Immunoglobulin Genomics in the Guinea Pig (*Cavia porcellus*)

Yongchen Guo¹, Yonghua Bao²*, Qingwen Meng³, Xiaoxiang Hu¹, Qingyong Meng¹, Liming Ren¹, 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, Xinxing Medical University, Xinxing, People’s Republic of China, ³ National Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin, People’s Republic of China, ⁴ College of Animal Science and Technology, Qingdao Agricultural University, Qingdao, People’s Republic of China

Abstract

In science, the guinea pig is known as one of the gold standards for modeling human disease. It is especially important as a molecular and cellular biology model for studying the human immune system, as its immunological genes are more similar to human genes than are those of mice. The utility of the guinea pig as a model organism can be further enhanced by further characterization of the genes encoding components of the immune system. Here, we report the genomic organization of the guinea pig immunoglobulin (Ig) heavy and light chain genes. The guinea pig IgH locus is located in genomic scaffolds 54 and 75, and spans approximately 6,480 kb. 507 V₄₇ segments (94 potentially functional genes and 413 pseudogenes), 41 D₄₇ segments, six J₄₇ segments, four constant region genes (μ, γ, ε, and δ), and one reverse δ remnant fragment were identified within the two scaffolds. Many V₄₇ pseudogenes were found within the guinea pig, and likely constituted a potential donor pool for gene conversion during evolution. The Igκ locus mapped to a 4,029 kb region of scaffold 37 and 24 is composed of 349 V₅₅ (111 potentially functional genes and 238 pseudogenes), three J₅₅, and one C₅₅ genes. The Igλ locus spans 1,642 kb in scaffold 4 and consists of 142 V₆₆ (58 potentially functional genes and 84 pseudogenes) and 11 J₆₆-C₆₆ clusters. Phylogenetic analysis suggested the guinea pig’s large germine V₄₇ gene segments appear to form limited gene families. Therefore, this species may generate antibody diversity via a gene conversion-like mechanism associated with its pseudogene reserves.

Introduction

The guinea pig (*Cavia porcellus*), also called the cavy, is a species of rodent belonging to the family *Caviidae*. This animal has been used in scientific experimentation since the 17th century. During the 19th and early 20th centuries, the guinea pig was a popular experimental animal for studying prevalent bacterial diseases such as tuberculosis and diphtheria [1], resulting in the epithet “guinea pig” being used to describe a test subject. Guinea pigs are currently still used in research, primarily as models for human diseases, including juvenile diabetes, tuberculosis, scurvy, and pregnancy complications [2,3,4,5,6,7,8].

Immunoglobulins (Igs) are only expressed by jawed vertebrates [9,10,11] and are usually composed of two identical heavy (H) chains and two identical light (L) chains. Exceptions include shark IgNAR and camelid IgGs, which are only comprised of heavy chains [9,10,11]. To date, mammalian Ig genes are organized into a “translocase” configuration [12]. In the heavy chain locus, multiple variable (V₄₇) diversity (D₄₇), and joining (J₄₇) gene segments are followed by μ, δ, γ, ε, and δ genes [13]. In the kappa encoding locus, multiple joining (J₆₆) region gene segments are present within a cluster, followed by a single constant (C₆₆) gene, whereas in the lambda encoding locus, joining (J₆₆) and constant (C₆₆) genes occur as J₆₆-C₆₆ blocks, which usually have multiple copies [14].

The word “guinea pig” is synonymous with scientific experimentation, but little is known about its Ig genes. We therefore used the recently available genome data of guinea pig provide as an opportunity to study the Ig genes of this species. Our study aimed to characterize the guinea pig IgH and IgL loci, in an effort to promote a better understanding of the immune system and evolutionary divergence of the Ig genes in placental mammals.

Materials and Methods

Guinea Pig Genome Sequence

Guinea pig (*C. porcellus*) genome data were obtained from the Ensembl database (http://www.ensembl.org), and the Broad Institute conducted genome sequencing and assembly (cavPor3, 6.79 x coverage, Jul 2008). High-coverage ensured increased accuracy of the genome analysis results.

Identification of the Guinea Pig Ig Genes

Guinea pig Ig constant region genes were retrieved on the basis of comparing guinea pig and human Ig gene sequences (http://
FUZZNUC, an online software (http://mobyle.pasteur.fr/cgi-bin/portal.py?forms::fuzznuc), was used to find adjacent recombination signal sequences (RSSs) for the identification of variable, diversity, and joining gene segments. Five or more mismatched bases were allowed to cover all genes.

Total RNA was extracted from spleen tissues of three male HARTLY guinea pigs using TRIzol Reagent (Invitrogen, USA) and pooled equally. Animals were treated in accordance with the China Agricultural University on the protection of animals used for experimental and other scientific purposes. The study was approved by animal welfare committee of China Agricultural University with approval number XK257.

Complementary deoxyribonucleic acid (cDNA) was synthesized from 1 μg of total RNA using 3’RACE-reverse transcription primers. The first polymerase chain reaction (PCR) amplification of each Ig heavy chain constant region gene (μ, γ, ε, and α) was performed using the corresponding sense primer and 3’RACE-antisense primer 1. While the second PCR amplification was conducted using the corresponding sense primer and 3’RACE-antisense primer 2. All primers are displayed in Table S1.

Southern Blotting Analysis of Genomic DNA Sample
Guinea pig liver genomic DNA was isolated according to the method as previously described [15]. The VH gene family-specific probes belonging to guinea pig VH1, VH2 and VH3 gene families were labeled using a PCR digoxigenin probe synthesis kit (Roche, Germany) with primers designed from VH1-95, VH2-17 and VH3-157 (The primer sequences are displayed in Table S1). Ten micrograms of genomic DNA digested separately with BamH I, EcoR I, Hind III and Xba I (New England Biolabs, USA) were loaded into each well of a 0.8% agarose gel and electrophoresed for 12 h, transferred to a positively charged nylon membrane (Roche, Germany). Hybridization and detection were conducted following the manufacturer’s instructions.
Figure 2. The guinea pig IgH locus in scaffold 75. The guinea pig IgH locus length is approximately 2,177 kb from the most VH (VH-217p) to the most end (VH-1p) in scaffold 75. Filled bars: potentially functional VH genes; open bars: VH pseudogenes. The scale does not apply the upper frame diagram.
doi:10.1371/journal.pone.0039298.g002

Figure 3. Alignment of the guinea pig IgD remnants with other mammalian IgD domains. 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.
doi:10.1371/journal.pone.0039298.g003
Figure 4. Phylogenetic analysis of mammalian immunoglobulin gamma genes. The phylogenetic tree was constructed from the amino acid sequences of the C_{H2} and C_{H3} exons with various mammalian species immunoglobulin gamma gene. The guinea pig IgG genes were clustered with rodents IgG genes.

doi:10.1371/journal.pone.0039298.g004
Sequence and Phylogenetic Analysis

Multiple sequence alignments were edited and handled with the Megalign software program [16], and the Clustal W and Clustal X algorithms [17,18], before being analysed using BioEdit [19]. Comparative phylogenetic trees were constructed using the PHYLIP 3.67 [20] software and TreeView [21] based on the final nucleotide alignment. The neighbor-joining algorithm was used for phylogenetic analysis and bootstrap support was provided by 1000 replicates. Sequences from other species used in our phylogenetic analyses and sequence alignments are presented in Figures S1, S2, and S3 and Table S2.

Dot Matrix Analysis

Pairwise dot matrix comparisons were made using DNAMAN software (window size = 30-bp, mismatch limit = 9-bp) to identify potential alignment of nucleotide bases between the sequences.

Definition of the VH/VL Gene Families

In mammals, germline VH and VL genes are categorized into different families according to their amino acid or nucleotide sequences similarity [22]. Sequences with greater than 75% similarity are generally considered to belong to the same family, while those with less than 70% similarity are placed in different gene families, and those possessing between 70% and 75% similarity are inspected on a case-by-case basis [23]. We placed potentially functional VH and VL gene segments sharing more than 70% similarity into the same family.

Results

Guinea Pig IgH Locus

Analysis of the genomic sequence revealed that the guinea pig IgH locus is located within genomic scaffolds 54 and 75 (Figure 1, Figure 2). The entire IgH locus spans approximately 6,480 kb of the two scaffolds (4,302 kb in scaffold 54 and 2,178 kb in scaffold 75). The total length is an estimate due to the existence of sequence gaps (Figure 1, Figure 2). Six JH segments, 507 VH, 4 DH, 4 constant region genes (μ, γ, ε, and α) and a reverse δ trace (marked with an arrow towards the left) were identified within the two scaffolds. Locations of the annotated IgH genes on the guinea pig genome are displayed in Table S3.
Figure 6. The alignment of deduced amino acid sequences of guinea pig 94 functional V<sub>H</sub> genes. The sequences of functional V genes were analyzed by Megalign software and BioEdit programe. The Designation of framework regions (FR) and complementarity determining regions (CDR) referred to IMGT numbering system, and the CDR region are indicated by grey-squared background. amino acid sequences that are the same as the top segment, VH1-75-54, are indicated with dots. Dashes mean gaps introduced to make the alignment. doi:10.1371/journal.pone.0039298.g006
Figure 7. Phylogenetic analysis of 94 guinea pig VH genes. A phylogenetic tree of nucleotide sequences of 94 guinea pig potentially functional VH segments was constructed. The three identified VH gene families are labeled with Arabic numerals.
doi:10.1371/journal.pone.0039298.g007
Analysis of the Guinea Pig Constant Region Genes

To acquire the cDNA sequence for four heavy chain isotypes (μ, γ, ε, and ι), 3’ RACE were performed on splenic RNA using specific primers. Using this strategy we successfully identified genes encoding the constant region exons for IgM, IgG, IgE, and IgA (Figure S4). As predicted from the analysis of the guinea pig Ig genomic genes, the domain structure for the four isotypes was typical of that for other mammalian species (Figure S5).

Within the guinea pig genome, C_{H4} and parts of the C_{H3} segments of the μ gene were determined to be missing because of the existence of gaps (Figure 1). By using the 3’ RACE method, the complete secreted IgM, including four exons, was successfully cloned (Figure S4–S5).

Most mammals also express the δ gene, with the exception of the rabbit [24] and opossum [25]. The area predicted to contain the IgD region, and in particular the 3’ region of the IgM exons between IgM and IgG, as well as the whole cavPor3 assembly, was thoroughly searched in two different orientations for coding sequences that might correspond to a putative IgD. These searches detected a trace fragment that is homologous with mammalian IgD (Figure 3) and showed an opposite transcription direction to the upstream μ gene. Based on sequence alignment, the three-segment trace was found to belong to the C_{H2} and C_{H3} domains (Figure 3).

Although two IgG isoforms (IgG1 and IgG2) were previously identified in domestic guinea pig serum [26], we only identified one IgG gene in the guinea pig genome perhaps due to sequence gaps. Our sequence alignment further showed that the recognized IgG in guinea pig genome shared the highest similarity with IgG1 (six amino acid difference) (Figure S6). Interestingly, we were just able to clone the IgG2 mRNA transcript by 3’ RACE. To address this question, a further Southern blotting experiment using the C_{H2} exon (high similarity between IgG1 and IgG2) as a probe, which showed that there were more than one γ genes in the guinea pig genome (Figure S7). Taken together, these data suggested that the guinea pig had two γ genes in its genome, but only one was preferentially expressed.

A phylogenetic tree constructed with C_{H2} and C_{H3} exons of IgG from different mammalian species revealed that the guinea pig IgG genes were clustered with rodents IgG genes (Figure 4). IgG2 and IgA also exhibit a hinge region, which is thought to have evolved by condensation of the C_{H2} exon in an ancestral isotype such as IgY in birds [25,27]. The hinge region of IgA is encoded by the 3’ end of the C_{H2} exon, as observed in other eutherian mammals [28,29,30] (Figure S4–S5).

In mammals, especially in humans and mice, a pentameric tandem repeat sequence is found upstream of the heavy chain constant region, which acts as a switch or S region. S regions have previously been mapped and sequenced, and are relevant in Ig class switch recombination. Such characteristic tandem repeats were also found within the upstream C_{H4}, C_{E} and C_{ε} gene sequences of the guinea pig. Three putative S regions span 2.2 kb to 3 kb, and exhibit similar repeats (TAGCTG and GGGCTG) to those observed in humans and mice. However, the characteristic sequence of the switch regions could not be identified within the γ gene, most likely due to sequence gaps. Dot matrix analysis of the guinea pig γ region revealed substantial nucleotide similarity with those of humans and mice (Figure 5).

Analysis of the Guinea Pig V_{H} Gene Segments

A total of 507 V_{H} segments were identified in scaffolds 54 and 75 (Figure 1, Figure 2). Ninety-four of these appeared to be potentially functional because they contained leader exons (L), uninterrupted open reading frames (ORF), downstream RSS, and a V gene domain (framework regions and complementarity determining regions). The remaining 413 segments that contain either in-frame stop codons or are partial sequences were designated as pseudogenes (Figure 1, Figure 2 and Table S3). Given that gaps existed within the assembly, it is possible that as yet unidentified V_{H} genes are also present in the guinea pig genome.

Phylogenetic analysis and multiple sequence alignments, including all functional guinea pig germline V_{H} gene segments, revealed the V_{H} gene families 1, 2 and 3, which were comprised of 22, 17 and 55 members, respectively (Figure 6, Figure 7). The
Figure 9. Phylogenetic analysis of mammalian \( V_{\mu} \) genes. Guinea pig sequences are represented by red branches, other species sequences are represented by black branches. Three \( V_{\mu} \) families of guinea pig were clustered respectively with other mammalian \( V_{\mu} \) families, and the \( V_{\mu} \) families of guinea pig belong to Clan II and Clan III. The \( V_{\mu}3 \) family could be divided into two subfamilies.

doi:10.1371/journal.pone.0039298.g009
largest family, V₃, could be further divided into two subfamilies (Figure 6, Figure 7). We also performed Southern blotting to verify the multiple numbers of V genes of different families in the guinea pig genome (Figure 8).

We chose all potentially functional guinea pig V genes and VH sequences that represented previously reported gene families from other mammalian species to construct a neighbor-joining phylogenetic tree [31,32] (Figure 9). The phylogenetic tree indicated that the mammalian VH gene families were classified into three clans [33]. The guinea pig VH families 1 and 2 belonged to clan II, and family 3 belonged to clan III.

Analysis of the Guinea Pig DH and JH Gene Segments

Approximately 504 kb downstream from the last VH segment (VH1-1), we identified 41 DH segments (i.e., DH1–DH41). They spanned a 660 kb region of DNA in scaffold 54 (Figure 1). Each

Table: DH sequences

| DH | Nonamer | 12bp Spacer | Heptamer | Coding region of DH segment | Heptamer 12bp Spacer | Nonamer |
|----|---------|-------------|---------|-----------------------------|----------------------|---------|
| D1 | GTCGAGTC | TAATATTGTA | CACGTCG | ATCGATTAATGCTACCTTTCACTCTTCG | CACGTCG | TAATATTGTA |
| D2 | ATTGCTGC | TAATATTGTA | CACGTCG | ATCGATTAATGCTACCTTTCACTCTTCG | CACGTCG | TAATATTGTA |
| D3 | GCCTGGTTG | TAATATTGTA | CACGTCG | ATCGATTAATGCTACCTTTCACTCTTCG | CACGTCG | TAATATTGTA |
| D4 | GAGGCTGCC | TAATATTGTA | CACGTCG | ATCGATTAATGCTACCTTTCACTCTTCG | CACGTCG | TAATATTGTA |
| D5 | CTGAGCTCT | TAATATTGTA | CACGTCG | ATCGATTAATGCTACCTTTCACTCTTCG | CACGTCG | TAATATTGTA |

Figure 10. The nucleotide sequences of DH genes. DH coding region nucleotide sequences are represented together the RSS elements (nonamer and heptamer sequences).

Figure 11. The nucleotide sequences of JH genes. JH nucleotide sequences are represented together the RSS elements (nonamer and heptamer sequences) and RNA donor splicing site.

Figure 12. The nucleotide sequences of VH genes. VH coding region nucleotide sequences are represented together the RSS elements (nonamer and heptamer sequences) and RNA donor splicing site.

Analysis of the Guinea Pig Di and Jj Gene Segments

Approximately 504 kb downstream from the last VH segment (VH1-1), we identified 41 DH segments (i.e., DH1–DH41). They spanned a 660 kb region of DNA in scaffold 54 (Figure 1). Each

Table: Nonamer, Spacernonamer, and JH region

| Nonamer | Spacernonamer | JH region | DNA donor splicing site |
|---------|-------------|-----------|------------------------|
| JL1     | AGGTTTGT    | GCCGAGCA  | ATCGATTAATGCTACCTTTCACTCTTCG | CACGTCG | TAATATTGTA |
| JL2     | CGGTTGCT    | GTGCGGTG  | ATCGATTAATGCTACCTTTCACTCTTCG | CACGTCG | TAATATTGTA |
| JL3     | GGTTTTCG    | CACGTCG   | ATCGATTAATGCTACCTTTCACTCTTCG | CACGTCG | TAATATTGTA |
| JL4     | GCCTGGTTG   | CACGTCG   | ATCGATTAATGCTACCTTTCACTCTTCG | CACGTCG | TAATATTGTA |
| JL5     | GTGCGGTG    | CACGTCG   | ATCGATTAATGCTACCTTTCACTCTTCG | CACGTCG | TAATATTGTA |
| JL6     | GGTTTTCG    | CACGTCG   | ATCGATTAATGCTACCTTTCACTCTTCG | CACGTCG | TAATATTGTA |

Figure 13. The nucleotide sequences of Jj genes. Jj nucleotide sequences are represented together the RSS elements (nonamer and heptamer sequences) and RNA donor splicing site.

doi:10.1371/journal.pone.0039298.g010

doi:10.1371/journal.pone.0039298.g011

doi:10.1371/journal.pone.0039298.g012
Figure 12. The guinea pig Igκ locus in scaffold 37. The guinea pig Igκ locus is in scaffold 37, the potentially functional Vκ gene segments are shown in filled bars, and pseudogenes are represented by open bars. The unidirectional arrowheads below Vκ gene segments on scaffold 37 indicate that their transcriptional direction is opposite to downstream Jκ segments. Double slashes indicate gaps >10 kb. The scale does not apply the upper frame diagram.
doi:10.1371/journal.pone.0039298.g012

Figure 13. The guinea pig Igκ locus in scaffold 24. The guinea pig Igκ locus is in scaffold 24, the potentially functional Vκ gene segments are shown in filled bars, and pseudogenes are represented by open bars. The unidirectional arrowheads below Vκ gene segments merely indicate a transcriptional direction different from that of the remaining Vκ gene segments in scaffold 24. The scale does not apply the upper frame diagram.
doi:10.1371/journal.pone.0039298.g013
DH segment was flanked by conserved RSS elements composed of heptamers and nonamers separated by 12 bp spacers and existed within at least one alternative reading frame (Figure 10 and Figure S8), suggesting that they are potentially functional.

JH region contained six genes (designated JH1 to JH6) spanning approximately 2 kb. Each JH gene had an upstream RSS element with a 22–23 bp spacer, ORF, and a downstream RNA donor-splicing site at the 3’ end, suggesting that they are potentially functional (Figure 11).

Guinea Pig Ig\(k\) Locus

The guinea pig Ig\(k\) chain is located in scaffolds 37 and 24, and spans an approximately 4,029 kb genomic region (Figure 12, Figure 13 and Table S4). A total of 349 V\(k\) genes were identified. Further analysis revealed that 111 V\(k\) genes might be potentially functional genes, given that they contained an L sequence, ORF, RSS and V domain. The remaining 238 segments contain either in-frame stop codons or frameshifts, and are thus designated as pseudogenes. Based on sequence analysis, the 111 potentially functional V\(k\) genes were divided into six families (V\(k\)1–V\(k\)6), which contained 69, 15, 7, 17, 2, and 1 member/s, respectively (Figure 14). In addition, 222 V\(k\) genes were arranged in the same transcriptional orientation and exhibited downstream J\(k\) and C\(k\), and 61 V\(k\) segments and a reverse transcriptional direction in scaffold 37. Downstream of the V\(k\) genes, three J\(k\) gene segments were constructed. The six V\(k\) gene families are labeled with Arabic numerals.

doi:10.1371/journal.pone.0039298.g014

**Figure 14. Phylogenetic analysis of the 111 guinea pig V\(k\) genes.** A phylogenetic tree of the nucleotide sequences of 111 guinea pig V\(k\) segments were constructed. The six V\(k\) gene families are labeled with Arabic numerals.
were identified that spanned 0.6 kb. Furthermore, approximately 4 kb downstream from the last J\(k\), a single C\(k\) gene was identified.

Guinea Pig Ig\(\lambda\) Locus

Guinea pig Ig\(\lambda\) chain genes were identified in scaffold 4, and spanned an approximately 1,642 kb length (Figure 15, Table S5). Of 142 germline V\(\lambda\) genes identified, 58 segments were categorized as potentially functional genes, and the remaining 84 were differentiated as pseudogenes. Based on sequence similarity analysis, the potentially functional guinea pig V\(\lambda\) genes were assigned to twelve families (Figure 16), comprised of 3, 1, 13, 6, 6, 0, 1, 3, 2, 5, 3 and 7 member/s, respectively. In contrast to the V\(k\) genes, all the V\(\lambda\) genes have the same transcriptional orientation as the downstream J\(k\) and C\(k\) regions. At the 3\(^{9}\) end of this scaffold, 11 J segments and C segments were organized in tandem and spanned 44 kb, while the C\(\lambda\)2 exhibited less structural integrity owing to the presence of gaps. The sequence alignments of J\(\lambda\) and C\(\lambda\) gene segments are shown in Figure 17.

Discussion

Rodents are a ubiquitous group of species worldwide, representing nearly half of all mammalian species, which evolved from a common ancestor shared with the lagomorphs approximately 62–100 million years ago [34,35]. The classification of the guinea pig within the family Caviidae and genus Cavia is somewhat controversial because the origin of this rodent is poorly known, with current classification data mostly relying on fossils or genetic relationships [36,37,38,39,40,41,42,43,44]. We therefore analyzed the Ig genes sequences of the guinea pig, not only to better understand the immune system of this species, but also to provide data for comparative studies of mammalian Ig genes.

The IgH locus of mammals is arranged in a “translocon” configuration [32,45,46,47,48,49,50,51,52]. In the present study, we characterized the guinea pig Ig genes based on recently released genomic data and our experimental results. The guinea pig IgH locus in a configuration of VH (507)-DH (41)-JH (6)-C\(\mu\)-C\(\delta\)-C\(\gamma\)-C\(\varepsilon\)-C\(\alpha\) spanning at least 6,480 kb in two scaffolds may be largest in all mammalian species studied so far.

On the basis of sequence analysis, we identified a single \(\mu\), \(\gamma\), \(\varepsilon\), and \(\alpha\) gene within the guinea pig genome. We also found three fragments of \(\delta\) gene in an opposite direction downstream from the \(\mu\) gene. Due to sequence gaps, it is not certain if an additional functional \(\delta\) gene also exists in this species. We have also tried to confirm the sequence and orientation of the \(\delta\) gene fragments by genomic PCR to eliminate assembly error. The sequences of the \(\delta\) gene fragments were verified. Because two genomic fragments (approximately two kb and six kb) between \(\mu\) and \(\delta\) gene can not be successfully amplified, so the orientation of the \(\delta\) gene fragments remains a question.

Many placental mammals, such as human, cow, sheep, horse and dog, have a single functional \(\delta\) gene. Except for a functional \(\delta\) gene which consists of ten C\(\delta\) domains, a reverse \(\delta\) pseudogene was previously observed in the platypus IgH locus [53], while in the elephant genome, only one C\(\delta\)3 remnant fragment was identified [51]. In camels, C\(\delta\)3 exon appears to be highly mutated [54]. In the rabbit [24] and opossum (marsupial) [25], the \(\delta\) gene has clearly been shown to be missing in their genomes, indicating that the \(\delta\) gene might be not as essential as the \(\mu\) gene in the humoral immunity.
IgG is an important antibody molecule, which is believed to have initially evolved 600 million years ago [55]. The structure of \( \gamma \) gene usually contains three C\( \gamma \) domains and one hinge domain. Different IgG subclasses have been reported in the majority of mammalian species, ranging from one in the rabbit [56], two in sheep [57], three in cattle [58], four in human and rat [49,59], five in cattle [58], six in human and rat [49,59], five in mouse [60], six in pig [61], seven in horse [62], up to nine in elephant [51]. In the guinea pig, two IgG subclasses are identified, and they share about 73% amino acid similarity in C\( \gamma \) domains. It has been postulated that different \( \gamma \) subclasses are derived from gene duplications of ancestral \( \gamma \) gene in mammals [63]. Three ancestral \( \gamma \) genes of mouse and rat can evolve multiple \( \gamma \) genes by gene duplications [64,65], and the divergence of the \( \gamma \) genes of human depends on one ancestral \( \gamma \) gene and duplicated C\( \gamma \)-C\( \delta \)-C\( \alpha \) fragments [66].

In different mammalian species, the number of \( V_H \) genes and the ratio between \( V_H \) functional genes and \( V_H \) pseudogenes vary significantly, even between closely related species, like mice and rats. For example, there are 60 pseudogenes and 44 functional \( V_H \) genes in humans, whereas in cows, only 6 pseudogenes and 11 functional genes have been identified [49,67,68]. The guinea pig germline \( V_H \) repertoire contains at least 507 \( V_H \) gene segments, which is the largest number in mammals studied to date. A large number of these germline \( V_H \) genes may greatly contribute to guinea pig antibody diversity. Another notable feature is the number of guinea pig \( V_H \) pseudogenes, which amount to 81% of

---

**Figure 16. Phylogenetic tree analysis of the 58 guinea pig \( V_L \) genes.** A phylogenetic tree of the nucleotide sequences of 58 guinea pig \( V_L \) segments were constructed. The eleven \( V_L \) gene families are labeled with Arabic numerals.

doi:10.1371/journal.pone.0039298.g016
the total VH genes (413 pseudogenes vs. 507 total genes). The high frequency of gene duplication in variable region generated multiple VH copies, many of which became pseudogenes due to genomic drift [68]. The VH pseudogenes are not truly nonfunctional in some species, for example, in rabbits, the pseudogenes usually are used to generate immunoglobulin diversity by gene conversion [67].

Guinea pig VH gene segments were also divided into three gene families (families 1, 2 and 3), which were orthologous to human VH families 4, 2 and 3, respectively. The guinea pig exhibited a multiple gene family group, as observed in dogs (three families) [69], humans (seven families) [70,71], horses (seven families) [32,72] and mice (six families) [73], yet different from the single family group observed in sheep [57], rabbits [74], camels [75], swine [76] and cattle [72]. The guinea pig VH genes were also classified into different clans as in other mammalian species [50,52,67,72,74,77,78,79]. Some reports have revealed that representative VH genes belong to all three VH clans in the human, mouse, cat and dog [67,77]. This characteristic is different in cattle, horses and sheep [50,52,72,78,79], which have lost much of their ancestral repertoire, and the VH genes only belong to clan II. Pigs and rabbits only have clan III genes [73,78,79]. These features are also found in swine and rabbits.

The precise evolutionary relationship among mammalian lineages has not yet been resolved [81]. Results of the relevant study show that marsupials and monotremes were estimated to have separated from the common ancestor of present-day placental mammals more than 130 million years ago, while the major radiation of the placental mammals occurred approximately 70–120 million years ago [82,83,84,85,86]. Certain V gene families which descended from common ancestor genes are orthologues between nonplacental mammals and placental mammals. For example, platypus (Monotreme), dog and human share the same ancestral gene families (VH3 and VH4), while the American short-tailed opossum (Monodelphis domestica), swine, rabbit and human share ancestral VH3 family, and artiodactyl species share ancestral VH4 family. With an older evolutionary origin in present day mammals [86], platypus have two ancestral VH gene families, while other mammals share one or two ancestral VH gene families. These could be explained by an inactivation or loss of VH gene members in these species during evolution [25]. For new VH gene families in human or mouse, the divergence of VH genes probably occurred after speciation.

The ratio of functional Vk and Vl is variable within mammalian species, with the germline Vk genes being more abundant than Vl genes in humans (40 Vk genes vs. 30 Vl genes) and mice (Vk genes vs. Vl genes over 95%) [87]. This is also the case for the guinea pig.
pig, in which $V_k$ germline genes are more dominant than $V_l$ (84 functional $V_k$ genes vs. 30 functional $V_l$ genes). It has been proposed that the priority of use of the light-chain gene isotypes at the protein level may be connected with the overall number of $V_k$ vector genes [97]. It is possible that the $k$ chain preponderates over the $l$ chain at the protein level in guinea pigs. Also, multiple pseudogenes exist in the guinea pig $V_H$ (413), $V_C$ (238), and $V_k$ (34) loci, which may contribute to the Ig diversity in guinea pigs, similar to other species [47,88,89].

In conclusion, we have reported the characterization and annotation of the guinea pig Ig loci genomic maps for the first time. This information, together with the characterization of the guinea pig Ig germline gene repertoire currently being undertaken, should lay the foundations for further studies into the differentiation and structure of mammalian Ig genes, including those found in guinea pigs.

Supporting Information

Figure S1 Multiple sequence alignments of guinea pig $V_H$ genes.

Figure S2 Multiple sequence alignments of guinea pig $V_k$ genes.

Figure S3 Multiple sequence alignments of guinea pig $V_l$ genes.

Figure S4 Guinea pig immunoglobulin heavy chain constant region encoding gene amino acid sequences. Four guinea pig encoding gene amino acid sequences of constant region (IgM, IgG, IgE and IgA) are cloned by 3’RACE. (TIF)

Figure S5 Analysis results of the guinea pig $\gamma, \alpha$ and $\epsilon$ genes in scaffold 54. Four guinea pig constant region encoding genes of transmembrane type for IgM, IgG, IgE and IgA were identified by bioinformatics analysis. (TIF)

Figure S6 Alignment of the IgG amino acid sequences of guinea pig. Alignment of IgG1, IgG2 and IgG sequences. Dots indicate similar residues as in G1 and G, whereas dashes indicate gaps introduced for optimal alignment. The difference is represented by grey background between G and G1. (TIF)

Figure S7 Southern blotting analysis of guinea pig genomic DNA. Southern blotting analysis of guinea pig heavy chain constant region IgG genes. The genomic DNA was digested with BamHI (B), EcoRI (E), Hind III (H) and Xba I (X), and hybridized with probes for IgG-CH1 sequence. (TIF)

Figure S8 Guinea pig germline $D_H$ segments. The nonamers (9-mer) and heptamers (7-mer) are displayed. Heptamer components that are different from the consensus (5’: CAGCTGTG and 3’: CACAGTG) are shadowed. The deduced amino acids of all three reading frames of the coding region of $D$ segments are shown, all the $D_H$ segments were attached by 12 bp spacer. (RAR)

Table S1 Primers were designed for amplify guinea pig four classes immunoglobulin M, E, A, and G genes.

Table S2 GenBank accession numbers or references of the gene sequences from other species.

Table S3 The guinea pig immunoglobulin heavy chain DNA segments location in scaffolds 54 and 75.

Table S4 The guinea pig immunoglobulin $\kappa$ light chain DNA segments location in scaffolds 37 and 24.

Table S5 The guinea pig immunoglobulin $\lambda$ light chain DNA segments location in scaffold 4.

Author Contributions

Conceived and designed the experiments: YG YB YZ. Performed the experiments: YG YB YZ. Analyzed the data: YG YB YZ. Contributed reagents/materials/analysis tools: QM XH QM LR NL. Wrote the paper: YG YB YZ.

References

1. Paddilla-Carlin DJ, McMurray DN, Hickey AJ (2008) The guinea pig as a model of infectious diseases. Comp Med 58: 324–340.
2. Mizutani N, Nabe T, Shimazu M, Yoshino S, Kohno S (2011) Effect of Genodermatitis baritum on pollen-induced bighorn nasal blockage in a guinea pig model of allergic rhinitis. Phytother Res 26: 325–332.
3. Ikeda R, Nakaya K, Oshima H, Oshima T, Kawai T, et al. (2011) Effect of aspiration of periphray during stapes surgery on the endocochlear potential of guinea pig. Otolaryngol Head Neck Surg 145: 891–895.
4. Sturm EM, Schuligoi R, Konya V, Sturm GJ, Heinemann A (2011) Inhibitory effect of prostaglandin E2 on bone marrow kinetics of eosinophils in the guinea pig. Oto Nasus Larynx 38: 671–677.
5. Anderson KE, Soler R, Fullhase C (2011) Rodent models for urodynamic investigation. Neurourol Urodyn 30: 636–646.
6. Kajiwara D, Aoyagi H, Shigeno K, Togawa M, Tanaka K, et al. (2011) Role of hematopoietic prostaglandin D synthase in biphasic nasal obstruction in guinea pig. Otolaryngol Head Neck Surg 145: 891–895.
7. Clewley JP, Arnold C (1997) MEGALGN. The multiple alignment module of LASERGENE. Methods Mol Biol 70: 119–129.
8. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
9. Higgins DG, Thompson JD, Gibson TJ (1996) Using CLUSTAL for multiple sequence alignments. Methods Enzymol 266: 383–402.
10. Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98.
20. Felsenstein J (2007) PHYITL. (Phylogeny Inference Package) Ver. 3.6

[Computer software and manual]. Seattle, Washington: Department of Genome Sciences and Department of Biology, University of Washington.

21. Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. Comp Appl Biosci 12: 357–358.

22. Schroeder HW Jr, Hillson JL, Perlmutter RM (1996) Structure and evolution of mammalian VH families. Int Immunol 2: 41–50.

23. Broader PH, Rihet P (1984) The immunoglobulin heavy chain variable region (IgVH) locus in the mouse. One hundred IgVH genes comprise seven families of homologous genes. Eur J Immunol 14: 922–930.

24. Lanning DK, Zhai SK, Knight KL (2003) Analysis of the 3' Cmu region of the rabbit Ig heavy chain locus. Gene 309: 135–144.

25. Wang X, Ogilvie RE, Miller RD (2009) On the genomics of immunoglobulins in the gray, short-tailed opossum Monodelphis domestica. Immunogenetics 61: 581–596.

26. Pluzhnikov RP, Wishner BG, Bryant SH, Amzel LM, Lopez de Castro JA et al. (1979) Three dimensional structure of the pFc' fragment of guinea pig IgA. Mol Memol Biol 16: 841–850.

27. Warr GW, Magee KE, Higgins DA (1995) IgV: clues to the origins of modern antibodies. Immunol Today 16: 392–396.

28. Tucker PW, Slightom JL, Blattner FR (1981) Mouse IgA heavy chain gene sequence: implications for evolution of the immunoglobulin hinge axons. Proc Natl Acad Sci U S A 78: 7684–7686.

29. Kawamura S, Saitou N, Ueda S (1992) Concerted evolution of the primate IgA heavy-gene alpha through gene conversion. J Biol Chem 267: 7359–7367.

30. Osborne BA, Goldle TE, Schwartz RL, Rudikoff S (1988) Evolution of the IgA heavy chain gene in the genus Mus. Genetics 119: 925–931.

31. Nitsukawa T, Sato K, Inoue Y (1988) Coevolution of the murine IgA high- and light-chain variable-region gene families. Mol Biol Evol 15: 617–625.

32. Almagro JC, Martinez L, Smith SL, Alagon A, Estevé J et al. (2006) Analysis of the mouse VH repertoire and comparison with the human IGHV germ line gene repertoire, and showing evolution of mouse VH genes. Mol Immunol 43: 1036–1045.

33. Tomlinson DM, Walter G, Marks JD, Lively MB, Winter G (1992) The repertoire of human germ line VH sequences reveals about fifty groups of structural segments with different evolutionary loops. J Mol Biol 227: 776–796.

34. Benton MJ, Donoghue PC (2007) Paleontological evidence to date the tree of immunoglobulins. Immunol Rev 217: 100–120.

35. Osborne BA, Goldreale AR, Bertke SK, Blattner FR (1988) Structure and evolution of the immunoglobulin Cmu region of the African elephant (Loxodonta africana). PLoS ONE 34: 1009–1020.

36. Horner DS, Lefkimmiatis K, Reyes A, Gissi C, Saccone C et al. (2007) A preliminary analysis of the immunoglobulin heavy chain locus in the platypus (Ornithorhynchus anatinus). Mol Memol Biol 46: 2513–2522.

37. Robinson-Rechavi M, Ponger L, Mouchiroud D (2000) Nuclear gene LCAT linkage to the Ig CH locus. J Immunol 164: 2001–2009.

38. Ellison JW, Berson BJ, Hood LE (1992) The nucleotide sequence of a human immunoglobulin C gamma1 gene. Nucleic Acids Res 10: 4071–4079.

39. Knight KL, Burzette RC, McNichola JM (1985) Organization and polymerization of rabbit immunoglobulin heavy chain genes. J Immunol 134: 1245–1250.

40. Dufour V, Malinge S, Nau F (1996) The sheep Ig gamma variable region consists of single VH family. J Immunol 156: 2163–2170.

41. Cao Y, Adachi J, Yano T, Hasegawa M (1994) Phylogenetic place of guinea pig. Immunol Today 15: 392–398.

42. Lin YH, McLenachan PA, Gore AR, Phillips MJ, Ota R et al. (2002) Four new segments with different hypervariable loops. J Mol Biol 227: 776–798.

43. Gruber AG, Ovchinnikov VA, Rabbitts TH (2000) Immunoglobulin heavy chain exon 2: implications for evolution of immunoglobulin hinge axons. PLoS ONE 5: 1009–1020.

44. Flanagan JG, Lefranc MP, Rabbitts TH (1984) Mechanisms of divergence and convergence of the human IgV heavy chain alpha and one alpha 2 constant region gene sequences. Cell 66: 681–688.

45. Shimizu A, Takahashi N, Yosa Y, Honjo T (1982) Organization of the constant-region gene family of the mouse immunoglobulin heavy chain. Cell 29: 499–506.

46. Butler JE, Sun J, Wertz N, Sinkora M (2006) Antibody repertoire development in viviparous mammals. Dev Comp Immunol 30: 199–221.

47. Graf R, Werner K, Pfeffer M, Klämbt FR (1981) Gene conversion in bovine Ig. Construction and characterization of hapten-binding bovine/marine chimeric IgE, IgA, IgG1, IgG2, and IgG3 molecules. J Immunol 140: 3654–3659.

48. Flanagan JG, Lefranc MP, Robbins TH (1981) Mechanisms of divergence and convergence of the human IgH heavy chain alpha and alpha 2 constant region gene sequences. Cell 66: 681–688.
84. Air GM, Thompson EO (1971) Studies on marsupial proteins. IV. Amino acid sequence of myoglobin from the red kangaroo, Megaleia rufa. Aust J Biol Sci 24: 75–95.
85. Retief JD, Winkfein RJ, Dixon GH (1993) Evolution of the monotremes. The sequences of the protamine P1 genes of platypus and echidna. Eur J Biochem 218: 457–461.
86. Janke A, Gemmell NJ, Feldmaier-Fuchs G, von Haeseler A, Paabo S (1996) The mitochondrial genome of a monotreme—the platypus (Ornithorhynchus anatinus). J Mol Evol 42: 153–159.
87. Almagro JC, Hernandez I, Ramirez MC, Vargas-Madrazo E (1998) Structural differences between the repertoires of mouse and human germline genes and their evolutionary implications. Immunogenetics 47: 355–363.
88. Lucier MR, Thompson RE, Waare J, Lin AW, Osborne BA, et al. (1996) Multiple sites of V lambda diversification in cattle. J Immunol 161: 5430–5444.
89. Hano 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.