



COVID-19 Research Tools

Defeat the SARS-CoV-2 Variants

InvivoGen



Cutting Edge: CATERPILLER: A Large Family of Mammalian Genes Containing CARD, Pyrin, Nucleotide-Binding, and Leucine-Rich Repeat Domains

This information is current as of August 4, 2022.

Jonathan A. Harton, Michael W. Linhoff, Jinghua Zhang and Jenny P.-Y. Ting

J Immunol 2002; 169:4088-4093; ;
doi: 10.4049/jimmunol.169.8.4088
<http://www.jimmunol.org/content/169/8/4088>

Supplementary Material <http://www.jimmunol.org/content/suppl/2002/10/09/169.8.4088.DC1>

References This article cites **37** articles, 16 of which you can access for free at:
<http://www.jimmunol.org/content/169/8/4088.full#ref-list-1>

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

**average*

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>



Cutting Edge: CATERPILLER: A Large Family of Mammalian Genes Containing CARD, Pyrin, Nucleotide-Binding, and Leucine-Rich Repeat Domains¹

Jonathan A. Harton,² Michael W. Linhoff,² Jinghua Zhang, and Jenny P.-Y. Ting³

Large mammalian proteins containing a nucleotide-binding domain (NBD) and C-terminal leucine-rich repeats (LRR) similar in structure to plant disease resistance proteins have been suggested as critical in innate immunity. Our interest in CIITA, a NBD/LRR protein, and recent reports linking mutations in two other NBD/LRR proteins to inflammatory disorders have prompted us to perform a search for other members. Twenty-two known and novel NBD/LRR genes are spread across eight human chromosomes, with multigene clusters occurring on 11, 16, and 19. Most of these are telomeric. Their N termini vary, but most have a pyrin domain. The genomic organization demonstrates a high degree of conservation of the NBD- and LRR-encoding exons. Except for CIITA, all the predicted NBD/LRR proteins are likely ATP-binding proteins. Some have broad tissue expression, whereas others are restricted to myeloid cells. The implications of these data on origins, expression, and function of these genes are discussed. *The Journal of Immunology*, 2002, 169: 4088–4093.

A number of genes with nucleotide-binding domain (NBD)⁴ and leucine-rich repeat (LRR) domains are rapidly emerging as important in apoptosis, immune, and inflammatory disorders. These include *CIITA*, *Nod1/CARD4*, *Nod2/CARD15*, *DEFCAP/CARD7/NALP1*, and *CIAS1/PYPAFI*. *CIITA*, *Nod2*, and *CIAS1* are linked to a number of immunologic disorders. *CIITA* is the master transcriptional regulator of class II MHC. Genetic lesions in *CIITA* cause an immunodeficiency, type II bare lymphocyte syndrome (group A) (1). In the past year, mu-

tations in *Nod2* and *CIAS1* have been linked to four immunologic and inflammatory disorders (2–5). This raises the intriguing possibility that other family members may be important in similar disorders. We have found, along with these known genes, a large family of genes coding for proteins of similar structure, many of which occur in clusters on individual chromosomes. Additionally, many of these genes contain a pyrin domain, a motif first described in a gene associated with familial Mediterranean fever (an episodic fever syndrome), now associated with apoptosis and inflammation (6, 7).

CIITA was isolated using a complementation cloning strategy to restore MHC II expression to a MHC II-deficient cell line (1). *CIITA* is a master regulator of transcription, responsible for both IFN- γ and constitutive expression of MHC II and related genes (8, 9). The N-terminal activation domain of *CIITA* is necessary for transcriptional activation (8). The centrally located NBD of *CIITA* contains a GTP-binding domain required for nuclear import (8). *CIITA* undergoes self-association involving sequences in its NBD, C-terminal LRRs, and N terminus (10).

When *CIITA* was first discovered, initial searches for *CIITA*-related genes produced no significant matches. *Nod1*, an activator of caspase-9-mediated apoptosis and NF- κ B, also having an NBD and C-terminal LRRs, was the first described protein similar to *CIITA* in domain organization (11, 12). *Nod2*, with functions similar to those of *Nod1*, has been strongly implicated in Crohn's disease (2, 3, 13) and in familial granulomatous synovitis (Blau syndrome) (14). Most recently, patients with familial cold autoinflammatory syndrome (familial cold urticaria) and Muckle-Wells syndrome were found to have mutations in a new gene called *CIAS1*, which has a pyrin domain, NBD, and LRR (4). These syndromes are associated with a *CIAS1* splice variant called cryopyrin. These proteins may be similar to plant disease resistance proteins (R proteins) which detect pathogens and initiate defense mechanisms including MAP kinase activation, oxygen radical formation, salicylate production, induced transcription of kinases and transcription factors, and rapid cell death (15). Many of these plant proteins have an NBD and LRRs and may represent the oldest examples of proteins using this *CIITA*-like domain arrangement.

The advent of the nearly complete human genome sequence facilitated a search for sequences related to these proteins. We describe the identification of additional putative mammalian NBD/LRR proteins similar to the known family members. Including those already known, this analysis predicts at least 22 NBD/LRR genes in the human genome, which we call the CATERPILLER

Lineberger Comprehensive Cancer Center and Department of Microbiology and Immunology, University of North Carolina, Chapel Hill, NC 27599

Received for publication July 23, 2002. Accepted for publication August 22, 2002.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ This work was supported by National Institutes of Health Grants AI29564, AI45580, AI41751, and DK38108 (to J.P.-Y.T.).

² J.A.H. and M.W.L. contributed equally to this work.

³ Address correspondence and reprint requests to Dr. Jenny P.-Y. Ting, Campus Box 7295, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599-7295. E-mail address: panyun@med.unc.edu

⁴ Abbreviations used in this paper: NBD, nucleotide-binding domain; LRR, leucine-rich repeat; est, expressed sequence tag; CATERPILLER, CARD, transcription enhancer, R (purine)-binding, pyrin, lots of leucine repeats.

(CARD, transcription enhancer, R(purine)-binding, pyrin, lots of leucine repeats) gene family. Determination of chromosomal location, genomic organization, and sequence information are also shown.

Materials and Methods

Databases and search strategies

Searches were performed using the published Celera human genome scaffold data (16), the National Center for Biotechnology Information (NCBI) "nr" database (containing GenBank, European Molecular Biology Laboratory, DNA Data Base in Japan, Protein Data Base, and completed phase 3 and 4 high-throughput genomic sequencing (HTGS) sequences), and the NCBI genome database, (17). Initial searches were performed using the B cell form of CIITA protein sequence (1) as a query using the BLAST search algorithms BLASTP and TBLASTN (see supplemental data Fig. 1).⁵ BLASTP identifies amino acid sequence similarities through query sequence comparison with database proteins and is more likely to find distant relationships than BLASTN (18). TBLASTN compares the query protein sequence with translations of all six reading frames of available nucleotide sequences and has the same advantages as BLASTP. We used analogous domains of the resultant sequences to identify additional sequences and/or confirm initial identities; this is known as DOUBLE-BLAST, inspired by the intermediate search sequence method of Park et al. (19, 20) and is comparable in homologue detection with Hidden Markov Methods. LRR sequences, the N-terminal pyrin domains of DEFCAP, and the CARD domains of Nod1 and Nod2 were used to perform similar searches. The N-terminal sequences of CIITA yielded no related sequences obviously belonging to an NBD/LRR protein.

Assembly of putative novel genes and construction of genomic maps

Pyrin and LRR sequences identified within contigs containing NBDs were examined for location and orientation to determine the likelihood of residing in the same operon as an identified NBD. Pyrin and LRR domains were considered contiguous with an NBD if they fell upstream and downstream of the NBD, respectively, in the same orientation. CARD domains occur both upstream (Nod1/2) and downstream (DEFCAP) of the NBD (21), but none of the novel sequences contained CARD domains. As sequence data became available for more than a single domain, a putative genomic organization was generated by comparing the cDNA sequence with the genome sequence.

Cell lines, preparation of RNA, and RT-PCR

HeLa, MCF7, Jurkat, RAJI, and RAMOS cell lines were cultured in either DMEM (high glucose) or RPMI 1640 with 10% FCS, L-glutamine, and penicillin/streptomycin. Peripheral blood leukocytes were obtained as buffy coats from the American Red Cross (Durham, NC). Total RNA was prepared using the SV Total RNA Isolation kit (Promega, Madison, WI). Total RNA was reverse transcribed to cDNA using Moloney murine leukemia virus reverse transcriptase and amplified in an MJ Thermocycler (MJ Research, Cambridge, MA) in a separate reaction with primers specific for each target sequence. Amplification products were electrophoresed on 0.8% agarose and visualized with ethidium bromide.

Results and Discussion

Identification of novel CIITA-related sequences

BLAST searches of the published Celera and NCBI genomic databases using the NBD and LRR of CIITA, Nod1, Nod2, DEFCAP, and resultant target sequences as queries revealed 22 potential genes and pseudogenes, including the presently known genes, unified by the presence of an NBD and downstream LRRs (Table I). New genes were assigned a name based on chromosome number and order of discovery (e.g., 19.1, first found on chromosome 19). Nod1, Nod2, and DEFCAP contain CARD domains that may be involved in recruiting caspases (12, 13, 21). DEFCAP also has an N-terminal pyrin domain with homology to the familial Mediterranean fever protein (7). BLAST searches were also per-

formed for the CARD domains of Nod1/2, the pyrin domain of DEFCAP, and resulting target sequences. CARD domain homologues were not found for any of the novel sequences. The majority of the putative genes have upstream pyrin domains, but the upstream N-terminal sequences of several remain unknown.

Conservation of intron-exon organization

We determined exon/intron sizes and positions for the known and some predicted NBD/LRR proteins by the location of the sequence corresponding to the mRNA/cDNA assuming intactness of the contig (Fig. 1A). The genomic organization is complex and remarkably similar for all the sequences examined, with large NBD exons (~1500 nt) and LRR exons of ~76 nt, 174 nt, or both depending of the gene. CARD and pyrin domains are ~300 nt long.

CATERPILLER domains

Table I highlights the distinct domains of each sequence. Nod1, Nod2, and CIITA have N-terminal CARD or CARD-like domains. Thirteen have N-terminal pyrin domains. CIITA is unique in having an N-terminal acidic *trans* activation domain. Five of these sequences do not have CARD, pyrin, or CIITA-like activation domains upstream of their NBDs. The diversity of these N-terminal sequences suggests multiple functional modes.

The predicted nucleotide specificity based on motifs found in the CATERPILLER genes is shown in Table I. This is compared with another family, containing plant and animal proteins, grouped on the basis of a NTPase domain and C-terminal repeats of either the LRR or WD40 type, called the NACHT family, which includes NAIP, CIITA, HET-E, and TP1 (22). Remarkably, the majority are predicted to be ATP-binding proteins, with the exception of CIITA, which binds GTP, and HET-E. A GTP-binding protein-like magnesium coordination (G3) motif (DXXG) occurs in a number of the other sequences, but excepting the more distantly related Apaf1, it follows the more typical kinase 2 site found in ATP-binding proteins.

We aligned the NBDs of these predicted proteins, each ~500 aa long, and observed 12 groupings of conserved residues (motifs) (Fig. 1B). The full protein alignment of the NBD domains is shown in supplemental data Fig. 2. Although the seven NACHT motifs are present, the larger number of compared sequences permits a refined definition of the NACHT domain that excludes WD40 repeat-containing members, thus distinguishing a CATERPILLER NBD from the broader NACHT family. These motif definitions also suggest a divergence between the majority of the NBDs that we describe and those like NAIP. Functionally important motifs likely include motif I, which contains the Walker A sequence found in most nucleotide-binding proteins (23), and motifs III and V that overlap or are adjacent to leucine-charged domain motifs (24). These motifs are important for CIITA function (8). Motif III contains the kinase 2 motif which coordinates magnesium ions in ATP-binding proteins (23).

The presence of LRR sequences downstream of the NBD was required for inclusion as a CATERPILLER family member. The LRR sequences following NBDs have two exon arrangements, a singlet (~74 nt) containing one motif iteration or a duplex (~180 nt) containing two (Table I (column 8), Fig. 1A, and supplemental data Fig. 3). The sole absolute requirement for inclusion as an LRR is conservation of the hydrophobic residues "leucines" comprising the motif. BLAST searches for LRRs may miss some sequences due to a greater likelihood of less similarity between non-LRR-motif residues. Thus, without actual cDNA clones, it is impossible to be highly confident that all of the LRR exons downstream of the

⁵ The on-line version of this article contains supplemental material.

Table I. Summary of domain characteristics^a

	N Terminus	P-Loop (Kinase 1/G1) ^b	GTP-Mg ²⁺ (G3) ^c	ATP-Mg ²⁺ (Kinase 2) ^d	Guanine Binding (G4) ^e	Predicted Nucleotide Specificity	LRR
1.1/CIAS1	Pyrin	GAAGIGKT		LFMD		ATP	Duplex
Nod1	CARD	GDAGVGKS		LFTFD		ATP	Single
11.1	Pyrin	GSAGTGKT		LFILD		ATP	Single
11.2	Pyrin	GAAGVGKT		LFID		ATP	Duplex
11.4	Pyrin	GPAGIGKT		LFILD		ATP	Duplex
11.3	?	GTVGTGKS					Nonuniform
<i>J2</i>	Pyrin	None		LFIMD			Single/duplex
CIITA	CARD, acidic	GKAGQGKS	DAYG	LLILD	SKAD	GTP ^g	Single
Nod2	CARD ×2	GEAGSGKS		LFTFD		ATP	Single
16.1	?	GKAGMGKT		LLIFD		ATP	Single
16.2	?	GVAGMGKT		LLILD		ATP	Single
Nalp1/DEFKAP	Pyrin	GAAGIGKS	DEPG ^f	LFILD		ATP	Single/duplex
Nalp2	Pyrin	GPAGLGKT	DELG ^f	LFVID		ATP	Duplex
19.1	?	GPDGIGKT		LFIMD		ATP	Duplex
19.2	Pyrin	GAPGIGKT		LLLLD		ATP	Duplex
19.3	Pyrin	GAAGIGKS		LFID		ATP	Duplex
19.4	Pyrin	GPAGVGKT	DICG ^f	LFVID		ATP	Duplex
19.5	Pyrin ×2	GPQGIGKT		LFVID		ATP	Duplex
19.6	Pyrin	GERASGKT		LFILD		ATP	Duplex
19.7	Pyrin	GRAGVGKT		LFID		ATP	Duplex
19.8	?	GKSGIGKS	DDLG ^f	LFID		ATP	Duplex
<i>X</i>	?	ACAGTGKT	DPVG ^f	LLILD			Duplex
Apaf1		GMAGCGKS	DKSG	LLILD		dATP ^g /ATP ^g	WD40
RPM1		GMGSGKT		IVVLD		ATP	LRR
NAIP		GEAGSGKT		LFLLD		ATP	LRR
HET-E		GDPGKKT	DHAG	YLIID	TKHD	GTP/ATP	WD40
TP1		QSGQGKT	DQNG ^f	VLIID		ATP	WD40
Gα 12		GAGESGKS	DKLG		SKQD	GTP ^g	

^a NAIP, CIITA, HET-E, and TP1 are the defining members of the NACHT family (see text). Apaf1, RPM1, NAIP, HET-E, TP1, and Gα 12 are shown for comparison purposes. Pseudogenes and suspected pseudogenes are shown in italics. ×2, Two copies.

^b Consensus P-loop motif, GXXXXGK(S/T).

^c Consensus Mg²⁺ site (G3), DXXG.

^d Consensus Mg²⁺ site (kinase 2), $\psi/n/n/n/\psi$, ψ = hydrophobic.

^e Consensus guanine-binding site (G4), (N/T/S)KXD.

^f G3 motif occurring after kinase 2.

^g Published nucleotide specificity.

NBD have been identified for each putative gene. Given this caveat, it appears that all of the genes on chromosome 19 have doublet LRR exons whereas those on chromosome 16 have singlets. DEFKAP and the potential pseudogene 12 have both singlet and doublet exons.

Phylogenetic analysis of the NBD and evolutionary issues

An analysis using protein alignment and tree generation software (Data Analysis in Molecular Biology and Evolution) (25) was performed to examine the potential phylogenetic relationship of the predicted NBD protein sequences (Fig. 2A). Apaf1 and RPM1 (Table I) were included because their NBD regions are similar to those of this family. Except for 11.3, the newly identified NBD sequences are more closely related to one another than Apaf1 (Fig. 2A), suggesting that NBD/WD40 repeat proteins are more distantly related. Interestingly, the NBD of RPM1, an NBD/LRR R protein of *Arabidopsis*, is most closely related to Apaf1. The novel NBD most closely related to RPM1 is 11.3 which has an NBD exon interrupted by an intron. Consistent with divergent evolution, the NBDs of the known and putative proteins with upstream CARD domains are more closely related to each other than to those NBDs with upstream pyrin domains which form their own grouping phylogenetically. Further analysis of NBD/LRR-type plant R proteins and other eukaryotic NBD/LRR proteins will help resolve issues of divergent vs convergent evolution.

The assignment of the CATERPILLER genes to chromosomal positions is shown in Fig. 2B. Most are found in clusters on chromosomes 11, 16, and 19. Three occur at 11p15, three more between 16p12 and 16p13, and nine at 19q13. Proximities of the six sequences on a single contig at 19q13.4 strongly suggest that gene duplication has occurred for these sequences. All except four of these sequences are near the telomere, suggesting that those found singly may have their origins in chromosomal recombination. Among those not at the telomeric end of chromosomes, one (X) is likely a pseudogene. In *Saccharomyces*, fermentation gene alleles are thought to have been generated by the duplication of genes close to the telomeric end and subsequent genomic dispersion by recombination (26). Comparative genomics studies will best address these questions.

The presence of multiple individual exons containing one or two LRRs implies that exon shuffling may occur and that natural selection may favor the maintenance or elimination of a given LRR sequence or pair while simultaneously preserving other aspects of the gene in question (see supplemental data Figs. 3 and 4A). The specificity of plant R proteins is principally dependent on the LRR, and these are targets for diversifying selection (15). In Flax, a 6-aa difference in the LRR of P vs P2 determines Rust R protein specificity (27). The LRRs of RPS2 contain a small stretch important for cooperation with host factors determining *Arabidopsis* resistance to *Pseudomonas*

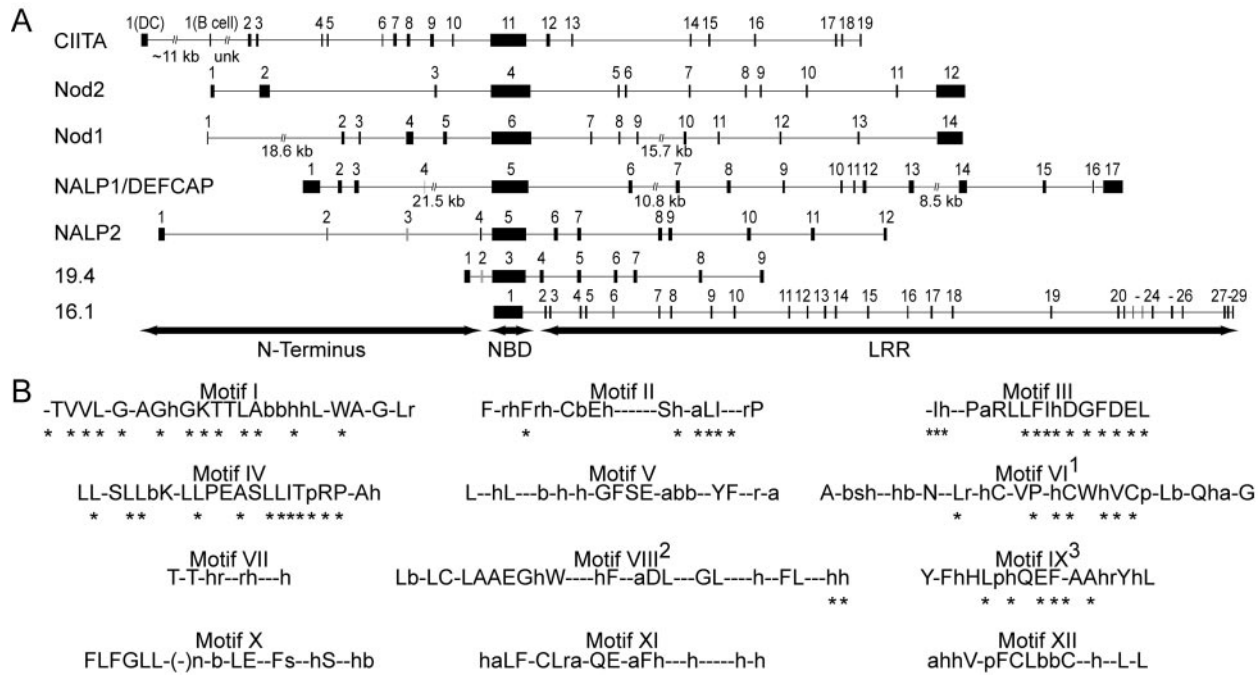


FIGURE 1. Motifs and genomic organization. *A*, Genomic organization for known and some predicted members of the CATERPILLER family are shown to scale. Black boxes represent exons. Unusually large introns are interrupted and their size indicated below in kilobase pairs. Exons with ambiguous positions are shown as gray boxes. The large 3' exons of Nod1 and Nod2 are 3' untranslated sequences. *B*, NBDs were aligned using Clustal with minor manual adjustments. Twelve motifs defining the CATERPILLER NBD are shown. Capital letters indicate residues (single letter code) that have a frequency of >50% or are invariant. Lower case letters indicate residues with frequency <50% but with a predominant characteristic (a = acidic, b = basic, h = hydrophobic, p = serine/threonine, r = aromatic). *, Those residues used to define the NACHT family. Superscripts 1, 2, and 3 indicate NACHT motifs V, VI, and VIII, respectively.

syringae (28). Unequal recombination, gene conversion, and accumulated mutations likely generate novel specificities for the NBD/LRR class of R proteins.

Evidence for expression of the CATERPILLER genes

In light of these data, the NBD/LRR protein family is larger than currently known. Significant information is available on the expression patterns of the known genes and this reflects their biologic role. CIITA has three different isoforms arising from three different promoters. Nod1 has a wide tissue distribution (12), whereas Nod2 and CIAS1 are restricted to monocytes, consistent with inflamma-

tory roles (4, 13). To begin to examine the expression of the other sequences, we have used the NCBI database to search for expressed sequence tags encoding at least part of the sequence (see Table II). UniGene sequence entries exist for *CIAS1*, *Nod1*, *Nod2*, *DEFCAP*, *Nalp2*, and 16.1. Fourteen of the genes are represented in GenBank human expressed sequence tag (est) database. The gene we identify as 19.3 has been previously described as a partial cDNA encoding a 344-aa protein (RNO2) composed of LRRs and is expressed in bone marrow, peripheral blood leukocytes, and nitric oxide-treated HL-60 cells (29). No est entry was found for 11.2, 12, 19.1, 19.2, 19.5, 19.8, or X. We have also conducted a

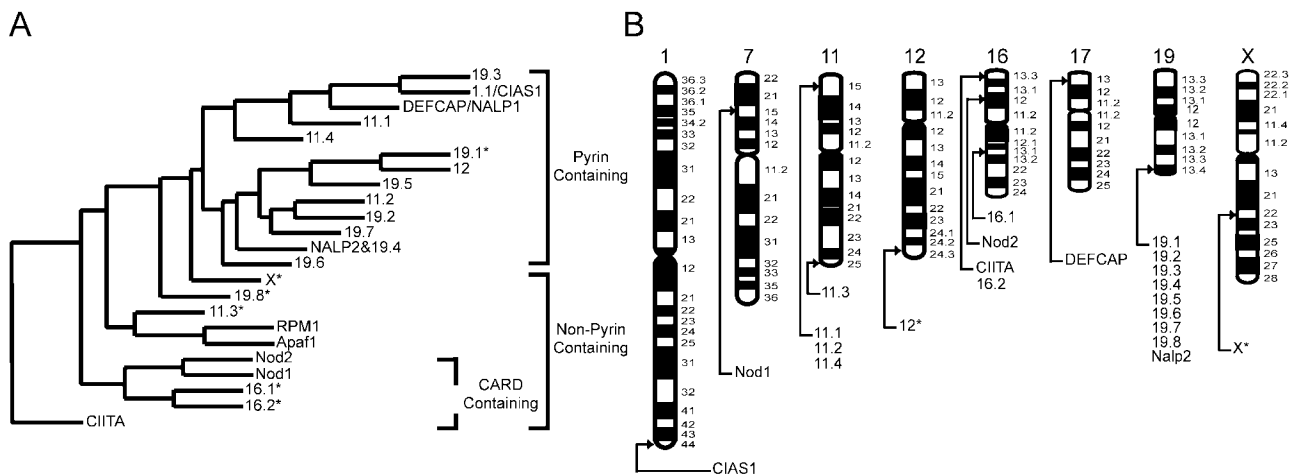


FIGURE 2. Phylogenetic tree for NBDs and chromosomal locations of CATERPILLER genes. *A*, Deduced amino acid sequences from NBD exons were compared with one another using alignment and tree generation software in the Data Analysis in Molecular Biology and Evolution software package (25). *, A predicted gene with unknown N-terminal sequences. *B*, The chromosomal location of each known or predicted sequence. For chromosomal locations with multiple sequences, the name order does not correspond to the ordering on the chromosome.

Table II. Expression pattern of CATERPILLER genes^a

Name	UniGene	GenBank est	Hemopoietic ^b	Somatic ^c
<i>1.1/CIAS1</i>	Hs.159483	+	+	–
<i>Nod1</i>	Hs.19405	+	+ ^d	+ ^d
<i>11.1</i>		+	+	+
<i>11.2</i>			+	–
<i>11.3</i>		+	+	+
<i>11.4</i>		+	+	–
<i>12</i>			NT	NT
<i>CIITA</i>		+	+ ^d	+ ^{d,e}
<i>Nod2</i>	Hs.135201	+	+ ^d	– ^d
<i>16.1</i>	Hs.10888	+	+	+
<i>16.2</i>		+	+	–
<i>DEFKAP</i>	Hs.104305	+	+	+
<i>19.1</i>			–	–
<i>19.2</i>			–	–
<i>19.3</i>		+	+	–
<i>19.5</i>			+	–
<i>19.6</i>		+	+	–
<i>19.7</i>		+	+	–
<i>19.8</i>			+	–
<i>Nalp2/19.4</i>	Hs.6844	+	+	–
<i>X</i>			NT	NT

^a For est searches, stretches of significant identity to translated est sequences were considered a positive match; expression was determined by RT-PCR using cDNA derived from the indicated sources. NT, not tested.

^b Primary human hemopoietic cells or cell lines.

^c HeLa and MCF7 (non-small cell lung carcinoma).

^d From published sources.

^e When induced.

preliminary survey of the expression of these new genes, summarized in Table II, and have detected message for every nonpseudogene except 19.1 and 19.2. Nearly all of the family members are expressed in hemopoietic cells and are likely restricted in that ubiquitous expression was uncommon.

Immunologic significance

Of the known genes, *CIITA*, *CIAS1*, and *Nod2* are clearly linked to immune function. *CIITA* directly controls MHC II gene expression, whereas *CIAS1* in familial cold urticaria and *Nod2* in Crohn's disease are likely regulating inflammatory responses. *DEFKAP* and *Nod1* both promote apoptosis and activate NF- κ B. Activation of NF- κ B is also observed for *Nod2*, and under appropriate conditions for *CIAS1*. These functions are reminiscent of plant R proteins that promote plant responses similar to innate immune functions (15).

Innate immune responses mediated by Toll in response to fungal pathogens in *Drosophila* highlight the importance of receptors recognizing specific pathogen-associated molecular patterns (30). LRR-containing proteins in plants and animals serve a similar function; this contention is supported by our threading result with selected LRRs suggesting that LRR structural features are conserved in the NBD/LRR family (supplemental data Fig. 4). Toll-like receptors have extracellular LRRs mediating recognition of a variety of microbial derivatives (31, 32). The LRRs of plant R proteins likewise recognize avirulence proteins from plant pathogens and provide specificity (33). Recent studies of *Nod1* and *Nod2* demonstrate that both require their LRRs for responses to various bacterial LPS (34). The LRRs of *CIITA* (although not known to interact with any pathogen-specific molecule) are functionally necessary, are involved in self-association and interaction with an endogenous protein, and regulate nuclear import (10). Thus, these LRRs likely serve as versatile recognition domains with specificity for self-interaction, protein/lipid/sugar recognition, or both, which seems probable. Deletion of the LRRs from

Nod1/2, *DEFKAP*, and *CIAS1* enhances their activities, suggesting that these LRRs are important sites of regulation.

As further evidence of the immunologic relatedness of this family of gene, we have recently studied the 19.3 gene product (named Monarch-1) and found it to be predominantly expressed by cells of the myeloid-monocytic-dendritic lineage. In addition, 19.3 expression is dramatically altered by bacterial products, and influences a number of immunologically relevant events.⁶

Related issues

The number of mammalian NBD/LRR sequences we were able to identify is significantly smaller than that occurring in some plants (35). The mammalian family may be larger than we describe as NAIP and Ipaf (*CARD12*), despite having NBDs and LRRs, were not detected using our parameters (except when using 16.2), likely due to the absence of some of the CATERPILLER motifs in their NBDs. Limited BLAST searches of translated nucleotide sequences from *Drosophila* and *Caenorhabditis elegans*, genomic databases failed to identify any NBD/LRR genes. A similar search of the *Danio rerio* (zebrafish) database did yield likely NBD/LRR sequences, and the mouse genome has at least as many genes in this family as did humans (J. A. Harton, unpublished observation). The preponderance of NBD/LRR proteins in plants is due to reliance on individual effector molecules for recognizing pathogen-specific products. Higher order eukaryotes have developed a highly complex adaptive immune system driving a staggering array of protein-specific immune responses with a limited number of genes.

N-terminal variation in the known and predicted genes suggests a subdivision of CATERPILLER proteins: group I, CARD-containing (e.g., *Nod1*); group II, pyrin-containing (e.g., *DEFKAP*); group III, *trans* activation domain (e.g., *CIITA*); and unknown (e.g., 16.1) (see Table I). However, these grouping may be oversimplified. For example, multiple cell type-specific forms of *CIITA* are known. The dendritic cell form has a CARD-like N terminus followed by the activation domain, although no caspase recruitment activity has been described (36). It is of interest that *Nod2* and cryopyrin are also expressed as multiple transcripts (4, 13). Whether these different transcripts code for proteins of somewhat different function is clearly of interest. Additionally, self-association has also been demonstrated for *CIITA* and *Nod1*, whereas heterodimerization of *CIAS1* with apoptotic protein ASC may involve the pyrin domain of *CIAS1* (5, 10, 12). Self- and heteroassociation might amplify and generate diversity necessary to mediate appropriate responses.

Genes coding proteins structurally related to *CIITA*, *Nod2*, and others in having an NBD, multiple C-terminal LRRs, and few different N-terminal domains abound in the human genome. The sequences and genomic organization of these genes suggest a high degree of relatedness, a common origin, and a potential link to the basic immune response genes of plants. Studies on *CIITA*, *CIAS1*, *DEFKAP*, *Nod1*, and *Nod2* reveal some interesting parallels with the plant proteins and strongly suggest that this family of proteins will likely influence mammalian immune responses.

Note added in proof. During the review of this manuscript a report describing the initial characterization of Pypaf7, which we refer to as 19.3/Monarch-1, was published. 2002, *J. Biol. Chem.* 277:29874.

⁶ K. L. Williams, D. J. Taxman, M. W. Linhoff, and J. P.-Y. Ting. Monarch-1: a Pyrin/NBD/LRR protein that broadly controls classical and non-classical class I MHC genes. Submitted for publication.

References

- Steimle, V., L. A. Otten, M. Zufferey, and B. Mach. 1993. Complementation cloning of an MHC class II *trans* activator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome). *Cell* 75:135.
- Ogura, Y., D. K. Bonen, N. Inohara, D. L. Nicolae, F. F. Chen, R. Ramos, H. Britton, T. Moran, R. Karaliuskas, R. H. Duerr, et al. 2001. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 411:603.
- Hugot, J. P., M. Chamaillard, H. Zouali, S. Lesage, J. P. Cezard, J. Belaiche, S. Almer, C. Tysk, C. A. O'Morain, M. Gassull, et al. 2001. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411:599.
- Hoffman, H. M., J. L. Mueller, D. H. Broide, A. A. Wanderer, and R. D. Kolodner. 2001. Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle-Wells syndrome. *Nat. Genet.* 29:301.
- Manji, G. A., L. Wang, B. J. Geddes, M. Brown, S. Merriam, A. Al-Garawi, S. Mak, J. M. Lora, M. Briskin, M. Jurman, et al. 2002. PYPAF1: a PYRIN-containing Apaf1-like protein that assembles with ASC and regulates activation of NF- κ B. *J. Biol. Chem.* 277:11570.
- Samuels, J., I. Aksentjevich, Y. Torosyan, M. Centola, Z. Deng, R. Sood, and D. L. Kastner. 1998. Familial Mediterranean fever at the millennium: clinical spectrum, ancient mutations, and a survey of 100 American referrals to the National Institutes of Health. *Medicine* 77:268.
- Bertin, J., and P. S. DiStefano. 2000. The PYRIN domain: a novel motif found in apoptosis and inflammation proteins. *Cell Death Differ.* 7:1273.
- Harton, J. A., and J. P. Ting. 2000. Class II *trans* activator: mastering the art of major histocompatibility complex expression. *Mol. Cell Biol.* 20:6185.
- Reith, W., and B. Mach. 2001. The bare lymphocyte syndrome and the regulation of MHC expression. *Annu. Rev. Immunol.* 19:331.
- Ting, J. P., and J. Trowsdale. 2002. Genetic control of MHC class II expression. *Cell* 109(Suppl.):S21.
- Bertin, J., W. J. Nir, C. M. Fischer, O. V. Tayber, P. R. Errada, J. R. Grant, J. J. Keilty, M. L. Gosselin, K. E. Robison, G. H. Wong, M. A. Glucksmann, and P. S. DiStefano. 1999. Human CARD4 protein is a novel CED-4/Apaf-1 cell death family member that activates NF- κ B. *J. Biol. Chem.* 274:12955.
- Inohara, N., T. Koseki, L. del Peso, Y. Hu, C. Yee, S. Chen, R. Carrio, J. Merino, D. Liu, J. Ni, and G. Nunez. 1999. Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor- κ B. *J. Biol. Chem.* 274:14560.
- Ogura, Y., N. Inohara, A. Benito, F. F. Chen, S. Yamaoka, and G. Nunez. 2001. Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF- κ B. *J. Biol. Chem.* 276:4812.
- Miceli-Richard, C., S. Lesage, M. Rybojad, A. M. Prieur, S. Manouvrier-Hanu, R. Hafner, M. Chamaillard, H. Zouali, G. Thomas, and J. P. Hugot. 2001. CARD15 mutations in Blau syndrome. *Nat. Genet.* 29:19.
- Dangl, J. L., and J. D. Jones. 2001. Plant pathogens and integrated defence responses to infection. *Nature* 411:826.
- Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, et al. 2001. The sequence of the human genome. *Science* 291:1304.
- Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860.
- Pearson, W. R. 2000. Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol. Biol.* 132:185.
- Park, J., K. Karplus, C. Barrett, R. Hughey, D. Haussler, T. Hubbard, and C. Chothia. 1998. Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. *J. Mol. Biol.* 284:1201.
- Karplus, K., C. Barrett, and R. Hughey. 1998. Hidden Markov models for detecting remote protein homologies. *Bioinformatics* 14:846.
- Hlaing, T., R. F. Guo, K. A. Dilley, J. M. Loussia, T. A. Morrish, M. M. Shi, C. Vincenz, and P. A. Ward. 2001. Molecular cloning and characterization of DEFCAP-L and -S, two isoforms of a novel member of the mammalian Ced-4 family of apoptosis proteins. *J. Biol. Chem.* 276:9230.
- Koonin, E. V., and L. Aravind. 2000. The NACHT family: a new group of predicted NTPases implicated in apoptosis and MHC transcription activation. *Trends Biochem. Sci.* 25:223.
- Traut, T. W. 1994. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. *Eur. J. Biochem.* 222:9.
- Heery, D. M., E. Kalkhoven, S. Hoare, and M. G. Parker. 1997. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387:733.
- Xia, X., and Z. Xie. 2001. DAMBE: software package for data analysis in molecular biology and evolution. *J. Hered.* 92:371.
- Charron, M. J., E. Read, S. R. Haut, and C. A. Michels. 1989. Molecular evolution of the telomere-associated MAL loci of *Saccharomyces*. *Genetics* 122:307.
- Dodds, P., G. Lawrence, and J. Ellis. 2001. Six amino acid changes confined to the leucine-rich repeat β -strand/ β -turn motif determine the difference between the P and P2 rust resistance specificities in flax. *Plant Cell* 13:163.
- Banerjee, D., X. Zhang, and A. F. Bent. 2001. The leucine-rich repeat domain can determine effective interaction between RPS2 and other host factors in *Arabidopsis* RPS2-mediated disease resistance. *Genetics* 158:439.
- Shami, P. J., N. Kanai, L. Y. Wang, T. M. Vreeke, and C. H. Parker. 2001. Identification and characterization of a novel gene that is upregulated in leukaemia cells by nitric oxide. *Br. J. Haematol.* 112:138.
- Medzhitov, R. 2001. Toll-like receptors and innate immunity. *Nat. Rev. Immunol.* 1:135.
- Poltorak, A., P. Ricciardi-Castagnoli, S. Citterio, and B. Beutler. 2000. Physical contact between lipopolysaccharide and Toll-like receptor 4 revealed by genetic complementation. *Proc. Natl. Acad. Sci. USA* 97:2163.
- Bauer, S., C. J. Kirschning, H. Hacker, V. Redecke, S. Hausmann, S. Akira, H. Wagner, and G. B. Lipford. 2001. Human TLR9 confers responsiveness to bacterial DNA via species-specific CpG motif recognition. *Proc. Natl. Acad. Sci. USA* 98:9237.
- Van Der Hoorn, R. A., R. Roth, and P. J. De Wit. 2001. Identification of distinct specificity determinants in resistance protein cf-4 allows construction of a cf-9 mutant that confers recognition of avirulence protein avr4. *Plant Cell* 13:273.
- Inohara, N., Y. Ogura, F. F. Chen, A. Muto, and G. Nunez. 2001. Human nod1 confers responsiveness to bacterial lipopolysaccharides. *J. Biol. Chem.* 276:2551.
- Pan, Q., J. Wendel, and R. Fluhr. 2000. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J. Mol. Evol.* 50:203.
- Nickerson, K., T. J. Sisk, N. Inohara, C. S. Yee, J. Kennell, M. C. Cho, P. J. Yannie, 2nd, G. Nunez, and C. H. Chang. 2001. Dendritic cell-specific MHC class II *trans* activator contains a caspase recruitment domain that confers potent *trans* activation activity. *J. Biol. Chem.* 276:19089.
- Wang, L., G. A. Manji, J. M. Grenier, A. Al-Garawi, S. Merriam, J. M. Lora, B. J. Geddes, M. Briskin, P. S. DiStefano, and J. Bertin. 2002. PYPAF7, a novel PYRIN-containing Apaf1-like protein that regulates activation of NF- κ B and caspase-1-dependent cytokine processing. *J. Biol. Chem.* 277:29874.