The Proto-MHC of Placozoans, a Region Specialized in Cellular Stress and Ubiquitination/Proteasome Pathways

Jaanus Suurväli, Luc Jouneau, Dominique Thépot, Simona Grusea, Pierre Pontarotti, Louis Du Pasquier, Sirje Rüütel Boudinot and Pierre Boudinot

*J Immunol* published online 11 August 2014
http://www.jimmunol.org/content/early/2014/08/09/jimmunol.1401177

Supplementary Material
http://www.jimmunol.org/content/suppl/2014/08/09/jimmunol.1401177.7.DCSupplemental

Why *The JI*? Submit online.
- **Rapid Reviews! 30 days*** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

Submit online.

Why *The JI*? Submit online.
- **Rapid Reviews! 30 days*** from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

Subscription
Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

Permissions
Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts

The Journal of Immunology is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852
Copyright © 2014 by The American Association of Immunologists, Inc. All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.
The Proto-MHC of Placozoans, a Region Specialized in Cellular Stress and Ubiquitination/Proteasome Pathways

Jaanus Suurvälä,* Luc Jouneau,† Dominique Thépot,‡ Simona Grusea,§ Pierre Pontarotti,¶ Louis Du Pasquier,‖ Sirje Rüütel Boudinot,* and Pierre Boudinot‡

The MHC is a large genetic region controlling Ag processing and recognition by T lymphocytes in vertebrates. Approximately 40% of its genes are implicated in innate or adaptive immunity. A putative proto-MHC exists in the chordate amphioxus and in the fruit fly, indicating that a core MHC region predated the emergence of the adaptive immune system in vertebrates. In this study, we identify a putative proto-MHC with archetypal markers in the most basal branch of Metazoans—the placozoan Trichoplax adhaerens, indicating that the proto-MHC is much older than previously believed—and present in the common ancestor of bilaterians (contains vertebrates) and placozoans. Our evidence for a T. adhaerens proto-MHC was based on macrosynteny and phylogenetic analyses revealing approximately one third of the multiple marker sets within the human MHC-related paralogy groups have unique counterparts in T. adhaerens, consistent with two successive whole genome duplications during early vertebrate evolution. A genetic ontologic analysis of the proto-MHC markers in T. adhaerens was consistent with its involvement in defense, showing proteins implicated in antiviral immunity, stress response, and ubiquitination/proteasome pathway. Proteasome genes psma, psmb, and psmd are present, whereas the typical markers of adaptive immunity, such as MHC class I and II, are absent. Our results suggest that the proto-MHC was involved in intracellular intrinsic immunity and provide insight into the primordial architecture and functional landscape of this region that later in evolution became associated with numerous genes critical for adaptive immunity in vertebrates. The Journal of Immunology, 2014, 193: 000–000.

The human MHC of jawed vertebrates is defined as a large genetic region of ∼4 megabases (Mb) encoding more than 100 genes, approximately half of which are implicated in immunity (1). It is divided into three major regions. Class I and class II regions encode the polymorphic Ag-presenting molecules class I, Ia, and β, factors such as B30.2 proteins, and genes involved in Ag processing pathways such as proteasome genes and TAP. The B30.2 domain (2) mediates defense and other functions in several families of proteins, such as butyrophilins and tripartite motif proteins (TRIM) (2–7). The gene-dense class III region encodes several complement components and other genes involved in inflammation (8). The architecture of vertebrate MHCs vary, from the comparatively small MHC of the chicken to the teleost fish, in which class I and II genes are not linked (1, 9–11).

However, the various genes of the complex and their basic functions have been generally conserved, and the elements of the human MHC represent the archetypal MHC genes found across jawed vertebrates (8).

The origin of the MHC is incompletely understood, but the MHC backbone is considered ancient and linked to innate immunity. Class I and class II MHC genes have been suggested to originate from families of molecules present in what is called a “proto-MHC region” for convenience but is involved in innate immunity (12, 13). This ancient backbone could be inherited from the invertebrate ancestors of deuterostomians and protostomians (14–17) (Fig. 1A), from a proto-MHC region. It has been proposed that cis-duplications and translocations of this single ancestral region occurred during evolution, leading to three original complexes on three different chromosome segments: a primordial MHC that gave rise to the set of genes involved in Ag presentation; a neutrophil complex that led to genomic regions comprising neutrophil receptors and the Leukocyte Receptor Complex; and a third genetic complex, the tunicate MHC-related complex JAM-NECTIN (JN) that is at the origin of paralogous regions containing many Ig superfamilly members. These proteins have cell adhesion properties (JAM, NECTINs, poliovirus receptors’ family members) and are often implicated in the biology of lymphocytes, NK cells, and other leukocytes (CD96, cd155, cd112, JAM B, and CTX family members) (15, 16, 18, 19). Pairs of structurally related molecules can act as receptor–ligand systems and mediate interactions between leukocytes and endothelia, linking this third proto-MHC derived complex to immunity.

According to the hypothesis originally proposed by Susumu Ohno (20), these three genetic regions were duplicated twice at an early stage of vertebrate evolution, leading to three independent tetras of paralogous regions (paralogs; see Fig. 1B for definitions) with some overlaps (e.g., B7 receptors were spread over those three regions) (21) and breaks. Ohno’s hypothesis of two
cycles of whole genome duplications in ancestral vertebrates is supported by the study of hox complexes, and it applies to the MHC-related complexes as well as other regions (22–30). In the human genome, there are three main paralogs of the MHC on chromosomes 1 (1q21-q25/1p11-1p33), 9 (9q32-q34), and 19 (19p13.1-p13.3) (22–24). In addition, there are smaller fragments, such as those located on 15q13-q26 and 5q11-q23, translocated from the MHC paralogs on chromosomes 6 and 9, respectively (25). Four paralogs of neurotrophin are on chromosomes 1 (1q32-44, 1p13), 14 (14q11-q32), and 19 (19q13-1q44) that contains the leukocyte receptor complex, and 11 (11q12-q13, 11q23-q24, 11p12-p15) (18). Additional fragments of the neurotrophin group are on chromosomes 2 (2p21-p23), 20 (20p11-p12), and 12 (12q22-q24, 12p11-p13) with the NK complex. Finally, linkage groups corresponding to the JN complex are on chromosomes 1 (1q22-24), 3 (3q13-21), 11 (11q23-25), 21 (21q21-22), and 19q (again with some leukocyte receptor complex elements) (15, 16, 31) (see Fig. 1C for a map of all human MHC related paralogs).

The concentration of genes involved in immunity linked to class I and II MHC regions might confer a selective advantage (32). It is unknown whether the conservation of macrosynergy across vertebrates might be partly explained by inheritance independent of constraints imposed by selective advantages of natural selection. However, the conservation of linked MHC/neurotrophin/JN complexes in invertebrates suggests a significant role of such constraints. In chordates, there is a proto-MHC in the amphioxus Branchiostoma floridae (33, 34) (Fig. 1B), and a genetic complex related to JN exists in the tunicate Ciona intestinalis (14–16), whereas a clearly MHC-related region was not identified (14). There is conserved synteny between human MHC paralogs and a genomic region of the protostome Drosophila melanogaster (17). There is scant information on MHC-related regions in other groups of invertebrates, especially in the most ancient ones.

Placozoans constitute one of the most basal branches of the evolutionary tree of Metazoa (35) (Fig. 1A); it is generally accepted that they are a sister group to bilaterians and chordians, such as corals and jellyfish (36, 37), and there is increasing evidence they are the most basal branches of metazoa (38–40). Since segments of large-scale genome organization are similar in vertebrates and the placozoan Trichoplax adhaerens (35), this species is a useful model to investigate the origin of the MHC and MHC-related genetic complexes in ancestral metazoa. T. adhaerens is the only recognized species of the phylum Placozoa, although a greater degree of diversity has been proposed for this phylum (41–44). T. adhaerens is a small (100–200 μm), disc-shaped marine organism consisting of two epithelial layers enclosing a layer of multinucleate fiber cells. Only four cell types have been identified; it lacks most of the elements associated with multicellular organisms, such as complex organ systems and extracellular matrix (43, 45). The genome of T. adhaerens is 100 megabases long with twelve pairs of chromosomes and ~11,500 predicted protein-coding genes (35), of which 6516 have been found expressed in the first comprehensive proteome analysis (46). The homologs of these genes are often associated with complex developmental and signaling pathways and are highly conserved in metazoa (35, 47). Importantly, T. adhaerens scaffolds have been mapped on the reconstruction of ancestral chordate linkage groups (ACLGs) and extensive phylogenetic analyses performed for conserved regions (35). The ACLGs are defined from conserved synteny between the amphioxus and vertebrates genomes, and they correspond to the gene content of a putative chromosome of their last common ancestor (48). Hence, the ACLGs are a key generic reference for studying the conservation of blocks of synteny among chordates, vertebrates (including humans), and other species.

In this study, we examined T. adhaerens for the presence of genomic regions enriched in archetypal markers of proto-MHC and proto-neurotrophin complexes. Identified markers of the proto-MHC region were evaluated by gene ontology analysis to test whether such markers contained genes suggestive of immune-related functions, such as stress response and ubiquitination/proteasome pathways.

**Materials and Methods**

*Identification and analysis of T. adhaerens counterparts of human genes located in MHC-, neurotrophin-, and JN-related paralogs*

Human protein-coding genes located within the genomic regions described as MHC-, neurotrophin-, and JN paralogs (21, 49), and within the MHC-related region defined by the R4 RGS genes (50) were extracted from Ensembl Genome Browser release 70 (human genome assembly GRCh37; 7957 genes). The sequence of the longest protein isoform encoded by each gene was extracted using Ensembl Biomart. Protein sequences (human queries) were blasted using Δ-BLAST (51) against the proteome of T. adhaerens (assembly ASM15027v1 from release 19 of Ensembl Genomes). The e-value cutoff was set at 1e-10, and 5,865 human sequences retrieved 2540 unique best blast hits in T. adhaerens scaffolds. To target counterparts of the MHC-related regions, we selected T. adhaerens scaffolds with highest density of best blast hits—that is, scaffolds in which the proportion of hits of at least one human gene was higher than a given threshold. A density threshold was set from the distribution at 25% to maximize the contrast between the groups, selecting the scaffolds 1, 2, 3, 6, 7, 9, 10, 12, 14, 15, 22, 28, 31, 34, 35, 40, and 42 for further analysis. Protein sequences corresponding to the 6721 genes located within these scaffolds were extracted using Ensembl Biomart and blasted back against the human proteome using Δ-BLAST with an e-value cutoff of 1e-10. This second blast analysis (using T. adhaerens queries) identified 1273 human genes for which the T. adhaerens best blast hit matched the initial query sequence when "back" blasted against the human proteome. Such human and T. adhaerens entries identified by the reciprocal best blast analyses were designated as human and T. adhaerens reciprocal best blast hits (RBBH).

*Identification of a putative T. adhaerens proto-MHC*

To analyze the distribution of RBBH across T. adhaerens scaffolds, the number of RBBH per total number of genes (RBBH density) was calculated within 1 Mb blocks across scaffolds (or per scaffold for scaffolds shorter than 1 Mb). A threshold was set from the distribution at 15% to optimize the contrast between the groups, and identified a set of 1-Mb regions with a high density of RBBHs. Regions connecting blocks with high density of RBBH were included in this set. These RBBH-enriched regions were further analyzed as to which of the three sets of paralogs (MHC-related, neurotrophin-related, or JN) of the RBBH were related. For each scaffold, the percentage of neurotrophin-related, MHC-related, and JN-related among all RBBH was calculated in two ways: 1) all RBBH with human counterparts on the respective paralogs were included, including regions where different paralogs overlap in the human genome; genes located in these overlapping regions were designated ambiguous and were considered for the percentage calculations for both overlapping paralogs; 2) RBBHs with a human counterpart in the ambiguous regions were omitted, and the percentages were calculated based on the number of genes with human counterpart located in nonambiguous neurotrophin-related, MHC-related, and JN-related regions. If the percentage of the dominant set among all RBBH exceeded 75% in a block, this block was considered MHC-related, neurotrophin-related, or JN-related. The putative T. adhaerens proto-MHC was identified as a collection of all RBBH-enriched regions (located on scaffolds 2, 3, 7, 9, 10, and 15) where MHC-related RBBH were dominant. Out the 1198 genes in these regions, 307 were MHC-related, neurotrophin-related, or JN-related. The putative proto-MHC region defined by the R4 RGS genes (33, 34) (Fig. 1B), and a genetic complex related to JN exists in the tunicate Ciona intestinalis (14–16), whereas a clearly MHC-related region was not identified (14).}

**Statistical analysis**

The significance of marker clustering in T. adhaerens proto-MHC region was tested using a statistical test developed in (52). This test is adapted to approaches in which one starts with a fixed, reference genomic region in the genome of a given species A (e.g., human) and searches for orthologous regions in the genome of another species B (e.g., T. adhaerens). The test analyzes the probability of finding a number of genes clustered in a given region rather than randomly distributed in the whole genome B,
given the size of the region and the size of the whole genome. Importantly, the size is computed as the number of genes not bp. Qualitatively, the statistical significance increases when the size of the region in which markers are clustered decreases and when the size of the whole genome investigated increases. Notably, the total numbers of genes in humans and T. adhaerens differ only by a factor of 2 (~20,000 versus 10,000, respectively). The test is based on a compound Poisson approximation for computing the p value of an orthologous gene cluster under the null hypothesis of random gene order. A critical feature of the method is that it accounts for the existence of multigene families—that is, the existence of multiple counterparts (co-orthologs) in the genome B for genes of the reference region (e.g., the human MHC paralogons); to do so, the co-orthologs of the target genome B are weighted in inverse proportion to the size of the multigene family to which they belong. If a gene from the starting region has a unique ortholog in the target genome, this ortholog has a weight of 1. If a gene from the starting region has k counterparts in the target genome, then each of these co-orthologs will be given a weight of 1/k. The weight of a given region of the genome B (e.g., the T. adhaerens proto-MHC region) is defined as the sum of the weights of the orthologs belonging to it. The p value of this region is the probability, under the null hypothesis of random gene order, of finding somewhere in the genome B a region of higher weight/smaller length, and is computed using a compound Poisson approximation, as explained previously (52).

Gene ontology enrichment analysis
BiNGO, a plugin of Cytoscape (53), was used to look for local enrichment of gene ontology (GO) terms relating to biological processes among human counterparts of the 307 MHC-related genes located in the T. adhaerens proto-MHC region. The list of all 3259 genes located within the human MHC paralogons according to Ensembl 70 was used as the reference set. As every GO term was subjected to statistical test for enrichment, Benjamini and Hochberg false discovery rate p value adjustment was applied to correct for multiple testing (54, 55) and performed by BiNGO.

Microsynteny analysis
Conserved microsynten between T. adhaerens proto-MHC and human MHC paralogons were defined as sets of RBBH genes that are located in close proximity to each other in both humans and T. adhaerens (i.e., separated by 30 genes or fewer). A linkage was considered a conserved microsynteny only when three or more such genes were linked in such a way in both humans and T. adhaerens. These microsynteny gene sets were then used to test for the presence of conserved DNA-binding transcription factors using a list of 1988 such genes from humans (56).

Analysis of noncoding regions of the T. adhaerens proto-MHC
To identify putative regulatory regions, we looked for noncoding sequences conserved between T. adhaerens proto-MHC and the human genome. Using the masked sequence of Trichoplax genome (from release 19 of Ensembl Genomes), we first masked all the exons by stretches of N. The masked sequence of the proto-MHC regions was scanned for stretches of poly(N) and all stretches of more than 10 Ns were deleted, leading to sequences corresponding to introns and intergenic DNA. Fragments of <30 bp were discarded, and Blastn and tblastx analyses were performed for the remaining 18,860 fragments.

Analysis of the Monosiga brevicollis genome
A publicly available list of Mus musculus best blast hits for the genes of M. brevicollis was obtained from the JGI Genome Browser. Human MHC paralogon genes corresponding to these mouse genes were determined using Ensembl Biomart. The list of 395 human-Monosiga gene pairs was compared with our set of human-T. adhaerens RBBH, including those with T. adhaerens counterpart located in the proto-MHC. Human genes with a counterpart in either one or both organisms were then analyzed for gene ontology enrichment as described above. The complete list of M. brevicollis genes with human counterparts located on MHC paralogons was further tested for microsynteny with humans as described above for T. adhaerens.

Results
Identification of putative proto-MHC and proto-neurotrophin genomic regions in T. adhaerens
To determine whether the most basal group of Metazoaons possessed MHC-related paralogons, we examined the genome of T. adhaerens for conserved blocks of syteny corresponding to such genomic segments (Fig. 1).

We recently found that the RGS1/RGS16 region, located next to the MHC paralogon on human chromosome 1, provides useful markers to investigate the origins and the evolution of the MHC in invertebrates (50). The best blast hits of most of these markers were clustered in one 9.7-Mb scaffold (scaffold 2) of T. adhaerens, which also contained homologs of typical MHC markers. We therefore undertook a systematic survey of T. adhaerens homologs of all 7957 human genes located within the 19 genomic regions corresponding to MHC-, neurotrophin-, and JN- paralogons (21, 25, 49, 50). Sequences of the longest protein isoforms encoded by each of these genes were extracted from the human genome assembly and were used as query for reciprocal Δ-BLAST (51) searches against the predicted proteome of T. adhaerens. Forward and reverse blast analyses identified 1273 pairs of markers for which the T. adhaerens gene was the best hit of the forward analysis, and the human sequence used as an initial query in the forward blast was retrieved as the best blast hit of the reverse analysis. Human T. adhaerens entries of such pairs identified by reciprocal blast will be designated below as human and T. adhaerens RBBH, respectively.

To analyze the distribution of T. adhaerens RBBH across genomic scaffolds, the density (ratio between the number of RBBH and the total number of genes) was calculated within 1 Mb blocks across scaffolds, or per scaffold for scaffolds shorter than 1 Mb. Among blocks containing more than 15% RBBH, we identified those specifically enriched in counterparts of human MHC, neurotrophin, and JN tetrads (Fig. 2), for which we hypothesized these regions represented T. adhaerens counterparts (Fig. 2). Many human genes from the MHC-related paralogons had a counterpart on the T. adhaerens scaffold 2 that we identified in our initial screen. However, a high density of such genes was also found on parts of scaffolds 3, 7, 9, 10, and 15 (Fig. 2). Human genes from the neurotrophin-related paralogons had counterparts in other regions of the T. adhaerens genome mostly scattered across scaffold 1, but also located on scaffolds 6, 9, 12, 22, 31, 34, 40, and 42. In contrast, we did not find any region enriched mostly in markers from the third set of paralogons (JN) identified by Du Pasquier et al. (15, 16), with the exception of the very small scaffold 35. This observation suggested that this tetrad might have been produced by a later duplication during the evolution of bilaterians. However, alternative scenarios cannot be excluded; these could not be properly tested in this study because our approach would miss markers that evolve very fast or do not fall into enriched regions.

To test the validity of our approach, we applied it to search for the counterparts of MHC, neurotrophin, and JN paralogons in a vertebrate species. A similar procedure as described above for T. adhaerens was followed for chicken (Gallus gallus), and it successfully identified MHC paralogons on chromosomes 8, 10, 17, 25, 28, and Z and putative neurotrophin and JN regions (Supplemental Fig. 1). These data show that a method that successfully maps MHC paralogons in the chicken also identified candidate counterparts of proto-MHC and proto-neurotrophin regions in the T. adhaerens genome.

Confirmation of a proto-MHC in T. adhaerens by macrosynteny, phylogenetic analyses, and a statistical test of gene distribution
The T. adhaerens genome shows extensive large-scale conservation in genomic organization between placozoans (T. adhaerens) and chordates (35). One of the most prominent conserved segments is located on scaffold 2, matching the region identified in this study as the main counterpart of the MHC and MHC ohnologs. For further analysis, we therefore selected each ACLG corresponding to the putative T. adhaerens proto-MHC regions from reference (35). The segments of human chromosomes matching
these ACLG are shown in Table I. Our candidate regions matched ACLG 8, 10, and 11, which contain essentially all segments of human MHC or MHC ohnologs as described in (49), namely 1p21.1-34.2, 1q23.3-32.1, 5p15.33-q13.2, 5q13.2-31.1, 6p21.2-22.2, 9q22.3-24.3, 9q22.31-34.3, 15q15.3-26.3, 19p13.1-13.2 (Table I). Hence, the common origin of the proto-MHC region identified in this study, and of the MHC set of paralogons described in vertebrates, is confirmed by the conserved macrosynteny mapping and the extensive phylogenetic analysis reported by Srivastava et al. (35).

In addition, we independently assessed the validity of the correspondence between T. adhaerens proto-MHC and human MHC paralogons using a statistical test for the genomic distribution of

FIGURE 1. Metazoans evolutionary relationships, definitions, and human MHC-related regions. (A) Phylogeny of metazoans based on GIGA Community of Scientists (37). (B) Definitions of phylogenetic terms used in the text. (C) Schematic representation of MHC-related regions belonging to the major histocompatibility, neurotrophin, and JN complexes of the human genome.
Ohno’s theory of whole genome duplications (20) predicts that many human MHC markers with paralogous versions on scaffolds would generally have a unique counterpart in the genome of a prototypic nonpolyploid invertebrate located in a proto-MHC region. We evaluated whether our results were consistent with this prediction by listing T. adhaerens RBBH of at least one gene from human sets of ohnologs with representatives in three or four paralogs. There were three unique T. adhaerens genes corresponding to sets of paralogs represented on four paralogs among the 33 such sets present in the human genome and 10 corresponding to sets of paralogs present on three paralogs, among the 106 such sets found in human (Fig. 3). These markers were mainly located in T. adhaerens scaffold 2 and include the homologs of several sets of MHC markers such as RXXR, PBX1, and VAV1 (25, 29, 49, 57, 58). Among 443 sets of paralogs with representatives on at least two of the MHC paralogs in human, 129 (~30%) had a unique counterpart in the conserved regions of T. adhaerens. Thirteen additional sets of paralogs had multiple targets in T. adhaerens due to secondary duplications. Overall, the observations indicated that approximately one third of the multiple marker sets have a unique counterpart in T. adhaerens, as predicted by Ohno’s model of whole genome duplications during early vertebrate evolution.

Table I. Phylogenetic analyses of the T. adhaerens proto-MHC support its relationship with human MHC and MHC paralogs

<table>
<thead>
<tr>
<th>Regions of the Putative T. adhaerens Proto-MHC (Scaffold:Location in Mb)</th>
<th>Corresponding ACLGs (11)</th>
<th>Location of the ACLG Blocks in Homo Sapiens (11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2:2,000,000–2,500,000</td>
<td>10 (1.43E-19)×a</td>
<td>1p22; 1q23-32; 6p21-22; 9q32-34×a</td>
</tr>
<tr>
<td>2:5,000,000–3,500,000</td>
<td>10 (1.24E-10)×a</td>
<td>1p22; 1q23-32; 6p21-22; 9q32-34×a</td>
</tr>
<tr>
<td>2:3,500,000–4,000,000</td>
<td>10 (4.04E-17)×</td>
<td>1p22; 1q23-32; 6p21-22; 9q32-34×</td>
</tr>
<tr>
<td>2:4,000,000–4,500,000</td>
<td>10 (7.55E-22)×v</td>
<td>1p22; 1q23-32; 6p21-22; 9q32-34×</td>
</tr>
<tr>
<td>2:4,500,000–5,000,000</td>
<td>10 (9.20E-17)×v</td>
<td>1p22; 1q23-32; 6p21-22; 9q32-34×</td>
</tr>
<tr>
<td>2:5,000,000–5,500,000</td>
<td>16 (3.59E-08)</td>
<td>2q13; 3p21; 7p11; 7q11; 7q36; 10p13; 12q13; 17q12; 17q23</td>
</tr>
<tr>
<td>2:5,500,000–6,000,000</td>
<td>5 (2.49E-04)</td>
<td>7p21; 16p11; 17p11; 17q21; 17q24; 22q12</td>
</tr>
<tr>
<td>3:1,500,000–2,000,000</td>
<td>16 (1.70E-05)</td>
<td>2q13; 3p21; 7p11; 7q11; 7q36; 10p13; 12q13; 17q12; 17q23</td>
</tr>
<tr>
<td>3:4,000,000–4,500,000</td>
<td>11 (1.76E-04)×v</td>
<td>1p12-q23; 5q13-31; 9p22-q22; 15q15-26; 19p13×</td>
</tr>
<tr>
<td>3:4,500,000–5,000,000</td>
<td>11 (8.43E-21)×</td>
<td>1p12-q23; 5q13-31; 9p22-q22; 15q15-26; 19p13×</td>
</tr>
<tr>
<td>3:5,000,000–5,500,000</td>
<td>11 (3.75E-08)×</td>
<td>1p12-q23; 5q13-31; 9p22-q22; 15q15-26; 19p13×</td>
</tr>
<tr>
<td>3:1,500,000–2,000,000</td>
<td>8 (6.42E-11)×</td>
<td>1p21-34; 5p15-q13; 9p22-24; 9q32-32; 19p13×</td>
</tr>
<tr>
<td>7:1–500,000</td>
<td>8 (1.51E-05)×</td>
<td>1p21-34; 5p15-q13; 9p22-24; 9q32-32; 19p13×</td>
</tr>
<tr>
<td>7:500,000–1,000,000</td>
<td>8 (5.16E-12)×</td>
<td>1p21-34; 5p15-q13; 9p22-24; 9q32-32; 19p13×</td>
</tr>
<tr>
<td>9:1–500,000</td>
<td>8 (8.16E-12)×</td>
<td>1p21-34; 5p15-q13; 9p22-24; 9q32-32; 19p13×</td>
</tr>
<tr>
<td>9:500,000–1,000,000</td>
<td>8 (4.11E-14)×</td>
<td>1p21-34; 5p15-q13; 9p22-24; 9q32-32; 19p13×</td>
</tr>
<tr>
<td>10:1,000,000–1,500,000</td>
<td>3 (2.72E-06)</td>
<td>4q12; 4q35; 5q31; 10q11</td>
</tr>
<tr>
<td>10:1,500,000–2,000,000</td>
<td>8 (4.85E-08)×</td>
<td>1p21-34; 5p15-q13; 9p22-24; 9q32-32; 19p13×</td>
</tr>
<tr>
<td>10:2,000,000–2,435,506</td>
<td>8 (3.07E-08)×</td>
<td>1p21-34; 5p15-q13; 9p22-24; 9q32-32; 19p13×</td>
</tr>
<tr>
<td>15:1–500,000</td>
<td>11 (8.71E-11)×</td>
<td>1p12-q23; 15q15-26; 5q13-31; 9p22-q22; 19p13×</td>
</tr>
<tr>
<td>15:500,000–1,000,000</td>
<td>16 (4.06E-05)</td>
<td>2q13; 3p21; 7p11; 7q11; 7q36; 10p13; 12q13; 17q12; 17q23</td>
</tr>
</tbody>
</table>

×Linkage groups corresponding to MHC paralogs in human.

The p values for significant numbers of ancestral genes are shared between the segments of T. adhaerens putative proto-MHC and corresponding ACLGs (11).
Microsynteny groups conserved between the human MHC ohnologs and the T. adhaerens proto-MHC do not systematically colocalize with key developmental master genes

In addition to large-scale synteny conservation between human MHC-related paralogons and T. adhaerens proto-MHC, there were nineteen conserved microsynteny gene sets implicating three or more genes with five having five or more genes. Marker genes involved in such sets were located in the same region, but the colinearity between humans and T. adhaerens was generally not conserved. A striking example was found on human chromosome 9q32-q34.3 region where 65 markers had a counterpart in T. adhaerens scaffold 2 (between 1.9 and 5.2 Mb, see Supplemental Table II), consistent with the MHC paralogon on human chromosome 9 having retained most of the ancestral configuration (i.e., the plesiomorphic organization) (33, 59). In addition, scaffold 2 has four microsyntenies with the RGS1/RGS16 region located on human chromosome 1 (50).

Large gene deserts with enhancers acting over long distances are mostly found in vertebrate genomes, whereas invertebrate metazoans generally have local regulatory controls of expression (60). However, it has been proposed that locked genomic regulatory blocks (GRBs) defined by key developmental transcription factors and their distal enhancers provide an explanation for the maintenance of long-range conserved synteny across vertebrate and invertebrate genomes (61, 62). It is proposed that bystander genes be trapped in the GRBs and thus form conserved syntenic blocks of genes (61, 62). These bystander genes are unrelated to the developmental transcription factor gene defining the GRBs in terms of function, regulation and phylogenetics. According to this hypothesis, observed syntenic regions of the genome are expected containing key developmental transcription factor genes. To test this hypothesis, we looked for such genes in the microsyntenic regions we identified and found no significant enrichment in transcription factors in the microsyntenies of T. adhaerens. Six transcription factor genes (PBX3 [pre-B-cell leukemia homeobox 3], RXRA [retinoid X receptor α], PRDM12 [PR domain-containing protein 12], GTF3C5 [general transcription factor 3C polypeptide 5], ED1 [endodermal differentiation-related factor 1], and COBRA1 [cofactor of BRCA1]) were found in the large conserved gene set (60 genes) from chromosome 9. However, among the other 18 gene sets from conserved microsyntenies, only 3 encompass a transcription factor (DMRTA2 = DMRT5 [double sex- and mab-3–related transcription factor 5], GTF2B [general transcription factor 2B], GTF2C [general transcription factor 2C], and GTF2D [general transcription factor 2D]).
and CHAF1A [chromatin assembly factor 1, subunit A]). With the exception of the large microsynteny group from chromosome 9q34, none of the five longest gene sets (five genes or more) encoded transcription factors.

We also looked for noncoding sequences conserved between the T. adhaerens proto-MHC and the human genome to identify putative regulatory regions. We did not find significant conserved noncoding sequence with obvious regulatory potential. Only two of the 94 sequences found with Blastn and tblastX showed significant conservation in at least one of selected genomes from representative metazoans: Amphimedon queenslandica (Porifera), Nematoscelis vectensis (Cnidaria), Capitella teleta and Helodella robusta (Annelida), Crassostrea gigas and Lottia gigantea (Mollusca), and Stronglylocentrotus purpuratus (Echinodermata). The first one, PCNC553, a 68 bp sequence from T. adhaerens scaffold 7, matched an exon of the human WDR65 gene located on human chromosome 1 in an MHC paralogon (e value = 8 × 10^-10 with exon 12 of transcript ENST00000372492) and the WDR65 gene in A. queenslandica (Aql1.205499; Aql1.217579), N. vectensis (NEMVEDRAFT_v1g213283), C. teleta (CapteG108742), L. gigantea (LotgG118162), and S. purpuratus (SPU_010131). A detailed analysis revealed that this motif represents a non-annotated exon flanking the gene TriadG27693, one of the two WDR65 genes found in T. adhaerens close to each other on Scaffold7. The second hit was PCNC589, a 38-bp sequence from T. adhaerens scaffold 3 that was found in H. robusta, C. gigas, and L. gigantea. This sequence appeared to be located in a non-annotated exon at the 5' end of a well-conserved gene, the sodium channel scn.

Taken together these results argue against a predominant role of transcription factor genes in the maintenance of the microsynteny gene sets conserved between humans and T. adhaerens.

The gene content of the T. adhaerens proto-MHC reveals that this region was involved in immunity in the placozoan-bilaterian ancestor

A tentative set of primordial MHC markers was inferred based on gene conservation between the proto-MHC of T. adhaerens and one or several human MHC paralogons. A functional ontology analysis of this gene set was then performed using BiNGO, a plugin of Cytoscape that maps the predominant functional themes of a gene set on the GO hierarchy (53), and computes enrichment calculations in comparison with a reference list. BiNGO identified the key GO terms that were overrepresented in the T. adhaerens proto-MHC, in reference to the whole set of genes present in all human MHC-related paralogons. Two groups of GO terms potentially linked to defense processes were found (Table II) (1): proteasome and ubiquitination (15 genes out of 307) and (2) stress response (25 genes of 307). This analysis also identified several enriched GO terms related to metabolic processes, especially RNA metabolism and gene expression control. We then analyzed the list of human genes present in MHC-related paralogons without counterpart within the T. adhaerens proto-MHC region. In contrast to those with counterpart, this list was significantly enriched in GO terms with function in adaptive immunity, indicating that many MHC (and MHC paralogon) genes important for immune pathways appeared late in the evolution of metazoans (e.g., MHC class I and class II). This list was also enriched in representatives related to receptors and cell-to-cell recognition, consistent with their absence in T. adhaerens where only four cell types have been identified (45).

The results of our GO analysis led us to look for genes encoding proteins with B30.2 domains in T. adhaerens. The B30.2 domain, a fusion of a PRY and a SPRY motif (2, 3), is often found in key proteins of immunity encoded in the MHC and MHC related regions of vertebrates and its presence in genes of T. adhaerens MHC-related regions would represent another link to immunity. In T. adhaerens, 12 proteins with a B30.2 domain were found (Fig. 4). Seven proteins containing a B30.2 domain were located within the proto-MHC or neurotrophin regions, including a TRIM protein with a Ring-B.Box-CC-FN-B30.2 domain structure similar to the one reported in Nematoscelis vectensis (63). This gene is located in the proto-neurotrophin region on scaffold 22, 200 kb and 400 kb from proteasome genes psmc1 and psmd8, respectively. Most genes with a B30.2 domain possess at least one putative human ortholog within MHC or neurotrophin paralogons. Fig. 4 shows that T. adhaerens B30.2 sequences cluster with their respective human orthologs with high bootstrap values, indicating that they were already specialized/differentiated into subfamily-specific domains in the common ancestor of placozoans and vertebrates.

Absence of a genomic region enriched in proteasome and stress response genes in the choanoflagellate Monosiga brevicollis

To investigate whether proto-MHC region was a metazoan innovation, we searched for counterparts of human MHC-related markers in the choanoflagellate Monosiga brevicollis (genome size 41.6 Mb, 9200 genes) (64). Choanoflagellates are highly

Table II. GO analysis identifies overrepresentation of terms related to proteasome/ubiquitination and stress response within genes conserved between T. adhaerens proto-MHC and human MHC paralogons

<table>
<thead>
<tr>
<th>Pathway</th>
<th>GO Category Accession</th>
<th>GO Category Description</th>
<th>Genes with a Counterpart in the Trichoplax Proto-MHC, Mapped within These GO Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteasome / ubiquitination</td>
<td>6511</td>
<td>Ubiquitin-dependent protein catabolic process</td>
<td>ARHI1, CDC34, CLPX, EDEM3, FAF1, FZR1, HSPA5, LONP1, PSMA4, PSMB7, PSMD5, RNF11, TOPORS, USP3, USP33</td>
</tr>
<tr>
<td></td>
<td>10498</td>
<td>Proteasomal protein catabolic process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19941</td>
<td>Modification-dependent protein catabolic process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>43161</td>
<td>Proteasomal ubiquitin-dependent protein catabolic process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>43632</td>
<td>Modification-dependent macromolecule catabolic process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>51603</td>
<td>Proteolysis involved in cellular protein catabolic process</td>
<td></td>
</tr>
<tr>
<td>Stress response</td>
<td>6281</td>
<td>DNA repair</td>
<td>ARNT, ATF6, ATG10, BLM, CCNH, CDK7, CHAF1A, COLAA3BP, DCLRE1B, DIXH9, FAN1, FZR1, GNL1, HSPA5, INT5S, LONP1, MAP2K7, MORF4L1, POLG, PRDX6, RXRA, TOPORS, USP1, USP3, XAB2</td>
</tr>
<tr>
<td></td>
<td>33554</td>
<td>Cellular response to stress</td>
<td></td>
</tr>
</tbody>
</table>

*GO categories that are also relatively underrepresented among MHC-related human markers without a counterpart in T. adhaerens proto-MHC.
similar to the choanocytes of sponges, and are considered the closest known relatives of metazoans (65–67). Among MHC-related markers with RBBH in T. adhaerens, approximately one third had a counterpart in M. brevicollis (Fig. 5A). Using BiNGO as above, we looked for enriched GO terms within the different gene subsets and found only markers for cellular response to stress for RBBH present in T. adhaerens, whereas markers for proteasome and ubiquitination were enriched both in T. adhaerens and M. brevicollis. We then examined whether these genes involved in proteasome and ubiquitination might be grouped in a single genomic region within M. brevicollis. Fig. 5B shows that these genes are largely dispersed among M. brevicollis scaffolds, resulting in the absence of a clustering of genes involved in proteolysis and ubiquitination in this species. However, a subset of MHC-related genes present on the human chromosome region 9q32-34 (59)—the most conserved region among MHC paralogons across vertebrates—appears to be located in the same synteny group in humans, T. adhaerens and M. brevicollis (Fig. 5C). These markers include two markers of human MHC tetrads, VAV and RABGAP1, but are not obviously related to immunity. A more extensive analysis of the genome of M. brevicollis will be required for a comprehensive picture of the distribution of the markers linked to the genes of metazoan MHC and proto-MHC related markers. However, our observations provided no evidence for a proto-MHC region containing genes involved in innate immunity, proteasome function, and stress responses in this species.

Discussion
The MHC of vertebrates is a large genetic region that determines Ag recognition by T lymphocytes, graft compatibility, and contains genes encoding receptors, cytokines, and effectors of innate immunity. One approach to understand the functional significance of the components of the vertebrate MHC is to reconstruct its evolutionary history. We report a proto-MHC with archetypal markers in one representative of the most basal branch of metazoans, the placozoan T. adhaerens. The presence of a proto-MHC exists in the common ancestor of deuterostomes and protostomes as revealed by studies in Drosophila. Our results show that it also exists in more primitive branches of the animal kingdom. Placozoan proto-MHC markers include good homologs of genes key to antiviral immunity, stress response, and ubiquitination/proteasome, suggesting that the appearance of class I and II molecules and Ag presentation pathway in vertebrates took advantage of the molecules encoded in this region. In contrast, we did not find evidence of a proto-MHC in the genome of a choanoflagellate, one of the closest known relatives of metazoans among unicellular organisms; it is therefore tempting to speculate that the proto-MHC as a genetic region is an innovation of metazoans, like other key features and pathways such as the TBCEL/coel-1 dependent microtubule function during development and neuronal differentiation (68). However, because the presence of a proto-MHC has not been evaluated in other opisthokonts (e.g., fungi), the apparent lack of proto-MHC in M. brevicollis could be due to the loss of the primordial linkage in this species and does not constitute definitive evidence for the absence of proto-MHC in unicellular organisms. A comprehensive survey of protozoan genomes will be necessary to clarify this issue.

The hypothesis of the presence of proteasomes—ancient components involved in the cellular stress response—in the primordial MHC (23) is consistent with our genetic analysis of T. adhaerens. The proteasome genes located in T. adhaerens are not the functional counterparts of the specialized immunoproteasomes found in vertebrates. Nonetheless, we speculate they are coregulated with genes involved in stress response, pathogen binding, and

**FIGURE 4.** Distance tree of B30.2 domains found in T. adhaerens. Protein sequences were aligned using Clustal W, and a distance tree was computed using Mega5 (neighbor joining, pairwise deletion, bootstrap value = 1000).
ubiquitination, and their presence in the region offers the possibility they have been co-opted and selected in the bona fide MHC during vertebrate evolution. In fact, the T. adhaerens psmb-like corresponds to the subset from which immunoproteasomes were derived. In addition, the early peptide presentation system may have lacked specialized and inducible proteasomes, as constitutive proteasomes can generate peptides that are presented by MHC I, although with lower efficiency (69, 70).

The association of genes of stress response, ubiquitination, and protein catabolism (proteasome) within the proto-MHC—possibly as a coregulated unit—in T. adhaerens is consistent with an ancient functional link between these pathways; stress response induces ubiquitination of pathogens or cellular proteins that are either redirected to new compartments or to degradation by the proteasome. Such a stress/ubiquitination/proteasome cascade has been described in C. elegans; DNA damage to germ cells induces a response that elicits resistance to stress as well as activation of the ubiquitination–proteasome system in somatic cells in various tissues (71). This fundamental inflammatory response can be involved in several processes, including aging, adaptation, and defense against pathogens. A specific implication of the proto-MHC–neurotrophin in immunity is further supported by the presence of several B30.2 proteins in these regions. We previously proposed that the association of B30.2 domains with key proteins of immunity (e.g., TRIM, butyrophilin) found in the MHC and MHC-related regions of vertebrates could have an ancient origin (63). The B30.2 domain structurally resembles a β-barrel (2, 3) and allows specific recognition of ligands via the loops at the top of the domain. Of note, several B30.2 domains found in T. adhaerens are associated with domains that possess E3 ubiquitin ligase activity, such as the RING in TRIMs. In particular, the T. adhaerens TRIM has multiple human co-orthologs within paralogs of the MHC, neurotrophi or JN sets, and all of them share the same domain structure (72). Among them are TRIM1 and TRIM9, two key modulators of the IFN pathway that strongly inhibit viral growth (73). In the absence of IFN, an antiviral activity of the T. adhaerens TRIM might proceed via direct binding (and ubiquitination) of viral proteins as for TRIM5 in primates, or via modulation of expression of other antiviral factors. Although TRIM genes and B30.2 domains are conservatively associated with vertebrate MHCs (and paralogs) (7, 63), it is striking that they are also present in the MHC/neurotrophin regions in the ancestor of placozoans and bilaterians. More generally, our data show a preferential location of B30.2 proteins within the proto-MHC linkage to the one of Hox/paraHox genes.

In addition to the stress/ubiquitination/proteasome system and B30.2 domains, we also found the T. adhaerens proto-MHC has several genes whose human counterparts are located on chromosome 1 in the “RGS1/RGS16 region,” including a typical rgs-r4–like gene. In humans, comparative and responsiveness quantitative trait loci analyses show that this region is critical for antiviral defense (50, 74). The markers from the human RGS1/RGS16 region that are conserved in the T. adhaerens proto-MHC might be components of an ancient antiviral pathway.

Our observations do not reveal the evolutionary mechanisms responsible for the linkage conservation of genes involved in immunity within MHC related regions during evolution. However, it is interesting to compare the evolution of the genes found conserved in the proto-MHC linkage to the one of Hox/paraHox genes.
genes. Hox /paraHox are clustered in most metazoans because they derive from old successor duplications and their sequential expression is highly regulated (75). In contrast, the constituents of the proto-MHC linkage are multiple, do not derive from the same ancestor gene, and are not found in the same order in different species, suggesting that they are subjected to different constraints.

In conclusion, T. adhaerens has retained a simple and primitive organization at both genetic (genomic) and organism levels, and its proto-MHC may reflect the primordial architecture and the functional landscape of this region, which later in evolution became associated with a large number of genes critical for the adaptive immunity in vertebrates.

Acknowledgments
We thank Drs. Franck Bourrat, Daniel Chourrout, Simon Fillatreau, Maria Leptin, and Jean-Pierre Levrard for helpful comments and critical reading of the manuscript, and Dr. Howard Etinger for improving the English usage.

Disclosures
The authors have no financial conflicts of interest.

References

Downloaded from http://www.jimmunol.org/ by guest on April 25, 2022


