

RESEARCH ARTICLE

Open Access

Rapid evolution of *BRCA1* and *BRCA2* in humans and other primates

Dianne I Lou¹, Ross M McBee¹, Uyen Q Le¹, Anne C Stone², Gregory K Wilkerson³, Ann M Demogines¹ and Sara L Sawyer^{1*}

Abstract

Background: The maintenance of chromosomal integrity is an essential task of every living organism and cellular repair mechanisms exist to guard against insults to DNA. Given the importance of this process, it is expected that DNA repair proteins would be evolutionarily conserved, exhibiting very minimal sequence change over time. However, *BRCA1*, an essential gene involved in DNA repair, has been reported to be evolving rapidly despite the fact that many protein-altering mutations within this gene convey a significantly elevated risk for breast and ovarian cancers.

Results: To obtain a deeper understanding of the evolutionary trajectory of *BRCA1*, we analyzed complete *BRCA1* gene sequences from 23 primate species. We show that specific amino acid sites have experienced repeated selection for amino acid replacement over primate evolution. This selection has been focused specifically on humans and our closest living relatives, chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*). After examining *BRCA1* polymorphisms in 7 bonobo, 44 chimpanzee, and 44 rhesus macaque (*Macaca mulatta*) individuals, we find considerable variation within each of these species and evidence for recent selection in chimpanzee populations. Finally, we also sequenced and analyzed *BRCA2* from 24 primate species and find that this gene has also evolved under positive selection.

Conclusions: While mutations leading to truncated forms of *BRCA1* are clearly linked to cancer phenotypes in humans, there is also an underlying selective pressure in favor of amino acid-altering substitutions in this gene. A hypothesis where viruses are the drivers of this natural selection is discussed.

Keywords: DNA damage response, Simian primates, Cell cycle, Positive selection

Background

Defects in the *BRCA1* or *BRCA2* genes are responsible for most hereditary forms of breast cancer and account for as many as 10% of all breast cancer cases [1]. Women with a strong family history of cancer who possess a harmful *BRCA1* or *BRCA2* allele are at high risk for developing breast cancer within their lifetime (80% and 60%, respectively) [2,3]. In addition, *BRCA1* mutation carriers have a 30-40% chance of developing ovarian cancer, while *BRCA2* mutations also increase the risk of ovarian, pancreatic, prostate, and male breast cancer [2]. Cancers occur when heterozygous individuals experience a somatic loss of heterozygosity event at the *BRCA1* or *BRCA2* locus, leaving

only the abnormal allele intact. Because both gene products play a critical role in key cellular processes such as DNA repair, cell cycle control, and transcriptional regulation, it is clear why inactivating mutations are so detrimental. The importance of these proteins is further evidenced by the fact that both *BRCA1* and *BRCA2* null mice are embryonic lethal [4].

Given their indispensable functions in maintaining the integrity of the genome, one might expect strict evolutionary conservation of *BRCA1* and *BRCA2* over time. Indeed, some regions of *BRCA1* have experienced purifying selection strong enough to operate even on synonymous mutations [5]. However, contrary to this line of reasoning, a number of groups have documented the rapid evolution of *BRCA1* [6-11] and *BRCA2* [10] in mammals. Rapid evolution occurs when a gene experiences positive natural selection for new, advantageous mutations that

* Correspondence: sarsaw2@gmail.com

¹Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712, USA

Full list of author information is available at the end of the article

arise in a population. Because advantageous mutations commonly involve a change in protein sequence (non-synonymous mutations), recurrent rounds of positive selection in a gene lead to rapid evolution of the encoded protein sequence over time. For *BRCA1*, the evolutionary rate was particularly elevated on the branches leading to humans and chimpanzees (*Pan troglodytes*) [6]. The identification of this signature in *BRCA1* suggests that some alleles and polymorphisms currently circulating within the human population may offer a selectable advantage. However, both the cause and consequence of this unexpected mode of evolution seen in *BRCA1* remain unknown.

Here, we report an extensive evolutionary analysis of the primate *BRCA1* gene. In previous studies of *BRCA1* evolution, only exon 11 was examined with a limited number of primate species included in the analyses [6-11]. To extend previous studies, we have generated full-length *BRCA1* sequences for 17 additional primate species. Using this more extensive dataset, we validate the finding of positive selection in humans and their closest ape relatives (in our study, chimpanzees and also bonobos (*Pan paniscus*)). We also show that specific codons in *BRCA1* have experienced recurrent positive selection over evolutionary time, both within and outside of exon 11, resulting in a small number of highly variable residue positions in an otherwise highly conserved protein. In addition, we sequenced exon 11 of *BRCA1* from populations of chimpanzee, bonobo, and rhesus macaque (*Macaca mulatta*) individuals and found that several unique polymorphisms exist within these populations. Two polymorphisms in the chimpanzee population were found to be in Hardy-Weinberg disequilibrium suggesting that selection may still be operating on this gene in modern times. Lastly, exon 11 of *BRCA2*, another important genetic determinant for hereditary breast and ovarian cancers, was also sequenced from diverse primate species. This gene also bears the surprising signature of positive selection. It is unclear why these critical genes bear this unusual evolutionary signature, but we present one possible hypothesis involving interactions between DNA repair proteins and viruses.

Results

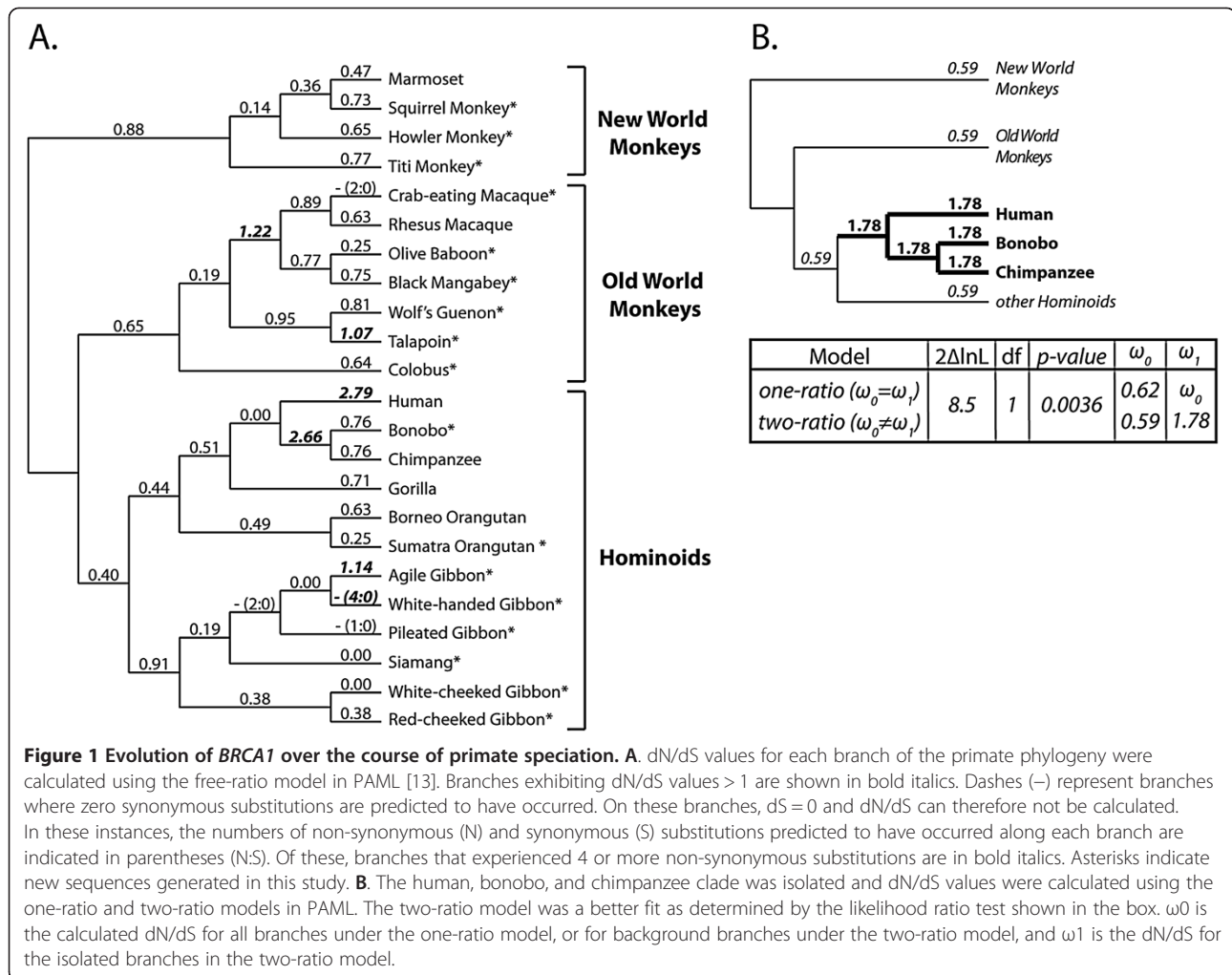
BRCA1 is evolving under positive selection in primates

To expand our understanding of the positive selection shaping *BRCA1* in primates, we obtained cell lines from 17 simian primate species, harvested total RNA, and created cDNA libraries. From these, the 5.6 kilobase full-length coding region of *BRCA1* was sequenced. These sequences were combined with full-length *BRCA1* sequences from six primate species with available genome projects, creating an alignment of 23 full-length *BRCA1* sequences. 17 out of the 23 full-length sequences have never before been analyzed (asterisks in Figure 1A).

The type of selection that a gene has experienced can be inferred from its rate of accumulation of non-synonymous (changing the encoded amino acid; denoted dN) and synonymous (silent; dS) substitutions over time. Protein-altering mutations are far less likely to be tolerated than synonymous mutations, and so $dN/dS < 1$ for the vast majority of genes encoded by human and other mammalian genomes [12]. Some genes, such as pseudogenes, evolve neutrally with $dN/dS \sim 1$ because there is not strong selection for or against new mutations in these genes. Finally, selection in favor of non-synonymous mutations results in a $dN/dS > 1$. These genes are classified as being under positive selection, and are experiencing continued selection for “innovation” at the protein sequence level. In these genes, not only has the penalty against protein-altering mutations been relaxed, but this very type of mutation is being selectively retained. Using PAML [13], we fit the full-length *BRCA1* alignment (Additional file 1) to models of positive selection where a subset of codons is allowed to evolve with $dN/dS > 1$ (M2a, M8) and to null models not allowing positive selection (M1a, M7, M8a). Likelihood ratio tests revealed that the dataset fit the positive selection models significantly better than the null models ($p < 0.05$, Table 1). Thus, *BRCA1* has experienced selection in favor of non-synonymous mutations over the speciation of simian primates.

We next estimated dN/dS values on each branch on the primate evolutionary tree using the free-ratio model in PAML. As expected, most branches exhibited a $dN/dS < 1$ (Figure 1A). The branch leading to humans had the most elevated signal with a dN/dS of 2.79. The second highest value of dN/dS on the *BRCA1* tree is found on the branch leading to the last common ancestor of bonobos and chimpanzees, with a dN/dS of 2.66. Because the free-ratio model is highly parameterized, we next compared one-ratio and two-ratio models to determine whether selection has differentially affected the human, chimpanzee, and bonobo clade. As shown in Figure 1B, our simian primate dataset fit the two-ratio model significantly better than the one-ratio model, with the human, chimpanzee, and bonobo clade exhibiting a dN/dS of 1.78, while all other branches had a dN/dS of 0.59. In summary, our extended primate dataset shows that *BRCA1* is experiencing positive selection, and that the most intense selection has operated on the human/chimpanzee/bonobo clade.

Based on a comparison of extant and predicted ancestral sequences, humans are estimated to have accumulated 25 substitutions in the *BRCA1* gene since their divergence from chimpanzees and bonobos six million years ago, 22 of which are non-synonymous (Figure 2A). In order to understand how unusual this is, we looked at the evolution of other genes, specifically ones encoding *BRCA1*-interacting proteins, along the branch leading to



humans. Because we do not have extended sequence sets for all of these genes, we took a simpler approach. For each gene, we aligned the human, chimpanzee, and gorilla sequences and manually counted the number of human-specific substitutions (any position where the human gene sequence differs from both the chimpanzee and gorilla gene sequence). These were categorized as non-synonymous (N) or synonymous (S) based on how they affected the codon in which they were found. When these values are normalized to gene size, *BRCA1* has the highest enrichment of non-synonymous substitutions [(N/kb)/(S/kb)]. Care must be taken in comparing this metric between genes, because different genes have different equilibrium codon frequencies, and therefore have different mutational opportunities for synonymous and non-synonymous mutations. However, the *BRCA1* gene has an enrichment ratio that is more than 4-fold higher than any of the other genes shown (Figure 2B).

BRCA1 encodes a 220 kDa protein with two conserved domains: an N-terminal RING domain and two tandem C-terminal BRCT domains (Figure 2C). The RING

domain has E3 ubiquitin ligase activity that is essential in the DNA damage response. The BRCT motifs function as a protein-protein interaction module that binds phosphorylated proteins involved in DNA repair, cell cycle control, chromatin remodeling, and transcription. There is also a coiled-coil region between these two domains. Interestingly, all but one of the non-synonymous substitutions predicted to have occurred in the human/bonobo/chimpanzee clade fall outside of these known structural motifs (Figure 2C).

Human variation at selected sites in *BRCA1*

The M8 model allows a class of codons to evolve under positive selection (dN/dS > 1). 10 codons were identified as belonging to this class with a high posterior probability (P = 0.85 or above). These codons do not lie in the region of *BRCA1* where it was previously reported that selection might be acting against synonymous mutations [5], potentially given rise to a false signature of dN/dS > 1. Instead, all 10 sites show high variability between primate species at the protein level, often encoding very

Table 1 PAML Analysis of BRCA1 and BRCA2

M1a-M2a	ω_0^a	codon freq. ^b	2 $\Delta\ln L^c$	df ^c	p-value ^c	M1a-M2a	ω_0^a	codon freq. ^b	2 $\Delta\ln L^c$	df ^c	p-value ^c
BRCA1	0.4	f61	10.0	2	0.0066	BRCA2	0.4	f61	21.3	2	<0.0001
	0.4	f3x4	6.1	2	0.0466		0.4	f3x4	16.1	2	0.0003
	1.6	f61	10.0	2	0.0066		1.6	f61	21.3	2	<0.0001
	1.6	f3x4	6.1	2	0.0466		1.6	f3x4	16.1	2	0.0003
M7-M8	ω_0^a	codon freq. ^b	2 $\Delta\ln L^c$	df ^c	p-value ^c	M7-M8	ω_0^a	codon freq. ^b	2 $\Delta\ln L^c$	df ^c	p-value ^c
BRCA1	0.4	f61	10.6	2	0.0049	BRCA2	0.4	f61	23.3	2	<0.0001
	0.4	f3x4	6.2	2	0.0447		0.4	f3x4	18.6	2	<0.0001
	1.6	f61	10.6	2	0.0049		1.6	f61	23.3	2	<0.0001
	1.6	f3x4	6.2	2	0.0447		1.6	f3x4	18.6	2	<0.0001
M8a-M8	ω_0^a	codon freq. ^b	2 $\Delta\ln L^c$	df ^c	p-value ^c	M8a-M8	ω_0^a	codon freq. ^b	2 $\Delta\ln L^c$	df ^c	p-value ^c
BRCA1	0.4	f61	10.1	1	0.0015	BRCA2	0.4	f61	19.9	1	<0.0001
	0.4	f3x4	6.2	1	0.013		0.4	f3x4	15.1	1	0.0001
	1.6	f61	10.1	1	0.0015		1.6	f61	19.9	1	<0.0001
	1.6	f3x4	6.2	1	0.013		1.6	f3x4	15.1	1	0.0001

^aInitial seed value for ω (dN/dS).

^bModel of codon frequency.

^cTwice the difference in the natural logs of the likelihoods ($2 \times \Delta\ln L$) of the two models being compared (a model that allows positive selection (M2a or M8) is compared to a null model (M1a, M7, M8a)). This value is used in a likelihood ratio test along with the degrees of freedom (df). The p-value indicates the confidence with which the null model can be rejected.

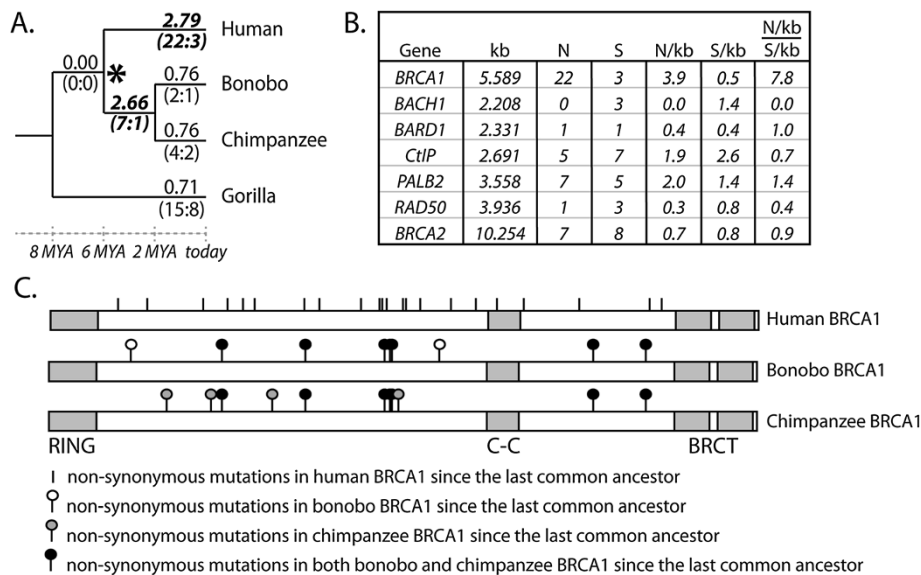


Figure 2 BRCA1 evolution in the human, bonobo, and chimpanzee clade. A. dN/dS values for *BRCA1* were calculated on each branch of the primate tree using the free-ratio model in PAML. dN/dS values > 1 are shown in bold italics. The numbers of non-synonymous (N) and synonymous (S) substitutions predicted to have occurred along each branch are indicated in parentheses (N:S). The asterisk represents the last common ancestor of humans, bonobos, and chimpanzees. MYA, million years ago. **B.** The number of human-specific non-synonymous (N) and synonymous (S) substitutions in *BRCA1* and other genes encoding *BRCA1*-interacting proteins. The length of each gene is shown in kilobases (kb). Non-synonymous and synonymous substitutions are shown as number of substitutions per kilobases (N/kb and S/kb, respectively). An "enrichment ratio" of N/kb over S/kb was also calculated. **C.** A domain diagram of *BRCA1* is shown with the RING domain, coiled-coil domain (C-C), and BRCT domains indicated. On this are superimposed all of the non-synonymous substitutions predicted to have occurred in the tree shown in panel A since the divergence of humans, bonobos, and chimpanzees from their last common ancestor (asterisk in A). Vertical lines indicate substitutions specific to humans, lines with white circles are substitutions specific to bonobos, and lines with grey circles are substitutions specific to chimpanzees. Lines with black circles indicate substitutions common to both bonobos and chimpanzees.

dissimilar amino acids (first four rows in Figure 3A). Next, these positively selected codon positions were examined for variability within the human population. The Breast Cancer Information Core (BIC, <http://research.nhgri.nih.gov/bic/>) is a repository of human *BRCA1* polymorphisms. Using this database, we identified single nucleotide polymorphisms (SNPs) at amino acid sites 170, 888, 890, 1203, and 1443 (Figure 3A). At four out of these five sites (position 888, 890, 1203, and 1443), we find that some human *BRCA1* alleles encode a unique amino acid not observed in any of our primate sequences. In addition, SNPs known to cause human disease occur in six out of 10 sites. In all cases, these disease-linked SNPs are not amino acid-altering mutations, but rather more radical frame-shifting or nonsense mutations (Figure 3A). In particular, nonsense mutations occurring in codon 1443 are among the most common mutations documented in the BIC. In Figure 3B, all 10 sites of positive selection were mapped onto a domain diagram of *BRCA1* (bottom) along with the most common human non-synonymous SNPs found in the BIC (top). As described previously for mutations accumulated in the human/chimpanzee/bonobo clade, all but one of the positively selected residues (1370S in the coiled-coil

domain) lie outside of any known structural motifs. In summary, the 10 codon positions identified in this analysis are highly variable between primate species and within the human population, and are involved in the etiology of cancers associated with this gene. Disease-associated SNPs at these sites tend to be radical, protein-truncating mutations. However, a presumably distinct phenomenon appears to be driving selection in favor of non-synonymous point mutations at these positions.

BRCA1 variation in other primate populations

So far, we have documented sequence differences between the *BRCA1* proteins of different primate species. We have shown that non-synonymous substitutions are accumulating in *BRCA1* faster than expected under constrained, or even neutral, evolution. We next wished to explore whether positive selection is still acting on *BRCA1* in modern populations. There is already evidence that this is true in the human population, because several *BRCA1* SNPs have been found to depart from Hardy-Weinberg equilibrium in European populations [14,15] and in Australia [6]. We wished to determine if the same might be true in bonobo and chimpanzee

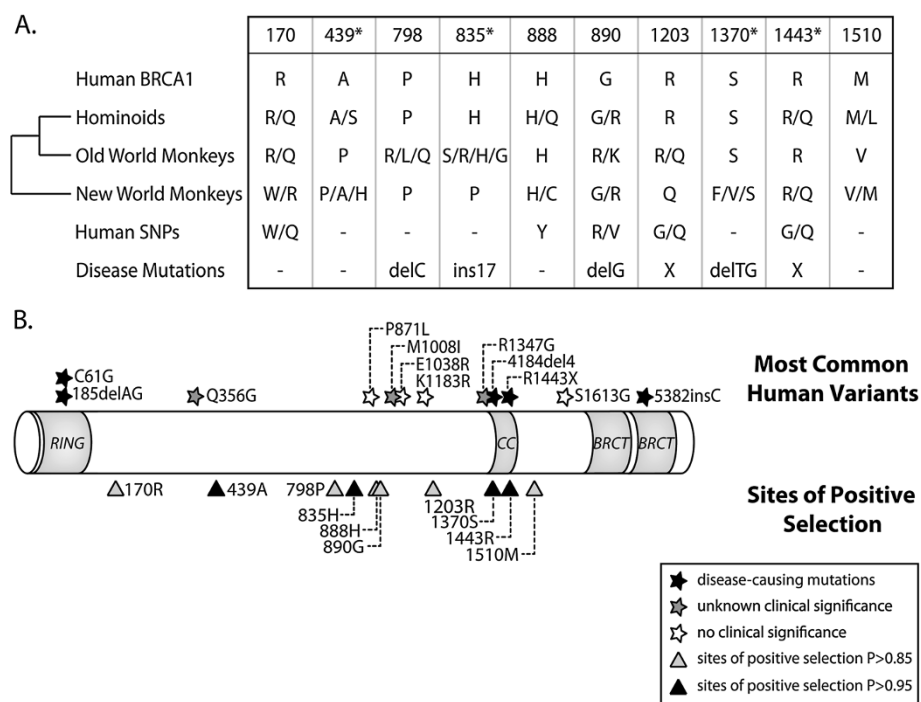


Figure 3 Specific codons in *BRCA1* have experienced positive selection during primate speciation. A. Shown are the ten codons that have evolved under positive selection ($dN/dS > 1$) in primates with a $P > 0.85$. Codons with a $P > 0.95$ are indicated with asterisks. The amino acids encoded at these positions in human *BRCA1* are shown, along with those found in hominoids, old world monkeys, and new world monkeys. In addition, human SNPs and disease mutations also found at these sites are listed. X refers to a single nucleotide mutation that results in a termination codon. **B.** A domain diagram of *BRCA1* is shown with the RING domain, coiled-coil domain (CC), and BRCT domains. The triangles at the bottom represent sites of positive selection (grey - $P > 0.85$, black - $P > 0.95$). The 12 most common human variants recorded in the BIC are shown at the top of the diagram as stars. The black stars indicate disease-causing mutations, white stars represent variants with no known clinical significance, and grey stars are those with unknown significance.

populations. We amplified and sequenced the largest *BRCA1* exon, exon 11 which is ~3.4 kilobases and comprises ~61% of the *BRCA1* coding region, from the genomic DNA of seven bonobo and 44 chimpanzee individuals (Table 2). In bonobos, we found nine polymorphic sites, eight of which were single nucleotide polymorphisms (SNPs), with three of these being non-synonymous. Eight of the SNPs were in Hardy Weinberg equilibrium.

Interestingly, one bonobo individual was also homozygous for a seven amino acid deletion ($\Delta 1058-1064$) (Table 2). Hardy-Weinberg equilibrium was rejected for this polymorphism, although the support was weak and did not survive correction for multiple testing (Table 2). The chimpanzee sequence set revealed nine SNPs, seven of which were non-synonymous. Interestingly, in this larger sample set ($n = 44$), three of the non-synonymous SNPs

Table 2 SNP Analysis of *BRCA1* in Bonobo, Chimpanzee, and Rhesus Macaque Individuals

Species	SNPs ^a	Genotype			p-value ^b	Human ^c	Human Polymorphisms	
		AA	AB	BB			BIC ^d	1000 genomes ^e
Bonobo n = 7	I493L	6	1	0	0.841	I		
	T582M	6	1	0	0.841	T		
	L833L	4	3	0	0.471	L	dupAAGTATCCAT*	
	V1047V	5	1	1	0.128	V		
	G1048G	5	1	1	0.128	G	G1048D, G1048V, G1048G	G1048D, G1048V, G1048G
	T1051I	5	1	1	0.128	T		
	$\Delta 1058-1064^{\text{HWD}}$	6	0	1	0.008			
	V1061V	6	1	0	0.841	I	delA*	delA*
	G1062G	6	1	0	0.841	G		
Chimpanzee n = 44	E309K ^{HWD}	19	14	11	0.023	K	K309T	K309Q, K309T
	E427K	34	9	1	0.663	E		
	S578S	40	4	0	0.752	S	S578Y	S578Y
	G590S ^{HWD}	20	12	12	0.004	S	S590G	S590G
	K731E	19	16	9	0.122	K	delAGAAG*	delAGAAG*
	I925T	34	9	1	0.663	I	I925L	I925V, I925L, insT*
	S1042S	41	3	0	0.823	S		
	G1077R ^{HWD}	42	1	1	1.4E-5	G		G1077W, G1077G
	G1100E	20	16	8	0.155	G		
Rhesus n = 44	A225A	42	2	0	0.888	A		
	N375S	43	1	0	0.920	N	delA*, N376S	delA*, N376S
	R466R	42	2	0	0.888	K	K467X*	K467X*
	T487S	43	1	0	0.920	T	insA*	insA*
	N684N	29	14	1	0.647	N		
	V739M	38	6	0	0.624	V	V740L	V740L, insA*
	D773G	29	15	0	0.173	G		
	D852D	40	4	0	0.752	D	insA*	insA*
	N923H	40	4	0	0.752	N		
	K936K	40	4	0	0.752	K		
	A1167E	40	4	0	0.752	A		
Q1203R	29	14	1	0.647	R	R1203Q, R1203G, R1203X*	R1203Q, R1203G, R1203X*	

^aNumbering refers to the amino acid position in the respective primates. In the case of rhesus macaques, amino acids 375 to 936 correspond to amino acids 376 to 937 in humans.

^bp-values were calculated using a chi-squared test with a $df = 1$. A p-value cutoff (after Bonferroni correction) < 0.0056 , 0.0056 , and 0.0042 for bonobos, chimpanzees, and rhesus macaque, respectively, was considered statistically significant. Tests that survived this correction have the p-value listed in italics.

^cAmino acid found in the human *BRCA1* protein at each of the positions listed.

^dHuman variants found at the positions indicated in the Breast Cancer Information Core.

^eHuman variants found at the positions indicated in the 1000 Genomes database.

*Known human disease-causing variant.

^{HWD} SNPs found to be in Hardy-Weinberg Disequilibrium.

were found to be in Hardy Weinberg disequilibrium, suggesting that selection is acting either for (E309K and G590S) or against (G1077R) these mutations. The support for one of these (E309K) was weak and did not survive correction for multiple testing (Table 2). It is particularly intriguing to see that humans also share with chimpanzees this same S/G SNP at position 590. In both the bonobo and chimpanzee populations, all synonymous SNPs were in Hardy-Weinberg equilibrium.

We also sequenced exon 11 from 44 rhesus macaque individuals. Rhesus macaques are not part of the human/chimpanzee/bonobo clade and are instead distantly-related members of the Old World monkey clade (Figure 1A). In these macaques, we found 12 SNPs in *BRCA1*, with seven being non-synonymous (Table 2). This includes a SNP found at position 1203, a site of positive selection in the inter-species dataset. This codon is also the site of a known disease-linked mutation in humans; however, the cancer-linked SNP at this position introduces a stop codon. Nonetheless, all of these are in Hardy-Weinberg equilibrium.

Caution must be used when interpreting signatures of selection acting on polymorphisms in primate populations. When sampling primates, it is not possible to get completely random and non-related population sets. Deviations from Hardy-Weinberg equilibrium may occur due to factors other than selection. Reasons for falsely rejecting Hardy Weinberg equilibrium include 1) non-random mating, 2) small population sizes which magnify the effects of genetic drift, 3) introduction of new alleles, 4) population subdivision or admixture, 5) biases in sequencing errors, and 6) linkage disequilibrium with another locus under selection. Because the chimpanzee population consists of individuals from two different subspecies, admixture could plausibly lead to rejection of Hardy Weinberg equilibrium.

We also performed the McDonald-Kreitman and Tajima's D tests on our datasets (data not shown). The tests were not significant and therefore do not support selection acting on any of these polymorphisms. False conclusions in this test can again result from a population with hidden structure. In summary, while the analyses using the simian primate dataset consisting of 23 species suggest that recurrent positive selection has been acting on *BRCA1* over the course of several million years, the Hardy-Weinberg equilibrium tests performed here and by others indicate that selection is acting on modern day humans, and possibly also chimpanzees.

***BRCA2* is also evolving under positive selection in primates**

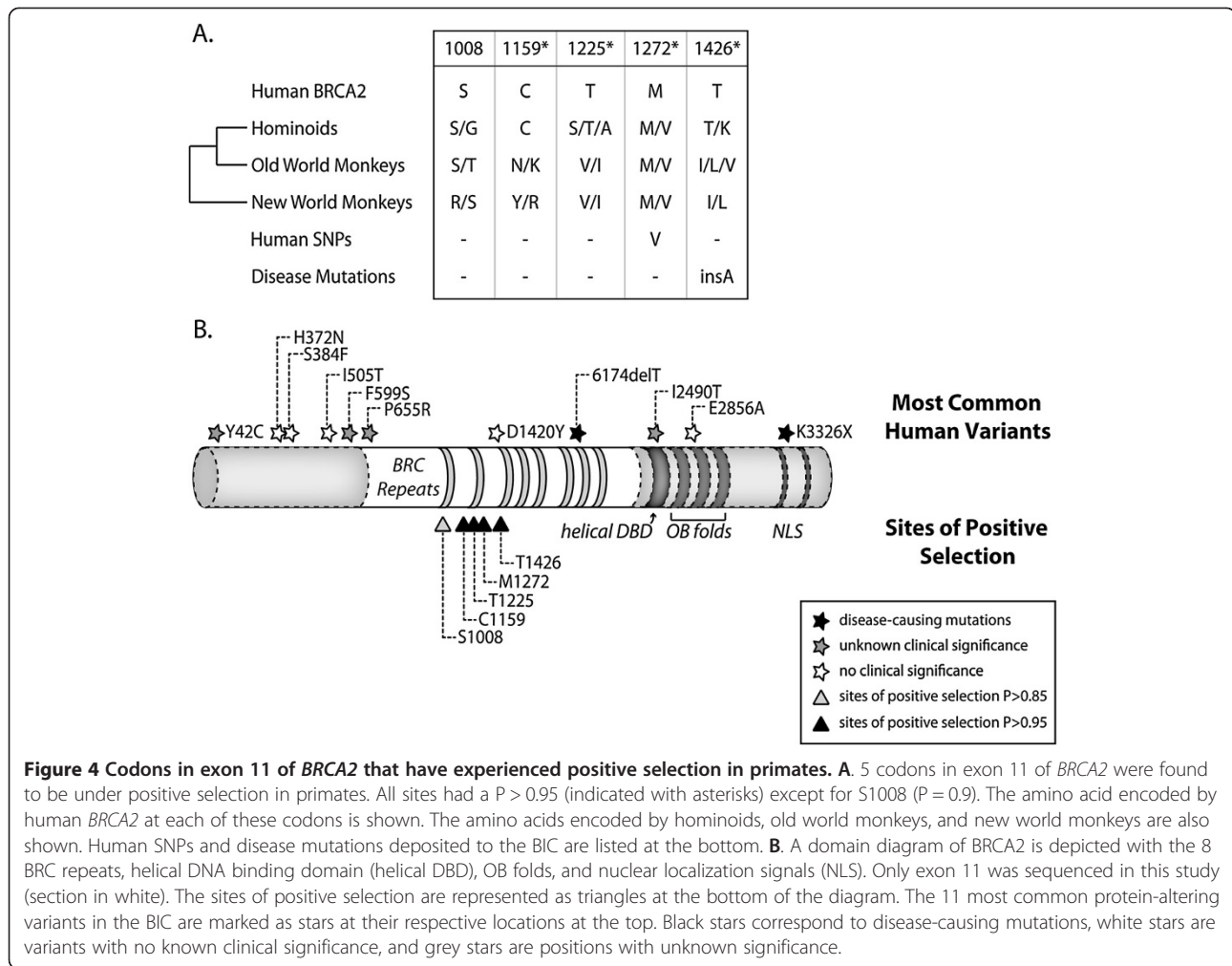
Because of the rapidly evolving nature of *BRCA1*, we also completed an evolutionary analysis of *BRCA2*, another strong determinant for hereditary breast and ovarian cancer. Although *BRCA2* has been shown to be under positive selection, only a small number of primate

species was included in this study [10]. We sequenced the ~5 kilobase exon 11 from 18 primate species. Exon 11 is the largest of 27 exons and encodes about 50% of the entire *BRCA2* protein. The sequences, along with six additional sequences from available genome projects, were assembled into a multiple alignment (Additional file 2). We fit the alignment to positive selection and null models as described above. The positive selection models were again a significantly better fit to the sequence set than the null models, with a p value ≤ 0.0003 (Table 1). In summary, *BRCA2* is under positive selection in primates as well, although this signature appears not to be concentrated on the human/chimpanzee/bonobo clade (Additional file 3).

In contrast to *BRCA1*, *BRCA2* is a 390 kDa nuclear protein that is exclusively involved in the homologous recombination pathway for repairing double-strand breaks. The eight BRC motifs and the extreme C terminus mediate interactions with and recruitment of Rad51, a protein that catalyzes strand invasion during homologous recombination [16-18]. All eight BRC repeats are encoded within exon 11. The M8 model estimates that five codons are evolving under positive selection with posterior probability > 0.85 (Figure 4A). Two of these positively selected sites were found to have a human polymorphism documented in the BIC (Figure 4A). When all five sites of positive selection are mapped onto a domain diagram of *BRCA2* (Figure 4B), they cluster within the first three BRC domains (1008, 1225, and 1426) and the intervening regions (1159 and 1272). To examine this further, we aligned the amino acid sequence of all eight BRC repeats of human *BRCA2* and highlighted sites 1008, 1225, and 1426 (Figure 5A). Surprisingly, all three sites of positive selection lie adjacent to a hydrophobic motif (FxxA) known to mediate interactions with Rad51 (Figure 5A red box). Since the co-crystal structure of the *BRCA2* BRC4 in complex with Rad51 is available, we mapped these three sites to their analogous positions in BRC4 and found that they are in close proximity to the Rad51 binding interface (Figure 5B, PDB: 1N0W) [19]. The clustering of these residues near this interface might provide a clue to the driver of natural selection at these sites.

Discussion

Nearly all known cases of recurrent positive selection in primate genomes involve genes in one of three categories: 1) immunity, 2) environmental perception (such as odorant and taste receptors), or 3) sexual selection and mate-choice [21,22]. This is due to the fact that ever-changing external stimuli (i.e. pathogens, environmental odors/tastes, etc.) drive the selection of new allelic variants. For example, immunity factors that are constantly challenged by pathogens exhibit some of the most striking signatures of positive selection seen in primate genomes [23-28].



Here, immunity genes will experience positive selection for protein-altering mutations that improve recognition of a relevant pathogen. Conversely, the pathogen will counter-evolve to escape detection, again placing selective pressure on the host population for new mutations that improve the immunity protein. This cycle can repeat itself indefinitely, resulting in an ever-escalating host-virus arms race. Therefore, it is surprising to see that *BRCA1* and *BRCA2*, genes that do not classically fit into any of the three categories listed above, are evolving in a similar manner to these highly adaptive immunity genes. In addition to the two described here, other DNA repair genes have also been shown to evolve under positive selection [29,30], but the driver behind this unusual finding remains to be identified.

An intense battle exists between host DNA repair machinery and viruses, and we propose that this could contribute to the evolutionary signatures documented here. Many viruses are known to interact with the DNA repair machinery and cell cycle regulators [31,32]. One fundamental issue is that the free ends of viral genomes are exposed, in contrast to the host's DNA, which is capped

by telomeres. Despite this, many viruses need to access the nucleus where the host's DNA repair machinery recognizes these un-capped viral genome ends as "damaged" cellular DNA, activating the DNA damage response. In order for productive infection to proceed, viruses must actively thwart these host repair pathways. For example, DNA repair proteins interfere with the adenovirus life-cycle by concatenating the ends of newly synthesized viral DNA, inhibiting efficient packaging into viral progeny [33]. In turn, adenovirus has evolved a way around this blockade by encoding proteins that mislocalize or degrade the specific host factors involved. Depending on the virus involved, host DNA repair factors can also be hijacked to facilitate viral replication. For instance, herpes simplex virus-1 simultaneously activates DNA repair constituents that aid in viral genome replication [34,35] and counteracts those that do not [36,37]. Human immunodeficiency virus 1 is also known to activate the DNA damage response and manipulate cell cycle checkpoints through the actions of its accessory protein Vpr [38,39]. Additionally, several studies have shown that specific DNA repair

A.

BRC-1	NHSFGG-SFRTASNKEIKLSEHNIKSKMFFKDIEE
BRC-2	-NEVGFRGFYSAHGIKLNVSTEALQKAVKLFSDIEN
BRC-3	-FETSDFFFQTASGKNISVAKESFNKIVNFFDOKPE
BRC-4	-KEPTLLGFHTASGKKVKIAKESLDKVKNLDFDEKEQ
BRC-5	-IENSALAFYTSCSRKTSVVSQTSLEAKKWLREGIF
BRC-6	-FEVGPPAFRIASGKIVCVSHETIKKVKDIFTDSFS
BRC-7	-SANTCGIFSTASGKSVQVSDASLQNAQVFSEIED
BRC-8	-NSSAFSGFSTASGKQVSILESSLHKVKGVLEEFDL

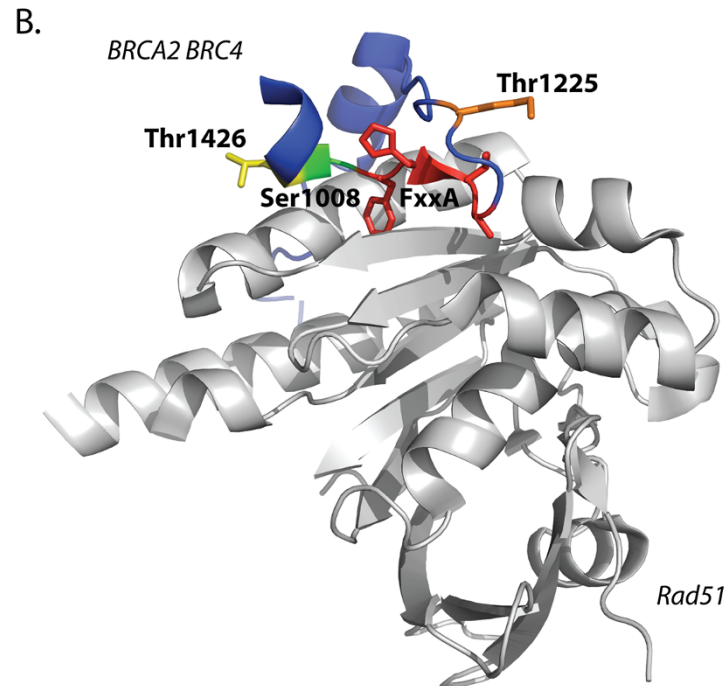


Figure 5 The sites of positive selection lying within the BRC repeats of BRCA2 are located adjacent to the Rad51 binding region. **A.** The 8 BRC repeats of the human BRCA2 protein were aligned using ClustalX. The red and peach colored boxes are the motifs within the BRC repeats thought to facilitate binding with Rad51 [20]. Residues 1008, 1225, and 1426 are colored in green, orange, and yellow, respectively. All three sites lie just adjacent to the FxxA motif which interacts with two hydrophobic pockets in the Rad51 oligomer. **B.** The co-crystal structure of BRC4 (blue) in complex with Rad51 (grey) is shown (PDB ID 1N0W [19]). The FxxA motif is depicted in red. Residues 1008, 1225, and 1426 are shown in green, orange, and yellow, respectively.

proteins play critical roles in retroviral genome integration [40-43] while others seem to decrease the efficiency of infection [44-46].

One can imagine that these and other viruses that access the nucleus during replication could feasibly interact with BRCA1 or BRCA2, driving the selection of variants that ultimately lead to decreased susceptibility to infection. However, it is possible that variant alleles selected for this purpose would have detrimental consequences to protein function in the context of host DNA repair. Most of the deleterious *BRCA1* and *BRCA2* variants characterized thus far introduce stop codons or frame-shifts that result in premature truncation of the protein, the consequences of which manifest as cancer at relatively early ages. The effects of non-synonymous

point mutations, such as those documented here, might be expected to be much more subtle. The effects of subtle mutations are more difficult to assess because the resulting genomic instability may only be realized later in life and can be confounded by other genetic or environmental influences. We therefore propose a hypothesis where viruses are driving the intriguingly rapid rate of evolution seen in *BRCA1* and *BRCA2*, potentially giving rise to antagonistic pleiotropy. This would be analogous to the malaria and sickle cell anemia trade-off that is well documented [47].

Conclusions

The BRCA1 and BRCA2 proteins play key roles in the repair of damage to chromosomal DNA. We have

expanded the analysis of the evolution of these genes, showing that both have been subject to recurrent positive selection during simian primate speciation. Although the force or forces driving the diversifying selection of these genes is unknown, the result is that the sequence of these proteins has been altered in humans and our closest living relatives. It remains to be seen whether this is an instance of antagonistic pleiotropy, where positive selection driven by one force causes functional consequences in another context, potentially the formation of cancers [48].

Methods

Non-human primate samples

Of the 44 chimpanzee samples evaluated in this study, 34 were obtained from the Chimpanzee Biomedical Research Resource (NIH8U42OD011197-13), which is supported through a cooperative agreement with the National Institutes of Health (NIH). This NIH-supported colony is housed at the MD Anderson Cancer Center's Michale E. Keeling Center for Comparative Medicine and Research (KCCMR) in Bastrop, TX. The origins of the chimpanzees comprising the KCCMR colony are highly diverse with only a few closely related (siblings/offspring) animals in the colony (Additional file 4). Blood from 34 chimpanzees was collected directly into PAXgene Blood RNA Tubes (PreAnalytix) at the same time other blood samples were obtained as part of the prescheduled annual veterinary exam for each animal. Another 10 chimpanzee genomic DNA samples were purchased from Coriell (Additional file 5).

All 44 rhesus macaque samples evaluated in this study were obtained from animals housed at the KCCMR in collaboration with researchers at this institution. The colony at the KCCMR is a closed breeding colony comprised of approximately 980 rhesus macaques of Indian-origin that originated from a colony of 286 founder animals in 1988 (degree of relatedness can be found in Additional file 6). Blood from these animals was collected directly into PAXgene Blood RNA Tubes (PreAnalytix) at the same time other blood samples were obtained as part of the prescheduled annual veterinary exam for each animal.

Bonobo genomic DNA samples were obtained from the integrated primate biomaterials and information resource (IPBIR) of the Coriell Institute or extracted from blood samples obtained from the Columbus zoo and the Language Research Center, Georgia State University. All seven individuals are unrelated (Additional file 7).

The remaining non-human primate samples were acquired as cell lines purchased from the Coriell Institute under a U.S. Fish and Wildlife Service permit (sources and unique identifiers are listed in Additional file 8). This study was approved by the University of Texas at Austin Institutional Review Board.

Primate *BRCA1* and *BRCA2* sequencing

Human *BRCA1* and *BRCA2* coding sequences were obtained from GenBank (accession number NM 007294 and NM 000059, respectively). *BRCA1* and *BRCA2* sequences from chimpanzee, gorilla, orangutan, rhesus macaque, and marmoset were obtained using the BLAT alignment tool on the UCSC genome database (<http://genome.ucsc.edu/>). For the remaining 18 primate sequences, primary or immortalized cell lines were grown in standard media supplemented with 15% fetal bovine serum at 37°C and 5% CO₂. Cells were collected and RNA was extracted using the AllPrep DNA/RNA kit (QIAGEN). cDNA libraries were generated using SuperScript III First-Strand Synthesis Kit (Invitrogen) using oligo dT or random hexamer primers. PCR products were generated using PCR SuperMix High Fidelity (Invitrogen) and directly sequenced or cloned into pCR4 for sequencing. Primers used for PCR and sequencing can be found in Additional files 9, 10, 11 and 12. These sequences have been deposited in GenBank (accession numbers KM017616-KM017652).

Blood from rhesus macaque and chimpanzee individuals was collected in PAXgene Blood RNA Tubes (PreAnalytix). RNA was extracted using the PAXgene Blood miRNA Kit (QIAGEN) and genomic DNA was obtained using the AllPrep DNA/RNA kit (QIAGEN). *BRCA1* Exon 11 was amplified from extracted genomic DNA (chimpanzee, bonobo, and rhesus macaque) using PCR SuperMix High Fidelity (Invitrogen) and sequenced. Details on PCR and sequencing primers can be found in Additional file 9 and 10.

PAML analysis

A multiple sequence alignment was generated for *BRCA1* and *BRCA2* using ClustalX2.1 [49]. The alignments are straight-forward with only a few small indels (Additional files 1 and 2). Gene sequences at each ancestral node were reconstructed using the codeml program in PAML 4.3 [50]. dN/dS values along each branch of the phylogenetic tree were calculated using the free-ratio model. Substitution counts given along specified branches are the estimates made in the free ratio model, but were also calculated by directly comparing the predicted ancestral and the known extant sequences and counting differences manually. Both methods yielded the same values. The one-ratio and two-ratio models were performed as described previously [51]. To detect selection, multiple alignments were fit to the NSsites models M1a (null model, codon values of dN/dS are fit into two site classes, one with value between 0 and 1, and one fixed at dN/dS = 1), M2a (positive selection model, similar to M1a but with an extra codon class of dN/dS > 1), M7 (null model, codon values of dN/dS fit to a beta distribution bounded between 0 and 1), M8a (null model, similar to M7 except with an extra fixed codon class at dN/dS = 1), and M8 (positive selection model, similar to M7 but with

an extra class of dN/dS > 1). Model fitting was performed with multiple seed values for dN/dS (ω) and assuming either the f61 or f3x4 model of codon frequencies [52]. Likelihood ratio tests were performed to assess whether permitting some codons to evolve under positive selection gives a significantly better fit to the data than models where positive selection is not allowed [53,54]. These different model comparisons represent different trade-offs between power and accuracy [55]. In all cases the positive selection model was a significantly better fit ($p < 0.05$), and individual codons assigned to the dN/dS > 1 class with high posterior probabilities ($P > 0.85$ by Bayes Empirical Bayes [56]) were analyzed. The crystal structure was obtained from the RCSB Protein Data Bank (<http://www.pdb.org>) and residues under positive selection were mapped using MacPyMol (<http://www.pymol.org>).

Hardy-weinberg equilibrium test

Single nucleotide polymorphisms (SNPs) were annotated for each bonobo, chimpanzee, and rhesus macaque individual. Allele frequencies were calculated for each SNP and tested for departure from Hardy-Weinberg equilibrium (<http://www.oege.org>) [57]. Chi squared values were calculated using 1 degree of freedom. A p-value (after Bonferroni correction) < 0.0056, 0.0056, and 0.0042 for bonobos, chimpanzees, and rhesus macaque, respectively, was considered statistically significant.

Ethics

No new human data was generated or analyzed in this study.

Additional files

Additional file 1: Alignment of BRCA1 sequences. description – alignment of *BRCA1* sequences used in the PAML analyses.

Additional file 2: Alignment of BRCA2 sequences. description – alignment of *BRCA2* sequences used in the PAML analyses.

Additional file 3: Evolution of BRCA2 over the course of primate speciation. dN/dS values for each branch of the primate phylogeny were calculated using the free-ratio model in PAML [13]. Branches exhibiting dN/dS values > 1 are shown in bold italics. Dashes (–) represent branches where zero synonymous substitutions are predicted to have occurred. On these branches, dS = 0 and dN/dS can therefore not be calculated. In these instances, the numbers of non-synonymous (N) and synonymous (S) substitutions predicted to have occurred along each branch are indicated in parentheses (N:S). Of these, branches that experienced 4 or more non-synonymous changes are italicized.

Additional file 4: Degree of relatedness in Pan troglodyte (chimpanzee) individuals. description – sex, age, and relatedness of chimpanzee individuals used in this study.

Additional file 5: Sources and unique identifiers of Pan troglodyte genomic DNA used to generate BRCA1 exon 11 sequences. description – sources and unique identifiers of chimpanzee genomic DNA used in this study.

Additional file 6: Degree of relatedness in Macaca mulatta (rhesus macaque) individuals. description – relatedness of rhesus macaque individuals used in this study.

Additional file 7: Pan paniscus (bonobo) individuals information. description – sex and sources of bonobo samples used in this study.

Additional file 8: Sources and unique identifiers of cell lines used to generate primate cDNA libraries and sequences. description – sources and unique identifiers of cell lines used in this study.

Additional file 9: Primers used for BRCA1 amplification and sequencing. description – Primers used to amplify and sequence *BRCA1*.

Additional file 10: Sequences of primers used for BRCA1 sequencing. description – sequences of primers used to amplify and sequence *BRCA1*.

Additional file 11: Primers used for BRCA2 amplification and sequencing. description – primers used to amplify and sequence *BRCA2*.

Additional file 12: Sequences of primers used for BRCA2 sequencing. description – sequences of primers used to amplify and sequence *BRCA2*.

Abbreviations

BIC: Breast Cancer Information Core; SNP: Single Nucleotide Polymorphism.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

DIL sequenced genes, carried out molecular genetic studies, analyzed data, and wrote the manuscript; RMM sequenced genes and analyzed data; UQL sequenced genes; ACS provided primate materials and performed statistical tests; GKW oversaw the collection of primate materials and edited the manuscript; AMD sequenced genes and edited the manuscript; SLS conceived the study, participated in its design and coordination, and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Tanya Paull for critical discussions. This work was supported by the National Institutes of Health [R01-GM-093086 to S.L.S., 8U42OD011197-13 to the Keeling Center]; and the National Science Foundation [BCS-07115972 to A.C.S]. S.L.S. holds a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund, and is an Alfred P. Sloan Research Fellow in Computational and Evolutionary Molecular Biology. D.I.L. is an NRSA fellow of the National Cancer Institute [F30 CA171715-01].

Author details

¹Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712, USA. ²School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85281, USA. ³Department of Veterinary Sciences, Michale E. Keeling Center for Comparative Medicine and Research, The University of Texas MD Anderson Cancer Center, Bastrop, TX 78602, USA.

Received: 8 April 2014 Accepted: 27 June 2014

Published: 11 July 2014

References

1. Mullen P, Miller WR, Mackay J, Fitzpatrick DR, Langdon SP, Warner JP: **BRCA1 5382insC mutation in sporadic and familial breast and ovarian carcinoma in Scotland.** *Br J Cancer* 1997, **75**:1377–1380.
2. O'Donovan PJ, Livingston DM: **BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair.** *Carcinogenesis* 2010, **31**:961–967.
3. Hemel D, Domchek SM: **Breast Cancer Predisposition Syndromes.** *Hematol Oncol Clin North Am* 2010, **24**:799–814.
4. Ludwig T, Chapman DL, Papaioannou VE, Efstratiadis A: **Targeted mutations of breast cancer susceptibility gene homologs in mice: lethal phenotypes of Brca1, Brca2, Brca1/Brca2, Brca1/p53, and Brca2/p53 nullizygous embryos.** *Genes Dev* 1997, **11**:1226–1241.
5. Hurst LD, Pál C: **Evidence for purifying selection acting on silent sites in BRCA1.** *Trends Genet* 2001, **17**:62–65.
6. Huttley GA, Eastale S, Southey MC, Tesoriero A, Giles GG, McCredie MR, Hopper JL, Venter DJ: **Adaptive evolution of the tumour suppressor BRCA1 in humans and chimpanzees.** *Australian Breast Cancer Family Study.* *Nat Genet* 2000, **25**:410–413.

7. Yang Z, Nielsen R: Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 2002, **19**:908–917.
8. Fleming MA, Potter JD, Ramirez CJ, Ostrander GK, Ostrander EA: Understanding missense mutations in the BRCA1 gene: an evolutionary approach. *Proc Natl Acad Sci U S A* 2003, **100**:1151–1156.
9. Burk-Herrick A, Scally M, Amrine-Madsen H, Stanhope MJ, Springer MS: Natural selection and mammalian BRCA1 sequences: elucidating functionally important sites relevant to breast cancer susceptibility in humans. *Mamm Genome* 2006, **17**:257–270.
10. O'Connell MJ: Selection and the Cell Cycle: Positive Darwinian Selection in a Well-Known DNA Damage Response Pathway. *J Mol Evol* 2010, **71**:444–457.
11. Pavlicek A, Noskov V, Kouprina N, Barrett JC, Jurka J, Larionov V: Evolution of the tumor suppressor BRCA1 locus in primates: implications for cancer predisposition. *Hum Mol Genet* 2004, **13**:2737–2751.
12. Meyerson NR, Sawyer SL: Two-stepping through time: mammals and viruses. *Trends Microbiol* 2011, **19**:286–294.
13. Yang Z: PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007, **24**:1586–1591.
14. Durocher F, Shattuck-Eidens D, McClure M, Labrie F, Skolnick MH, Goldgar DE, Simard J: Comparison of BRCA1 polymorphisms, rare sequence variants and/or missense mutations in unaffected and breast/ovarian cancer populations. *Hum Mol Genet* 1996, **5**:835–842.
15. Dunning AM, Chiano M, Smith NR, Dearden J, Gore M, Oakes S, Wilson C, Stratton M, Peto J, Easton D, Clayton D, Ponder BA: Common BRCA1 variants and susceptibility to breast and ovarian cancer in the general population. *Hum Mol Genet* 1997, **6**:285–289.
16. Mizuta R, LaSalle JM, Cheng HL, Shinohara A, Ogawa H, Copeland N, Jenkins NA, Lalonde M, Alt FW: RAB22 and RAB163/mouse BRCA2: proteins that specifically interact with the RAD51 protein. *Proc Natl Acad Sci U S A* 1997, **94**:6927–6932.
17. Wong AKC, Pero R, Ormonde PA, Tavtigian SV, Bartel PL: RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene brca2. *J Biol Chem* 1997, **272**:31941–31944.
18. Holloman WK: Unraveling the mechanism of BRCA2 in homologous recombination. *Nat Struct Mol Biol* 2011, **18**:748–754.
19. Pellegrini L, Yu DS, Lo T, Anand S, Lee M, Blundell TL, Venkitaraman AR: Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature* 2002, **420**:287–293.
20. Rajendra E, Venkitaraman AR: Two modules in the BRC repeats of BRCA2 mediate structural and functional interactions with the RAD51 recombinase. *Nucleic Acids Res* 2009, **38**:82–96.
21. Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferreira S, Wang G, Zheng X, White TJ, Sninsky JJ, Adams MD, Cargill M: Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. *Science* 2003, **302**:1960–1963.
22. Vallender EJ, Lahn BT: Positive selection on the human genome. *Hum Mol Genet* 2004, **13**:R245–R254. Spec No 2.
23. Sawyer SL, Emerman M, Malik HS: Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol* 2004, **2**:E275.
24. Sawyer SL, Wu LI, Emerman M, Malik HS: Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc Natl Acad Sci U S A* 2005, **102**:2832–2837.
25. Elde NC, Child SJ, Geballe AP, Malik HS: Protein kinase R reveals an evolutionary model for defeating viral mimicry. *Nature* 2009, **457**:485–489.
26. Lim ES, Malik HS, Emerman M: Ancient Adaptive Evolution of Tetherin Shaped the Functions of Vpu and Nef in Human Immunodeficiency Virus and Primate Lentiviruses. *J Virol* 2010, **84**:7124–7134.
27. Laguette N, Rahm N, Sobhian B, Chable-Bessia C, Münch J, Snoeck J, Sauter D, Switzer WM, Heneine W, Kirchhoff F, Delsuc F, Telenti A, Benkirane M: Evolutionary and Functional Analyses of the Interaction between the Myeloid Restriction Factor SAMHD1 and the Lentiviral Vpx Protein. *Cell Host and Microbe* 2012, **11**:205–217.
28. Lim ES, Fregoso OL, McCoy CO, Matsen FA, Malik HS, Emerman M: The Ability of Primate Lentiviruses to Degrade the Monocyte Restriction Factor SAMHD1 Preceded the Birth of the Viral Accessory Protein Vpx. *Cell Host and Microbe* 2012, **11**:194–204.
29. Demogines A, East AM, Lee J-H, Grossman SR, Sabeti PC, Paull TT, Sawyer SL: Ancient and Recent Adaptive Evolution of Primate Non-Homologous End Joining Genes. *PLoS Genet* 2010, **6**:e1001169.
30. Sawyer SL, Malik HS: Positive selection of yeast nonhomologous end-joining genes and a retrotransposon conflict hypothesis. *Proc Natl Acad Sci U S A* 2006, **103**:17614–17619.
31. Lilley CE, Schwartz RA, Weitzman MD: Using or abusing: viruses and the cellular DNA damage response. *Trends Microbiol* 2007, **15**:119–126.
32. Chaurushiya MS, Weitzman MD: Viral manipulation of DNA repair and cell cycle checkpoints. *DNA Repair (Amst)* 2009, **8**:1166–1176.
33. Stracker TH, Carson CT, Weitzman MD: Adenovirus oncoproteins inactivate the Mre11-Rad50-NBS1 DNA repair complex. *Nature* 2002, **418**:348–352.
34. Lilley CE, Carson CT, Muotri AR, Gage FH, Weitzman MD: DNA repair proteins affect the lifecycle of herpes simplex virus 1. *Proc Natl Acad Sci U S A* 2005, **102**:5844–5849.
35. Mohni KN, Mastrocola AS, Bai P, Weller SK, Heinen CD: DNA mismatch repair proteins are required for efficient herpes simplex virus 1 replication. *J Virol* 2011, **85**:12241–12253.
36. Lees-Miller SP, Long MC, Kilvert MA, Lam V, Rice SA, Spencer CA: Attenuation of DNA-dependent protein kinase activity and its catalytic subunit by the herpes simplex virus type 1 transactivator ICP0. *J Virol* 1996, **70**:7471–7477.
37. Lilley CE, Chaurushiya MS, Boutell C, Everett RD, Weitzman MD: The intrinsic antiviral defense to incoming HSV-1 genomes includes specific DNA repair proteins and is counteracted by the viral protein ICP0. *PLoS Pathog* 2011, **7**:e1002084.
38. Zimmerman ES, Chen J, Andersen JL, Ardon O, Dehart JL, Blackett J, Choudhary SK, Camerini D, Nghiem P, Planelles V: Human immunodeficiency virus type 1 Vpr-mediated G2 arrest requires Rad17 and Hus1 and induces nuclear BRCA1 and gamma-H2AX focus formation. *Mol Cell Biol* 2004, **24**:9286–9294.
39. Nakai-Murakami C, Shimura M, Kinomoto M, Takizawa Y, Tokunaga K, Taguchi T, Hoshino S, Miyagawa K, Sata T, Kurumizaka H, Yuo A, Ishizaka Y: HIV-1 Vpr induces ATM-dependent cellular signal with enhanced homologous recombination. *Oncogene* 2006, **26**:477–486.
40. Daniel R, Katz RA, Skalka AM: A Role for DNA-PK in Retroviral DNA Integration. *Science* 1999, **284**:644–647.
41. Daniel R, Greger JG, Katz RA, Taganov KD, Wu X, Kappes JC, Skalka AM: Evidence that stable retroviral transduction and cell survival following DNA integration depend on components of the nonhomologous end joining repair pathway. *J Virol* 2004, **78**:8573–8581.
42. Smith JA, Wang F-X, Zhang H, Wu K-J, Williams KJ, Daniel R: Evidence that the Nijmegen breakage syndrome protein, an early sensor of double-strand DNA breaks (DSB), is involved in HIV-1 post-integration repair by recruiting the ataxia telangiectasia-mutated kinase in a process similar to, but distinct from, cellular DSB repair. *J Virol* 2008, **5**:11.
43. Zhong Q, Chen C-F, Chen P-L, Lee W-H: BRCA1 Facilitates Microhomology-mediated End Joining of DNA Double Strand Breaks. *J Biol Chem* 2002, **277**:28641–28647.
44. Lau A, Kanaar R, Jackson SP, O'Connor MJ: Suppression of retroviral infection by the RAD52 DNA repair protein. *EMBO J* 2004, **23**:3421–3429.
45. Lloyd AG, Tateishi S, Bieniasz PD, Muesing MA, Yamaizumi M, Mulder LCF: Effect of DNA Repair Protein Rad18 on Viral Infection. *PLoS Pathog* 2006, **2**:e40.
46. Cosnefroy O, Tocco A, Lesbats P, Thierry S, Calmels C, Wiktorowicz T, Reigadas S, Kwon Y, De Cian A, Desfarges S, Bonot P, San Filippo J, Litvak S, Le Cam E, Rethwilm A, Fleury H, Connell PP, Sung P, Delelis O, Andreola ML, Parissi V: Stimulation of the Human RAD51 Nucleofilament Restricts HIV-1 Integration In Vitro and in Infected Cells. *J Virol* 2011, **86**:513–526.
47. Carter AJ, Nguyen AQ: Antagonistic pleiotropy as a widespread mechanism for the maintenance of polymorphic disease alleles. *BMC Med Genet* 2011, **12**:160.
48. Crespi BJ, Summers K: Positive selection in the evolution of cancer. *Biol Rev* 2006, **81**:407.
49. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: Clustal W and Clustal X version 2.0. *Bioinformatics* 2007, **23**:2947–2948.
50. Yang Z: PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 1997, **13**:555–556.
51. Yang Z: Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 1998, **15**:568–573.
52. Goldman N, Yang Z: A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 1994, **11**:725–736.

53. Yang Z, Nielsen R, Goldman N, Pedersen AM: **Codon-substitution models for heterogeneous selection pressure at amino acid sites.** *Genetics* 2000, **155**:431–449.
54. Nielsen R, Yang Z: **Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene.** *Genetics* 1998, **148**:929–936.
55. Wong WSW, Yang Z, Goldman N, Nielsen R: **Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites.** *Genetics* 2004, **168**:1041–1051.
56. Yang Z, Wong WSW, Nielsen R: **Bayes empirical bayes inference of amino acid sites under positive selection.** *Mol Biol Evol* 2005, **22**:1107–1118.
57. Rodriguez S, Gaunt TR, Day INM: **Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies.** *Am J Epidemiol* 2009, **169**:505–514.

doi:10.1186/1471-2148-14-155

Cite this article as: Lou et al.: Rapid evolution of *BRCA1* and *BRCA2* in humans and other primates. *BMC Evolutionary Biology* 2014 **14**:155.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

