J Mol Evol (2004) 59:498–506
DOI: 10.1007/s00239-004-2641-8

JOURNAL OF MOLECULAR EVOLUTION

Molecular Evolution of *Daphnia* Immunity Genes: Polymorphism in a *Gram-Negative Binding Protein* Gene and an α-*2-Macroglobulin* Gene

Tom J. Little,[1] John K. Colbourne,[2] Teresa J. Crease[3]

[1] Institute for Cell, Animal and Population Biology, University of Edinburgh, Kings Buildings, West Mains Rd, Edinburgh EH9 3JT, UK
[2] Center for Genomics and Bioinformatics, Indiana University, 915 East Third Street, Bloomington, IN 47405, USA
[3] Department of Zoology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Received: 4 February 2004 / Accepted: 15 April 2004 [*Reviewing Editor:* Dr. Martin Kreitman]

Abstract. Studies of DNA polymorphism have shown that some immune system genes of mammals and plants are exceptionally diverse, indicating that coevolution between these taxa and their parasites mediates positive selective sweeps and/or balancing selection. The genes of the arthropod immune system remain comparatively unstudied. We isolated two putative immune system genes from the cladoceran crustacean *Daphnia* and examined DNA sequence diversity. For one gene, encoding a putative gram-negative binding protein, we found evidence of only purifying selection, indicating that this gene is under strong functional constraint and that selection acts to eliminate amino acid variation. For another gene, encoding a putative α-2-macroglobulin, we found evidence of positive selection, indicating the possible involvement of this gene in a host–parasite arms race. We discuss the assumed function of these genes and offer speculation regarding which components of the arthropod immune system might experience diversifying adaptive evolution.

Key words: Selection — Arms — Race — Innate immunity — Pathogen — Host — Parasite

Correspondence to: Tom J. Little; *email:* tom.little@ed.ac.uk

Introduction

Reciprocal antagonism might lock host and parasite populations into a coevolutionary process that promotes genetic diversity. This view has been substantiated by molecular studies, which have observed that some vertebrate immune system genes are exceptionally diverse, as are some cell surface antigens of pathogens (Ford 2002; Hughes 1991, 1997; Hurst and Smith 1999; Olson 2002). A comparable situation is exhibited in plant–pathogen systems where some resistance genes and pathogen ellicitor molecules show considerable polymorphism (Stahl and Bishop 2000; Stahl et al. 1999).

Host–parasite coevolution can promote diversity in at least two ways, and these may act together (Bergelson et al. 2001). The first is through balancing selection, where allelic variants are maintained by frequency-dependent or overdominant selection. The maintenance of polymorphism through balancing selection is evident as the deep divergence of alleles at single loci and heterozygosity within populations, as seen in the MHC alleles of vertebrates (Hill et al. 1991, 1992; Hughes and Nei 1992). Second, host–pathogen interactions can result in an arms race, whereby variation is fostered by selection in favor of new pathogen mutants that elude the immune system, followed by selection on hosts for improved ability to recognize and destroy these new variants, followed by a further evolutionary response in the pathogen, and so on. Under an arms race, natural selection proceeds as a series of directional selective sweeps, and this is

evident as an elevated rate of amino acid substitutions in comparisons among populations or species, accompanied by a lack of heterozygosity within populations (Ford 2002; Hurst and Smith 1999; Olson 2002). Both arms races and balancing selection patterns can be contrasted with purifying selection, whereby new mutations perform poorly and are pruned from the population, resulting in gene sequences that show relatively little amino acid polymorphism.

Molecular evidence of selection is scarce for the immunity genes of arthropods, although there has been considerable progress in elucidating the genetic and functional basis of arthropod defense systems (e.g., Christophides et al. 2002; Hoffmann et al. 1999). The arthropod immune system (summarized in Aderem and Ulevitch 2000; Du Pasquier 2001; Hoffmann et al. 1999; Janeway and Medzhitov 2002; Roitt 1997) appears to be relatively nonspecific (but see Carius et al. 2001; Dimopoulos et al. 1998) and lacks memory (but see Arala-Chaves and Sequeira 2000; Kurtz and Franz 2003; Little et al. 2003; Moret and Schmid-Hempel 2001). Arthropod immune systems comprise cellular responses (phagocytes, etc.), a variety of antibacterial peptides (reviewed in Engstron 1999), and a phenoloxidase cascade (e.g., Soderhall and Cerenius 1998) that produces melanin (used, for example, to encase parasitoids). Other important components include nitric oxide synthase (Dimopoulos et al. 1998), clotting reactions (Iwanaga 2002), and serine protease inhibitors (Armstrong and Quigley 1999; Kanost 1999; Oduol et al. 2000). So far, immune-related genes have been examined for the footprint of natural selection only in *Drosophila*. Studies of the gene encoding Relish (a transcription factor of antimicrobial peptides), as well as a genomewide study comparing *D. melanogaster* to *D. simulans*, indicated that immune system genes are subject to positive selection to a greater extent than are other parts of the genome (Begun and Whitley 2000; Schlenke and Begun 2003). However, mixed results were obtained by pioneering studies of particular antimicrobial peptides (Clark and Wang 1997; Date et al. 1998; Lazzaro and Clark 2003; Ramos-Onsins and Aguade 1998), and a study of recognition molecules in the peptidoglycan family gave evidence of purifying selection (Jiggins and Hurst 2003).

Here, we present the first data on molecular evolution of putative immune system genes from a crustacean, the cladoceran *Daphnia*. By analogy with plant and vertebrate systems, host proteins that recognize pathogens are prime candidates for diversifying selection, and we therefore targeted such genes. Specifically, we present DNA polymorphism data from a putative *gram-negative binding protein* (*GNBP*) gene and a putative *α-2-macroglobulin* (*A2M*) gene. In other invertebrates, the products of both genes are pattern recognition receptors (Janeway and Medzhitov 2002; Medzhitov and Janeway 1997), but they differ in the pathogen-associated molecular patterns they recognize; GNBPs recognize polysaccharides on pathogen surfaces, while A2Ms bind pathogen serine proteases. Our analyses of polymorphism were largely directed toward the detection of positive directional selection. Our motivation for this study rests on the hypothesis that variation in infection rates and disease symptoms are attributable, in part, to genetic variation maintained through pathogen-mediated selection. Therefore, the identification of host genetic polymorphism involved in arms races in natural populations will aid understanding of variation in disease prevalence.

Materials and Methods

Initial Gene Capture and Characterization

From GenBank we obtained DNA sequences of arthropod immunity genes and aligned them using the MegAlign v4.03 program of the Lazergene software package. Amino acid and nucleotide alignments were made principally with the horseshoe crab (*Limulus* sp.), the crayfish (*Pacifastacus leniusculus*), several species of *Drosophila*, and *Anopheles* mosquitos. From these alignments, we identified conserved regions of the genes and designed degenerate oligonucleotide primers that had the potential to amplify DNA fragments in *Daphnia* via the polymerase chain reaction (PCR). This approach proved successful for two *Daphnia* genes: a putative *α-2-macroglobulin* (*A2M*) and a putative *gram-negative binding protein* (*GNBP*). Numerous primer combinations were tested for each gene on *Daphnia* DNA, but amplified products were obtained using the following pairs of primers: an *A2M* forward primer based on the amino acid sequence FQPF (5′ TTC CAG CCN TTC TTC NT), an *A2M* reverse primer based on the the amino acid sequence CGEQ (5′ CCA TGT TTT GYT CNC CAC), a *GNBP* forward primer based on the amino acid sequence GGGN (5′ GGN GGN GGN AAY TGG GAR TTC CA), and a *GNBP* reverse primer based around the amino acid sequence MAPF (5′ AAC TTY TGR TCR AAN GGN GCC AT).

In each case, the degenerate primers amplified three or four products of variable molecular weights. Each product was electrophoresed and excised from an agarose gel, purified using the Qiagen Gel Extraction kit, cloned using the TOPO TA 2.1 vector (Invitrogen Life Technologies), and then sequenced with the ABI Prism BigDye Terminator Cycle Sequencing Kit, Version 3.0. Sequences were compared to the nonredundant protein sequence database using the Blastx alignment program (Altschul et al. 1997) to reveal similarity to genes archived in GenBank (NCBI). From those sequences that resembled published immunity genes with a high probability, we designed *Daphnia*-specific PCR primers to amplify genomic DNA from a range of species and populations. When using these primers for PCR, electrophoresis of the amplicons produced a single discrete band in each case. Primers for the *A2M* locus, which produced a 650-bp product, were 5′ TTGCCTTACTCGGTGAAACG and 5′ GCAGGTCCCATGAG GTCACC. This region of the *Daphnia A2M* is nearby the sequence coding for the thiol ester bond common to A2M proteins. Indeed the degenerate reverse primer used to initially capture the gene fragment within *Daphnia* was based on the amino acid sequence coding for the highly conserved thiol ester bond (CGEQ). Thus, this fragment is hereafter referred to as the "thiol ester" fragment. *Daphnia*-specific primers for the *GNBP* locus, which produced a

650-bp PCR product, were 5′ GGAGTTCCAGTATTACGACAA and 5′ TCA ACA ATG TCA ATC TCT CC.

With these specific primers, we gathered sequence information from three populations of *D. pulex* (subgenus *Daphnia*), from five populations of *D. rosea* (subgenus *Hyalodaphnia*), and from seven populations of *D. magna* (subgenus *Ctenodaphnia*) collected mostly in Southern Scotland, U.K. The sole exception was the inclusion of a population of *D. magna* from Northern Germany. Thus, 15 ponds (populations) were studied in total and our data analyses include one sequence from each population. It was possible to study only a subset of these populations for *GNBP*; one sequence from each of 11 populations was analyzed, 7 from *D. magna* and 2 each from *D. pulex* and *D. rosea*. In all cases DNA was extracted from freshly collected live individuals or from those that had been frozen immediately after collection. For both genes, we chose two populations from which to sequence an additional five individuals to test for variation within a pond (population). Polymorphism within ponds was not detected and we therefore restricted our analyses to a single representative from each pond.

Genomic DNA from individual *Daphnia* was extracted using a modification of the CTAB method (Colbourne et al. 1998). PCR amplification conditions for all *Daphnia*-specific primers used a 52°C annealing temperature for 30 s and an extension time of 45 s. Magnesium chloride concentration was 2.5 m*M* in all reactions. Sequencing utilized the BIGDYE sequencing reaction kit and the ABI automated sequencing facility at the University of Edinburgh, Institute of Cell, Animal and Population Biology. Full details of PCR and sequencing protocols are available from the corresponding author.

Chromosome Walking

To analyze regions of *A2M* not recovered in our initial PCR amplifications, i.e., outside of the "thiol ester" fragment, we used Invitrogen TOPO Walker technology to "walk" along the *Daphnia* chromosome. Efforts to sequence out toward the 3′ end of the coding strand gained 1200 bp, but these were not analyzed for polymorphism. Efforts to sequence out toward the 5′ end of the coding strand successfully revealed an additional 1400 bp further into the gene. New primers were then developed to amplify an ~1000-bp fragment within this region closer to the 5′ end for additional analyses of polymorphism. For *D. magna*, successful PCR amplification was obtained from the following: a forward primer based on the amino acid sequence MARG (5′ GGC GAC GTG ATG CAT ACA GG) and a reverse primer based on the amino acid sequence PHTIT (5′ CAT GTG GTG ATC GTG TGA GG). For *D. pulex* and *D. rosea*, successful PCR amplification was obtained from the following: a forward primer based on the amino acid sequence GYSVVD (5′ GAA TTT GAA AGG TAG CGC TGG) and a reverse primer based on the amino acid sequence PHTIT (5′ CAT GTG GTA ATC GTG TGA GG). This PCR fragment likely included sequence encoding the "bait" region of the A2M. We hereafter refer to this PCR product as the "bait" fragment.

Reverse Transcription (rt)PCR

To test for the expression of our study genes, and to acquire sequences to delimit intron/exon boundaries, we performed rtPCR on each of the gene fragments. Single individuals of *D. pulex*, *D. rosea*, or *D. magna* were allowed to propagate clonally to abundance in 1-L jars with approximately 1 cm of natural pond sediment at the bottom. Pond sediments typically contain sporebanks of parasites, and by keeping *Daphnia* in contact with these we aimed to induce the transcription of immunity genes. After about 2 weeks of exposure to sediment, total RNA was extracted from ~100 individuals of each species using the RNeasy plant extraction kit (Qiagen). Reverse transcription of RNA to cDNA was performed with the omniscript rtPCR kit (Qiagen) and using the complement of the *Daphnia*-specific primers listed above, as well as an oligo(dT) (18mer). Standard PCR was then carried using the same conditions that amplified genomic DNA.

Summary of Main Analyses

Both *A2Ms* and *GNBPs* exist as multigene families within insects (Christophides et al. 2002). Whether this is the case for *Daphnia* is important for the present study because our analyses of polymorphism and divergence assume that we are comparing orthologues. We are confident that *Daphnia A2M* exists as a single gene copy. First, although PCR with the highly degenerate primers used to initially capture this gene amplified a number of fragments, all were sequenced and only one of these was revealed to be *A2M*. Second, we have used PCR to probe an arrayed fourfold coverage *Daphnia* cosmid library (see http://daphnia.cgb.indiana.edu/tools/) with the *Daphnia*-specific *A2M* primers and identified a single cosmid clone that contained *A2M*. Thus, for A2M we are reasonably certain that our comparisons among populations and species involved orthologues. For *GNBP*, the highly degenerate primers used to initially capture this gene amplified a number of fragments, all were sequenced and three of these were revealed to be *GNBPs*. However, there were clear size differences between the gene copies, and the *Daphnia*-specific primers we developed amplified only one (the shortest) of these. We failed to amplify GNBP from the cosmid library. Thus, for GNBP, we are less certain that our comparisons involved orthologues, and the relatively high amount of silent divergence among species, compared to A2M (see Results), may perhaps further fuel this uncertainty.

Basic polymorphism data, e.g., the nucleotide diversity (π) for each gene, was generated using the program DNAsp (Roza and Rozas 1999). Beyond this, we were primarily interested in recovering signatures of positive selection. An important concept underlying such analyses is the notion that positive selection on gene sequences results in rates of nonsynonymous nucleotide substitution (substitutions which result in an amino acid substitutions) that exceed levels of synonymous substitution (which do not result in amino acid substitutions). We estimated variables K_a (the number of nonsynonymous substitutions per nonsynonymous site) and K_s (the number of synonymous substitutions per synonymous site) using DNAsp. An approximation of the neutral rate of nucleotide divergence is obtained from K_s, thus K_a/K_s ratios $\ll 1$ indicate that the gene is under purifying selection, but higher values may be caused by positive selection. Although simply examining K_a/K_s ratios provides a notoriously conservative estimate of positive selection (e.g., Yang et al. 2000), we declined to use more powerful analyses (e.g., the maximum likelihood procedures in PAML) of K_a/K_s ratios because many of our data were drawn from multiple populations of the same species, and the occurrence of recombination would violate a central assumption of these phylogenetic-based analyses (Anisimova et al. 2000).

We performed MacDonald–Kreitman (MK) tests (1991) which detect selection by finding statistical discrepancy in the relative level of synonymous and nonsynonymous substitutions within and between species. The basic assumption of MK analysis is that the ratio of replacement to synonymous substitution between species will be the same as the ratio of replacement to synonymous polymorphism within species *if* divergence and polymorphism are due solely to random drift acting on neutral mutations. The use of the MK test overcomes issues concerning the ability to distinguish sites of adaptive evolution from sites evolving relatively quickly because of low selective constraints. To test for selection by comparing polymorphism between the two genes studied, as well as that between the two studied regions *A2M*, we employed the Hudson,

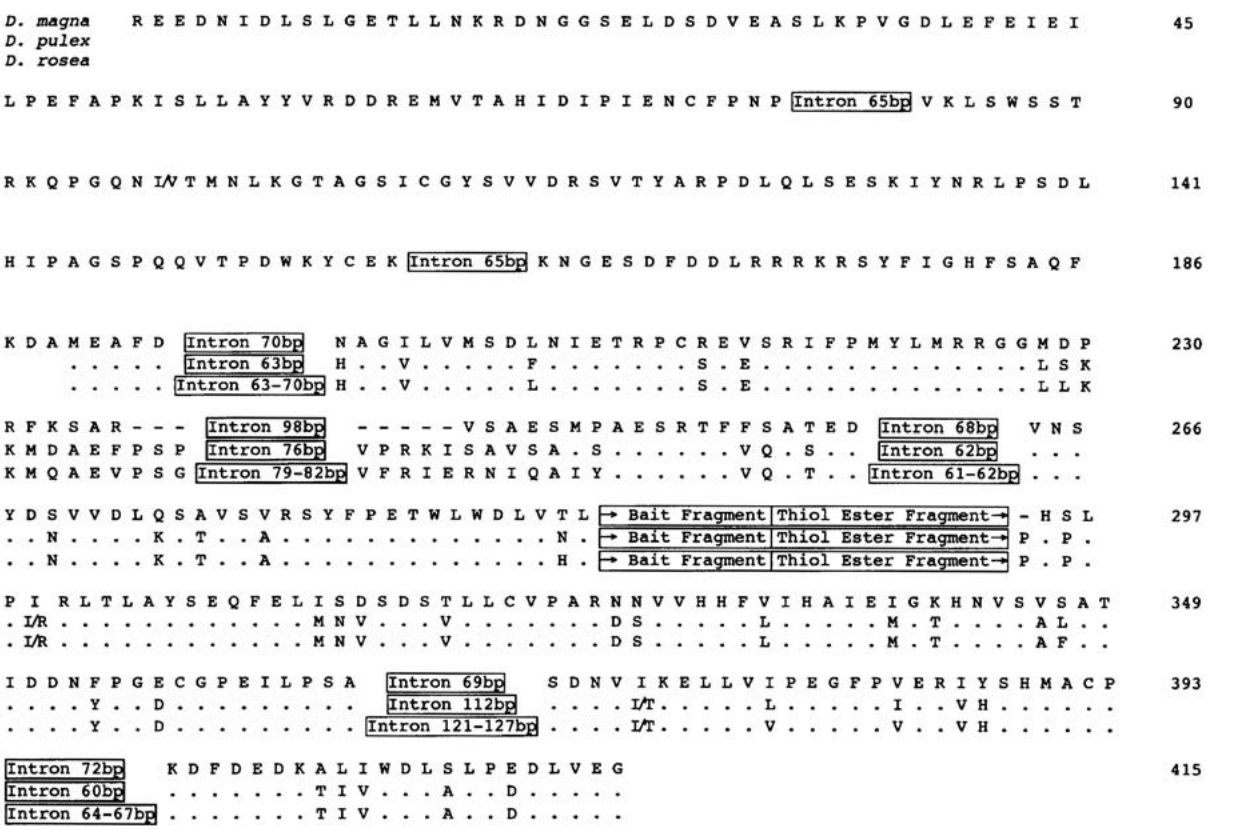

Fig. 1. The deduced amino acid sequences from DNA sequences presumed to encode α-2-macroglobulin from *Daphnia magna*, *D. pulex*, and *D. rosea*. Intron/exon boundaries were determined by comparing the DNA sequence obtained through sequencing fragments obtained by rtPCR to sequences obtained from genomic DNA. Details on the most polymorphic region, spanning amino acid residues 229 to 261, are given in Fig. 3.

Kreitmam, and Aquade (HKA) test (1987) as implemented in DNAsp.

Results

General Patterns of Polymorphism and Divergence

For *A2M*, we obtained a total of 1735 bp of sequence information from *D. magna* (Fig. 1). Of these nucleotides, 1224 bp was from the "bait" fragment and the remaining bases were from the "thiol ester" fragment. Slightly less sequence information was obtained for *D. pulex* and *D. rosea* (Fig. 1): 540 bp in the bait region and 564 bp in the thiol ester fragment (Fig. 1). For *GNBP*, we obtained approximately 450 bp of sequence information from each of *D. magna* and *D. pulex* (Fig. 2). For *D. rosea* we obtained an additional 279 bp, but comparable sequence data for this gene were unavailable from the other species for analyses of polymorphism. Fragments produced by rtPCR were shorter for both genes than fragments from PCR of genomic DNA due to the presence of introns.

As expected from studies of variation in the mitochondrial 12S rRNA gene (Colbourne and Hebert 1996), the sequences from *D. magna* were highly divergent from those of the other two species used in this study. We therefore declined to compare, especially for the tests of selection (below), *D. magna* to the other species due to the possibility that saturated sites would bias the estimated levels of substitution. By contrast, *D. pulex* and *D. rosea* showed sufficient nucleotide similarity that it seemed reasonable to compare these two species (Figs. 1 and 2). For example, no indels were found between *D. pulex* and *D. rosea* within coding regions, and introns were easily aligned. Across all sites, *D. pulex* and *D. rosea* showed 5% sequence divergence compared to over 20% between *D. magna* and either *D. pulex* or *D. rosea*.

General measures of polymorphism (all species) and divergence (*D. pulex* vs *D. rosea* only) are given in Tables 1 and 2. Comparing among populations and species, nucleotide polymorphism in *GNBP* averaged less than half of that found in *A2M* (Table 2). Although the overall divergence between species was higher for *GNBP* than it was for *A2M*, this was almost exclusively due to high divergence at silent or noncoding sites (Table 2).

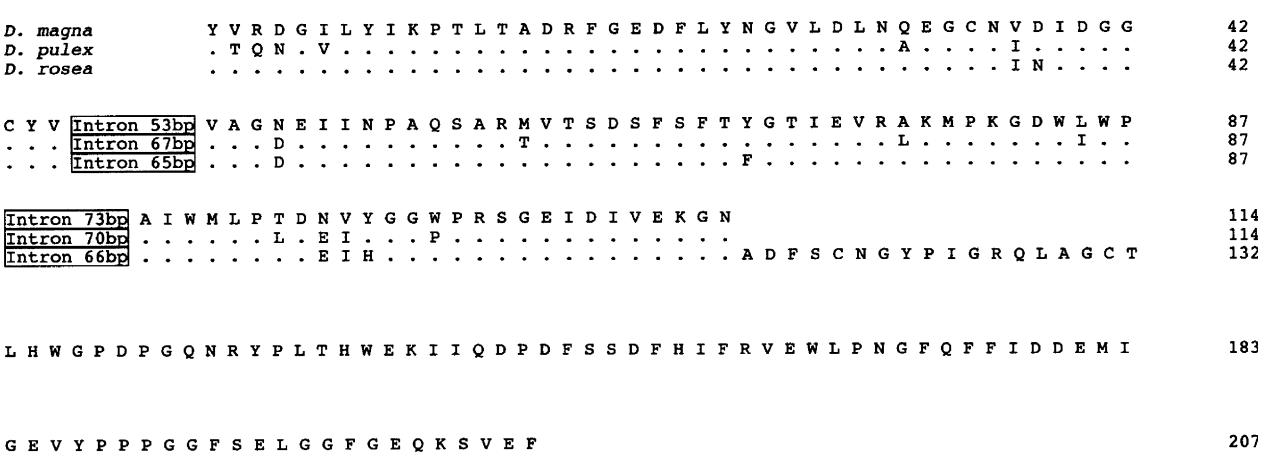

Fig. 2. The deduced amino acid sequences from DNA sequences presumed to encode a gram-negative binding protein from *Daphnia magna*, *D. pulex*, and *D. rosea*. Intron/exon boundaries were determined by comparing the DNA sequence obtained through sequencing fragments obtained by rtPCR to sequences obtained from genomic DNA. Comparison of *D. rosea* and *D. pulex* revealed no amino acid substitutions with more than one nonsynonymous substitution.

Table 1. Polymorphism parameters for DNA sequences from an α-*2-macroglobulin* (*A2M*) gene and a *gram-negative binding protein* (*GNBP*) gene from 15 populations of European *Daphnia pulex*, *D. rosea*, and *D. magna*

					Polymorphic sites			
Gene	Species	Total length (bp)	Exon length (bp)	*n*	Total	Noncoding	Coding—silent	Replacement
A2M								
Total	*D. magna*	1735	1224	7	9	6	2	1
	D. pulex	1104	693	3	6	3	1	2
	D. rosea	1104	693	5	33	30	7	3
TE	*D. magna*	508	366	7	4	3	1	0
	D. pulex	564	366	3	4	2	0	2
	D. rosea	564	366	5	11	4	4	3
Bait	*D. magna*	1227	858	7	5	3	1	1
	D. pulex	540	327	3	2	1	1	0
	D. rosea	540	327	5	9	6	3	0
GNBP	*D. magna*	468	341	7	0	0	0	0
	D. pulex	446	309	2	2	1	1	0
	D. rosea	440	309	2	1	1	0	0

Tests for Selection

Considering only coding regions, an HKA test to compare the level of polymorphism in *GNBP* to that in the entire *A2M* sequence produced a marginally significant result ($\chi^2 = 3.65$, df = 1, $p = 0.056$). An HKA test comparing the two regions of *A2M* to one another revealed that the observed patterns of polymorphism are inconsistent with a neutral model of evolution ($\chi^2 = 4.45$, df = 1, $p = 0.035$). This indicates that the evolutionary pressure differs between the two regions of *A2M*. Therefore, in the other tests for selection below, we treat the two *A2M* regions separately.

The MK test, performed on the *A2M* bait fragment from *D. pulex* and *D. rosea*, indicated that there is a significant difference between the ratio of replacement and synonymous differences within and between species (Fisher's exact test, df = 1, $p = 0.01$; Table 3), which is consistent with adaptive fixation of amino acid differences between species. Similar analyses of the *A2M* thiol ester fragment produced a nonsignificant result ($G = 0.64$, df = 1, $p = 0.77$). An MK analysis (Table 3) of *GNBP* also did not reveal evidence of positive selection.

Of the 28 replacement differences between *D. pulex* and *D. rosea*, 25 were in the bait fragment and 23 occurred within a relatively short stretch of nucleotides (Fig. 3). Nine sites (shaded in Fig. 3) in the "bait" fragment correspond to amino acid differences representing two nonsynonymous nucleotide substitutions, whereas all other amino acid differences separating *D. pulex* and *D. rosea* involved just one nonsynonymous substitution. For eight of these nine sites, both substitutions are nonsynonymous in whichever order they occur, while for one site (amino acid 229, serine/leucine), the two-step evolutionary pathway could have travelled a step that resulted in a synonymous substitution. In this case, there would be only 24 nonsynonymous and 11 synonymous substitutions entered in Table 3. However, even this more conservative test yielded a significant result (Fisher's exact test, $p = 0.015$).

Table 2. Nucleotide diversity (π) and divergence of DNA sequences from an *α-2-macroglobulin* gene (*A2M*) and a *gram-negative binding protein* (*GNBP*) gene from 15 populations of European *Daphnia pulex*, *D. rosea*, and *D. magna*

		π			
Gene	Species	Total	Noncoding	Coding—silent	Replacement
Polymorphism					
A2M					
Total	*D. magna*	0.0020	0.0044	0.0031	0.0004
	D. pulex	0.0038	0.0054	0.0042	0.0026
	D. rosea	0.0156	0.0305	0.0255	0.0026
TE	*D. magna*	0.0034	0.0086	0.0062	0.0000
	D. pulex	0.0050	0.0079	0.0000	0.0048
	D. rosea	0.0107	0.0123	0.0265	0.0050
Bait	*D. magna*	0.0014	0.0028	0.0017	0.0005
	D. pulex	0.0025	0.0033	0.0088	0.0000
	D. rosea	0.0104	0.0187	0.0243	0.0000
GNBP	*D. magna*	0.0000	0.0000	0.0000	0.0000
	D. pulex	0.0045	0.0072	0.0136	0.0000
	D. rosea	0.0023	0.0076	0.0000	0.0000
Divergence					
A2M					
Total	*D. pulex* vs *D. rosea*	0.0932	0.1245	0.1257	0.0636
TE	*D. pulex* vs *D. rosea*	0.0939	0.2246	0.1277	0.0158
Bait	*D. pulex* vs *D. rosea*	0.1244	0.1296	0.1236	0.1208
GNBP	*D. pulex* vs *D. rosea*	0.2123	0.3870	0.4710	0.0737

Note. Two separate fragments (the "bait" and "thiol ester" fragments) of *A2M* were studied.

Table 3. Summary of polymorphism data used in the McDonald–Kreitman test comparing *Daphnia pulex* and *D. rosea*

	Intraspecific polymorphism		Interspecific fixed differences		
	Synonymous coding	Replacement	Synonymous coding	Replacement	*p*
A2M					
Total	8	3	19	28	0.11
TE	4	3	9	3	0.77
Bait	4	0	10	25	0.01
GNBP	1	0	26	19	1.00

Note. The relative ratio of replacement to synonymous substitutions was compared within and between species using a contingency table analysis and a *G*-test or Fisher's exact test.

The average number of synonymous substitutions per synonymous site (K_s) between species of *D. pulex* and *D. rosea* in the entire *A2M* region was 0.13, while the average number of nonsynonymous substitutions per nonsynonymous site (K_a) was 0.065, giving a K_a:K_s ratio of 0.5. There were, again, apparent differences in the K_a:K_s ratios between the "bait" fragment to the "thiol ester" fragment. Between *D. pulex* and *D. rosea*, the average K_a:K_s ratio in the thiol ester region was 0.14. Yet this ratio was over six times higher in the bait fragment, 0.95. The average K_a:K_s ratio at *GNBP* was 0.05.

Discussion

This study documents the first DNA polymorphism data from two putative immune system genes of a crustacean. One of these genes, a putative *GNBP*, showed evidence of purifying selection. The other gene, a putative member of the *A2M* family of serine protease inhibitor genes, was studied in two regions. The first region was near the site coding for the thiol ester domain which, following proteolytic attack, is involved in irreversible inhibition of the pathogen serine protease. The "thiol ester" fragment of *Daphnia* A2M showed evidence of purifying selection. By contrast, the "bait" fragment of A2M appeared to be under positive selection, as there was a high number of amino acid substitutions among species. The ratio of replacement to silent substitutions between *D. pulex* and *D. rosea* was significantly higher than expected from levels of within-species polymorphism (by the MK test), and nine codons in the bait fragment showed multiple replacement substitutions. The bait fragment was so named because it encodes amino acid sequence analogous to the A2M bait region that, in functional studies of other taxa, is the target of

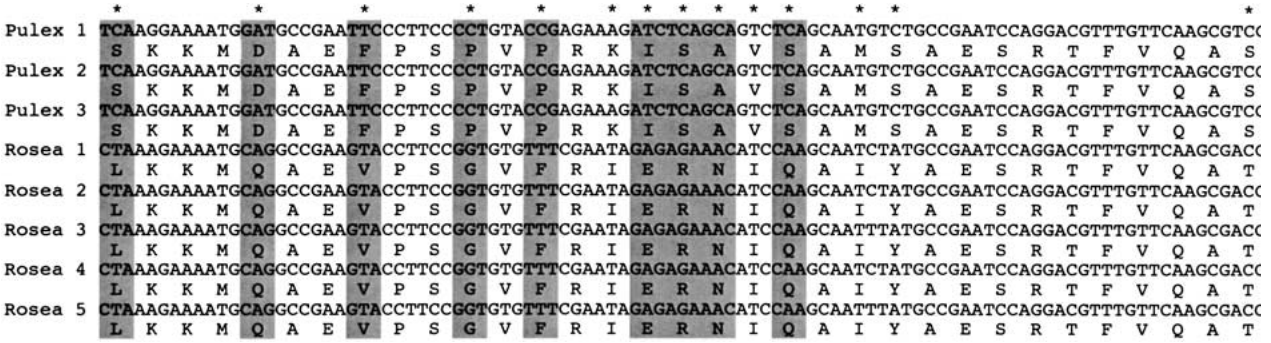

Fig. 3. Nucleotide and deduced amino acid sequences from 99 base pairs of the α-*2-macroglobulin* bait fragment as sequenced in eight populations of *D. pulex* and *D. rosea*. Asterisks denote fixed differences at nonsynonymous nucleotide sites. Shaded codons are amino acid replacements represented by two nonsynonymous nucleotide substitutions.

pathogen serine proteases (Armstrong and Quigley 1999).

An important concept in the current understanding of innate immunity is that of the PAMP–PRR interaction (Janeway and Medzhitov 2002; Medzhitov and Janeway 1997). We believe that population genetic predictions are possible from the PAMP/PRR paradigm and that the present data, although hardly a thorough test, are consistent with these predictions. PAMPs are pathogen-associated molecular patterns, and PRRs are the host-produced pattern recognition receptors that detect them. Many PAMPs are conserved molecules, often polysaccharides, that are essential for the survival of the pathogen and, as such, cannot easily be modified to conceal their recognition by the host. If PAMP escape mutants are unlikely, then PRRs are also unlikely candidates for an arms race. Thus, *Daphnia* GNBP appears to be under purifying selection and shows, like GNBPs from other organisms, particular similarity to the putative catalytic region of B-1-3 glucanases. This region is thought to be important for binding of polysaccharides (Yahata et al. 1999). Other work in our lab concerns a peptidoglycan-recognizing protein from *Anopheles* mosquitoes, which also shows evidence of purifying selection (unpublished data).

Immune system components involved in protein–protein interactions seem more promising candidates for an arms race. Indeed, in antagonistic interactions involving vertebrate hosts, this is precisely where stunning diversity has been observed; for example, in the cell surface proteins of pathogens and in the host proteins that recognize them. A2Ms are involved in the recognition of pathogen products, but unlike many arthropod PRRs, they recognize pathogen proteins. A2Ms recognize and bind pathogen serine proteases, which are both important virulence factors and are used as part of pathogen strategies to gain entry into hosts (Han et al. 2000). We tentatively suggest that the bait region of *Daphnia* A2M shows evidence of positive selection because it has been subject to a series of directional selective sweeps (an arms race) in which host A2M evolves to inhibit parasite serine proteases, and parasite serine proteases evolve to become unrecognizable by hosts. The lack of polymorphism within species probably indicates that balancing selection is not prevalent at this gene.

Serpins, also serine protease inhibitors, provide another example of elevated amino acid evolution based on studies of mammals and of parasitic nematodes (Barbour et al. 2002; Gill and Mock 1985; Zang and Maizels 2001). Notably, high polymorphism is observed only in the reactive centers which bind proteases, of the serpins. Host serpins may interact with pathogen serine proteases in a manner analogous to A2Ms (Barbour et al. 2002; Kanost 1999) or regulate host serine proteases involved in immune cascades (Oduol et al. 2000; Soderhall and Cerenius 1998). Given the latter function, adaptive evolution of host serpins might suggest that these sites are the target of manipulation strategies by pathogens, which will evade the immune response if host serpins are prevented from performing a role in the immune response. Another example of an arthropod immune gene showing evidence of elevated amino acid replacement comes from a transcription factor: the NF-κB/IκB protein Relish from *Drosophila*. In mammals, IκB proteins are inhibitors of the NF-κB proteins which upregulate components of the innate immune system. The functional details of the NF-κB/IκB protein in *Drosophila* are less clear, but an elevated amino acid replacement rate in the IκB region has led to speculation that this sequence is the target of manipulation strategies by pathogens (Begun and Whitley 2000).

This discussion is speculative, as presently there are too few available data on polymorphism in arthropod immune system genes to draw general conclusions and to guide future studies. Indeed, while genome projects and functional analyses of immunity genes are rapidly elucidating the genetic basis of

arthropod host–pathogen interactions (Holt et al. 2002; Leulier et al. 2003; Oduol et al. 2000), the tempo and mode of selection on immune-related genes have hardly been studied. For some taxa, e.g., *Anopheles* or *Drosophila*, choosing candidate genes for population genetic analyses is daunting, given the amount of genomic information available. The predictions drawn from functional understanding of innate immunity might help with this choice. This situation is, of course, opposite to the problem we face in finding candidate genes in *Daphnia*, where genomic information is limited. Ultimately, however, it will be necessary to assess the consequences of polymorphism in immune system genes by associating genotypic with phenotypic variation in the laboratory and by establishing links with field patterns. In this regard we feel that *Daphnia* ought to remain the target of studies such as the present one, given their amenability to both laboratory and field studies of parasitism (e.g., Little and Ebert 2000).

Acknowledgments. Thanks to Mike Lynch and members of the Lynch Lab (Eugene and Bloomington) for assisting TL during the early phase of this work. In Edinburgh, laboratory support was provided by Andrew Read, Brian Chan, Josephine Pemberton, and members of Pemberton lab. This work was financially supported by The Wellcome Trust UK and by the School of Biology, University of Edinburgh. Peter Andolfatto, Brian Charlesworth, Stephen Wright, Mark Blaxter, and Francis Jiggins offered advice on a draft version of this paper. All authors acknowledge the assistance of the *Daphnia* Genomics Consortium framework. A2M sequences used in this study correspond to GenBank accession numbers AY540086–AY540100, and GNBP sequences used in this study correspond to GenBank accession numbers AY540101–AY540103.

References

Aderem A, Ulevitch RJ (2000) Toll-like receptors in the induction of the innate immune response. Nature 406:782–787

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Anisimova A, Nielsen R, Yang Z (2000) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–1236

Arala-Chaves M, Sequeira T (2000) Is there any kind of adaptive immunity in invertebrates? Aquaculture 191:247–258

Armstrong PB, Quigley JP (1999) alpha(2)-macroglobulin: An evolutionarily conserved arm of the innate immune system. Dev Comp Immunol 23:375–390

Barbour KW, Goodwin RL, Guillonneau F, Wang YP, Baumann H, Berger FG (2002) Functional diversification during evolution of the murine alpha(1)-proteinase inhibitor family: Role of the hypervariable reactive center loop. Mol Biol Evol 19:718–727

Begun DJ, Whitley P (2000) Adaptive evolution of relish, a *Drosophila* NF-kappa B/I kappa B protein. Genetics 154:1231–1238

Bergelson J, Kreitman M, Stahl EA, Tian DC (2001) Evolutionary dynamics of plant R-genes. Science 292:2281–2285

Carius H-J, Little TJ, Ebert D (2001) Genetic variation in a host-parasite association: Potential for coevolution and frequency dependent selection. Evolution 55:1136–1145

Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, Brey PT, Collins FH, Danielli A, Dimopoulos G, Hetru C, Hoa N, Hoffmann JA, Kanzok SM, Letunic I, Levashina EA, Loukeris TG, Lycett G, Meister S, Michel K, Muller HM, Osta MA, Paskewitz SM, Reichhart JM, Rzhetsky A, Troxler L, Vernick KD, Vlachou D, Volz J, von Mering C, Xu JN, Zheng LB, Bork P, Kafatos FC (2002) Immunity-related genes and gene families in *Anopheles gambiae*. Science 298:159–165

Clark AG, Wang L (1997) Molecular population genetics of Drosophila immune system genes. Genetics 147:713–724

Colbourne JK, Crease TJ, Weider LJ, Hebert PDN, Dufresne F, Hobaek A (1998) Phylogenetics and evolution of a circumarctic species complex (Cladocera: Daphnia pulex). Biol J Linn Soc 65:347–365

Colbourne JK, Hebert PDN (1996) The systematics of North American *Daphnia* (Crustacea: Anomopoda): a molecular phylogenetic approach. Philos Trans R Soc Lond Ser 351:349–360

Date A, Satta Y, Takahata N, Chigusa SI (1998) Evolutionary history and mechanism of the *Drosophila* cecropin gene family. Immunogenetics 47:417–429

Dimopoulos G, Seeley D, Wolf A, Kafatos FC (1998) Malaria infection of the mosquito *Anopheles gambiae* activates immune-responsive genes during critical transition stages of the parasite life cycle. EMBO J 17:6115–6123

Du Pasquier L (2001) The immune system of invertebrates and vertebrates. Comp Biochem Physiol B Biochem Mol Biol 129:1–15

Engstron Y (1999) Induction and regulation of antimicrobial peptides in *Drosophila*. Dev Comp Immunol 23:345–358

Ford MJ (2002) Applications of selective neutrality tests to molecular ecology. Mol Ecol 11:1245–1262

Gill DE, Mock BA (1985) Ecological and evolutionary dynamics of parasites: The case of *Trypanosoma diemyctyli* in the red-spotted newt *Notophtalmus viridescens*. In: Rollinson D, Anderson RM (eds) Ecology and genetics of host–parasite interactions. Linnean Society of London, London, pp 157–183

Han YS, Thompson J, Kafatos FC, Barillas-Mury C (2000) Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: The time bomb theory of ookinete invasion of mosquitoes. EMBO J 19:6030–6040

Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM (1991) Common West African HLA antigens are associated with protection from severe malaria. Nature 352:595–600

Hill AVS, Kwiatkowski D, McMichael AJ, Greenwood BM, Bennett S (1992) Maintenance of Mhc polymorphism—Reply. Nature 355:403–403

Hoffmann JA, Kafatos FC, Janeway CA, Ezekowitz RAB (1999) Phylogenetic perspectives in innate immunity. Science :1313–1318

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JMC, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai ZW, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, Cai S, Center A, Chatuverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu ZP, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke ZX, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo

NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao HG, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun JT, Thomasova D, Ton LQ, Topalis P, Tu ZJ, Unger MF, Walenz B, Wang AH, Wang J, Wang M, Wang XL, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang HY, Zhao Q et al. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298:129

Hudson RR, Kreitman M, Aguade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159

Hughes AL (1991) Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): Evidence for positive selection on immunogenic regions. Genetics 127:345–353

Hughes AL (1997) Rapid evolution of immunoglobulin superfamily C2 domains expressed in immune system cells. Mol Biol Evol 14:1–5

Hughes AL, Nei M (1992) Maintenance of Mhc polymorphism. Nature 355:402–403

Hurst LD, Smith NGC (1999) Do essential genes evolve slowly? Curr Biol 9:747–750

Iwanaga S (2002) The molecular basis of innate immunity in the horseshoe crab. Curr Opin Immunol 14:87–95

Janeway CA, Medzhitov R (2002) Innate immune recognition. Annu Rev Immunol 20:197–216

Jiggins FM, Hurst GDD (2003) The evolution of parasite recognition genes in the innate immune system: Purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. J Mol Evol 57:598–605

Kanost MR (1999) Serine proteinase inhibitors in arthropod immunity. Dev Comp Immunol 23:291–301

Kurtz J, Franz K (2003) Evidence for memory in invertebrate immunity. Nature 425:37–38

Lazzaro BP, Clark AG (2003) Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. Mol Biol Evol 20:914–923

Leulier F, Parquet C, Pili-Floury S, Ryu JH, Caroff M, Lee WJ, Mengin-Lecreulx D, Lemaitre B (2003) The *Drosophila* immune system detects bacteria through specific peptidoglycan recognition. Nature Immunol 4:478–484

Little TJ, Ebert D (2000) The cause of parasitic infection in natural populations of *Daphnia*: The role of host genetics. Proc R Soc Lond Ser B Biol Sci 267:2037–2042

Little TJ, O'Connor B, Colegrave N, Watt K, Read AF (2003) Maternal transfer of strain-specific immunity in an invertebrate. Curr Biol 13:489–492

McDonald J, Kreitman M (1991) Adaptive protein evolution at the *adh* locus in *Drosophila*. Nature 351:652–654

Medzhitov R, Janeway CA (1997) Innate immunity: The virtues of a nonclonal system of recognition. Cell 91:295–298

Moret Y, Schmid-Hempel P (2001) Immune defence in bumble-bee offspring. Nature 414:506–506

Oduol F, Xu JN, Niare O, Natarajan R, Vernick KD (2000) Genes identified by an expression screen of the vector mosquito *Anopheles gambiae* display differential molecular immune response to malaria parasites and bacteria. Proc Nat Acad Sci USA 97:11397–11402

Olson S (2002) Seeking the signs of selection. Science 298:1324–1325

Ramos-Onsins S, Aguade M (1998) Molecular evolution of the Cecropin multigene family in *Drosophila*: Functional genes vs. pseudogenes. Genetics 150:157–171

Roitt IM (1997) Essential immunology. Blackwell Science, Oxford

Rozas J, Rozas R (1999) DNAsp version 3: and integrated program for population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Schlenke TA, Begun DJ (2003) Natural selection drives *Drosophila* immune system evolution. Genetics 164:1471–1480

Soderhall K, Cerenius L (1998) Role of the prophenoloxidase-activating system in invertebrate immunity. Curr Opin Immunol 10:23–28

Stahl EA, Bishop JG (2000) Plant-pathogen arms races at the molecular level. Curr Opin Plant Biol 3:299–304

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*. Nature 400:667–671

Yang Z, Nielsen R, Goldman N, Pedersen A-MK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Zang XX, Maizels RM (2001) Serine proteinase inhibitors from nematodes and the arms race between host and pathogen. Trends Biochem Sci 26:191–197