Current Biology 20, 1122–1127, June 22, 2010 ©2010 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2010.04.050

Report

Hydractinia Allodeterminant *alr1* Resides in an Immunoglobulin Superfamily-like Gene Complex

Sabrina F.P. Rosa,[1,4,5] Anahid E. Powell,[1,7] Rafael D. Rosengarten,[2] Matthew L. Nicotra,[5] Maria A. Moreno,[2] Jane Grimwood,[6] Fadi G. Lakkis,[5] Stephen L. Dellaporta,[2] and Leo W. Buss[1,3,*]

[1]Department of Ecology and Evolutionary Biology

[2]Department of Molecular, Cellular and Developmental Biology

[3]Department of Geology and Geophysics

Yale University, New Haven, CT 06511, USA

[4]Institute for Molecular Biology and Medicine, Université Libre de Bruxelles, 6041 Gosselies, Belgium

[5]Thomas E. Starzl Transplantation Institute, Department of Surgery and Department of Immunology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA

[6]HudsonAlpha Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA

Summary

Allorecognition, the ability to discriminate between self and nonself, is ubiquitous among colonial metazoans and widespread among aclonal taxa [1–3]. Genetic models for the study of allorecognition have been developed in the jawed vertebrates [4], invertebrate chordate *Botryllus* [5, 6], and cnidarian *Hydractinia* [7]. In *Botryllus*, two genes contribute to the histocompatibility response, *FuHC* [5, 8] and *fester* [6]. In the cnidarian *Hydractinia*, one of the two known allorecognition loci, *alr2*, has been isolated [7], and a second linked locus, *alr1*, has been mapped to the same chromosomal region, called the allorecognition complex (ARC) [9, 10]. Here we isolate *alr1* by positional cloning and report it to encode a transmembrane receptor protein with two hypervariable extracellular regions similar to immunoglobulin (Ig)-like domains. Variation in the extracellular domain largely predicts fusibility within and between laboratory strains and wild-type isolates. *alr1* was found embedded in a family of immunoglobulin superfamily (IgSF)-like genes, thus establishing that the ARC histocompatibility complex is an invertebrate IgSF-like gene complex.

Results and Discussion

Hydractinia symbiolongicarpus is a colonial hydrozoan that grows as a surface encrustation, often on hermit crab shells. Colonies encounter one another by growing into contact, whereupon they display one of three responses: fusion, rejection, or transitory fusion [10–14]. The genetic basis of these phenotypes was first determined by classical breeding experiments [9, 15]. Near-congeneic lines segregating for two allorecognition haplotypes (*f* and *r*) were established, and two loci (*alr1* and *alr2*) were mapped to a single chromosomal interval, the allorecognition complex (ARC). Within the congeneic lines, *Hydractinia* colonies fuse if they share at least one allele at both *alr1* and *alr2*. The colonies reject if no alleles are shared at either locus. The transitory fusion (TF) response, characterized by an initial fusion followed by rejection [10], occurs as a dosage effect. In laboratory lines, transitory fusion arises when colonies share one or more alleles at one locus but no alleles at the other.

Positional Cloning and Identification of an *alr1* Candidate

A chromosome walk was initiated at a marker tightly linked to *alr1* (marker 194, Figure 1A; [10]), using a bacterial artificial chromosome (BAC) library constructed with DNA from the laboratory line homozygous for the *f* allele [7]. The *alr1*-containing interval was circumscribed using a large pool of recombinants (Xs in Figure 1B) available between the markers flanking *alr1* (markers 194 and 18, Figure 1A) [9, 10]. Detailed recombination breakpoint mapping defined a minimal *alr1-f* interval of 300.8 kb. The syntenic *alr1-r* interval was obtained from a BAC library constructed from a colony homozygous for the *r* allele [7] and found to span a region of at least 458.2 kb. The minimum tiling paths of both haplotypes were sequenced (Figure 1A).

Allorecognition determinants are expected to be membrane-bound proteins bearing polymorphic extracellular recognition domains. *alr2*, the linked allodeterminant previously identified from *Hydractinia*, is prototypical. *alr2* was determined to be an immunoglobulin superfamily (IgSF)-like molecule bearing three extracellular domains similar to Ig-like domains, a transmembrane domain, and a cytoplasmic tail bearing an ITIM-like signaling domain. The *alr1*-containing interval was analyzed for gene content, and putative coding sequences (pCDSs) were identified (Figure 1C; see also Table S1 available online). Three pCDSs (pCDS1, 4, and 7) raised immediate interest because they displayed domains similar to Ig-like domains. For clarity in nomenclature, we hereafter refer to these domains as noncanonical Ig-like domains and provide details of their differences from canonical Ig-like domains in the Supplemental Results and Discussion (Figure S2) where necessary. Moreover, pCDS1 and pCDS4 bore sequence similarity to the *alr2* allorecognition molecule. The full gene complement of the interval is listed in the Supplemental Results and Discussion and Table S1.

Polymorphism between our two inbred histoincompatible lines *f* and *r* was determined for the three pCDSs containing noncanonical Ig-like domains (Table S1). Only pCDS1 and pCDS4 were polymorphic, eliminating pCDS7 as an *alr1* candidate. The pCDS1 allele displayed structural variation between the pCDS1-*f* and pCDS1-*r* alleles. Specifically, the pCDS1-*f* allele was found to contain a partial gene duplication, and the pCDS1-*r* homolog sequence was missing the first exon of the predicted gene (Figure S1). Thus, only pCDS4 fulfilled all of the required criteria to be a plausible allorecognition receptor.

Full-length cDNA sequences were generated from homozygous *f* and *r* colonies, and the genomic organization of CDS4 was defined (Figure 1D). CDS4 covered a 22.4 kb interval and contained eight exons. The CDS4 predicted protein was a 537 aa type I transmembrane protein and contained a 21 aa

*Correspondence: leo.buss@yale.edu

[7]Present address: Institute of Biology/Applied Genetics, Freie Universität Berlin, 14195 Berlin, Germany

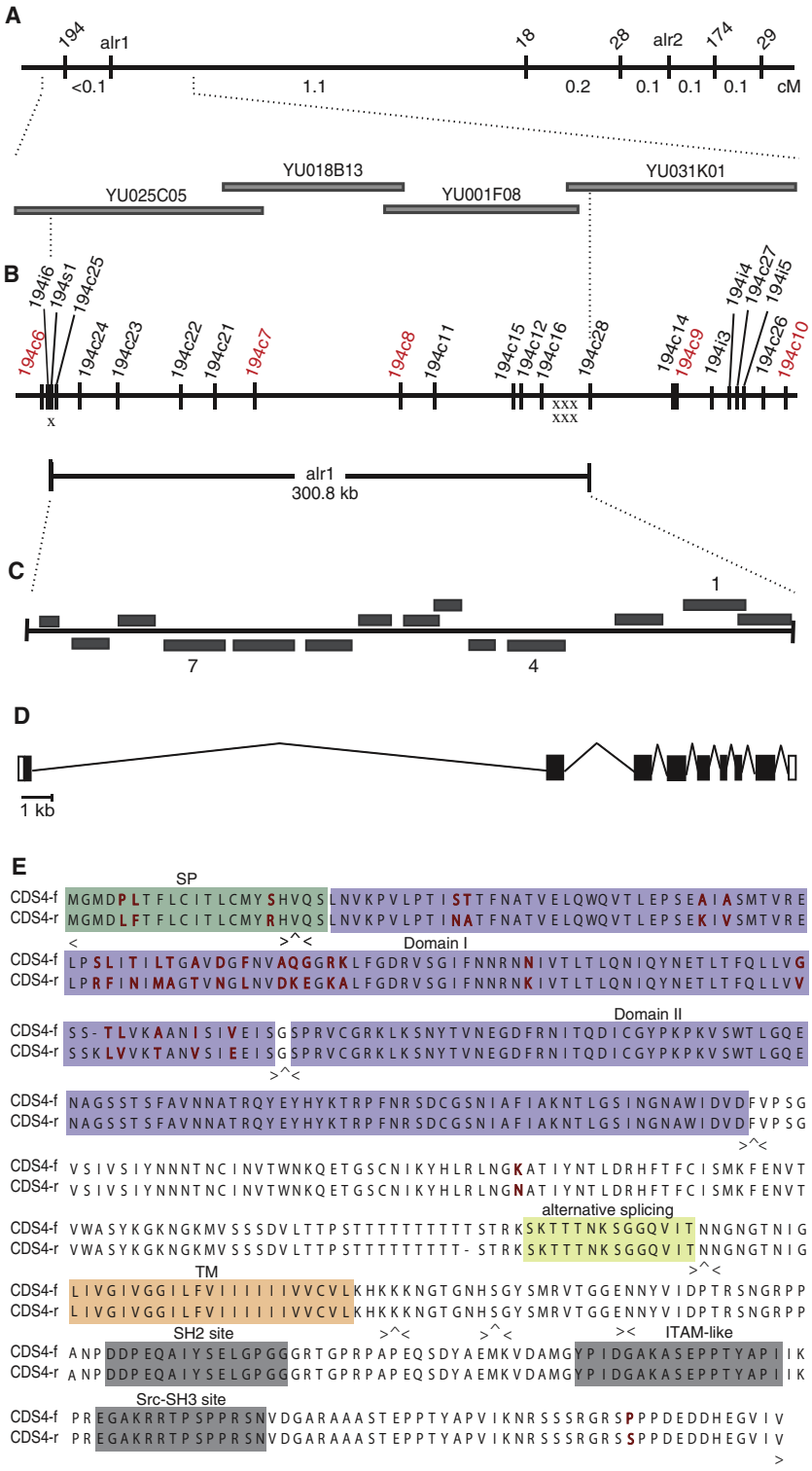

Figure 1. Positional Cloning of *alr1*

(A) Genetic map of the *Hydractinia* allorecognition complex (ARC). The *alr1* minimum tiling path was distributed over four bacterial artificial chromosome (BAC) clones (heavy gray lines).

(B) Physical map of the genomic *alr1* interval (ARC-*f* haplotype). Breakpoints are designated as Xs. BAC end markers are in red, and markers designed to map the recombination breakpoints are in black. Markers 194s1 and 194c28 define the 300.8 kb minimum *alr1* interval.

(C) Predicted coding sequences in the *alr1-f* interval (see also Table S1 and Figure S1). Boxes above and below the line represent predicted genes on the plus and minus strand, respectively. Boxes numbered 1, 4, and 7 represent coding sequences bearing sequence similarity to Ig-like domains (see text and Figure S2).

(D) Genomic organization of CDS4. White boxes, black boxes, and bent lines represent untranslated regions, exons, and introns, respectively.

(E) Alignment of the CDS4 ARC-*f* and *-r* amino acid sequences. Polymorphic residues are shown in red. Domain organization of the predicted CDS4 protein is shown within colored boxes. The following labels are used: SP, signal peptide; domains I and II, noncanonical Ig-like domains; TM, transmembrane domain; SH2, ITAM-like, and Src-SH3 sites, potential signal transduction motifs. The pale green box indicates the segment of exon 4 missing in alternatively spliced variants. Exon organization is shown below the alignment. < and > indicate amino acids encoded by the last and first complete codon of the exon, respectively, and ∧ indicates a codon spanning two exons.

signal peptide, encoded mostly by exon 1; a 324 aa extracellular region with two domains bearing similarity to Ig-like domains (domain I and II), encoded by exons 2 and 3, respectively; a 23 aa transmembrane (TM) domain, encoded by exon 5; and a 169 aa cytoplasmic domain comprising exons 6, 7, and 8 (Figure 1E). In silico analyses detected a Src homology region 2 (SH2) and a Src homology region 3 (SH3) in the cytoplasmic tail. The cytoplasmic tail also contained an immunoreceptor tyrosine-based activation motif (ITAM)-like motif (Y-x-I-x10-Y-xx-I). The CDS4 ITAM-like motif contained all of the conserved ITAM residues (Figure 1E). Polymorphism across CDS4 was assessed by aligning the CDS4-*f* and *-r* peptides (Figure 1E). A total of 29 aa substitutions were found between the two alleles, 24 of which were located in the 118 aa region spanning domain I.

BLASTP searches with CDS4 using the NCBI nonredundant protein database identified diverse IgSF proteins. The most significant alignments were to two predicted proteins of *Hydra magnipapillata*. The first alignment was between 286 aa of the extracellular region of CDS4 and the predicted *Hydra* protein LOC100215989 (e = 5×10^{-10}; 31% identity; XP_ 002157348.1 on scaffold NW_002101519). The second match occurred between the CDS4 TM domain and part of the cytoplasmic tail to the *Hydra* hypothetical protein LOC100206843 (e = 2×10^{-07}; 25% identity; XP_002166513 on scaffold NW_002085647). The next most significant alignment to CDS4 was the *Hydractinia* allodeterminant *alr2* (e = 4×10^{-06}; 27% identity; ACN91134.1). Alignments of CDS4 and *alr2* spanned extracellular regions encoded by the second

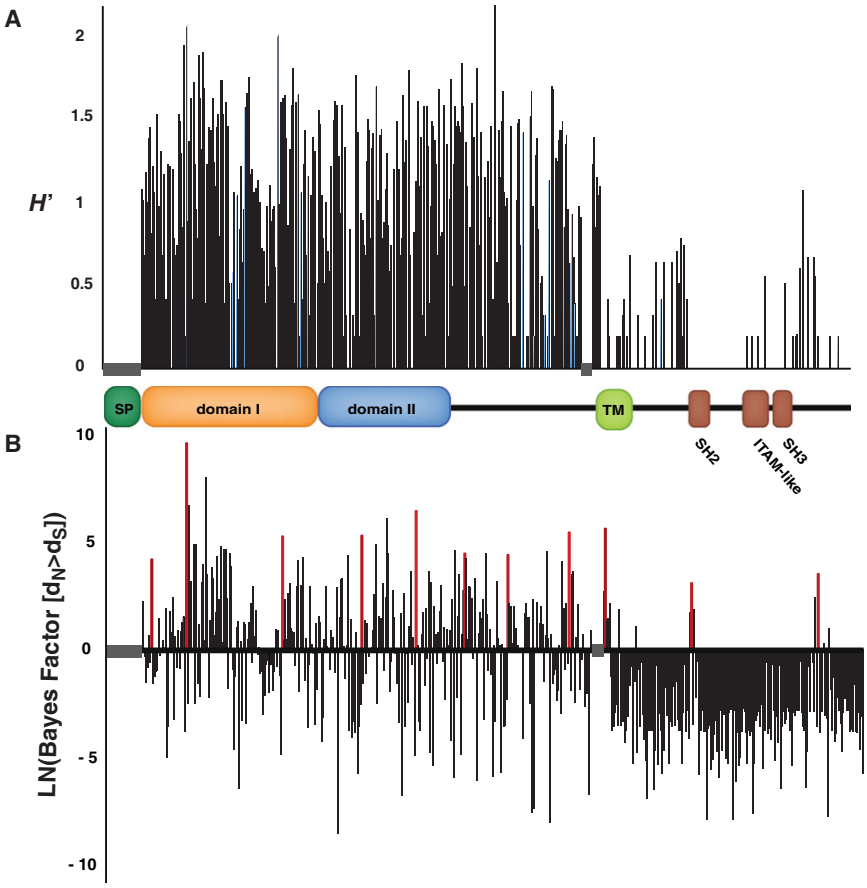

Figure 2. CDS4 Sequence Variation

(A) Distribution of variability in 20 CDS4 alleles (see also Figure S4). Per-site sequence variability was estimated using the Shannon diversity index (Shannon entropy [25], *H'*). *H'* ranges from 0 (only one amino acid type is present at one site) to 4.322 (the 20 amino acids are present at one site). CDS4 domains are shown below the plot (see Figure 1D). Gray boxes along the x axis represent regions where reliable alignments could not be generated.

(B) Distribution of synonymous and nonsynonymous mutations. The per-site difference between the number of nonsynonymous mutations (d_N) and synonymous mutations (d_S) is shown along the y axis and was measured as LN (Bayes factor [$d_N > d_S$]) using the REL method available in the HyPhy statistical package [33]. Red lines indicate the 11 sites predicted to be under positive selection by at least two of the three methods (see also Table S3).

half of exon 2 (domain I) and exon 3 (domain II). Similarity searches against the cnidarian expressed sequence tag (EST) database showed a highly significant alignment to *Hydractinia echinata* EST tah98e03.x1 and y1. The alignment spanned the entire intracellular domain of CDS4 and the entire EST length. Similarity to allorecognition proteins from other organisms such as the vertebrate major histocompatibility complex (MHC) or the *Botryllus Fu/HC* and *fester* was not observed.

Domain II was consistently and significantly predicted as being most similar to the Ig I-set fold in conserved domain searches and tertiary structure predictions (Table S2). Domain I was predicted to belong to the IgSF by CD-Search. Its fold was most similar to the fold of VCBP3 (variable chitin binding protein 3) (95% confidence), which belongs to variable immune-related IgSF proteins found in amphioxus [16, 17]. The similarity between domains I and II and Ig-like domains was further examined by performing alignments between the CDS4 domains and a set of canonical Ig-like I- and V-set domains. Canonical residues were largely conserved, and, when a residue was not conserved, the chemical properties of the residues were in most cases similar (Figure S2).

Quantitative polymerase chain reaction (PCR) assays showed that CDS4 was expressed in tissues capable of allorecognition, whole adult colony, and mat tissue (Figure S3). Expression was also observed in the tissues representative of early developmental stages: unfertilized and fertilized embryos, blastulae, and early planula larvae, as observed for *alr2* [7]. Newly metamorphosed polyps are known to be prone to allogeneic encounters in natural populations [18, 19], and an early ontogenetic activation of the allorecognition system was expected.

Polymorphism and Positive Selection

Allorecognition loci are expected to be highly polymorphic [3]. In addition to the *f* and *r* alleles, shown to be polymorphic (Figure 1E) in laboratory lines, cDNA sequences were obtained for an additional 18 CDS4 wild-type alleles (Table S4). The extracellular region of the CDS4 peptide was hypervariable (Figure 2A; Figure S4). Amino acid sequences from different alleles could be highly divergent, with identities as low as 30%. Only of 123 and 11 of 100 sites were invariant between the 20 alleles in the exons encoding CDS4 extracellular domains I and II, respectively. In contrast, the cytoplasmic tail of wild-type alleles shared an average aa identity of 94%. Splice variants were found in exon 4 for 16 of the 20 alleles obtained (see also Figure 1E). In the short splice variant of exon 4, only 13 of 109 amino acid positions were conserved.

Allorecognition genes are expected to be under positive selection [20, 21]. The hypervariable CDS4 extracellular region displays strikingly higher levels of nonsynonymous to synonymous substitutions than the more conserved intracellular region (Figure 2B). Statistical analysis of d_N/d_S ratios identified 11 codons under positive selection, 8 of which were located in the extracellular domains (Table S4).

CDS4 was identified as the only intact polymorphic gene within a tight chromosomal interval defined by recombination breakpoints, leading us to conclude that CDS4 was the locus previously identified as *alr1*. *alr1* was determined to be a transmembrane protein bearing a hypervariable extracellular region under positive selection and was found to be expressed in appropriate tissues.

Phenotype Prediction in Wild-Type Animals

Within inbred lines, *alr1* and *alr2* genotypes predict observed phenotypes in a dose-dependent fashion [10]. Colonies that match an allele at both loci fuse, colonies that match at neither loci reject, and colonies that match one locus but not the other undergo a form of transitory fusion. The range of transitory

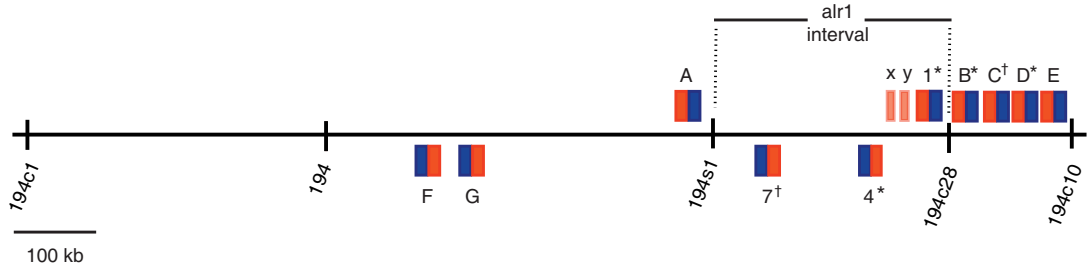

Figure 3. IgSF-like Gene Complex

Coding sequences showing sequence similarity to Ig-like domains in a 1.3 Mb region including the *alr1-f* interval. Orange boxes represent regions bearing similarity to V-set Ig-like domains, and blue boxes represent regions bearing similarity to I-set Ig-like domains. Letters and numbers represent the IgSF-like genes identified outside and within the *alr1*-containing interval, respectively. Light orange boxes are regions showing similarity to Ig V-set domains but not predicted as belonging to a coding sequence. Asterisks (*) and daggers (†) indicate putative coding sequences whose extracellular domains were polymorphic and monomorphic between the two laboratory haplotypes, respectively. Markers 194c1 and 194c10 represent the ends of the chromosome walk. Markers 194s1 and 194c28 represent the *alr1* interval limits.

fusion phenotypes displayed in wild-types is substantially more diverse than that displayed by inbred lines [9, 12, 13, 22, 23]. It is of interest, therefore, to determine the extent to which knowledge of the sequences of the two *alr* loci is sufficient to explain fusibility in wild-type animals. Any departure from the expectations based on the inbred animals would suggest the existence of either modifying loci or additional allodeterminants not detected within the inbred lines.

Two tests of whether *alr1* sequence variation could predict wild-type fusibility were performed. In the first, the fusibility phenotype of a wild-type animal to a homozygous laboratory strain is known, and the test is to determine whether the wild-type animal shares the laboratory *alr1* allele. This test is particularly stringent in that fusion frequencies in nature are very low [7, 15, 19, 23]. Specifically, a prior screen of 535 animals for fusible wild-type colonies identified only two fusible *Hydractinia* colonies (LH06-082 and LH06-416), both of which displayed transitory fusion against the ARC *f/f* inbred line [7]. An *alr1* allele was obtained from both of these colonies that shared 99% amino acid identity to the extracellular domains of the *alr1-f* allele. For comparison, the average amino acid variation between extracellular domains of wild-type *alr1* alleles was less than 50%. The hypervariable domain of the *alr2* allele was 100% in the first pair [7], whereas partial *alr2* sequences were mismatched in the second pair.

The second test derived from sequencing of 18 wild-type *alr1* alleles. Here, we identified animals sharing a common *alr1* allele and tested whether the colonies displayed a fusible phenotype. Three pairs of colonies were found to share an *alr1* allele. Colonies LH06-058 and LH07-014 shared an allele bearing 99.7% aa identity in the extracellular domain and were found to fuse. This pair also shared an *alr2* allele showing 100% identity in the hypervariable domain. Similarly, LH06-003 and LH06-049 displayed 100% identity and transitory fusion. This pair shared an *alr2* allele bearing 100% identity in the hypervariable domain. The remaining pair, LH06-050 and LH07-014, displayed 100% identity in the extracellular domain and rejected. This pair was mismatched in *alr2* hypervariable domain. Amino acid alignments of the hypervariable domains I of *alr1* and *alr2* from colonies tested for fusibility are shown in Figure S5.

These results show that knowledge of *alr* genotype is largely, but not completely, predictive of phenotype. As noted above, the probability of detecting a wild-type colony that fuses to laboratory lines is < 1%. Yet in the two tests applied here, 80% of the pairs (4 of 5) expected to display fusible phenotypes did so. Although *alr* sequence variation is predictive, in two wild-type genetic backgrounds the form of transitory fusion differed from that expected in inbred lines, and in one case rejection was observed where fusion was expected. This latter pair is of special interest because it indicates that other allodeterminants, such as those described below, may contribute to the allorecognition phenotype in the *Hydractinia* system (see Supplemental Results and Discussion).

alr1 Resides within an IgSF-like Gene Complex

A total of ten predicted genes bearing noncanonical Ig-like domains were identified in a 700 kb region surrounding *alr1-f* (Figure 3). The extent of the gene family is unknown, and more IgSF-like genes may be identified as the region is extended. All IgSF-like genes were structurally similar and contained two extracellular domains, one most similar to a V-set Ig-like domain and one most similar to an I-set Ig-like domain. A similarity search against the National Center for Biotechnology Information (NCBI) nonredundant protein database yielded alignments only with conserved Ig-like domains. The most significant alignments returned for seven of the IgSF-like genes were to the *Hydractinia alr2* gene. Sequences of six predicted IgSF-like genes were available for both laboratory haplotypes, and four were polymorphic (Figure 3). Variable genes within this complex are plausible candidates for explaining the disparity in fusibility identified between LH06-050 and LH07-014 mentioned above, a topic of active investigation.

Comparison of the extracellular domains of *alr1* wild-type alleles with the members of the IgSF-like gene complex identified in the ARC-*f* haplotype revealed a striking pattern. When the *alr1* wild-type alleles were used as queries in similarity searches against the 700 kb IgSF-like gene complex, the *alr1* wild-type intracellular domains were in each case most similar to the originally identified *alr1-f* allele. Conversely, the wild-type *alr1* extracellular domains often displayed greater sequence similarity to extracellular domains from linked genes within the complex than to *alr1-f* (Figure 4). This finding suggests that members of this complex could serve as sequence donors.

These results complete the genetic characterization of the allorecognition loci previously defined in *Hydractinia* and identify them as members of an IgSF-like gene complex. Histocompatibility complexes, such as the natural killer (NK) or MHC complexes, are known in jawed vertebrates and have also

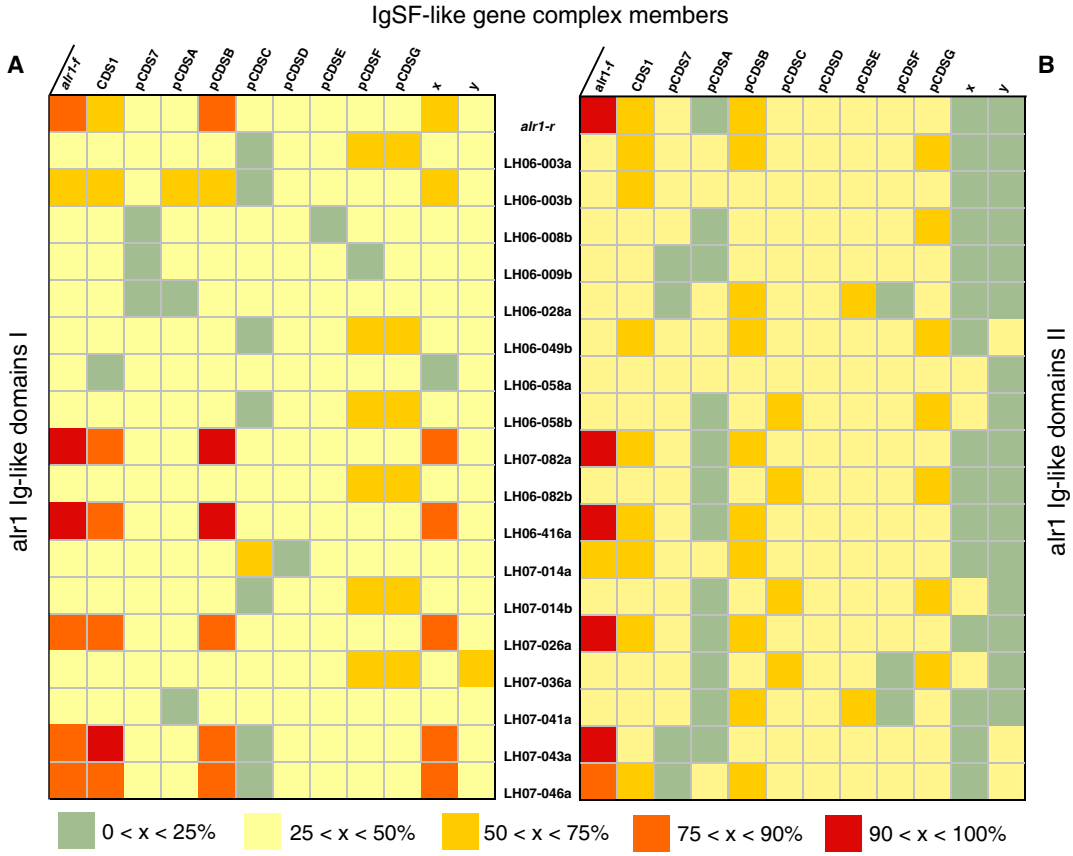

Figure 4. Pairwise Comparisons between Extracellular Domains I and II of *alr1* Alleles and IgSF-like Gene Complex Members

Horizontal rows contain CDS4 (*alr1*) alleles; vertical columns contain IgSF-like gene complex members from the ARC-*f* haplotype. x and y are as designated in Figure 3. Colored squares indicate the percentage amino acid similarity, as indicated by the legend at bottom.

(A) Extracellular domain I.

(B) Extracellular domain II.

been recognized in ascidians [4, 24–27]. Given the finding that both placozoan and cnidarian genomes have been shown to display surprisingly large tracks of synteny with chordates [28–30], the question of whether the *Hydractinia alr*-containing chromosomal intervals display synteny with vertebrate IgSF gene complexes is of immediate interest. Unfortunately, variability of the newly identified *Hydractinia* IgSF-like genes and the absence of other conserved genes (Table S1) within the *alr1* interval prohibit the identification of regions of synteny with other organisms with currently available sequences. This question, along with that of the role of chromosomal organization in the generation of the observed hypervariability in *alr1*, awaits future investigation.

Experimental Procedures

Animal Collection, Maintenance, and Fusibility Assays

Animal maintenance and fusibility assays were performed as described in [9]. Two wild-type colonies (LH06-082 and LH06-416) used to identify fusing pairs and matching CDS4 alleles were the same as those used in [7]. The third pair (LH06-058 and LH07-014) was identified from a collection of colonies made in 2006 and 2007 from the intertidal at Lighthouse Point, New Haven, CT.

Positional Cloning

The positional cloning strategy was as described previously [7]. Once the *alr1-f* minimum interval had been circumscribed by proximal and distal recombination breakpoints along the chromosome walk, the ARC-*r* BAC library [7] was screened with probes designed to the extremities of the *alr1-f* interval in order to identify the corresponding *alr1-r* interval (GenBank accession numbers AC234855, AC234857, and AC234859). Note that for the *f* haplotype, the entire interval obtained by the bidirectional chromosome walk initiated at marker 194 (1.3 Mb) was sequenced (GenBank accession numbers AC234863–AC234869, AC234871, AC234877, AC234878, AC234883, AC234887, and AC234888).

Sequence Analysis and Determination of the *alr1* Candidates

The sequenced *alr1-f* and *-r* genomic intervals were submitted to similarity searches against NCBI protein databases (nonredundant protein sequences and Swiss-Prot databases) and the NCBI Expressed Sequence Tags database (with search set on the phylum Cnidaria) and to ab initio gene prediction algorithms to identify the potential homologs and potential coding sequences, as detailed in the Supplemental Experimental Procedures. Quantitative expression of CDS4 was assessed by real-time PCR as described in Figure S3 and Supplemental Experimental Procedures.

Polymorphism and Selection Analysis

Full-length coding sequences (signal peptide to stop codon) were obtained by rapid amplification of cDNA ends and RT-PCR for 18 alleles from 14 wild-type colonies, as detailed in the Supplemental Experimental Procedures. Predicted peptide sequences were aligned using PRANK [31] and MUSCLE [32] algorithms. The regions corresponding to the long splice variant of exon 4 were available for only 16 of the 20 alleles and were not included in the alignments. Regions corresponding to the signal peptides (26 aa positions) and to the junction of exon 4 and 5 (8 aa positions) produced alignment artifacts as a result of structural differences and were removed from the analyses. Analyses for positive selection were performed using the SLAC, FEL, and REL methods from the HyPhy statistical package [33].

Accession Numbers

Coordinates for the *alr1* and *alr2* sequences have been deposited in GenBank with the accession numbers HM070427–HM070448 (*alr1*) and HM013606, -10, -12, -14, -15, FJ207405, FJ207408, and EU219736 (*alr2*).

Supplemental Information

Supplemental Information includes Supplemental Results and Discussion, five figures, four tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.cub.2010.04.050.

Acknowledgments

We thank K. Altland, S. Kirschner, J. Hayase, and A. Gloria-Soria for technical assistance and discussions. L. Cadavid provided the ARC-*r* BAC library. This work was supported by National Institutes of Health (NIH) grants 1R21-A1066242 and 1R56AI079103-01 (F.G.L., S.L.D., and L.W.B.) and National Science Foundation (NSF) grant IOS-0818295 (L.W.B., S.L.D., and M.A.M.). The US Department of Energy Joint Genome Institute provided sequencing and analyses under the Community Sequencing Program supported by the Office of Science of the US Department of Energy under contract DE-AC02-05CH11231 (L.W.B. and S.L.D.). A.E.P. was supported by NIH Training Grant T32-GM07499. R.D.R. was supported by a NSF Graduate Research Fellowship.

Received: January 23, 2010
Revised: April 14, 2010
Accepted: April 22, 2010
Published online: May 27, 2010

References

1. Buss, L.W. (1982). Somatic cell parasitism and the evolution of somatic tissue compatibility. Proc. Natl. Acad. Sci. USA *79*, 5337–5341.
2. Boehm, T. (2006). Quality control in self/nonself discrimination. Cell *125*, 845–858.
3. Grosberg, R.K. (1988). The evolution of allorecognition specificity in clonal invertebrates. Q. Rev. Biol. *63*, 377–412.
4. Kulski, J.K., Shiina, T., Anzai, T., Kohara, S., and Inoko, H. (2002). Comparative genomic analysis of the MHC: The evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol. Rev. *190*, 95–122.
5. De Tomaso, A.W., Nyholm, S.V., Palmeri, K.J., Ishizuka, K.J., Ludington, W.B., Mitchel, K., and Weissman, I.L. (2005). Isolation and characterization of a protochordate histocompatibility locus. Nature *438*, 454–459.
6. Nyholm, S.V., Passegue, E., Ludington, W.B., Voskoboynik, A., Mitchel, K., Weissman, I.L., and De Tomaso, A.W. (2006). *fester*, a candidate allorecognition receptor from a primitive chordate. Immunity *25*, 163–173.
7. Nicotra, M.L., Powell, A.E., Rosengarten, R.D., Moreno, M., Grimwood, J., Lakkis, F.G., Dellaporta, S.L., and Buss, L.W. (2009). A hypervariable invertebrate allodeterminant. Curr. Biol. *19*, 583–589.
8. De Tomaso, A.W., and Weissman, I.L. (2003). Initial characterization of a protochordate histocompatibility locus. Immunogenetics *55*, 480–490.
9. Cadavid, L.F., Powell, A.E., Nicotra, M.L., Moreno, M., and Buss, L.W. (2004). An invertebrate histocompatibility complex. Genetics *167*, 357–365.
10. Powell, A.E., Nicotra, M.L., Moreno, M., Lakkis, F.G., Dellaporta, S.L., and Buss, L.W. (2007). Differential effects of allorecognition loci on phenotype in *Hydractinia*. Genetics *177*, 2101–2107.
11. Buss, L.W., McFadden, C.S., and Keene, D.R. (1984). Biology of hydractiniid hydroids. 2. Histocompatibility effector system competitive mechanism mediated by nematocyst discharge. Biol. Bull. *167*, 139–158.
12. Hauenschild, C. (1954). Genetische und entwichlungphysiologische Untersuchungen über Intersexualität und Gewebeverträlichkeit bei *Hydractinia echinata* Flem. Rouxs Arch. Dev. Biol. *147*, 1–41.
13. Hauenschild, C. (1956). Über die Vererbung einer Gewebeverträglichkeitseigenschaft bei dem Hydroidpolypen *Hydractinia echinata*. Z. Naturforsch. Pt. B *11*, 132–138.
14. Müller, A.W. (1964). Experimentelle Untersuchungen über Stockentwicklung, Pelypendifferenzierung und Sexualchimären bei *Hydractinia echinata*. Rouxs Arch. *155*, 182–268.
15. Mokady, O., and Buss, L.W. (1996). Transmission genetics of allorecognition in *Hydractinia symbiolongicarpus* (Cnidaria:Hydrozoa). Genetics *143*, 823–827.
16. Cannon, J.P., Haire, R.N., and Litman, G.W. (2002). Identification of diversified genes that contain immunoglobulin-like variable regions in a protochordate. Nat. Immunol. *3*, 1200–1207.
17. Dishaw, L.J., Mueller, M.G., Gwatney, N., Cannon, J.P., Haire, R.N., Litman, R.T., Amemiya, C.T., Ota, T., Rowen, L., Glusman, G., et al. (2008). Genomic complexity of the variable region-containing chitin-binding proteins in amphioxus. BMC Genet. *9*, 78.
18. Buss, L.W., and Yund, P.O. (1988). A comparison of recent and historical populations of the colonial hydroid *Hydractinia*. Ecology *69*, 646–654.
19. Nicotra, M.L., and Buss, L.W. (2005). A test for larval kin aggregations. Biol. Bull. *208*, 157–158.
20. Grosberg, R.K., and Hart, M.W. (2000). Mate selection and the evolution of highly polymorphic self/nonself recognition genes. Science *289*, 2111–2114.
21. Humphreys, T., and Reinherz, E.L. (1994). Invertebrate immune recognition, natural immunity and the evolution of positive selection. Immunol. Today *15*, 316–320.
22. Shenk, M.A., and Buss, L.W. (1991). Ontogenic changes in fusibility in the colonial hydroid *Hydractinia symbiolongicarpus*. J. Exp. Zool. *257*, 80–86.
23. Grosberg, R.K., Levitan, D.R., and Cameron, B.B. (1996). Evolutionary genetics of allorecognition in the colonial hydroid *Hydractinia symbiolongicarpus*. Evolution *50*, 2221–2240.
24. Litman, G.W., Hawke, N.A., and Yoder, J.A. (2001). Novel immune-type receptor genes. Immunol. Rev. *181*, 250–259.
25. Kelley, J., Walter, L., and Trowsdale, J. (2005). Comparative genomics of natural killer cell receptor gene clusters. PLoS Genet. *1*, 129–139.
26. Osman, N., Kozak, C.A., McKenzie, I.F., and Hogarth, P.M. (1992). Structure and mapping of the gene encoding mouse high affinity Fc gamma RI and chromosomal location of the human Fc gamma RI gene. J. Immunol. *148*, 1570–1575.
27. Zucchetti, I., De Santis, R., Grusea, S., Pontarotti, P., and Du Pasquier, L. (2009). Origin and evolution of the vertebrate leukocyte receptors: The lesson from tunicates. Immunogenetics *61*, 463–481.
28. Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V., et al. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science *317*, 86–94.
29. Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L., et al. (2008). The *Trichoplax* genome and the nature of placozoans. Nature *454*, 955–960.
30. Chapman, J.A., Kirkness, E.F., Simakov, O., Hampson, S.E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P.G., Borman, J., Busam, D., et al. (2010). The dynamic genome of *Hydra*. Nature *464*, 592–596.
31. Loytynoja, A., and Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. Proc. Natl. Acad. Sci. USA *102*, 10557–10562.
32. Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.
33. Pond, S.L., Frost, S.D.W., and Muse, S.V. (2005). HyPhy: Hypothesis testing using phylogenies. Bioinformatics *21*, 676–679.