**RESEARCH ARTICLE**　　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

CrossMark

# Characterization of the peripheral blood transcriptome and adaptive evolution of the MHC I and TLR gene families in the wolf (*Canis lupus*)

Guangshuai Liu, Honghai Zhang[*], Guolei Sun, Chao Zhao, Shuai Shang, Xiaodong Gao, Tian Xia and Xiufeng Yang

## Abstract

**Background:** The wolf (*Canis lupus*) is one of the most widely distributed terrestrial mammals, because it is well adapted to various ecological niches and their corresponding pathogen environments. Immunological competence is a crucial factor involved in adapting to a changing environment and fighting pathogen infection in animals. In this study, the peripheral blood transcriptome of wolves was generated via RNA-seq to advance understanding of the wolf immunome, with a special focus on the major histocompatibility complex class I (MHC I) and toll-like receptor (TLR) gene families, which are involved in pathogen recognition and defense.

**Results:** The blood transcriptomic libraries of eight wolves originating from Tibet and Inner Mongolia were sequenced, and approximately 383 million reads were generated. Using a genome-guided assembly strategy, we obtained 123,851 unigenes, with a mean length of 845 bp and an N50 length of 1121 bp. On the basis of BLAST searches against the NCBI non-redundant protein database (Nr), a total of 36,192 (29.22%) unigenes were annotated. For functional classification, 24,663 unigenes were assigned to 13,016 Gene Ontology (GO) terms belonging to 51 sub-categories of the three main GO categories. Additionally, 7682 unigenes were classified into 6 Kyoto Encyclopedia of Genes and Genomes (KEGG) categories, in which the most represented functional sub-categories were signal transduction and the immune system, and 16,238 unigenes were functionally classified into 25 Eukaryotic Orthologous Groups (KOG) categories. We observed an overall higher ω ($d_N/d_S$) value at antigen-binding sites (ABSs) than at non-ABS regions as well as clear evidence of intergenic/intragenic recombination events at wolf MHC I loci. Additionally, our analysis revealed that carnivorous TLRs were dominated by purifying selection, with mean ω values at each TLR locus ranging from 0.173 to 0.527. However, we also found significant instances of positive selection that acted on several codons in pathogen recognition domains and were linked to species-specific differences in pathogen recognition.

**Conclusions:** This study represents the first attempt to characterize the blood transcriptome of the wolf and to highlight the value of investigating the immune system. Balancing selection and recombination have contributed to the historical evolution of wolf MHC I genes. Moreover, TLRs in carnivores have undergone adaptive evolution against the background of purifying selection, and a high level of adaptive evolution was detected in the wolf TLR system.

**Keywords:** Wolf, Blood transcriptome, RNA-seq, Immune system, MHC, TLR

* Correspondence: zhanghonghai67@126.com
Qufu Normal University, Jingxuan Street No. 57, Qufu, Shandong province, China

Liu *et al. BMC Genomics* (2017) 18:584

Page 2 of 14

## Background

The gray wolf (*Canis lupus*) is one of the world's most widely distributed mammals and is found in a wide range of ecologically different habitats [1]. This distribution suggests that the wolf is exposed to a complex pathogenic environment, which might drive the evolution of immune-related genes. Immunological competence is crucial for animals to adapt to a changing environment and to fight pathogen infection. Hence, illustrating the nature of the immune system is essential to explore the immunological responses and disease resistance ability of wolves, and effective conservation efforts often make use of immunological insights [2, 3]. Furthermore, dogs are an increasingly important model system for understanding the genetic basis of phenotypic diversity [4] and the pathogenesis of many human diseases [5, 6]. Given the unusual evolutionary relationships between wolves and dogs, documenting the immune architecture of wolves can also provide insights into the adaptive changes that occurred in dogs during the domestication process.

Blood acts as the pipeline of the immune system, carrying various immunomodulatory factors and specific antibodies that are involved in innate immune defenses as well as in the initiation of adaptive immune responses [7]. Fluctuations in the expression profiles of genes in the blood recapitulate information about genetic, epigenetic, cellular and environmental factors [8]. Therefore, blood constitutes a particularly attractive source of surrogate tissue for assessing the status of the immune system, which best monitored by profiling transcript abundance in blood. Indeed, blood transcriptome profiling has been developed into a mainstream tool for exploring the pathogenesis of a wide range of human diseases [8, 9]. Many molecular and cellular profiling approaches are currently available for the trancriptome profiling analyses. As compared with other technologies, sequence-based transcriptomic assays, such as RNA-seq, enable genome-wide analysis of the complexity of whole transcriptional landscapes with the greatest breadth and robustness [10]. Blood transcriptome profiling has been successfully performed on several wild mammal species, such as koala [11], giant panda [12], polar bear and brown bear [13]. However, no studies addressing the wolf blood transcriptome have been published to date.

Pathogens are among the most remarkable selective forces acting on wildlife during their evolutionary history. Detailed investigation of evolutionary patterns in genes involved in pathogen recognition is crucial for understanding coevolutionary processes between pathogens and their hosts [14]. Thus, two pathogen recognition receptor gene families, the major histocompatibility complex (MHC) and the toll-like receptors (TLRs), which play central roles in the adaptive and innate immune systems of vertebrates, respectively, were selected for analysis in the present study. The MHC loci constitute a highly polymorphic region of the mammal genome, and variants of these genes have been associated with pathogen resistance and the long-term survival of species [15]. Three major groups of MHC genes have been distinguished in mammals [16]. The MHC class I and II genes encode cell-surface glycoproteins that present pathogen-derived antigens to effector cells and are involved in triggering the downstream adaptive immune cascade [15, 17]. Balancing selection is commonly accepted to be responsible for the maintenance of polymorphism at these loci in a wide range of taxa [18]. Recombination is also presumed to be an important evolutionary mechanism driving MHC allelic diversity and polymorphism [19]. A considerable amount of research has previously been conducted on wolf MHC II loci [20–22]. However, no assessment of the adaptive pattern of wolf MHC I loci has been performed to date. Seven MHC class I genes have been identified in the dog genome, among which four are transcribed and are designated functional genes: DLA-12, DLA-64, DLA-79 and DLA-88 [23, 24].

TLRs encode type I transmembrane glycoproteins that recognize a wide diversity of pathogens by sensing pathogen-associated molecular patterns (PAMPs), and they can be classified as either non-viral (TLR1, TLR2, TLR4, TLR5, TLR6 and TLR10) or viral TLRs (TLR3, TLR7, TLR8 and TLR9) according to ligand specificity in many mammals [25]. The general structure of TLRs is characterized by a variable number of pathogen-recognition extracellular leucine-rich repeats (LRRs), a transmembrane region and a conserved intracellular toll/interleukin-1 receptor (TIR) signaling domain [26]. Because TLRs lie directly at the host-pathogen interface, they are potentially subject to coevolutionary dynamics. TLRs are evolutionarily conserved across vertebrates. However, some studies on primates [27], rodents [28] and birds [29] have revealed different degrees of species-specific positive selection acting on TLR repertory. At present, the role of pathogen-mediated selection in the evolution of TLRs among carnivores remains unknown.

To obtain a comprehensive understanding of the immunological properties of wolves, we first performed genome-wide transcriptome sequencing of peripheral blood by using RNA-seq. Then, a powerful genome-guided de novo assembly strategy was implemented in Trinity software [30] and used to generate a high-quality transcriptome. We utilized the obtained sequence dataset to perform a general characterization of the blood transcriptome with regard to gene functional annotation, immune pathway identification and the expression levels of immune-related genes. Furthermore, we studied the evolutionary patterns of two immune receptor gene families (MHC I and TLRs) and attempted to reveal the

Liu *et al. BMC Genomics* (2017) 18:584

Page 3 of 14

adaptive mechanism of the wolf response to a complex pathogen environment. To the best of our knowledge, this study is the first attempt to characterize the peripheral blood transcriptome of the wolf.

## Results

### Illumina sequencing and assembly

A total of 383 million paired reads with a length of 125 bp were generated from eight blood transcriptome libraries. After trimming and quality filtering, 373 million (97.32%) high-quality clean reads were retained and were aligned to the dog reference genome (CanFam3.1). In total, 146,640 transcripts were collected and assembled into 123,851 unigenes with an average length of 845 bp and an N50 length of 1121 bp (Table 1). All of the clean reads were then aligned back to the unigenes to assess the read content of the assembly. The results showed that up to 83.7% of the reads mapped back to the unigenes, thus indicating that the assembled unigenes represented the transcriptome reads well.

### Annotation of unigenes

All of the unigenes were annotated by using BLAST to Nr, NCBI non-redundant nucleotide database (Nt), KOG, a manually annotated and reviewed protein sequence database (Swiss-Prot), KEGG, and GO databases. Results showed that the Nt database has the largest matches with 81,101 (65.5%) unigenes were annotated. In total, 44,576 (36.0%) unigenes had matches in the five public protein databases, and the 36,192 (29.22%) unigenes BLAST hits in the Nr database were studied as the Nr database had the maximum protein reference sequences (Additional file 1: Fig. S1A). The species distribution of the best BLAST hits was shown in Additional file 1: Fig. S1B, and expectedly, most annotated unigenes had hits to the canine (22,573, 62.4%). The E-value distribution of matches showed that 13,271 (36.7%) unigenes have a highly significant homology ($\leq$ 1.0E-100) and 3409 (9.4%)

**Table 1** Summary of the wolf blood transcriptome sequencing and assembly

| Category | Numbers |
| --- | --- |
| Total raw reads | 383,315,270 |
| Total clean reads | 373,038,916 |
| Total bases (bp) | 46,629,864,500 |
| Total assembled transcripts | 146,640 |
| Total assembled unigenes | 123,851 |
| Total length of unigenes (bp) | 104,598,095 |
| Max length of unigenes (bp) | 18,148 |
| Min length of unigenes (bp) | 301 |
| Average length of unigenes (bp) | 845 |
| Contig N10 (bp) | 4836 |
| Contig N50 (bp) | 1121 |

unigenes have E-value greater than 1E-10 (Additional file 1: Fig. S1C). The identity distribution of unigenes revealed that 28,208 (77.9%) unigenes have a high similarity (>80%) and 32,886 (90.9%) unigenes have a similarity greater than 60% (Additional file 1: Fig. S1D).

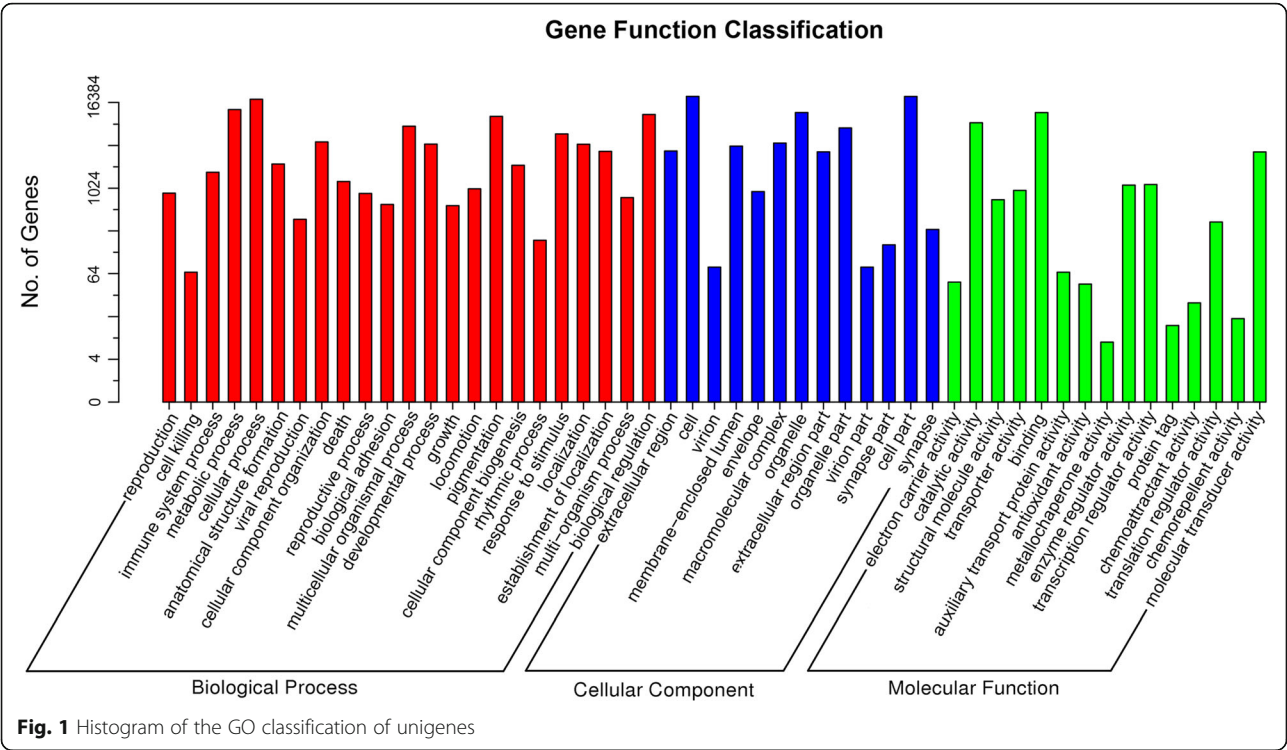### Functional annotation and classification results

On the basis of sequence homology against the Nr database, a total of 24,663 unigenes were annotated to 13,016 GO terms, belonging to 51 GO sub-categories (Fig. 1). There were 23 sub-categories in BP (Biological Process), 13 sub-categories in CC (Cellular Component), and 15 sub-categories in MF (Molecular Function), respectively. Cellular process (GO: 0009987) and the metabolic process (GO: 0008152) were the most represented BP categories (Additional file 2: Table S1).

KEGG analysis showed that a total of 13,665 unigenes had significant matches in the database, and 7682 unigenes were functionally classified into 6 KEGG categories, including 335 KEGG Orthology (KO) pathways (Additional file 3: Table S2). Among the 31 identified sub-categories within the KEGG categories excluding human diseases, signal transduction (1853) and the immune system (969) were the most highly represented categories (Fig. 2). Within the organismal systems functional category, it was notable that the immune system represented 969 of the 2230 unigenes involved in 15 immune-related KEGG pathways (Table 2, Additional file 4: Table S3). With the exception of the platelet activation pathway, the annotated immune-related pathways contained 79.0% mapped genes among the total known genes in the pathways. Among the identified immune pathways, the chemokine signaling pathway (ko04062) was the most represented. Pathways involving the MHC and TLR gene families were also identified, which included antigen processing and presentation (ko04612), the intestinal immune network for IgA production (ko04672) and natural killer cell-mediated cytotoxicity (ko04650) for MHC and the toll-like receptor signaling pathway (ko04620) for TLRs. Pathways acting as a functional bridge between the innate and adaptive immune responses were also identified, such as Fc gamma R-mediated phagocytosis, NOD-like receptor and RIG-I-like receptor signaling pathways [31–33] (Table 2, Additional file 4: Table S3).

By KOG classifications, 16,238 unigenes were annotated into 25 categories (Fig. 3). The cluster of general function prediction only exhibited 4296 (26.46%) unigenes, which constituted the most represented functional group. Signal transduction mechanism (2676, 16.48%) was in the second place, followed by posttranslational modification, protein turnover, chaperones at 1409 (8.68%) and function unknown at 1376 (8.47%). (Additional file 5: Table S4).

Liu *et al. BMC Genomics* (2017) 18:584

Page 4 of 14



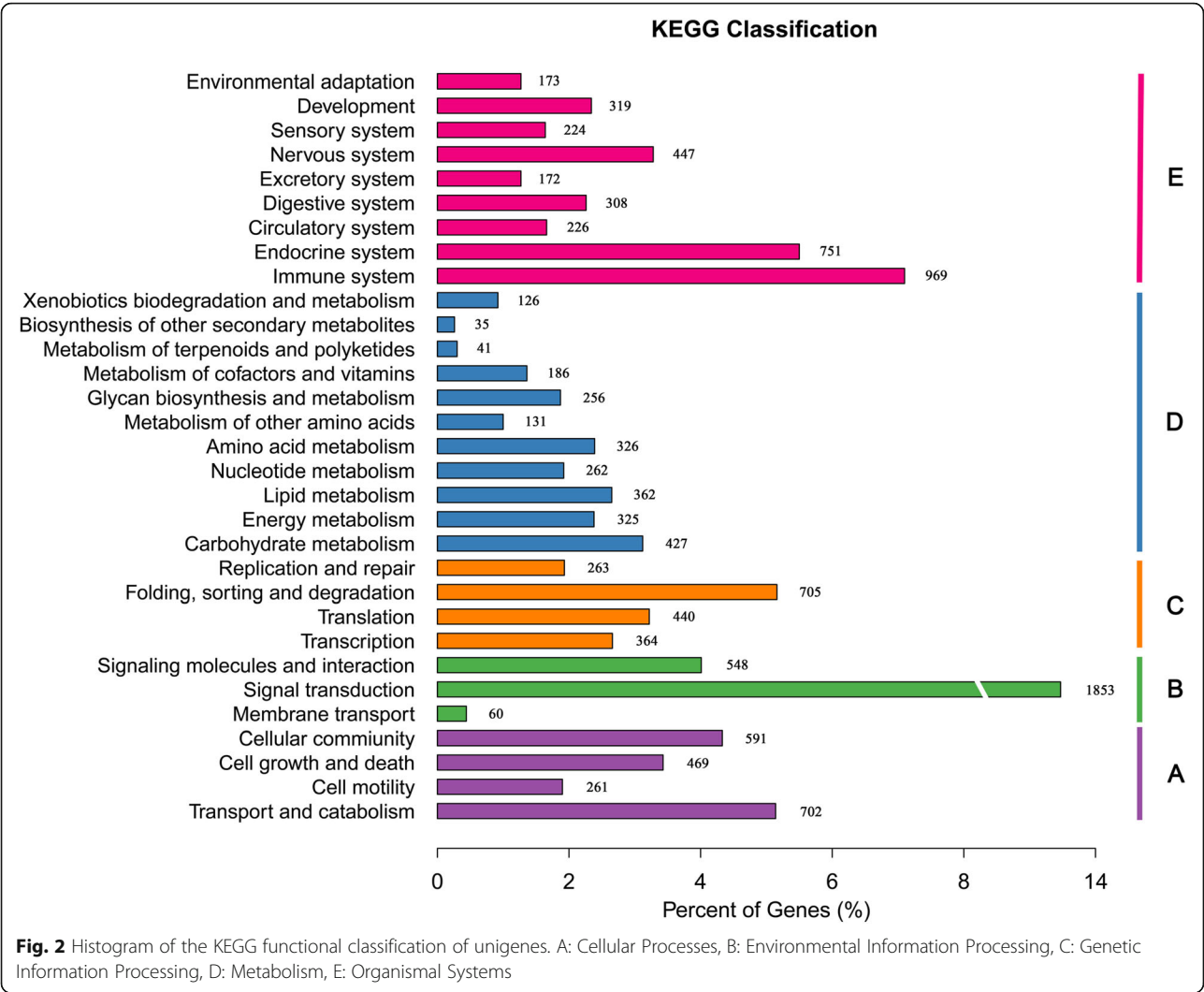**Fig. 1** Histogram of the GO classification of unigenes

## Evidence of positive selection and recombination in the MHC I gene family

The complete sequences encoding the antigen-binding domains of 4 MHC I genes were amplified from the cDNA of 4 blood samples that were also used in the transcriptome analyses. A total of 17 alleles (exon 2-exon 3) were identified among the 4 MHC I loci, and the number of alleles at each locus was a follows: 6 (DLA-12), 2 (DLA-64), 2 (DLA-79) and 7 (DLA-88) [GenBank: KX583612-KX583628; Additional file 6: Fig. S2]. Among these alleles, 2 DLA-12, 1 DLA-64, 1 DLA-79 and 6 DLA-88 alleles have not been identified previously in any canid species. On the basis of the BLAST results in terms of Nr and FPKM (fragments per kilobase of transcript per million fragments sequenced) values, we identified 4 and 12 transcripts of MHC I and MHC II genes, respectively, in the wolf blood transcriptome. Among the identified MHC genes, DLA-64 and DLA-DRB1 were the most highly expressed MHC I and MHC II genes, with FPKM values of 371 and 291, respectively (Additional file 7: Table S5).

Evidence of historical positive selection was detected in the ABSs (encoded by exon 2 and exon 3) of DLA-12 and DLA-88 (Table 3), corresponding to a pattern expected for MHC genes under balancing selection. Regarding the putative ABS of DLA-88, the rate of nonsynonymous substitutions ($d_N$) was significantly higher than the rate of synonymous substitutions ($d_S$) by more than 3-fold ($\omega = 3.3$; $p = 0.002$). A $\omega$ value greater than 1 was also detected at non-ABS positions, but this difference showed no significant deviation from neutrality (Table 3). Moreover, $d_N$ was 13.6 times higher in ABSs than in non-ABSs. Surprisingly, the $\omega$ value of the ABSs of DLA-12 was not significant ($p = 0.313$) contrary to the non-ABSs, although $d_N$ was higher than $d_S$ in both the ABSs and the non-ABSs, probably because of a high level of nonsynonymous activity in non-ABSs or a conservative ABS position classification method. Notably, the variants in DLA-12 alleles were all nonsynonymous changes (Additional file 6: Fig. S2). CODEML analysis showed that the M8 model fit the data better than the M7 model (Table 3). In a comparison of M8 and M7 site models, we detected 15 positively selected sites from MHC I genes, among which 10 sites were located at ABSs. Moreover, 8 and 11 codons were inferred to be positively selected sites in DLA-12 and DLA-88, respectively, among which 1 and 7 codons were located in ABSs (Table 3, Additional file 6: Fig. S2).

Screening for recombination with RDP yielded two DLA-88 alleles as intralocus recombinants, with one DLA-64 allele and one DLA-79 allele as interlocus recombinants (Additional file 6: Fig. S2). Analysis using GARD identified two significant breakpoints (268, 444; $p < 0.01$), and the first breakpoint was located at the boundary of exon 2 and exon 3 (Additional file 6: Fig. S2). RDP4 or the GARD method did not provide significant evidence of intralocus recombination within the other three MHC I genes. These combined findings suggested that positive selection and recombination (intergenic/intragenic recombination) have shaped the evolutionary patterns of the MHC I genes of the wolf.

Liu *et al. BMC Genomics* (2017) 18:584

Page 5 of 14



**Fig. 2** Histogram of the KEGG functional classification of unigenes. A: Cellular Processes, B: Environmental Information Processing, C: Genetic Information Processing, D: Metabolism, E: Organismal Systems

## Signatures of positive selection against a background of purifying selection across the carnivorous TLR gene family

By combining assembled unigenes and clone-sequencing results, we obtained the full CDS of 10 TLRs in the wolf [GenBank: KX756657-KX756666]. Comparison of the ratio of $d_N$ and $d_S$ revealed distinct overall purifying selection for all TLRs across the carnivorous lineage, with mean ω values ranging from 0.173 to 0.527 (Table 4).

Among the ten TLRs, TLR3, TLR5, TLR7, TLR8 and TLR9 possessed lower ω values than did the other 5 TLRs. Interestingly, with the exception of TLR5, which presented an ω value of 0.252, the former TLR category corresponded to receptors known to recognize nucleic acids from viruses, thus indicating stronger purifying selection in viral TLRs. However, despite a global trend of purifying selection acting on TLRs in carnivores, evidence of statistically significant positive selection was detected across most loci via codon-based maximum likelihood (ML)

methods (as indicated by the PAML M8, fixed-effect likelihood (FEL) and random effect likelihood (REL) methods), with a total of 62 codons (Table 4). Furthermore, 51 codons (82.3%) were observed in the extracellular LRR domain, thus suggesting that positive selection has acted mainly on pathogen recognition domains (Additional file 8: Tables S7–S16). The nature and strength of selection varied among 10 TLRs. For instance, the nonviral TLR4 exhibited a remarkable pattern of adaptive evolution, with 27 positively selected sites being identified. TLR7 appeared to show the strongest evolutionary constraint among the four virus TLRs, with no positively selected sites being identified by all three ML methods. We also noted that no codons under statistically significant positive selection were found in non-viral TLR10, although the ω value of this gene was greater than that of all virus TLRs (Table 4).

Species-specific positively selected sites were detected with the Mixed Effects Model of Evolution (MEME) algorithm.

Liu *et al. BMC Genomics* (2017) 18:584

Page 6 of 14

**Table 2** Immune-related pathways identified in the wolf blood transcriptome

| KO identifier | Pathway name | Annotated unigenes | Mapped genes | Genes in pathway |
|---|---|---|---|---|
| ko04062 | Chemokine signaling pathway | 216 | 114 | 146 |
| ko04660 | T cell receptor signaling pathway | 166 | 77 | 85 |
| ko04670 | Leukocyte transendothelial migration | 155 | 60 | 75 |
| ko04620 | Toll-like receptor signaling pathway | 136 | 67 | 76 |
| ko04650 | Natural killer cell mediated cytotoxicity | 132 | 66 | 81 |
| ko04666 | Fc gamma R-mediated phagocytosis | 121 | 50 | 58 |
| ko04640 | Hematopoietic cell lineage | 98 | 63 | 77 |
| ko04621 | NOD-like receptor signaling pathway | 93 | 39 | 51 |
| ko04662 | B cell receptor signaling pathway | 92 | 52 | 57 |
| ko04612 | Antigen processing and presentation | 91 | 36 | 41 |
| ko04664 | Fc epsilon RI signaling pathway | 87 | 39 | 45 |
| ko04622 | RIG-I-like receptor signaling pathway | 79 | 46 | 53 |
| ko04610 | Complement and coagulation cascades | 75 | 50 | 78 |
| ko04623 | Cytosolic DNA-sensing pathway | 62 | 41 | 51 |
| ko04672 | Intestinal immune network for IgA production | 59 | 28 | 37 |

Eleven codon positions in 7 wolf TLRs (TLR2–6, TLR8–9) were found to be affected by episodic selection, and 7 of the TLRs were located directly in the LRRs. Within the 7 sites in LRRs, site 309 in TLR2, 491 in TLR3, 306 in TLR4, and 87 and 504 in TLR9 showed radical amino acid changes in terms of size, polarity, or electric charge (Additional file 9: Table S17).

## Discussion

### Overview of the wolf peripheral blood transcriptome

The gray wolf is one of the most widely distributed predators in the world. A recent study of wolves of different ecotypes has identified statistical signatures of selection in immune-related genes, thus indicating that the immune system has been involved in the adaptive evolution of the wolf to various habitats [34]. In this study, we established the genetic architecture of the wolf peripheral blood transcriptome to advance understanding of the wolf immunome, with a special focus on the adaptive evolution patterns of two pathogen recognition receptor gene families (MHC I and TLR). We obtained approximately 373 million high-quality clean reads and assembled them into 123,851 unigenes by using Trinity. The unigenes exhibited a maximum length of 18,148 bp
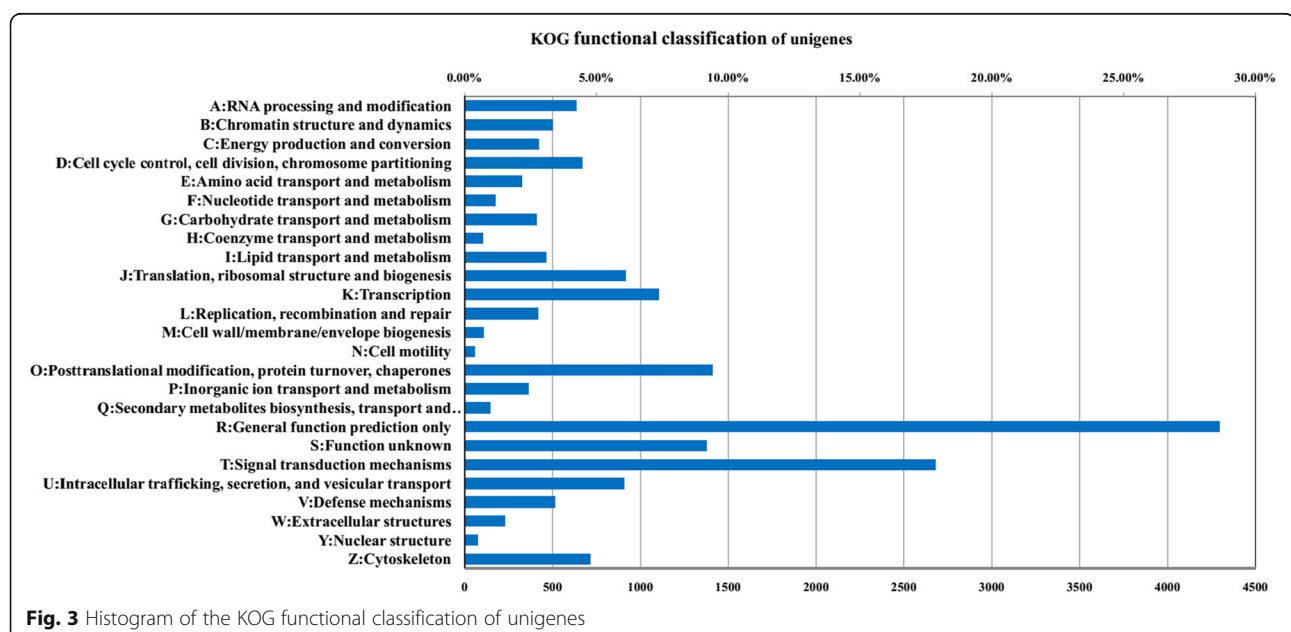


**Fig. 3** Histogram of the KOG functional classification of unigenes

Liu *et al. BMC Genomics* (2017) 18:584

Page 7 of 14

**Table 3** Rates of $d_N$ and $d_S$ substitutions in the ABSs and non-ABSs of wolf MHC I genes

| | | N | $d_N$ | $d_S$ | Z | p | -2lnΔL | PAML M8 |
|---|---|---|---|---|---|---|---|---|
| MHC I | ALL sites | 181 or 182 | 14.5 (1.5) | 10.2 (1.4) | **2.304** | **0.023** | 104.28** | 44, 61, <u>65</u>, <u>66</u>, <u>72</u>, <u>73</u>, <u>95</u>, <u>97</u>, <u>99</u>, 114, <u>116</u>, 138, <u>152</u>, <u>156</u>, 157 |
| | ABS | 29 | 41.6 (8.7) | 20.0 (6.2) | **2.69** | **0.008** | | |
| | Non-ABS | 152 or 153 | 10.6 (1.3) | 8.6 (1.4) | 1.063 | 0.29 | | |
| DLA-12 | All sites | 181 | 0.9 (0.3) | 0 | **2.834** | **0.005** | 14.15** | 49, 51, 61, 114, <u>116</u>, 127, 141, 154 |
| | ABS | 29 | 0.6 (0.5) | 0 | 1.013 | 0.313 | | |
| | Non-ABS | 152 | 0.9 (0.3) | 0 | **2.66** | **0.009** | | |
| DLA-88 | All sites | 182 | 4.8 (1.0) | 2.1 (0.6) | **2.67** | **0.009** | 81.47** | 61, <u>65</u>, <u>66</u>, <u>72</u>, <u>95</u>, <u>97</u>, 114, <u>116</u>, 138, <u>152</u>, 157 |
| | ABS | 29 | 21.8 (6.4) | 6.6 (2.7) | **3.12** | **0.002** | | |
| | Non-ABS | 153 | 2.7 (0.7) | 1.6 (0.6) | 1.17 | 0.245 | | |

N is the number of codons in each category. The $d_N$ and $d_S$ values are given as the percentage per site. The standard errors in parentheses were computed using 1000 bootstrap replicates. The inferred positively selected codons located at ABSs are underlined. Significant results are highlighted in bold. **$p < 0.01$

and an N50 value of 1121 bp, and up to 83.7% clean reads mapped to the assembled unigenes, thus indicating a high assembly quality of transcriptome sequences.

We then compared the unigenes against public databases to perform further functional annotation and classification. GO annotation showed that cellular process and metabolic process were the most represented BP categories, in agreement with the results reported in research in pandas [12]. We also noted that 100 unigenes were annotated for antigen processing and presentation (Additional file 2: Table S1), representing a high level of genetic diversity in immune-related genes in the wolf. This result was consistent with recent studies of MHC genetic diversity in the wolf [20–22]. These heterogeneous genes involved in antigen processing and presentation may lead to the recognition of a wide spectrum of pathogens and epidemics in the wolf. Signal transduction and the immune system were the most common KEGG sub-categories, thus

highlighting the primary functionality of intravascular blood in responding to dynamic environmental signals and in defenses against invading pathogens [8, 35]. Among the 15 identified immune pathways, the chemokine signaling pathway was the most represented, with all 5 known chemokine subfamilies being identified [36]. Interestingly, several studies have demonstrated that a range of genes in this pathway are involved in the wound healing process [37–39]. Consequently, it is reasonable to assume that the chemokine signaling pathway plays an important role in inflammation and host immune surveillance in the wolf, which is known to frequently suffer from a variety of injuries primarily resulting from scuffles with intended prey [40]. Pathways involving the MHC and TLR gene families were identified, thus suggesting that peripheral blood plays important roles in both the innate and adaptive immune systems. Furthermore, pathways functionally bridging the innate and adaptive immune responses were

**Table 4** Positively selected sites across carnivorous TLRs as inferred by different methods

| Locus | Size (aa) | -2lnΔL | PAML M8 | FEL | REL | Total No. of sites | Sites located in LRRs | ω |
|---|---|---|---|---|---|---|---|---|
| TLR1 | 789 | 11.57** | <u>629</u> | 69, <u>289</u>, <u>291</u>, <u>316</u>, <u>629</u> | <u>289</u>, <u>291</u>, <u>316</u>, 324, 326, 366, 395, 561, <u>629</u> | 10 | 8 | 0.527 |
| TLR2 | 766 | 5.88 | None | <u>557</u>, <u>609</u> | 487, <u>557</u>, <u>609</u> | 3 | 1 | 0.348 |
| TLR3 | 905 | 2.02 | None | <u>13</u>, <u>270</u>, <u>293</u> | <u>13</u>, 69, <u>270</u>, 279, <u>293</u>, 344, 587 | 7 | 6 | 0.307 |
| TLR4 | 833 | 7.24* | <u>193</u>, <u>295</u>, <u>341</u> | 46, <u>193</u>, <u>319</u>, <u>562</u>, <u>564</u> | 46, 58, 91, 139, <u>193</u>, <u>295</u>, 306, 308, <u>319</u>, 321, 339, <u>341</u>, 347, 364, 369, 384, 393, 394, 464, <u>505</u>, 510, 514, 517, <u>562</u>, <u>564</u>, 620, 832 | 27 | 24 | 0.493 |
| TLR5 | 858 | 1.60 | None | 63, 137, 609 | None | 3 | 2 | 0.252 |
| TLR6 | 797 | 2.34 | None | 109, 631 | None | 2 | 1 | 0.491 |
| TLR7 | 1050 | N.A. | None | None | None | - | - | 0.173 |
| TLR8 | 1038 | 2.42 | None | <u>234</u>, <u>457</u>, <u>683</u>, <u>689</u> | 202, <u>234</u>, 390, 416, <u>457</u>, <u>683</u>, <u>689</u> | 7 | 7 | 0.226 |
| TLR9 | 1032 | 10.70** | <u>5</u> | <u>5</u>, 457, 627 | None | 3 | 2 | 0.163 |
| TLR10 | 807 | N.A. | None | None | None | - | - | 0.439 |

Positively selected codons inferred by more than one ML method are underlined. The overall ω for each TLR was calculated by the SLAC method. N.A. indicates not available because of a low -2lnΔL value. *$p < 0.05$, **$p < 0.01$

Liu *et al. BMC Genomics* (2017) 18:584

Page 8 of 14

also identified (Table 2, Additional file 4: Table S3). These results revealed the complexity of immune mechanisms in wolf peripheral blood.

### MHC class I gene family

MHC genes constitute a central component of the adaptive immune response and are thought to be associated with pathogens resistance and the long-term survival of animals [15]. Therefore, research into this genetic system should provide valuable information about the underlying immune response mechanisms in animals. Although only 4 individuals were included in this study, we identified 17 alleles (exons 2–3) among the 4 MHC I genes. DLA-88 was found to be more polymorphic than the other three genes, in agreement with results from previous studies in dogs [23, 24]. DLA-12 also displayed a relatively high level of polymorphism, with 6 alleles identified. However, the variability of DLA-12 in dogs appears to be much more limited, and only 1 allele has been observed among 18 individuals [24]. Therefore, the DLA-12 polymorphic and genetic type may be related to the existent of divergent immunological properties between wolves and dogs. We also examined the expression levels of MHC genes and found that DLA-64 was the most highly expressed MHC gene (Additional file 7: Table S5). We successfully cloned DLA-12 and DLA-88 from wolf blood cDNA libraries, and the results were consistent with previous research on dogs [23, 41], thus suggesting that these two genes are expressed in canine blood. However, these two genes were not identified through the Nr BLAST results. We speculate that the abundant alleles and high sequence variability made it difficult to assemble the transcripts of DLA-12 and DLA-88, and quantitative real-time PCR may be a useful method for quantifying the expression levels of these two genes. Furthermore, wolf MHC class II genes, which were also expressed in wolf blood (Additional file 7: Table S5), showed greater polymorphism relative to that in other carnivores [21, 42]; this polymorphism may be necessary for wolves to cope with the complex pathogen environments of different habitats.

Previous studies have concluded that balancing selection maintains allelic diversity and an excess of $d_N$ over $d_S$ at MHC in vertebrates [43]. A higher value of ω at ABSs than non-ABSs supports the influence of balancing selection on the MHC I loci of the wolf, in accordance with published data on MHC I loci in other mammals [44–46]. Additionally, strong positive selection on ABSs has been reported for MHC II genes in wolves [21, 22]. Because the MHC genes are involved in antigen presentation, the increased number of nonsynonymous mutations in ABSs is likely to increase the wolf's ability to recognize a diverse range of pathogens. Evidence of positive selection was detected in both DLA-12 and DLA-

88, although the excess of nonsynonymous substitutions in the ABS of DLA-12 was not significant. Fifth codons were identified as positively selected sites by CODEML. Among these sites, 10 were located in ABSs, which are involved in peptide binding, thus suggesting that these sites may be of functional importance. Additionally, 1 and 7 positively selected sites were detected in the ABSs of DLA-12 and DLA-88 (Table 3, Additional file 6: Fig. S2), implying the existence of a conservative antigen-binding function for DLA-12 compared with DLA-88. In conclusion, different intensities of polymorphic and divergent selection might be indicative of distinct functional roles for wolf MHC I genes.

We detected clear evidence of intergenic/intragenic recombination that has shaped the evolution of wolf MHC I genes. An intragenic recombination effect was observed at DLA-88, but not at the other 3 MHC I loci, a result indicating the existence of distinct recombination patterns for different MHC I loci. Furthermore, we detected a significant breakpoint located near the boundary of exon 2 and exon 3 (Additional file 6:Fig. S2), which supports recombination through the mechanism of exon shuffling [47]. This mechanism has been identified in many lower vertebrates as well as several mammals, such as badgers and marmosets [48–50]. Collectively, our analyses supported a role of recombination (intra –/interlocus) and balancing selection in the generation and maintenance of wolf MHC I diversity. The complex evolutionary mechanisms of MHC I loci may indicate that the wolf has evolved its adaptive immunity along the history of host-pathogen coevolution.

### Toll-like receptor gene family

As a key family of innate immunity receptors, TLRs lie directly at the host-pathogen interface and might be involved in coevolutionary dynamics with their microbial counterparts [14]. This is the first study to thoroughly characterize the entire TLR repertoire of the wolf, and it provides the most extensive survey of TLR evolution in the carnivorous clade to date. We examined all 10 known TLRs and found a clear signature of purifying selection that was consistent with previous reports for primates and rodents [27, 28]. This conservative mode is presumably due to the need to preserve the well-established pathogenic sensing function of TLRs. We also found that viral TLRs were subject to stronger purifying selection than non-viral TLRs, except for TLR5. This dichotomy has been suggested to be due to the different nature of the microorganisms targeted by the two groups of TLRs [51]. Specifically, we suggest that the greater evolutionary flexibility of non-viral TLRs is compatible with more diverse PAMPs of non-viral pathogens. Among the 10 TLRs, TLR5 is the only member that recognizes an exclusively proteinaceous ligand of

Liu *et al. BMC Genomics* (2017) 18:584

Page 9 of 14

flagellins [52], and it is therefore reasonable to predict the existence of functional constraint for TLR5. Interestingly, TLR5 was the TLR under the strongest purifying selection among non-viral TLRs ($\omega = 0.252$), a pattern essentially the same as that reported in non-human primates [27]. In humans, however, TLR5 has been suggested to be functionally redundant. A loss-of-function TLR5 stop mutation is distributed nearly worldwide and exhibits a high frequency in some human populations (up to 23%) [53]. Another nonsynonymous mutation in human TLR5, leading to diminished NF-κB signaling, has recently been reported to confer a selective advantage [54]. Although the accessory mechanism of flagellin-recognition in the case of invalid TLR5 remains unclear, the different evolutionary pattern of TLR5 may reflect the different flagellin pathogen environments experienced by human and wildlife. TLR9 also appears to have evolved under strong stabilizing selection, as indicated by its lowest overall $\omega$ value. Furthermore, TLR9 is expressed at far higher levels than other TLRs (Additional file 7: Table S6). These results for TLR5 and TLR9 demonstrated their essential, non-redundant biological roles in carnivorous immune performance.

Despite pervasive purifying selection, we found evidence of positive selection acting upon a few codons, primarily within the pathogen recognition domains (i.e., LRRs) for most TLRs. This finding could be in agreement with species-specific differences in TLRs observed during related ligand recognition [55]. Carnivorous TLR4 displays a relatively high accumulation of codons exhibiting positive selection, a result consistent with those previously reported in other mammals [27, 28]. In addition to a well-established function in LPS recognition (i.e., the glycolipid in the outer membrane of Gram-negative bacteria), TLR4 has been implicated in the responses to other pathogens, including fungi, protozoans, and even viruses [56]. The broad spectra of targeted pathogens and the species-specific pathogenic environment could be the reasons for the remarkable pattern of adaptive evolution observed in TLR4. Further analysis has revealed that some of the sites (295, 339, 341 and 364) (Table 4) are located in or adjacent to the regions responsible for the interactions between TLR4 and LPS [57]. In particular, site 341, which was identified by both the PAML and REL methods, directly participates in the binding of LPS to TLR4 [57]. TLR1 acts in association with TLR2 in the recognition of microbial lipoproteins or lipopeptides [58] and was found to exhibit the second-highest number of positively selected codons. Several codons (289, 291, 316, 324, 326 and 366) (Table 4) under selection are located in LRRs that participate in interactions between TLR1, TLR2, and lipoproteins [58]. Two sites (316 and 326) were shown to directly participate in the binding of ligands to the TLR1-TLR2-lipopeptide complex, and two sites (324 and 366) were located directly at the dimerization interface between the TLR1 and TLR2 proteins [58]. With respect to TLR7 and TLR8, although they are thought to defense against most common viruses [26, 51], the selective pattern was discordant between these TLRs (Table 4). However, no positive selection codon was identified in TLR7, and the identified positively selected sites of TLR8 did not fall within functional ligand binding regions [59], thus suggesting functional constraint of these two genes among carnivores.

Documenting the genetic and selective landscape of TLRs can provide insights into the existing differences in infectious disease susceptibility and pathogenic resistance observed among different species or populations. For example, the signature of positive selection in the human TLR3 555I codon has been functionally characterized as mediating the recognition of the herpes simplex virus type-1 (HSV-1), which is naturally benign to humans, whereas non-human primates exhibit high susceptibility [60, 61]. In contrast, TLR4 D299G has been linked to different levels of susceptibility to bacterial infections as well as a higher prevalence of atopic asthma in Swedish children [62]. We used the MEME method in a wolf lineage to detect specific sites that showed positive selection, which can be missed by codon-based methods [63]. Overall, we detected 11 positively selected sites in 7 wolf TLRs, which may result in a broad spectrum of pathogen recognition in the wolf. Unfortunately, owing to the present insufficient understanding of the differences in susceptibility to infectious diseases between wolves and humans, it was not possible for us to associate these diseases with amino acid variations in wolf TLRs. However, some of the positively selected sites identified in this study were located in regions that participate in the binding of ligands to TLRs (Additional file 9: Table S17) and may affect pathogen recognition and defenses in the wolf.

## Conclusions

Utilizing Illumina sequencing technology and a genome-guided assembly strategy, we assembled and characterized the first peripheral blood transcriptome for the wolf. The wolf blood transcriptome provided a large-scale view of gene content for the characterization of wolf immunological properties and revealed an adaptive mechanism related to a complex pathogen environment. We found that many immune-related genes were expressed in wolf blood, and signal transduction and the immune system were the most represented KEGG functional groups. We also investigated the evolutionary patterns of two pathogen recognition gene families (MHC I and TLR), which might be involved in coevolutionary processes with pathogens. Balancing selection and

Liu *et al. BMC Genomics* (2017) 18:584

Page 10 of 14

recombination together have driven the historical evolution of wolf MHC I genes. Moreover, carnivorous TLRs have undergone adaptive evolution against a background of purifying selection, and a high level of adaptive evolution was detected in wolf TLR repertoire. Finally, we advocate for further study of the evolutionary and selective landscape of immune-related genes of mammals to characterize the genetic mechanisms underlying differences in infectious disease susceptibility and pathogenic resistance observed among humans and wildlife.

## Methods

### Animals and samples collection

We collected blood samples from eight unrelated adult wolves from Tibet and Inner Mongolia in China. The habitat types of the two regions were desert plateau and steppe, respectively. Among the eight wolves, four individuals (one male and three females) originated from Tibet and were raised in Luobulingka Zoo in Tibet, and four individuals (two males and two females) came from Inner Mongolia, two of which were raised in the Dailake National Nature Reserve in Inner Mongolia, and two of which were raised in the Yantai Zoo in Shandong. Blood samples were collected from the leg vein of each wolf after the wolves were anesthetized. The fresh blood samples were immediately stored in RNAprotect® Animal Blood Tubes (QIAGEN, Germany) and frozen in liquid nitrogen until use. All experimental procedures were approved by the Qufu Normal University Institutional Animal Care and Use Committee (Permit Number: QFNU2015–005).

### RNA extraction and transcriptome sequencing

Total RNA was extracted using RNeasy® Protect Animal Blood Kit following the manufacturer's protocol (QIAGEN, Germany). RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). RNA concentration was assessed using Qubit® RNA Assay Kit in the Qubit® 2.0 Fluorometer (Life Technologies, CA, USA).

A total amount of 3 μg total RNA per sample was used as input material for the RNA sample preparations. Eight samples showed quality values with RNA integrity numbers between 8.0 and 9.2 as determined with the Bioanalyzer. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's recommendations. At last, after purified using AMPure XP system, the library products were assessed on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq SR Cluster Kit v3-cBot-HS (Illumia) according to the manufacturer's instructions. After cluster generation, the prepared libraries were sequenced on an Illumina Hiseq 2500 platform and 125 bp paired-end reads were generated.

### Genome-guided de novo transcriptome assembly

Reads from eight wolves were combined to increase the accuracy and completeness of the assembly. Clean reads were obtained after removal of adapters and low-quality reads that might have negatively affected the analysis results using in-house Perl scripts. The retained reads were first aligned to the Ensembl dog reference genome (boxer genome, CanFam3.1) using TopHat v2.0.13 [64]. On the basis of the alignment results, the high-quality reads were used for genome-guided de novo transcriptome assembly using Trinity v2.1.1 [30, 65], with the parameter –genome_guided_bam. In this process, Trinity first portioned reads according to locus and then performed de novo transcriptome assembly at each locus. This strategy has been showen to be a powerful method for generating a high-quality transcriptome and covering genes missing from the genome [66, 67]. After assembly, the longest transcription sequence within the same locus was taken and defined as a unigene. FPKM [68] values were calculated using RSEM software [69] to quantify the unigenes.

### Functional annotation and classification of the blood transcriptome

All of the unigenes were used for BLAST searches with annotation against the Nr database (ftp://ftp.ncbi.nih.gov/blast/db/) using an E-value cut-off of 1E-5. Unigene sequences were also aligned and annotated by BLAST searches to other databases in the following order: Nt (ftp://ftp.ncbi.nih.gov/blast/db/), KOG (ftp://ftp.ncbi.nih.gov/pub/COG/KOG/), Swiss-Prot (http://www.uniprot.org/), KEGG (http://www.genome.jp/kegg/), and GO (http://www.geneontology.org/).

The results of BLAST searching against Nr database were imported into Blast2GO 3.2 [70] for GO term mapping. The outputs of GO mapping were submitted to the WEGO [71] for GO classification under three main domains: BP (Biological Process), CC (Cellular Component), and MF (Molecular Function). KEGG annotation was performed using the KOBAS 2.0 [72] based on BLAST results in KEGG database. Moreover, the unigenes were also aligned to the KOG database to predict and classify possible functions.

### Characterization of MHC I and TLR gene families

Previous studies and BLAST searches against the Nt database in this study have revealed that the 4 MHC I genes [23, 41] and 10 TLR genes are all expressed in canine blood. We used a cloning-sequencing method to obtain the above genes from blood complementary DNA (cDNA) libraries to improve sequence reliability. cDNA was synthesized using a PrimeScript™ Double Strand cDNA Synthesis Kit (TaKaRa, Dalian, China) according to the manufacturer's instructions. For MHC I, the full-length sequences of exons 2–3 of 4 genes were amplified

Liu *et al. BMC Genomics* (2017) 18:584

Page 11 of 14

in 4 individuals with locus-specific primer sets. Full details of these primer sets are provided in Additional file 10: Table S18. The PCR products were purified with a QIAquick Gel Extraction kit (QIAGEN, Germany) and cloned using a pMD™18-T Vector Cloning kit (TaKaRa), according to the manufacturer's instructions. A total of 21 clones for the highly polymorphic DLA-88 and 8–16 clones for the moderately polymorphic DLA-12, 64 and 79 were sequenced per individual to decrease the possibility of false homozygotes being detected as a result of amplification bias. We used the term "allele" for unique exon 2–3 variants of MHC I genes. Additionally, we recognized a clone sequence as an MHC I allele if it was detected in 2 independent PCR amplifications from a single individual or appeared in at least 2 individuals. To amplify the complete CDS of 10 TLRs, we used a series of overlapping primers designed on the basis of corresponding dog sequences in GenBank (http://www.ncbi.nlm.nih.gov/genbank/) and the assembled related unigenes. The PCR primers for the 10 TLRs are presented in Additional file 11: Table S19.

### Analysis of adaptive evolution in the MHC I and TLR gene families

The obtained sequences from the two gene families were examined and assembled using DNASTAR4 (http://www.dnastar.com/). Nucleotide sequence alignment and amino acid translation were performed using MEGA v5.0 [73].

For the MHC I genes, we used the modified Nei–Gojobori method with the Jukes-Cantor correction to compute the average $d_S$ and $d_N$ at all codons, ABSs and non-ABSs. Significant differences in the $d_N/d_S$ ratio were determined with a codon-based Z-test after 1000 bootstrap replicates in MEGA (test statistic: Z). The putative ABS locations of DLA genes were inferred from information of human MHC I genes [74]. Because the ABS positions in 3 human class I genes (HLA, B, C) are slightly different, we conservatively treated the consensus ABSs of the 3 HLA genes as the ABSs of the DLA genes. Moreover, the site model implemented in CODEML from PAML v4.8 [75] was used to perform analyses of selective pressure at individual codons for MHC I genes. Two alternative models were implemented in CODEML, one of which (M7) restricted sites to neutral evolution or evolution under purifying selection ($\omega \leq 1$), whereas M8 allowed a class of sites to evolve under positive selection ($\omega > 1$). A likelihood ratio test (LRT) with 2 degrees of freedom was then applied to compare the two nested models, and the codons under positive selection were identified using the Bayes Empirical Bayes approach (BEB), with a posterior probability higher than 95%. The recombination signal was tested by using seven algorithms (RDP [76], GENECONV [77], Chimaera [78], MaxChi [79], BootScan [80], SiScan [81], and 3Seq [82]) implemented in the Recombination Detection Program version 4 (RDP4) program suite [83] with default settings. Moreover, the

online GARD tool [84] from the Datamonkey web server (http://www.datamonkey.org/) was used for assessing the presence of historical recombination signals.

Other carnivorous TLR sequences used in this study were retrieved from GenBank and Ensembl. For each TLR subset, 6–9 species from some of the most representative families of carnivores (Canidae, Ursidae, Odobenidae, Phocidae, Mustelidae and Felidae) were used. Odobenidae and Phocidae were not included in the TLR5 and TLR9 subsets, owing to the lack of data. Detailed species and sequence accession numbers are given in Additional file 12: Table S20. Selective pressure analyses at individual codons of each TLR were first performed by using the M7 and M8 models in CODEML. All sequences of each TLR were then analyzed through two other ML methods (FEL and REL), which were implemented in the HyPhy package [85] available from the Datamonkey server. For these analyses, we used the automatic model selection tool on the server to determine the best-fitting nucleotide substitution model. Codons with *p* values <0.1 for FEL and a Bayes Factor > 50 for REL were considered candidates for selection. In addition, the single likelihood ancestor counting (SLAC) method was used to obtain the overall $\omega$ for each TLR.

The imprint of natural selection on a functional gene is often difficult to detect because selection is frequently transient and episodic (i.e., it affects only a subset of lineages) [63]. We subsequently searched for evidence of episodic diversifying selection at individual sites along the branches of the trees by using the powerful MEME method on the Datamonkey server. This method allows the $\omega$ value to vary among sites along branches in the tree and can be used to detect positive selection that may occur only in restricted lineages. We primarily focused our attention on positively selected sites detected in the wolf branch, and the default $\alpha = 0.1$ was used as the significance threshold for MEME. Ancestral amino acid sequences were reconstructed via a Bayesian method using ANCESTOR [86], and the amino acid changes at positively selected sites were then determined on the basis of ancestral and extant amino acid sequences. We assessed the functional significance of positively selected sites by mapping these sites to the structural domains of each protein, which were predicted using the software LRRfinder [87] with upper and lower boundaries fixed at 95% and 80%, respectively.

## Additional files

**Additional file 1: Fig. S1.** Transcriptome assembly of wolf blood BLAST statistical analysis. (DOCX 188 kb)

**Additional file 2: Table S1.** GO functional classification of wolf blood transcriptome. (XLSX 29 kb)

**Additional file 3: Table S2.** KEGG functional classification of wolf blood transcriptome. (XLSX 159 kb)

Liu *et al. BMC Genomics* (2017) 18:584

Page 12 of 14

### References
1. Ripple WJ, Estes JA, Beschta RL, Wilmers CC, Ritchie EG, Hebblewhite M, Berger J, Elmhagen B, Letnic M, Nelson MP, et al. Status and ecological effects of the world's largest carnivores. Science. 2014;343(6167):1241484.
2. Knobel DL, Fooks AR, Brookes SM, Randall DA, Williams SD, Argaw K, Shiferaw F, Tallents LA, Laurenson MK. Trapping and vaccination of endangered Ethiopian wolves to control an outbreak of rabies. J Appl Ecol. 2007;45(1):109–16.
3. Gordon CH, Banyard AC, Hussein A, Laurenson MK, Malcolm JR, Marino J, Regassa F, Stewart AM, Fooks AR, Sillero-Zubiri C. Canine distemper in endangered Ethiopian wolves. Emerg Infect Dis. 2015;21(5):824–32.
4. Cadieu E, Neff MW, Quignon P, Walsh K, Chase K, Parker HG, Vonholdt BM, Rhue A, Boyko A, Byers A, et al. Coat variation in the domestic dog is governed by variants in three genes. Science. 2009;326(5949):150–3.
5. Rankin KS, Starkey M, Lunec J, Gerrand CH, Murphy S, Biswas S. Of dogs and men: comparative biology as a tool for the discovery of novel biomarkers and drug development targets in osteosarcoma. Pediatr Blood Cancer. 2012;58(3):327–33.
6. Zhao Y, Sheng Z, Huang J. A systematic analysis of heart transcriptome highlights divergent cardiovascular disease pathways between animal models and humans. Mol BioSyst. 2012;8(2):504–10.
7. Hartenstein V. Blood cells and blood cell development in the animal kingdom. Annu Rev Cell Dev Biol. 2006;22:677–712.
8. Chaussabel D, Pascual V, Banchereau J. Assessing the human immune system through blood transcriptomics. BMC Biol. 2010;8:84.
9. Mohr S, Liew CC. The peripheral-blood transcriptome: new insights into disease and risk assessment. Trends Mol Med. 2007;13(10):422–32.
10. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18(9):1509–17.
11. Abts KC, Ivy JA, DeWoody JA. Immunomics of the koala (Phascolarctos Cinereus). Immunogenetics. 2015;67(5–6):305–21.
12. Du L, Li W, Fan Z, Shen F, Yang M, Wang Z, Jian Z, Hou R, Yue B, Zhang X. First insights into the giant panda (Ailuropoda Melanoleuca) blood transcriptome: a resource for novel gene loci and immunogenetics. Mol Ecol Resour. 2015;15(4):1001–13.
13. Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao F, Kim HL, Burhans RC, Drautz DI, Wittekindt NE, et al. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proc Natl Acad Sci U S A. 2012;109(36):E2382–90.
14. Fornuskova A, Bryja J, Vinkler M, Macholan M, Pialek J. Contrasting patterns of polymorphism and selection in bacterial-sensing toll-like receptor 4 in two house mouse subspecies. Ecol Evol. 2014;4(14):2931–44.
15. Sommer S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool. 2005;2:16.
16. Kelley J, Walter L, Trowsdale J. Comparative genomics of major histocompatibility complexes. Immunogenetics. 2004;56(10):683–95.
17. van den Elsen PJ. Expression regulation of major histocompatibility complex class I and class II encoding genes. Front Immunol. 2011;2:48.
18. Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci. 1997;94(15):7799–806.
19. Yeager M, Hughes AL. Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. Immunol Rev. 1999;167(1):45–58.
20. Galaverni M, Caniglia R, Fabbri E, Lapalombella S, Randi E. MHC variability in an isolated wolf population in Italy. J Heredity. 2013;104(5):601–12.
21. Niskanen AK, Kennedy LJ, Ruokonen M, Kojola I, Lohi H, Isomursu M, Jansson E, Pyhajarvi T, Aspi J. Balancing selection and heterozygote advantage in major histocompatibility complex loci of the bottlenecked Finnish wolf population. Mol Ecol. 2014;23(4):875–89.
22. Arbanasic H, Huber D, Kusak J, Gomercic T, Hrenovic J, Galov A. Extensive polymorphism and evidence of selection pressure on major histocompatibility complex DLA-DRB1, DQA1 and DQB1 class II genes in Croatian grey wolves. Tissue Antigens. 2013;81(1):19–27.
23. Ross P, Buntzman AS, Vincent BG, Grover EN, Gojanovich GS, Collins EJ, Frelinger JA, Hess PR. Allelic diversity at the DLA-88 locus in golden retriever and boxer breeds is limited. Tissue Antigens. 2012;80(2):175–83.
24. Graumann MB, DeRose SA, Ostrander EA, Storb R. Polymorphism analysis of four canine MHC class I genes. Tissue Antigens. 1998;51(4 Pt 1):374–81.
25. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell. 2006;124(4):783–801.
26. Gay NJ, Gangloff M. Structure and function of toll receptors and their ligands. Annu Rev Biochem. 2007;76:141–65.
27. Wlasiuk G, Nachman MW. Adaptation and constraint at toll-like receptors in primates. Mol Biol Evol. 2010;27(9):2172–86.
28. Fornuskova A, Vinkler M, Pages M, Galan M, Jousselin E, Cerqueira F, Morand S, Charbonnel N, Bryja J, Cosson JF. Contrasted evolutionary histories of two toll-like receptors (Tlr4 and Tlr7) in wild rodents (MURINAE). BMC Evol Biol. 2013;13:194.
29. Alcaide M, Edwards SV. Molecular evolution of the toll-like receptor multigene family in birds. Mol Biol Evol. 2011;28(5):1703–15.
30. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. De novo transcript sequence

Liu *et al. BMC Genomics* (2017) 18:584

Page 13 of 14

reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–512.

31. Molfetta R, Quatrini L, Gasparrini F, Zitti B, Santoni A, Paolini R. Regulation of fc receptor endocytic trafficking by ubiquitination. Front Immunol. 2014;5:449.

32. Shaw PJ, Lamkanfi M, Kanneganti TD. NOD-like receptor (NLR) signaling beyond the inflammasome. Eur J Immunol. 2010;40(3):624–7.

33. Bruns AM, Horvath CM. Activation of RIG-I-like receptor signal transduction. Crit Rev Biochem Mol Biol. 2012;47(3):194–206.

34. Schweizer RM, BM vH, Harrigan R, Knowles JC, Musiani M, Coltman D, Novembre J, Wayne RK. genetic subdivision and candidate genes under selection in north American grey wolves. Mol Ecol. 2016;25(1):380–402.

35. Hickey MJ, Kubes P. Intravascular immunity: the host-pathogen encounter in blood vessels. Nat Rev Immunol. 2009;9(5):364–75.

36. Nomiyama H, Osada N, Yoshie O. The evolution of mammalian chemokine genes. Cytokine Growth Factor Rev. 2010;21(4):253–62.

37. Kubo H, Hayashi T, Ago K, Ago M, Kanekura T, Ogata M. Temporal expression of wound healing-related genes in skin burn injury. Legal Med. 2014;16(1):8–13.

38. Ishida Y, Gao JL, Murphy PM. Chemokine receptor CX3CR1 mediates skin wound healing by promoting macrophage and fibroblast accumulation and function. J Immunol. 2007;180(1):569–79.

39. Engelhardt E, Toksoy A, Goebeler M, Bröcker E-B, Gillitzer R. Differential expression of chemokines is correlated with phase-specific infiltration of leukocyte subsets, keratinocyte migration/proliferation and neoangiogenesis during human wound healing. J Dermatol Sci. 1998;16:S177.

40. Busch RH. The wolf almanac. Guilford: The Lyons Press; 1995.

41. Burnett RC, DeRose SA, Wagner JL, Store R. Molecular analysis of six dog leukocyte antigen class I sequences including three complete genes, two truncated genes and one full-length processed gene. Tissue Antigens. 1997; 49(5):484–95.

42. Kennedy LJ, Randall DA, Knobel D, Brown JJ, Fooks AR, Argaw K, Shiferaw F, Ollier WE, Sillero-Zubiri C, Macdonald DW, et al. Major histocompatibility complex diversity in the endangered Ethiopian wolf (Canis Simensis). Tissue Antigens. 2011;77(2):118–25.

43. Bernatchez L, Landry C. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15years? J Evol Biol. 2003;16(3):363–77.

44. Babik W, Kawalko A, Wojcik JM, Radwan J. Low major histocompatibility complex class I (MHC I) variation in the European bison (Bison Bonasus). J Heredity. 2012;103(3):349–59.

45. Kuduk K, Babik W, Bojarska K, Sliwinska EB, Kindberg J, Taberlet P, Swenson JE, Radwan J. Evolution of major histocompatibility complex class I and class II genes in the brown bear. BMC Evol Biol. 2012;12:197.

46. Pokorny I, Sharma R, Goyal SP, Mishra S, Tiedemann R. MHC class I and MHC class II DRB gene variability in wild and captive Bengal tigers (Panthera Tigris Tigris). Immunogenetics. 2010;62(10):667–79.

47. Jakobsen IB, Wilson SR, Easteal S. Patterns of reticulate evolution for the classical class I and II HLA loci. Immunogenetics. 1998;48(5):312–23.

48. Zhao M, Wang Y, Shen H, Li C, Chen C, Luo Z, Wu H. Evolution by selection, recombination, and gene duplication in MHC class I genes of two Rhacophoridae species. BMC Evol Biol. 2013;13(1):113.

49. Sin YW, Dugdale HL, Newman C, Macdonald DW, Burke T. Evolution of MHC class I genes in the European badger (Meles Meles). Ecol Evol. 2012;2(7): 1644–62.

50. Doxiadis GG, van der Wiel MK, Brok HP, de Groot NG, Otting N, t hart BA, van rood JJ, Bontrop RE. reactivation by exon shuffling of a conserved HLA-DR3-like pseudogene segment in a new world primate species. Proc Natl Acad Sci U S A. 2006;103(15):5864–8.

51. Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, Bouchier C, Tichit M, Neyrolles O, Gicquel B, et al. Evolutionary dynamics of human toll-like receptors and their different contributions to host defense. PLoS Genet. 2009;5(7):e1000562.

52. Smith KD, Andersen-Nissen E, Hayashi F, Strobe K, Bergman MA, Barrett SL, Cookson BT, Aderem A. Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. Nat Immunol. 2003;4(12):1247–53.

53. Wlasiuk G, Khan S, Switzer WM, Nachman MW. A history of recurrent positive selection at the toll-like receptor 5 in primates. Mol Biol Evol. 2009; 26(4):937–49.

54. Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, Yen A, Park DJ, Griesemer D, Karlsson EK, Wong SH, et al. Identifying recent adaptations in large-scale genomic data. Cell. 2013;152(4):703–13.

55. Werling D, Jann OC, Offord V, Glass EJ, Coffey TJ. Variation matters: TLR structure and species-specific pathogen recognition. Trends Immunol. 2009; 30(3):124–30.

56. Kumar H, Kawai T, Akira S. Pathogen recognition in the innate immune response. Biochem J. 2009;420(1):1–16.

57. Park BS, Song DH, Kim HM, Choi BS, Lee H, Lee JO. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nat. 2009; 458(7242):1191–5.

58. Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik SG, Lee H, Lee JO. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. Cell. 2007;130(6):1071–82.

59. Wei T, Gong J, Jamitzky F, Heckl WM, Stark RW, Rossle SC. Homology modeling of human toll-like receptors TLR7, 8, and 9 ligand-binding domains. Protein Sci. 2009;18(8):1684–91.

60. Quach H, Wilson D, Laval G, Patin E, Manry J, Guibert J, Barreiro LB, Nerrienet E, Verschoor E, Gessain A, et al. Different selective pressures shape the evolution of toll-like receptors in human and African great ape populations. Hum Mol Genet. 2013;22(23):4829–40.

61. Lefaux B, Duprez R, Tanguy M, Longeart L, Gessain A, Boulanger E. Letter to the editor. Vet Pathol. 2004;41(3):302–4.

62. Fageras Bottcher M, Hmani-Aifa M, Lindstrom A, Jenmalm MC, Mai XM, Nilsson L, Zdolsek HA, Bjorksten B, Soderkvist P, Vaarala O. A TLR4 polymorphism is associated with asthma and reduced lipopolysaccharide-induced interleukin-12(p70) responses in Swedish children. J Allergy Clin Immunol. 2004;114(3):561–7.

63. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8(7):e1002764.

64. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14(4):R36.

65. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7): 644–52.

66. Hassan MA, Melo MB, Haas B, Jensen KD, Saeij JP. De novo reconstruction of the toxoplasma gondii transcriptome improves on the current genome annotation and reveals alternatively spliced transcripts and putative long non-coding RNAs. BMC Genomics. 2012;13:696.

67. Marchant A, Mougel F, Mendonca V, Quartier M, Jacquin-Joly E, da Rosa JA, Petit E, Harry M. Comparing de novo and reference-based transcriptome assembly strategies by applying them to the blood-sucking bug Rhodnius prolixus. Insect Biochem Mol Biol. 2016;69:25–33.

68. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5.

69. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.

70. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36(10):3420–35.

71. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, et al. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 2006;34(Web Server issue):W293–7.

72. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39(Web Server issue):W316–22.

73. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731–9.

74. Reche PA, Reinherz EL. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. J Mol Biol. 2003;331(3):623–41.

75. Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.

76. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. Bioinformatics. 2000;16(6):562–3.

77. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. Virol. 1999;265(2):218–25.

Liu *et al. BMC Genomics* (2017) 18:584

Page 14 of 14

78. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc Natl Acad Sci U S A. 2001; 98(24):13757–62.

79. Smith JM. Analyzing the mosaic structure of genes. J Mol Evol. 1992;34(2): 126–9.

80. Martin DP, Posada D, Crandall KA, Williamson C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retrovir. 2005;21(1):98–102.

81. Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics. 2000;16(7):573–82.

82. Boni MF, Posada D, Feldman MW. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics. 2007;176(2):1035–47.

83. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 2015;1(1): vev003-vev003.

84. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol. 2006;23(10):1891–901.

85. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005;21(5):676–9.

86. Zhang J, Nei M. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J Mol Evol. 1997;44(Suppl 1): S139–46.

87. Offord V, Coffey TJ, Werling D. LRRfinder: a web application for the identification of leucine-rich repeats and an integrative toll-like receptor database. Dev Comp Immunol. 2010;34(10):1035–41.

# Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;

2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;

3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;

4. use bots or other automated methods to access the content or redirect messages

5. override any security feature or exclusionary protocol; or

6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com