Evolution and functional divergence of NLRP genes in mammalian reproductive systems

Xin Tian, Géraldine Pascal, Philippe Monget

BMC Evolutionary Biology

Research article

Evolution and functional divergence of *NLRP* genes in mammalian reproductive systems

Xin Tian, Géraldine Pascal and Philippe Monget*

Address: Physiologie de la Reproduction et des Comportements, UMR 6175 INRA-CNRS-Université François Rabelais de Tours-Haras Nationaux, 37380 Nouzilly, France

Email: Xin Tian - xin.savois@tours.inra.fr; Géraldine Pascal - geraldine.pascal@tours.inra.fr; Philippe Monget* - Philippe.Monget@tours.inra.fr

* Corresponding author

Abstract

Background: NLRPs (Nucleotide-binding oligomerization domain, Leucine rich Repeat and Pyrin domain containing Proteins) are members of NLR (Nod-like receptors) protein family. Recent researches have shown that *NLRP* genes play important roles in both mammalian innate immune system and reproductive system. Several of *NLRP* genes were shown to be specifically expressed in the oocyte in mammals. The aim of the present work was to study how these genes evolved and diverged after their duplication, as well as whether natural selection played a role during their evolution.

Results: By using *in silico* methods, we have evaluated the evolution and functional divergence of *NLRP* genes, in particular of mouse reproduction-related *Nlrp* genes. We found that (1) major *NLRP* genes have been duplicated before the divergence of mammals, with certain lineage-specific duplications in primates (*NLRP7* and *11*) and in rodents (*Nlrp1*, *4* and *9* duplicates); (2) tandem duplication events gave rise to a mammalian reproduction-related *NLRP* cluster including *NLRP2*, *4*, *5*, *7*, *8*, *9*, *11*, *13* and *14* genes; (3) the function of mammalian oocyte-specific *NLRP* genes (*NLRP4*, *5*, *9* and *14*) might have diverged during gene evolution; (4) recent segmental duplications concerning *Nlrp4* copies and vomeronasal 1 receptor encoding genes (*V1r*) have been undertaken in the mouse; and (5) duplicates of *Nlrp4* and *9* in the mouse might have been subjected to adaptive evolution.

Conclusion: In conclusion, this study brings us novel information on the evolution of mammalian reproduction-related *NLRP*s. On the one hand, *NLRP* genes duplicated and functionally diversified in mammalian reproductive systems (such as *NLRP4*, *5*, *9* and *14*). On the other hand, during evolution, different lineages adapted to develop their own *NLRP* genes, particularly in reproductive function (such as the specific expansion of *Nlrp4* and *Nlrp9* in the mouse).

Background

The innate immune system is an ancestral and ubiquitous system of defense against microbial infection and other potential threats to the host. The first mammalian molecules shown to be involved in innate immune recognition

of, and defense against, microbial pathogens are the Toll-like receptors (TLRs), which constitute the main sensors for detection of extracellular microbes [1]. Recent research has shown that Nod-like receptors (NLRs) act as intracellular regulators of bacterial-induced inflammation [2-4].

The NLR protein family is separated into several sub-families based on the different N-terminal effecter domains [5]. The NLRP subfamily (known also as NALP family) is a new identified NLR group characterized by a PYRIN domain. Initially, research on the function of NLRP proteins focused on their roles in apoptotic and inflammatory signaling pathways via the formation of a large signaling platform (named inflammasome) and the activation of caspases in innate immunity [4,6,7]. Interestingly, more and more researches have recently revealed that some *NLRP* genes play roles in mammalian reproduction. For instance, the mouse *Nlrp5* (known also as *MATER*) was one of the first identified mammalian maternal effect genes, *i.e.* it encodes mRNA required for successful development of a fertilized oocyte [8]. In particular, the targeted invalidation of mouse oocyte-specific *Nlrp5* leads to the sterility of females due to a blockage at the two-cell stage in the development of embryos [8]. *In vitro* knockdown experiments in mouse fertilized eggs also revealed that a decrease of mouse germ-cell-specific *Nlrp14* transcripts results in an arrest of development between the one-cell and eight-cell stages of the embryos [9,10]. In the mouse, exclusive duplications of *Nlrp4* and *Nlrp9* have been detected with specific expression profiles restricted to the oocyte [9], whereas the mouse *Nlrp4a* and the cattle *NLRP9* have also been reported to be expressed in the testis [11,12]. In humans, the mutations of *NLRP7* are found to cause recurrent hydatidiform moles, spontaneous abortions, stillbirths and intrauterine growth retardation [13,14]. Additionally, the recent expression analyses of *NLRP* genes in the human and the rhesus macaque (*Macaca mulatta*) have shown that most if not all *NLRP* genes are expressed in primate gametes and early embryos, suggesting a role of *NLRPs* in primate pre-implantation development [15,16].

Overall, it seems that the *NLRP* genes play roles not only in innate immunity but also in the reproductive system of mammals. Moreover, *Nalp5* and *Nalp14* invalidation data in the mouse suggests that there is no functional compensation of each of these oocyte-specific genes. In this context, one question is to study how these reproduction-related *NLRP* genes evolved and diverged in function after their duplication, as well as whether natural selection played a role during their evolution.

The availability of several completely sequenced vertebrate genomes allows us to use a phylogenomic approach to identify the *NLRP* orthologues from different mammals and to evaluate the evolutionary features of *NLRPs* in different lineages. In this paper, we focus on the evolution and functional divergence of reproduction-related *NLRPs*, as well as lineage-specific expansion of *Nlrp4* and *Nlrp9* in rodents. The aim of the present work is to provide new information in understanding how *NLRP* genes evolve in

reproductive system and how different mammals adapted to develop their own *NLRP* gene copies particularly in reproductive biology.

Results
Phylogeny and syntenic comparison of NLRPs
A total of 83 mammalian NLRP amino acid sequences with chicken NLRP3-like as outgroup were used for the reconstruction of phylogenetic trees. A total of 122 positions are included in the final dataset.

In the consensus phylogenetic tree (Figure 1), each NLRP protein, except for NLRP2, is shown to be monophyletic by all four methods with high bootstrap values. NLRP2 encompasses the primate-specific NLRP7, suggesting the origin of *NLRP7* from a duplication of the *NLRP2/7* ancestor in primates. NLRP11 is also primate-specific and might be related to the reproduction-related NLRP4 and 9. NLRP4 proteins are only identified in primates and rodents, indicating the potential birth of the ancestral *NLRP4* gene before the divergence between primates and rodents. Otherwise, the other *NLRP* genes seem to have duplicated before the divergence of mammals.

In rodents, three *Nlrp* genes show lineage-specific duplications including *Nlrp1*, *Nlrp4* and *Nlrp9* (Figure 1). More specifically, *Nlrp1* exclusively duplicated in the mouse, *Nlrp9* duplicated independently in the rodents after the separation of mouse and rat, whereas the duplication of *Nlrp4* is likely to have occurred both before and after this separation.

Unexpectedly, *NLRP8* and *NLRP13* seem to be universal in mammals except for rodents. We speculate that these two genes might have been lost during evolution of the rodents. Unfortunately, our tBLASTn search (using the human *NLRP8* and *13* respectively as queries against the mouse genome) failed to find any trace of pseudogene of neither *NLRP8* nor *NLRP13* within the mouse conserved syntenic region. We found that, on the human chromosome 19, these two genes are located side by side between *NLRP4* and *NLRP5*, whereas the mouse *Nlrp4* gene has been expansively duplicated and the duplicates are located around *Nlrp5* on the mouse chromosome 7. In rodents, the biological significance of the probable loss of *NLRP8* and *13* on one hand, and the parallel expansion of *Nlrp4* on the other hand, remains unknown.

Remarkably, our phylogenetic analyses identified a well supported (bootstrap values >90% by ME, NJ and ML methods) reproduction-related clade including nine NLRP proteins: NLRP2, 4, 5, 7, 8, 9, 11, 13 and 14 (Figure 1). All these *NLRP* genes have been shown to be expressed in the human oocyte and pre-implantation embryos [16]. *NLRP4*, *5*, *8*, *9* and *14* are exclusively expressed in germ-
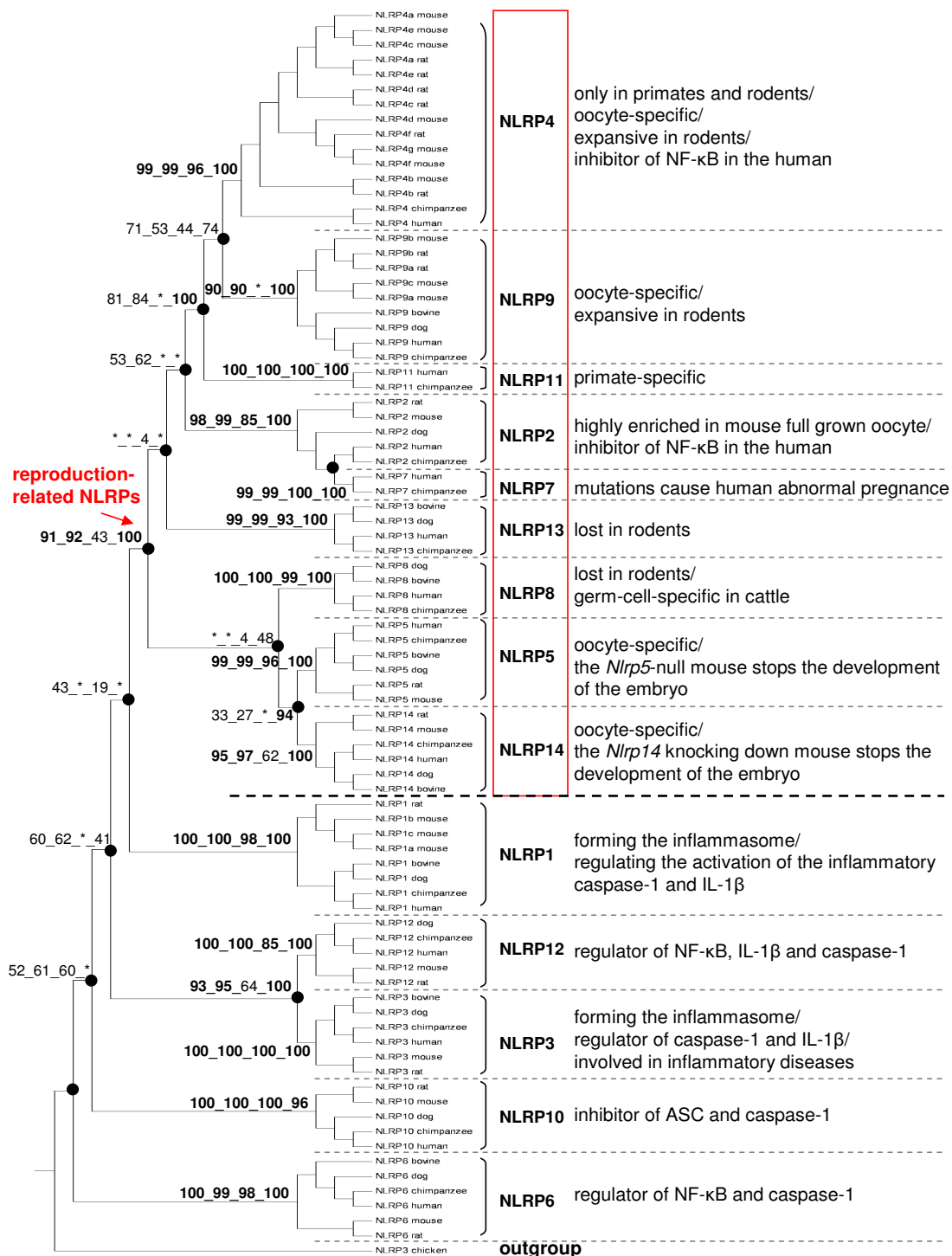
Figure 1
Consensus phylogenetic tree of NLRPs reconstructed by the fusion of four separate methods (NJ, ME, MP and ML). A total of 83 NLRP protein sequences from 6 mammalian species (cattle, dog, human, chimpanzee, mouse and rat) are used with chicken as outgroup. A total of 122 amino acid sites are included in the analyses. The bootstrap values are labeled on the main branches for the four different methods (followed the order of NJ, ME, MP and ML methods). The symbol * means that the branch is not supported by the corresponding method.

cells, especially in the oocyte of mammals [8-12,17,18]. The mutations of *NLRP7* are associated with abnormal embryo development in the human [13,14]. The expression of mouse *Nlrp2* has been found to be highly enriched in fully grown oocytes but diminished in the 2-cell embryos upon embryo genome activation [19]. In both the human [16] and the rhesus macaque [15], *NLRP11* and *13* have been found specifically or preferentially expressed in the oocyte, with a similar expression pattern to other oocyte-specific *NLRP* genes, *i.e.*, enriched in maturing oocytes and then progressively diminished in embryos, indicating an exclusive maternal origin of these transcripts. Unfortunately, the biological study of these two genes is not sufficient to address their putative roles in reproduction.

Moreover, this reproduction-related clade is also supported by the syntenic analysis (Figure 2). In the human genome, all the reproduction-related *NLRP* genes, except for *NLRP14*, are tandemly located on the chromosome 19q13.42, suggesting that multiple tandem duplication events might have given birth to these genes. The similar arrangements of the reproduction-related *NLRPs* are also found in cattle and dog genomes (Figure 2). In the mouse genome, the reproduction-related *Nlrps* (except for *Nlrp14*) are not located side by side, but interrupted by some other types of genes, especially and mainly by certain *V1r* genes (discussed below) on chromosome 7A1 and 7A3. Although *NLRP14* is not located in the syntenic region encompassing other reproduction-related *NLRPs* in all the studied mammalian genomes (involving cattle, dog, human, chimpanzee, mouse and rat), our phylogenetic result and the published expression profile show that this gene is close to other reproduction-related *NLRPs*. This suggests that early (before the divergence of mammals) genomic rearrangements might have resulted in the separation of *NLRP14* from its relatives.

We identified the other *NLRP* members (*NLRP1*, *3*, *6*, *10* and *12*) as non-reproduction-related *NLRP* genes, because they are not in the reproduction-related clade in the phylogenetic trees and ubiquitously expressed. These genes have been shown to participate in inflammatory and immune responses by regulating the activation of other signaling factors, such as NF-κB, caspase-1 and IL-1β [20-24]. In contrast, little is known about the function of the reproduction-related NLRP proteins in inflammation and immunity, except that the human NLRP2, 4 and 7 may be able to inhibit IL-1β and/or NF-κB [25-28], and thus may contribute to modulate the inflammatory response. It is known that inflammation and bacterial infection can cause infertility, ectopic pregnancy and abortion [29]. Thus, NLRP proteins might play roles both in innate immunity and in reproductive biology.
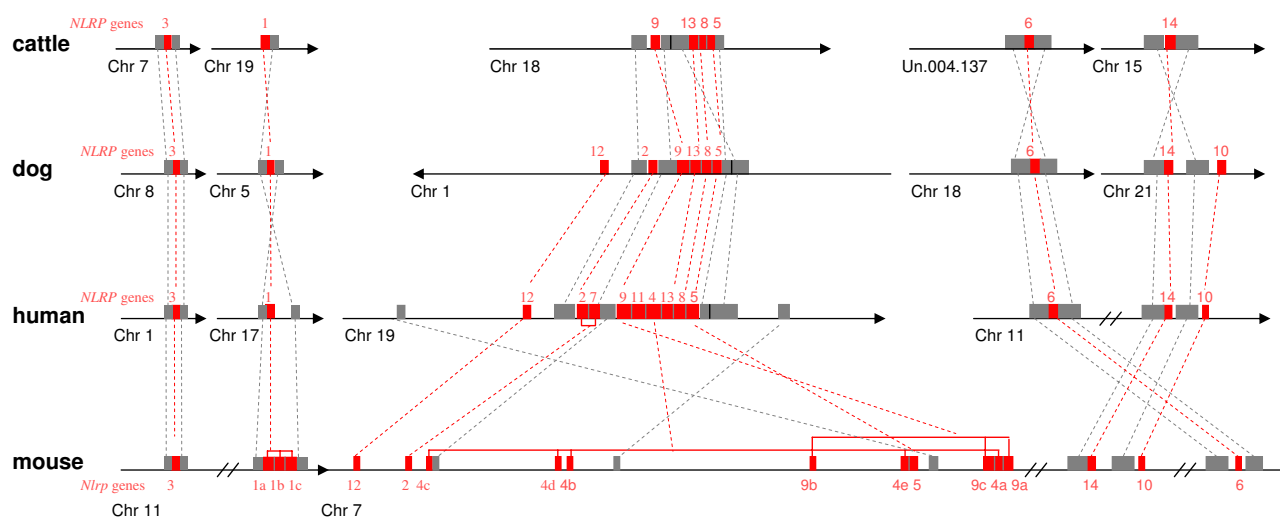
Figure 2
Syntenic comparison of *NLRP* genes among the cattle, the dog, the human and the mouse. The *NLRP* genes are marked in red, the orthologues are indicated by discontinuous lines. The order of *NLRPs* in each genome is listed as following: **The cattle:** *NLRP3* (chr 7); *NLRP1* (chr 19); *NLRP9, 13, 8* and *5* (chr 18); *NLRP6* (Un.004.137); *NLRP14* (chr 15). **The dog:** *NLRP3* (chr 8); *NLRP1* (chr 5); *NLRP12, 2, 9, 13, 8* and *5* (chr 1); *NLRP6* (chr 18); *NLRP14* and *10* (chr 21). **The human:** *NLRP3* (chr 1); *NLRP1* (chr 17); *NLRP12, 2, 9, 13, 8* and *5* (chr 19); *NLRP6, NLRP14* and *10* (chr 11). **The mouse:** *Nlrp3, 1a, 1b* and *1c* (chr 11); *Nlrp12, 2, 4c, 4d, 4b, 9b, 4e, 5, 9c, 4a, 9a, 14, 10* and *6* (chr 7).

Functional divergence

In the reproduction-related *NLRP* cluster referred by the phylogenetic analysis above, there are several oocyte-specific *NLRP* genes including *NLRP4*, *5*, *8*, *9* and *14*. To investigate whether functional divergence have occurred following the duplications of these oocyte-specific *NLRPs*, four oocyte-expressed *NLRP* genes (*NLRP4*, *5*, *9* and *14*) were estimated for functional divergence by using the DIVERGE program [30]. *NLRP8*, which was not found in rodents, was therefore not included in this analysis in order to compare the level of divergence across all candidate species.

After alignment and removal of the sites with gaps, a total of 650 amino acid sites are included in the analysis. Pairwise comparisons of each of the four oocyte-expressed *NLRP* gene clusters (Additional file 1) were carried out and the rate of amino acid evolution at each sequence position was estimated. As shown in Table 1, functional divergence is significant between each comparison ($\theta > 0$ with $p < 0.001$) by both MFE and MLE methods, indicating that site-specific altered selective constraints should contribute to the functional evolution of the oocyte-specific *NLRP* genes after their duplication.

Furthermore, the important amino acid residues, responsible for functional divergence, were predicted by calculating the site-specific profile based on posterior analysis for all pairs of clusters with functional divergence. In order to extensively reduce false positive, cutoff values, *i.e.*, the minimal posterior probabilities for RFD (Residues with predicted Functional Divergence) were established empirically by progressively removing the highest scoring residues from the alignment until θ dropped to zero. As shown in Table 1, the least RFD (33 residues, covering 5.1% of a total of 650 aligned sites) are observed between *NLRP5* and *NLRP14*, and the most RFD (118 residues, covering 18.2% of a total of 650 aligned sites) are observed between *NLRP4* and *NLRP14*. In general, the RFD are detected in all three functional domains (NACHT, PYRIN and LRRs) of NLRP proteins, implying

that shifted functional constrains might have acted on each protein domain.

Segmental duplications concerning **Nlrp4** and **V1r** genes in the mouse genome

In the mouse, there are seven *Nlrp4* gene copies (named from *Nlrp4a* to *4g*) which are specifically expressed in oocytes [9,11]. Our lab has also found the expression of *Nlrp4a* in the testis by RT-PCR [11]. *Nlrp4a-4e* genes are located on chromosome 7A, but *Nlrp4f* and *4g* are located on chromosomes 13B3 and chromosome 9 respectively. By the tBLASTn method, we identified six additional putative pseudogenes on chromosome 7A (named here *ΨNlrp4h-4m*, Figure 3a). Based on the phylogenetic tree (Figure 3a), multiple duplications of mouse *Nlrp4* could be depicted as follows. *Nlrp4b* might be the earliest *Nlrp4* copy in the mouse genome with an orthologue present in the rat genome, suggesting an origin before the separation of mouse and rat. The duplication of ancestral *Nlrp4b* is supposed to have given rise to *Nlrp4d*, and the subsequent duplications might have resulted in *Nlrp4f/4g* and *Nlrp4a/4c*. Subsequent chromosomal rearrangements might have caused the independent locations of *Nlrp4f* and *Nlrp4g* with other *Nlrp4* duplicates. All the putative *Nlrp4h-4m* pseudogene sequences show high identities with each other as well as with *Nlrp4e*. In fact, in all the pseudogenes, a common frameshift of 1-bp deletion was detected at the position corresponding to 358aa of *Nlrp4e* and it caused a subsequent premature stop codon at the position of 434aa, suggesting that the pseudogenization had happened before duplications of these copies.

In the mouse genome, *Nlrp4* duplicates (except for *Nlrp4f* and *4g*) are located between 6.00 Mb and 27.32 Mb on the chromosome 7. Interestingly, all these *Nlrp4* duplicates are found next to the vomeronasal 1 receptor encoding genes (*V1r*), such as *V1rd* and *V1re* genes (Figure 3a). The phylogeny and genomic location of these *V1r* genes reveal that *V1r* genes might have a similar duplication model to their neighbors, *Nlrp4* genes (Figure 3a). Thus, one may hypothesize that *Nlrp4* may have not duplicated one by

Table 1: Pairwise comparison of functional divergence of mouse oocyte-expressed *Nlrp* genes

Gene cluster	MFE θ	MLE θ	p value	RFD No.	Cutoff
Nlrp4/9	0.65 ± 0.07	0.61 ± 0.05	< 0.0001	91	0.78
Nlrp4/14	0.96 ± 0.08	0.89 ± 0.06	< 0.0001	118	0.94
Nlrp4/5	0.68 ± 0.08	0.67 ± 0.06	< 0.0001	61	0.84
Nlrp9/14	0.76 ± 0.08	0.65 ± 0.06	< 0.0001	59	0.80
Nlrp5/9	0.64 ± 0.09	0.58 ± 0.07	< 0.0001	53	0.75
Nlrp5/14	0.49 ± 0.10	0.45 ± 0.07	< 0.0001	33	0.70

θ, coefficient of functional divergence; *MFE*, Model-Free Method; *MLE*, Maximum-Likelihood Estimation under two-state model; *p* value, significance level calculated by the method of Fisher's transformation on z scale, and from chi square on LRT; *RFD No.*, number of residues with predicted functional divergence;-*cutoff*, the minimal posterior probability for amino acids causing functional divergence, established empirically by progressively removing the highest scoring residues from the alignment until θ dropped to zero.
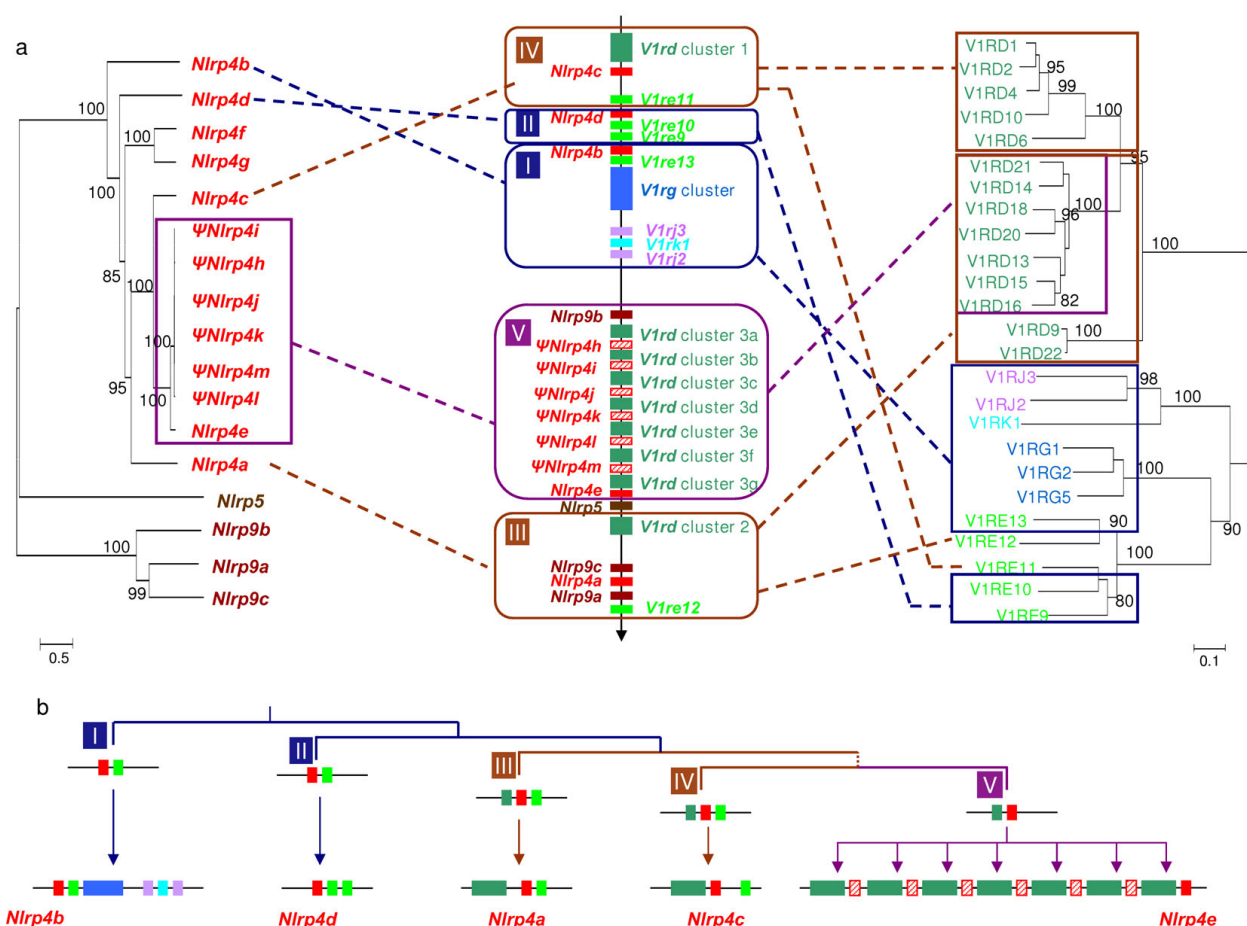
Figure 3
Segmental duplications of *Nlrp4* duplicates and *V1r* genes in the mouse genome. Figure 3a Comparison of gene relationships in the phylogenetic tree and their genomic position. Left: Phylogenetic tree of mouse *Nlrp4* duplicates (including pseudogenes) and its closely related *Nlrp5* and *Nlrp9* genes. A total of 439 nucleotide sites are included in the final dataset. Right: Phylogenetic tree of 25 mouse *V1r* proteins whose encoding genes are located in the same genomic region as *Nlrp4* genes. A total of 271 amino acid sites are included in the final dataset. The statistical confidence of each branch was estimated by the bootstrap method with 1000 replications, and only the values ≥ 80% are labeled in the trees. Middle: Genomic organization of related *Nlrp* and *V1r* genes. Five putative regions (I V) might be concerned in segmental duplication. Figure 3b A possible scenario of segmental duplication deduced from synteny combining phylogeny: Genomic region I (involving *Nlrp4b*) and region II (involving *Nlrp4d*) might have been resulted from an early duplication, and the subsequence duplication events have given birth to region III (involving *Nlrp4a*) and region IV (involving *Nlrp4c*). Region V might have been duplicated later, and in this genomic region, other more recent segmental duplications (concerning pseudogenized *Nlrp4* copy and *V1rd* genes) have been undertaken several times. We note that the order of duplication events in this potential scenario is mainly based on the topology of *Nlrp4* phylogenetic tree. There might be other alternative itinerary to explain the present complex pattern of this genomic region.

one, but were rather involved in segmental duplications encompassing several different genes, such as *V1r* genes.

Vomeronasal organ, which is well developed in rodents in comparison with human, detects pheromones and other chemical signals implicated in innate reproductive and social behaviours between the members of the same spe-

cies [31,32]. The *V1r* gene family is supposed to have emerged during the divergence of placental mammals (80 110 MYA), but many subfamilies including *V1rd* and *V1re* originated only in rodents after their separation from the primates [33], which is consistent with the rodent-specific duplication of *Nlrp4*. Thus, the phylogenetic relationship and genomic location (Figure 3a) of these *V1r* genes,

that are located in the segmental duplication region, could help us to better understand the segmental duplications. Overall, as shown in Figure 3b, a presumed itinerary of segmental duplication was deduced by combining phylogeny with genomic location of both *Nlrp4* duplicates and *V1r* genes. Although we can not confirm the order of duplication events by the present data, we can state that segmental duplication events characterize the evolution of this genomic region.

Positive selection on **Nlrp4** *and* **Nlrp9** *paralogues in the mouse*

Several examples have shown that genes in duplication blocks could be maintained by positive selection [34-37]. Here, we implemented PAML4 [38] to investigate the model of selective force acting on mouse *Nlrp4* and *Nlrp9* duplicates during their evolution.

By using Site Models, estimates of the parameter values under M2a and M8 (Table 2) indicate that a fraction of sites (about 56%) may be under positive selection in both datasets (*Nlrp4*_mouse and *Nlrp9*_mouse). LRTs of all comparisons (including M1a versus M2a, M7 versus M8 and M8a versus M8) are significant ($p < 0.001$) for both datasets (Table 2), implying that selective forces varied among amino acid sites and that the mouse *Nlrp4* and *9*

genes might have been subjected to positive selection after their expansion. Furthermore, the comparison of model MEC (AICc = 15913.11 for *Nlrp4*_mouse and AICc = 14196.22 for *Nlrp9*_mouse) versus model M8a (AICc = 16003.68 for *Nlrp4*_mouse and AICc = 14293.43 for *Nlrp9*_mouse) also supports positive selection in these two gene clusters. The AICc scores of MEC are lower than those of M8a, indicating that MEC better fits the data, *i.e.* positive selection is significant.

Subsequently, we wanted to know if the target of positive selection had changed during evolution in these two gene clades. We compared the potential sites subjected to positive selection between the two datasets (Table 2). In *Nlrp4*_mouse dataset, 2 sites (133C and 677V) are identified as positively selected sites at $p > 95\%$ level by both M2a and M8 models, and 2 more sites (153S and 657N) are identified only by the M8 model. In *Nlrp9*_mouse dataset, 10 sites (106H, 123 E, 356F, 360L, 378R, 513E, 536V, 571V, 661L and 948A) are identified as potential targets of positive selection by both M2a and M8 models, and 1 more sites (278L) is identified only by the M8 model. We note that no common sites are shared by the two datasets as positively selected sites at $p > 95\%$ level, suggesting a dramatic shift in the target of positive selection between the *Nlrp4* and *Nlrp9* genes in the mouse.

Table 2: Parameter estimates and likelihood scores for site models in PAML

model	*l*	Estimates of parameters	$2\Delta l$	Positively selected sites (BEB)
*Nlrp4*_mouse				
M1a	-7999.63	$\rho_0 = 0.33$		Not allowed
M2a	-7986.58	$\rho_0 = 0.28, (\rho_1 = 0.67), \rho_s = 0.05$ $\omega_s = 3.93$	26.09* (M2a vs M1a)	2 sites > 95% 133C, 677V
M7	-7999.85	$p = 0.04, q = 0.02$		Not allowed
M8a	-7999.55	$\rho_0 = 0.34, (\rho_1 = 0.66),$ $p = 0.81, q = 6.69$	27.56* (M8 vs M7)	Not allowed
M8	-7986.07	$\rho_0 = 0.94, \rho_s = 0.06,$ $p = 0.20, q = 0.09,$ $\omega_s = 3.64$	26.98* (M8 vs M8a)	4 sites > 95% 133C, 153S, 657N, 677V
*Nlrp9*_mouse				
M1a	-7140.29	$\rho_0 = 0.36043$		Not allowed
M2a	-7107.99	$\rho_0 = 0.28, \rho_1 = 0.66, \rho_s = 0.06,$ $\omega_2 = 10.17$	64.60* (M2a vs M1a)	10 sites > 95%, 2 sites > 99% 106H, 123 E, 356F, 360L, **378R**, 513E, **536V**, 571V, 661L, 948A
M7	-7140.98	$p = 0.02, q = 0.01$		Not allowed
M8a	-7140.29	$\rho_0 = 0.36, (\rho_1 = 0.64),$ $p = 0.01, q = 98.97$	65.99* (M8 vs M7)	Not allowed
M8	-7107.99	$\rho_0 = 0.94, \rho_s = 0.06,$ $p = 0.01, q = 0.01,$ $\omega_s = 10.20$	64.60* (M8 vs M8a)	11 sites > 95%, 6 sites > 99% 106H, **123 E**, 278L, 356F, **360L**, **378R**, 513E, **536V**, **571V**, 661L, **948A**

NOTE. Log-likelihood values (*l*) are given for each model. The value ω_S is the average d_N/d_S ratio for sites under positive selection in the models M2a and M8, p and q are the shape parameters for the beta distribution of ω in M7 and M8. ρ_0, ρ_1, and ρ_S are the proportions of codons subject to purifying, neutral, and positive selection, respectively. * p value is significant at 0.001. Predicted positively selected sites with posterior probabilities > 95% are listed, and the sites with $p > 99\%$ are in boldface. Site numbers and amino acids refer to the reference sequences *Nlrp4a* (NP_766484) and *Nlrp9a* (NP_001041684) for the two datasets, respectively.

When the 3D structures or models of these proteins are available in the future, we could investigate whether these positively selected sites are located in functionally pivotal regions.

Discussion

Independent evolution of reproduction-related NLRP *genes in different mammalian lineages*

Phylogenetic analyses show that *NLRP* genes have been originated and duplicated before the divergence of mammals. Certain genes are well conserved during their evolution, such as *NLRP3* (82% identity between the human and the mouse), whereas many genes involved in reproduction have rapidly evolved resulting in higher sequence divergence among lineages, such as *NLRP5* (50% identity between the human and the mouse) and *NLRP14* (62% identity between the human and the mouse). The pairwise estimates of evolutionary divergence between the reproduction-related paralogues are ranged from 61% to 75% (except that the pairwise distance between *NLRP2* and its recent primate-specific duplicate, *NLRP7*, is 45%), which is obviously higher than the one between the non-reproduction-related paralogues (ranged from 43% to 69%). Moreover, gene duplication and/or gene loss is found to have occurred independently in different mammalian lineages. On the one hand, *Nlrp4* (originated before the divergence of primates and rodents) and *Nlrp9* have been extensively duplicated in rodents, and other lineage-specific gene duplication events concern the specific duplication of *Nlrp1* (three paralogues) in the mouse as well as the unique origin of *NLRP7* and *NLRP11* in primates. On the other hand, *Nlrp8* and *Nlrp13* are seemed to be lost from the genomes of rodents. Interestingly, the major gene duplication and gene loss events are found to be associated with the reproduction-related genes (such as *NLRP4*, *7*, *8*, *9*, *11* and *13*), implying that reproduction-related genes have undertaken a fast and diverged evolution among different mammalian lineages. Given note that major reproduction-related *NLRP* genes are germ-cell specific, whereas the other non-reproduction-related genes are usually expressed in multiple tissues from a single organism. Recent large-scale gene expression studies have shown that the tissue specificity of genes is correlated positively with gene evolution rates [39-41]. Furthermore, our result reinforces the hypothesis that reproduction-related genes are most highly divergent and evolve more rapidly than genes expressed in other organs [42-45].

Functional divergence of the oocyte-specific genes

Gene duplication is thought to be the essential source of gene novelty, with new or altered functions, as shown by widespread existence of gene families. Among *NLRP* genes, there are at least four oocyte specifically expressed in mammals, including *NLRP4*, *5*, *9* and *14*. These genes have been found to be restricted expressed in oocytes and

early embryos [9,15], suggesting their important roles in oogenesis and preimplantation embryo development. Additionally, their temporal expression in oocytes and degradation during preimplantation development coincide with the timing of gene expression transition from maternal to zygotic control, indicating their roles as maternal effect genes.

In the present work, by using an *in silico* method (with the help of the program DIVERGE), we demonstrate that the functional divergence could have occurred between each pair of oocyte-specific *NLRP* genes. This result is supported by certain published experimental data. For example, the knock-out of mouse *Nlrp5* induced female infertility due to a blockage of early embryonic development [8]; and the injection of siRNA against *Nlrp14* into fertilized eggs resulted in arrested development of embryos between 1-cell and 8-cell stages [9]. These results suggest that the expression of other oocyte-specific *NLRP* genes is not able to compensate the absence of neither *NLRP5* nor *NLRP14* genes. Additionally, recent research has shown that *NLRP14* is also expressed in the testis and its mutation might cause spermatogenic failure in the human [46]. Thus, we presume that the expressional divergence of *NLRP5* and *NLRP14* during the development of mouse germ-cells and embryos may lead to functional specialization. On the other hand, target invalidation/inhibition of *NLRP5* and *14* did not reveal any deleterious effect on ovarian folliculogenesis but rather in early embryonic development, suggesting other genes might play the similar roles during folliculogenesis and meiotic maturation. For these latter functions, plausible candidates could be other oocyte-specific *NLRP* genes, such as *NLRP4* and *NLRP9*. To investigate this possibility, further functional studies such as knock-out models or other targeted inhibition experiments on these genes should be carried out.

Gene duplication is thought to be a major driving force in enabling the evolution of tissue specialization. By using microarray gene expression data from mammals [47], fruit flies and yeast [48], it has recently shown that as gene family size enlarges, there is a general trend for paralogous genes with decreased breadth and increased specificity of expression. In particular, by studying the relationship between gene family size and expression breadth of 1249 protein families in the mouse, Freilich and his colleagues [49] have recently demonstrated that duplicates that arose through post-multicellularity duplication events show a tendency to become more specifically expressed, supporting the view suggested by the subfunctionalization model [50]. In our case, especially with the evidence that *NLRP5* and *NLRP14* genes are significantly involved in different developmental stages of embryos, we hypothesize that the functional divergence of oocyte-specific *NLRP* duplicates

might be derived by the expression specialization of duplicate genes, which could provide the mammals reproductive advantages such as in adaptation.

Segmental duplication in the mouse genome

The phylogenetic research indicates the rodent-specific expansion of *Nlrp4* and *NLRP9*, and the furthermore genomic analyses find that these *Nlrp* duplicates, especially *Nlrp4*, have been duplicated together with other genes, such as *V1r* copies. This segmental duplication is not rare during the mouse and human genomic evolution. It has recently revealed that the segmental duplication constitutes about 5% of mouse genome, and its distribution is in a highly non-random fashion [51]. Interestingly, the duplication blocks account for 32% of the first 50 Mb of the chromosome 7, where the segmental duplications concerning *Nlrp4* and *V1r* genes are located. *Nlrp9* has been also duplicated in this region, but its duplication scenario is not clear and might be due to subsequent recombination.

As found in the human [52], the segmental duplication regions in the mouse might also be "hot spots" for the occurrence of non-allelic homologous recombination, leading to genomic mutations such as deletion, duplication, inversion or translocation [53-55]. Thus, these instable regions might have a biological significance in form of genome evolution [56,57].

Adaptive diversification after gene duplications in **Nlrp4** and **Nlrp9**

In the mouse, *Nlrp4* and *Nlrp9* are specifically extensive in gene copies. By using Site-Models implemented in PAML4, we evaluated the variation of selective pressure acting on *Nlrp4* and *Nlrp9* duplicates, respectively. Our result shows that positive selection is significant in both datasets, which consistent with other reports referring that the rapid divergence of the reproductive genes may be driven by positive selection [43,58]. In particular, in the mouse, several gene cluster concerning lineage-specific expansion of reproduction-related genes have been found to be under adaptive evolution (positive selection), such as the *Sva* (seminal vesicle autoantigen) gene cluster [59], the *Prame* (or oogenesin) gene cluster [60], the *Psg* (pregnancy-specific glycoprotein) gene cluster [61], the *Rhox4* gene cluster [62] and the *Svs* (seminal vesicle secretion proteins) gene cluster [63]. The lineage-independent expansion and subsequently rapid evolution of such genes might contribute to speciation or adaptation, such as genetic barriers between species and hybrid incompatibilities, or provide the species specification in reproductive processes such as sperm competition, host immunity to pathogens, and manipulation of female/male reproductive physiology and behavior. In the case of *NLRP4* and *9* genes, their biological functions are still unclear, so

we can not evaluate the significance of their specific duplications in the reproduction of the mouse. However, the similar evolutionary mechanism driven by positive selection suggests that after duplication, these *NLRP* gene copies diverged and acquired abilities (probably by subfunctionalization) to adapt the new environment. In the subsequent research, the adaptive evolution of the other reproduction-related *NLRP* genes could be further evaluated. Hopefully, when the biological functions of these *NLRP* genes are elucidated, the nature of the selective pressure acting on *NLRPs* will be better understood.

Interestingly, the evolution of *V1r* genes, the other genes involved in the segmental duplication with *Nlrp4* and *9*, is also driven by positive selection in rodents [33]. V1R proteins are thought to be responsible for the detection of pheromones that induce innate reproductive behaviors between members of the same species [31,32]. Thus, it is presumed that the adaptive evolution of the *V1r* genes might play an important role in reinforcing pre-zygotic barriers among species of rodents [33].

Conclusion

NLRP genes have originated and duplicated mainly before the divergence of mammals. During evolution, *NLRPs*, especially the reproduction-related *NLRP* genes, have undergone a fast and independent diversification in different mammalian lineages. The expansion of reproduction-related *NLRP* genes has been associated with functional divergence after duplication, suggesting that each *NLRP* oocyte-specific gene might play an essential role in oogenesis and early preimplantation embryo development. The mouse-specific expansion of *Nlrp4* and *Nlrp9*, concerning segmental duplication events, has been driven by positive evolution. Our founding suggests that the duplication and functional divergence of *NLRP* might provide mammals advantages in reproductive biology.

Methods

Molecular phylogenetic and syntenic analyses

The protein sequences of all the 14 known human NLRPs were retrieved from the GenBank http://www.ncbi.nlm.nih.gov/. The NLRP proteins from other species were searched by PSI-BLAST [64] with human NLRP protein sequences as queries against the protein databases (NCBI: RefSeq protein databases) of chimpanzee (*Pan troglodytes*), dog (*Canis familiaris*), cattle (*Bos taurus*), mouse (*Mus musculus*), rat (*Rattus norvegicus*) and chicken (*Gallus gallus*). The predicted coding sequences of the best hit proteins were retrieved when the hits presented more than 80% in length to be aligned with the query sequence (with E values < 10^{-100}). These settings could effectively detect the potential NLRP members from different species but avoid involving the relative NLR proteins from other protein subfamilies. After removal of

redundant sequences, the initial data set (Additional file 2) for NLRP phylogenetic studies includes 83 protein sequences from 6 mammals and 1 sequence from chicken.

Analyses of the orthologous and paralogous relationships among different species were carried out by combining the phylogenetic reconstruction of the gene family with the syntenic comparison. Multiple sequence alignments were performed using the Clustal W algorithm [65] and then manually edited. The alignment was submitted as Online Additional file 3. All alignment gap sites were eliminated before phylogenetic analyses. The phylogenetic trees were reconstructed with Neighbor Joining (NJ), Minimum-Evolution (ME) and Maximum Parsimony (MP) methods implemented in MEGA4 [66], as well as with Maximum Likelihood (ML) method implemented in PhyML [67]. The consensus phylogenetic tree [68] was generated by the fusion of the independent trees reconstructed by the four methods. In all analyses, chicken NLRP protein (Q5F3J4) was treated as the root of all the mammalian NLRPs. The bootstrap values [69] were estimated with 1000 replications. The syntenic comparison is based on Ensembl http://www.ensembl.org/index.html utilities such as "orthologue prediction" and "view syntenic regions".

Test for functional divergence of oocyte-specific NLRP genes

The program DIVERGE [30] was used to estimate type I functional divergence [70,71] between oocyte-expressed *NLRP* paralogues (*NLRP4*, *5*, *9* and *14*). Type I sites represent amino acid residues conserved in one clade (gene cluster) but highly variable in another, suggesting that these residues have been subjected to different functional constraints. Statistically, this functional divergence between two clades is measured by the coefficient of functional divergence, θ, ranging from 0 to 1. A null hypothesis of $\theta = 0$ indicates that the evolutionary rate is virtually the same between two duplicate genes at each site [70,71]. If the null hypothesis was rejected, a site-specific profile was then used to predict the critical amino acid residues most likely to be responsible for the detected functional divergence.

A set of 33 protein sequences from mammalian NLRP4, 5, 9 and 14 is included in the analysis. Sequences of rat NLRP4d, 9a and 14 were excluded from the alignment because of their shorter lengths. Before the use of DIVERGE, our sequence dataset was examined to satisfy the conditions recommended by Gu et Vander Velden (2002), which permit us to obtain higher efficiency of detecting functional divergence-related residues: 1) each cluster has more than four amino acid sequences; (2) all pairwise sequence identities are <90%; and (3) multiple alignment is reliable. The phylogenetic tree used for

DIVERGE was reconstructed by MEGA4 [66] with ME method. The coefficients of functional divergence (θ) between gene clusters are calculated by Model-Free Estimation (MFE) and Maximum-Likelihood Estimation under two-state model (MLE). To detect amino acid residues reflecting functional divergence, all four reproduction-related *NLRP* gene clades were compared to each other.

Investigation of mouse specific segmental duplications

In order to investigate the predicted segmental duplications concerning *Nlrp4* duplicates on the mouse chromosome 7, a tBLASTn search [64] was implemented by using the mouse *Nlrp4a* (the longest *Nlrp4* sequence) as a query against the mouse genome. This method permitted us to identify both expressed *Nlrp4* genes and traces of the pseudogenes. Then we combined phylogenetic analyses and genomic mapping to identify the segmental duplication [35]. The phylogenetic tree of mouse *Nlrp4* genes and pseudogenes was reconstructed by using nucleotide sequences instead of protein sequences. The phylogenetic tree was also reconstructed for vomeronasal 1 receptor proteins (V1R), because their encoding genes are located paralleling to *Nlrp4* duplicates in the mouse genome and they are presumed to have duplicated together with *Nlrp4*. The accession numbers of the V1R protein sequences used here are listed in Additional file 4. The genomic organization of all the concerned genes and pseudogenes in the mouse is inferred from Ensembl (release 49).

Evolutionary analyses

To examine whether the duplicates of *Nlrp4* and *Nlrp9* in the mouse have been submitted to adaptive evolution, an analysis of variation in selective pressure following gene duplication events was carried out with the CODEML program implemented in PAML4 [38]. The alignments were resulted from Clustal W and PAL2NAL [72]. The shorter mouse *Nlrp4g* was excluded in order to obtain more informative sites.

Two datasets named *Nlrp4*_mouse (including 7 sequences) and *Nlrp9*_mouse (including 3 sequences) were investigated for different Site Models [73] implemented in PAML4. In this study, we employed three pairs of models including M1a (NearlyNeutral: $0 < \omega_0 < 1$ and $\omega_1 = 1$) versus M2a (PositiveSelection: $0 < \omega_0 < 1$, $\omega_1 = 1$ and $\omega_s > 1$) [73], M7 (beta: $0 < \omega < 1$) versus M8 (beta and ω: $0 < \omega < 1$ and $\omega_s > 1$) [74], and M8a (beta & $\omega_s = 1$: $0 < \omega < 1$ and $\omega_s = 1$) versus M8 [75]. LRTs (Likelihood Ratio Tests) were used to test for significant differences in the fit of the models incorporating selection relative to their (nested) counterparts that did not allow positive selection [73]. These tests provide a useful series of metrices for interpreting the significance of the results and a degree of protection against false positives [76]. Bayes empirical Bayes (BEB)

method [77] implemented in PAML4 was used to estimate posterior probabilities of selection on each codon. Furthermore, a newly described model MEC [78] was also employed on the Selecton Server [79] to compare with the results from other models for positive selection. The advantage of the MEC model over the other models used here is that by treating specifically each amino-acid replacement, *Ka* is computed differently. So under the MEC model, a position with radical replacements will obtain a higher *Ka* value than a position with more moderate replacements [78].

Authors' contributions
XT performed the main data collection and analyses. GP helped to guide in the bioinformatic analyses. PM conceived the study and helped to guide in the general analyses. All coauthors participated in manuscript preparation and review.

Additional material

Additional file 1
The phylogenetic tree of oocyte-expressed NLRPs *for DIVERGE analyses*. *The figure shows the phylogenetic tree of oocyte-specific* NLRPs *derived from amino acid sequences for DIVERGE analyses.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-202-S1.pdf]

Additional file 2
The NLRP *sampling for all the analyses*. *The table shows the* NLRP *sampling (including the gene symbol, their genomic location and the accession number) for all the analyses.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-202-S2.pdf]

Additional file 3
The Clustal W alignment of 84 NLRP protein sequences for phylogenetic analyses. *The data shows the alignment result of 84 NLRP protein sequences used for phylogenetic analyses (Figure 1).*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-202-S3.pdf]

Additional file 4
The mouse V1r sampling used for phylogeny. *The Table shows the mouse* V1r *sampling used for phylogenetic analysis (Figure 3).*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-202-S4.pdf]

Acknowledgements

References
1. Barton GM, Medzhitov R: **Toll-like receptors and their ligands.** *Current topics in microbiology and immunology* 2002, **270:**81-92.
2. Chamaillard M, Girardin SE, Viala J, Philpott DJ: **Nods, Nalps and Naip: intracellular regulators of bacterial-induced inflammation.** *Cellular microbiology* 2003, **5(9):**581-592.
3. Fritz JH, Ferrero RL, Philpott DJ, Girardin SE: **Nod-like proteins in immunity, inflammation and disease.** *Nature immunology* 2006, **7(12):**1250-1257.
4. Meylan E, Tschopp J, Karin M: **Intracellular pattern recognition receptors in the host response.** *Nature* 2006, **442(7098):**39-44.
5. Ting JP, Lovering RC, Alnemri ES, Bertin J, Boss JM, Davis BK, Flavell RA, Girardin SE, Godzik A, Harton JA, *et al.*: **The NLR gene family: a standard nomenclature.** *Immunity* 2008, **28(3):**285-287.
6. Martinon F, Gaide O, Petrilli V, Mayor A, Tschopp J: **NALP inflammasomes: a central role in innate immunity.** *Seminars in immunopathology* 2007, **29(3):**213-229.
7. Tschopp J, Martinon F, Burns K: **NALPs: a novel protein family involved in inflammation.** *Nat Rev Mol Cell Biol* 2003, **4(2):**95-104.
8. Tong ZB, Gold L, Pfeifer KE, Dorward H, Lee E, Bondy CA, Dean J, Nelson LM: **Mater, a maternal effect gene required for early embryonic development in mice.** *Nature genetics* 2000, **26(3):**267-268.
9. Hamatani T, Falco G, Carter MG, Akutsu H, Stagg CA, Sharov AA, Dudekula DB, VanBuren V, Ko MS: **Age-associated alteration of gene expression patterns in mouse oocytes.** *Human molecular genetics* 2004, **13(19):**2263-2278.
10. Horikawa M, Kirkman NJ, Mayo KE, Mulders SM, Zhou J, Bondy CA, Hsu SY, King GJ, Adashi EY: **The mouse germ-cell-specific leucine-rich repeat protein NALP14: a member of the NACHT nucleoside triphosphatase family.** *Biology of reproduction* 2005, **72(4):**879-889.
11. Dade S, Callebaut I, Paillisson A, Bontoux M, Dalbies-Tran R, Monget P: **In silico identification and structural features of six new genes similar to MATER specifically expressed in the oocyte.** *Biochemical and biophysical research communications* 2004, **324(2):**547-553.
12. Ponsuksili S, Brunner RM, Goldammer T, Kuhn C, Walz C, Chomdej S, Tesfaye D, Schellander K, Wimmers K, Schwerin M: **Bovine NALP5, NALP8, and NALP9 genes: assignment to a QTL region and the expression in adult tissues, oocytes, and pre-implantation embryos.** *Biology of reproduction* 2006, **74(3):**577-584.
13. Murdoch S, Djuric U, Mazhar B, Seoud M, Khan R, Kuick R, Bagga R, Kircheisen R, Ao A, Ratti B, *et al.*: **Mutations in NALP7 cause recurrent hydatidiform moles and reproductive wastage in humans.** *Nature genetics* 2006, **38(3):**300-302.
14. Qian J, Deveault C, Bagga R, Xie X, Slim R: **Women heterozygous for NALP7/NLRP7 mutations are at risk for reproductive wastage: report of two novel mutations.** *Human mutation* 2007, **28(7):**741.
15. McDaniel P, Wu X: **Identification of oocyte-selective NLRP genes in rhesus macaque monkeys (Macaca mulatta).** *Molecular reproduction and development* 2009, **76(2):**151-159.
16. Zhang P, Dixon M, Zucchelli M, Hambiliki F, Levkov L, Hovatta O, Kere J: **Expression analysis of the NLRP gene family suggests a role in human preimplantation development.** *PLoS ONE* 2008, **3(7):**e2755.
17. Dalbies-Tran R, Papillier P, Pennetier S, Uzbekova S, Monget P: **Bovine mater-like NALP9 is an oocyte marker gene.** *Molecular reproduction and development* 2005, **71(4):**414-421.
18. Tong ZB, Bondy CA, Zhou J, Nelson LM: **A human homologue of mouse Mater, a maternal effect gene essential for early embryonic development.** *Human reproduction (Oxford, England)* 2002, **17(4):**903-911.
19. Evsikov AV, Graber JH, Brockman JM, Hampl A, Holbrook AE, Singh P, Eppig JJ, Solter D, Knowles BB: **Cracking the egg: molecular dynamics and evolutionary aspects of the transition from the fully grown oocyte to embryo.** *Genes & development* 2006, **20(19):**2713-2727.
20. Grenier JM, Wang L, Manji GA, Huang WJ, Al-Garawi A, Kelly R, Carlson A, Merriam S, Lora JM, Briskin M, *et al.*: **Functional screening of five PYPAF family members identifies PYPAF5 as a novel regulator of NF-kappaB and caspase-1.** *FEBS letters* 2002, **530(13):**73-78.

21. Manji GA, Wang L, Geddes BJ, Brown M, Merriam S, Al-Garawi A, Mak S, Lora JM, Briskin M, Jurman M, *et al.*: **PYPAF1, a PYRIN-containing Apaf1-like protein that assembles with ASC and regulates activation of NF-kappa B.** *The Journal of biological chemistry* 2002, **277(13)**:11570-11575.

22. Martinon F, Burns K, Tschopp J: **The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta.** *Molecular cell* 2002, **10(2)**:417-426.

23. Wang L, Manji GA, Grenier JM, Al-Garawi A, Merriam S, Lora JM, Geddes BJ, Briskin M, DiStefano PS, Bertin J: **PYPAF7, a novel PYRIN-containing Apaf1-like protein that regulates activation of NF-kappa B and caspase-1-dependent cytokine processing.** *The Journal of biological chemistry* 2002, **277(33)**:29874-29880.

24. Wang Y, Hasegawa M, Imamura R, Kinoshita T, Kondo C, Konaka K, Suda T: **PYNOD, a novel Apaf-1/CED4-like protein is an inhibitor of ASC and caspase-1.** *International immunology* 2004, **16(6)**:777-786.

25. Bruey JM, Bruey-Sedano N, Newman R, Chandler S, Stehlik C, Reed JC: **PAN1/NALP2/PYPAF2, an inducible inflammatory mediator that regulates NF-kappaB and caspase-1 activation in macrophages.** *The Journal of biological chemistry* 2004, **279(50)**:51897-51907.

26. Fiorentino L, Stehlik C, Oliveira V, Ariza ME, Godzik A, Reed JC: **A novel PAAD-containing protein that modulates NF-kappa B induction by cytokines tumor necrosis factor-alpha and interleukin-1beta.** *The Journal of biological chemistry* 2002, **277(38)**:35333-35340.

27. Fontalba A, Gutierrez O, Fernandez-Luna JL: **NLRP2, an inhibitor of the NF-kappaB pathway, is transcriptionally activated by NF-kappaB and exhibits a nonfunctional allelic variant.** *J Immunol* 2007, **179(12)**:8519-8524.

28. Kinoshita T, Wang Y, Hasegawa M, Imamura R, Suda T: **PYPAF3, a PYRIN-containing APAF-1-like protein, is a feedback regulator of caspase-1-dependent interleukin-1beta secretion.** *The Journal of biological chemistry* 2005, **280(23)**:21720-21725.

29. Slim R, Mehio A: **The genetics of hydatidiform moles: new lights on an ancient disease.** *Clinical genetics* 2007, **71(1)**:25-34.

30. Gu X, Velden K Vander: **DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family.** *Bioinformatics (Oxford, England)* 2002, **18(3)**:500-501.

31. Halpern M: **The organization and function of the vomeronasal system.** *Annual review of neuroscience* 1987, **10**:325-362.

32. Keverne EB: **The vomeronasal organ.** *Science (New York, NY)* 1999, **286(5440)**:716-720.

33. Shi P, Bielawski JP, Yang H, Zhang YP: **Adaptive diversification of vomeronasal receptor 1 genes in rodents.** *Journal of molecular evolution* 2005, **60(5)**:566-576.

34. Samonte RV, Eichler EE: **Segmental duplications and the evolution of the primate genome.** *Nature reviews* 2002, **3(1)**:65-72.

35. Darbo E, Danchin EG, Mc Dermott MF, Pontarotti P: **Evolution of major histocompatibility complex by "en bloc" duplication before mammalian radiation.** *Immunogenetics* 2008, **60(8)**:423-438.

36. Bailey JA, Eichler EE: **Primate segmental duplications: crucibles of evolution, diversity and disease.** *Nature reviews* 2006, **7(7)**:552-564.

37. Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M, Eichler EE: **Positive selection of a gene family during the emergence of humans and African apes.** *Nature* 2001, **413(6855)**:514-519.

38. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Molecular biology and evolution* 2007, **24(8)**:1586-1591.

39. Duret L, Mouchiroud D: **Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate.** *Molecular biology and evolution* 2000, **17(1)**:68-74.

40. Winter EE, Goodstadt L, Ponting CP: **Elevated rates of protein secretion, evolution, and disease among tissue-specific genes.** *Genome research* 2004, **14(1)**:54-61.

41. Zhang L, Li WH: **Mammalian housekeeping genes evolve more slowly than tissue-specific genes.** *Molecular biology and evolution* 2004, **21(2)**:236-239.

42. Singh RS: **Patterns of species divergence and genetic theories of speciation.** In *Topics in Population Biology and Evolution* Edited by: Wohrmann K, Jain S. New York: Springer-Verlag; 1990:231-265.

43. Swanson WJ, Vacquier VD: **The rapid evolution of reproductive proteins.** *Nature reviews* 2002, **3(2)**:137-144.

44. Torgerson DG, Kulathinal RJ, Singh RS: **Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes.** *Molecular biology and evolution* 2002, **19(11)**:1973-1980.

45. Torgerson DG, Singh RS: **Rapid evolution through gene duplication and subfunctionalization of the testes-specific alpha4 proteasome subunits in Drosophila.** *Genetics* 2004, **168(3)**:1421-1432.

46. Westerveld GH, Korver CM, van Pelt AM, Leschot NJ, Veen F van der, Repping S, Lombardi MP: **Mutations in the testis-specific NALP14 gene in men suffering from spermatogenic failure.** *Human reproduction (Oxford, England)* 2006, **21(12)**:3178-3184.

47. Huminiecki L, Wolfe KH: **Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse.** *Genome research* 2004, **14(10A)**:1870-1879.

48. Gu Z, Rifkin SA, White KP, Li WH: **Duplicate genes increase gene expression diversity within and between species.** *Nature genetics* 2004, **36(6)**:577-579.

49. Freilich S, Massingham T, Blanc E, Goldovsky L, Thornton JM: **Relating tissue specialization to the differentiation of expression of singleton and duplicate mouse proteins.** *Genome biology* 2006, **7(10)**:R89.

50. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* 1999, **151(4)**:1531-1545.

51. She X, Cheng Z, Zollner S, Church DM, Eichler EE: **Mouse segmental duplication and copy number variation.** *Nature genetics* 2008, **40(7)**:909-914.

52. Emanuel BS, Shaikh TH: **Segmental duplications: an 'expanding' role in genomic instability and disease.** *Nature reviews* 2001, **2(10)**:791-800.

53. Giglio S, Broman KW, Matsumoto N, Calvari V, Gimelli G, Neumann T, Ohashi H, Voullaire L, Larizza D, Giorda R, *et al.*: **Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements.** *American journal of human genetics* 2001, **68(4)**:874-883.

54. Saglio G, Storlazzi CT, Giugliano E, Surace C, Anelli L, Rege-Cambrin G, Zagaria A, Jimenez Velasco A, Heiniger A, Scaravaglio P, *et al.*: **A 76-kb duplicon maps close to the BCR gene on chromosome 22 and the ABL gene on chromosome 9: possible involvement in the genesis of the Philadelphia chromosome translocation.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99(15)**:9882-9887.

55. Shaw CJ, Bi W, Lupski JR: **Genetic proof of unequal meiotic crossovers in reciprocal deletion and duplication of 17p11.2.** *American journal of human genetics* 2002, **71(5)**:1072-1081.

56. Armengol L, Marques-Bonet T, Cheung J, Khaja R, Gonzalez JR, Scherer SW, Navarro A, Estivill X: **Murine segmental duplications are hot spots for chromosome and gene evolution.** *Genomics* 2005, **86(6)**:692-700.

57. Jiang Z, Tang H, Ventura M, Cardone MF, Marques-Bonet T, She X, Pevzner PA, Eichler EE: **Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution.** *Nature genetics* 2007, **39(11)**:1361-1368.

58. Clark NL, Aagaard JE, Swanson WJ: **Evolution of reproductive proteins from animals and plants.** *Reproduction (Cambridge, England)* 2006, **131(1)**:11-22.

59. Osawa M, Horiuchi H, Tian W, Kaneko M: **Divergent evolution of the prolactin-inducible protein gene and related genes in the mouse genome.** *Gene* 2004, **325**:179-186.

60. Birtle Z, Goodstadt L, Ponting C: **Duplication and positive selection among hominin-specific PRAME genes.** *BMC genomics* 2005, **6**:120.

61. McLellan AS, Zimmermann W, Moore T: **Conservation of pregnancy-specific glycoprotein (PSG) N domains following independent expansions of the gene families in rodents and primates.** *BMC evolutionary biology* 2005, **5**:39.

62. Jackson M, Watt AJ, Gautier P, Gilchrist D, Driehaus J, Graham GJ, Keebler J, Prugnolle F, Awadalla P, Forrester LM: **A murine specific expansion of the Rhox cluster involved in embryonic stem cell biology is under natural selection.** *BMC genomics* 2006, **7**:212.

63. Karn RC, Clark NL, Nguyen ED, Swanson WJ: **Adaptive evolution in rodent seminal vesicle secretion proteins.** *Molecular biology and evolution* 2008, **25(11):**2301-2310.
64. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic acids research* 1997, **25(17):**3389-3402.
65. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic acids research* 1994, **22(22):**4673-4680.
66. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Molecular biology and evolution* 2007, **24(8):**1596-1599.
67. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Systematic biology* 2003, **52(5):**696-704.
68. Felsenstein J: **PHYLIP (Phylogeny Inference Package) version 3.6.** Distributed by the author Department of Genome Sciences, University of Washington, Seattle; 2004.
69. Felsenstein J: **Confidence limits on phylogenies: An approach using the bootstrap.** *Evolution* 1985, **39:**783-791.
70. Gu X: **Maximum-likelihood approach for gene family evolution under functional divergence.** *Molecular biology and evolution* 2001, **18(4):**453-464.
71. Gu X: **Statistical methods for testing functional divergence after gene duplication.** *Molecular biology and evolution* 1999, **16(12):**1664-1674.
72. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments.** *Nucleic acids research* 2006:W609-612.
73. Nielsen R, Yang Z: **Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene.** *Genetics* 1998, **148(3):**929-936.
74. Yang Z, Nielsen R, Goldman N, Pedersen AM: **Codon-substitution models for heterogeneous selection pressure at amino acid sites.** *Genetics* 2000, **155(1):**431-449.
75. Swanson WJ, Nielsen R, Yang Q: **Pervasive adaptive evolution in mammalian fertilization proteins.** *Molecular biology and evolution* 2003, **20(1):**18-20.
76. Anisimova M, Nielsen R, Yang Z: **Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites.** *Genetics* 2003, **164(3):**1229-1236.
77. Yang Z, Wong WS, Nielsen R: **Bayes empirical bayes inference of amino acid sites under positive selection.** *Molecular biology and evolution* 2005, **22(4):**1107-1118.
78. Doron-Faigenboim A, Pupko T: **A combined empirical and mechanistic codon model.** *Molecular biology and evolution* 2007, **24(2):**388-397.
79. Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T: **Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach.** *Nucleic acids research* 2007:W506-511.