OPEN ACCESS Freely available online

PLoS one

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

Yongchen Guo[1], Yonghua Bao[2]*, Qingwen Meng[3], Xiaoxiang Hu[1], Qingyong Meng[1], Liming Ren[1], Ning Li[1], Yaofeng Zhao[1,4]*

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** 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 V_H segments (94 potentially functional genes and 413 pseudogenes), 41 D_H segments, six J_H segments, four constant region genes (μ, γ, ε, and α), and one reverse δ remnant fragment were identified within the two scaffolds. Many V_H 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 $V_κ$ (111 potentially functional genes and 238 pseudogenes), three $J_κ$ and one $C_κ$ genes. The Igλ locus spans 1,642 kb in scaffold 4 and consists of 142 $V_λ$ (58 potentially functional genes and 84 pseudogenes) and 11 $J_λ$ -$C_λ$ clusters. Phylogenetic analysis suggested the guinea pig's large germline V_H 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.

Citation: Guo Y, Bao Y, Meng Q, Hu X, Meng Q, et al. (2012) Immunoglobulin Genomics in the Guinea Pig (*Cavia porcellus*). PLoS ONE 7(6): e39298. doi:10.1371/journal.pone.0039298

Editor: Sebastian D. Fugmann, National Institute on Aging, United States of America

Received December 29, 2011; **Accepted** May 17, 2012; **Published** June 22, 2012

Funding: The authors are supported by the National Science Fund for Distinguished Young Scholars (30725029), the Tai shan Scholar Foundation of Shan dong Province, the National Basic Research Program of China (973 Program-2010CB945300), Program for New Century Excellent Talents in University of China and the Swedish Research Council. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: yaofengzhao@cau.edu.cn (YZ); baoyonghua2005@yahoo.com.cn (YB)

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 'translocon' configuration [12]. In the heavy chain locus, multiple variable (V_H), diversity (D_H), and joining (J_H) gene segments are followed by μ, δ, γ, ε, and α genes [13]. In the kappa encoding locus, multiple joining ($J_κ$) region gene segments are present within a cluster, followed by a single constant ($C_κ$) gene, whereas in the lambda encoding locus, joining ($J_λ$) and constant ($C_λ$) genes occur as $J_λ$-$C_λ$ 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× 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://

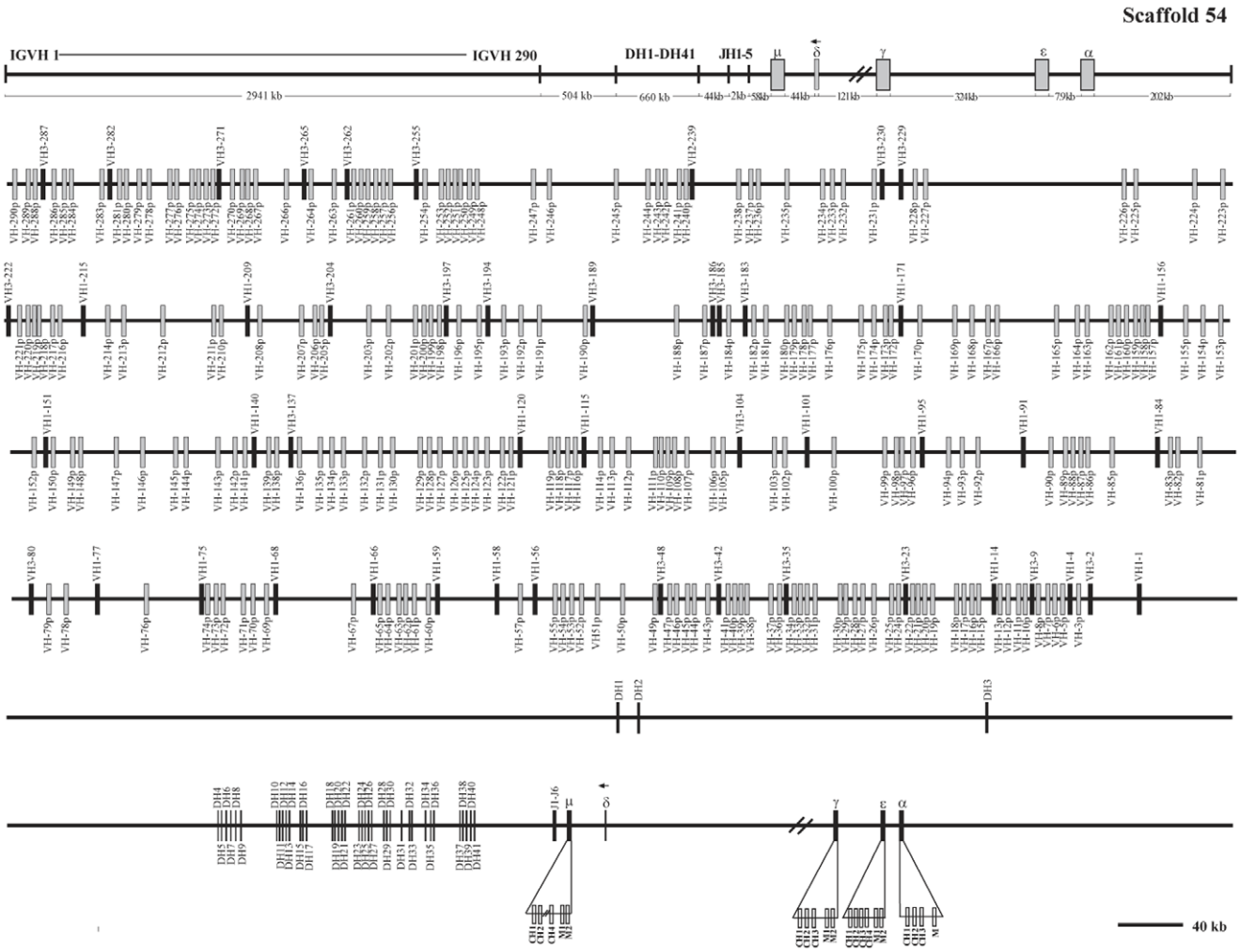

Figure 1. The guinea pig IgH locus in scaffold 54. The guinea pig IgH locus length is approximately 4,302 kb from the most 5′ V_H segment (V_H-290p) to the most 3′ end of α gene in genomic scaffold 54. Filled bars: potentially functional V_H genes; open bars: V_H pseudogenes and V-p; D_H: diversity genes; J_H: joining genes; μ: IgM coding gene; δ: IgD fragment; γ: IgG coding gene; ε: IgE coding gene and α: IgA coding gene. V_H and D_H genes are numbered based on the families and their order. The number before the dash (V_Hn) indicates the family, while the number after dash represents the genomic position. J_H genes are numbered based on the order of their locations in the locus. Double slashes indicate gaps >10 kb. The unidirectional arrowheads above δ gene segments in scaffold 54 indicate that its direction is opposite to the other heavy chain constant genes transcriptional direction. The scale does not apply the upper frame diagram.
doi:10.1371/journal.pone.0039298.g001

genome.ucsc.edu/). 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 Ribonucleic Acid (RNA) Extract and 3′ RACE

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 V_H gene family-specific probes belonging to guinea pig V_H1, V_H2 and V_H3 gene families were labeled using a PCR digoxygenin probe synthesis kit (Roche, Germany) with primers designed from V_H1-95, V_H2-17 and V_H3-157 (The primer sequences are displayed in Table S1). Ten micrograms of genomic DNA digested separately with *Bam*H I, *Eco*R I, *Hin*d 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.

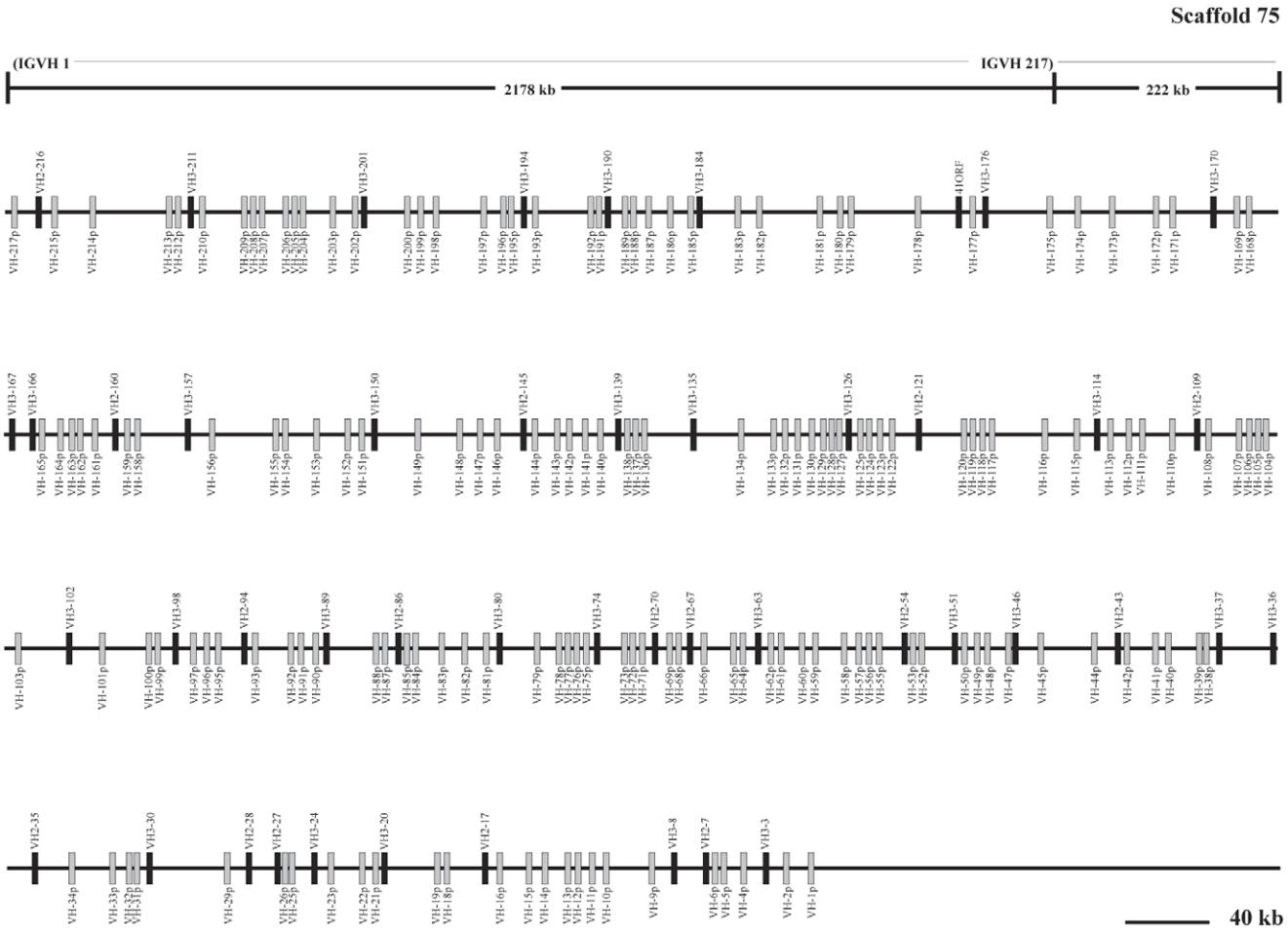

Figure 2. The guinea pig IgH locus in scaffold 75. The guinea pig IgH locus length is approximately 2,177 kb from the most V_H (V_H-217p) to the most end (V_H-1p) in scaffold 75. Filled bars: potentially functional V_H genes; open bars: V_H pseudogenes. The scale does not apply the upper frame diagram.
doi:10.1371/journal.pone.0039298.g002

```
Equus_caballus-CH2          LRKSEPSTRHTQPETQKPRIPVDTPLKECQSHTHPPSIYLLHPPLQGLWLKGEATFTCLVVGDDLKDAH
Cercocebus_torquatus-CH2    .S.AT-------------.P.ATNRNT..P...Q.LGV...P.A..D..FRDKV....F...S..Q...
Macaca_mulatta-CH2          .S.AT-------------.P.ATNRNT..P...Q.LGV...P.A..D..FQDKV....F...S..Q...
Canis lupus-CH2             A.V.V.P.S...TQA.E.GC....I.R..WN......L.M.R...R.P..Q...A..........QK..
Sus scrofa-CH2              -------.VS.KAL.-T....------A..N..R...V...L........RA.......A..Q..QE.R
Guinea pig                  .A.TR-------------CFSSGPL..A..A....S.TM....LC.ED..FQK..-------.G------

Sus_scrofa-CH3              LLCEVSSFSPLDILLTWLEGQQEVDPSWFATARPAAQPGNPTFRTWSVLRVPASPGHQDATYTCVVGHE
Equus_caballus-CH3          .Q.K..G...PE.V.......R.............T.....T..Q...I.L..TI..PPT..........
Bos_taurus-CH3              ......G...V.............E..Q....HTT..A.RASSH.......SSPLD.AG.......S..
Ovis_aries-CH3              ......G...V.............E..Q....HTT..S.HAS.H.....H.SSPLD.VGS......S..
Canis_lupus_familiaris-CH3  F.....G...P......IKD.I..........P.M....SG..Q...L...L.PQ.PHPP......R..
Macaca_mulatta-CH3          ......D...PN...M...N.R..NT.....TH.TP...STM.WA.......GPTSP.P.......S..
Papio_anubis-CH3            ......D...PN...M...D.R..NT.....TH.TP...STM.WA.......GPTSP.P.......S..
Cercocebus_torquatus-CH3    ......D...PN...M...D.R..NT.....TH.TP...STM.WA.......GPTSP.P.......S..
Guinea pig                  FCVKFPE..H.-N.FFMC..............YQDR.------CHFPN.QCST--------------.S

Equus_caballus-CH3          QREVDPSWFATARPTAQPGNTTFQTWSILLVPTIPGPPTATYTCVVGHEASRQL---LNTSWSLDTGGP
Sus_scrofa-CH3              .Q............A.....P..R...V.R..AS..HQD.............T.----..A..R....PL
Ailuropoda_melanoleuca-CH3  ----.ATG....S.V.............V.R.LAAQ..LP.......R.D...KV---..........PL
Canis_lupus_familiaris-CH3  .I..........P.M....SG......L.R.LAPQ..HPP......R.....K.---.........S.LT
Eptesicus_fuscus-CH3        P.Q..A......A...EQ.TS..R...V.R..AAS...P.....E.R.P..GR.---.........A--
human-CH3                   ISSSGTTYY.P...PP..RS...WA..V.R..AP.S.QP.......S..D..T.----..A.R..EVSCE
mouse-CH3                   HSKMKSTN.V..N......-G......V.RL.VALSSSLD......E....KTK----..A.K..AISED
Myotis_lucifugus-CH3        P.Q..S......V...E..KS..R...V.R..AA....R.....E.R.T..GR.---.....R...---
Guinea pig                  ----T.PGLP---LLTKI..A....C.V.HT.ATSSS.M.AHI....Y...WKMSRS.D.GSESQPE.S
```

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

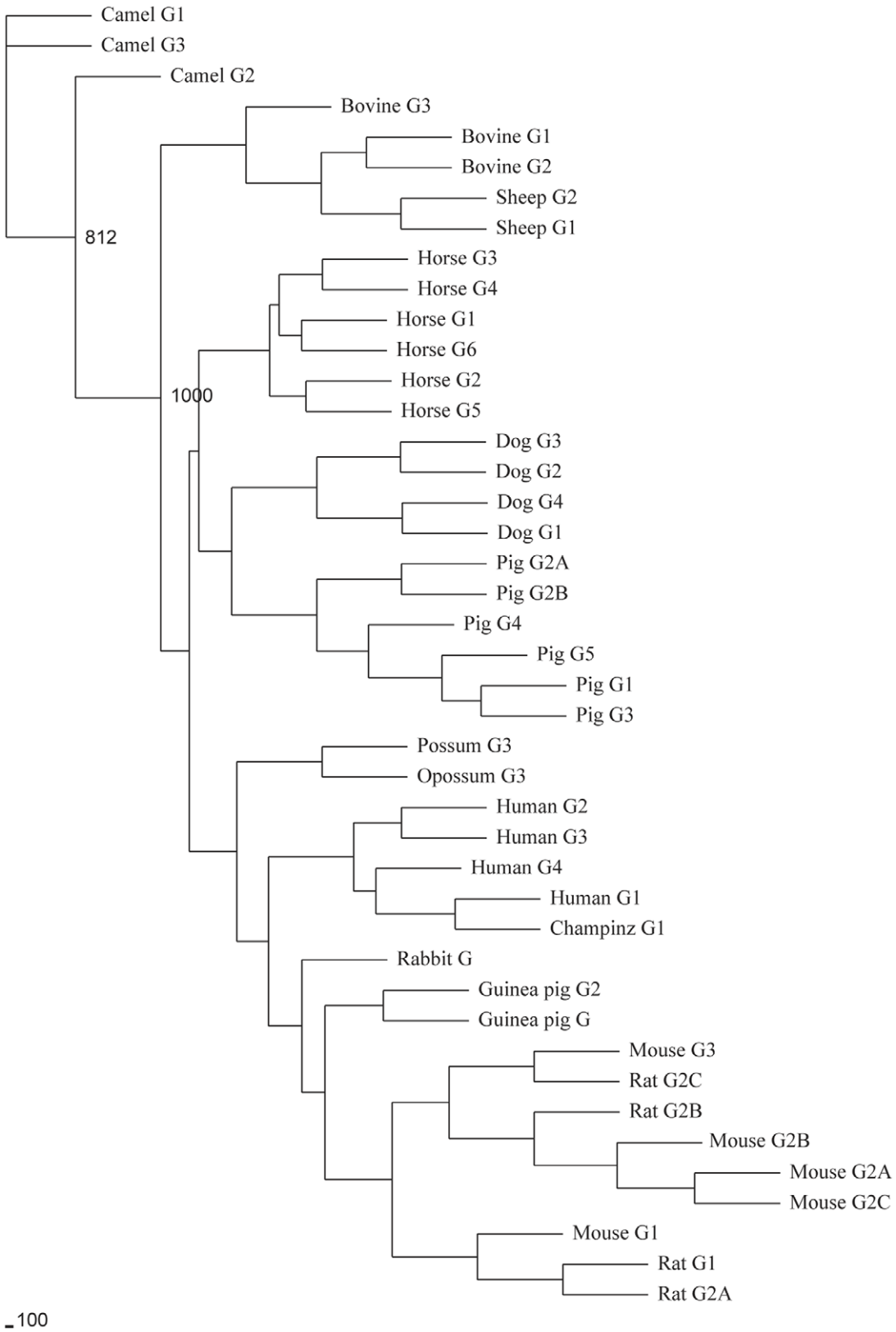

Figure 4. Phylogenetic analysis of mammalian immunoglobulin gamma genes. The phylogenetic tree was constructed from the amino acid sequences of the C_H2 and C_H3 exons with various mammalian species immunoglobulin gamma gene. The guinea pig IgG genes were clustered with rodents IgG genes.
doi:10.1371/journal.pone.0039298.g004

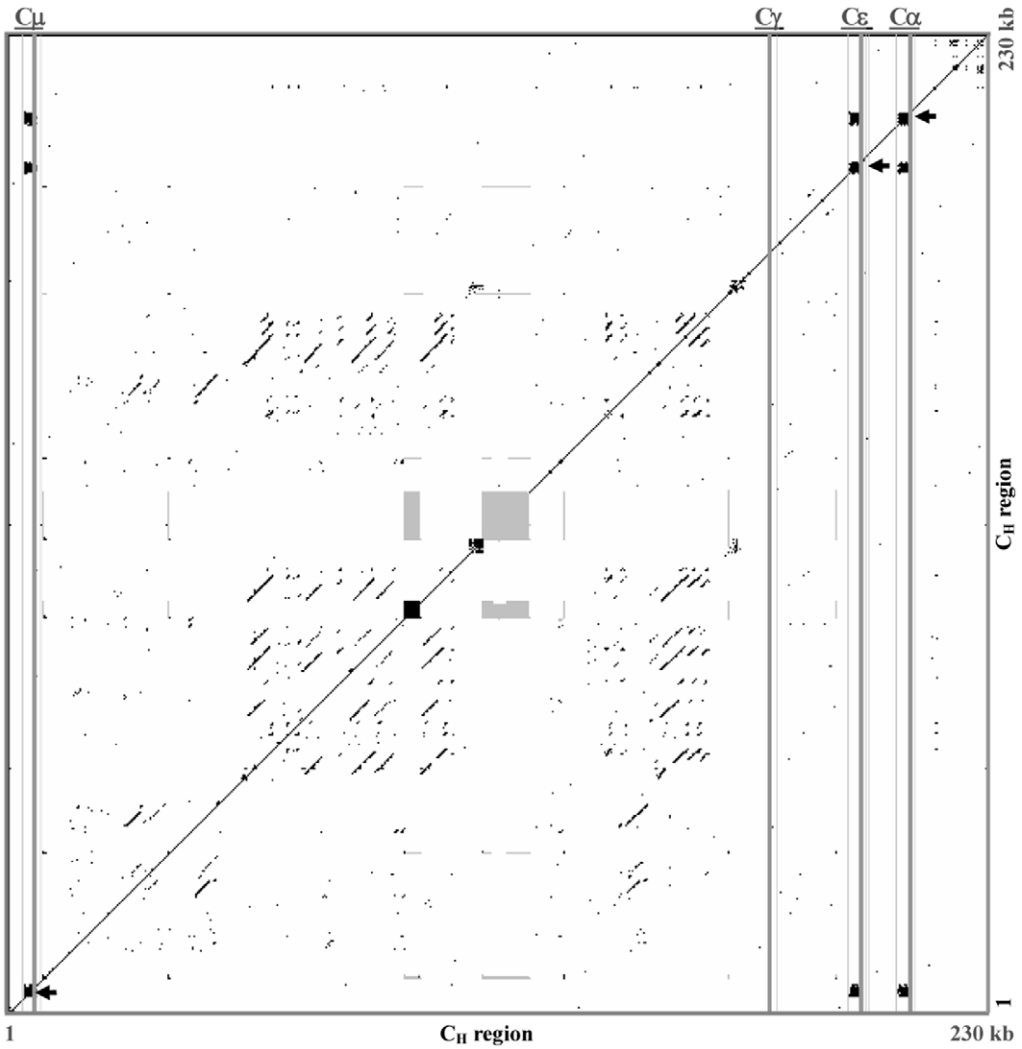

Figure 5. Dot plot comparison of the guinea pig C_H (μ, ε, γ and α) region. A dot matrix representing repetitive sequences of guinea pig C_H (μ, ε, γ and α) genes. Switch regions are indicated by black-squared boxes and marked with arrowheads, and gaps are indicated by grey-squared boxes. Positions of C_H genes are indicated as grey vertical lines. The dots represent homologies with a search length of 30 bp and maximum of 9 bp mismatches.
doi:10.1371/journal.pone.0039298.g005

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 V_H/V_L Gene Families

In mammals, germline V_H and V_L genes are categorized into different families according to their amino acid or nucleotide sequences similarity [22]. Sequences with greater than 75% similarity are general 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 V_H and V_L 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 J_H segments, 507 V_H, 41 D_H, four 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.

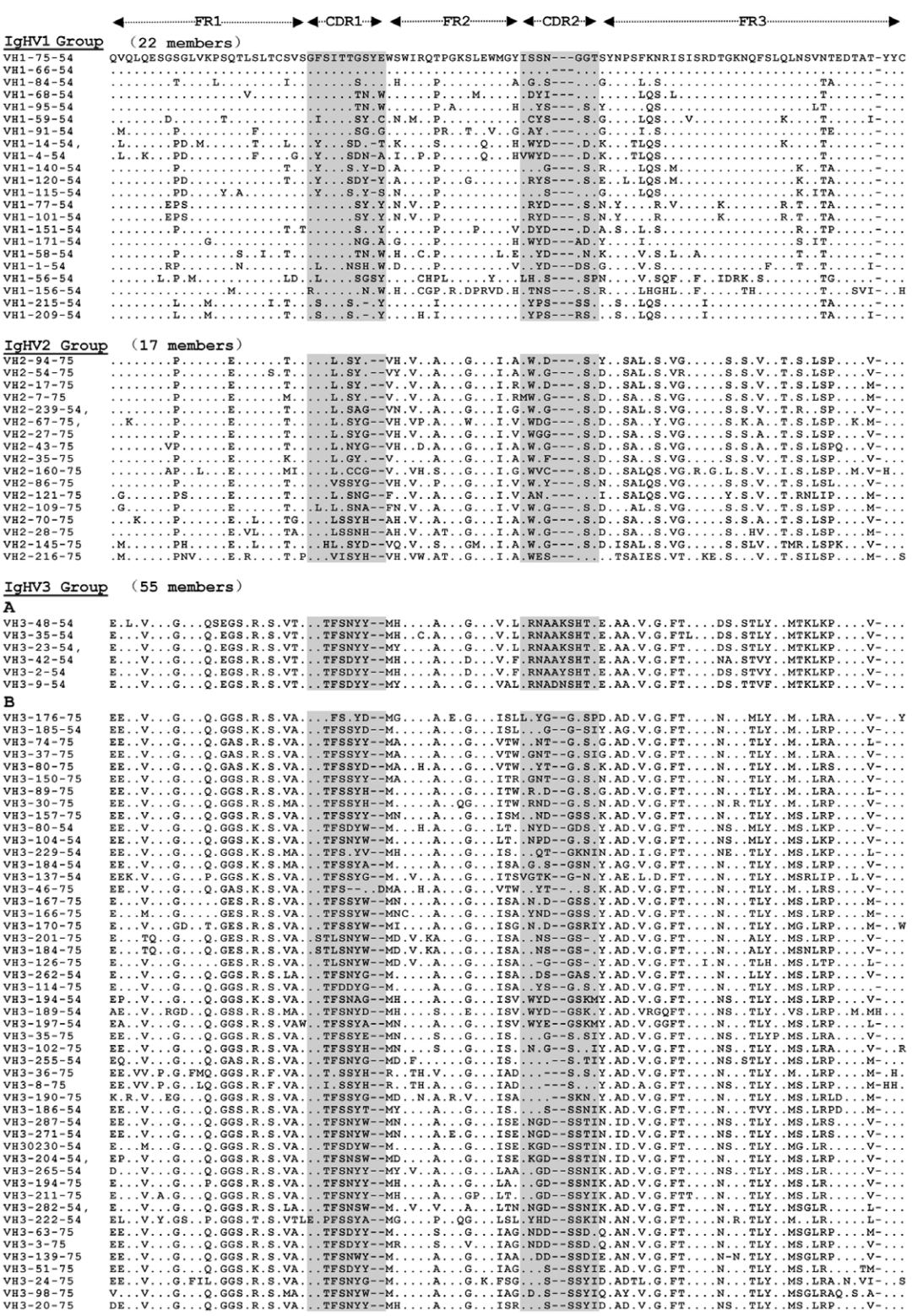

Figure 6. The alignment of deduced amino acid sequences of guinea pig 94 functional V_H genes. The sequences of functional V genes were analyzed by Megalign software and BioEdit programe. The Designation of framework regions (FR) and complementarity determining regions (CDR) referred to IMGT numbering system, and the CDR region are indicated by grey-squared background. amino acid sequences that are the same as the top segment, V_H1-75-54, 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 V_H genes. A phylogenetic tree of nucleotide sequences of 94 guinea pig potentially functional V_H segments was constructed. The three identified V_H gene families are labeled with Arabic numerals.
doi:10.1371/journal.pone.0039298.g007

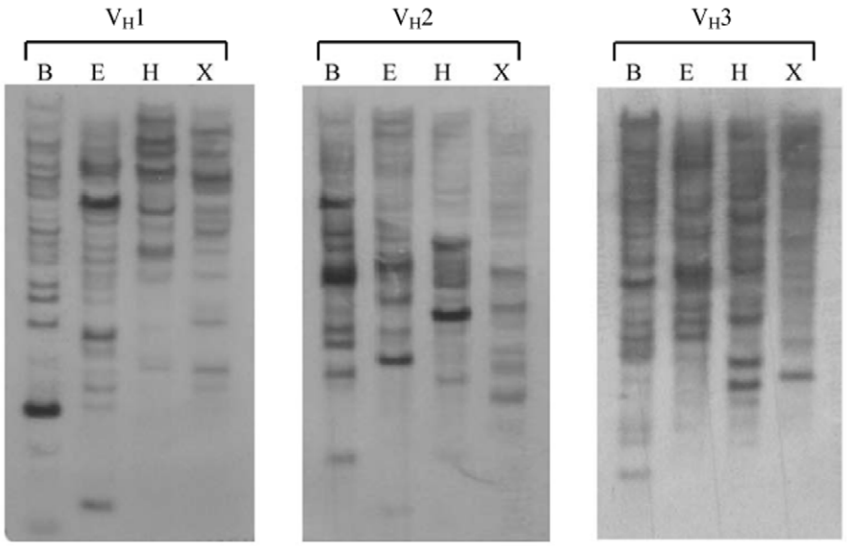

Figure 8. Southern blotting analysis of guinea pig genomic DNA. Southern blotting analysis of the guinea pig heavy chain variable region genes. Genomic DNA was digested with *Bam*H I (B), *Eco*R I (E), *Hin*d III (H) and *Xba* I (X), and hybridized with probes for each of three guinea pig families.
doi:10.1371/journal.pone.0039298.g008

Analysis of the Guinea Pig Constant Region Genes

To acquire the cDNA sequence for four heavy chain isotypes (μ, γ, ε, and α), 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_H3 and parts of the C_H4 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 an IgD C 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 δ trace fragments that are homologous with mammalian IgD (Figure 3) and showed an opposite transcription direction to the upstream μ gene. Based on sequence alignment, the three-fragment δ trace was found to belong to the C_H2 and C_H3 domains (Figure 3).

Although two IgG isotypes (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_H1 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 5′ 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 Cμ, Cε and Cα gene sequences of the guinea pig. Three putative S regions span 2.2 kb to 3 kb, and exhibit similar repeats (GAGCT and GGGCT) 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 S 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

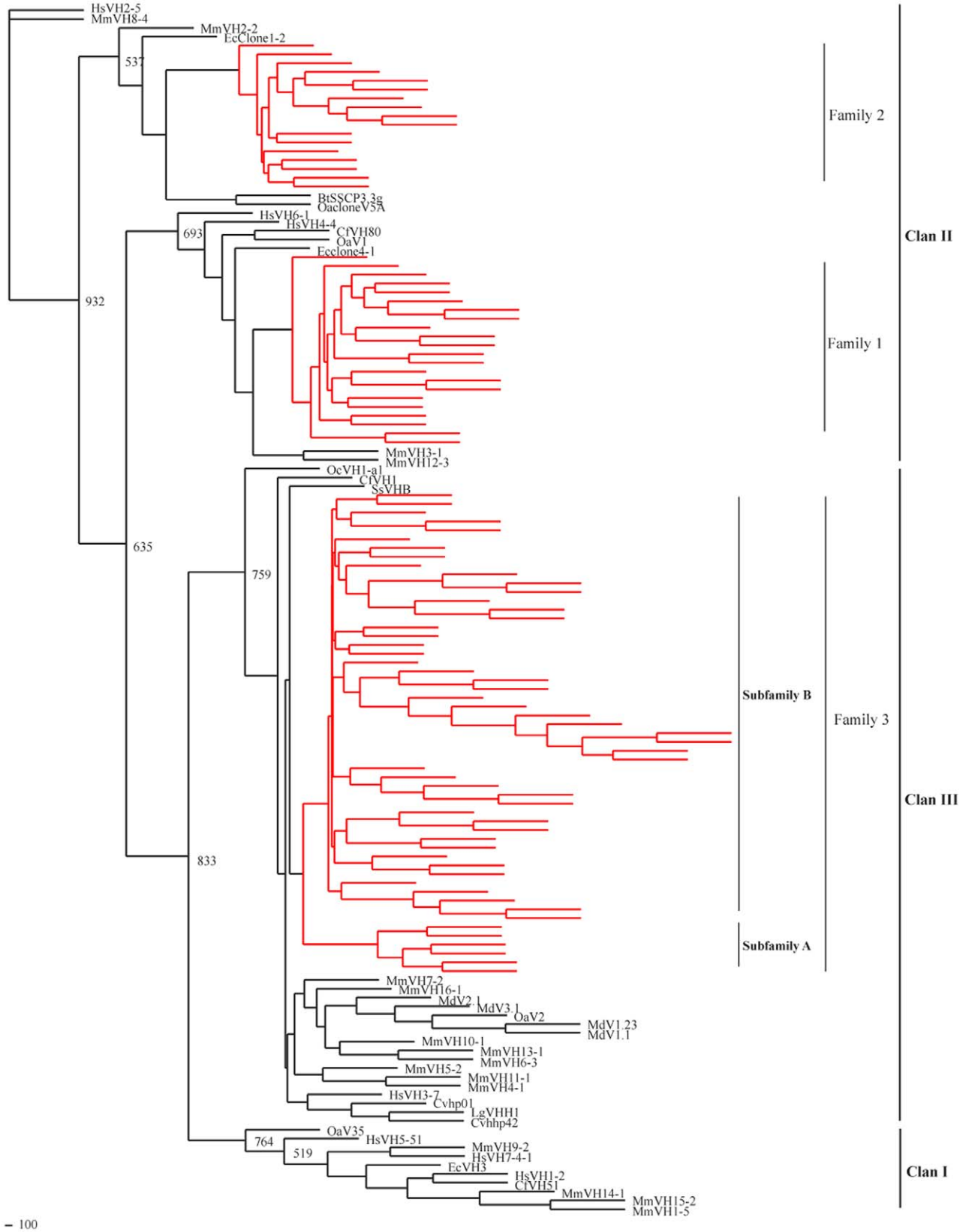

Figure 9. Phylogenetic analysis of mammalian V_H genes. Guinea pig sequences are represented by red branches, other species sequences are represented by black branches. Three V_H families of guinea pig were clustered respectively with other mammalian V_H families, and the V_H families of guinea pig belong to ClanII and ClanIII. The V_H3 family could be divided into two subfamilies.
doi:10.1371/journal.pone.0039298.g009

DH	Nonamer	12bp Spacer	Heptamer	Coding region of DH segment	Heptamer	12bp Spacer	Nonamer
D1	GTCTGATGC	TAAATATTATGA	CACTGTG	ATCTGAATATGCATAACTCTT	CACAGTG	TCTTTATTTCCA	AAACTAAAA
D2	AATGTGTGG	CCAGCATGGCTA	CACTGTG	ACTTCAAAGCTAGCCAGAGCAG	CATAGTG	AAACCCTATCTC	ACAACAAAT
D3	AGTTCATGC	TGGCCTTAAGCT	CACTTTG	TAGCCCAGGCTGGTCTCGAACTCTCAATGATCCTTCTACCTCAGCCTT	CCCAGTG	CTAGGTTTATAG	GCGTGAACC
D4	GGGATTTGG	TAGGGTCTGTGT	CACTGTG	GTATGAGTGTGATGGACATGGCTGCCTCACCACAGCAGGAATAC	CACAGCC	TCACAGCACGCA	TCCAAACCC
D5	AGGGCTTTT	TGTGAAGTCTCC	CACTGTG	TACCCTGATAACTAT	CACAGTG	AGGAACCCAGTA	GGAAAGCCA
D6	CCATATTGT	CAAGGCTTCTGT	CACTGTG	TATCTGGAGTAGCAGCCAC	CACAGTT	ATGGGCTCTGTG	CTCAAAACA
D7	GGTTTTTGC	TGATGTCAGTGT	CCCTGTG	TACTGAGATAGCTGGTTC	CTCAGTA	ATACCCAGTCAT	CCAGAAACC
D8	GAAATTTGT	GCAGGCCTGAAT	CACAGTG	ATACTATAGTGATGTGTTTGGTAGTTCC	CACAGTG	ATACAGGGCACA	CATAAAATT
D9	AGGGCTTTT	TGTGAAGTCCCC	CACTGTG	TACTCTGGTAGCTAC	CACAATG	AGGAACCTGGAA	GGAAAGCCC
D10	AATTTTTGT	GCTGGCCTGTGT	CACTGTG	AGAGTATTATAGCTACCATTATGGTGCTTTTATTCC	CACAGTG	ACACAGGCCCAC	TTCAAAAGC
D11	CATATTTGT	CACGGCTTCTGT	CACTGTG	TATATGGGTTGGAACTAC	CACAGTG	GTGTGCACTGTG	CTCAAAACA
D12	GGTTTTTGT	TGTAGTCTGTGT	CACTGTG	GATTGGTATAGCTGGACC	CACAGTG	ATACCCAGTCAG	ACAGAAACC
D13	GGAATTTGT	GCAGGCCTGAGA	CACAGTG	CTTATTTAGTAGATGGTTCT	CACAGTG	ACACAGATCACA	CATAAAACT
D14	CGATTCTGC	ATAGCTCCAGAT	CACAGTG	GGTATGGTGCTTTACTAC	CACAGTG	ACGGAGTTCACG	TGCAAAACA
D15	AATTTTTGT	GGTGCCATGTGT	CACCGTG	GGGTTATTATAGCTGGGGTACTTACTATATTCT	CACAGTG	ACACAGCCCTTC	TCTAAAACT
D16	GAGATTTAT	CATGGTCTGTGT	CACTGTG	GTAAGTGTGATGGATAC	CACAGCC	TCACAGCACCCA	TCGAAACCC
D17	GCTTTTTGT	GAAGGGCTCCCC	CACTGTG	GTACTACTGGAACTAC	CACACTG	ATGCATCCAGCA	GCAAAAACC
D18	AATTTTTGT	GCTGGACTGTGT	CACTGTG	AGGATATTATAGATATATGTGGGGTGGACCTTACTATACC	CACATTG	ACACAGCCCTAC	TCCAAAAGC
D19	GGAATTTGG	CAGGGTCTGTGT	TGCTGTG	GTATGAATGTGATTGCATATGGCTGCTTTACCACAGCAGGAATGC	CACAGTG	TCATAGCACCCA	TTCAAACCT
D20	AGGGCTTTT	TGTGAAGTCCCC	CACTGTG	TATTCTGGTAGCTAC	CACAGTG	AGGGACCTGGCA	GGAAAGCCC
D21	CCATATTGT	CAGGGTTTCTGT	TACTGTG	TATATGGGGTAGCAGCTGC	CACAGTG	ATATACACTCTG	CTGAAAACG
D22	GGAATTTGT	GTGGGCCTGACA	CACAGTG	GATCATATACAGTAGTTGGTACTATAGTCT	CGCAGTG	ATGCAGGCCAAG	CATAAAACT
D23	AATTTTTGT	GCTGGCCTGTGT	CACTGTG	AGATTATTATGGCTACGATTATACC	CACAGTG	ACAGGGACCCAC	TCCAAAAGC
D24	CCATATTGT	CAGGGCTTCTGT	CACTGTG	TGTATGGTATAGCAACTAC	CACCGTG	GTGTGCACTGTG	CTCAAAACA
D25	GGTTTTTGT	TGCAGTCTGTGT	CACTGTG	TATTGGGGTAGCTGGAAC	CACAGTG	ATACCTAGTCAG	ACAGAAACC
D26	GGAATTTGT	GCAGGCCTGAGA	CACAGTG	GTTATTATAGTGGTAGTGGATGGGTCT	CACTGTG	ACACAGGTCACA	CATAAAACT
D27	AGATTCTGC	ATAGCTCCAGTT	CACAGTG	GGTATGGTGGTTACTAC	CACAGTG	AGAGAGGTCAGG	TGCAAAACA
D28	AATTTTTGT	GGTGCCATGTGT	CACCGTG	GGGTTATTATAGCTCGGATGGTTACTATAGTCT	CACAGTG	ACACAGCCCCTC	TCTAAAACT
D29	GCTTTTTGT	GAAGGGCTCCCC	CACTGTG	GTACTACGATGACTAC	CACACTG	ATGCATCCAGCA	GCAAAAACC
D30	GCAATTTGG	TAGGGTCTGTGT	CAGTGTG	GTATGAGCGTGATGGACATGGCTGCCTTACCACAGCAGGAATAC	CACAGCT	TCACAGCACCCA	TCAAAACCC
D31	AATTTTTGT	GCTGGCCTGTGT	CACTGTG	AAGATATTATAAATATATGTGGGGTGGATCTTACTACAAC	CACAGTG	ACACAGCCTTAC	TCCAAAAGT
D32	CCATATTGT	CAGGGCTTCTAT	CACTGTG	TACATTGGGTAGCATCTGC	CACAGTG	ATGTGCACTCTC	CTGAAAAGG
D33	AGATTCTGC	CTATCTCCCAAT	CACAGTG	GGTATATTGGTAGCTAA	TGCAGTG	AAAGAGGTCAGA	TCCAAAACT
D34	AATTTTTGT	GCTGGCCTGTGT	CACTGTG	AGAGTATTATAGC	TACAGTG	ATGGTACTTACT	ATTATACCC
D35	AGTTTTTGT	TGCAGTCTGTGT	CACTGTG	GATTGGAGTACCTGGACC	CACAGTG	ATACCCAGTCAG	ACAGAAATC
D36	GGAATTTGT	GCAGGCCTGAGA	CACAGTG	GTTATTATATTGGTGGTGCATGGTTCT	CACAGTG	ACACAGGTCACA	CATAAAACT
D37	AATTTTTGT	GGTGCCATGTGT	CACCGTG	TGGTTATTATAGCTCGGGTGTTTACTATATTCT	CACAGTG	ACACAGCCCCTC	TCTAAAACT
D38	GAGATTTAT	CAGGGTCTGTGT	CACTGTG	GTAAGTGTGATGGATAC	CACAGCC	TCACAGCACCCA	TCGAAACCC
D39	GCTTTTTGT	GAAGGGCTCCCC	CACTGTG	GTACTACGATGACTAC	CACACTG	ATGCATCCAGCA	GCAAAAACC
D40	GGAATTTGT	GTAGGCTTGAGA	CACAGTG	GGTATGGTGGTACTATGGTCT	CACAGTG	ATGCAAGCCACA	CATAAAACG
D41	GTATTCTGC	AAAGCTCCAAAT	CACAGTG	TATAGCTAC	CACAGTG	AGATAAGCCAGG	TCCAAAACT

Figure 10. The nucleotide sequences of D_H genes. D_H coding region nucleotide sequences are represented together the RSS elements (nonamer and heptamer sequences).
doi:10.1371/journal.pone.0039298.g010

largest family, V_H3, could be further divided into two subfamilies (Figure 6, Figure 7). We also performed Southern blotting to verify the multiple numbers of V_H genes of different families in the guinea pig genome (Figure 8).

We chose all potentially functional guinea pig V_H sequences of and V_H 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 V_H gene families were classified into three clans [33]. The guinea pig V_H families 1 and 2 belonged to clan II, and family 3 belonged to clan III.

Analysis of the Guinea Pig D_H and J_H Gene Segments

Approximately 504 kb downstream from the last V_H segment (V_H1-1), we identified 41 D_H segments (i.e., D_H1–D_H41). They spanned a 660 kb region of DNA in scaffold 54 (Figure 1). Each

	Nonamer	Spacer	Heptamer	JH region	RNA donor splicing site
JH1	AGTTTTTGT	gtggacaatggcagagcaagtgt	CTCAGTG	AATGCCTTTGGTTTCTGGGGCCAGGGCACCTCAGTCACCGTCACGTCA N A F G F **W** G Q G T S V T V T S	**GGTAA**
JH2	GGTTTTTGT	atactccataagggagcctacag	CATTGTG	AACTTTCAATACTGGGGCCAGGGGACCCTGGTCACCGTCTCCTCA N F Q Y **W** G Q G T L V T V S S	**GGTGA**
JH3	GGGTTTTGC	ctggggtcttgccagggttgtta	CAATGTG	AACCAATTTAACCACTGGGGCCCTCCTATCCTGGTCACTATCTCCTCA N Q F N H **W** G P P I L V T I S S	**GGTGA**
JH4	GGTTTTTGT	ataccctctaaagaggcccataa	CAGTGTG	TACTTTGATGTATGGGGCGCTGGAACACTGGTCACCGTCTCCTCA Y F D V **W** G A G T L V T V S S	**GGTGA**
JH5	AGTTCTTGT	ttgggatcctgtcatggttgtca	CAAGGTG	AACTGGTTTGACAACTGGGGCCGAGGGGTCTCGGTCACCGTCTCCTCA N W F D N **W** G R G V S V T V S S	**GGTGA**
JH6	GGGTTTCTG	tggggtgtggatggagcgtgcac	TACTGTG	TACTATGCTTTGGATATCTGGGGCCCTGGCACCCTGGTCACCGTCTCCTCA Y Y A L D I **W** G P G T L V T V S S	**GGTAA**

Figure 11. The nucleotide sequences of J_H genes. J_H nucleotide sequences are represented together the RSS elements (nonamer and heptamer sequences) and RNA donor splicing site.
doi:10.1371/journal.pone.0039298.g011

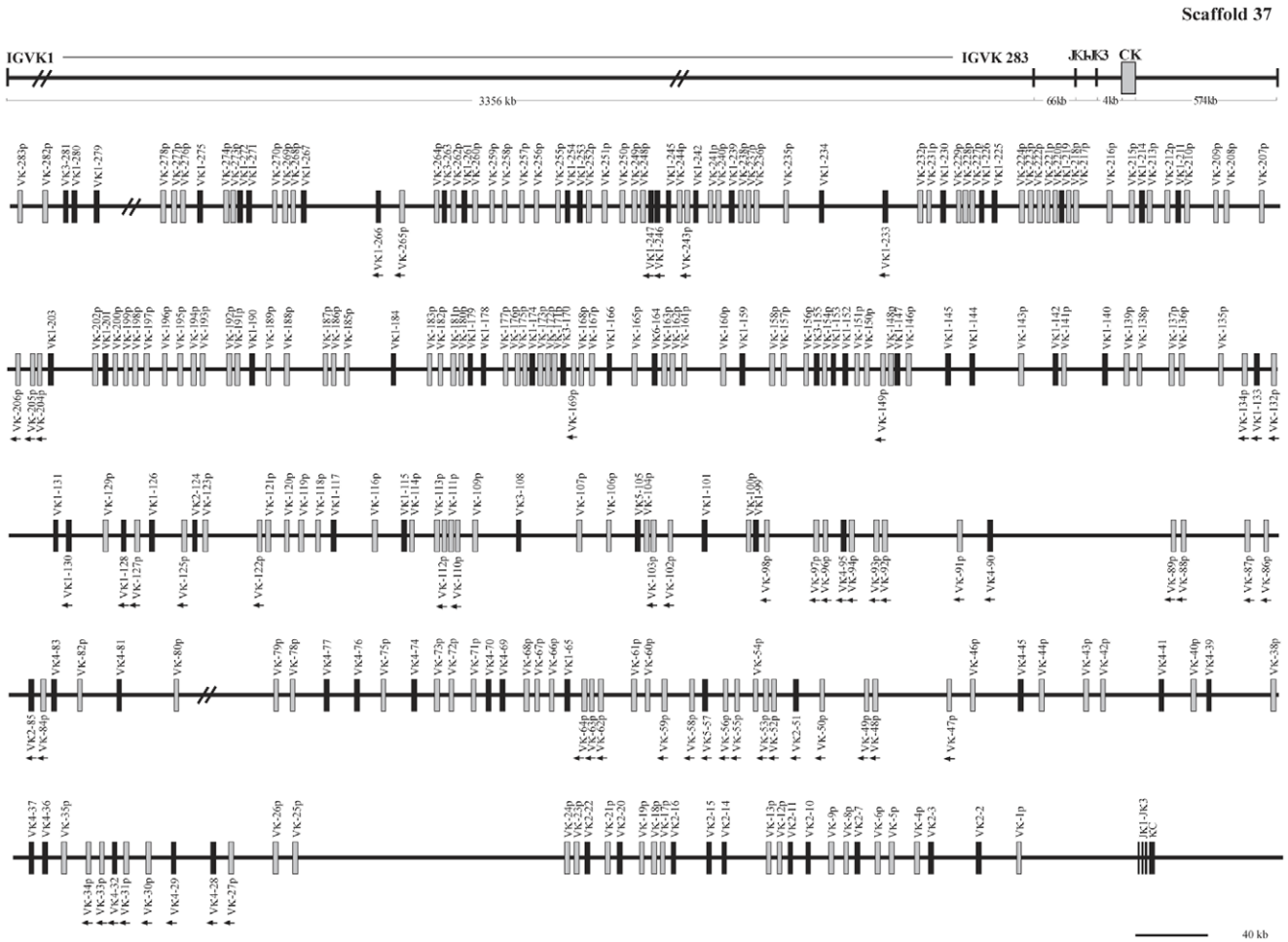

Figure 12. The guinea pig Igκ locus in scaffold 37. The guinea pig Igκ locus is in scaffold 37, the potentially functional V_κ gene segments are shown in filled bars, and pseudogenes are represented by open bars. The unidirectional arrowheads below V_κ gene segments on scaffold 37 indicate that their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

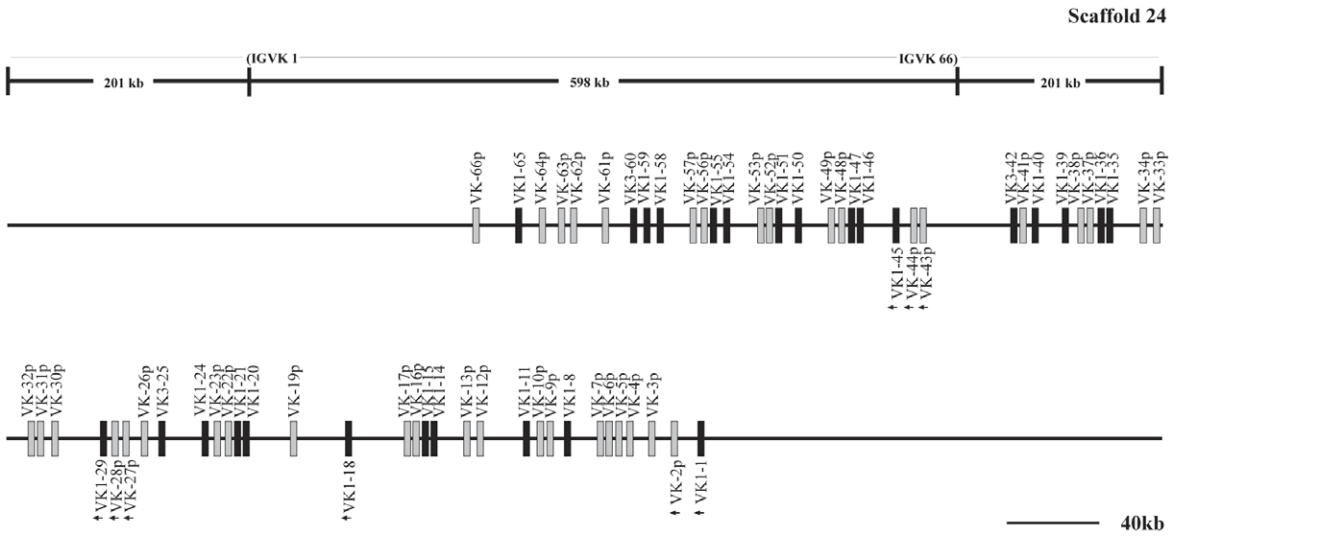

Figure 13. The guinea pig Igκ locus in scaffold 24. The guinea pig Igκ locus is in scaffold 24, the potentially functional V_κ gene segments are shown in filled bars, and pseudogenes are represented by open bars. The unidirectional arrowheads below V_κ gene segments merely indicate a transcriptional direction different from that of the remaining V_κ gene segments in scaffold 24. The scale does not apply the upper frame diagram.
doi:10.1371/journal.pone.0039298.g013

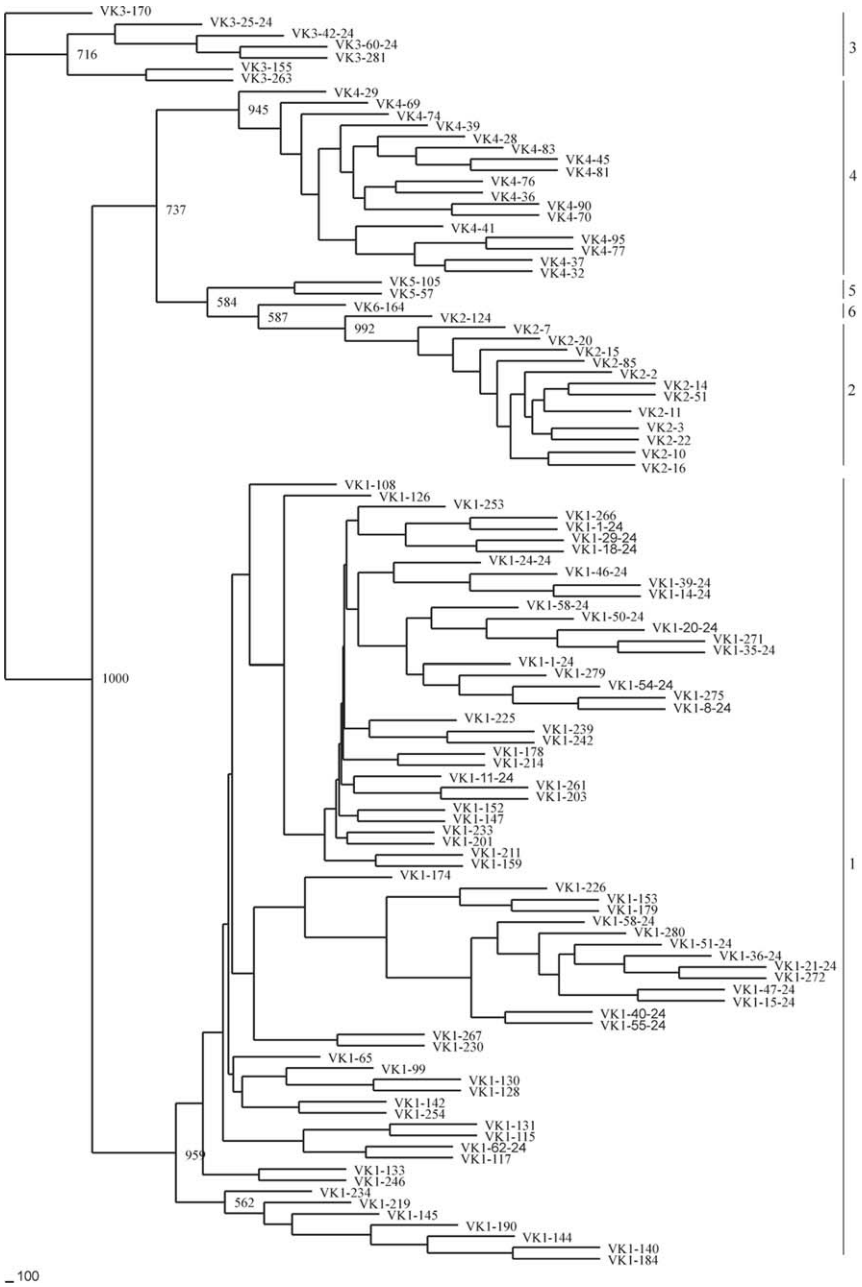

Figure 14. Phylogenetic analysis of the 111 guinea pig Vκ genes. A phylogenetic tree of the nucleotide sequences of 111 guinea pig V_κ segments were constructed. The six V_κ gene families are labeled with Arabic numerals.
doi:10.1371/journal.pone.0039298.g014

D_H segment was flanked by conserved RSS elements composed of heptamers and nonamers separated by 12 bp spacers and existed within at last one alternative reading frame (Figure 10 and Figure S8), suggesting that they are potentially functional.

J_H region contained six genes (designated J_H1 to J_H6) spanning approximately 2 kb. Each J_H 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 Igκ Locus

The guinea pig Igκ 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_κ genes were identified. Further analysis revealed that 111 V_κ 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_κ genes were divided into six families ($V_\kappa1$–$V_\kappa6$), which contained 69, 15, 7, 17, 2, and 1 member/s, respectively (Figure 14). In addition, 222 V_κ genes were arranged in the same transcriptional orientation and exhibited downstream J_κ and C_κ, and 61 V_κ segments and a reverse transcriptional direction in scaffold 37. Downstream of the V_κ genes, three J_κ gene segments

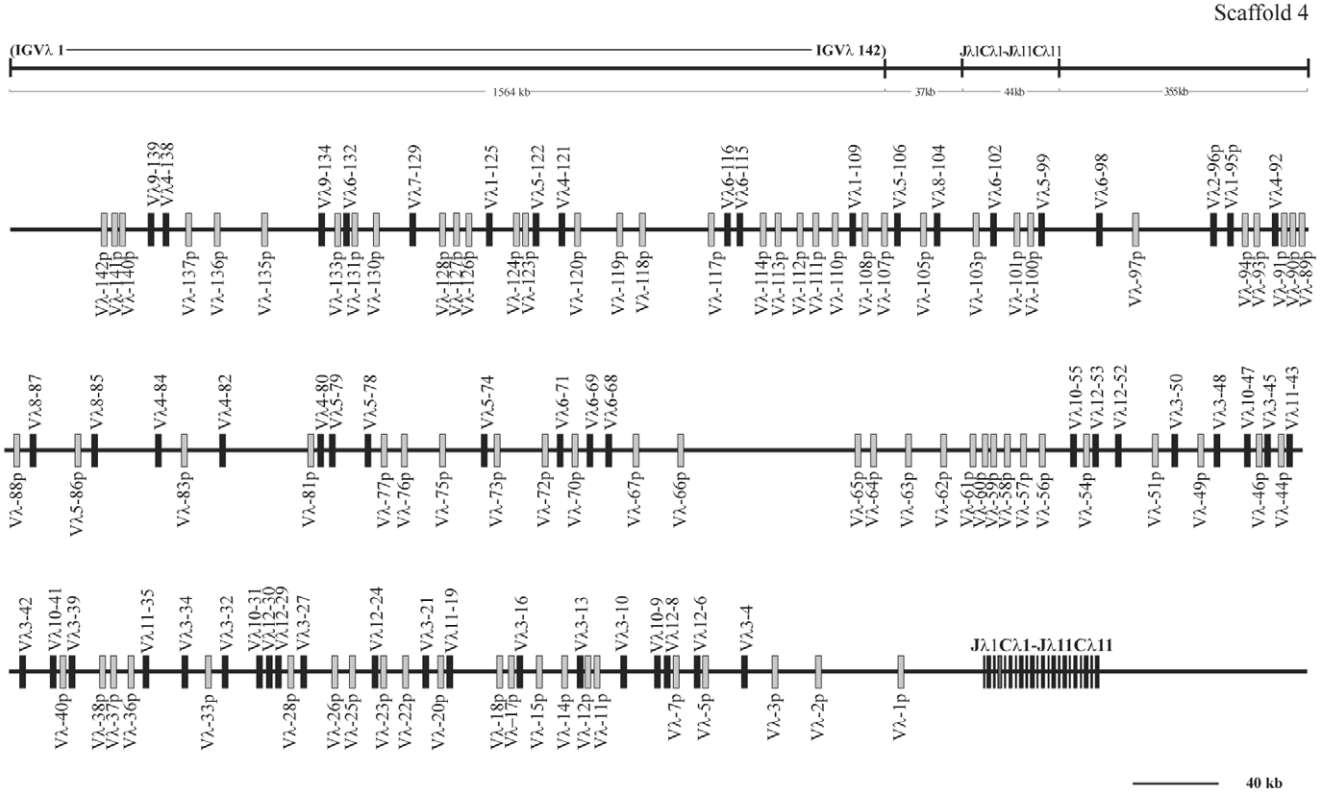

Figure 15. The guinea pig Igλ locus. The guinea pig Igλ locus is in scaffold 4, the potentially functional V_λ gene segments are shown in filled bars, and pseudogenes are represented by open bars.
doi:10.1371/journal.pone.0039298.g015

were identified that spanned 0.6 kb. Furthermore, approximately 4 kb downstream from the last J_κ, a single C_κ gene was identified.

Guinea Pig Igλ Locus

Guinea pig Igλ chain genes were identified in scaffold 4, and spanned an approximately 1,642 kb length (Figure 15, Table S5). Of 142 germline V_λ 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_λ genes were assigned to twelve families (Figure 16), comprised of 3, 1, 13, 6, 6, 8, 1, 3, 2, 5, 3 and 7 member/s, respectively. In contrast to the V_κ genes, all the V_λ genes have the same transcriptional orientation as the downstream J_κ and C_κ regions. At the 3′ 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_λ and C_λ 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 V_H (507)-D_H (41)-J_H (6)-Cμ-ψδ-Cγ2-Cε-Cα 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 μ, γ, ε and α gene within the guinea pig genome. We also found three fragments of δ gene in an opposite direction downstream from the μ gene. Due to sequence gaps, it is not certain if an additional functional δ gene also exists in this species. We have also tried to confirm the sequence and orientation of the δ gene fragments by genomic PCR to eliminate assembly error. The sequences of the δ gene fragments were verified. Because two genomic fragments (approximately two kb and six kb) between μ and δ gene can not be successfully amplified, so the orientation of the δ gene fragments remains a question.

Many placental mammals, such as human, cow, sheep, horse and dog, have a single functional δ gene. Except for a functional δ gene which consists of ten C_H domains, a reverse δ pseudogene was previously observed in the platypus IgH locus [53], while in the elephant genome, only one Cδ3 remnant fragment was identified [51]. In camelids, Cδ3 exon appears to be highly mutated [54]. In the rabbit [24] and opossum (marsupial) [25], the δ gene has clearly been shown to be missing in their genomes, indicating that the δ gene might be not as essential as the μ gene in the humoral immunity.

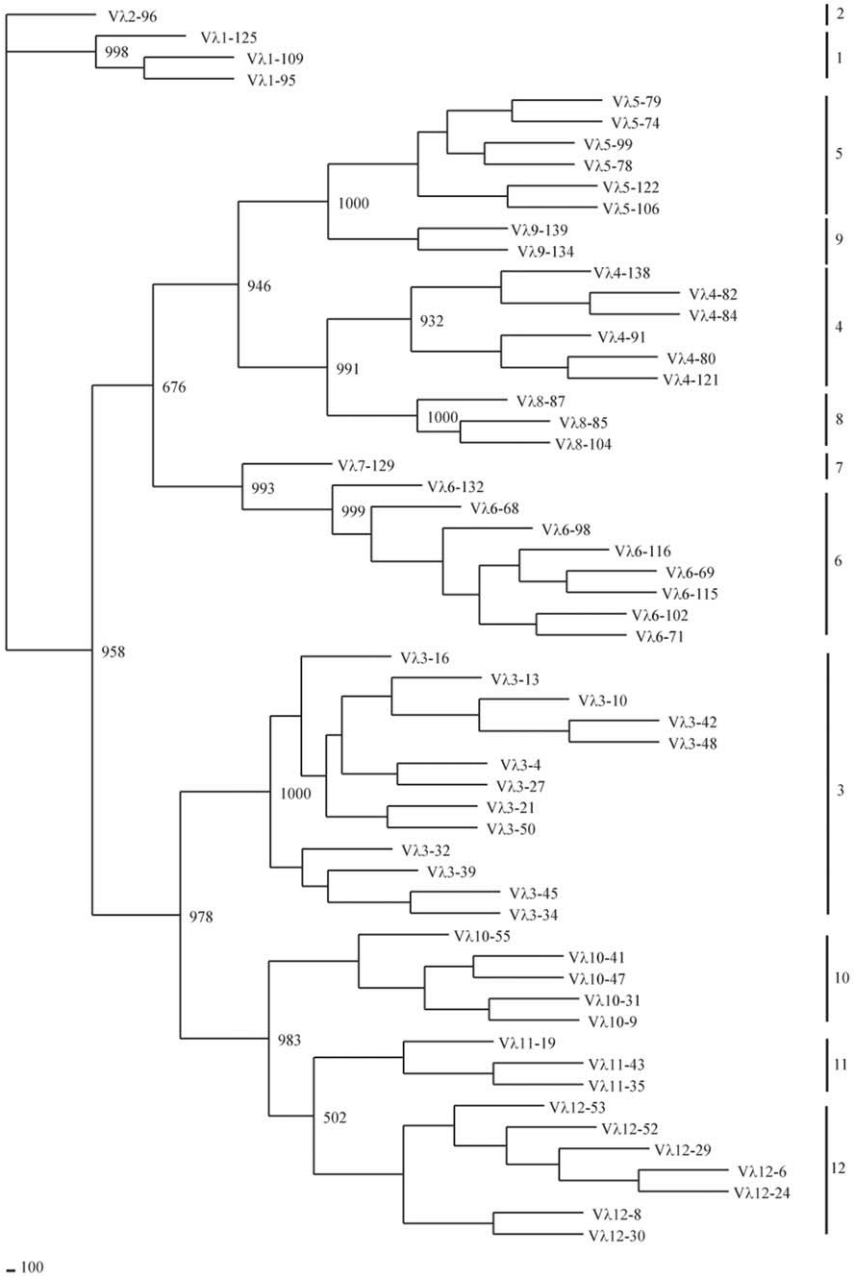

Figure 16. Phylogenetic tree analysis of the 58 guinea pig V_λ genes. A phylogenetic tree of the nucleotide sequences of 58 guinea pig V_λ segments were constructed. The eleven V_λ gene families are labeled with Arabic numerals.
doi:10.1371/journal.pone.0039298.g016

IgG is an important antibody molecule, which is believed to have initially evolved 600 million years ago [55]. The structure of γ gene usually contains three C_H 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 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_H domains. It has been postulated that different γ subclasses are derived from gene duplications of ancestral γ gene in mammals [63]. Three ancestral γ genes of mouse and rat can evolve multiple γ genes by gene duplications [64,65], and the divergence of the γ genes of human depends on one ancestral γ gene and duplicated Cγ-Cγ-Cε-Cα 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

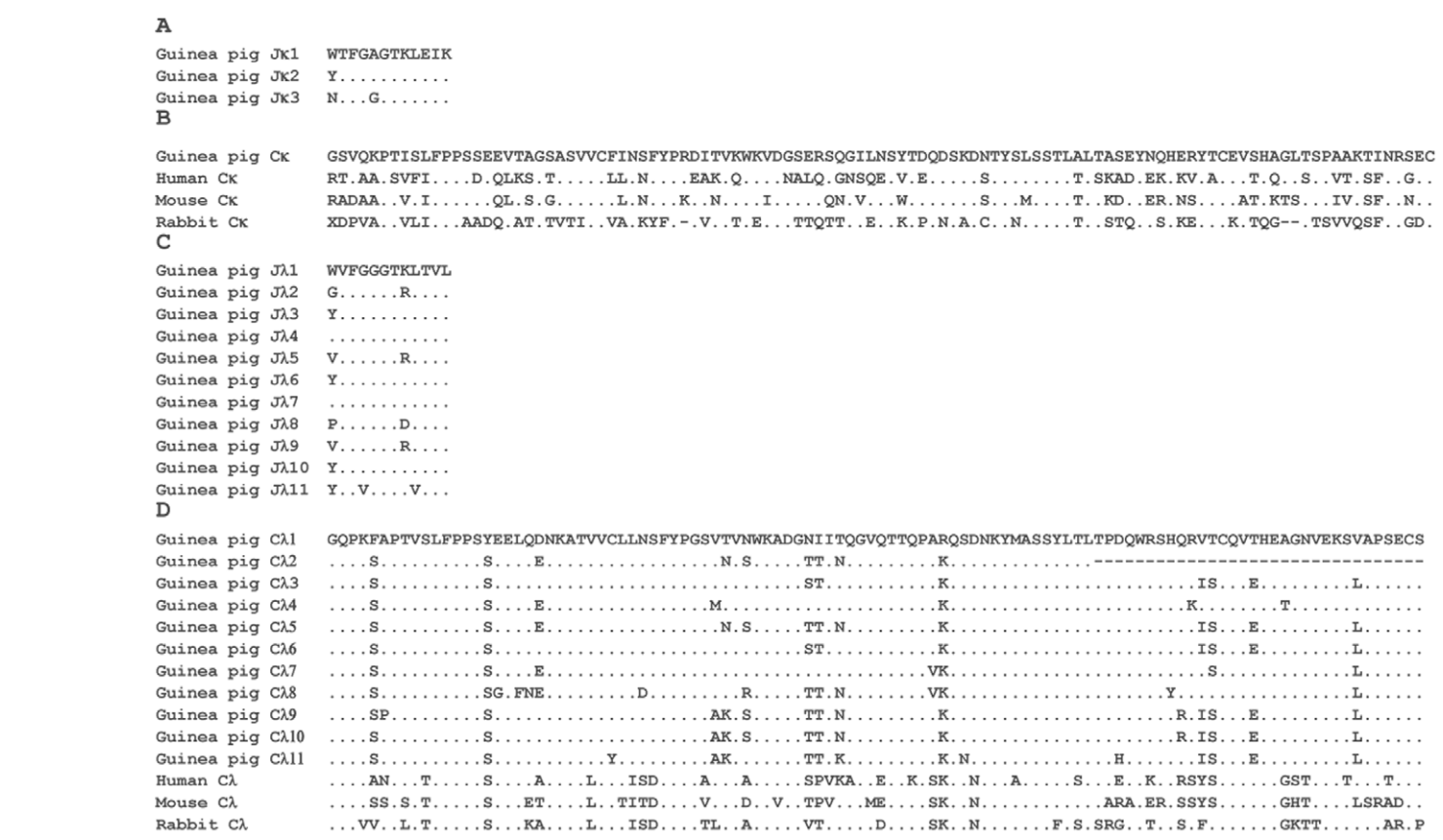

Figure 17. The alignment of amino acid sequences of J and C genes from guinea pig IgL chains. A, Alignment of the deduced amino acid sequences of the three guinea pig J_κ gene segments. B, Alignment of the amino acid sequences of the C_κ proteins from human, mouse and rabbit. C, Alignment of the deduced amino acid sequences of the eleven guinea pig J_λ gene segments. D, Alignment of the amino acid sequences of the C_λ proteins from human, mouse and rabbit. 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.g017

the total V_H genes (413 pseudogenes vs. 507 total genes). The high frequency of gene duplication in variable region generated multiple V_H copies, many of which became pseudogenes due to genomic drift [68]. The V_H 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 V_H gene segments were also divided into three gene families (families 1, 2 and 3), which were orthologous to human V_H 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 (sixteen 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 V_H 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 V_H genes belong to all three V_H 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 V_H genes only belong to clan II. Pigs and rabbits only have clan III genes [74,77,79]. V_H genes of guinea pig are distributed in clans II and III, and the majority of V_H genes (family 3) are most closely related to human V_H3 (clan III), which has been proposed as the ancestral V_H gene family [78,80]. These features are also found in swine and rabbits [74,76].

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 (V_H3 and V_H4), While the American short-tailed opossum (*Monodelphis domestica*), swine, rabbit and human share ancestral V_H3 family, and artiodactyl species share ancestral V_H4 family. With an older evolutionary origin in present day mammals [86], platypus have two ancestral V_H gene families, while other mammals share one or two ancestral V_H gene families. These could be explained by an inactivation or loss of V_H gene members in these species during evolution [25]. For new V_H gene families in human or mouse, the divergence of V_H genes probably occurred after speciation.

The ratio of functional V_κ and V_λ is variable within mammalian species, with the germline V_κ genes being more abundant than V_λ genes in humans (40 V_κ genes *vs.* 30 V_λ genes) and mice (V_κ genes *vs.* V_λ genes over 95%) [87]. This is also the case for the guinea

pig, in which V_κ germline genes are more dominant than V_λ (84 functional V_κ genes *vs.* 58 functional V_λ 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 gene segments [87]. It is possible that the κ chain preponderates over the λ chain at the protein level in guinea pigs. Also, multiple pseudogenes exist in the guinea pig V_H (413), V_κ (238), and V_λ (84) 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.
(DOC)

Figure S2 Multiple sequence alignments of guinea pig V_κ genes.
(DOC)

Figure S3 Multiple sequence alignments of guinea pig V_λ genes.
(DOC)

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 Ig μ, γ, ε and α genes in scaffold 54. Four guinea pig constant region encoding genes of tansmembrane type for IgM, IgG, IgE and IgA were identified by bioinformatics analysis.
(TIF)

Figure S6 Alignment of the IgG amino acid sequences of the guinea pig. Alignment of IgG1, IgG2 and IgG squence of guinea pig. Dots indicate similar residues as in G1 and G, whereas dashes indicate gaps introduced for optimal alignment. The diference 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 *Bam*H I (B), *Eco*R I (E), *Hin*d 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′: CACTGTG 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.
(DOC)

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

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

Table S4 The guinea pig immunoglobulin κ light chain DNA segments location in scaffolds 37 and 24.
(RAR)

Table S5 The guinea pig immunoglobulin λ light chain DNA segments location in scaffold 4.
(RAR)

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. Padilla-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 Ganoderma lucidum on pollen-induced biphasic nasal blockage in a guinea pig model of allergic rhinitis. Phytother Res 26: 325–332.
3. Ikeda R, Nakaya K, Oshima H, Oshima T, Kawase T, et al. (2011) Effect of aspiration of perilymph during stapes surgery on the endocochlear potential of guinea pig. Otolaryngol Head Neck Surg 145: 801–805.
4. Sturm EM, Schuligoi R, Konya V, Sturm GJ, Heinemann A (2011) Inhibitory effect of prostaglandin I2 on bone marrow kinetics of eosinophils in the guinea pig. J Leukoc Biol 90: 285–291.
5. Helling K, Wodarzcyk K, Brieger J, Schmidtmann I, Li H, et al. (2011) Doxycycline reduces nitric oxide production in guinea pig inner ears. Auris Nasus Larynx 38: 671–677.
6. Andersson KE, Soler R, Fullhase C (2011) Rodent models for urodynamic investigation. Neurourol Urodyn 30: 636–646.
7. 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 model of experimental allergic rhinitis. Eur J Pharmacol 667: 389–395.
8. Andrianova EP, Krementsugskaia SR, Lugovskaia NN, Mayorova TK, Borisov VV, et al. (2011) Foot and mouth disease virus polyepitope protein produced in bacteria and plants induces protective immunity in Guinea pigs. Biochemistry (Mosc) 76: 339–346.
9. Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, et al. (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374: 168–173.
10. Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, et al. (1993) Naturally occurring antibodies devoid of light chains. Nature 363: 446–448.
11. Freedman MH (1998) Limit, logic, and computation. Proc Natl Acad Sci U S A 95: 95–97.
12. Zhang YA, Nie P, 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: 81–90.
13. Thomas RK, Re D, Wolf J, Diehl V (2004) Part I: Hodgkin's lymphoma–molecular biology of Hodgkin and Reed-Sternberg cells. Lancet Oncol 5: 11–18.
14. 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: 156–165.
15. Sambrook J, Russell DW (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
16. Clewley JP, Arnold C (1997) MEGALIGN. The multiple alignment module of LASERGENE. Methods Mol Biol 70: 119–129.
17. 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.
18. Higgins DG, Thompson JD, Gibson TJ (1996) Using CLUSTAL for multiple sequence alignments. Methods Enzymol 266: 383–402.
19. Hall T (1990) 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) PHYLIP (Phylogeny Inference Package) Ver. 3.67 [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. Comput Appl Biosci 12: 357–358.
22. Schroeder HW, Jr., Hillson JL, Perlmutter RM (1990) Structure and evolution of mammalian VH families. Int Immunol 2: 41–50.
23. Brodeur PH, Riblet R (1984) The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V 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, Olp JJ, Miller RD (2009) On the genomics of immunoglobulins in the gray, short-tailed opossum Monodelphis domestica. Immunogenetics 61: 581–596.
26. Phizackerley RP, Wishner BC, Bryant SH, Amzel LM, Lopez de Castro JA, et al. (1979) Three dimensional structure of the pFc' fragment of guinea pig IgG1. Mol Immunol 16: 841–850.
27. Warr GW, Magor KE, Higgins DA (1995) IgY: clues to the origins of modern antibodies. Immunol Today 16: 392–398.
28. Tucker PW, Slightom JL, Blattner FR (1981) Mouse IgA heavy chain gene sequence: implications for evolution of immunoglobulin hinge axons. Proc Natl Acad Sci U S A 78: 7684–7688.
29. Kawamura S, Saitou N, Ueda S (1992) Concerted evolution of the primate immunoglobulin alpha-gene through gene conversion. J Biol Chem 267: 7359–7367.
30. Osborne BA, Golde TE, Schwartz RL, Rudikoff S (1988) Evolution of the IgA heavy chain gene in the genus Mus. Genetics 119: 925–931.
31. Sitnikova T, Su C (1998) Coevolution of immunoglobulin heavy- and light-chain variable-region gene families. Mol Biol Evol 15: 617–625.
32. Almagro JC, Martinez L, Smith SL, Alagon A, Estevez J, et al. (2006) Analysis of the horse V(H) repertoire and comparison with the human IGHV germline genes, and sheep, cattle and pig V(H) sequences. Mol Immunol 43: 1836–1845.
33. Tomlinson IM, Walter G, Marks JD, Llewelyn MB, Winter G (1992) The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops. J Mol Biol 227: 776–798.
34. Benton MJ, Donoghue PC (2007) Paleontological evidence to date the tree of life. Mol Biol Evol 24: 26–53.
35. Churakov G, Sadasivuni MK, Rosenbloom KR, Huchon D, Brosius J, et al. (2010) Rodent evolution: back to the root. Mol Biol Evol 27: 1315–1326.
36. Horner DS, Lefkimmiatis K, Reyes A, Gissi C, Saccone C, et al. (2007) Phylogenetic analyses of complete mitochondrial genome sequences suggest a basal divergence of the enigmatic rodent Anomalurus. BMC Evol Biol 7: 16.
37. Graur D, Hide WA, Li WH (1991) Is the guinea-pig a rodent? Nature 351: 649–652.
38. Li WH, Hide WA, Zharkikh A, Ma DP, Graur D (1992) The molecular taxonomy and evolution of the guinea pig. J Hered 83: 174–181.
39. D'Erchia AM, Gissi C, Pesole G, Saccone C, Arnason U (1996) The guinea-pig is not a rodent. Nature 381: 597–600.
40. Reyes A, Pesole G, Saccone C (2000) Long-branch attraction phenomenon and the impact of among-site rate variation on rodent phylogeny. Gene 259: 177–187.
41. Cao Y, Adachi J, Yano T, Hasegawa M (1994) Phylogenetic place of guinea pigs: no support of the rodent-polyphyly hypothesis from maximum-likelihood analyses of multiple protein sequences. Mol Biol Evol 11: 593–604.
42. Kuma K, Miyata T (1994) Mammalian phylogeny inferred from multiple protein data. Jpn J Genet 69: 555–566.
43. Robinson-Rechavi M, Ponger L, Mouchiroud D (2000) Nuclear gene LCAT supports rodent monophyly. Mol Biol Evol 17: 1410–1412.
44. Lin YH, McLenachan PA, Gore AR, Phillips MJ, Ota R, et al. (2002) Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. Mol Biol Evol 19: 2060–2070.
45. Warr GW (1997) The adaptive immune system of fish. Dev Biol Stand 90: 15–21.
46. Bengten E, Wilson M, Miller N, Clem LW, Pilstrom L, et al. (2000) Immunoglobulin isotypes: structure, function, and genetics. Curr Top Microbiol Immunol 248: 189–219.
47. 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.
48. Chevillard C, Ozaki J, Herring CD, Riblet R (2002) A three-megabase yeast artificial chromosome contig spanning the C57BL mouse Igh locus. J Immunol 168: 5659–5666.
49. Hendricks J, Terpstra P, Dammers PM, Somasundaram R, Visser A, et al. (2010) Organization of the variable region of the immunoglobulin heavy-chain gene locus of the rat. Immunogenetics 62: 479–486.
50. Berens SJ, Wylie DE, Lopez OJ (1997) Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. Int Immunol 9: 189–199.
51. Guo Y, Bao Y, Wang H, Hu X, Zhao Z, et al. (2011) A preliminary analysis of the immunoglobulin genes in the African elephant (Loxodonta africana). PLoS One 6: e16889.
52. 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.
53. Gambon-Deza F, Sanchez-Espinel C, Magadan-Mompo S (2009) The immunoglobulin heavy chain locus in the platypus (Ornithorhynchus anatinus). Mol Immunol 46: 2515–2523.
54. Achour I, Cavelier P, Tichit M, Bouchier C, Lafaye P, et al. (2008) Tetrameric and homodimeric camelid IgGs originate from the same IgH locus. J Immunol 181: 2001–2009.
55. Ellison JW, Berson BJ, Hood LE (1982) The nucleotide sequence of a human immunoglobulin C gamma1 gene. Nucleic Acids Res 10: 4071–4079.
56. Knight KL, Burnett RC, McNicholas JM (1985) Organization and polymorphism of rabbit immunoglobulin heavy chain genes. J Immunol 134: 1245–1250.
57. Dufour V, Malinge S, Nau F (1996) The sheep Ig variable region repertoire consists of a single VH family. J Immunol 156: 2163–2170.
58. Knight KL, Suter M, Becker RS (1988) Genetic engineering of bovine Ig. Construction and characterization of hapten-binding bovine/murine chimeric IgE, IgA, IgG1, IgG2, and IgG3 molecules. J Immunol 140: 3654–3659.
59. Flanagan JG, Lefranc MP, Rabbitts TH (1984) Mechanisms of divergence and convergence of the human immunoglobulin alpha 1 and alpha 2 constant region gene sequences. Cell 36: 681–688.
60. Shimizu A, Takahashi N, Yaoita Y, Honjo T (1982) Organization of the constant-region gene family of the mouse immunoglobulin heavy chain. Cell 28: 499–506.
61. Butler JE, Sun J, Wertz N, Sinkora M (2006) Antibody repertoire development in swine. Dev Comp Immunol 30: 199–221.
62. Wagner B, Miller DC, Lear 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.
63. Honjo T, Matsuda F (1995) Immunoglobulin heavy chain loci of mouse and man. Immunoglobulin Genes. Academic Press New York: Pp. 145–171.
64. Morgado MG, Cam P, Gris-Liebe C, Cazenave PA, Jouvin-Marche E (1989) Further evidence that BALB/c and C57BL/6 gamma 2a genes originate from two distinct isotypes. EMBO J 8: 3245–3251.
65. Bruggemann M (1988) Evolution of the rat immunoglobulin gamma heavy-chain gene family. Gene 74: 473–482.
66. Bensmana M, Huck S, Lefranc G, Lefranc MP (1988) The human immunoglobulin pseudo-gamma IGHGP gene shows no major structural defect. Nucleic Acids Res 16: 3108.
67. Das S, Nozawa M, Klein J, Nei M (2008) Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. Immunogenetics 60: 47–55.
68. Nei M (2007) The new mutation theory of phenotypic evolution. Proc Natl Acad Sci U S A 104: 12235–12242.
69. Bao Y, Guo Y, Xiao S, Zhao Z (2010) Molecular characterization of the VH repertoire in Canis familiaris. Vet Immunol Immunopathol 137: 64–75.
70. Berman JE, Mellis SJ, Pollock R, Smith CL, Suh H, et al. (1988) Content and organization of the human Ig VH locus: definition of three new VH families and linkage to the Ig CH locus. EMBO J 7: 727–738.
71. van Dijk KW, Mortari F, Kirkham PM, Schroeder HW, Jr., Milner EC (1993) The human immunoglobulin VH7 gene family consists of a small, polymorphic group of six to eight gene segments dispersed throughout the VH locus. Eur J Immunol 23: 832–839.
72. Sinclair MC, Gilchrist J, Aitken R (1997) Bovine IgG repertoire is dominated by a single diversified VH gene family. J Immunol 159: 3883–3889.
73. Johnston CM, Wood AL, Bolland DJ, Corcoran AE (2006) Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region. J Immunol 176: 4221–4234.
74. Currier SJ, Gallarda JL, Knight KL (1988) Partial molecular genetic map of the rabbit VH chromosomal region. J Immunol 140: 1651–1659.
75. Nguyen VK, Muyldermans S, Hamers R (1998) The specific variable domain of camel heavy-chain antibodies is encoded in the germline. J Mol Biol 275: 413–418.
76. Sun J, Kacskovics I, Brown WR, Butler JE (1994) Expressed swine VH genes belong to a small VH gene family homologous to human VHIII. J Immunol 153: 5618–5627.
77. Butler JE (1997) Immunoglobulin gene organization and the mechanism of repertoire development. Scand J Immunol 45: 455–462.
78. Charlton KA, Moyle S, Porter AJ, Harris WJ (2000) Analysis of the diversity of a sheep antibody repertoire as revealed from a bacteriophage display library. J Immunol 164: 6221–6229.
79. Baker ML, Tachedjian M, Wang LF (2010) Immunoglobulin heavy chain diversity in Pteropid bats: evidence for a diverse and highly specific antigen binding repertoire. Immunogenetics 62: 173–184.
80. Tutter A, Riblet R (1989) Conservation of an immunoglobulin variable-region gene family indicates a specific, noncoding function. Proc Natl Acad Sci U S A 86: 7460–7464.
81. Johansson J, Aveskogh M, Munday B, Hellman L (2002) Heavy chain V region diversity in the duck-billed platypus (Ornithorhynchus anatinus): long and highly variable complementarity-determining region 3 compensates for limited germline diversity. J Immunol 168: 5155–5162.
82. Kullander K, Carlson B, Hallbook F (1997) Molecular phylogeny and evolution of the neurotrophins from monotremes and marsupials. J Mol Evol 45: 311–321.
83. Janke A, Xu X, Arnason U (1997) The complete mitochondrial genome of the wallaroo (Macropus robustus) and the phylogenetic relationship among Monotremata, Marsupialia, and Eutheria. Proc Natl Acad Sci U S A 94: 1276–1281.

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, Waire J, Lin AW, Osborne BA, et al. (1998) Multiple sites of V lambda diversification in cattle. J Immunol 161: 5438–5444.
89. 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.