1 (2e-13)	ISCW007589-PA (2e-13)	ADDG01114070.1-Varroa.1 (6e-13)	Mocc_0020f7 (3e-13)	EFX84867 (2e-13)
intersex (AAF58653.1)	DEPT_05843 (7e-14)	7.6	DEFA_024410 (2e-15)	0.0	KPM02401.1 (1e-11)	tetur20g02150.1 (4e-16)	*Not Detected*	*Not Detected*	Mocc_0012af (4e-19)	EFX80823 (2e-25)

Supplemental Table 4: Microbiome Sequences

Genbank ID	**Contigs**	**Species**	**Percentage**
NC_020211.1	644	*Serratia marcescens*	86.7
NC_010161.1	17	*Bartonella tribocorum*	2.3
NC_014932.1	16	*Bartonella clarridgeiae*	2.2
NC_012846.1	13	*Bartonella grahamii*	1.7
NC_008783.1	12	*Bartonella bacilliformis*	1.6
NC_020301.1	11	*Bartonella vinsonii*	1.5
NC_005956.1	10	*Bartonella henselae*	1.3
NC_021659.1	8	*Serratia plymuthica*	1.1
NC_017639.1	7	*Escherichia coli*	0.9
NC_018533.1	5	*Bartonella quintana*	0.7

Supplemental Table 5: Transfer RNA Data

D. pteronyssinus								
Sequence		tRNA	Bounds	tRNA	Anti	Intron Bounds	Cove	
Name	tRNA #	Begin	End	Type	Codon	Begin	End	Score
--------	------	----	------	----	-----	-----	----	------
NODE_121_length_140161_cov_6.2817	1	126340	126267	Arg	ACG	0	0	67.22
NODE_82_length_215458_cov_9.55141	2	62551	62478	Arg	ACG	0	0	67.22
NODE_126_length_134430_cov_7.0447	1	43937	44010	Asn	GTT	0	0	75.36
NODE_256_length_24894_cov_4.54867	1	16154	16225	Asn	GTT	0	0	86.94
NODE_50_length_327162_cov_6.56616	3	66794	66721	Asn	GTT	0	0	75.36
NODE_256_length_24894_cov_4.54867	3	15889	15819	Cys	GCA	0	0	59
NODE_6_length_1335509_cov_10.0386	1	11311	11381	Gly	GCC	0	0	81.49
NODE_6_length_1335509_cov_10.0386	5	224443	224373	Gly	GCC	0	0	81.49
NODE_204_length_59224_cov_4.57969	1	52116	52033	Leu	CAA	0	0	61.56
NODE_256_length_24894_cov_4.54867	2	17200	17282	Leu	TAA	0	0	50.05
NODE_174_length_82258_cov_5.66399	1	63290	63363	Met	CAT	0	0	68.14
NODE_201_length_60701_cov_4.25767	1	4941	5014	Met	CAT	0	0	74.21
NODE_174_length_82258_cov_5.66399	2	69448	69534	Ser	CGA	0	0	55.42
NODE_238_length_36967_cov_2.44137	1	33794	33708	Ser	GCT	0	0	61.02
NODE_162_length_93710_cov_4.52414	1	83966	83894	Thr	CGT	0	0	78.58
NODE_37_length_405049_cov_10.3973	2	53431	53360	Val	AAC	0	0	67.67
NODE_21_length_537647_cov_8.12628	2	312772	312700	Val	CAC	0	0	78.81
NODE_123_length_138814_cov_7.81013	1	9505	9577	Val	TAC	0	0	84.07
NODE_234_length_40232_cov_7.50959	1	10682	10808	Leu	TAG	10720	10766	55.32
NODE_62_length_270443_cov_8.00326	1	14503	14598	Ala	AGC	14540	14562	64.64
NODE_109_length_160280_cov_7.72601	1	16931	17083	Ser	TGA	16968	17038	72.1
NODE_91_length_193628_cov_8.07474	1	20183	20061	Leu	AAG	20145	20105	59.77
NODE_76_length_224417_cov_12.1388	1	25459	25566	Gly	GCC	25495	25531	75.23
NODE_104_length_167383_cov_11.8693	1	33142	33047	Ala	AGC	33105	33083	62.35

NODE_166_length_89217_cov_9.86487	1	34688	34824	His	GTG	34725	34789	58.58
NODE_48_length_337033_cov_8.5389	1	39564	39698	Gln	TTG	39601	39663	71.93
NODE_147_length_107316_cov_4.0667	1	42624	42492	Gln	TTG	42587	42527	65.4
NODE_76_length_224417_cov_12.1388	2	43693	43596	Gly	GCC	43657	43631	74.68
NODE_45_length_355045_cov_8.99001	1	50847	50731	Glu	TTC	50810	50766	66.67
NODE_163_length_90635_cov_4.07421	1	59664	59575	Met	CAT	59627	59610	59.57
NODE_128_length_133363_cov_5.80674	1	74807	74936	Gln	CTG	74844	74901	60.53
NODE_124_length_138038_cov_8.88327	1	85759	85935	Pro	TGG	85795	85899	61.87
NODE_48_length_337033_cov_8.5389	3	96982	96844	Cys	GCA	96946	96880	67.75
NODE_88_length_198073_cov_9.89486	2	97171	97023	Arg	CCG	97134	97059	51.58
NODE_22_length_535680_cov_8.55032	1	110104	109998	Asp	GTC	110067	110033	69.04
NODE_34_length_414532_cov_10.0175	1	114015	114135	Met	CAT	114051	114099	61.83
NODE_50_length_327162_cov_6.56616	1	121543	121670	Pro	TGG	121579	121634	67.18
NODE_96_length_185906_cov_8.08312	1	122128	122254	Lys	TTT	122166	122219	57.93
NODE_36_length_407118_cov_8.18279	1	126606	126721	Glu	CTC	126643	126686	66.97
NODE_133_length_128442_cov_10.2256	1	127703	127805	Lys	TTT	127741	127769	70.97
NODE_24_length_497719_cov_10.2388	1	139538	139681	Thr	CGT	139574	139645	65.64
NODE_107_length_162135_cov_6.38695	1	140373	140501	Ile	AAT	140410	140465	74.02
NODE_88_length_198073_cov_9.89486	1	149096	148981	Val	AAC	149059	149017	83.88
NODE_82_length_215458_cov_9.55141	1	149100	149201	Thr	AGT	149138	149165	74.95
NODE_57_length_295260_cov_9.30942	1	167121	166936	Ser	GCT	167084	166981	76.25
NODE_48_length_337033_cov_8.5389	2	176992	177161	Thr	TGT	177028	177125	69.11
NODE_66_length_258107_cov_10.3146	1	180678	180540	Pro	AGG	180642	180576	69.2
NODE_37_length_405049_cov_10.3973	1	197074	196870	Leu	CAA	197036	196916	61.57
NODE_67_length_256035_cov_7.74746	1	200931	201058	Tyr	GTA	200969	201022	67.35
NODE_20_length_548211_cov_8.31221	1	213954	214180	Leu	CAG	213992	214133	60.68
NODE_16_length_614163_cov_8.46035	1	220679	220574	Arg	CCT	220642	220610	62.28
NODE_62_length_270443_cov_8.00326	2	231561	231744	His	GTG	231598	231710	48.14
NODE_50_length_327162_cov_6.56616	2	325252	325378	Ile	AAT	325289	325342	71.91
NODE_21_length_537647_cov_8.12628	1	378876	378753	Gly	TCC	378839	378788	59.61

NODE_20_length_548211_cov_8.31221	2	391253	391497	Leu	CAA	391291	391451	56.45
NODE_2_length_1922611_cov_10.9849	2	424212	424105	Asp	GTC	424175	424140	66.82
NODE_14_length_617406_cov_10.4945	1	453426	453294	Pro	CGG	453390	453330	74.41
NODE_8_length_793838_cov_10.483	1	459641	459764	Leu	TAA	459679	459718	67.76
NODE_19_length_556627_cov_7.84241	1	553179	553292	Phe	GAA	553216	553256	61.1
NODE_19_length_556627_cov_7.84241	2	553638	553521	Phe	GAA	553601	553557	61.88
NODE_13_length_645311_cov_7.24388	1	602947	602792	Ala	CGC	602911	602828	69.74
NODE_6_length_1335509_cov_10.0386	2	691687	691809	Tyr	GTA	691725	691773	63.78
NODE_3_length_1472047_cov_11.5404	2	694153	694044	Ala	TGC	694117	694080	62.94
NODE_6_length_1335509_cov_10.0386	3	987082	987184	Lys	TTT	987120	987148	68.42
NODE_6_length_1335509_cov_10.0386	4	1094531	1094414	Lys	CTT	1094492	1094450	75.59
NODE_3_length_1472047_cov_11.5404	1	1115042	1115216	Ser	CGA	1115079	1115171	76
NODE_2_length_1922611_cov_10.9849	1	1601266	1601157	Trp	CCA	1601230	1601193	68.37
NODE_1_length_2378295_cov_10.3433	1	1905847	1905716	Glu	TTC	1905810	1905751	65.73

D. farinae								
Sequence		tRNA	Bounds	tRNA	Anti	Intron Bounds	Cove	
Name	tRNA #	Begin	End	Type	Codon	Begin	End	Score
--------	------	----	------	----	-----	-----	----	------
KN266225.1	2	36125	36198	Ala	TGC	0	0	75.26
KN266225.1	21	502233	502160	Ala	GGC	0	0	74.47
KN266225.1	17	648874	648947	Arg	TCG	0	0	66.34
KN266225.1	22	467628	467557	Arg	CCT	0	0	64.81
KN266225.1	23	348408	348335	Arg	ACG	0	0	67.51
KN266225.1	15	545844	545917	Asn	GTT	0	0	86.68
KN266282.1	1	113959	114032	Asn	GTT	0	0	79.68
KN266371.1	1	25771	25844	Asn	GTT	0	0	79.68
KN266225.1	20	649744	649671	Asp	GTC	0	0	79.35
KN266225.1	1	26260	26331	Gln	TTG	0	0	59.88
KN266324.1	2	148163	148092	Glu	TTC	0	0	62.59

KN266225.1	5	308954	309026	Gly	GCC	0	0	74.25
KN266225.1	9	349839	349911	Gly	TCC	0	0	71.55
KN266225.1	25	211631	211558	His	GTG	0	0	69.72
KN266225.1	6	309247	309329	Leu	TAA	0	0	70.51
KN266225.1	12	441024	441105	Leu	GAG	0	0	40.13
KN266225.1	4	145415	145487	Lys	TTT	0	0	78.08
KN266225.1	7	340394	340466	Met	CAT	0	0	74.03
KN266225.1	14	511421	511495	Met	CAT	0	0	72.13
KN266225.1	24	337101	337028	Met	CAT	0	0	84.38
KN266225.1	18	718258	718330	Phe	GAA	0	0	63.16
KN266297.1	1	66738	66809	Pro	TGG	0	0	73.8
KN266297.1	2	56493	56422	Pro	TGG	0	0	73.8
KN266225.1	3	81436	81524	Ser	GGA	0	0	59.51
KN266225.1	26	32507	32419	Ser	CGA	0	0	42.33
KN266324.1	1	65671	65757	Ser	TGA	0	0	56.3
KN266225.1	10	349928	349999	Thr	GGT	0	0	77.58
KN266225.1	19	654135	654063	Thr	TGT	0	0	79.18
KN266225.1	11	351373	351445	Trp	CCA	0	0	61.77
KN266225.1	8	349727	349807	Tyr	GTA	0	0	53.52
KN266225.1	13	441306	441380	Val	GAC	0	0	68.83
KN266225.1	16	547030	547103	Val	TAC	0	0	77.92
KN266417.1	1	46590	46518	Val	TAC	0	0	79.88
KN266694.1	1	426	661	Leu	TAG	464	619	50.01
KN266659.1	1	2074	2192	Cys	GCA	2110	2156	61.88
KN266642.1	1	4654	4753	Gly	GCC	4690	4718	72.02
KN266295.1	1	6317	6208	Ala	AGC	6280	6244	61.3
KN266352.1	1	6397	6552	Pro	CGG	6433	6516	67.47
KN266378.1	1	8540	8399	Met	CAT	8504	8435	57.9
KN266713.1	1	11500	11600	Trp	CCA	11538	11564	65.18
KN266442.1	1	11529	11740	Glu	TTC	11566	11705	59.37

KN266703.1	1	13311	13411	Trp	CCA	13349	13375	67.65
KN266554.1	1	14403	14538	Ile	AAT	14440	14502	79.73
KN266571.1	1	15442	15326	Arg	ACG	15404	15362	66.25
KN266389.1	1	16219	16025	Leu	AAG	16181	16069	51.95
KN266495.1	1	23441	23576	Lys	TTT	23479	23540	67.05
KN266387.1	1	36554	36703	Gln	CTG	36592	36668	52.05
KN266380.1	1	60210	60348	Glu	CTC	60247	60313	56.24
KN266471.1	1	62288	62395	Arg	CCT	62325	62359	65.54
KN266246.1	1	72145	72015	Thr	TGT	72109	72050	64.12
KN266240.1	1	72431	72600	Arg	TCT	72469	72564	76.44
KN266434.1	1	76579	76461	Ile	TAT	76541	76497	59.9
KN266427.1	1	84943	85042	Asp	GTC	84980	85007	62.35
KN266302.1	1	93468	93671	Thr	CGT	93504	93635	65.8
KN266231.1	1	102875	102791	Arg	TCG	102838	102827	65.64
KN266239.1	1	103273	103087	Ser	GCT	103236	103132	79.05
KN266268.1	2	104272	104181	Val	CAC	104236	104217	57.43
KN266282.1	2	113362	113214	Ser	AGA	113325	113259	77.59
KN266286.1	1	128566	128440	Ala	AGC	128529	128476	65.37
KN266314.1	1	173391	173559	Val	AAC	173427	173523	58.84
KN266268.1	1	173958	173823	Gly	TCC	173921	173858	74.64
KN266267.1	1	206995	206873	His	GTG	206958	206908	48.5
KN266232.1	1	248870	248965	Phe	GAA	248908	248929	68.45
KN266236.1	1	279918	279819	Asp	GTC	279881	279854	64.15
KN266226.1	1	432427	432613	Leu	CAA	432466	432567	61.44

```
==================================================
Repeat Masker Statistics  (Supplemental Table 5 continued)
sequences:                    834
total length:          52471908 bp  (52470720 bp excl N/X-runs)
GC level:                  29.03%
bases masked:           4846039 bp ( 9.24 %)
==================================================
                       number of    length       percentage
                       elements*    occupied     of sequence
--------------------------------------------------
SINEs:                         2    123 bp           0.00%
ALUs                           0    0 bp             0.00%
MIRs                           0    0 bp             0.00%
LINEs:                       122    7753 bp          0.01%
LINE1                         12    722 bp           0.00%
LINE2                         18    1032 bp          0.00%
L3/CR1                        48    3049 bp          0.01%
LTR elements:                  7    558 bp           0.00%
ERVL                           1    42 bp            0.00%
ERVL-MaLRs                     0    0 bp             0.00%
ERV_classI                     6    516 bp           0.00%
ERV_classII                    0    0 bp             0.00%
DNA elements:                 81    6421 bp          0.01%
hAT-Charlie                   31    2374 bp          0.00%
TcMar-Tigger                   9    586 bp           0.00%
Unclassified:                 10    807 bp           0.00%
Total interspersed repeats:         15662 bp         0.03%
Small RNA:                    72    9807 bp          0.02%
Satellites:                    0    0 bp             0.00%
Simple repeats:           111737    4252224 bp       8.10%
Low complexity:            12161    570131 bp        1.09%
==================================================

* most repeats fragmented by insertions or deletions
  have been counted as one element


The query species was assumed to be homo sapiens
RepeatMasker version open-4.0.5 , default mode

run with rmblastn version 2.2.27+
RepBase Update 20140131, RM database version 20140131
```

Specifications MISA Statistics (Supplemental Table 5 Continued)	
==============	
Microsatellites (unit size / minimum number of repeats):	(1/10) (2/6) (3/5) (4/5) (5/5) (6/5)
Maximal number of bases interrupting 2 SSRs in a compound microsatellite:	100
RESULTS OF MICROSATELLITE SEARCH	
================================	
Total number of sequences examined:	834
Total size of examined sequences (bp):	52471908
Total number of identified SSRs:	103106
Number of SSR containing sequences:	399
Number of sequences containing more than 1 SSR:	338
Number of SSRs present in compound formation:	28539
Distribution to different repeat type classes	
Unit size	Number of SSRs
1	39128
2	12715
3	46858
4	4139
5	222
6	44