Supplemental material for:

Transferred interbacterial antagonism genes augment eukaryotic innate immune function

Seemay Chou[1,*], Matthew D. Daugherty[2,3,*], S. Brook Peterson[1], Jacob Biboy[4], Youyun Yang[5], Brandon L. Jutras[6,7], Lillian K. Fritz-Laylin[8], Michael A. Ferrin[1], Brittany N. Harding[1], Christine Jacobs-Wagner[6,7,9,10], X. Frank Yang[5], Waldemar Vollmer[4], Harmit S. Malik[2,3], and Joseph D. Mougous[1,]**

[1]Department of Microbiology, University of Washington School of Medicine, Seattle, WA 98195, USA.

[2]Division of Basic Sciences and [3]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

[4]Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, NE2 4AX, UK.

[5]Department of Microbiology and Immunology, Indiana University School of Medicine, Indianapolis, IN 46202, USA.

[6]Microbial Diversity Institute, [7]Howard Hughes Medical Institute, and Departments of [9]Microbial Pathogenesis and [10]Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06516, USA.

[8]Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158, USA.

[*] These authors contributed equally to this work.
[**] To whom correspondence should be addressed: mougous@u.washington.edu

Methods

Computational searches. Homologs of *tae* were searched for using iterative PSI-BLAST[29]. First, bacterial homologs were assembled using PSI-BLAST searches limited to bacterial sequences in the non-redundant (NR) protein database. Sequences with e-values <1e-10 (for *tae2)* or <1e-20 (for *tae1*, *tae3* and *tae4*) and greater than 50% query coverage were included in successive rounds until no new homologs were identified. Position-specific scoring matrix (PSSM) was used to query the NR database limited to eukaryotic sequences. Eukaryotic homologs with e-values <1e-5 were used to initiate an iterative PSI-BLAST search for eukaryotic proteins, as described above (e-value cut-off 1e-5). Eukaryotic homologs were validated by the presence of introns or flanking eukaryotic genes. To validate the *Oxytricha trifallax dae3* gene identified in the macronucleus genome, we searched the unpublished micronucleus genome (http://oxy.ciliate.org/blast/) for evidence of a fragmented *dae3* gene that would be consistent with gene rearrangement in this species[13]. EST, whole-genome sequence (WGS) and transcriptome databases were searched with tBlastN[29] using validated *dae* genes. We gratefully acknowledge the deposition of unpublished data into these databases from multiple sources, including Baylor College of Medicine Human Genome Sequencing Center (https://www.hgsc.bcm.edu), The Genome Institute at Washington University School of Medicine (http://genome.wustl.edu), the U.S. Geological Survey (http://www.usgs.gov), the Functional Genomics Center Zurich (http://www.fgcz.ch), the Joint Genome Institute (http://jgi.doe.gov) and the Broad Institute (http://www.broadinstitute.org). Hits from EST or transcriptome databases were accepted in cases where the hit was more closely related to a validated *dae* than a

bacterial *tae*. When gene predictions based on genomic sequences differed from experimental data from EST or transcriptome databases, we utilized experimental data to confirm or modify the predicted protein sequence. For instance, the predicted *dae2* from *I. scapularis* (gi|242000170) lacks a secretion signal, whereas the sequence from EST data (gi|156264544) is different in the first exon, resulting in a strongly predicted secretion signal similar to the other tick sequences.

Phylogenetic and evolutionary analysis. Bacterial and eukaryotic sequences were aligned using MUSCLE[30] and edited using Geneious[31]. Regions encompassing the catalytic domain were used for phylogenetic analyses; sequences with >98% identity were excluded. The best-fitting evolutionary model was determined by Prottest[32]. Maximum likelihood phylogenetic trees were generated with PhyML[33] using 500 bootstrap replicates. To validate phylogenetic inferences, Bayesian Markov chain Monte Carlo (MCMC) analyses were performed in MrBayes[34] sampling every 500 generations until the standard deviation of split frequencies was <0.01 or 10^6 generations were sampled. Tests for purifying selection were performed on aligned and degapped nucleotide sequences of *dae* or *tae* genes. Whole gene non-synonymous:synonymous (dN/dS) ratio calculations, as well as statistical tests for purifying or positive selection for individual codons, were performed using SLAC in the HyPhy software suite[35]. Additional statistical tests in the HyPhy software suite (REL and FUBAR) confirmed that several *tae* and *dae* codons display statistically significant signatures of purifying selection. No codons demonstrate signatures of positive selection. N-terminal eukaryotic secretion signals were predicted using SignalP[36] using default cut-off values. Sequence logos were constructed using Geneious.

DNA libraries. For *N. gruberi* cDNA libraries, strain NEG was grown on *Klebsiella* and differentiated using standard protocols[37]. Synchrony was estimated by percentage of flagellates after fixing in Lugol's iodine (n>100 per time-point)[38] . 10^7 cells were harvested per sample. For *I. scapularis* cDNA libraries, ticks from the Tick-Rearing Center at Oklahoma State University, Stillwater, OK were homogenized by grinding in liquid nitrogen. RNA and DNA was purified from *I. scapularis* and *N. gruberi* samples with Trizol reagent (Invitrogen) according to the manufacturer's instructions. Contaminating genomic DNA in RNA samples was removed by treatment with Turbo DNase (Invitrogen) for one hour at 37 ^{o}C, followed by a second Trizol purification. DNA contamination was checked by PCR using actin- or GAPDH-specific primers for *I. scapularis* and *N. gruberi*, respectively. cDNA libraries were synthesized using the iScript cDNA synthesis kit (Biorad).

Expression of Dae proteins. The codon-optimized *dae* genes from *I. scapularis* (*dae2*), *N. gruberi* (*dae3*), and *B. floridae* (*dae4*) with predicted signal sequences removed were synthesized by Genscript and cloned into the pHis-sumo expression vector. Shuffle T7 pLysS cells were transformed with plasmids, and expression was induced at an optical density (OD_{600}) of 0.6 with 0.1 mM isopropyl-beta-D-thiogalactopyranoside for 20 h at 18 ^{o}C. Cells resuspended in 20 mM HEPES pH 7.5, 0.5 M NaCl, 25 mM imidazole were lysed by sonication. Lysate was cleared by centrifugation for 1 h at 18,000 x *g*, and proteins were purified with a metal-chelating affinity column. The tag was proteolytically removed with the H3C and separated from proteins using a second affinity column and size exclusion chromatography (GE Healthcare).

Sacculus analysis. PG sacculi from *E. coli* D456 (Δ*dacA* Δ*dacB* Δ*dacC*) were purified as previously described[39,40]. Preparations (300 μg) were incubated with Dae2 (1 μM), Dae3 (10 μM) or Dae4 (10 μM) in 300 μl of 20 mM HEPES pH 7.5, 100 mM NaCl for 4 h at 37 °C. PG sacculi from *B. subtilis* 168 (300 μg) or from *S. pneumoniae* R6 (120 μg) were incubated with Dae2 (6 μM) for 4 h at 37 °C. The samples were digested with cellosyl, reduced and analyzed by HPLC using published methods[40]. For preparations from *B. burgdorferi*, the B31-MI-16 strain, an infectious clone of the sequenced type strain B31, was cultured at 34 °C to early mid-log exponential growth[41-43]. Cultures were chilled on ice for 10 min and gently harvested by centrifugation at 3,250 x *g* for 15 min. Pelleted cells were washed 3 times and resuspended in cold PBS. Cell suspensions were added, drop-wise, to 6 ml of 8% SDS and boiled for 30 min. PG was prepared, incubated with Dae2 (1 μM) for 4 h at 37 °C, and analyzed as previously described[40], with the exception of HPLC buffer B, which contained 30% methanol.

Western Blot Analysis. Tissues were dissected from *I. scapularis* ticks purchased from the Tick-Rearing Center at Oklahoma State University, Stillwater, OK. A rabbit polyclonal antibody specific for *I. scapularis* Dae2 was generated by GenScript using a synthetic peptide corresponding to Dae2 amino acids 123-136 (RYGNTGKPNYNGDN, Lot #195690-4). Mouse anti-actin antibody from Abcam (GR14272-8) and anti-Rabbit (A6154) and Anti-Mouse (A4416) horseradish peroxidase (HRP)-conjugated secondary antibodies from Sigma were used. Western blot analyses and imaging were performed as previously described[44]. Four replicate analyses of tissues were performed; a representative blot is shown in Figure 3a.

Growth curves. *E. coli* growth curves were generated as previously described[11]. The vector pSCHRAB2 was used for expression of cytoplasmic *I. scapularis* Dae2, and the pSCRHAB2 vector with a *pelB* leader sequence inserted was used for expression of periplasmic Dae2. Curves are representative of three biological experiments and contain technical triplicates.

Lysis assays. Assays were performed as previously described[45]. *E. coli* reactions were carried out at enzyme concentrations of 1mM; *B. subtilis* reactions were carried out at concentrations of 1 mM (lysozyme) and 50 mM (Dae2). Curves are representative of three biological experiments and contain technical triplicates.

Bacterial killing assays. Colonies of *E. coli* or *B. subtilis* cells grown on solid LB media were washed in 0.2X PBS pH 6 and resuspended to an OD_{600} of 0.1 and 0.01, respectively. Cells were incubated with recombinant Dae2 enzyme (WT or C43A) at RT for 3 hr, and serial dilutions were plated on solid LB. Viability was quantified by enumeration of colony forming units. Curves contain four technical replicates.

Mouse and RNAi experiments. All animal experiments and tick protocols were approved by the Institutional Animal Care and Use Committee at Indiana University. The low-passage, virulent *B. burgdorferi* strain 5A4NP1, a derivative of B31-MI, was a gift from H. Kawabata and S. Norris, University of Texas Health Science Center at Houston. The strain was cultivated in Barbour-Stoenner-Kelly (BSK-II) medium supplemented with 6% normal rabbit serum (Pel Freez

Biologicals, Rogers, AR) at 35 °C with 5% CO_2. Kanamycin was added to the culture at 300 μg/ml. The mouse feeding experiments were conducted in the Vector-borne Diseases Laboratory at Indiana University School of Medicine, Indianapolis, IN. Briefly, C3H/HeN mice were needle-infected with *B. burgdorferi* (10^5 spirochetes per mouse). Two weeks post-inoculation, mouse infection was confirmed by cultivation of ear-punch biopsy specimens to assess spirochaete growth. A single growth-positive culture was used as the criterion for infection of each mouse.

RNAi in nymphal ticks was performed using previously described protocols[46]. To generate double-stranded RNA (dsRNA), 374 bp of *I. scapularis dae2* and 356 bp of the green fluorescent protein gene (*gfp*) were amplified using the following primers containing the T7 promoter (gfp_RNAi_F: gagctcTAATACGACTCACTATAGGGAGAGTGTGAGTTATAGTTGTATTCCAAT; gfp_RNAi_R: ggtaccTAATACGACTCACTATAGGGAGAGTGGAGAGGGTGAAGGTGATGCAAC; dae2_RNAi_F: ctagtcgagctcTAATACGACTCACTATAGGGAGACGCTCGTGGTCCTGGGAT; dae2_RNAi_R: ctagtcggtaccTAATACGACTCACTATAGGGAGAGTTGTAGTTGGGCTTCCCTGTA). dsRNA was synthesized and purified from PCR products using a commercial kit (Megascript RNAi Kit, Ambion, Inc.), and resuspended into elution buffer (10 mM Tris-HCl pH 7, 1 mM EDTA), aliquoted, and stored into −20°C until further use.

RNAi experiments were performed on a randomized pool of nymphs reared from three engorged female ticks collected from the wild. Five microliters of the *dae2* or *gfp* dsRNA (3 μg/μl) was loaded into capillary tubes, and 0.5 ml dsRNA was microinjected into the gut of each unfed nymph, as recently described[47]. Microinjected ticks were allowed to rest in a temperature-

controlled humidity chamber for 16 hr and ~100 nymphs were subsequently fed on *B. burgdorferi* –infected female 14-day old C3H/HeN mice. Two mice were included per RNAi treatment to account for potential variability in *B. burgdorferi* infection loads. Ticks were allowed to feed to repletion (3-5 days) and collected within 24 hours (t=0). Fed ticks were maintained in a temperature-controlled incubator until the indicated time point (2 weeks). Knockdown efficiency was analyzed by qRT-PCR analysis of *dae2* levels in engorged nymphs. The RNAi treatment groups were blinded from the time of dsRNA injections through qPCR analyses.

qPCR analyses. qPCR was performed on cDNA samples using the SsoAdvanced Universal SYBR Green Supermix (Biorad). Expression levels for *dae* genes were normalized to *actin* or *GAPDH* expression levels in *I. scapularis* and *N. gruberi*, respectively. Analyses of *dae* gene expression include three technical replicates for *N. gruberi* and technical duplicates of three biological replicates for *I. scapularis*. Populations of *B. burgdorferi* and total bacteria were quantified by qPCR in tick DNA samples using primers targeted to *flaB* (*B. burgdorferi*-specific), the 16S rRNA gene[48], or *trospa* (tick-specific). Biological replicates are shown for qPCR analyses of *B. burgdorferi* and total bacterial levels. Transcript or DNA copy numbers were calculated using a standard curve.

Methods and Extended Data References

11 Russell, A. B. *et al.* Type VI secretion delivers bacteriolytic effectors to target cells. *Nature* **475**, 343-347, doi:10.1038/nature10244 (2011).

13 Swart, E. C. *et al.* The Oxytricha trifallax macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. *PLoS Biol* **11**, e1001473, doi:10.1371/journal.pbio.1001473 (2013).

14 Jeyaprakash, A. & Hoy, M. A. First divergence time estimate of spiders, scorpions, mites and ticks (subphylum: Chelicerata) inferred from mitochondrial phylogeny. *Exp Appl Acarol* **47**, 1-18, doi:10.1007/s10493-008-9203-5 (2009).

28 Keeling, P. J. *et al.* The tree of eukaryotes. *Trends Ecol Evol* **20**, 670-676, doi:10.1016/j.tree.2005.09.005 (2005).

29 Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**, 3389-3402 (1997).

30 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797 (2004).

31 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)* **28**, 1647-1649, doi:10.1093/bioinformatics/bts199 (2012).

32 Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics (Oxford, England)* **21**, 2104-2105, doi:10.1093/bioinformatics/bti263 (2005).

33 Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **59**, 307-321, doi:10.1093/sysbio/syq010 (2010).

34 Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* **19**, 1572-1574 (2003).

35 Pond, S. L., Frost, S. D. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics (Oxford, England)* **21**, 676-679, doi:10.1093/bioinformatics/bti079 (2005).

36 Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**, 785-786, doi:10.1038/nmeth.1701 (2011).

37 Fulton, C. Amebo-flagellates as research partners: the laboratory biology of Naegleria and Tetramitus. *Methods Cell Physiol* **4**, 341-476 (1970).

38 Fulton, C. & Dingle, A. D. Appearance of the flagellate phenotype in populations of Naegleria amebae. *Dev Biol* **15**, 165-191 (1967).

39 Edwards, D. H. & Donachie, D. W. in *Bacterial Growth and Lysis: Metabolism and Structure of the Bacterial Sacculus* (eds M. A. de Pedro, J. V. Holtje, & W. Loffelhardt) (Plenum Press, 1993).

40 Glauner, B. Separation and quantification of muropeptides with high-performance liquid chromatography. *Anal Biochem* **172**, 451-464 (1988).

41 Casjens, S. *et al.* A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete Borrelia burgdorferi. *Molecular microbiology* **35**, 490-516 (2000).

42 Fraser, C. M. *et al.* Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. *Nature* **390**, 580-586, doi:10.1038/37551 (1997).

43 Jutras, B. L., Chenail, A. M. & Stevenson, B. Changes in bacterial growth rate govern expression of the Borrelia burgdorferi OspC and Erp infection-associated surface proteins. *Journal of bacteriology* **195**, 757-764, doi:10.1128/JB.01956-12 (2013).

44 Hood, R. D. *et al.* A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. *Cell host & microbe* **7**, 25-37 (2010).

45 Chou, S. *et al.* Structure of a peptidoglycan amidase effector targeted to Gram-negative bacteria by the type VI secretion system. *Cell Reports* **1**, 656-664, doi:10.1016/j.celrep.2012.05.016 (2012).

46 Ramamoorthi, N. *et al.* The Lyme disease agent exploits a tick protein to infect the mammalian host. *Nature* **436**, 573-577, doi:10.1038/nature03812 (2005).

47 Pal, U. *et al.* TROSPA, an Ixodes scapularis receptor for Borrelia burgdorferi. *Cell* **119**, 457-468, doi:10.1016/j.cell.2004.10.027 (2004).

48 Nadkarni, M. A., Martin, F. E., Jacques, N. A. & Hunter, N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology (Reading, England)* **148**, 257-266 (2002).