

Different Patterns of Evolution in the Centromeric and Telomeric Regions of Group *A* and *B* Haplotypes of the Human Killer Cell Ig-Like Receptor Locus

Chul-Woo Pyo¹, Lisbeth A. Guethlein², Quyen Vu¹, Ruihan Wang¹, Laurent Abi-Rached², Paul J. Norman², Steven G. E. Marsh³, Jeffrey S. Miller⁴, Peter Parham², Daniel E. Geraghty^{1*}

1 Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, **2** Department of Structural Biology, Stanford University School of Medicine, Stanford, California, United States of America, **3** Anthony Nolan Research Institute, London, United Kingdom, **4** University of Minnesota, Minneapolis, Minnesota, United States of America

Abstract

The fast evolving human *KIR* gene family encodes variable lymphocyte receptors specific for polymorphic HLA class I determinants. Nucleotide sequences for 24 representative human *KIR* haplotypes were determined. With three previously defined haplotypes, this gave a set of 12 group *A* and 15 group *B* haplotypes for assessment of *KIR* variation. The seven gene-content haplotypes are all combinations of four centromeric and two telomeric motifs. *2DL5*, *2DS5* and *2DS3* can be present in centromeric and telomeric locations. With one exception, haplotypes having identical gene content differed in their combinations of *KIR* alleles. Sequence diversity varied between haplotype groups and between centromeric and telomeric halves of the *KIR* locus. The most variable *A* haplotype genes are in the telomeric half, whereas the most variable genes characterizing *B* haplotypes are in the centromeric half. Of the highly polymorphic genes, only the *3DL3* framework gene exhibits a similar diversity when carried by *A* and *B* haplotypes. Phylogenetic analysis and divergence time estimates, point to the centromeric gene-content motifs that distinguish *A* and *B* haplotypes having emerged ~6 million years ago, contemporaneously with the separation of human and chimpanzee ancestors. In contrast, the telomeric motifs that distinguish *A* and *B* haplotypes emerged more recently, ~1.7 million years ago, before the emergence of *Homo sapiens*. Thus the centromeric and telomeric motifs that typify *A* and *B* haplotypes have likely been present throughout human evolution. The results suggest the common ancestor of *A* and *B* haplotypes combined a *B*-like centromeric region with an *A*-like telomeric region.

Citation: Pyo C-W, Guethlein LA, Vu Q, Wang R, Abi-Rached L, et al. (2010) Different Patterns of Evolution in the Centromeric and Telomeric Regions of Group *A* and *B* Haplotypes of the Human Killer Cell Ig-Like Receptor Locus. PLoS ONE 5(12): e15115. doi:10.1371/journal.pone.0015115

Editor: Hiroaki Matsunami, Duke University, United States of America

Received: August 5, 2010; **Accepted:** October 25, 2010; **Published:** December 29, 2010

Copyright: © 2010 Pyo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by National Institutes of Health grants CA111412 to J.S.M. and RR018669 to D.E.G. 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: geraghty@fhcrc.org

Introduction

Humans, apes and monkeys have expanded families of genes encoding killer cell immunoglobulin-like receptors (*KIR*) (reviewed in [1]). Through the recognition of MHC class I, *KIR* regulate the development and response of natural killer (NK) cells (reviewed in [2]). *KIR* are also expressed by subpopulations of $\alpha\beta$ and $\gamma\delta$ T cells [3]. In their function, genetics and variegated expression, primate *KIR* are very similar to rodent Ly49, although these common properties are the result of convergent evolution [4]. Because both receptor and ligand are highly polymorphic, the interactions between *KIR* and HLA class I that regulate NK cell function are extraordinarily diverse. And this is further increased by the influence of variable peptides that bind to HLA class I and are contacted by *KIR* [5,6]. As a consequence, variability in *KIR* genotype in human populations is associated with as wide-ranging a collection of diseases as HLA. Notably, they include susceptibility to infection [7,8] and autoimmunity [9,10,11,12,13,14,15], the outcome of hematopoietic cell transplantation [16,17] and the success of placental reproduction [18].

Genetic diversity in the human *KIR* gene family arises from two factors: variability in *KIR* gene content and allelic polymorphism [19]. Because of the propensity for asymmetric recombination within the *KIR* gene family, the simple definitions of genes and alleles do not always apply [20], and as a consequence the reported number of human *KIR* genes varies. By conservative account, the human *KIR* family consists of 11 genes (*2DL1*, *2/3*, *4* and *5*; *2DS1*, *2*, *4* and *3/5*; and *3DL1/S1*, *3DL2* and *3*) and two pseudogenes (*2DP1* and *3DP1*). Through the combination of gene-content diversity and allelic polymorphism, the variability in *KIR* genotype is such that most pairs of unrelated human individuals have different *KIR* genotypes [21], as is also the case for HLA class I [22]. A unique feature of the human *KIR* system, and one not mirrored in other higher primates, is the segregation of two distinctive groups of haplotypes (*A* and *B*) [23], which are present in all the >150 human populations examined and are maintained by balancing selection [24]. The group *A* haplotypes have a simple and constant gene content, dominated by genes encoding inhibitory receptors. In contrast, the group *B* haplotypes have variable and greater gene content, involving genes encoding

distinctive inhibitory receptors and a variety of activating receptors [23].

Although the structures of numerous *KIR* haplotypes have been indirectly inferred and deduced from population analysis and family studies [19,25,26,27,28,29,30,31], complete determination of human *KIR* haplotype structures using direct methods has so far been limited to one *A* haplotype and two *B* haplotypes [25,32,33]. Given the decisive role of *KIR* variability in modulating the human NK cell response during infection, pregnancy and allogeneic transplantation, it became essential that the structures of the common *KIR* haplotypes be unambiguously defined. To this end we determined the sequences of a balanced selection of 24 common group *A* and *B* *KIR* haplotypes.

Results

Based upon *KIR* gene content, cell lines derived from 12 individuals were chosen for complete sequence analysis of *KIR* haplotypes. In choosing these cells, two selection criteria were applied: first, that each cell line was inferred to carry both an *A* and *B* *KIR* haplotype; second, that all common gene haplotypes in the *B* group were represented in the panel of cell lines. As predicted, eleven members of the panel had both an *A* and *B* *KIR* haplotype. The twelfth member of the panel (from cell line GRC212) had two *B* haplotypes, one sharing the centromeric part of the locus with the *A* haplotype and the other sharing the telomeric part with *A*. Combining these haplotypes with three other *KIR* haplotypes deposited in Gen-Bank [32,33,34] gave a data set of 27 *KIR* haplotypes in which seven gene-content

haplotypes were represented (Fig. 1). Twelve haplotypes are of group *A*, and 15 of group *B*. Four of the six group *B* gene-content haplotypes were represented more than once. Common to the 27 haplotypes were the three framework regions of the *KIR* locus originally defined by Trowsdale and colleagues [33]: *KIR3DL3* at the centromeric end, *KIR3DL2* at the telomeric end, and the combination of *KIR3DP1* and *KIR2DL4* in the middle of the locus.

In order to determine the representative frequency of each of the sequenced haplotypes, including an assessment of how inclusive our sequenced haplotypes are of those common in populations, we examined a panel of 192 DNAs including 48 each of Caucasian, Asian, African-American, and Hispanic. Using a panel of STS assays specific for each of the *KIR* genes and pseudogenes as described in Methods we were able to provide a total gene content analysis for each individual. A subset of these assays included PCR primers that extended from the end of one gene to the start of an adjacent gene allowing us to determine *cis* and *trans* relationships among some pairs of genes. These data altogether allowed us to deduce the pair of haplotypes contained within each individual, with the exceptions of two pairs of ambiguities - cA01|tA01-cB01|tB01 versus cA01|tB01-cB01|tA01 and cA01|tA01-cB02|tB01 versus cA01|tB01-cB02|tA01. We used estimated frequencies of each of these haplotypes calculated by linkage disequilibrium estimates of the haplotype frequencies to adjust for these ambiguities. Of the 7 haplotypes sequenced, 6 were relatively common in all populations examined while the seventh was found in only 2 individuals of African descent (Table 1). In addition, a total of 21 chromosomes did not match any of the sequenced haplotype patterns, including 13 chromosomes in the African-American samples. However, of the 12 distinct

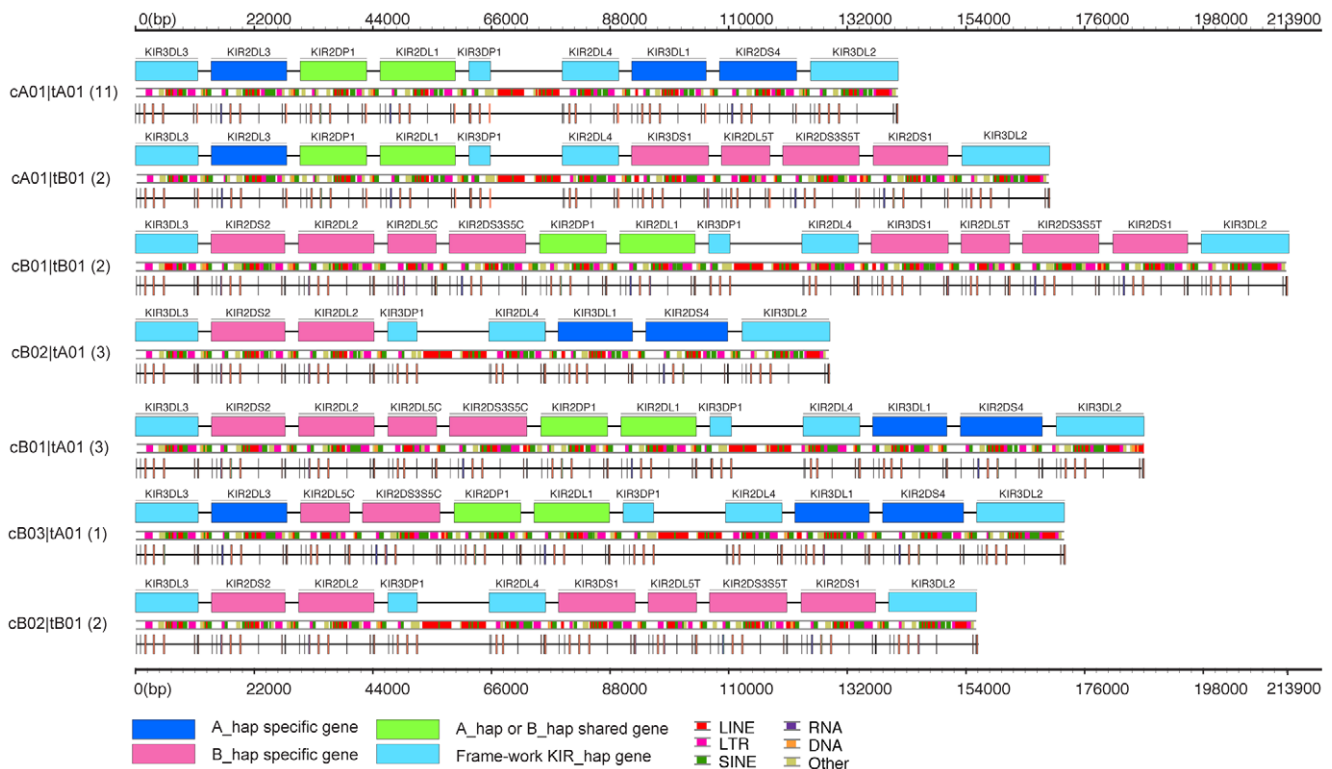


Figure 1. Genomic sequencing of seven gene-content *KIR* haplotypes represented in a panel of 11 *A* and 13 *B* *KIR* sequences. *KIR* gene content haplotype structures found from the phased sequences of 24 haplotypes. Each haplotype is named according to the WHO nomenclature as indicated to the left of each sequence, with the numbers of each haplotype resequenced indicated in parenthesis. The key at the bottom designates the color-coding used to distinguish the genes according to type, and the repeat sequence structures found within each sequence. The structure for each gene and pseudogene is indicated beneath with bars indicating exons.
doi:10.1371/journal.pone.0015115.g001

Table 1. KIR haplotype frequencies in 192 reference samples.

KIR Haplotype	AFA n=96	ASI n=96	HIS n=96	CAU n=96	Total n=384 (%)
cA01 tA01	52	61	59	60	232 (60.4)
cA01 tB01	5	15	15	9	45 (11.7)
cB01 tB01	2	1	1	6	10 (2.6)
cB02 tA01	6	5	8	15	34 (8.9)
cB01 tA01	14	2	8	4	28 (7.3)
cB02 tB01	2	6	4	1	13 (3.4)
cB03 tA01	2	0	0	0	2 (0.5)
other	13	6	1	1	21 (5.5)

doi:10.1371/journal.pone.0015115.t001

haplotypes that made up this group, none were present at more than 1% of the total. We did not fully characterize these haplotypes other than to exclude them from among the 7 sequenced haplotypes and to distinguish them among themselves and presume they may include examples of the reported rare KIR haplotypes [20,35,36].

A diversity of KIR haplotypes is formed from few centromeric and telomeric gene content motifs

Dividing the KIR haplotypes into centromeric and telomeric regions separated by the *3DP1-2DL4* framework [33], showed that the seven gene content KIR haplotypes are all combinations of four centromeric and two telomeric gene-content motifs (Figure 1). Motif *Cen-B3* is present on a single haplotype, whereas the other motifs are represented on 5 to 19 haplotypes (Table 2). Excluding *Cen-B3*, all possible combinations of the remaining three centromeric and two telomeric motifs are represented in the 27 KIR haplotypes. Because much of KIR diversity arises from recombinatorial association of centromeric and telomeric motifs, a logical and adaptable nomenclature for haplotypes has been based on this principle and is introduced in Figure 1.

Although the twelve group A haplotypes have identical KIR gene content, they all have different combinations of alleles for the constituent A haplotype genes (Figure 2). This demonstrates the extent to which allelic polymorphism diversifies group A haplotypes. Allelic polymorphism is also seen to distinguish group B haplotypes of identical gene content. However, two haplotypes having the identical *Cen-B1* and *Tel-B1* gene content motif also have identical combinations of KIR alleles.

Most polymorphic of the motifs is *Tel-A* with twelve allotype combinations, followed by *Cen-A* with ten allotype combinations. There is considerably less variation in the motifs that define B

Table 2. KIR motif frequencies.

MOTIF	Number of allele motifs	Number of haplotypes
Cen	A1	13
	B1	4
	B2	6
	B3	1
Tel	A1	13
	B1	4

doi:10.1371/journal.pone.0015115.t002

haplotypes, the most diverse being *Cen-B2* with six allele combinations. This hierarchy of allelic variation agrees with that observed in population studies [27,31]. Fifteen new alleles encoding ten new proteins and fourteen new pseudogene alleles were defined in this study (Fig. 2). All but one of the new variants (2DS5*004) were either alleles of framework genes or genes characteristic of the A haplotypes. By applying a similar hierarchical approach to that used for HLA nomenclature [37], the proposed KIR haplotype nomenclature can accommodate the higher resolution achieved with the combination of allele sequences, as shown in Figure 2.

Nucleotide diversity of the KIR region varies with haplotype groups and the two halves of the KIR locus

To first assess the diversity of the KIR region as a whole, we calculated the nucleotide diversity (Π) by applying DnaSP [38] to an alignment of KIR sequences including all of the sequence data comprising the centromeric and telomeric regions subdivided into 5 of the 6 motifs, each compared separately (Figure 3A, File S1). The cB03 motif was not included at this level of analysis since only one copy was available. Two distinguishing features are the extremely low levels of diversity over *2DL4* and *3DS1* in the KIR-tB01 segment and the substantially higher diversity in the *2DP1-2DL1* segment in *KIR-cB01* versus *KIR-cA01*. Most of the diversity found spanning the 2DS3S5 locus results from the differences between the 2DS3 and 2DS5 alleles and therefore they were analyzed separately in the individual gene analysis immediately following.

To assess the allelic variation of individual KIR genes we similarly calculated their nucleotide diversity to an alignment of KIR sequences individually, each beginning 250 bp upstream of the start codon and extends throughout the gene to end at the polyadenylation site. In this alignment, the *3DP1* sequences naturally terminate at the end of exon 5. Comparison of variants characterizing the A and B haplotypes was performed (Figure 3B, C). Thus, subgroups of the framework genes, *2DL1* and *2DP1* were analyzed separately according to their presence on either the A or B haplotypes. In addition, the *2DL5* sequences were subdivided according to their linkage, either to *2DS3* or *2DS5*, as well as their presence in either the centromeric or telomeric part of the locus (Figure 3B). Of the eight genes and KIR3DP1 common to A and B haplotypes, only *3DL3* and *2DP1* exhibited similar variability in the two haplotype groups. In contrast, *2DL1* and *3DP1* are more variable in the B haplotypes and the other five genes are more variable in the A haplotypes. Further distinguishing the haplotype groups, the most variable A haplotype genes (*3DL1* and *2DS4*) are in the telomeric part of the locus, whereas the most variable B haplotype genes (*2DL5* and *2DS5*) are in the centromeric part.

The overall diversity in the centromeric and telomeric parts of the A and B haplotypes was assessed from averages obtained from the Π values for their constituent genes (Figure 3C). The *3DL1* and *2DS4* genes of the *Tel-A* segments have the highest Π , a diversity that extends to the neighboring framework genes and makes *Tel-A* significantly more diverse than the other three segments. Although the average nucleotide diversity for *Cen-A* and *Cen-B* is similar, the *Cen-B* segment genes form a bimodal distribution in which *2DS2*, *2DL2L3*, and *2DS3S5* have little diversity, whereas *2DL1*, *2DS5*, and *2DL5* have substantial diversity. Apart from *3DL3*, the *Cen-A* segment genes are of low diversity, which falls between the values for the two groups of *Cen-B* genes. This pattern of variation points to the minimally diverse *Cen-B* genes having either been more recently formed or more recently subject to selection.

To investigate further the observation that *3DL3* is highly polymorphic in both *Cen-A* and *Cen-B*, we performed domain-by-domain phylogenetic analysis of the alleles of this gene (Figure 4). The analysis revealed a recombination break point between the

Haplotype	Haplotype motif		Centromere										Telomere						Source
	Cen	Tel	3DL3	2DS2	2DL2L3	2DL5C	2DS35C	2DP1	2DL1	3DP1	2DL4	3DL1S1	2DL5T	2DS35T	2DS1	2DS4	3DL2		
<i>cA01 A01</i>	<i>A01</i>	<i>A01</i>	3DL3	2DS2	2DL2L3	2DL5C	2DS35C	2DP1	2DL1	3DP1	2DL4	3DL1S1	2DL5T	2DS35T	2DS1	2DS4	3DL2		
<i>cA01:001 tA01:001</i>			*00201	*001				*00302	*00201	*00302	*00801	*00101			*003	*005	*001	AC011501	
<i>cA01:001:02 tA01:002:01</i>			*00206		*001			N1	*00302	N1	*00102	*01502			*00101	*002		LUCE	
<i>cA01:002 tA01:002:01</i>			*00802		*001			*00201	*00302	*00302	*00102	*01502			*00101	*002		T7526	
<i>cA01:003:01:01 tA01:003</i>			*00901		*001			*00201	*00302	*00302	*00103	*025			*00101	part		ABC08	
<i>cA01:003:01:01 tA01:004</i>			*00901		*001			*00201	*00302	*00302	*011	*00501			N1 007L	N1*001		G085	
<i>cA01:003:01:02 tA01:005</i>			*00901		*001			N2	*00302	*00302	*00103	*008			*003	*00901		FH06	
<i>cA01:003:01:03 tA01:006</i>			*00901		*001			*00201	*00302	N2	*00103	*01701			*00101	N2 012L		RSH	
<i>cA01:006 tA01:007</i>			*012		*001			*00201	*00302	N3	*00103	*01501			*00101	*013		FH13	
<i>cA01:007 tA01:008</i>			N1 010L		*001			*00201	*00302	*00302	*00102	*029			*0101	*002		G248	
<i>cA01:008 tA01:009</i>			*00101		*002			N3	*002	*006	*00103	N1*020			*00101	N4*009		FH05	
<i>cA01:009 tA01:010</i>			*019		*002			*003	*002	*006	*00802	*00401			*006	*005		FH15	
<i>cA01:010 tA01:002:02</i>			*005		*006			N4	*00303	*00302	*00103	*01502			*00101	*002		FH08	
<i>cA01 tB01</i>	<i>A01</i>	<i>B01</i>	3DL3	2DS2	2DL2	2DL5C	2DS35C	2DP1	2DL1	3DP1	2DL4	3DL1	2DL5T	2DS35T	2DS1	2DS4	3DL2		
<i>cA01:004 tB01:004</i>			*00901		*001			*00201	N1 003L	*00302	*00103						N3 001L	ABC08	
<i>cA01:005 tB01:002</i>			*010		*001			*00201	*00302	*00302	*00501	*01301	*00501	3*002	*00201		*00701	GRC212	
<i>cB01 tA01</i>	<i>B01</i>	<i>A01</i>	3DL3	2DS2	2DL2	2DL5C	2DS35C	2DP1	2DL1	3DP1	2DL4	3DL1	2DL5T	2DS35T	2DS1	2DS4	3DL2		
<i>cB01:002 tA01:011</i>			*00301	*00101	*001	*00201	3*00103	*00102	*007	*00301	*00602	*007				*004	*008	G348	
<i>cB01:003 tA01:007</i>			*01402	*00101	*001	*00801	5*007	N5	*00401	N4	*011	*00501				N2 007L	*010	FH13	
<i>cB01:004 tA01:007</i>			*00402	*00101	*001	*004	5*006	N6	N2 001L	N4	*00801	*00501				N2 007L	*010	RSH	
<i>cB01 tB01</i>	<i>B01</i>	<i>B01</i>	3DL3	2DS2	2DL2	2DL5C	2DS35C	2DP1	2DL1	3DP1	2DL4	3DL1	2DL5T	2DS35T	2DS1	2DS4	3DL2		
<i>cB01:001 tB01:001</i>			*00301	*00101	*001	*00201	3*00103	*00102	*00401	*00301	*00501	*01301	*00101	5*002	*00201		*00701	AY320039	
<i>cB01:001 tB01:002</i>			*00301	*00101	*001	*00201	3*00103	*00102	*00401	*00301	*00501	*01301	*00501	3*002	*00201		*00701	FH05	
<i>cB01:001 tB01:002</i>			*00301	*00101	*001	*00201	3*00103	*00102	*00401	*00301	*00501	*01301	*00501	3*002	*00201		*00701	FH15	
<i>cB02 tA01</i>	<i>B02</i>	<i>A01</i>	3DL3	2DS2	2DL2	2DL5C	2DS35C	2DP1	2DL1	3DP1	2DL4	3DL1	2DL5T	2DS35T	2DS1	2DS4	3DL2		
<i>cB02:002 tA01:008</i>			*00402	*00101	*003				N5	*00102	*029					*00101	*002	GRC212	
<i>cB02:004 tA01:011</i>			*01402	*00101	*001				*001	*00602	*007					*004	*008	G085	
<i>cB02:006 tA01:001</i>			N2 020L	*00101	*003				N6	*00801	*00101					*003	*00101	FH06	
<i>cB02 tB01</i>	<i>B02</i>	<i>B01</i>	3DL3	2DS2	2DL2	2DL5C	2DS35C	2DP1	2DL1	3DP1	2DL4	3DL1	2DL5T	2DS35T	2DS1	2DS4	3DL2		
<i>cB02:001 tB01:001</i>			*00301	*00101	*003				*002	*00501	*01301	*00101	5*002	*00201		part		AL133414	
<i>cB02:003 tB01:002</i>			*00801	*00101	*003				N6	*00501	*01301	*00501	3*002	*00201		*00701		T7526	
<i>cB02:005 tB01:003</i>			*01402	*00101	*003				N6	*00501	*01301	*00101	5*002	*00201		*018		LUCE	
<i>cB03 tA01</i>	<i>B03</i>	<i>A01</i>	3DL3	2DS2	2DL2	2DL5C	2DS35C	2DP1	2DL1	3DP1	2DL4	3DL1	2DL5T	2DS35T	2DS1	2DS4	3DL2		
<i>cB03:001 tA01:012</i>			part			N1	*00601	N5*004	N7	*010	N7	*00802	N2*004			*006	N5*002	FH08	
Summary																			
Alleles (exons and introns)			24	9	24	6	5	17	18	20	15	21	6	6	6	15	22		
Alleles (exons)			16	1	6	4	4	10	8	12	7	13	2	2	1	5	13		
Allotypes			14	1	6	4	4	4	7	7	5	12	2	2	1	5	10		

Figure 2. Four centromeric and two telomeric gene-content motifs combined with allelic variation diversifies KIR haplotypes. The names of the centromeric (*Cen*) and telomeric (*Tel*) motifs are given at the left under the heading ‘Haplotype’. The structures of each of seven gene content haplotypes are indicated above each group of sequenced haplotypes that contained the respective motif. Gene names are given in boxes: colored gray for framework genes, red for *Cen-A* and *Tel-A* genes, blue for *Cen-B* and *Tel-B* genes, and green for genes found in both *Cen-A* and *Cen-B*. The gene 3DL1S1 is distinguished within each haplotype according to its respective motif residence as red or blue as are 2DS1 and 2DS4. Beneath each gene name are listed the respective allele number within each sequenced haplotype with pink color indicating newly discovered alleles. Yellow shading indicates that the gene/allele sequence was incomplete. Shown to the right is the source, either the cell line name or GenBank entry, for each haplotype. The haplotype names for each sequenced haplotype are based on the allele combination in the centromeric and telomeric gene motifs. For example, the haplotype AC011501 at the top of the table has the *Cen-A1* and *Tel-A1* motifs. It was assigned 001 for each of the allelic combinations comprising these motifs and has thus been designated *cA01:001|tA01:001*. Only two of the 27 haplotypes share identical allele content (*cB01:001|tB01:002*). Under the chart of haplotypes, the numbers of different alleles and allotypes for each KIR gene are given. doi:10.1371/journal.pone.0015115.g002

exons encoding the D2 domain and the transmembrane region. This point of recombination defines two KIR3DL3 lineages that differ in their transmembrane domains and cytoplasmic tails. One of the 3DL3 lineages is exclusively associated with *Cen-A*, whereas the other is exclusively associated with *Cen-B*. The recombination breakpoint also defines two lineages for the extracellular domains of KIR3DL3, but these do not segregate with *Cen-A* and *Cen-B* group but are evenly distributed between them. However, no 3DL3 allele is common to *Cen-A* and *Cen-B*. One exception (3DL3*0801 found on a *Cen-B* segment) appears to be the result of a recombination that occurred 3’ of the 3DL3 gene in the intergenic region between it and 2DS2.

KIR2DS3, 2DS5, (2DS35S) and 2DL5 can be present in the centromeric and telomeric parts of the KIR locus

Previous studies have shown that two forms of *KIR2DL5* locate to the centromeric (*KIR2DL5B*) and telomeric (*KIR2DL5A*) parts of the *KIR* gene locus [39]. Based upon linkage disequilibrium between 2DL5B or 2DL5A, it has also been suggested that 2DS3 and 2DS5 can similarly be present in either centromeric or telomeric locations [31,40,41]. The structures of *KIR* haplotypes defined here prove this hypothesis to be true. The centromeric motifs of seven haplotypes contain 2DL5B, which can associate with 2DS3*00103 or one of three 2DS5 alleles (*new, *005 or *006). Likewise, the telomeric motifs of seven haplotypes contain 2DL5A, which can associate with either 2DS3*002 or 2DS5*002. Thus 2DL5A and 2DL5B both associate with 2DS3 and 2DS5. In all

combinations, the allele of 2DL5 uniquely defines the associated 2DS3 or 2DS5 allele, and vice versa. Particularly variable are the three pairs of associated 2DL5B and 2DS5 alleles (Fig.3A).

We aligned the seven sequences of genomic segments containing 2DL5A or 2DL5B and the neighboring 2DS3 or 2DS5 (Figure 5A), and performed domain-by-domain phylogenetic analysis with construction of neighbor-joining and parsimony trees. Domains were defined as individual introns or exons, recombination breakpoints were determined both by visual inspection of the alignment and use of RDP. The results showed that these segments divide into three parts having different evolutionary histories (Figure 5B). Sequences of the region extending from 5’ of the 2DL5 start codon through intron 2 of 2DL5, formed two groups corresponding to 2DL5A and 2DL5B. In this 1.5 kb region, 2DL5B*00601 is divergent, differing by 21 unique substitutions from other 2DL5 and other KIR. In the proximal region, extending from exon 3 of 2DL5 through to intron 6 of 2DS3 or 2DS5, the sequences form two groups, corresponding to the presence of either 2DS3 or 2DS5. This grouping is independent of the genomic location. Analysis of the region extending from intron 6 to the end of the 2DS3 or 2DS5 gene showed that 2DS3*00103 forms an outgroup to the other 2DS3 2DS5 variants. In the latter group, three substitutions distinguish sequences derived from the centromeric and telomeric regions.

Taken together, these data are consistent with two alternative models for the evolution of the segments containing 2DL5 and either 2DS3 or 2DS5 (Figure 5C). Common to both models, the progenitor haplotype had 2DL5, and the common ancestor of

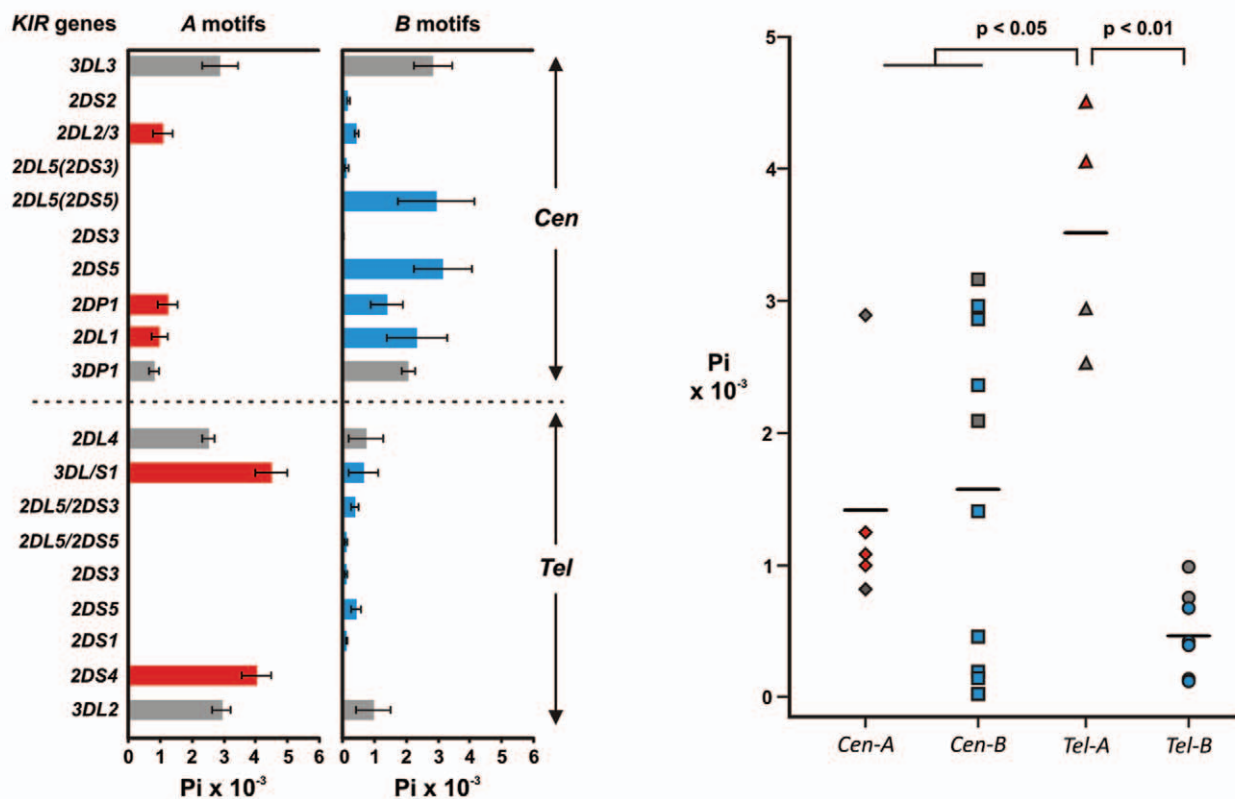
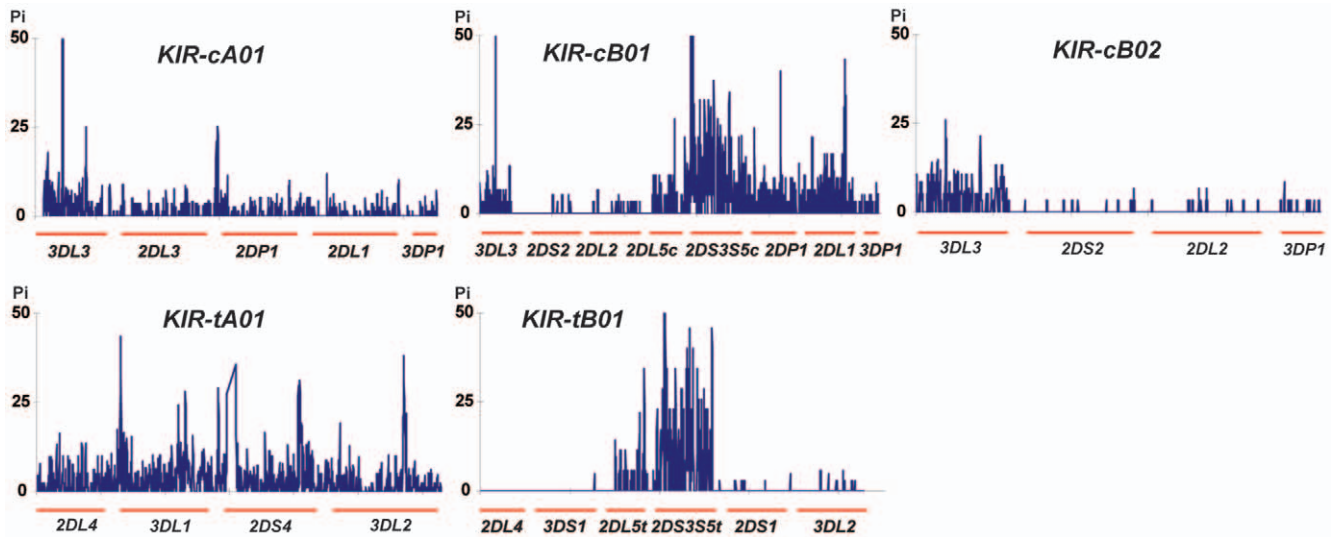


Figure 3. *Tel-A* genes exhibit the highest level of nucleotide variation. Nucleotide diversity, P_i , was calculated for each genomic segment and for each locus separately. Alignment of *KIR* sequences comprising the centromeric and telomeric regions subdivided into 5 of the 6 motifs was examined. *KIR* genes were also examined individually and alleles for framework genes were sorted according to their location on A or B motifs. *KIR2DL1* and *2DP1* alleles were similarly analyzed according to their presence on A or B motifs. *KIR2DL5* alleles were subdivided according to their linkage with either *2DS3* or *2DS5*. (A) Shows plots the P_i values over each genomic region for each of the 5 motifs as indicated. (B) Shows the P_i values for the individual *KIR* genes. Characteristic genes of the A haplotype are shown in red, characteristic genes of the B haplotype are shown in blue, and framework genes are shown in grey. Overall the *Tel-A* genes have the highest diversity. Certain *Cen-B* genes exhibit moderate diversity, with lowest diversity evident for the *Tel-B* genes. (C) Individual data points are the P_i values for each of the genes of *Cen-A*, *Cen-B*, *Tel-A*, and *Tel-B* segments and their mean values (horizontal lines). P_i values for *Cen-A* genes are indicated by diamonds, *Cen-B* by squares, *Tel-A* by triangles, and *Tel-B* by circles. Color coding is the same as in panel B. doi:10.1371/journal.pone.0015115.g003

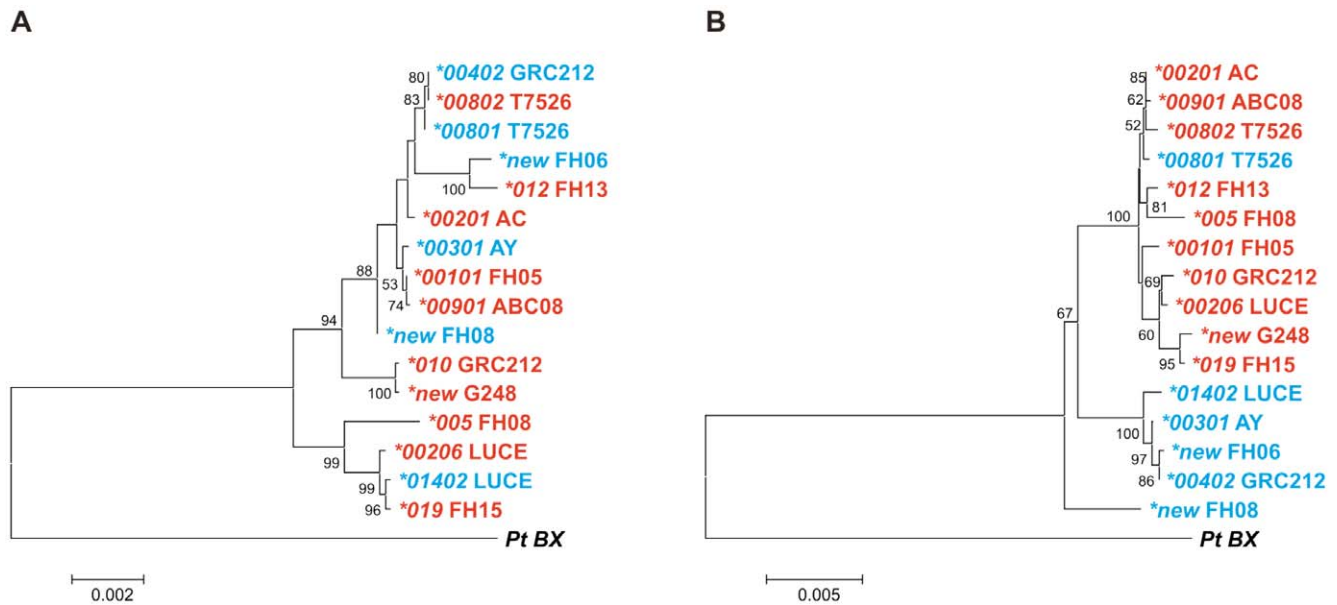


Figure 4. A recombination point in 3DL3 shuffles Ig domain encoding exons with different tail-encoding exons. (A) Shows the NJ trees for the 5' region of 3DL3 beginning 250 bp 5' of the start codon and ending 1 kb 3' of exon 7 (encoding the transmembrane domain). Sequence names are colored according to their presence on *Cen-A* or *Cen-B* segments. (B) Shows the NJ trees for the 3' region of 3DL3 beginning 1 kb 5' of exon 7 and ending at the polyadenylation signal. Color coding is as in panel A. The alleles divide according to their presence on *Cen-A* or *Cen-B* segments for the 3' region and are mixed in the 5' region. The exception (3DL3*0801 found on a *Cen-B* segment) is a result of a recombination that occurred 3' of the 3DL3 gene in the intergenic region between it and 2DS2. doi:10.1371/journal.pone.0015115.g004

2DS3 and 2DS5, in the centromeric part of the *KIR* locus. This proposition is supported by presence in the centromeric region, and absence from the telomeric region, of chimpanzee and orangutan *2DL5* orthologs. Subsequently this segment duplicated, with transposition of the daughter locus to the telomeric region. These two segments then diverged independently to form 2DS3 and 2DS5, which now differ by ~250 nucleotide substitutions in 15kb. Subsequent recombination and/or gene conversion events led to 2DS3 and 2DS5 being present in both the centromeric and telomeric locations. The two models differ in the locations where 2DS3 and 2DS5 arose (Figure 5C).

Model 1 proposes that 2DS5 arose in the centromeric region and 2DS3 arose in the telomeric region; in contrast, model 2 proposes that 2DS3 arose in the centromeric region and 2DS5 arose in the telomeric region (Figure 5C). Model 1 requires three recombination events to generate the haplotypes present in the panel (Figure 5C left), the two proposed above and an additional recombination within the *Tel* segment. Supporting model 1 is the greater diversity of 2DS5 in the centromeric location, suggesting that 2DS5 has existed there for the longer period of time. Alternatively, the increased 2DS5 diversity may not be a consequence of age, but of recent selection. Model 2 is the more parsimonious model, because it requires only two recombination events to generate the haplotypes seen in the panel (Figure 5C, right). In summary, there is no compelling evidence for ruling out either of the two models. What is clear, however, is that genetic diversity arising in the two locations now containing 2DS3 and 2DS5 genes has subsequently been moved between them by recombination.

The centromeric and telomeric motifs of A and B haplotypes diverged at very different times

As all modern human populations have A and B haplotypes, whereas this distinction has not been described in other primates, it was of interest to know when and how the A and B haplotype

motifs diverged from a common ancestor. To estimate divergence times, we performed phylogenetic analysis on three potentially informative regions, shown in black in Figure 6A. Region I is a 5.5 kb segment of the 3DL3 gene that starts 250 bp upstream of the start codon and ends 400 bp 5' of exon 5 at the recombination breakpoint in the orangutan sequence [42]; Region II comprises the 14 kb intergenic region between *3DP1* and *2DL4*. Region III is a 16.8 kb segment beginning 100 bp 5' of exon 3 of the *2DL5* gene and extending to 500 bp 3' of exon 5 of the neighbouring *2DS3* or *2DS5* gene, including the intergenic region between them.

Phylogenetic analysis was performed on datasets that combined orthologous/paralogous sequences from human and other primate species, and from which recombinant sequences were excluded. Regions I and II have counterparts in rhesus macaque, orangutan, gorilla, and chimpanzee, whereas region III is limited to humans and great apes. In the phylogenetic trees, the deeper branches segregate the *KIR* from different species (Figure 6B). The positions of these deeper branch points (nodes) were used to calculate times for the divergence of humans from macaque, orangutan, gorilla and chimpanzee. These divergence times, plotted in Figure 6B, correspond well with estimates based on the fossil record, giving confidence in the validity of the analysis.

The more recent branch points in the phylogenetic trees distinguish groups of human sequences and mark events in *KIR* locus evolution that occurred in the human lineage since separation from the chimpanzee lineage. The two main lineages in region I, correspond to 3DL3 alleles that segregate with the *Cen-A* and *Cen-B* motifs. Their divergence time is estimated to be 6-7.2 million years ago, contemporaneous with the divergence time of the human and chimpanzee lineages (Figure 6C). Further supporting this conclusion, the chimpanzee and gorilla genes form a distinct branch of the tree (Fig. 6B), suggesting that several *Cen* lineages were present at the time of separation of humans from chimpanzees, whereupon chimpanzees retained a lineage distinct

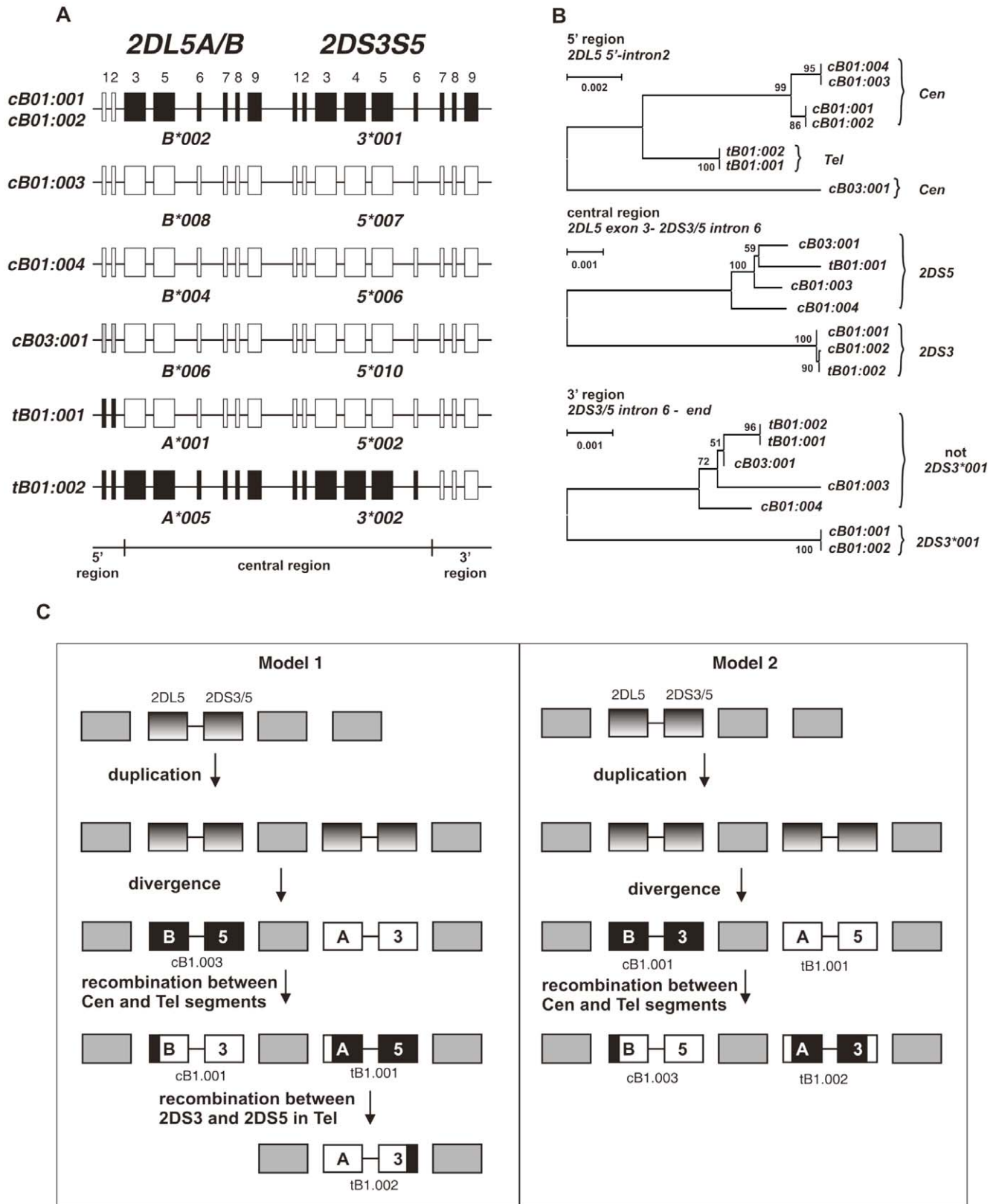


Figure 5. *2DS3* and *2DS5* descend from a common ancestor and diversified by interlocus recombination. (A) Shows the six combinations of *2DL5* and *2DS3/5* present in the 27 haplotypes, along with the motif to which they belong. Shading denotes their relatedness as determined by the phylogenetic analyses shown in panel B. (B) Genomic sequences extending from 5' of *2DL5* through the end of the neighboring *2DS3/5* gene were aligned, divided into three regions (5', central, and 3') and submitted to phylogenetic analysis. Neighbor joining trees are shown, with Bootstrap values for the nodes. For the 5' region (from 250 bp 5' of *2DL5* exon 1 through to the end of intron 2), the variants divide according to their location in the centromeric or telomeric half of the *KIR* locus. The central region (from exon 3 of *2DL5* to 1.9 kb 5' of exon 6 in *2DS3/5*) divides according to

association with *2DS3* or *2DS5*. In the 3' segment (from intron 6 to the polyadenylation signal at the end of *2DS3S5*) the variants containing *2DS3*001* form an outgroup. (C) Shows two models for the evolution of *2DS3* and *2DS5*. Both models start with a single *2DL5A/B-2DS3/5* progenitor in the centromeric region, the only site where chimpanzee and orangutan *2DL5* has been found ([65] and AC220148). This progenitor then duplicated, with one daughter being transposed to the telomeric region. Subsequent diversification at the two loci resulted in distinct *2DL5A*, *2DL5B*, *2DS3* and *2DS5* genes. The two models differ in the sites where *2DS3* and *2DS5* arose: in model 1, *2DS5* arose in the centromeric site, *2DS3* in the telomeric site, whereas in model 2, *2DS3* arose in the centromeric site, *2DS5* in the telomeric site. Subsequent recombination between variants in the centromeric and telomeric sites gave rise to the variants observed in the modern KIR haplotypes. Model 2 requires only two such recombinations and is more parsimonious than model 1, which requires three. However, the greater diversity observed in *2DL5B-2DS5* than *2DL5A-2DS5*, favors model 1 over model 2.

doi:10.1371/journal.pone.0015115.g005

from those present in modern day humans. This analysis clearly shows that *Cen-A* and *Cen-B* motifs have been present throughout much, if not all, of human evolution.

The two main lineages in region II, the intergenic region between *3DP1* and *2DL4*, segregate with the *Tel-A* and *Tel-B* motifs. Their estimated divergence time is ~ 1.7 million years ago (Figure 6C), several million years after the human-chimpanzee separation, but before the estimated emergence of the modern human species (150,000-190,000 years ago, [43]). The two main lineages in region III represent the duplication of *2DL5* and the ancestor of *2DS3* and *2DS5*, and their divergence in the centromeric and telomeric regions of the *KIR* locus. The estimated divergence time, and thus the time of the duplication event, is also estimated at ~ 1.7 million years ago. These independent analyses of a non-coding intergenic region, and of a segment carrying two characteristic *B* haplotype genes are concordant in showing that the *Cen-A/Cen-B* dichotomy existed for several million years before emergence of the *Tel-A/Tel-B* dichotomy. The results also indicate that all four of the *KIR* gene motifs (*Cen-A*, *Cen-B*, *Tel-A*, and *Tel-B*) have been present throughout the evolution of the modern human species.

Discussion

The distinctive organization of the human *KIR* locus drives the generation of gene-content diversity. Conserved genes are situated at the middle (*3DP1* and *2DL4*) and ends (*3DL3* and *3DL2*) of the locus, creating a framework around two regions of variability, in which highly homologous *KIR* genes are packed close together in head-to-tail configuration and separated by short and highly conserved intergenic regions [33]. These properties have facilitated the numerous asymmetric recombinations that duplicated *KIR* genes, deleted *KIR* genes, and formed new hybrid *KIR* genes with novel ligand-binding and signaling functions [44]. The propensity for recombination was further appreciated from comparison of humans with apes and monkeys, from which the conserved framework was further reduced to the extremities of the locus—the 5' part of *3DL3* and the 3' part of *3DL2*—plus part of *3DP1* but all of *2DL4* in the central region [42]. The only site of unique sequence in the *KIR* locus is in the 14kb intergenic region that separates *3DP1* from *2DL4* and divides the locus into centromeric and telomeric parts of similar size [45]. This unique sequence has been the site for events of reciprocal recombination that allowed centromeric and telomeric gene-content motifs to reassort in different combinations and form new variant *KIR* haplotypes [34,46]. That seven of the eight possible combinations of four common centromeric and two telomeric motifs are represented in the 27 *KIR* haplotypes studied here, testifies to the importance of this mechanism.

Well represented in the *KIR* haplotype panel are two highly divergent gene-content haplotypes. One of these is the 'long' cB01|tB01 haplotype [25] that contains all *KIR* genes except *2DS4*, and has all the *B* haplotype specific genes and alleles. The second is the *A* haplotype (cA01|tA01) that has *2DS4* and none of

the *B* haplotype specific genes and alleles. The extent of the differences between these two haplotypes is further emphasized at the allele level, because no allele for any gene is held in common. The other five gene content haplotypes are all *B* haplotypes that lack some of the *B*-specific genes, either as a consequence of deletion (eg: cB02|tB01), or recombination that introduced either the centromeric (eg: cA01|tB01) or telomeric (eg: cB01|tA01) motif of the *A* haplotype. From epidemiological studies, a variety of disease associations have been made with differences between *A* and *B* haplotypes [47]. The prevalence of natural recombinants should allow further examination to test the contributions of centromeric and telomeric *A* and *B* motifs to these associations.

Natural division of *KIR* haplotypes, and their constituent centromeric and telomeric motifs, into two distinctive groups appears to be unique to the human species. Moreover, a mixture of *A* and *B* haplotypes is present in all human populations ($N > 150$) that have been genotyped for *KIR*. Thus the combination of *A* and *B* haplotypes appears to confer selective advantage, as is most clearly illustrated by the dominance and equal frequency of two, maximally divergent *KIR* haplotypes (one *A* and one *B*) in the Yucpa population of Venezuelan Amerindians [24,48]. From phylogenetic comparisons, we have been able to explore the evolution of the *A* and *B* haplotypes and the differences between them. Of the genes typifying either *A* or *B* haplotypes, only *2DS4* and *2DL5*, respectively, are present in chimpanzee, but in that species they are usually found linked on the same haplotypes in the centromeric region, the region containing all chimpanzee *KIR* except the *2DL4* and *3DL* framework genes (L. Abi-Rached, manuscript in preparation). In human *KIR* haplotypes, a comparable number of genes are found in the centromeric ($N = 8$) and telomeric ($N = 7$) regions, indicating that recent gene expansion in the telomeric part of the *KIR* locus is specific to the human lineage. As part of that expansion both *2DS4* and *2DL5* were moved into the telomeric region, but as part of different and mutually exclusive gene-content motifs. Consistent with the view that colonization of the telomeric region with lineage Ib and lineage III genes is human specific, the telomeric region of orangutan *KIR* haplotypes comprises only the framework genes.

Our phylogenetic and divergence-time analyses indicate that evolution of the *A*, *B* haplotype difference occurred in the centromeric region around the time of the human-chimpanzee separation some 6 million years ago. As *2DL5* is present on $\sim 50\%$ of human and chimpanzee *KIR* haplotypes, it is likely that the common ancestor also had both *2DL5*⁺ and *2DL5*⁻ haplotypes. This difference could have provided the foundation on which the *Cen-A/Cen-B* difference was built (Figure 7). Leading to the more recent evolution of the *Tel-A/Tel-B* difference was duplication of *2DL5* and its neighboring lineage III *2DS*, with movement of one of daughter pair to the telomeric region. The difference between *Tel-A* and *Tel-B*, as assessed from the time of duplication of *2DL5*, emerged ~ 1.7 million years ago, >4 million years after the *Cen-A/Cen-B* difference (Fig. 6) and >1 million years before the origin of the modern human species, in the last 200,000 years [43]. Another key event in forming the *Tel-A/Tel-B* difference was movement of

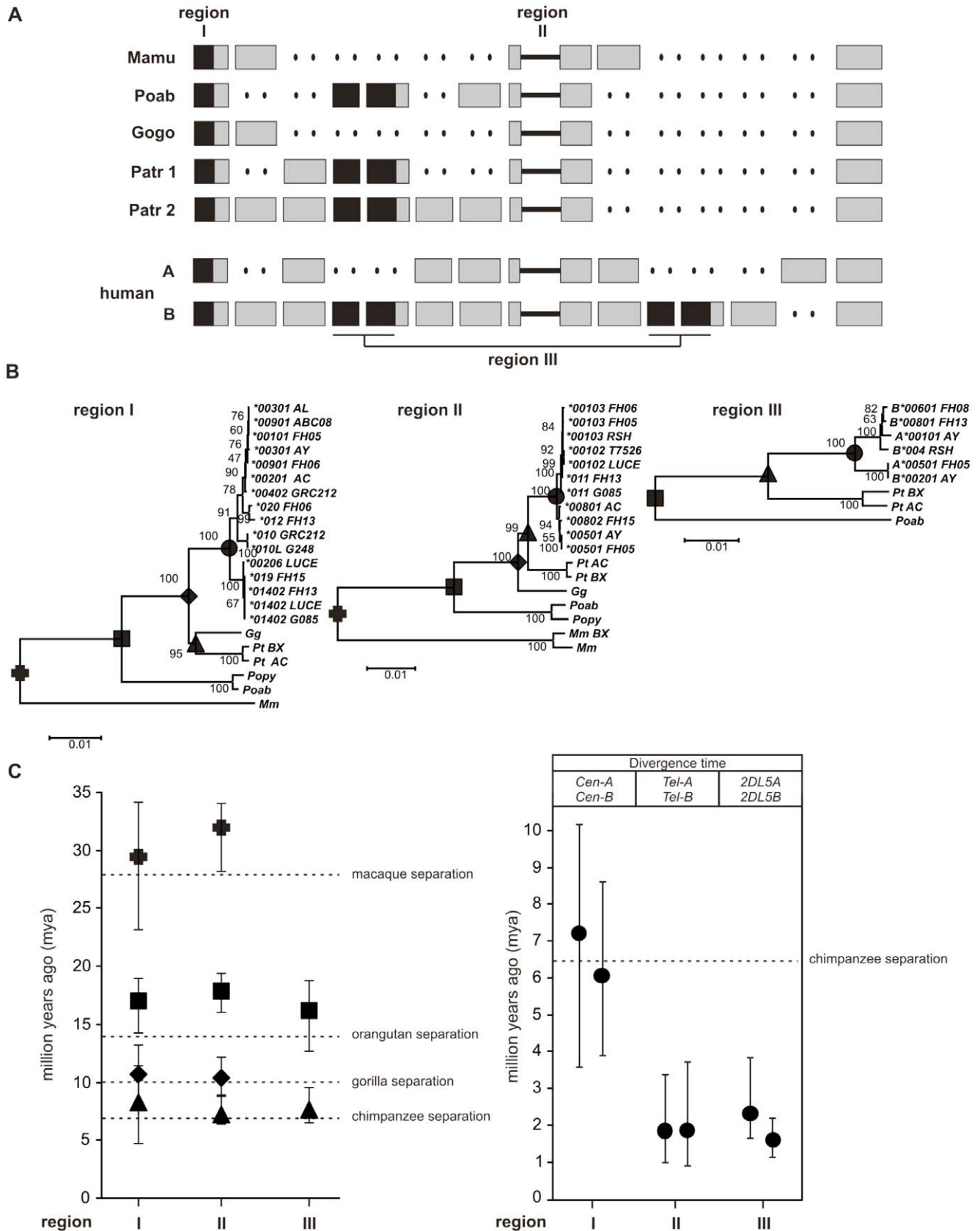


Figure 6. The *Cen-A* and *Cen-B* difference arose long before that between *Tel-A* and *Tel-B*. Three genomic segments with orthologs or paralogues in other higher primates were chosen for divergence time estimation. (A) Shows the three regions used for the analysis and the haplotypes from which they were obtained (human (AC006293/AC011501, AL133414, AY320039, and this study) chimpanzee (BX842589, AC155174), gorilla (CU92894), orangutan (EF014479, AC200148) and rhesus macaque (BX842590, BX842591)). They comprise region I, a 5.5 kb segment including the 5'

part of *3DL3* excluding intron 1, a 14 kb segment from the intergenic region between *3DP1* and *KIR2DL4* (region II), and a 16.8 segment beginning in intron 3 of *2DL5A/B* extending into the neighboring *2DS3/5* gene (region III). For the latter, both the centromeric and telomeric variants were included in the analysis. (B) Shows the phylogenetic trees for regions I, II and III, with bootstrap values for the nodes. The nodes denoted by dark shaded symbols were those used for divergence time estimates. (C) Plotted here are the divergence time estimates using the same symbols as in panel B. For comparison, the dotted horizontal lines indicate the lower limits for divergence times from the human lineage of rhesus macaque, orangutan, gorilla and chimpanzee as assessed from the fossil record [73]. The left panel examines the divergence of human KIR from KIR in other primates, the right panel examines the divergence of different forms of human KIR and the results from two independent runs of the program are shown. Analysis of region I estimates the divergence time for *Cen-A* and *Cen-B*, whereas analysis of region II estimates the divergence of *Tel-A* and *Tel-B*, and analysis of region III estimates the time of duplication for *2DL5A* and *2DL5B* as well as *2DS3* and *2DS5*. doi:10.1371/journal.pone.0015115.g006

2DS4 from a centromeric location (as in chimpanzee *KIR* haplotypes) to the telomeric location observed in human *KIR* haplotypes. Because of these species-specific locations, we could not calculate when these two forms diverged and thus distinguish between the following two possibilities: the first model is that *2DS4* moved from centromeric to telomeric location specifically on the human lineage; the second, that both types of *2DS4*⁺ haplotypes existed in the common ancestor, but those containing centromeric *2DS4* were lost on the human lineage, whereas haplotypes containing telomeric *2DS4* were lost on the chimpanzee lineage. Either way, the establishment of alternative gene content motifs in both the centromeric and telomeric regions, provided the basic structure within which increasingly elaborate *A* and *B* motifs could evolve. A key finding from this analysis is that the *A/B* haplotype dichotomy had been long in place when modern humans emerged and has existed throughout their evolution, consistent with the presence of *A* and *B* haplotypes in all modern human populations. NK cells play essential roles in both immune defense and reproduction, physiological functions that are essential for the survival of populations and species [47,49,50,51]. Because *A KIR* haplotypes favor elimination of infection [7], whereas *B KIR* favor reproductive success [18], a simple model is that balancing selection on the *A* and *B* haplotypes stems from the *A* haplotypes having evolved under pressure on the immune system, whereas the *B* haplotypes evolved under pressure on the reproductive system [48].

With the result of this study, 12 *A KIR* haplotypes and 15 *B KIR* haplotypes have been defined at the highest level of resolution. The gene-content, and allele-content motifs contained in the centromeric and telomeric regions of these 27 haplotypes are likely to account for many common haplotypes in many human populations, and should provide a strong base from which to investigate the rarer and more population-specific haplotypes. Several of these have already been investigated and include ones in which framework genes are duplicated [20,35,52,53] deleted [21,54] or fused with another gene to collapse the telomeric region [20]. There is increasing evidence that taking account of the immunogenetics of KIR can improve the outcome of allogeneic hematopoietic cell transplantation as therapy for acute myelogenous leukemia [16,55,56,57]. The results of this study will further the practical application of KIR genotyping to this valuable clinical goal.

Materials and Methods

Cell lines/source DNA

The DNA used for library construction was extracted from a panel of cell lines chosen to encompass major ethnic groups. In addition, the cell lines chosen were inferred to have both *A* and *B KIR* haplotype and to contain representatives of the most common *B* haplotypes (Table 3). DNA was prepared from B-LCLs using a Qiagen (Valencia, CA) genomic DNA extraction kit according to the manufacturers instructions. DNAs forming a diversity panel of 48 each of Caucasian, Asian, African-American,

and Hispanic were obtained from the Research Cell Bank (RCB) at the Fred Hutchinson Cancer Research Center and are commercially available (File S1).

Fosmid libraries, cloning, typing, sequencing

Fosmid Isolation was carried out as described in Raymond *et al.* [58] with modifications. Library constructions used the Epicentre copy-control vector pCC1. Sheared, end-repaired inserts were size-selected to be 30–50 kbp by pulsed-field-gel electrophoresis. To produce 10⁶-clone libraries we typically started with 20 μg of genomic DNA. Packaging was carried out with the Epicentre MaxPlax extracts and transfection was into Epicentre EPI300-T1^R *E. coli* cells. This cloning system allows induction of the fosmid copy number to ~50 copies/cell, which was critical to simplification of library screening and downstream sequencing. The high-copy-number origin was induced when growing saturated cultures from replicas of the master-library plates; lysed aliquots of induced cultures were used directly as a source of PCR template during the STS-content mapping of 3000-fosmid pools and at all subsequent stages of screening. The 3000-clone pool size was optimal as we found more complex pools reduced the number of PCR assays required for STS-content mapping, but they also decreased the reliability of initial tiling-path choices (by increasing the density of positive wells on the plate) which increased the work involved in finding the positive clone in the pool. Fosmids were derived from libraries ABC8 and G248 by computational screening of end sequences as described [59].

STS design. For KIR screening we used a panel of unlabeled primers designed locally and ordered from generic vendors that amplified each of the known KIR genes. File S1 includes the sequences and gene specificity for the primer pairs used for all of the fosmid screening carried out in this report. Samples are scored as positive or negative for a particular PCR assay based on SYBR green fluorescence. In some cases when allele-specific information was required, the PCR products were sequenced. Data were collected on an ABI 7900 instrument, operating in real-time (as opposed to end-point) mode. With 3000-clone pools, single-clone isolation proceeded in two steps: preparation and screening of 100-clone subpools. Typically, a single 384-well plate of subpools were prepared from each positive 3000-clone pool. A single positive subpool was then plated for single colonies, ~1000 of which were picked robotically to produce cultures employed for the final stage of screening. One positive-single-colony isolate was chosen for re-streaking, re-testing, and subsequent characterization.

Sequencing. Fosmids were sequenced using shotgun-sequencing protocols as outlined previously [60]. Steps included (1) subcloning of target DNA into pUC19 after shearing to ~4 kbp with a Genomic Solutions Hydroshear instrument, size-selection to 2–7 kbp on an agarose gel, and end repair using a mixture of T4 DNA polymerase and the Klenow fragment of DNA polymerase I; (2) robotic colony picking, growth of a saturated culture, and chemical lysis of a 1-μl aliquot of culture; (3) amplification of released DNA with the Amersham TempliPhi reagent; (4) thermocycling dideoxy sequencing reactions based on

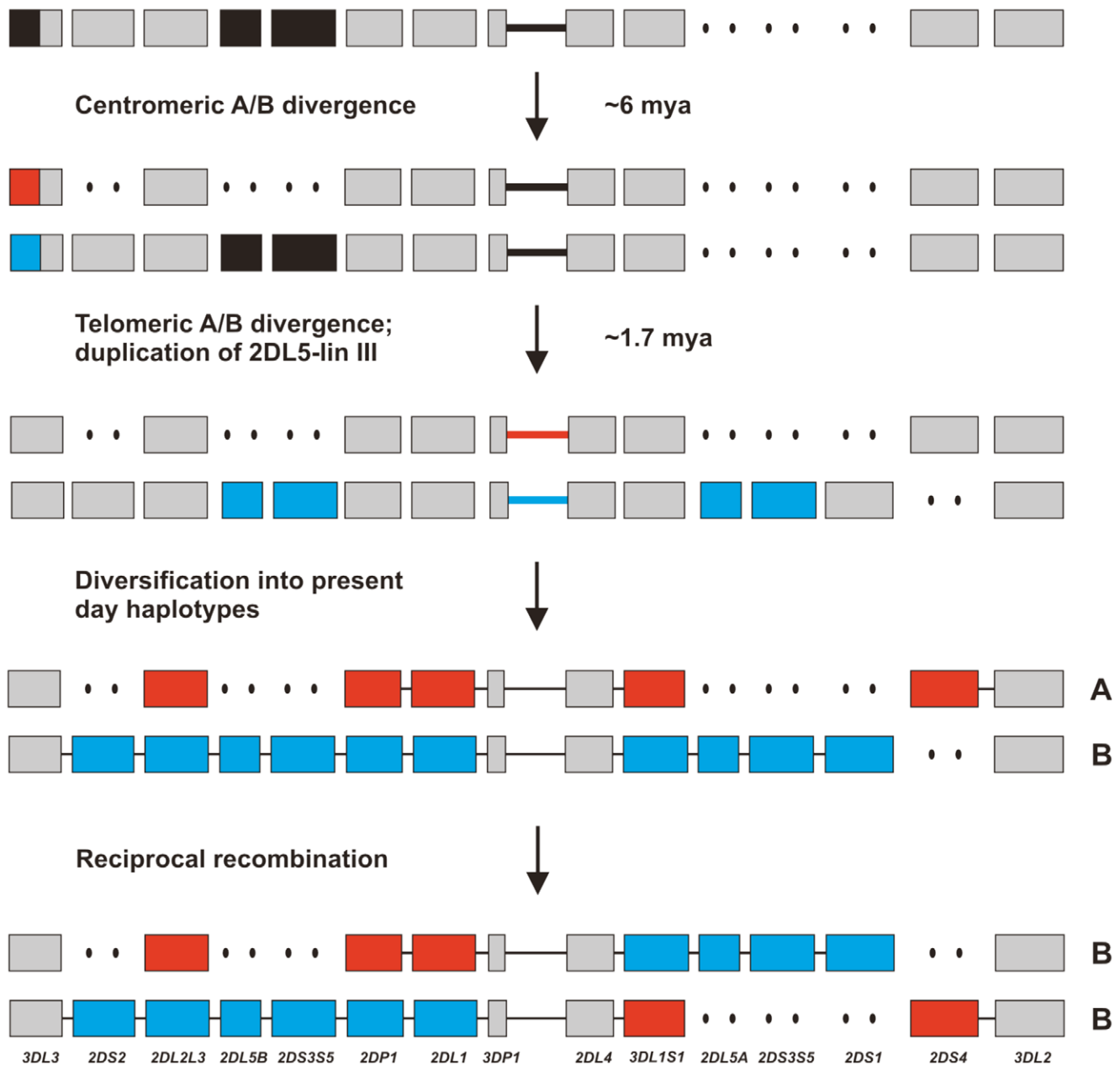


Figure 7. Model for the evolution of the A and B haplotypes. The progenitor haplotype contained the *2DL5-lineage III* segment in the centromeric region. The diversification that resulted in the formation of the *Cen-A* and *Cen-B* lineages occurred ~ 6 mya, contemporaneous with human:chimpanzee speciation. The *Tel-A:Tel-B* divergence is much younger occurring ~ 1.7 mya coincident with the timing for the duplication of the *2DL5-lineage III* progenitor that gave rise to *2DL5-2DS3* and *2DL5-2DS5*. This duplication could have produced either a *Cen-B:Tel-B* or *Cen-A:Tel-B* haplotype. Subsequent recombinations have given rise to the haplotype structures present today. doi:10.1371/journal.pone.0015115.g007

ABI BigDye chemistry; (5) separating the products on ABI 3730xl capillary-sequencing instruments; (6) assembling sequencing reads with the phred/phrap system; (7) carrying out one round of finishing using software to support primer design (average 1 finishing read per 2 fosmids). PCR-resequencing was currently carried out according to established procedures [61,62]. All data from PCR-resequencing assays were automatically interpreted by software developed in house (GeMS), which accurately interprets sequence traces from heterozygous DNAs [62,63].

Sequences have been deposited in GenBank with the following accession numbers: FH05_A_hap = GU182338; FH05_B_

hap = GU182339; FH06_A_hap = GU182340; FH06_BA1_hap = GU182341; FH08_A_hap = GU182342; FH08_BAX_hap = GU182343; FH13_A_hap = GU182344; FH13_BA2_hap = GU182345; FH15_A_hap = GU182346; FH15_B_hap = GU182347; G085_A_hap = GU182348; G085_BA1_hap = GU182349; G248_Ahap = GU182350; G248_BA2hap = GU182351; GRC212_AB_hap = GU182352; GRC212_BA1_hap = GU182353; LUCE_A_hap = GU182354; LUCE_Bdel_hap = GU182355; RSH_A_hap = GU182356; RSH_BA2_hap = GU182357; T7526_A_hap = GU182358; T7526_Bdel_hap = GU182359; ABC08_A1hap = GU182360; ABC08_AB_

Table 3. Fosmid clone resources.

Source	Ethnicity	Library density(colonies/ml)	Host Cell
ABC08	Yoruba	a	a
FH05	Caucasian	2500	EPI300
FH06	Hispanic	2170	EPI300
FH08	American-Black	1500/1800	XL1_B/EPI300
FH13	American-Black	1500/1800	XL1_B/EPI300
FH15	Caucasian	1640	XL1_B
G085	Spanish Gypsy	3800	EPI300
G248	unknown	a	a
GRC212	Guarani South Amerindian	3000	XL1_B
LUCE	unknown	3600	EPI300
RSH	South African Zulu	3500	EPI300
T7526	Chinese	3850	EPI300

^aclones were obtained from end sequenced fosmid libraries described in [59]. doi:10.1371/journal.pone.0015115.t003

hap_central_partial = GU182361; ABC08_AB_hap_telomere_partial GU182362. Alleles and haplotypes have been named according to the guidelines established by the KIR Nomenclature Committee [64] and deposited into IPD-KIR (<http://www.ebi.ac.uk/ipd/kir/>).

Computational analyses

Phylogenetic analysis. The dataset included the 24 newly sequenced haplotypes and the three human haplotypes deposited in GenBank (AC006293/AC011501, AL133414, AY320039) [32,33,34]. For the divergence time estimation chimpanzee (BX842589, AC155174) [65], gorilla (CU92894), orangutan (EF014479, AC200148) [42] and rhesus macaque (BX842590, BX842591) [65] sequences were used. Sequences of the individual genes were aligned using CLUSTAL X [66] or MAFFT [67] and manually corrected in BIOEDIT (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). The alignment was then divided into domains generally following intron-exon boundaries, except for intron 6 which was divided further into 3 regions. The first of these (intron 6a) starts at the beginning of the intron and ends at the beginning of the deletion common to *MmKIR3DL1* and *MmKIR3DL10* (approx. 750 bp), the second (intron 6b) begins here and ends at the beginning of the LINE insertion common to *KIR3DL2* and *PtKIR3DL1/2* (approx. 2.9 kb) and the third (intron 6c) starts after the LINE insertion and ends at the end of the intron (approx. 600 bp). Each of these alignments was used for neighbor-joining (NJ) and parsimony analyses. The NJ analysis was performed using MEGA version 4 (<http://www.megasoftware.net/>) [68] with 500 replicates, pairwise deletion, midpoint rooting, and the Tamura-Nei method. PAUP*4.0b10 (<http://paup.csit.fsu.edu/>) and the tree bisection-reconnection branch-swapping algorithm were used for parsimony analyses with 500 replicates and a heuristic search. Comparison of the resulting trees revealed no differences. Only neighbor-joining trees are presented in the figures.

Nucleotide diversity. The average number of nucleotide differences per site between two sequences, or nucleotide diversity, Π , and its sampling variance and standard error [69] were calculated using DNASP (<http://www.ub.es/dnasp/>) [38]. Genes were examined individually and framework genes were divided based on their location on either *A* or *B* segments. Also, *2DL1* and *2DPI* were analyzed separately according to their presence on *Cen-A* vs. *Cen-B* segments. Finally, the *2DL5* sequences were subgrouped according to their linkage to either *2DS3* or *2DS5* and their presence in either the centromeric or telomeric interval.

Divergence time estimation. Divergence time estimation was completed using MCMCTREE in the PAML package [70,71]. Starting kappa and alpha values were estimated using baseml in the PAML package [70,71] and rgene was estimated from neighbor joining trees from MEGA [68]. Three datasets including both human and non-human primate sequences were analyzed, a 5.5 kb segment extending from 250 bp 5' of the start codon to 450 bp 3' of exon 5 (excluding intron 1) of *KIR3DL3*, a 14 kb segment from the region between *3DPI* and *KIR2DL4*, and a 16.9 kb segment beginning 100bp 5' of exon 3 of *KIR2DL5* extending to 550 bp 3' of exon 5 of the neighboring lineage III gene. Each dataset was analyzed for recombinants using RDP [72] and recombinants were excluded from the analysis. NJ trees were constructed in MEGA and parsimony trees were constructed in PAUP as described above. The calibration times used for the analysis were human-chimpanzee split 6.5-10 mya, gorilla speciation >10 mya, orangutan speciation <18 mya, and rhesus macaque speciation 23-34 mya [73]. Datasets and control files are available upon request to the authors.

Supporting Information

File S1 This file includes data arranged in three tables, including supplementary Table 1 - KIR haplotype summary statistics, supplementary, Table 2 - KIR gene PCR-SSP for library screening and haplotyping, and supplementary Table 3 - KIR diversity cell panel. (DOC)

Supporting Information

File S1 This file includes data arranged in three tables, including supplementary Table 1 - KIR haplotype summary statistics, supplementary, Table 2 - KIR gene PCR-SSP for library screening and haplotyping, and supplementary Table 3 - KIR diversity cell panel. (DOC)

Author Contributions

Conceived and designed the experiments: DEG CWP. Performed the experiments: CWP QV RW. Analyzed the data: CWP LAG QV RW PP DEG. Contributed reagents/materials/analysis tools: LA PJN SGEM JSM. Wrote the paper: PP LAG CWP DEG.

References

- Sambrook JG, Beck S (2007) Evolutionary vignettes of natural killer cell receptors. *Curr Opin Immunol* 19: 553–560. Epub 2007 Sep 2007.
- Joncker NT, Raullet DH (2008) Regulation of NK cell responsiveness to achieve self-tolerance and maximal responses to diseased target cells. *Immunol Rev* 224: 85–97.
- Vely F, Peyrat M, Couedel C, Moret J, Halary F, et al. (2001) Regulation of inhibitory and activating killer-cell Ig-like receptor expression occurs in T cells after termination of TCR rearrangements. *J Immunol* 166: 2487–2494.
- Barten R, Torkar M, Haude A, Trowsdale J, Wilson MJ (2001) Divergent and convergent evolution of NK-cell receptors. *Trends Immunol* 22: 52–57.
- Hansasuta P, Dong T, Thananchai H, Weekes M, Willberg C, et al. (2004) Recognition of HLA-A3 and HLA-A11 by KIR3DL2 is peptide-specific. *Eur J Immunol* 34: 1673–1679.
- Thananchai H, Gillespie G, Martin MP, Bashirova A, Yawata N, et al. (2007) Cutting Edge: Allele-specific and peptide-dependent interactions between KIR3DL1 and HLA-A and HLA-B. *J Immunol* 178: 33–37.
- Khakoo SI, Thio CL, Martin MP, Brooks CR, Gao X, et al. (2004) HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* 305: 872–874.
- Martin MP, Gao X, Lee JH, Nelson GW, Detels R, et al. (2002) Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat Genet* 31: 429–434.
- Luszczek W, Manczak M, Cislo M, Nockowski P, Wisniewski A, et al. (2004) Gene for the activating natural killer cell receptor, KIR2DS1, is associated with susceptibility to psoriasis vulgaris. *Hum Immunol* 65: 758–766.

10. Martin MP, Nelson G, Lee JH, Pellett F, Gao X, et al. (2002) Cutting edge: susceptibility to psoriatic arthritis: influence of activating killer Ig-like receptor genes in the absence of specific HLA-C alleles. *J Immunol* 169: 2818–2822.
11. Momot T, Koch S, Hunzelmann N, Krieg T, Ulbricht K, et al. (2004) Association of killer cell immunoglobulin-like receptors with scleroderma. *Arthritis Rheum* 50: 1561–1565.
12. Nelson GW, Martin MP, Gladman D, Wade J, Trowsdale J, et al. (2004) Cutting edge: heterozygote advantage in autoimmune disease: hierarchy of protection/susceptibility conferred by HLA and killer Ig-like receptor combinations in psoriatic arthritis. *J Immunol* 173: 4273–4276.
13. Suzuki Y, Hamamoto Y, Ogasawara Y, Ishikawa K, Yoshikawa Y, et al. (2004) Genetic polymorphisms of killer cell immunoglobulin-like receptors are associated with susceptibility to psoriasis vulgaris. *J Invest Dermatol* 122: 1133–1136.
14. van der Slik AR, Koeleman BP, Verduijn W, Bruining GJ, Roep BO, et al. (2003) KIR in type 1 diabetes: disparate distribution of activating and inhibitory natural killer cell receptors in patients versus HLA-matched control subjects. *Diabetes* 52: 2639–2642.
15. Yen JH, Moore BE, Nakajima T, Scholl D, Schaid DJ, et al. (2001) Major histocompatibility complex class I-recognizing receptors are disease risk genes in rheumatoid arthritis. *J Exp Med* 193: 1159–1167.
16. Cooley S, Trachtenberg E, Bergemann TL, Sacternum K, Klein J, et al. (2009) Donors with group B KIR haplotypes improve relapse-free survival after unrelated hematopoietic cell transplantation for acute myelogenous leukemia. *Blood* 113: 726–732.
17. Moretta A, Pende D, Locatelli F, Moretta L (2009) Activating and inhibitory killer immunoglobulin-like receptors (KIR) in haploidentical haemopoietic stem cell transplantation to cure high-risk leukaemias. *Clin Exp Immunol* 157: 325–331.
18. Hiby SE, Walker JJ, O'Shaughnessy K M, Redman CW, Carrington M, et al. (2004) Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. *J Exp Med* 200: 957–965.
19. Shilling HG, Guethlein LA, Cheng NW, Gardiner CM, Rodriguez R, et al. (2002) Allelic polymorphism synergizes with variable gene content to individualize human KIR genotype. *J Immunol* 168: 2307–2315.
20. Norman PJ, Abi-Rached L, Gendzekhadze K, Hammond JA, Moesta AK, et al. (2009) Meiotic recombination generates rich diversity in NK cell receptor genes, alleles, and haplotypes. *Genome Res* 19: 757–769.
21. Norman PJ, Carrington CV, Byng M, Maxwell LD, Curran MD, et al. (2002) Natural killer cell immunoglobulin-like receptor (KIR) locus profiles in African and South Asian populations. *Genes Immun* 3: 86–95.
22. Beatty PG, Boucher KM, Mori M, Milford EL (2000) Probability of finding HLA-mismatched related or unrelated marrow or cord blood donors. *Hum Immunol* 61: 834–840.
23. Uhrberg M, Valiante NM, Shum BP, Shilling HG, Lienert-Weidenbach K, et al. (1997) Human diversity in killer cell inhibitory receptor genes. *Immunity* 7: 753–763.
24. Gendzekhadze K, Norman PJ, Abi-Rached L, Layrisse Z, Parham P (2006) High KIR diversity in Amerindians is maintained using few gene-content haplotypes. *Immunogenetics* 58: 474–480.
25. Hsu KC, Liu XR, Selvakumar A, Mickelson E, O'Reilly RJ, et al. (2002) Killer Ig-like receptor haplotype analysis by gene content: evidence for genomic diversity with a minimum of six basic framework haplotypes, each with multiple subsets. *J Immunol* 169: 5118–5129.
26. Martin MP, Single RM, Wilson MJ, Trowsdale J, Carrington M (2008) KIR haplotypes defined by segregation analysis in 59 Centre d'Etude Polymorphisme Humain (CEPH) families. *Immunogenetics* 60: 767–774.
27. Middleton D, Meenagh A, Gourraud PA (2007) KIR haplotype content at the allele level in 77 Northern Irish families. *Immunogenetics* 59: 145–158.
28. Norman PJ, Cook MA, Carey BS, Carrington CV, Verity DH, et al. (2004) SNP haplotypes and allele frequencies show evidence for disruptive and balancing selection in the human leukocyte receptor complex. *Immunogenetics* 56: 225–237.
29. Uhrberg M, Parham P, Wernet P (2002) Definition of gene content for nine common group B haplotypes of the Caucasoïd population: KIR haplotypes contain between seven and eleven KIR genes. *Immunogenetics* 54: 221–229.
30. Whang DH, Park H, Yoon JA, Park MH (2005) Haplotype analysis of killer cell immunoglobulin-like receptor genes in 77 Korean families. *Hum Immunol* 66: 146–154.
31. Yawata M, Yawata N, Abi-Rached L, Parham P (2002) Variation within the human killer cell immunoglobulin-like receptor (KIR) gene family. *Crit Rev Immunol* 22: 463–482.
32. Martin AM, Kulski JK, Gaudieri S, Witt CS, Freitas EM, et al. (2004) Comparative genomic analysis, diversity and evolution of two KIR haplotypes A and B. *Gene* 335: 121–131.
33. Wilson MJ, Torkar M, Haude A, Milne S, Jones T, et al. (2000) Plasticity in the organization and sequences of human KIR/ILT gene families. *Proc Natl Acad Sci U S A* 97: 4778–4783.
34. Hsu KC, Chida S, Geraghty DE, Dupont B (2002) The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. *Immunol Rev* 190: 40–52.
35. Martin MP, Bashirova A, Traherne J, Trowsdale J, Carrington M (2003) Cutting edge: expansion of the KIR locus by unequal crossing over. *J Immunol* 171: 2192–2195.
36. Bashirova AA, Martin MP, McVicar DW, Carrington M (2006) The Killer Immunoglobulin-Like Receptor Gene Cluster: Tuning the Genome for Defense (*). *Annu Rev Genomics Hum Genet* 7: 277–300.
37. Marsh SG, Albert ED, Bodmer WF, Bontrop RE, Dupont B, et al. (2005) Nomenclature for factors of the HLA system, 2004. *Tissue Antigens* 65: 301–369.
38. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
39. Gomez-Lozano N, Gardiner CM, Parham P, Vilches C (2002) Some human KIR haplotypes contain two KIR2DL5 genes: KIR2DL5A and KIR2DL5B. *Immunogenetics* 54: 314–319.
40. Du Z, Sharma SK, Spellman S, Reed EF, Rajalingam R (2008) KIR2DL5 alleles mark certain combination of activating KIR genes. *Genes Immun* 9: 470–480.
41. Ordonez D, Meenagh A, Gomez-Lozano N, Castano J, Middleton D, et al. (2008) Duplication, mutation and recombination of the human orphan gene KIR2DS3 contribute to the diversity of KIR haplotypes. *Genes Immun* 9: 431–437.
42. Guethlein LA, Older Aguilar AM, Abi-Rached L, Parham P (2007) Evolution of killer cell Ig-like receptor (KIR) genes: definition of an orangutan KIR haplotype reveals expansion of lineage III KIR associated with the emergence of MHC-C. *J Immunol* 179: 491–504.
43. Campbell MC, Tishkoff SA (2008) African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* 9: 403–433.
44. Abi-Rached L, Parham P (2005) Natural selection drives recurrent formation of activating killer cell immunoglobulin-like receptor and Ly49 from inhibitory homologues. *J Exp Med* 201: 1319–1332.
45. Trowsdale J, Barten R, Haude A, Stewart CA, Beck S, et al. (2001) The genomic context of natural killer receptor extended gene families. *Immunol Rev* 181: 20–38.
46. Yawata M, Yawata N, McQueen KL, Cheng NW, Guethlein LA, et al. (2002) Predominance of group A KIR haplotypes in Japanese associated with diverse NK cell repertoires of KIR expression. *Immunogenetics* 54: 543–550.
47. Parham P (2005) MHC class I molecules and KIRs in human history, health and survival. *Nat Rev Immunol* 5: 201–214.
48. Gendzekhadze K, Norman PJ, Abi-Rached L, Graef T, Moesta AK, et al. (2009) Co-evolution of KIR2DL3 with HLA-C in a human population retaining minimal essential diversity of KIR and HLA class I ligands. *Proc Natl Acad Sci U S A* 106: 18692–18697.
49. Cooper MA, Colonna M, Yokoyama WM (2009) Hidden talents of natural killers: NK cells in innate and adaptive immunity. *EMBO Rep* 10: 1103–1110.
50. Moffett A, Loke C (2006) Immunology of placentation in eutherian mammals. *Nat Rev Immunol* 6: 584–594.
51. Sun JC, Lanier LL (2009) Natural killer cells remember: an evolutionary bridge between innate and adaptive immunity? *Eur J Immunol* 39: 2059–2064.
52. Gomez-Lozano N, Estefania E, Williams F, Halfpenny I, Middleton D, et al. (2005) The silent KIR3DP1 gene (CD158c) is transcribed and might encode a secreted receptor in a minority of humans, in whom the KIR3DP1, KIR2DL4 and KIR3DL1/KIR3DS1 genes are duplicated. *Eur J Immunol* 35: 16–24.
53. Williams F, Maxwell LD, Halfpenny IA, Meenagh A, Sleator C, et al. (2003) Multiple copies of KIR 3DL/S1 and KIR 2DL4 genes identified in a number of individuals. *Hum Immunol* 64: 729–732.
54. Gomez-Lozano N, de Pablo R, Puente S, Vilches C (2003) Recognition of HLA-G by the NK cell receptor KIR2DL4 is not essential for human reproduction. *Eur J Immunol* 33: 639–644.
55. Hsu KC, Keever-Taylor CA, Wilton A, Pinto C, Heller G, et al. (2005) Improved outcome in HLA-identical sibling hematopoietic stem-cell transplantation for acute myelogenous leukemia predicted by KIR and HLA genotypes. *Blood* 105: 4878–4884.
56. Locatelli F, Pende D, Maccario R, Mingari MC, Moretta A, et al. (2009) Haploidentical hemopoietic stem cell transplantation for the treatment of high-risk leukemias: how NK cells make the difference. *Clin Immunol* 133: 171–178.
57. Ruggeri L, Capanni M, Urbani E, Perruccio K, Shlomchik WD, et al. (2002) Effectiveness of donor natural killer cell alloreactivity in mismatched hematopoietic transplants. *Science* 295: 2097–2100.
58. Raymond CK, Subramanian S, Paddock M, Qiu R, Deodato C, et al. (2005) Targeted, haplotype-resolved resequencing of long segments of the human genome. *Genomics* 86: 759–766.
59. Bovee D, Zhou Y, Haugen E, Wu Z, Hayden HS, et al. (2008) Closing gaps in the human genome with fosmid resources generated from multiple individuals. *Nat Genet* 40: 96–101.
60. Daza-Vamenta R, Glusman G, Rowen L, Guthrie B, Geraghty DE (2004) Genetic divergence of the rhesus macaque major histocompatibility complex. *Genome Res* 14: 1501–1515.
61. Geraghty DE, Daza R, Williams LM, Vu Q, Ishitani A (2002) Genetics of the immune response: identifying immune variation within the MHC and throughout the genome. *Immunol Rev* 190: 69–85.
62. Pyo CW, Moore Y, Williams L, Hyodo H, Li S, et al. (2006) HLA-E, F, and G polymorphism: genomic sequence defines haplotype structure and variation spanning the nonclassical class I genes. *Immunogenetics* in press.
63. Smith WP, Vu Q, Li S, Hansen JA, Zhao LP, et al. (2006) Towards Understanding MHC Disease Associations: Partial Resequencing of 46 Distinct HLA Haplotypes. *Genomics* 87: 561–571.

64. Marsh SG, Parham P, Dupont B, Geraghty DE, Trowsdale J, et al. (2003) Killer-cell immunoglobulin-like receptor (KIR) nomenclature report, 2002. *Immunogenetics* 55: 220–226.
65. Sambrook JG, Bashirova A, Palmer S, Sims S, Trowsdale J, et al. (2005) Single haplotype analysis demonstrates rapid evolution of the killer immunoglobulin-like receptor (KIR) loci in primates. *Genome Res* 15: 25–35.
66. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23: 403–405.
67. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059–3066.
68. Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9: 299–306.
69. Nei M (1987) *Molecular Evolutionary Genetics*.
70. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13: 555–556.
71. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586–1591.
72. Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 21: 260–262.
73. Benton MJ, Donoghue PC (2007) Paleontological evidence to date the tree of life. *Mol Biol Evol* 24: 26–53.