J Mol Evol (2014) 78:194–201
DOI 10.1007/s00239-014-9614-3

REVIEW

The Evolution of Reproduction-Related *NLRP* Genes

Edgar A. Duéñez-Guzmán · David Haig

Received: 17 January 2013 / Accepted: 19 February 2014 / Published online: 11 March 2014

Abstract NLRP proteins are important components of inflammasomes with a major role in innate immunity. A subset of *NLRP* genes, with unknown functions, are expressed in oocytes and early embryos. Mutations of *Nlrp5* in mice are associated with maternal-effect embryonic lethality and mutations of *NLRP7* in women are associated with conception of biparental complete hydatidiform moles (biCHMs), suggesting perturbed processes of genomic imprinting. Recessive mutations on *NLRP2/7* in humans are associated with reproductive disorders and appear to be induced by a demethylation of the maternal pronucleus. In this study, we find that radiation of *NLRP* genes occurred before the common ancestor of Afrotheria and Boreoeutheria, with the clade of oocyte-expressed genes originating before the divergence of marsupial and eutherian mammals. There have been multiple independent duplications of *NLRP2* genes one of which produced the *NLRP7* gene associated with biCHMs.

Keywords Mammals · Phylogeny · Genetics · Molecular evolution · Reproductive disorders

Supported by an Innovative Collaboration Award by the Howard Hughes Medical Institute to Catherine Dulac.

Electronic supplementary material The online version of this article (doi:10.1007/s00239-014-9614-3) contains supplementary material, which is available to authorized users.

E. A. Duéñez-Guzmán (✉)
Laboratory of Socioecology and Social Evolution, KU Leuven, Naamsestraat 59, Leuven 3000, Belgium
e-mail: eaduenez@gmail.com

D. Haig
Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

Introduction

NLRP proteins (Nod-like receptors with a pyrin domain) have attracted recent attention because of their role in innate immunity and inflammation (Kufer and Sansonetti 2011; Strowig et al. 2012). A subset of *NLRP* genes are expressed in mammalian oocytes (Hamatami et al. 2004; Ponsuksili et al. 2006; Zhang et al. 2008) and maternal deficiency of some of these "reproduction-related" *NLRPs* (*rNLRPs*) have been shown to cause embryonic lethality in mice (Hamatami et al. 2004; Tong et al. 2000; Peng et al. 2012) and perturbations of genomic imprinting in human oocytes (Murdoch et al. 2006; Kou et al. 2008; Parry et al. 2011).

Tian et al. (2009) undertook a phylogenetic analysis of the *NLRP* genes from humans, chimpanzees, rats, mice, cattle, and dogs. These species had similar sets of *NLRP* genes, with the *rNLRP* genes forming a monophyletic group. Therefore, the major radiation of *NLRP* genes was already present in the most recent common ancestor of Boreoeutheria. These authors observed that *rNLRPs* were more evolutionarily labile, in both copy number and sequence, than other *NLRPs*.

Evidence that *rNLRPs* might play a role in the establishment or maintenance of genomic imprinting motivated us to re-investigate the evolutionary history and possible functions of *rNLRPs*. We extend the analysis of Tian et al. (2009) using *NLRP* sequences from an afrotherian (African elephant *Loxodonta africana*), marsupial (gray short-tailed opossum *Monodelphis domestica*), and monotreme (platypus *Ornithorhynchus anatinus*). These sequences allow inferences about the *NLRP* repertoire of the last common ancestor of eutherian mammals, the last common ancestor of marsupials and eutherians, and the last common ancestor of all extant mammals. Moreover, advances in knowledge

of the effects of *rNLRPs* allows us to speculate about the role of these genes in genomic imprinting and reproductive disorders of mice and humans.

Results

Mammalian *NLRP* Repertoires

NLRP genes occur at eight locations in the human genome: 11p15.5 (*NLRP6*); 11p15.4 (*NLRP10*); 11p15.4 (*NLRP14*, 1 Mb from *NLRP10*); 19q13.42 (*NLRP12*); 19q13.42 (*NLRP2*, *NLRP7*); 19q13.43 (*NLRP4*, *5, 8, 9, 11, 13*); 1q44 (*NLRP3*); and 17p13.2 (*NLRP1*). *NLRP* genes were found in the elephant genome at seven of these locations (based on conserved flanking markers). The exception was elephant *NLRP6* which occurs on an unassembled fragment without flanking markers. The elephant genome contains an additional *NLRP* gene between *MID2* and *VSIG1* that we have provisionally named *NLRPX* (currently annotated as *NLRP12-like* in elephant, macaque, and marmoset). An *NLRP* pseudogene is located between these markers on the human X chromosome (*NLRP3P*). *NLRPX*-related sequences are sister to a clade containing *NLRP10* genes in our phylogenetic tree.

Four *NLRP* genes are currently annotated in the opossum genome. *NLRP10* is located on opossum chromosome 4 between *GVIN1* (corresponding to the *GVINP1* pseudogene at human 11p15.4) and *CSMD2* (human ortholog at 1p34). Two other opossum *NLRPs* also map to chromosome 4: one appears orthologous to *NLRP12* whereas the other (provisionally named *NLRPa*, currently annotated as *NLRP12*) appears in our phylogeny as sister to the *rNLRP* clade albeit with weak support. The fourth gene (provisionally named *NLRP5-like*, currently annotated as *NLRP14-like*) occurs next to *EPN1* on an unassembled fragment. Human *EPN1* neighbors the cluster of six *rNLRPs* at 19q13.43.

Five *NLRP* genes were found in the platypus genome. Four appear to be orthologs of human *NLRP3* (currently annotated as *NLRP12-like*), *NLRP6*, *NLRP10* and *NLRP12*. The fifth (provisionally named ψ *NLRP10*, currently annotated as *NLRP3-like*) occurs on the same fragment as *NLRP10* and is possibly a monotreme-specific duplicate of that gene. None of the platypus genes group with the *rNLRPs*.

Figure 2 summarizes the *NLRP* repertoires of the mammalian species used in this study. We represent the data as a table of species against *NLRP* genes with phylogenies depicted as a reference for both. The cells of the table are shaded according to the presence or absence of a given *NLRP* gene in a particular species.

Phylogenetic Tree Root

Our tree places *NLRP6* as sister to the other eutherian *NLRP* genes using chicken "*NLRP3*" to root the tree, similar to the analysis of Tian et al. (2009) who used the same root. A sister relation between *NLRP6* and other mammalian *NLRPs* has been found repeatedly in phylogenetic analyses. These include analyses of NACHT (Hughes 2006; Laing et al. 2008) and LRR domains (Ng et al. 2011) that rooted the *NLRP* clade using non-NLRP proteins. A slightly different root was proposed in a phylogenetic analysis of PYD domains that placed the root between a clade containing *NLRP6* and *NLRP10* and the other *NLRP* genes (Kersse et al. 2011).

Diversification of *NLRP2* and *NLRP7*

Figure 3 presents a Bayesian phylogenetic tree for current genes and pseudogenes labeled *NLRP2* or *NLRP7* in the databases as of May, 2012. The primate-specific duplication noted by Tian et al. (2009) is confirmed but independent duplications have also occurred in pigs, cattle, horses, and elephants. We suggest, as an interim measure, that use of *NLRP7* be restricted to genes of that name in the primate clade.

Syntenic Analyses

Online Resource 3 shows the probable orthologies of all *NLRP* genes from humans against the other species of mammals, as well as chicken. These figures use the latest genome assemblies (as of May, 2012), which is why they differ slightly from the data-set used for the phylogenetic tree in Fig. 1.

Discussion

Evolutionary History of *NLRP* Genes

The similarity of the *NLRP* repertoires of humans and elephants shows that the major radiation of *NLRP* genes, including *rNLRP* genes, had already occurred in the common ancestor of Afrotheria and Boreoeutheria.

Distal chromosome 4 of opossum contains genes with orthologs on human chromosomes 11p and 19q. Thus, the data are compatible with a scenario in which the opossum genome maintains ancestral linkage of *NLRP* genes that have been dispersed onto the equivalents of human 11p and 19q in eutherian mammals. Unfortunately, the evolutionary inference is weak because opossum, platypus, and chicken genomes are only partially assembled for these regions.

The existence of *NLRPa* and *NLRP5-like* in opossum suggests that *rNLRP* genes evolved before the last common ancestor of Metatherian and Eutherian mammals. Our phylogenetic analysis (Fig. 1) confirms that opossum *NLRP5-like* belongs to the *rNLRP* clade. Whether it is expressed in opossum oocytes is unknown. Conservation of synteny suggests *NLRP5-like* is probably located on opossum chromosome 4 with the other *NLRP* genes. Four out of four opossum *NLRP* genes plausibly map to a region of chromosome 4 with orthologs on human chromosomes 11p and 19q (where 12 out of 14 human *NLRP* genes are located).

Platypus *NLRP3* is located adjacent to *RPS5* and *SLC27A5*, genes whose orthologs are located near the telomere of human 19q (close to the major cluster of *rNLRPs*), whereas human *NLRP3* is located near the telomere of chromosome 1q as part of a small, recent addition to an otherwise ancient linkage group (Haig 2005). Therefore, *NLRP3* may have been linked to *rNLRP* genes in the most recent common ancestor of monotremes and therian mammals. This leaves *NLRP1* at human 17p13.2 as the only eutherian *NLRP* whose ortholog cannot be provisionally assigned to this ancestral linkage group.

However, a scenario that is consistent with our data is that *rNLRPs* were absent in the common ancestor of Monotremes and Metatherians. Then, a single *rNLRP* gene evolved in a common ancestor of Metatherians and Eutherians. This ancestral *rNLRP* is orthologous to possum *NLRPa* and it later diversified to give rise to the clusters *NLRP2/7* and *NLRP9/11/4/13/8/5*, as well as *NLRP14*. *NLRP2/7* may be still present in possum, but the current assembly does not contain them or their flanking markers (see Online Resource 3).

Lineage-Specific Duplications and Deletions

Our analysis provides evidence of lineage-specific duplications and deletions of *NLRP* genes in eutherian mammals. *NLRP11* was previously proposed to be restricted to primates and *NLRP4* to be restricted to Euarchontoglires (Tian et al. 2009) but we find orthologs of *NLRP11* in pig and of *NLRP4* in pig and horse (see Fig. 2 and Online Resource 4). Pig *NLRP11* represents an interesting example, because were it absent, parsimony would indicate a single loss of *NLRP11* in Laurasiatheria. Its presence indicates independent losses in mouse, dog, horse, and cattle.

The data also suggests an independent loss of *NLRP8* and *NLRP13* in marmoset, mouse, and elephant. In particular, the marmoset genome has *NLRP11* and *NLRP4* alone in an unplaced fragment, while conservation of synteny suggests these genes should be placed between *NLRP9* and *NLRP5*. A future assembly, however, might reconstruct this area and place the fragment inside it, potentially also reconstructing *NLRP8* and *NLRP13*.

A second phylogenetic reconstruction without any of the reconstructed pseudogenes is available in Online Resource 5. There are three minor differences between the phylogeny in Fig. 1 (which includes pseudogenes) and Online Resource 5. First, the reversal of the divergence of the *NLRP8-13* and *NLRP5-14* gene pairs. Second, the *NLRP10-X* gene pair changing from being a monophyletic group with *NLRP3* and *NLRP12*, to being its outgroup. Third and finally, possum *NLRP5-like* clustering with *NLRP8* instead of *NLRP5*.

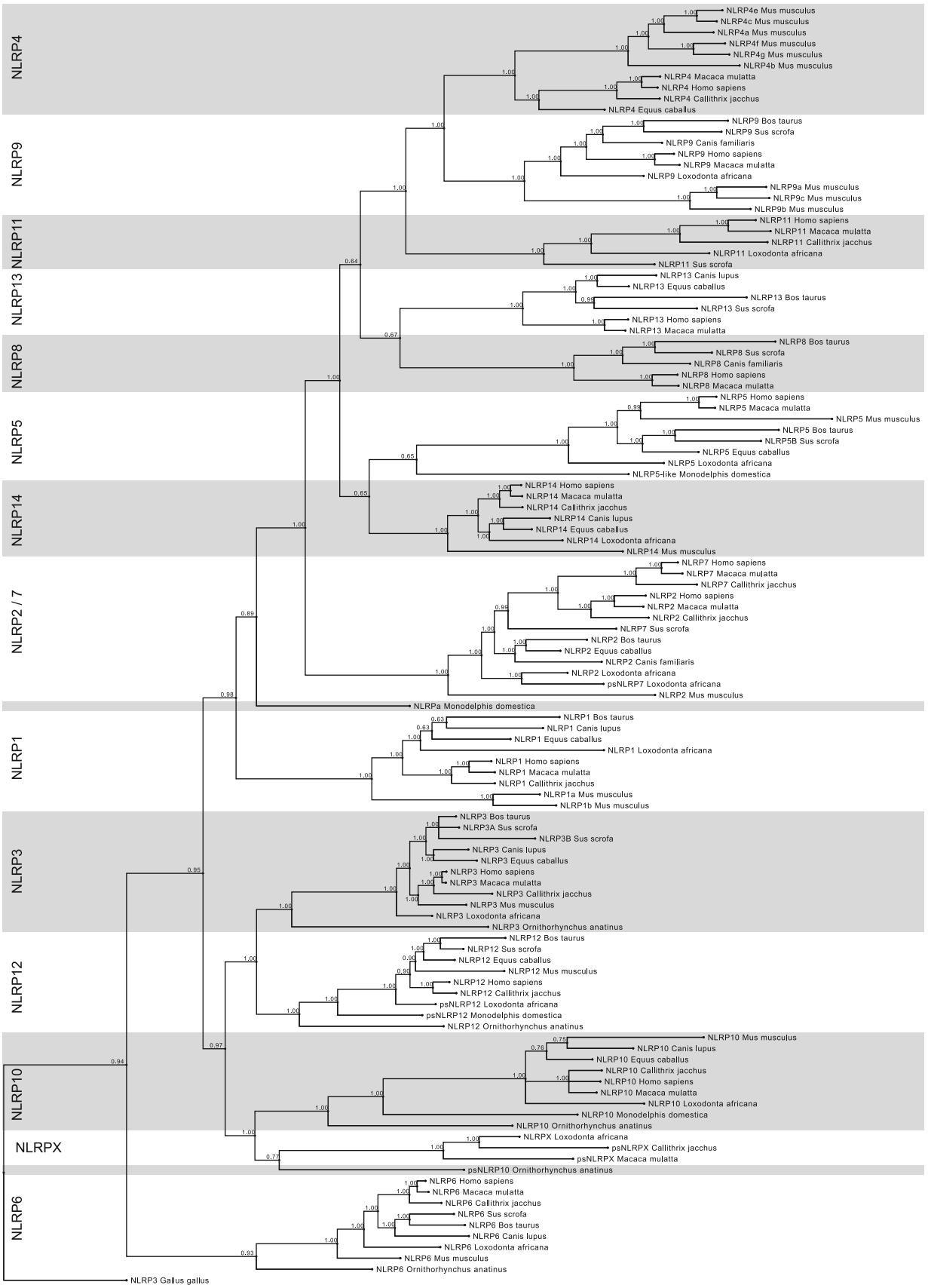
Maternal-Effect Lethality in Mice

Inactivation of *Nlrp5* (also known as *Mater*) in mouse mothers causes arrested development of embryos at the two-cell stage whether or not an embryo inherits a functional copy of *Nlrp5* from its father (Tong et al. 2000). Nlrp5 protein is associated with the cytoplasmic lattice of mouse oocytes and appears to be essential for the formation and/or stability of the lattice (Kim et al. 2010). Mitochondria of *Nlrp5*-deficient oocytes are scattered throughout the cytoplasm, rather than concentrated in the subcortical layer (Fernandes et al. 2012). Knockdown of *Nlrp2* or *Nlrp14* mRNA in mouse oocytes similarly causes arrested development of embryos during early cleavage (Hamatami et al. 2004; Peng et al. 2012).

Nlrp5 protein associates with Ecat1/Filia protein in mouse oocytes and early embryos (Ohsugi et al. 2008; Zheng and Dean 2009). Mutations in *Filia* cause maternal-effect embryonic lethality with apparent defects in the assembly of mitotic spindles (Zheng and Dean 2009). *Filia* belongs to a family of genes expressed in oocytes and early embryos (Pierre et al. 2007) that includes *C6orf221* (see below).

Perturbations of Genomic Imprinting in Humans

During normal embryonic development, the maternal allele is methylated and paternal allele unmethylated at most imprinting control regions (ICRs). The *H19* ICR is an exception to this generalization, with an unmethylated maternal allele and methylated paternal allele (Reik and Walter 2001; Schulz et al. 2010). DNA in the sperm pronucleus of recently fertilized mammalian oocytes undergoes conversion of most 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) whereas 5mC is protected from this activity in the egg pronucleus (Iqbal et al. 2011; Wossidlo et al. 2011). Paternally derived chromosomes then progressively lose 5hmC during early embryonic development (Inoue and Zhang 2011). Taken together these results suggest that most ICRs are methylated in the


NLRP4
NLRP4e Mus musculus
NLRP4c Mus musculus
NLRP4a Mus musculus
NLRP4f Mus musculus
NLRP4g Mus musculus
NLRP4b Mus musculus
NLRP4 Macaca mulatta
NLRP4 Homo sapiens
NLRP4 Callithrix jacchus
NLRP4 Equus caballus
NLRP9
NLRP9 Bos taurus
NLRP9 Sus scrofa
NLRP9 Canis familiaris
NLRP9 Homo sapiens
NLRP9 Macaca mulatta
NLRP9 Loxodonta africana
NLRP9a Mus musculus
NLRP9c Mus musculus
NLRP9b Mus musculus
NLRP11
NLRP11 Homo sapiens
NLRP11 Macaca mulatta
NLRP11 Callithrix jacchus
NLRP11 Loxodonta africana
NLRP11 Sus scrofa
NLRP13
NLRP13 Canis lupus
NLRP13 Equus caballus
NLRP13 Bos taurus
NLRP13 Sus scrofa
NLRP13 Homo sapiens
NLRP13 Macaca mulatta
NLRP8
NLRP8 Bos taurus
NLRP8 Sus scrofa
NLRP8 Canis familiaris
NLRP8 Homo sapiens
NLRP8 Macaca mulatta
NLRP5
NLRP5 Homo sapiens
NLRP5 Macaca mulatta
NLRP5 Mus musculus
NLRP5 Bos taurus
NLRP5B Sus scrofa
NLRP5 Equus caballus
NLRP5 Loxodonta africana
NLRP5-like Monodelphis domestica
NLRP14
NLRP14 Homo sapiens
NLRP14 Macaca mulatta
NLRP14 Callithrix jacchus
NLRP14 Canis lupus
NLRP14 Equus caballus
NLRP14 Loxodonta africana
NLRP14 Mus musculus
NLRP2 / 7
NLRP7 Homo sapiens
NLRP7 Macaca mulatta
NLRP7 Callithrix jacchus
NLRP2 Homo sapiens
NLRP2 Macaca mulatta
NLRP2 Callithrix jacchus
NLRP7 Sus scrofa
NLRP2 Bos taurus
NLRP2 Equus caballus
NLRP2 Canis familiaris
NLRP2 Loxodonta africana
psNLRP7 Loxodonta africana
NLRP2 Mus musculus
NLRPa Monodelphis domestica
NLRP1
NLRP1 Bos taurus
NLRP1 Canis lupus
NLRP1 Equus caballus
NLRP1 Loxodonta africana
NLRP1 Homo sapiens
NLRP1 Macaca mulatta
NLRP1 Callithrix jacchus
NLRP1a Mus musculus
NLRP1b Mus musculus
NLRP3
NLRP3 Bos taurus
NLRP3A Sus scrofa
NLRP3B Sus scrofa
NLRP3 Canis lupus
NLRP3 Equus caballus
NLRP3 Homo sapiens
NLRP3 Macaca mulatta
NLRP3 Callithrix jacchus
NLRP3 Mus musculus
NLRP3 Loxodonta africana
NLRP3 Ornithorhynchus anatinus
NLRP12
NLRP12 Bos taurus
NLRP12 Sus scrofa
NLRP12 Equus caballus
NLRP12 Mus musculus
NLRP12 Homo sapiens
NLRP12 Callithrix jacchus
psNLRP12 Loxodonta africana
psNLRP12 Monodelphis domestica
NLRP12 Ornithorhynchus anatinus
NLRP10
NLRP10 Mus musculus
NLRP10 Canis lupus
NLRP10 Equus caballus
NLRP10 Callithrix jacchus
NLRP10 Homo sapiens
NLRP10 Macaca mulatta
NLRP10 Loxodonta africana
NLRP10 Monodelphis domestica
NLRP10 Ornithorhynchus anatinus
NLRPX
NLRPX Loxodonta africana
psNLRPX Callithrix jacchus
psNLRPX Macaca mulatta
psNLRP10 Ornithorhynchus anatinus
NLRP6
NLRP6 Homo sapiens
NLRP6 Macaca mulatta
NLRP6 Callithrix jacchus
NLRP6 Sus scrofa
NLRP6 Bos taurus
NLRP6 Canis lupus
NLRP6 Loxodonta africana
NLRP6 Mus musculus
NLRP6 Ornithorhynchus anatinus
NLRP3 Gallus gallus

◀**Fig. 1** Consensus phylogenetic tree of the *NLRP* gene family. The phylogenetic tree was inferred using MrBayes for 10,000,000 time steps. We used 115 amino-acid sequences of *NLRP* genes and "pseudogenes" from 11 mammalian species (human, macaque, marmoset, mouse, pig, cattle, horse, dog, elephant, possum, and platypus) as well as one sequence from chicken as an outgroup

maternal germ line with paternal ICRs unmethylated because of demethylation of the sperm pronucleus after fertilization, with the methylated paternal *H19* ICR somehow protected from this process.

Women homozygous for mutations of *NLRP7* exhibit maternal-effect embryonic lethality in the form of the repeated conception of biparental complete hydatidiform moles (biCHMs) (Murdoch et al. 2006; Kou et al. 2008; Parry et al. 2011). These "embryos" exhibit biparental non-methylation of most imprinting control regions (ICRs), with the notable exception of the *H19* ICR (Kou et al. 2008; El-Maarri et al. 2003). Women with mutations of both alleles of *C6orf221* also produce biCHMs with loss of differential methylation at most ICRs except *H19* (Parry et al. 2011; Judson et al. 2002). Thus, loss-of-function of *NLRP7* and *C6orf221* cause similar perturbations of imprinting. *C6orf221* belongs to the *Ecat*/*Filia* gene family whose members are expressed in oocytes and early embryos and are known to interact with NLRP proteins (Zheng and Dean 2009).

The methylation pattern of biCHMs would be explained if maternal ICRs become demethylated in embryos because of a failure to protect the maternal pronucleus from conversion of 5mC to 5hmC. One possibility is that maternal deficiency of *NLRP7* or *C6orf221* results in a disruption of

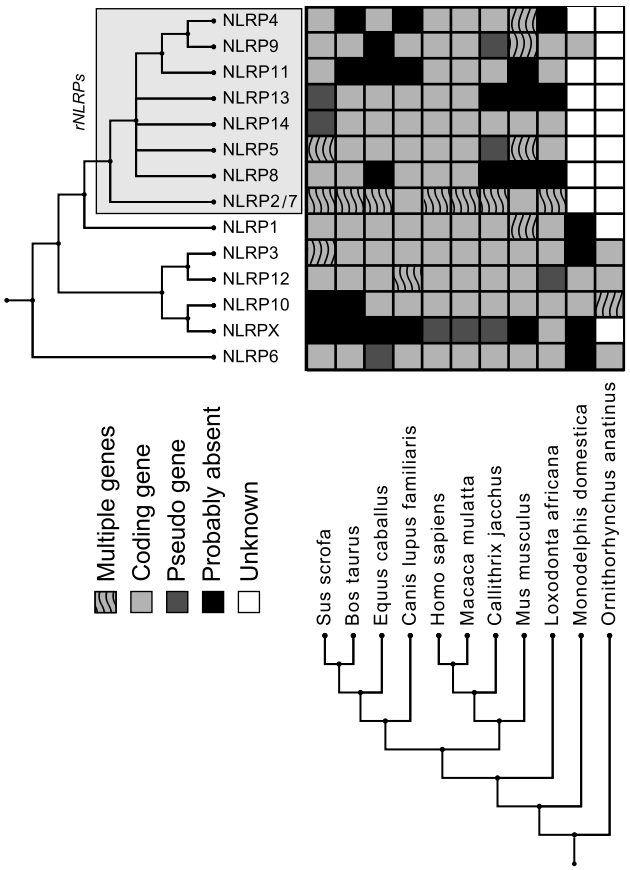


Fig. 2 Presence and absence of *NLRP* genes in mammalian species. This table summarizes the known information about the distribution of *NLRP* genes across the mammalian species included in our analyses. The cells of the table are shaded according to the presence or absence of a given *NLRP* gene in a particular species. *Light gray* is used to indicate the presence of a single *NLRP* gene identified in the databases as "protein coding." *Light gray with a wave pattern* is used if multiple copies of the gene are present, either identified as "protein coding" or "pseudogene." *Dark gray* indicates that a single gene is present and identified as "pseudogene." *Black* is used to denote the probable absence of the gene; in general, we consider a gene absent if it is not currently annotated in the species, absent from *BLAST* searches, and syntenic examination on the other mammalian species. We use *white* to denote that there is not sufficient evidence based on flanking markers to consider a gene absent

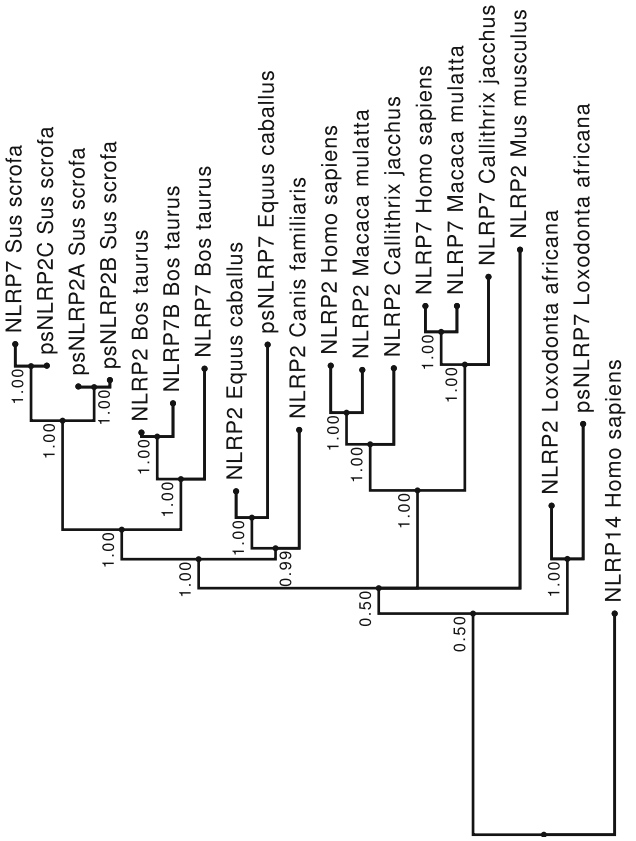


Fig. 3 Consensus phylogenetic tree of *NLRP2* and *NLRP7*. The phylogenetic tree was inferred using MrBayes for 1,000,000 time steps. We used 19 amino-acid sequences of all available *NLRP* genes and "pseudogenes" denoted as *NLRP2* or *NLRP7* in the mammalian species used in our other analyses, as well as human *NLRP14* as an outgroup. Potentially independent duplications of *NLRP2/7* can be seen in Laurasiatheria, primates, and elephant

spatial relations in the fertilized oocyte causes the egg pronucleus to be subject to the same epigenetic modifications as the sperm pronucleus.

Human *NLRP7* is the product of a recent duplication of *NLRP2*. Maternal homozygosity for a frame-shift mutation in *NLRP2* has been associated with loss of "maternal" methylation at 11p15.5 and a clinical diagnosis of Beckwith–Wiedemann syndrome (BWS) in offspring (Meyer et al. 2009). Finally, two individuals with phenotypic features of BWS have been reported with independent breakpoints in *ZNF215* (*ZKSCAN11*), a gene immediately adjacent to *NLRP14* (Alders et al. 2000).

Conclusions

Our analysis shows that all the major *NLRP* genes are present in eutherian mammals and that *rNLRP*-related genes are present in marsupials. The *rNLRPs* have a complex history of independent duplications in several eutherian lineages. While strong inferences about the origin of *rNLRP* genes remain elusive, the following scenario is consistent with the data. The appearance and diversification of *rNLRP* genes in the common ancestor of Metatherians (where opossum chromosome 4 maintains ancestral linkage of *NLRP* genes) and Eutherians, followed by breakage of linkage and consequent dispersal to their present locations in eutherian mammals.

Mutations in *NLRP2* and *NLRP7* in humans have been associated with disorders of genomic imprinting, Beckwith–Wiedemann syndrome and biparental complete hydatidiform moles. A plausible explanation is that deficiencies of *NLRP2* and *NLRP7* result in the maternal pronucleus being subject to the programmed demethylation that is normally restricted to the paternal pronucleus. Because human *NLRP2* and *NLRP7* are the product of a primate-specific duplication, mice are perhaps not the best model organisms to study these reproductive disorders.

Methods

Data-Set and Syntenic Analyses

A total of 116 amino-acid sequences for NLRP proteins were obtained from GenBank (Benson et al. 2011) and used in the phylogenetic analyses. We used a combination of genes annotated as NLRP or NLRP-like, identified as protein coding or as pseudogenes. Online Resource 2 contains the accession numbers of the sequences used in our analyses.

We focused on genes from human (*Homo sapiens*), macaque (*Macaca mulatta*), marmoset (*Callithrix jacchus*), cattle (*Bos taurus*), pig (*Sus scrofa*), dog (*Canis lupus familiaris*), horse (*Equus caballus*), mouse (*Mus musculus*), elephant (*Loxodonta africana*), opossum (*Monodelphis domestica*), platypus (*Ornithorhyunchus anatinus*), and chicken (*Gallus gallus*). Blastp (Altschul et al. 1997) was used with the 14 human NLRPs as queries to the non-redundant protein sequences (NCBI: nr) to assure no other NLRP proteins were undetected from these taxa. The sequence data used in the main phylogeny (Fig. 1) was collected on or before September 2011, using the current genome assemblies at the time. While new assemblies for dog, cattle, and pig are available at the time of this writing, there are no new assemblies for elephant, possum, or platypus. Therefore, we chose not to redo our analyses, for the new data do not pertain the initial diversification of *NLRP* genes.

A total of 112 protein sequences were available for *NLRP* genes from the species listed above and identified as protein coding. A gene annotated as *NLRP1*-like in chicken [GenBank:XP_422818.3] and one of the two *NLRP5s* annotated in pig [GenBank:NP_001156879.1] were outliers in our preliminary alignments and phylogenetic trees and were excluded from further analyses. The *NLRP1*-like gene of chicken is on chromosome 9 in a region of conserved synteny with human chromosome 3 from which no *NLRP* or *NLRP*-like has been reported. The results of blastp using chicken *NLRP1*-like as query shows that the most likely ortholog is an uncharacterized protein with *IFT80* and *IL12A* as flanking markers in mammalian species. The characterization of this protein as NLRP1-like is likely due to the presence of FIIND and DEATH domains, which are also present in mammalian *NLRP1* genes.

Six "pseudogenes" were included in our analysis: two from elephant, and one each from macaque, marmoset, opossum, and platypus. Pseudogene sequences were transformed into amino acid sequences by identifying exon boundaries and removing frame shifts. These reconstructed "proteins" aligned well with proteins encoded by orthologous *NLRP* genes.

Blastn was performed to look for other pseudo-genes or gene traces in the genomes of possum and platypus, but no significant hits were found. Additionally, orthologous and paralogous relationships were established by inspecting flanking markers of all *NLRP* genes in the genome assemblies for all species used. We considered enough evidence for the absence of an *NLRP* gene/pseudogene when neither Blast nor syntenic analyses could find suitable candidates (see Online Resource 3).

When a gene was suspected absent, we performed a tblastn search with the corresponding human *NLRP* gene(s) as query. A cut-off E-value of 10^{-10} was used throughout. The only non-*NLRP* hits in non-human mammals were

CARD4, *CARD15*, *CIITA*, *NOD1*, *NOD2*, *NOD4*, *NLRC3*, *NLRC5*, *NLRX1*, and *RNH1*. All of these hits are due to the presence of shared domains with *NLRP* genes. Furthermore, all these non-human hits were found in synteny with the corresponding human genes and their flanking genes.

Sequence Alignment and Phylogenetic Analyses

We used UGENE as our working platform (Okonechnikov et al. 2012). Multiple sequence alignments were performed with T-Coffee (Notredame et al. 2000) and are available upon request. The phylogenetic trees were reconstructed with Maximum Likelihood (ML) methods using Garli (Zwickl 2006) and Bayesian estimation (BA) using MrBayes (Ronquist and Huelsenbeck 2003; Huelsenbeck and Ronquist 2001).

For the Bayesian reconstructions we used 2 runs, each with 4 chains and a burn-in fraction of 0.25. The runs were carried for an initial one million iterations, with increments of one million iterations until convergence. We declared a run to converge when the differences in the Log-Likelihood were less than 0.1 % for the last million iterations and the *Potential Scale Reduction Factor* convergence criterion was within 1 % of 1. For the Maximum Likelihood reconstruction we used 2 search repetitions running for 5 million generations with 100 bootstrap iterations from an initially random tree. The JTT+G+F model of amino-acid evolution using the observed aminoacid frequencies (Jones et al. 1992) was selected based on the best Log-Likelihood score provided by ProtTest (Darriba et al. 2011).

The consensus tree (Online Resource 1) was constructed by merging the results from ML and BA using DendroPy (Sukumaran and Holder 2010). We chose to present only the Bayesian tree in Fig. 1 because it is more resolved.

In contrast to Tian et al. (2009), we included all amino-acids available from all sequences into the analysis, which included several partial and low-quality sequences. It is generally accepted that using partial sequences diminishes the accuracy of the inferred phylogeny. This detrimental effect is, however, still not well understood, with researchers suggesting detrimental effects ranging from insignificant to severe Wiens (2003); Hartmann and Vision (2008); Burleigh et al. (2009); Kück et al. (2010). Several approaches have been suggested to increase phylogenetic inference accuracy, ranging from masking (i.e., removal) of problematic data to the statistical simulation of missing data Hartmann and Vision (2008); Kück et al. (2010). However, Bayesian and, to a lesser extent, Maximum Likelihood approaches seem to be particularly resilient to missing data, with consistently high accuracy being achieved even in the absence of up to 95 % of some of the samples (Wiens and Moen 2008; Wiens and Morrill 2011). In our case, only 11 out of the 116 sequences (less than 10%) are partial (*NLRP6* of elephant and platypus, *NLRPb* of possum, *NLRP3* and *NLRP11* of marmoset) or low quality (*NLRP5* and *NLRP10* of elephant and horse, respectively, *NLRP1* of elephant, and *NLRP11* of pig). Also, the missing data of partial sequences is presumably small and is therefore unlikely it has a strong effect on accuracy. In particular, the shortest partial sequence corresponds to the *NLRP6* of platypus which has a length of 417 amino-acids. The longest *NLRP6* is in macaques and has 1045 amino-acids, while the longest *NLRP* sequence overall is *NLRP1*, also in macaques, which has 1475 amino-acids. Therefore, we can conservatively estimate less than 30 % missing data in less than 10 % of the sequences.

In Online Resource 5, we performed a phylogenetic recontructions of the *NLRP* genes in Fig. 1 leaving out all reconstructed pseudogene sequences.

References

Alders M, Ryan A, Hodges M, Bliek J, Feinberg AP, Privitera O, Westerveld A, Little PFR, Mannens M (2000) Disruption of a novel imprinted zinc-finger gene, ZNF215, in Beckwith–Wiedemann syndrome. Am J Hum Genet 66:1473–1484

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

Benson D, Karsch-Mizrachi I, Lipman D, Ostell J, Wheeler D (2011) Genbank. Nucleic Acids Res 39:32–37

Burleigh JG, Hilu K, Soltis D (2009) Inferring phylogenies with incomplete data sets: a 5-gene, 567-taxon analysis of angiosperms. BMC Evol Biol 9:61

Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165

El-Maarri O, Seoud M, Coullin P, Herbiniaux U, Oldenburg J, Rouleau G, Slim R (2003) Maternal alleles acquiring paternal methylation in biparental complete hydatidiform moles. Hum Mol Genet 12:1405–1413

Fernandes R, Tsuda C, Perumalsamy AL, Naranian T, Chong J, Acton BM, Tong ZB, Nelson LM, Jurisicova A (2012) NLRP5 mediates mitochondrial function in mouse oocytes and embryos. Biol Reprod 86:138

Haig D (2005) The complex history of distal human chromosome 1q. Genomics 86:767–770

Hamatami T, Falco G, Carter MG, Akutsu H, Stagg CA, Sharov AA, Dudekula DB, VanBuren V, Ko MSH (2004) Age-associated alteration of gene expression patterns in mouse oocytes. Hum Mol Genet 13:2263–2278

Hartmann S, Vision T (2008) Using ests for phylogenomics: can one accurately infer a phylogenetic tree from a gappy alignment? BMC Evol Biol 8:95

Huelsenbeck JP, Ronquist F (2001) Mrbayes: Bayesian inference of phylogeny. Bioinformatics 17:754–755

Hughes AL (2006) Evolutionary relationships of vertebrate nacht domain-containing proteins. Immunogenetics 58:785–791

Inoue A, Zhang Y (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science 334:194

Iqbal K, Jin SG, Pfeifer GP, Szabó PE (2011) Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci USA 108:3642–3647

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comp Appl Biosci 8:275–282

Judson H, Hayward B, Sheridan E, Bonthron D (2002) A global disorder of imprinting in the human female germ line. Nature 416:539–42

Kersse K, Versputen J, Vanden Berghe T, Vandenabeele P (2011) The death-fold superfamily of homotypic interaction motifs. Trends Biochem Sci 36:541–552

Kim B, Kan R, Anguish L, Nelson LM, Coonrod SA (2010) Potential role for mater in cytoplasmic lattice formation in murine oocytes. PLoS ONE 5:e12587

Kou YC, Shao L, Peng HH, Rosetta R, del Gaudio D, Wagner AF, Al-Hussaini TK, van der Veyver IB (2008) A recurrent intragenic genomic duplication, other novel mutations in NLRP7 and imprinting defects in recurrent biparental hydatidiform moles. Mol Hum Reprod 14:33–40

Kück P, Meusemann K, Dambach J, Thormann B, von Reumont BM, Wägele JW, Misof B (2010) Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. Frontiers Zool 7:10

Kufer TA, Sansonetti PJ (2011) NLR functions beyond pathogen recognition. Nat Immunol 12:121–128

Laing KJ, Purcell MK, Winton JR, Hansen JD (2008) A genomic view of nod-like receptor family in teleost fish: identification of a novel NLR subfamly in zebrafish. BMC Evol Biol 8:42

Meyer E, Lim D, Pasha S, Tee LJ, Rahman F, Yates JR, Woods CG, Reik W, Maher ER (2009) Germline mutation in NLRP2 (NALP2) in a familial imprinting disorder (Beckwith–Wiedemann syndrome). PLoS Genet 85:e1000423

Murdoch S, Djuric U, Mazhar B, Seoud M, Khan R, Kuick R, Bagga R, Kircheisen R, Ao A, Ratti B, Hanash S, Rouleau GA, Slim R (2006) Mutations in NALP7 cause recurrent hydatidiform moles and reproductive wastage in humans. Nat Genet 38:300–301

Ng ACY, Eisenberg JM, Heath RJW, Huett A, Robinson CM, Nau GJ, Xavier RJ (2011) Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. Proc Natl Acad Sci USA 108:4631–4638

Notredame C, Heringa J, Higgins DG (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217

Ohsugi M, Zheng P, Baibakov B, Li L, Dean J (2008) Maternally-derived filia–mater complex localizes asymmetrically in cleavage-stage mouse embryos. Development 135:259–269

Okonechnikov K, Golosova O, Fursov M, The UGENE Team (2012) Unipro ugene: a unified bioinformatics toolkit. Bioinformatics 28:1166–1167

Parry DA, Logan CV, Hayward BE, Shires M, Landolsi H, Diggle C, Carr I, Rittore C, Touitou I, Philibert L, Fisher RA, Fallahian M, Huntriss JD, Picton HM, Malik S, Taylor GR, Johnson CA, Bonthron DT, Sheridan EG (2011) Mutations causing familial biparental hydatidiform mole implicate c6orf221 as a possible regulator of genomic imprinting in the human oocyte. Am J Hum Genet 89:451–458

Peng H, Chang B, Lu C, Su J, Wu Y, Lv P, Wang Y, Jun L, Zhang B, Quan F, Guo Z, Zhang Y (2012) NLRP2, a maternal effect gene required for early embryonic development in the mouse. PLoS ONE 7:e30344

Pierre A, Gautier M, Callebaut I, Bontoux M, Jeanpierre E, Pontarotti P, Monget P (2007) Atypical structure and phylogenomic evolution of the new eutherian oocyte- and embryo-expressed KHDC1/DPPA5/ECAT1/OOEP gene family. Genomics 90:583–594

Ponsuksili S, Brunner RM, Goldammer T, Kühn C, Walz C, Chomdej S, Tesfaye D, Schellander K, Wimmers K, Schwerin M (2006) Bovine NALP5, NALP8, and NALP9 genes: assignment to a QTL region and the expression in adult tissues, oocytes, and preimplantation embryos. Biol Reprod 74:577–584

Reik W, Walter J (2001) Evolution of imprinting mechanisms: the battle of the sexes begins in the zygote. Nat Genet 27:255–256

Ronquist F, Huelsenbeck JP (2003) Mrbayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Schulz R, Proudhon C, Bestor TH, Woodfine K, Lin CS, Lin SP, Prissette M, Oakey RJ, Bourc'his D (2010) The parental non-equivalence of imprinting control regions during mammalian development and evolution. PLoS Genet 6:e1001214

Strowig T, Henao-Mejia J, Elinav E, Flavell R (2012) Inflammasomes in health and disease. Nature 481:278–286

Sukumaran J, Holder M (2010) Dendropy: a python library for phylogenetic computing. Bioinformatics 26:1569–1571

Tian X, Pascal G, Monget P (2009) Evolution and functional divergence of NLRP genes in mammalian reproductive systems. BMC Evol Biol 9:202

Tong ZB, Gold L, Pfeifer KE, Dorward H, Lee E, Bondy CA, Dean J, Nelson LM (2000) Mater, a maternal effect gene required for early embryonic development in mice. Nat Genet 26:267–268

Wiens JJ (2003) Missing data, incomplete taxa, and phylogenetic accuracy. Syst Biol 52:528–538

Wiens JJ, Moen DS (2008) Missing data and the accuracy of Bayesian phylogenetics. J Syst Evol 46:307–314

Wiens JJ, Morrill MC (2011) Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. Syst Biol 60:719–731

Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun 2:241

Zhang P, Dixon M, Zucchelli M, Hambiliki F, Levkov L, Hovatta O, Kere J (2008) Expression analysis of the NLRP gene family suggests a role in human preimplantation development. PLoS ONE 3:e2755

Zheng P, Dean J (2009) Role of filia, a maternal effect gene, in maintaining euploidy during cleavage-stage mouse embryogenesis. Proc Natl Acad Sci USA 106:7473–7478

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. PhD Dissertation, The University of Texas at Austin, Austin