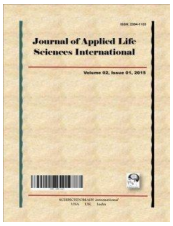

Journal of Applied Life Sciences International

18(2): 1-8, 2018; Article no.JALSI.43129
ISSN: 2394-1103

A Review on the Evolutionary Trajectories of mRNA BRCA1/2 Genes in Primates and the Implications of Cancer Susceptibility Variants within Immediate Human Populations

Tommy Rodriguez[1*]

[1]Pangaea Biosciences, Department of Research and Development, Miami, Florida, USA.

Authors' contribution

The sole author designed, analyzed and interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/JALSI/2018/43129
Editor(s):
(1) Malakeh. Z. Malak, Community Health Nursing, Al-Zaytoonah University of Jordan, Jordan.
Reviewers:
(1) Jose Carlos Rosa Pires De Souza, Mato Grosso Do Sul State University, Brazil.
(2) Bruno Ricardo Barreto Pires, Instituto Nacional de Câncer, Brazil.
(3) Jennifer Schroeder, Millikin University, USA.
Complete Peer review History: http://www.sciencedomain.org/review-history/25865

Original Research Article

Received 2nd June 2018
Accepted 6th August 2018
Published 13th August 2018

ABSTRACT

Researchers have shown that cancer susceptibility in BRCA1/2 occurs at disproportional rates and statistical degrees across a wide range of species, particularly non-human primates. This study incorporates two primary scopes for examining BRCA 1/2 cancer susceptibility among closely related lineages: (1) phylogenetic reconstruction of mRNA BRCA1/2 genes (both cancer susceptible and non-cancer susceptible) in variously distinct primate families (including *Homo sapiens*); and (2) pairwise comparative analysis of breast cancer 1 early onset BRCA1 mRNA partial cds within immediate human populations. The results generated by phylogenetic reconstruction together with pairwise comparative analysis revealed that cancer-causing alterations in BRCA1/2 appear to originate within localized gene pools at separate junctures throughout evolutionary time. This supports the explanation that BRCA1/2 genes may be undergoing rapid evolution, as revealed by unusually high proportion of dissimilarities between cancer susceptibility sequences among members of each group or species.

**Corresponding author: E-mail: trodriguez@pangaeabio.com, bio_nerd@outlook.com;*

Keywords: Breast cancer; ovarian cancer; phylogenetics; multiple sequence alignment; pairwise sequence alignment; BRCA genes.

1. INTRODUCTION

Heredity accounts for 5% to 10% of all cancer-related cases [1]. Some individuals are born with an increased risk of cancer because they inherit an altered gene important for cell growth or for repair of damaged DNA [2]. Over the past couples of decades, researchers have identified a number of gene alterations that predispose those individuals to various cancer types [3]. A well-documented example is the BRCA1/2 mutations that increase the risk of breast, ovarian and other cancers in those that inherit an altered gene [4,5,6].

Most cancers do not occur in patients with a hereditary risk, but inheritance could make the risk many times higher for developing breast or ovarian cancer during the course of a lifetime [2], [7]. Researchers have also shown that cancer susceptibility in BRCA1/2 occurs at disproportional rates across a wider range of species, particularly non-human primates [8]. In one such study, Puente et al. revealed that human cancer alterations present in non-human primates, contain intact similar open reading frames, and showed a high degree of conservation between closely related species [8]. However, it was also shown that the incidence of cancer in non-human primates was very low compared to humans [8]. While the prevailing view remains that [breast & ovarian] cancer is best and most often attributed to the damaging effects of epigenetic mechanisms that lead to abnormal cell growth [9], one might make a case for correlating perspectives in terms of evolutionary history. Positive selection has been previously cited for its involvement in rapidly evolving regions of the BRCA1/2 variants in primate populations [10,11,12].

This investigation incorporates two primary scopes for examining BRCA 1/2 cancer susceptibility among closely related lineages: (1) phylogenetic reconstruction of mRNA BRCA1/2 genes (both cancer susceptible and non-cancer susceptible) in variously distinct primate families (including *Homo sapiens*); and (2) pairwise comparative analysis of breast cancer 1 early onset BRCA1 mRNA partial cds within immediate human populations. By evaluating the results obtained from pairwise identity ratios, consensus comparisons, and phylogenetic reconstruction of 10 primate families, this paper seeks to address whether altered BRCA genes (or the specific mutation types) that disrupt error-free DNA repair mechanisms occur more frequently beyond interspecies taxonomy or originate spontaneously within a localized gene pool.

2. METHODS

2.1 Sequence Selection & Consensus Comparison

Two independent data types were used in this study. Each set of raw genomic data pertain to a particular experiment design. Sequence selection toward phylogenetic reconstruction compares non-cancer associated sequences against cancer susceptibility variants of the same mRNA type, among 10 primate families. As described above, tumor suppressing BRCA1/2 genomic datasets were selected due to their involvement in a number of different cancers in human populations and closely related others. Moreover, breast cancer 1 early onset BRCA1 mRNA variants were selected for consensus comparisons among *Homo sapiens* sequences.

The NCBI nucleotide databank was the repository where each mRNA BRCA1/2 sequence was acquired. [1] BLAST similarity searches helped identify homologous sequence candidates among closely related groups. Each set of raw sequences are referenced in 5 primary studies: (1) Emerging roles of BRCA1 alternative splicing [13]; (2) Evidence of a warfarin-sensitive cancer procoagulant in V2 carcinoma [14]; (3) Rapid evolution of BRCA1 and BRCA2 in humans and other primates [10]; (4) Growth retardation and tumour inhibition by BRCA1 [15]; and (5) Association of BRCA1/2 mutations with ovarian cancer prognosis: An updated meta-analysis [16]. From these collective findings, 3 distinct FASTA files containing a combination of 36 genomic sequences were compiled. Breast cancer 1 early onset BRCA1 mRNA variants were also obtained via NCBI nucleotide databank, [1] and were appropriated toward consensus/dissimilarity statistics.

It should be noted that the University of Utah, BRCA Mutation Database was an important reference repository in this study. [2] See additional notes for references, annotation numbers, and sequence descriptions [1 - 3].

2.2 Multiple Sequence Alignment, Pairwise Alignment & Phylogenetic Reconstruction

As it corresponds to the following procedures, Kalign for multiple sequence alignment (MSA) and pairwise alignment were utilized; whereas PHYLIP neighbor-joining method, a distance matrix algorithm, was employed toward phylogenetic reconstruction. [3] An accurate and fast MSA algorithm, Kalign is a dependable algorithmic selection for purposes of obtaining highly-robust alignments [17]. Kalign is an extension of Wu-Manber approximate pattern-matching algorithm, based on Levenshtein distances. This strategy enables Kalign to estimate sequence distances faster and more accurately than other popular iterative methods. Lassmann and Sonnhammer [17] show that Kalign is about 10 times faster than ClustalW and, depending on the alignment size, up to 50 times faster than other iterative methods; Kalign also delivers better overall resolution [17].

PHYLIP neighbor-joining can generate highly probable diagrams amid scenarios involving low degrees of variance, regardless of alignment size. Selected for these tree-building exercises, PHYLIP neighbor-joining is an accurate and statically consistent polynomial-time algorithm that does not assume that all lineages evolve at the same rate, and it constructs a tree by successive clustering of lineages, setting branch lengths as the lineages join [where a set of n taxa requires $n - 3$ iterations; each step is repeated by $(n - 1) \times (n - 1)$] [18,19]. This method utilizes a set of default parameters for distance matrix model F84. Additional bootstrapping compilers were not required for this operation, and transition ratios are generated automatically under default settings. [4] For reference purposes, the following formula demonstrates a standard neighbor-joining Q-matrix algorithm:

$$Q(i,j) = (n - 2)\, d(i,j) - \sum \{n,\, k = 1\}\, d(i,k) - \sum \{n,\, k = 1\}\, d(j,k) \quad (1)$$

Pair to node (distances):

$$(f,u) = \tfrac{1}{2}\, d(f,g) + \tfrac{1}{2}(n - 2)\, [\, \sum \{n,\, k = 1\}\, d(f,k) - \sum \{n,\, k = 1\}\, d(g,k)\,] \quad (2)$$

Taxa to node (distances):

$$d(u,k) = \tfrac{1}{2}\, [\, d(f,k) + d(g,k) - d(f,g)\,] \quad (3)$$

3. RESULTS AND DISCUSSION

3.1 Evolutionary Trees

Genetic mutation is the raw material needed for biological evolution to occur. The amount of time during which mutations accumulate to generate diversity results in higher or lesser degrees of genetic variation between different populations. It is widely held that most mutations play no significant role in the evolutionary process. Only those mutations that occur to the germline are significant in terms of evolutionary change. We assume that most mutations – something in range of 90 percent or higher – are either neutral or harmful to a host organism [20,21,22].

Germline mutations in the BRCA1/2 genes predispose affected individuals to breast and ovarian cancer syndromes [12]. The National Cancer Institute cites three reported founder mutations in BRCA1/2: (1) BRCA1:c.68_69delAG, (2) BRCA1:c.5266dupC; and (3) BRCA3:c.5946delT [23]. Both BRCA1 mutations are known founders in the Ashkenazi Jewish population, with c.68_69delAG being the most frequent with approximately 0.9% of all Ashkenazi Jewish individuals being carriers [23]. The BRCA1:c.68_69delAG was found most frequently in individuals of Ashkenazi Jewish descent but was also observed in some Hispanic populations, likely owing to historical gene flow between these two populations in Europe and America [23,24,25].

Previously cited studies have detected evolutionary changes in coding and non-coding regions of BRCA1/2 genes within primate populations, and to extend the number of predicted amino acid changes that would affect gene function [8,10,12]. One such study found high-risk elements to be remarkably stable in hominoid primates, having been conserved in chimpanzee, gorilla, orangutan and rhesus macaque [12]. As Pavlicek et al. noted, the majority of insertion mutations took place in the ancestral lineage leading to hominoid primates after the split of Hominidae (25–14 MYA) and the rhesus macaque branch; more recent hominoid lineages acquired mostly deletions [12]. Furthermore, Puente et al. showed disproportionately lower incidences of breast and ovarian cancer in closely related hominids; such as chimpanzees, our closest living relatives [8].

Kalign for MSA generated two sets of alignments pertaining to each gene: (1) BRCA1, 8,207 bp; and (2) BRCA2, 12,520 bp. From these separate alignments, two horizontal cladograms were reconstructed to depict evolutionary relatedness. Fig. 1 & Fig. 2 both illustrate mRNA BRCA1/2 cancer susceptibility sequences nested within the branches of intergroup lineages or clades, in accordance with each respective species. In all cases, I found that cancer susceptibility variants precede a divergent event between two different species; outgroup taxons are represented by non-cancer variants, in each vertical instance (see Fig. 1 & Fig. 2). These results depart from the conventional interpretation of cross-species inheritance of a single mutation type and support the conclusions reported by others noted previously. My findings further suggest that specific mutation types (insertions or deletions) leading to altered gene expression in error-free DNA repair mechanisms vary widely among separate taxonomical rankings. I can then infer from these results that cancer-causing alterations appear to originate within localized gene pools at separate junctures throughout evolutionary time.

3.2 Pairwise Identity Ratios

In order to align a pair of sequences, a scoring system is required to score matches and mismatches [26]. This helps identify variation in rates and degrees of dissimilarity between two or more sets of homologous sequences. Variation itself is a consequence of different factors, including the mutation process, genetic drift, and natural selection. Substitutions, insertions, and deletions may occur at different rates over evolutionary time [27].

For mRNA BRCA 1/2 sequences, the relative rates of different substitutions can be empirically determined by comparing each sequence to a specific biomarker; *Homo sapiens* was chosen in this particular instance. These empirical measurements can then form the basis for determining various degrees of evolutionary change. Here, pairwise alignment revealed an unusually high proportion of dissimilarities between cancer susceptibility sequences among members of each distinct group or species. Overall, pairwise identity in the aligned segments ranged from 93% to 99% for non-cancer variants. Identity ratios dropped from 76% to 96% when cancer susceptibility or early onset sequences were included (see Table 1).

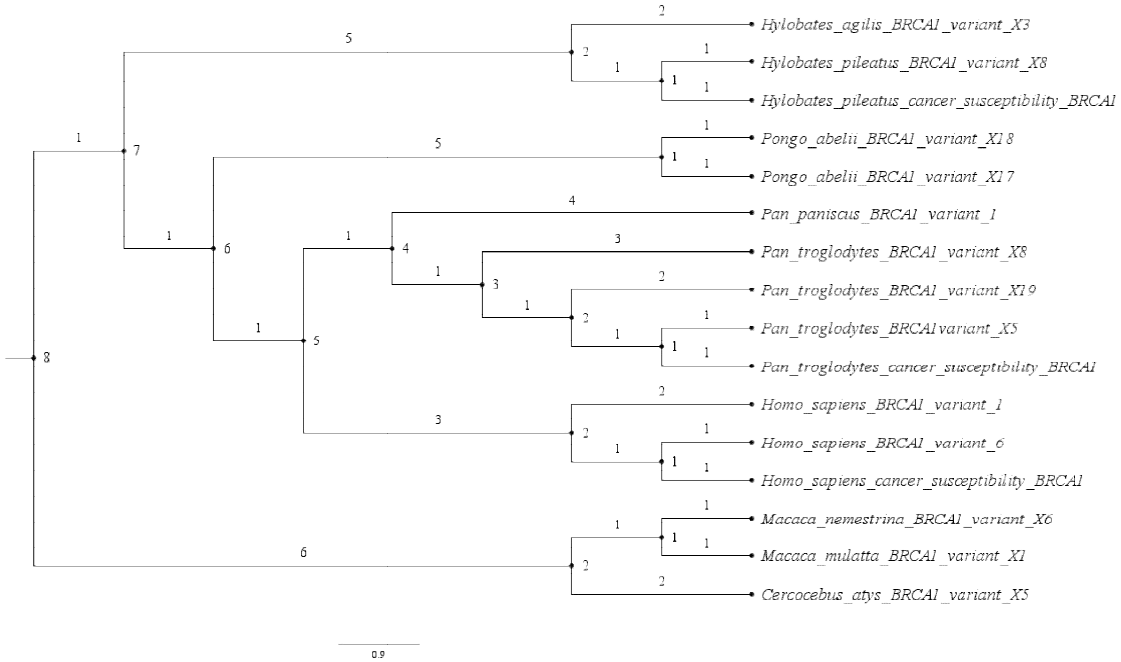

Fig. 1. Phylogenetic Tree of Sixteen Primate mRNA BRCA1 cancer susceptibility and non-cancer susceptibility sequences; including [in alphabetical order]: *Cercocebus atys*, *Homo sapiens*, *Hylobates agilis*, *Hylobates pileatus, Macaca nemestrina* , *Macaca mulatta*, *Pan paniscus*, *Pan troglodytes*, *Pongo abelii*

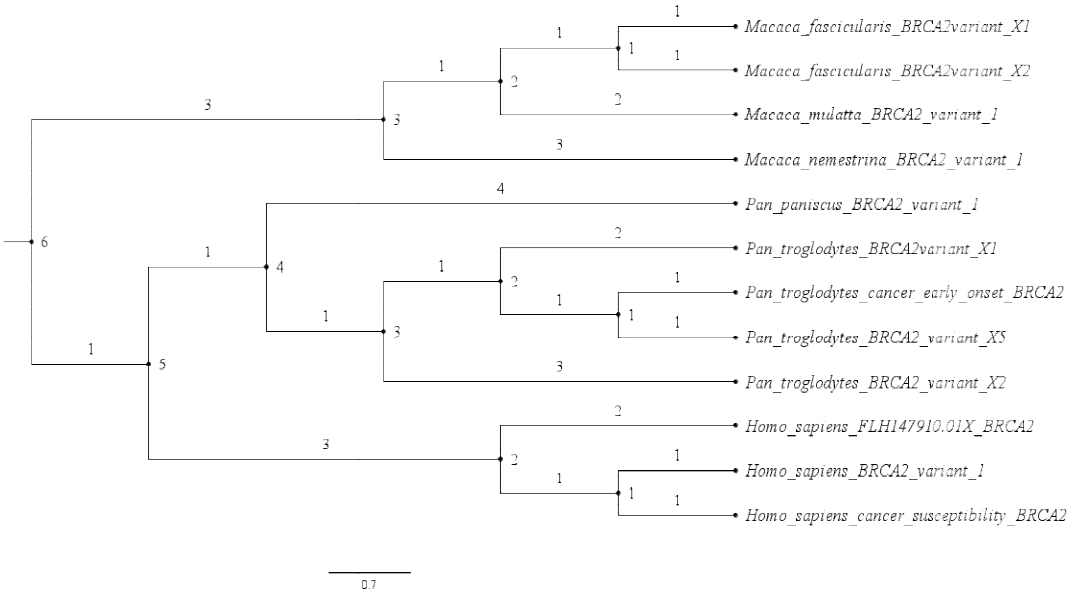

Fig. 2. Phylogenetic Tree of Twelve Primate mRNA BRCA2 cancer susceptibility and non-cancer susceptibility sequences; including [in alphabetical order]: *Homo sapiens*, *Macaca fascicularis*, *Macaca nemestrina* , *Macaca mulatta*, *Pan paniscus*, *Pan troglodytes*.

Table 1. Pairwise comparison of 10 primate family mRNA BRCA1/2 sequences against *Home sapiens*

Annotation Number	BRCA1	BRCA2
NR_027676.1 *Homo sapiens (7128 bp)*	0.99	-
NM_007294.3 *Homo sapiens (7224 bp)*	0.99	-
AF005068.1* *Homo sapiens (5693 bp)*	0.79	-
FLH147910.01X *Homo sapiens (10257 bp)*	-	0.99
NM_000059.3 *Homo sapiens (11386 bp)*	-	0.98
U43746.1* *Homo sapiens (10987 bp)*	-	0.96
XM_009432096.2 *Pan paniscus (7263 bp)*	0.98	-
XM_016930482.1 *Pan troglodytes (6743 bp)*	0.91	-
XM_016930483.1 *Pan troglodytes (6743 bp)*	0.91	-
NM_001301758.1* *Pan paniscus (5583 bp)*	0.77	
XM_003826866.2 *Pan paniscus (11661 bp)*	-	0.98
XM_016925134.1* *Pan troglodytes (11162 bp)*	-	0.91
XM_019039974.1 *Pan troglodytes (11324 bp)*	-	0.97
XM_002827484.2 *Pongo abelii (7103 bp)*	0.97	-
XM_009251713.1* *Pongo abelii (5794 bp)*	0.79	-
KM017622.1* *Pongo pygmaeus (5592 bp)*	0.76	-
XM_002824150.2 *Pongo abelii (11612 bp)*	-	0.97
XM_003913740.3 *Macaca nemestrina (11564 bp)*	-	0.95
XM_005585600.2 *Macaca fascicularis (11630 bp)*	-	0.95
XM_011748633.1* *Macaca nemestrina (11602 bp)*	-	0.89
XM_015119744.1 *Macaca mulatta (7489 bp)*	0.94	-
XM_011725007.1 *Macaca nemestrina (7483 bp)*	0.94	-
XM_003270250.3 *Nomascus leucogenys (11616 bp)*	-	0.97
XM_010367067.1 *Rhinopithecus roxellana (11382 bp)*	-	0.95
XM_012064631.1 *Cercocebus atys (11793 bp)*	-	0.95
XM_011997808.1* *Mandrillus leucophaeus (10873 bp)*	-	0.91
XM_003279499.1 *Nomascus leucogenys (7109 bp)*	0.97	-
XM_012046793.1 *Cercocebus atys (8151 bp)*	0.95	-
KM017619.1* *Hylobates pileatus (5586 bp)*	0.76	-
KM017620.1* *Hylobates agilis (5586 bp)*	0.76	-
KM017626.1* *Symphalangus syndactylus (5586 bp)*	0.76	-

** represents cancer susceptibility or early onset sequences*

3.3 Consensus Results of Breast Cancer 1 Early Onset BRCA1 mRNA Partial cds

Pairwise comparisons reveal quantifiable degrees of variation between two sets of sequences; namely, non-cancer variants versus cancer susceptibility. To further examine the question of statistical degrees of genetic alteration within a particular species, consensus values were generated between two or more individual sequences in human populations. As highlighted in Table 2, proportions of dissimilarity in *Homo sapiens* (BRCA1) ranges from 21% to 24%. Incidentally, these discrepancies are generally found to be lower in other primates (ranging from 7% to 22%); excluding *Macaca mulatta* (BRCA1). Coupled with the results from pairwise comparison, consensus results showed *Homo sapiens* having the highest proportion of dissimilarity among the entire dataset of primate families.

Table 2. Consensus analysis of BRCA1 cancer susceptibility against non-cancer variant *(Homo sapiens)*

Annotation number	Consensus %
NM_007294.3 *non-cancer variant*	0.99
AY890755.1*	0.76
AY888184.1* *(5592 bp)*	0.76
U14680.1* *(5711 bp)*	0.78
AB385129.1* *(5606 bp)*	0.76
AY751490.1* *(5524 bp)*	0.75
AF005068.1* *(5693 bp)*	0.77

** represents cancer susceptibility or early onset sequences*

4. CONCLUSION

On various levels, interspecies comparisons revealed the existence of unexpectedly high degrees of variation among the selected mRNA BRCA1/2 sequences. As the results from phylogenetic reconstruction demonstrate, cancer susceptibility variants are distinct to their respective clades and do not occur on outgroups of other species. Predisposition toward tumor suppression crosses taxonomical boundaries, but cancer-causing alterations in BRCA1/2 appear to originate within localized gene pools at separate junctures throughout evolutionary time. This supports the explanation that BRCA1/2 genes may be undergoing rapid evolution, as revealed by unusually high proportion of dissimilarities between cancer susceptibility sequences among members of each distinct group or species; *Homo sapiens* having the highest proportion of dissimilarity among the 10 primate families. The genomic datasets used in the study were limited relative to larger, more comprehensive investigations. Further examination is needed to validate my findings.

Supplementary Material

brca1-mrna-FL.aln, brca1-mrna-FL.nwk, brca1-mrna-FL.txt, brca2-mrna-FL.aln, brca2-mrna-FL.nwk, brca2-mrna-FL.txt

Additional Notes

[1] Annotations: (set 1) NR_027676.1, NM_007294.3, XM_002827484.2, XM_009251713.1, KM017622.1, XM_003279499.1, KM017619.1, KM017620.1, U43746.1, KM017626.1, XM_011725007.1, XM_015119744.1, XM_012046793.1, NM_001301758.1, XM_009432096.2, XM_016930482.1, XM_016930483.1, AF005068.1, NM_000059.3, XM_003826866.2, XM_016925134.1, XM_019039974.1, XM_002824150.2, XM_005585600.2, XM_011748633.1, XM_003913740.3, XM_011997808.1, XM_012064631.1, XM_010367067.1, XM_003270250.3; (set 2) NM_007294.3, AY890755.1, AY888184.1, U14680.1, AB385129.1, AY751490.1, AF005068.1

[2] BRCA Mutation Database, Department of Pathology, University of Utah. Web URL: http://arup.utah.edu/database/BRCA/

[3] UGENE was used in comparative sequence analysis. The DNA sequences noted above are in FASTA format. They were obtained from the NCBI database archives.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Wideroff L, Vadaparampil ST, Greene MH, Taplin S, Olson L, Freedman AN. Hereditary breast/ovarian and colorectal cancer genetics knowledge in a national sample of US physicians. Journal of Medical Genetics. 2005;42(10):749-755.

2. Burke W, Daly M, Garber J, Botkin J, Kahn MJE, Lynch P, Thomson E. Recommendations for follow-up care of individuals with an inherited predisposition to cancer: II. BRCA1 and BRCA2. Jama. 1997;277(12):997-1003.
3. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nature Medicine. 2004;10(8):789.
4. Fan S, Meng Q, Auborn K, Carter T, Rosen EM. BRCA1 and BRCA2 as molecular targets for phytochemicals indole-3-carbinol and genistein in breast and prostate cancer cells. British Journal of Cancer. 2006;94(3):407-426.
5. Gayther SA, de Foy KA, Harrington P, Pharoah P, Dunsmuir WD, Edwards S M, Ford D. The frequency of germ-line mutations in the breast cancer predisposition genes BRCA1 and BRCA2 in familial prostate cancer. Cancer Research. 2000;60(16):4513-4518.
6. Mitra A, Fisher C, Foster CS, Jameson C, Barbachanno Y, Bartlett, Easton D. Prostate cancer in male BRCA1 and BRCA2 mutation carriers has a more aggressive phenotype. British Journal of Cancer. 2008;98(2):502-507.
7. King MC, Marks JH, Mandell JB. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science. 2003;302(5645):643-646.
8. Puente XS, Velasco G, Gutiérrez-Fernández A, Bertranpetit J, King MC, López-Otín C. Comparative analysis of cancer genes in the human and chimpanzee genomes. BMC Genomics. 2006;7(1):15.
9. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010;31(1):27-36.
10. Lou DI, McBee RM, Le UQ, Stone AC, Wilkerson GK, Demogines AM, Sawyer SL. Rapid evolution of BRCA1 and BRCA2 in humans and other primates. BMC Evolutionary Biology. 2014;14(1): 155.
11. Huttley GA, Easteal S, Southey MC, Tesoriero A, Giles GG, McCredie MR, Venter DJ. Adaptive evolution of the tumour suppressor BRCA1 in humans and chimpanzees. Nature Genetics. 2000; 25(4):410-413.
12. Pavlicek A, Noskov VN, Kouprina N, Barrett JC, Jurka J, Larionov V. Evolution of the tumor suppressor BRCA1 locus in primates: Implications for cancer predisposition. Human Molecular Genetics. 2004;13(22):2737-2751.
13. Orban TI, Olah E. Emerging roles of BRCA1 alternative splicing. Molecular Pathology. 2003;56(4):191.
14. Roncaglioni MC, Falanga A, D'Alessandro AP, Alessio MG, Casali B, Donati MB. Evidence of a warfarin-sensitive cancer procoagulant in V2 carcinoma. Haematologica. 1989;74(2):143-147.
15. Holt JT, Thompson ME, Szabo C, Robinson-Benion C, Arteaga CL, King M. C, Jensen RA. Growth retardation and tumour inhibition by BRCA1. *Nature genetics.* 1996;12(3):298.
16. Huang YW. Association of BRCA1/2 mutations with ovarian cancer prognosis: An updated meta-analysis. Medicine. 2018;97(2):e9380.
17. Lassmann T, Sonnhammer EL. Kalign–an accurate and fast multiple sequence alignment algorithm. BMC Bioinformatics. 2005;6(1):1.
18. Felsenstein J. Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal of Molecular Evolution. 1981;17(6):368-376.
19. Ragan MA. Phylogenetic inference based on matrix representation of trees. Molecular Phylogenetics and Evolution. 1992;1(1):53-58.
20. Orr HA. The distribution of fitness effects among beneficial mutations. Genetics. 2003;163(4):1519-1526.
21. Perfeito L, Fernandes L, Mota C, Gordo I. Adaptive mutations in bacteria: High rate and small effects. Science. 2007; 317(5839):813-815.
22. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993;134(4):1289-1303.
23. Foulkes WD, Knoppers BM, Turnbull C. Population genetic testing for cancer susceptibility: Founder mutations to genomes. Nature Reviews Clinical Oncology. 2016;13(1):41.
24. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Cortes ML. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012; 486(7403):405-409.

25. Gillet NA, Willems L. Whole genome sequencing of 51 breast cancers reveals that tumors are devoid of bovine leukemia virus DNA. Retrovirology. 2016;13(1):75.
26. Higgins DG, Sharp PM. CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. Gene. 1988;73(1):237-244.
27. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution. 1980;16(2):111-120.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history/25865