A Preliminary Analysis of the Immunoglobulin Genes in the African Elephant (Loxodonta africana)

Yongchen Guo¹, Yonghua Bao²*, Hui Wang³, Xiaohai Hu¹, Zhihui Zhao³, Ning Li¹, Yaofeng Zhao¹,4*

¹ State Key Laboratory of AgroBiotechnology, College of Biological Sciences, China Agricultural University, Beijing, People’s Republic of China, 2 Department of Basic Immunology, Xinxiang Medical University, Xinxiang, People’s Republic of China, 3 Agricultural Division, College of Animal Science and Veterinary Medicine, Jilin University, Changchun, People’s Republic of China, 4 College of Animal Science and Technology, Qingdao Agricultural University, Qingdao, People’s Republic of China

Abstract

The genomic organization of the IgH (Immunoglobulin heavy chain), Igκ (Immunoglobulin kappa chain), and Igλ (Immunoglobulin lambda chain) loci in the African elephant (Loxodonta africana) was annotated using available genome data. The elephant IgH locus on scaffold 57 spans over 2,974 kb, and consists of at least 112 V_h gene segments, 87 D_h gene segments, six J_h gene segments, a single μ_h, a δ remnant, and eight γ genes (γ_a and γ_e genes are missing, most likely due to sequence gaps). The Igκ locus, found on three scaffolds (202, 50 and 86), contains a total of 153 V_κ gene segments, three J_κ segments, and a single C_κ gene. Two different transcriptional orientations were determined for these V_κ gene segments. In contrast, the Igλ locus on scaffold 68 includes 15 V_λ gene segments, all with the same transcriptional polarity as the downstream J_λ-C_λ cluster. These data suggest that the elephant immunoglobulin gene repertoire is highly diverse and complex. Our results provide insights into the immunoglobulin genes in a placental mammal that is evolutionarily distant from humans, mice, and domestic animals.

Introduction

The elephant is the biggest terrestrial placental mammal alive today. It belongs to the order Proboscidea and the family elephantidae, which contains only two existing species: the Asian elephant (Elephas maximus) and the African elephant (Loxodonta africana). The three lineages of this family: Loxodonta, Elephas, and Mammutthus are thought to have originated 4–6 million years ago. Whereas some species of the former two lineages are still alive today, the last representative of the Mammutthus lineage, the woolly mammoth (Mammutthus primigenius), became extinct very recently (about 3.7 thousand years ago) [1]. Phylogenetic analysis suggest that the elephant is most closely related to living mammals of Trichechus (such as the West Indian Manatee, Trichechus manatus) and Procavia (such as the Rock Hyrax, Procavia capensis) [2].

Elephants are reported to be susceptible to a wide variety of infections caused by bacteria [3,4], viruses [5–13], and parasites [14–17]. However, there have been very few studies previously performed on the elephant immune system. In addition, little is known about the elephant immunoglobulins, except for serological testing for IgM [18], IgG [19,20], and IgA [21]. It was reported that there were at least five subclasses of IgG in African elephant sera, with no apparent IgM or IgA [20].

Immunoglobulins are the antigen-recognition molecules of B cells of jawed vertebrates, which usually consist of two identical heavy (H) and two identical light (L) chains. In some exceptional cases, such as shark IgNAR and selected subclasses of camelid IgGs, only heavy chains are used [22–24]. Variable regions in the N-terminus of H/L chains are encoded by V_H/V_L, D_H, and J_H/J_L genes to determine the antigen binding site and antibody specificity. However, constant regions in the C-terminus of H/L chains are encoded by IGHC/C_κ or C_λ genes and are responsible for the immunoglobulin classes and functional activities [25,26].

In the mammals studied so far, the locus of unique immunoglobulin heavy chain genes and loci of λ and κ light chain genes are commonly organized in a “translocon” pattern [27,28]. In the heavy chain locus, multiple V_H, D_H, and J_H gene segments are followed by consecutive μ, δ, γ, ε, and α gene segments [29]. In the λ light chain locus, a cluster of V_λ gene segments is followed by multiple sets of clustered J_λ gene segments, each linked to a single C_λ gene. Differentially, the cluster of V_κ gene segments is followed by a cluster of J_κ gene segments, and then by a single C_κ gene [30].

IgH and IgL loci have been characterized in different mammalian species [31–48]. Although the genomic organization of immunoglobulin genes in mammals has remained relatively constant, variation exists in the number of variable, diversity, joining, and constant region genes. Here, we present the genomic organization of the IgH, Igκ, and Igλ loci of the African elephant, annotated on a basis of its genome data.

Materials and Methods

The elephant genome sequence

The genome sequence of the African Elephant (Loxodonta africana), provided by the Broad Institute via whole genome
shotgun, can be obtained from the Ensembl database (http://www.ensembl.org). LoxAf3, an assembly of the genome of African Elephant, has been sequenced to 7× coverage (loxAf3, 7× coverage, July 2009). The elephant immunoglobulin gene sequences were retrieved from the UCSC genome browser (http://genome.ucsc.edu/).

Identification of the elephant Ig genes

Human immunoglobulin gene sequences were used as queries to search the elephant genome scaffolds that contained immunoglobulin genes. A conventional TBLASTN approach was used to identify constant region genes of the elephant immunoglobulins. FUZZNUG, an online software (http://embossgui.sourceforge.net/demo/fuzznuc.html) was used to find adjacent recombination signal sequences (RSSs) for identification of variable, diversity, and joining gene segments. Five or more mismatched bases were allowed to cover all genes. The locations of the annotated elephant gene sequences on the elephant genome are shown in Table S1 (S1-1~S1-5).

Sequence alignments

Editing and comparison of sequences were carried out using the DNastar program. Alignment of multiple sequences was performed using the Clustal W algorithm, then aligned with Clustal X software, and exported by BioEdit software with view conservation by plotting identities to a standard as a dot.

Dot matrix analysis

A dot matrix analysis (window size 30 bp and mismatch limit 9 bp) was used for comparing two sequences to identify a possible alignment of characters between the sequences.

Phylogenetic analysis

Phylogenetic studies were carried out using MrBayes3.1 and viewed with the TreeView package. All the trees were obtained with 1 million generations for the chains, a sample frequency of 100, and a burn in of 2,500 (ngen = 1000000; SampleFreq = 100; burnin = 2,500). The site by site rate variation was set to a gamma distribution (rates = gamma) for all the Bayesian trees and a General Time-Reversible (GTR) (nst = 6) model of substitution was chosen. The sequences from other species used in phylogenetic analyses are presented in Table S2 (S2-1~S2-2).

Definition of the VH/VL gene families

In mammals, germline VH and VL gene segments can be grouped into families based on their nucleotide sequence similarity [49]. The established criteria are that the same family members share more than 80% nucleotide similarity, those with less than 70% similarity are put into different families, and those possessing between 70% and 80% similarity are inspected on a case-by-case basis [50]. In our analysis, we placed VH and VL segments having similarity greater than 70% into the same family.

Results

Elephant immunoglobulin heavy chain genes

IgH locus. The public elephant genome assembly used in this study was loxAf3, which is an assembly of the genome of the African Elephant (Loxodonta africana), sequenced to 7× coverage. The high genome coverage of this assembly confers a high reliability on the gene analysis. BLAST searching localized the elephant IgH locus to genomic scaffold 57. It spans approximately 2,974 kb from the most 5’ VH segment (VH2-112p) to the most 5’ γ gene (Fig. 1). A single μ and eight γ genes were identified in this scaffold. Neither ε nor α genes could be found, most likely due to sequence gaps.

Constant region genes. Like other mammalian species, the elephant μ gene contains four CH and two transmembrane exons. A sequence comparison of μ genes among thirteen vertebrate species demonstrated that the critical amino acids for immunoglobulin folding, Cysteine (C) and Triptophan (W) [51], were highly conserved in elephants (Figure S1). In addition, the elephant IgM constant region showed the highest amino acid sequence identity to human (63.8%), and the least to echidna (50.8%).

Most mammals also express a δ gene, which is always situated immediately downstream of the μ, and the distance between μ and δ usually does not exceed 7 kb. A BLAST search against the elephant whole genome using both DNA and amino acid sequences of the δ genes of other mammalian species showed no intact δ gene. However, approximately 10 kb downstream of the elephant μ (no sequence gaps for 90 kb downstream), we identified a short fragment encoding a polypeptide (Figure S2) homologous to the IgD CH3 domain of other mammals. This was done by a thorough examination of amino acid sequences encoded by the DNA sequences between μ and γ1 (based on translation of all reading frames of both sense and anti-sense sequences). An alignment of the elephant IgD remnant and the IgD CH3 domains of several mammalian species is presented in Figure S2. This indicates that the gene has been highly mutated and pseudogenized in the elephant.

In addition to the eight γ genes (γ1 to γ8) in scaffold 57 (Fig. 1), an additional γ gene (tentatively named as γ9) was identified in scaffold 495 (data not shown), which spans 77 kb. Scaffold 495 is not assembled together with scaffold 57; therefore, γ9 could potentially be either an additional subclass encoding gene or an allelic variant. The identification of multiple IgG subclass-encoding genes is in accordance with a previous report, which indicated that there were at least five subclasses of IgG in African elephant sera [20]. Sequence analysis showed no additional Ig genes in genomic scaffold 495, except for the γ9 gene. The greatest variation among mammalian IgG subclasses is usually concentrated in their hinge regions [52–54]. However, no elephant IgG cDNA sequences have been sequenced, it is very difficult to accurately assess the hinge regions of the elephant IgG heavy chains. The hinge region is usually encoded on a separate exon that could not be identified in the elephant due to the low level of conservation and the absence of cDNA sequences. An amino acid alignment of the nine elephant IgG subclasses is presented in Fig. 2. The first exons (CH1) of γ1 and γ2 are missing because of gaps. The CH3 exon of γ3 is pseudogenized because of a premature stop codon (marked with a star in Fig. 2), and a frameshift mutation (marked with shadowing in Fig. 2) caused by nucleotide (adenine) insertions at positions 148 and 158, respectively. To clarify the relationship among γ chains from mammalian species, a phylogenetic tree of IgG CH2 and CH3 exons was constructed and is shown in Figure S3. The elephant γ genes form a distinct cluster. This is consistent with previous analysis, which showed that the divergence of IgG subclasses occurred after speciation [52].

Dot matrix analysis of the elephant IgH locus showed there are switch regions upstream of the μ gene and six γ genes (γ1, γ4, γ5, γ6, γ7, and γ8), as in humans and mice [55,56]. The switch regions of γ2, γ3, and γ9 could not be identified, most likely due to sequence gaps. Structurally, the switch regions, as in other species, are all composed of pentameric repeats (GGGCT) and GAGCT). The elephant Sp region shows substantial nucleotide similarity.
The six elephant S regions are similar, but share little sequence similarity with human and mouse S (Fig. 4 and data not shown).

VH gene segments. A total of 112 VH segments were identified in the elephant IgH locus. 51 of them appear to be potentially functional, because they have leader exons, normal open reading frames (ORF), downstream RSSs, and V gene domain (framework regions (FRs) and complementarity determining regions (CDRs)). The remaining 61 segments contain either in-frame stop codons or frameshifts, and are thus designated as pseudogenes. In addition, there are 17 partial segments of about 200 bp in length, which are regarded as truncated VH sequences. There are gaps above 10 kb in the elephant genome among the VH gene segments (Fig. 1), suggesting that there might be more VH segments. To examine the relationships among the elephant germline VH segments, pseudogenes as well as functional genes were used to construct a phylogenetic tree (Fig. 5). The seven identified VH gene families (1, 2, 3, 4, 5, 6, and 7) were confirmed to be homologous with the corresponding human VH gene families. The elephant VH family contains the most members (72 VH segments), which could be further divided into three groups (Fig. 5). We chose representative VH sequences from elephant and other mammals, covering almost all VH families identified, to construct phylogenetic trees (Fig. 6).

The elephant VH genes clearly fall into the three previously known VH clans.

DH gene segments. In the elephant IgH locus, 87 DH segments were identified and are presented in Figure S4 (S4-1 to S4-10). It should be noted that there might be more DH segments because of the existence of sequence gaps. Except for DH76, which has a 10 bp spacer, all the DH segments are flanked by characteristic heptamers and nonamers separated by 12-bp spacers. The potential coding regions of DH segments are 10–37 bp in length (Figure S4, S4-1 to S4-10). It has been suggested that coding regions of DH segments of humans can be described by the characteristics of their amino acids [57]. Inspection showed that a great number of polar/hydrophobic amino acids or stop codons occur widely in elephant DH coding regions (data not shown). In humans and mice, the germline DH segments can be classified into families based on the extent of sequence similarity [58,59]. Analysis of nucleotide similarity in the coding regions and flanking RSSs indicated that the 87 elephant DH segments could be divided into seven families. Members within the same family share at least 70% nucleotide identity (data not shown), while some members in a family have completely identical sequences (these are shadowed in Fig. 7). We present the sequence alignment of the seven families in
Fig. 7, which shows that each family contains characteristic sequence intervals that are distinct from other families.

**JH gene segments.** There were six germline JH gene segments found in the elephant IgH locus (Fig. 8). All the JH segments had conserved nucleotide sequences at the 3' end. JH1 was pseudogenized by replacement of a Tryptophan (W) residue by a stop codon.

Elephant immunoglobulin light chains

**κ chain.** Immunoglobulin κ chain genes of elephant were identified on three scaffolds: 202, 50, and 86. A schematic diagram is shown in Fig. 9. Of the 153 germline Vκ segments from the three scaffolds, 53 were regarded as potentially functional genes and 100 as pseudogenes. Based on sequence similarity analysis, 142 of the Vκ segments can be assigned to eight families (Vκ1~Vκ8) (Table...
S3), which contain 2, 31, 2, 102, 1, 1, 2, and 1 members, respectively. The remaining 11 \(V_k\) pseudogenes could not be assigned to any family because they share less than 70% nucleotide similarity with any other \(V_k\) gene segment. A phylogenetic tree of the elephant \(V_k\) functional genes is shown in Fig. 10. The six elephant \(V_k\) families (\(V_k1\sim V_k6\)) correspond to the six human \(V_k\) gene families. In addition, scaffold 86 includes 24 \(V_k\) segments showing the same transcriptional orientation as the \(J_k\) and \(C_k\), and 18 \(V_k\) segments showing a reverse transcriptional direction. Three \(J_k\) segments and one \(C_k\) gene on scaffold 86 are displayed in Figure S5. In addition, \(V_k\) segments located on scaffolds 202 and 50 also possess two different transcriptional directions.

\(\lambda\) chain. Scaffold 68 was determined to contain the elephant \(\lambda\) light gene complex (Fig. 9). Sequences analysis revealed that the 12 elephant \(V_\lambda\) gene segments belonged to six families (Fig. 11), which were homologous with the human \(V_\lambda\) 1, 3, 4, 7, 9 and 10 families. The remaining three \(V_\lambda\) pseudogenes could not be assigned to any family because they share less than 70% nucleotide similarity with any other \(V_\lambda\) gene segment. The three elephant \(V_\lambda\) families consists of seven members. In contrast to \(V_k\), all the \(V_\lambda\) segments possess an identical transcriptional polarity to the downstream \(J_\lambda\) segments. In addition, only \(V_\lambda3-3\) and \(V_\lambda3-7\) are identified as potentially functional genes. At the 3’ end of the locus, three constant region genes are organized in tandem, where both \(C_\lambda2\) and \(C_\lambda3\) are preceded by a \(J_\lambda\). The \(J_\lambda\) segment before \(C_\lambda1\) is missing because of a sequence gap. Three \(C_\lambda\) genes show approximately 90% amino acid identity. The sequences of two \(J_\lambda\) segments and three \(C_\lambda\) genes are presented in Figure S5.

Discussion

In this study, we have made a preliminary analysis of the immunoglobulin genes in the elephant using the recently released elephant genome, revealing that the elephant IgH locus conforms
Figure 5. Phylogenetic analysis of the 112 elephant VH genes. A phylogenetic tree of nucleotide sequences of 112 elephant VH segments was constructed. The seven identified VH gene families are labeled with Arabic numerals. The credibility value for each node is shown. doi:10.1371/journal.pone.0016889.g005
to the “translocon” pattern. Compared with human IgH locus, which occupies a 1.25 Mb region [60], elephant IgH locus appears to span larger genomic region (approximately 3 Mb).

We translated the nucleotide sequences between the m and γ1 genes in all three reading frames in both the positive and negative directions. By blasting the nucleotide and corresponding amino acid sequences against the NCBI database, only the IgD-CH3 remnant was identified.

With the exception of marsupials [61,62], most placental and even monotreme mammals studied so far have been shown to have multiple IgG subclasses encoded by independent sets of exons [63]. The elephant genome contains nine IgG genes, although it is not known whether all of them are functional. This number is larger than that in any placental mammals so far examined (ranging from 1 to 7) [43,64–69], providing another remarkable example for IgH chain constant region diversity in mammals.

Our analysis also suggested a high degree of complexity in the elephant IgVH locus. At least 112 V_H segments constitute the elephant germ-line V_H repertoire. According to the number of V_H gene families, placentals studied so far could be divided into two groups. The multiple gene families group includes mice (16 families), human (seven families), and horse (seven families). The few gene families or single gene family group includes dog (three families), rabbits (one family), cattle (one family), camel (one family), and swine (one family) [70–78]. The elephant, having 7 V_H gene families, should be put into the first group. The

---

Figure 6. Phylogenetic analysis of mammalian V_H genes. Representatives of the seven elephant V_H families are clustered with their human counterparts. The three mammalian V_H clans are labeled with Roman numerals. The credibility value for each node is shown. doi:10.1371/journal.pone.0016889.g006
### Figure 7. Alignment of nucleotide sequences of seven elephant germline D_{λ} families.

Seven families representing elephant 87 germline D_{λ} segments are aligned. Nucleotides that are the same as the top segment, D_{λ}34, are indicated with dots. Dashes mean gaps introduced to make the alignment. D_{λ}57 and 62, D_{λ}50 and 38, D_{λ}74 and 66, and D_{λ}47 and 36 are shadowed as they share identical sequences. Coding regions of D_{λ} segments are separated from recombination signal sequences (RSSs) (nonamer, spacer, and heptamer).

| Family 1 | Nonamer | Spacer | Heptamer | Coding region of D_{λ} segment | Heptamer | Spacer | Nonamer |
|----------|---------|--------|----------|-------------------------------|---------|--------|---------|
| D_{λ}34  | GAGTTGTAT | AGAGCACATGAT | CCGTGG | ACA-ATATAAAGAAAGTATCTCTGATTAGTAC | CACAGTG | ACACACCATGTC | CCCAARRAC |
| D_{λ}35  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}53  | .G. .G. | AG. .A. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}55  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}68  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}70  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}62  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}64  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}76  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}86  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |

### Family 2

| D_{λ}35  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}38  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}44  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}55  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |

### Family 3

| D_{λ}35  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}38  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |

### Family 4

| D_{λ}35  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}38  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |

### Family 5

| D_{λ}35  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}38  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |

### Family 6

| D_{λ}35  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}38  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |

### Family 7

| D_{λ}35  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |
| D_{λ}38  | .G. .G. | AG. CA. | . . . | . . . | . . . | . . . | . . . |

---

**Figure 7.** Alignment of nucleotide sequences of seven elephant germline D_{λ} families. Seven families representing elephant 87 germline D_{λ} segments are aligned. Nucleotides that are the same as the top segment, D_{λ}34, are indicated with dots. Dashes mean gaps introduced to make the alignment. D_{λ}57 and 62, D_{λ}50 and 38, D_{λ}74 and 66, and D_{λ}47 and 36 are shadowed as they share identical sequences. Coding regions of D_{λ} segments are separated from recombination signal sequences (RSSs) (nonamer, spacer, and heptamer). doi:10.1371/journal.pone.0016889.g007
Figure 8. The six elephant germline $J_\alpha$ gene segments. Nucleotide and deduced amino acid sequences of six $J_\alpha$ segments, along with RSSs, are shown. The amino acid residue W is replaced by a stop codon in the $J_\alpha1$ segment. doi:10.1371/journal.pone.0016889.g008

Figure 9. The elephant IgL locus. The elephant IgL locus is distributed over three scaffolds (202, 50, and 86), and the IgL locus is located on scaffold 68. Overall configurations are drawn approximately to scale. The potentially functional $V_\kappa$ and $V_\lambda$ genes are shown as filled bars, while pseudogenes are represented by open bars and indicated with the letter p. Double slashes indicate gaps >10 kb. The unidirectional arrowheads below $V_\kappa$ gene segments on scaffold 86 indicate that their transcriptional direction is opposite to downstream $J_\kappa$ segments. However, the unidirectional arrowheads on scaffolds 202 and 50 do not represent different transcriptional directions from the identified $J_\kappa$ gene segment; they merely indicate a transcriptional direction different from that of the remaining $V_\kappa$ gene segments in the scaffold. doi:10.1371/journal.pone.0016889.g009
mammalian V_H families can be further classified into three clans: I, II, and III, which have co-existed in the genome for more than 400 Myr [79]. Similar to those of humans, the elephant V_H families also conform to three clans: families 1, 5, and 7 form clan I, families 2, 4, and 6 form clan II, and family 3 forms clan III. The largest group of elephant V_H genes is the V_H^4 family of clan II. It has been demonstrated that the unique V_H family identified in cattle belonged to clan II [75,77]. In sheep, most V_H genes are also categorized into clan II [80]. Based on a recent report, clan II also appeared to be the largest group in the horse [41], indicating that the herbivore animals may prefer to use the clan II V_H genes.

Close attention should also be paid to the elephant D_H locus, where at least 87 germline D segments could be mapped to a 450-kb DNA region; the largest number in mammals examined so far. The presence of more D_H segments may greatly increase the Ig diversity generated through DNA rearrangement. The size of the elephant D_H coding regions ranges from 10 to 37 bp, similar to that of human (11 to 37 bp) [57]. Further inspection revealed that the elephant D_H segments were translated in three reading frames abundant in polar/hydrophobic amino acids, which is different to dog [78], horse [41], mouse [81], rabbit [82], and chicken [83], which show preferences for neutral [polar/hydrophilic] amino acids.

For the light chain genes, elephant V_K germline genes are more abundant than V_L (53 functional V_K genes vs. 2 functional V_L genes). Different mammalian species possess different ratios of V_K and V_L. In humans, roughly 60% of the variable light chain
repertoire is \( \kappa \) (40 functional \( \kappa \) genes vs. 30 functional \( \lambda \) genes). The germline \( \kappa \) genes of mice are dominant by as much as 95% or more [84]. It has been proposed that the preferential use of light chain isotypes at the protein level may be correlated with the overall number of V gene segments [84]. It is thus possible that the \( \kappa \) chain predominates over the \( \lambda \) chain at the protein level in elephants.

Interestingly, a great number of pseudogenes exist in the elephant \( V_H \) (61/112), \( V_{\kappa} \) (100/155), and \( V_\lambda \) (13/15) loci. In some species, the base-pair changes could be inferred using an existing pseudogene or germline gene as a template, and therefore pseudogenes in the V loci constitute a potential donor pool for gene conversion to generate immunoglobulin diversity [85–88]. A great number of V pseudogenes may contribute to the immunoglobulin diversity in elephants.

The study of structure and organization of the immunoglobulin gene loci is vital to the understanding of the nature of antibody molecules. This study provides information for comparative studies of mammalian Ig genes, as well as data for further studies of the elephant immunoglobulin genes.

---

Figure 11. Phylogenetic analysis of the 12 elephant \( V_\lambda \) genes. A phylogenetic tree of the nucleotide sequences of the 12 elephant \( V_\lambda \) segments was constructed. The 12 elephant \( V_\lambda \) gene segments belong to six families, which are clustered with the human \( V_\lambda \) 1, 3, 4, 7, 9, and 10 families, respectively. The credibility value for each node is shown.

doi:10.1371/journal.pone.0016889.g011
Supporting Information

Figure S1  Alignment of IgM amino acid sequences from several vertebrate species.  Elephant IgM was compared with a panel of vertebrate IgM sequences. Dots indicate similar residues as in elephant µ, whereas dashes indicate gaps introduced for optimal alignment. The cysteine residues C and W important for intra-domain disulfide bonds are shown on the first line of the alignment. (TIF)

Figure S2  Alignment of the elephant IgD remnant with the IgD CH3 domains of several mammalian species. Amino acid residues that are identical to the top counterpart in every panel are shown as dots; Gaps and missing data are indicated by hyphens. Stop codons are indicated by stars. (TIF)

Figure S3  Phylogenetic tree of the immunoglobulin gamma heavy chains of some mammalian species. The phylogenetic tree was constructed from the amino acid sequences of the CH3 exons of the immunoglobulin gamma heavy chains of various mammalian species. The credibility value for each node is shown. (TIF)

Figure S4  Alignment of the deduced amino acid sequences of the three elephant Jk genes in scaffolds 202, 50, and 86. (TIF)

Table S1  The eight elephant Vk gene families from scaffolds 202, 50, and 86. (TIF)

Table S2  GenBank accession numbers or references of the gene sequences from other species used in this paper. (RAR)

Table S3  Analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence in sharks. Comp Biochem Physiol B Biochem Mol Biol 112: 569–572. (RAR)

Author Contributions
Conceived and designed the experiments: YG YB YZ. Performed the experiments: YG YB HW XZ NL YZ. Analyzed the data: YG YB YZ. Contributed reagents/materials/analysis tools: HW XZ NL. Wrote the paper: YG YB YZ.

References
1. Thomas MG, Hagelberg E, Jones HB, Yang Z, Lister AM (2000) Molecular and morphological evidence on the phylogeny of the Elephantidae. Proc Biol Sci 267: 2493–2500.
2. Murphy WJ, Eazizik E, Johnson WE, Zhang YP, Ryder OA, et al. (2001) Molecular phylogenetics and the origins of placental mammals. Nature 409: 614–618.
3. Mikota SK, Peddie L, Peddie J, Izaza R, Dunker F, et al. (2001) Epidemiology and diagnosis of Mycobacterium tuberculosis in captive Asian elephants (Elephas maximus). J Zoo Wildl Med 32: 1–16.
4. Lyashchenko KP, Greenwald R, Esfandiari J, Olsen JH, Ball R, et al. (2006) Tuberculosis in elephants: antibodies to defined antigens of Mycobacterium tuberculosis, potential for early diagnosis, and monitoring of treatment. Clin Vaccine Immunol 13: 722–732.
5. Howell PG, Young E, Hedger RS (1973) Foot-and-mouth disease in the african elephant (Loxodonta africana). Onderstepoort J Vet Res 40: 41–52.
6. Davies FG, Otieno S (1977) Elephants and zebras as possible reservoir hosts for African horse sickness virus. Vet Rec 100: 291–292.
7. Metzler AE, Osent P, Giustetti F, Rubel A, Lang EM (1990) Serological evidence of herpesvirus infection in captive Asian elephants (Elephas maximus). J Wildl Dis 26: 41–49.
8. Turnbull PC, Bell RH, Saigawa K, Munyenyembe FE, Mulenga CK, et al. (1998) Anthrax in wildlife in the Luangwa Valley, Zambia. Vet Rec 128: 399–403.
9. Inepral VS, Warun BN, Davies FG, Sei R, Olubayo R (1992) An attempt to define the host range for African horse sickness virus ( Orbivirus, Reoviridae) in east Africa, by a serological survey in some Equidae, Camelidae, Loxodontaefidae and Carnivore. Vet Microbiol 30: 19–23.
10. Formenty P, Domenech J, Lanuque J, Osautara M, Diauara S, et al. (1994) [Epidemiologic study of bluetongue in sheep, cattle and different species of wild animals in the Ivory Coast]. Rev Sci Tech 13: 737–753.
11. Bhat MN, Manickam R, Kumanan K (1997) Serological evidence of bovine herpesviruses 1 and 2 in Asian elephants. J Wildl Dis 33: 919–920.
12. Evans A, Gakaya F, Pavaska JT, Rostal M, Akoole L, et al. (2008) Prevalence of antibodies against Rift Valley fever virus in Kenyan wildlife. Epidemiol Infect 136: 1261–1269.
13. Oni O, Wajuwaka W, Bodele O, Chunsing W (2006) Canine distemper virus antibodies in the Asian elephant (Elephas maximus). Vet Rec 159: 420–421.
14. Howe T, Dubey JP (1999) Prevalence of Toxoplasma gondii antibodies in sera of domestic pigs and some wild game species from Zimbabwe. J Parasitol 85: 372–373.
15. Riemann GP, Burridge MJ, Behymur DE, Franti CE (1975) Toxoplasma gondii antibodies in free-living African mammals. J Wildl Dis 11: 529–533.
16. Tantusavan D, Muikkaew K, Dubey JP (2001) Seroprevalence of Toxoplasma gondii in elephants (Elephas maximus indicus) in Thailand. J Parasitol 87: 229–230.
17. Dangolla A, Ekamayake DK, Rajapakse RP, Dubey JP, Silva ID (2006) Seroprevalence of Toxoplasma gondii antibodies in captive elephants (Elephas maximus maximus) in Sri Lanka. Vet Parasitol 137: 172–174.
18. King DP, Sanders JL, Nomura CT, Stoddard RA, Ortiz CL, et al. (1998) Ontogeny of humoral immunity in northern elephant seal (Mirounga angustirostris) neonates. Comp Biochem Physiol B Biochem Mol Biol 121: 363–368.
19. Kelly PJ, Tappeira M, Matthewman L, Mason PR, Wright EP (1993) Reactions of sera from laboratory, domestic and wild animals in Africa with protein A and a recombinant chimeric protein AG. Comp Immunol Microbiol Infect Dis 16: 299–305.
20. Kelly PJ, Carter SD, Arzwi SM, Cadman HF (1998) Isolation and characterisation of immunoglobulin g and IgG subclasses of the African elephant (Loxodonta africana). Comp Immunol Microbiol Infect Dis 21: 63–73.
21. Marquez ME, Slobodianik NH, Rownay de Ferrer PA, Carlini AR, Vergani DF, et al. (1995) Immunoglobulin A levels in southern elephant seal (Mirounga leonina) milk during the suckling period. Comp Biochem Physiol B Biochem Mol Biol 112: 569–572.
22. Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, et al. (1993) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374: 168–173.
23. Hansen-Casterman C, Aarhus T, Muyllemans S, Robinson G, Hamers G, et al. (1993) Naturally occurring antibodies devoid of light chains. Nature 363: 446–448.
24. Roux KH, Greenberg AS, Greene L, Streltes L, Avila D, et al. (1998) Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. Proc Natl Acad Sci 95.
25. Edelman GM, Gall WE (1969) The antibody problem. Annu Rev Biochem 38: 415–466.
26. Porter RR (1967) The structure of the heavy chain of immunoglobulin and its relevance to the nature of the antibody-combining site. The Second CIBA Medal Lecture. Biochem J 105: 417–426.

27. Honjo T, Frederick A, Neuberger M, eds (2004) Molecular Biology of B Cells: immunocytes. Academic Press. 600 p.

28. Zhang YA, Nie F, Luo HY, Wang YP, Sun YH, et al. (2003) Characterization of cDNA encoding immunoglobulin light chain of the mandarin fish (Siniperca chuatsi). Vet Immunol Immunopathol 95: 31–41.

29. Thomas RK, Re D, Wolf J, Doll V (2004) Part I: Hodgkin’s lymphoma—molecular biology of Hodgkin and Reed-Sternberg cells. Lancet Oncol 5: 11–18.

30. Qin T, Ren L, Hu X, Guo Y, Fei J, et al. (2008) Genomic organization of the immunoglobulin light chain gene loci in Xenopus tropicalis: evolutionary implications. Dev Comp Immunol 32: 156–163.

31. Gambon-Deza F, Sanchez-Espinol C, Magadan-Mompo S (2009) The immunoglobulin heavy chain locus in the platypus (Ornithorhynchus anatinus). Mol Immunol 46: 2315–2323.

32. Choudhary CL, Hering CD, Rihlet R (2002) A three-megabase yeast artificial chromosome contig spanning the C5BL mouse Ig locus. J Immunol 168: 5650–5666.

33. Fripaip JP, Williams SC, Tomlinson IM, Cook GP, Gheri D, et al. (1993) Organization of the human immunoglobulin lambda light chain locus on chromosome 22q11.2. Hum Mol Genet 4: 983–991.

34. Hendricks J, Terpstra P, Dammers PM, Somasundaram R, Visser A, et al. (2003) Characterization of the variable region of the immunoglobulin heavy-chain gene loci of the rat. Immunogenetics 56: 473–486.

35. Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, et al. (2001) Evolutionary dynamics of the human immunoglobulin kappa locus and the germline repertoire of the Vkappa genes. Eur J Immunol 31: 1017–1028.

36. Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, et al. (1997) One-megabase sequence analysis of the human immunoglobulin lambda gene locus. Genome Res 7: 250–261.

37. Kirschbaum T, Jaenichen R, Zachau HG (1996) The mouse immunoglobulin kappa locus contains about 140 variable gene segments. Eur J Immunol 26: 1613–1620.

38. Kirschbaum T, Pourrajabi S, Zochau HG (1997) Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. Int Immunol 9: 1009–1020.

39. Mage RG, Lanning D, Knight KL (2006) B cell and antibody repertoire evolution in rabbits: the requirement of gut-associated lymphoid tissues. Dev Comp Immunol 30: 137–154.

40. Peich M, Sonola H, Polheinz HD, Straubinger B, Gerl R, et al. (1985) A large section of the gene encoding human immunoglobulin variable regions of the kappa type is duplicated. J Mol Biol 183: 291–299.

41. Sun Y, Wang C, Wang Y, Zhang T, Ren L, et al. (2010) A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. Dev Comp Immunol 34: 1009–1020.

42. Tang L, Sampson C, Dreitz MJ, McCall G (2001) Cloning and characterization of cDNAs encoding four different canine immunoglobulin gamma chains. Vet Immunol Immunopathol 80: 243–250.

43. Wagner B, Müller DG, Lean TL, Antczak DF (2004) The complete map of the Ig heavy chain constant gene region reveals evidence for seven IgG isotypes and for IgD in the horse. J Immunol 173: 3230–3242.

44. Wang X, Olpe H, Polheinz HD (2009) On the genetics of immunoglobulins in the gray, short-tailed opossum Monodelphis domestica. Immunogenetics 61: 399–404.

45. Ye J (2004) The immunoglobulin IGHD gene locus in C5BL/6 mice. Immunogenetics 56: 479–481.

46. Zhao Y, Cui H, Whitington CM, Wei Z, Zhang X, et al. (2009) Ornithorhynchus anatinus (platypus) links the evolution of immunoglobulin genes in eutherian mammals and nonmammalian tetrapods. J Immunol 183: 3263–3269.

47. Zhao Y, Kacskovics I, Rabbani H, Hammarstrom L (2003) Physical mapping of the bovine immunoglobulin heavy chain constant region gene. J Biol Chem 278: 35024–35032.

48. Zhao Z, Xue Y, Pan-Hammarstrom Q, Liu W, Zhao Y, et al. (2002) Evolution of the six horse IGHG genes and corresponding immunoglobulin VH families. Int Immunol 2: 41–50.

49. Almagro JC, Martinez L, Smith SL, Alagon A, Estevez J, et al. (2006) Analysis of the somatic repertoire in Canis familiaris. Vet Immunol Immunopathol 137: 64–75.

50. Berman JE, Mellis SJ, Pollock R, Smith SI, Suh H, et al. (1988) Content and organization of the human immunoglobulin lambda light-chain locus on chromosome 16q22.1. Human Mol Genet 4: 983–991.

51. Berman JE, Mellis SJ, Pollock R, Smith SI, Suh H, et al. (1988) Characterization of the human immunoglobulin heavy (IGH) gene family. J Exp Med 168: 5659–5666.
86. Almagro JC, Dominguez-Martinez V, Lara-Ochoa F, Vargas-Madrazo E (1996) Structural repertoire in human VL pseudogenes of immunoglobulins: comparison with functional germline genes and amino acid sequences. Immunogenetics 41: 92–96.

87. Haino M, Hayashida H, Miyata T, Shin EK, Matsuda F, et al. (1994) Comparison and evolution of human immunoglobulin VH segments located in the 3’ 0.8-megabase region. Evidence for unidirectional transfer of segmental gene sequences. J Biol Chem 269: 2619–2626.

88. Vargas-Madrazo E, Almagro JC, Lara-Ochoa F (1995) Structural repertoire in VH pseudogenes of immunoglobulins: comparison with human germline genes and human amino acid sequences. J Mol Biol 246: 74–81.