CME LIVE OPEN ACCESS

CME Journal of Clinical Case Report

Research Article

Explore Selection Patterns in Cancer-Associated Genes in Various Species: The Comparative Analysis.

Sahar Madani[1*], Irine Gotsiridze[1]

[1]*Faculty of Informatics and Control Systems, Biomedical Engineering department, Georgian Technical University, Tbilisi, Georgia.*

Corresponding Author: Sahar madani, Faculty of Informatics and Control Systems, Biomedical Engineering department, Georgian Technical University, Tbilisi, Georgia.

Received: 2024 Apl 02 **Accepted:** 2024 Apl 15 **Published:** 2024 May 23

Abstract

Cancer has long been a complex and puzzling disease for humans to understand. It is primarily caused by the mutation of two groups of genes known as oncogenes and tumor suppressor genes (TSGs). Certain inherited medical conditions can increase the risk of developing more common forms of thyroid cancer associated with these genes. Uncommon genetic conditions have also been linked to higher rates of thyroid cancer. The genes associated with thyroid cancer are an ideal model for investigating the effects of selection at different stages of thyroid cancer.

A study is currently underway to identify patterns of positive selection in seven highly mutated genes associated with thyroid cancer through a detailed phylogenetic analysis. The nucleotide sequences for these genes were obtained from the Ref Seq database at the National Center for Biotechnology Information (NCBI) and aligned using the Muscle software. The best-fit model for nucleotide substitution was chosen using JMODELTEST, and the corresponding phylogenetic trees were constructed using PAML. The CODMEL software was used to calculate the non-synonymous to synonymous ratio and to select the most plausible evolutionary model for each gene.

The results of the analysis revealed that three genes (BDP1, TG, and TNN) showed strong signatures for positive selection, with p-values of 8.49E-18, 0.002, and 1.75E-87, respectively. The study is ongoing, and the analysis will be extended to the non-coding part to increase the evidence of positive selection on these genes. The aim is to better understand the evolutionary history of thyroid cancer-related genes, enrich the evidence data for the positively selected sites, and improve our understanding of the functionalities of these proteins, domains, and motifs. This could potentially provide valuable insights into drug-targeting sites.

Keywords: Positive Selection, Thyroid Cancer, Synonymous, Non-Synonymous, Phylogenetic Tree, Multiple Alignment, Comparative Analysis.

1. Introduction

1. 1 Sequences Data Collection: The detected signals of positive selection within thyroid genes in this study are either determined by ancient mammalian bouts of selection or more recent mammalian group variegation with no clear evidence to suggest any role for sexual selection during the diversification of these genes in mammals. The protein-coding and noncoding sequences for 20 mammalian species used in this study were obtained from NCBI database [The National Center for Biotechnology Information] and the ENSEMBL database. For genes with multiple protein sequences representing spliced forms, we selected the longest sequences as representatives for further analysis. The gene associated with thyroid cancer for mammalian were retrieved from GDC Data Portal (https://portal.gdc.cancer.gov/). The data set included 7 genes including BRAF, TG, BDP1, HMCN1, HRAS, NRAS, and TTN genes for 20 species [1,2].

1.2 Phylogenetic Tree Reconstruction.

The multiple sequence alignment (MSA) for *all 7 genes* was built using the retrieved coding sequences translated to amino acids and further back-translated to nucleotides and muscle implemented in SEAVIEW V4. We reduplicate the multiple sequences alignment in G BLOCKS to reduce the false positive resulting from incorrectly aligned positions. After the filtered MSA we used MrAIC to detect the best evolutionary models from the filtered MSA and we used AICc (Akaike information criterion correction) for *the* models' comparison. We choose the model with the lowest AIC (2thyroid information criterion) value as it takes into account both the likelihood of the model and the number of parameters contained within a model. PAML is a package of programs for phylogenetic Analyses of DNA and protein sequences using maximum likelihood (ML). PhyML v4 was used to get Phylogenetic gene-based tree reconstructions.

The phylogenetic tree for the individual thyroid genes was reconstructed using Maximum Likelihood (ML) in MEGA 5.05 and the reliability of the tree branching was assessed using 1,000 replicate bootstraps [4,5].

1.3 Tests of Selection.

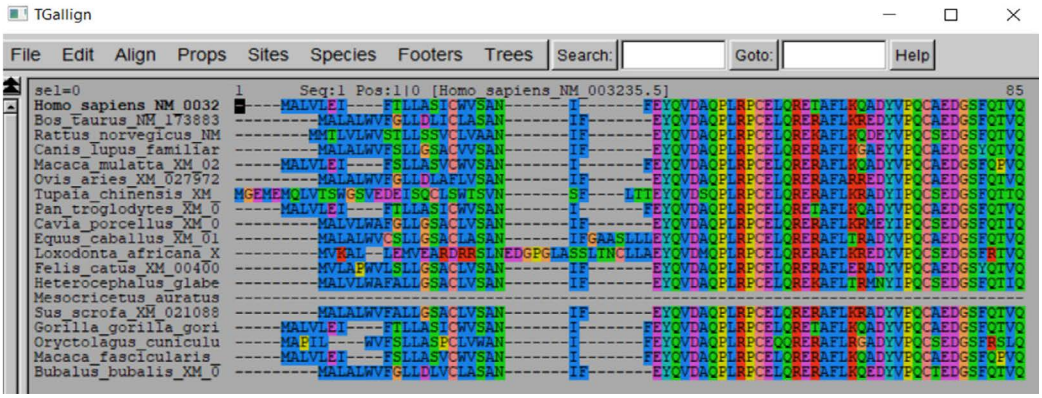
Maximum probability is a customary statistical method for estimating unknown parameters of a probability model. A parameter is some descriptor of the model. An acquainted model may be the normal distribution of a population with two parameters: the mean and variance. In phylogenetics, there are many parameters, which include rates, differential transformation costs, and, most importantly, the tree itself. Maximum likelihood is the third method used to build trees. Likelihood provides probabilities of the sequences given a model of their evolution on a particular tree. The more probable the sequences given the tree, the more the tree is preferred. Phylogenetic methods for comparative evaluation of DNA and protein sequences are becoming ever more vital with the rapid accumulation of molecular sequence data, spearheaded by way of numerous genome projects [5-11].

It is now frequent for phylogeny reconstruction to be carried out with the usage of massive record sets involving hundreds or even thousands of genes. Similarly, phylogenetic strategies are widely used to estimate the evolutionary quotes of genes and genomes to observe footprints of natural selection, and the evolutionary information is used to interpret genomic data. For example, both evolutionary conservations indicating terrible purifying selection and accelerated evolution driven via nice Darwinian resolution have been employed to discover functionally good-sized regions of the genome Tests of selection were executed using CODEML within Phylogenetic Analysis by Maximum Likelihood (PAML; v 4.4e). For each thyroid gene, the Likelihoods of the M7 and M8 models were calculated to recognize phylogenetic groups that had experienced positive selection. The M7 model postulates that ω values lie between 0 and 1 across the sequences, which is indicative of purifying selection or neutral evolution. The M8 model lets ω to override1, which is distinguishing of positive selection. To ensure that relaxed selection was not misinterpreted as positive selection, the likelihood of the null model M8a (where ω is fixed at 1) was also compared to that of M8. When codons within an alignment were identified as being under positive selection, the Bayes Empirical Bayes (BEB) method was used to identify the specific codon sites under positive selection. Codon sites with a posterior probability of positive selection higher than 90% were further evaluated based on the nature of the amino acid substitution occurring between species (e.g., cysteine residues influencing the secondary structure and phosphorylation sites that can change the function or localization of a protein). For each gene, the number of nonsynonymous substitutions per nonsynonymous site (dN) and the number of synonymous substitutions per synonymous site (dS) were calculated using a maximum-likelihood method CODEML implemented in PAML v4.6 [7-27].

The site model was used to identify positive selection. Which is assessed based on the nonsynonymous/synonymous substitution ratio (dn/ds = ω). If there is no election, the nonsynonymous and synonymous substitution ratio are fixed with the same probability (ω =1). Furthermore, because of selective advantage, the selection can increase fixation probabilities for nonsynonymous mutations (positive selection, ω > 1), or because of selective coerce the probabilities are decreased (purifying/negative selection, ω < 1). Tests of selection were executed using CODEML within Phylogenetic Analysis by Maximum Likelihood (PAML; v 4). For each thyroid gene, the Likelihoods of the M7 and M8 models were calculated to recognize phylogenetic groups that had experienced positive selection [8]. The M7 model postulates that ω values lie between 0 and 1 across the sequences, which is indicative of purifying selection or neutral evolution. The M8 model lets ω to override1, which is distinguishing of positive selection. To ensure that relaxed selection was not misinterpreted as positive selection, the likelihood of the null model M8a (where ω is fixed at 1) was also compared to that of M8. When codons within an alignment were identified as being under positive selection, the Bayes Empirical Bayes (BEB) method was used to identify the specific codon sites under positive selection [11-25].

For each gene the number of nonsynonymous substitutions per nonsynonymous site (dN) and the number of synonymous substitutions per synonymous site (dS) were calculated using a maximum-likelihood method CODEML implemented in PAML v4.6. Estimations of dN, dS and dN/dS, were obtained using six different models (Model 0, 7, 8 and 8a. Model 0 (M0, one-ratio) was used to estimate global dN/dS, dN and dS. Models 7 (M7, beta) and 8 (M8, beta +o > 1), approximate the dN/dS variation over sites through a beta distribution, estimating the proportion and the dN/dS ratio of the positively selected sites, whereas M8 only includes site-classes above neutrality. Comparisons between models M8 and M8a were used to identify deviations from neutrality. This pairwise comparison focuses on testing whether sites belonging to a site-class with a dN/dS> 1 are evolving differently from near neutrality (dN/dS&1). For each pairwise comparison M7 versus M8, M8 versus M8a, the LRT obtained were compared against a 2 distribution. The degrees of freedom, used to obtain the 2 critical values, were the difference in the number of parameters in the null and alternate model for each pairwise test. For each P value, we also estimated the corresponding q value. When the q value was below, the P value obtained for the LRT value the gene was considered to be under positive selection (1), and when above, the gene was considered negatively selected (0).

Figure 1: Alignment of Integrin Thyroid



genes for 20 mammalian species. GenBank accession numbers: Human (Homo sapiens), NM_018429.3; Chimpanzee (Pan troglodytes), XM_024354641.1 ; Rhesus monkey (Macaca mulata), XM_028830889.1; Mouse (Mus musculus), NM_010575.2; Rat (Ratus norvegicus), XM_017600986; Rabbit (Oryctolagus cuniculus), XM_01734; Dog (Canis lupus familiaris), XM_022414800.1; Horse (Equus caballus),XM_023622070.1; Pig (Sus scrofa), XM_021075901.1; Cattle (Bos taurus), XM_024979882.1; sheep (Ovis_aries),XM_027964793.1; (Tupaia chinensis),XM_027773115; Chinese hamster (Cricetulus griseus),XM_02742010; horse (Equus caballus),XM_023622070.1;guinea pig (Cavia porcellus), XM_023561141.1; cat (Felis catus), XM_023259509.1;mole-rat (Heterocephalusglaber), XM_02124; golden hamster (Mesocricetus auratus), XM_021233; (Gorilla gorilla), XM_019; monkey (Macaca fascicularis), XM_01543241; buffalo (Bubalus bubalis), XM_025274121.1

Citation: Madani, S., Gotsiridze, I., (2024). Explore Selection Patterns in Cancer-Associated Genes in Various Species: The Comparative Analysis. CME Journal of Clinical Case Report,1(1), 1-12.

1.4 How to Calculate Positive Selection in Genes?

SeaView - Multiplatform GUI for molecular phylogeny

Version 5.0

NEW: seaview performs reconcilation between gene and species trees using **Treerecs**
NEW: bootstrap support optionally with the "Transfer Bootstrap Expectation" method
NEW: trimming-rule to shorten long sequence names in phylogenetic trees
NEW: 64-bit version for the MS Windows platform
NEW: multiple-tree windows
NEW: seaview uses **PHYLIP v3.696** to compute parsimony trees
NEW: seaview can be run without GUI using a command line
NEW: seaview drives the **PhyML v3.1** program to compute maximum likelihood phylogenetic trees.
NEW: seaview drives the **Gblocks** program to select blocks of conserved sites.

SeaView is a multiplatform, graphical user interface for multiple sequence alignment and molecular phylogeny.

- SeaView reads and writes various file formats (NEXUS, MSF, CLUSTAL, FASTA, PHYLIP, MASE, Newick) of DNA and protein sequences and of phylogenetic trees.
- SeaView drives programs muscle or Clustal Omega for multiple sequence alignment, and also allows to use any external alignment algorithm able to read and write FASTA-formatted files.
- Seaview drives the Gblocks program to select blocks of evolutionarily conserved sites.
- SeaView computes phylogenetic trees by
  - parsimony, using PHYLIP's dnapars/protpars algorithm,
  - distance, with NJ or BioNJ algorithms on a variety of evolutionary distances,
  - maximum likelihood, driving program PhyML 3.1.
- Seaview can use the Transfer Bootstrap Expectation method to compute the bootstrap support of PhyML and distance trees.
- Seaview uses the Treerecs method to reconcile gene and species trees.
- SeaView prints and draws phylogenetic trees on screen, SVG, PDF or PostScript files.
- SeaView allows to download sequences from EMBL/GenBank/UniProt using the Internet.

Screen shots of the main alignment and tree windows. Dialog window to perform Maximum-Likelihood tree-building.
On-line help document.Old seaview version 3.2

Figure 2 : Aligned Using The Muscle Software As Implemented In The Seaview Package.

We collected nucleotide sequences for seven genes (BDP1, BRAF, HMCN1, HRAS, NRAS, TG and TTN) from NCBI, Ref Seq database. The sequences were then aligned using the Muscle software as implemented in the SEAVEIW package. We selected the best-fit model for nucleotide substitution using JMODELTEST, and then we constructed the corresponding phylogenetic trees using PAML. CODMEL was used for calculating non-synonymous to synonymous ratio and for selecting the most plausible evolutionary model for each gene (0, 7, 8 and 8a).

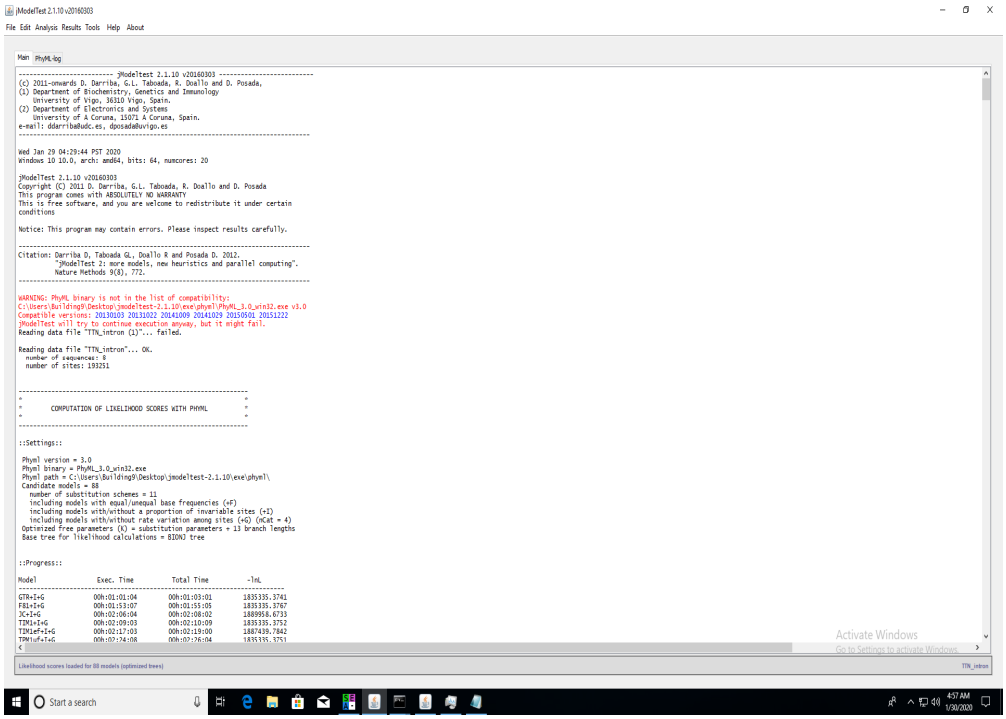


Figure 3: Selected The Best-Fit Model For Nucleotide Substitution Using Model test

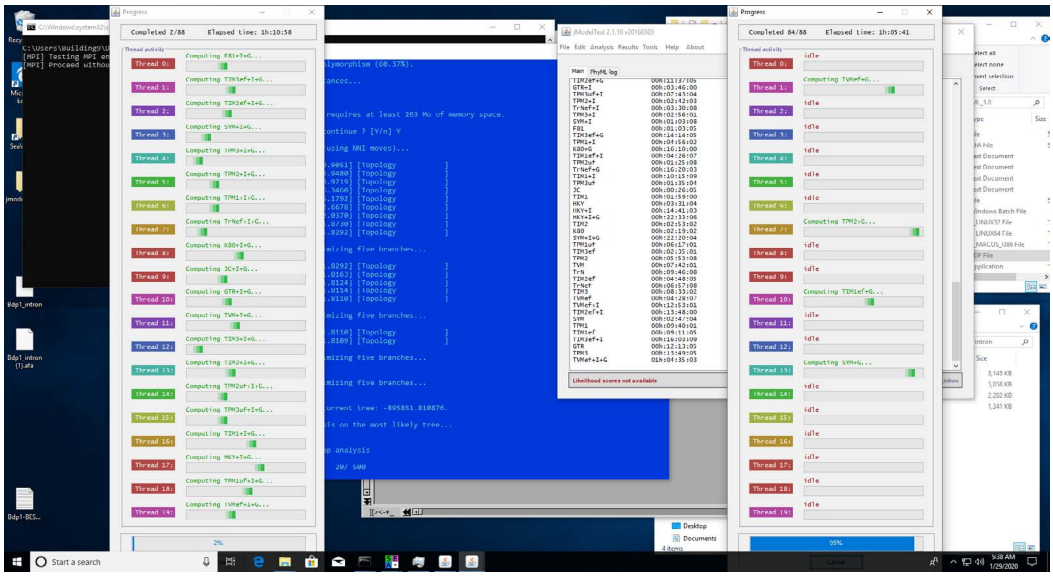


Figure 4: The Process Of Selecting The Best-Fit Model For Nucleotide Substitution Using Model Test

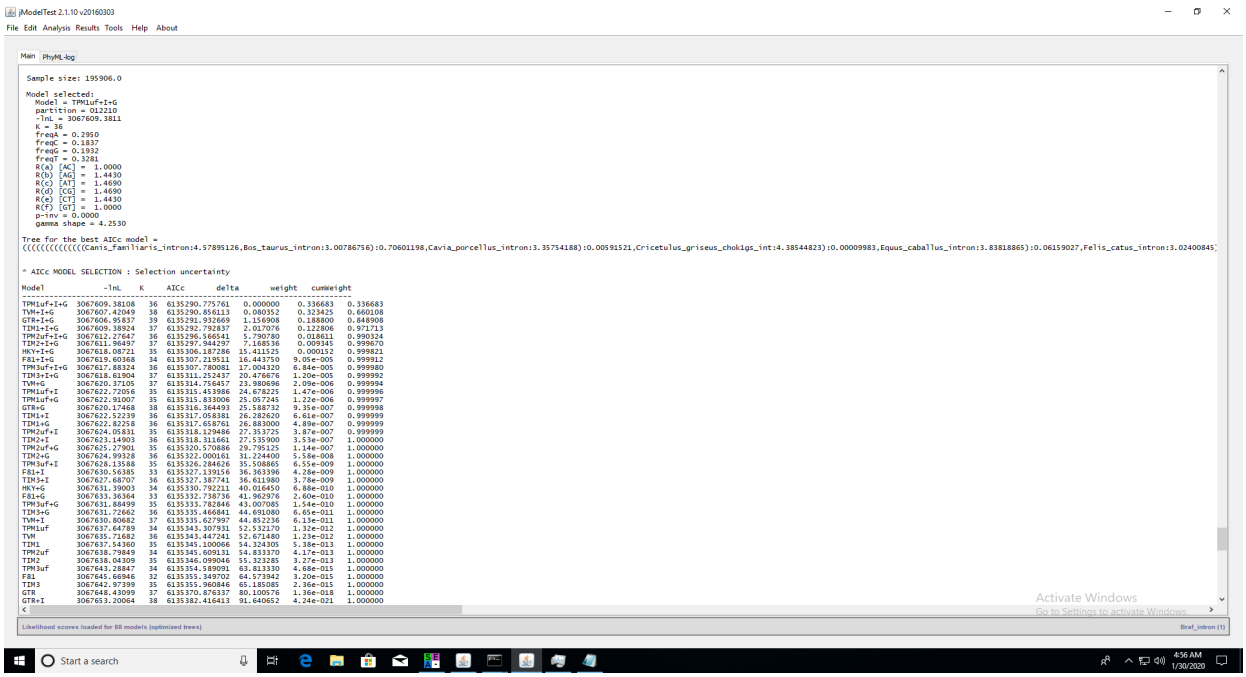


Figure 5: The Process Of Selecting The Best-Fit Model For Nucleotide Substitution Using Model Test.

We will try to understand the evolutionary history of thyroid cancer-related genes, and enrich the evidence data for the positively selected site. Understanding the role of positive selection in amino acid substitution will increase our understanding of the functionalities of these proteins, domains and motifs which could provide a recommendation for drug-targeting sites.

Citation: Madani, S., Gotsiridze, I., (2024). Explore Selection Patterns in Cancer-Associated Genes in Various Species: The Comparative Analysis. CME Journal of Clinical Case Report,1(1), 1-12.

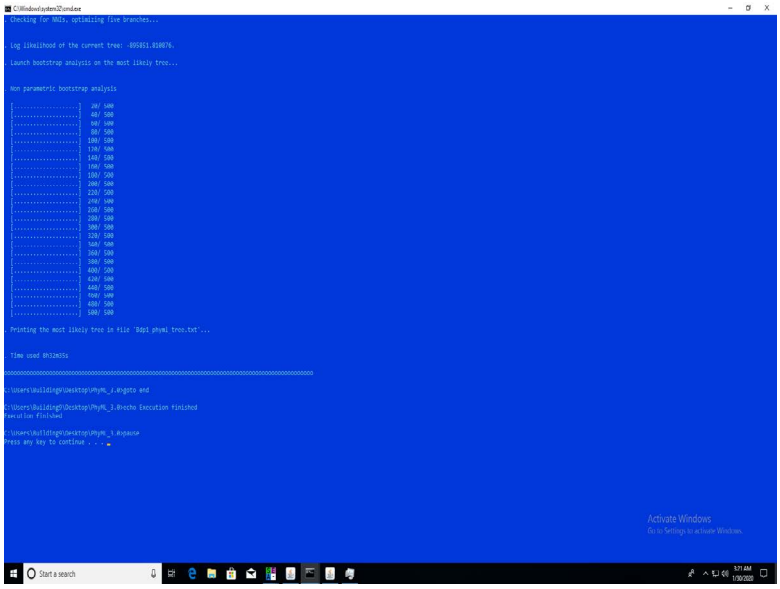


Figure 6: The Process Of Selecting The Best-Fit Model For Nucleotide Substitution Using Model Test.

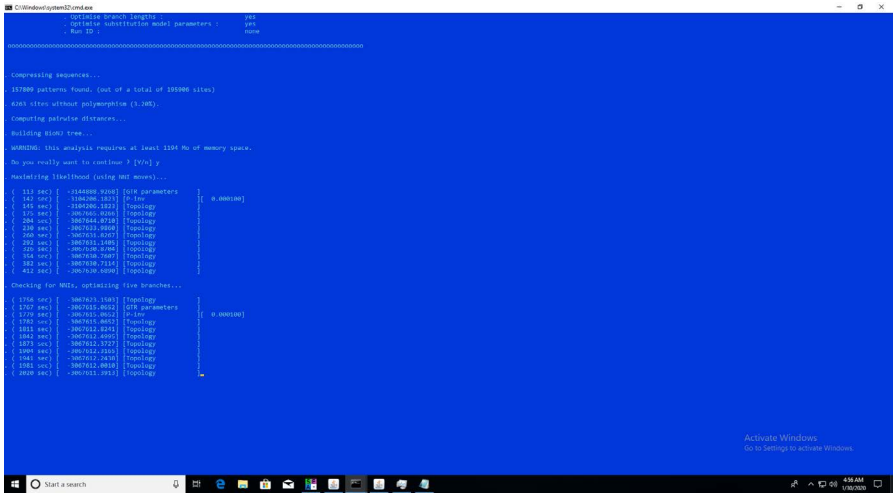


Figure 7: Process Of Selected The Best-Fit Model For Nucleotide Substitution Using Model Test.

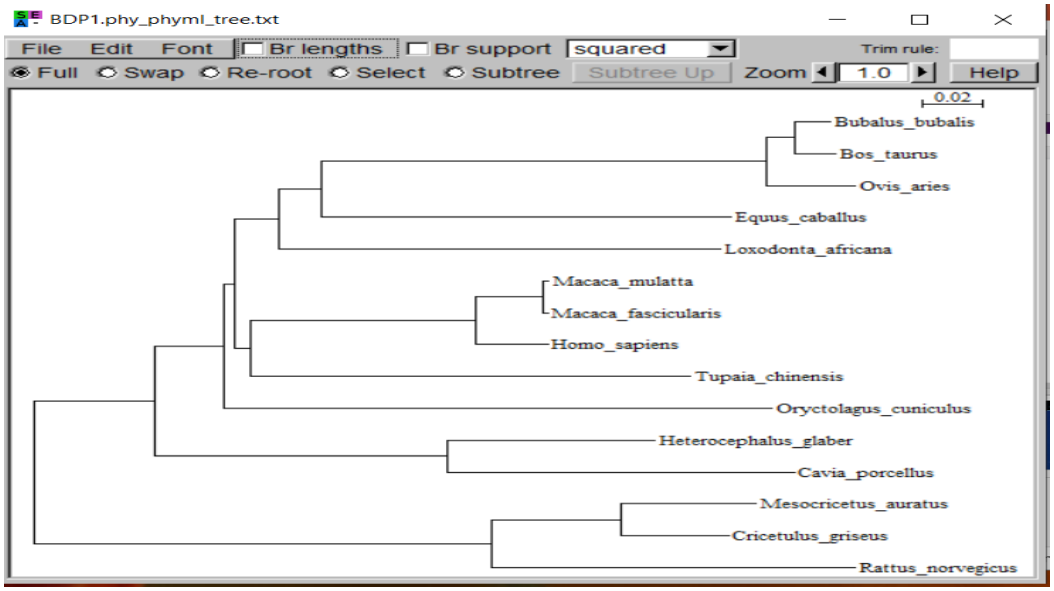


Figure 8: Bdp1 Phy-Phyml-Tree.

Citation: Madani, S., Gotsiridze, I., (2024). Explore Selection Patterns in Cancer-Associated Genes in Various Species: The Comparative Analysis. CME Journal of Clinical Case Report,1(1), 1-12.

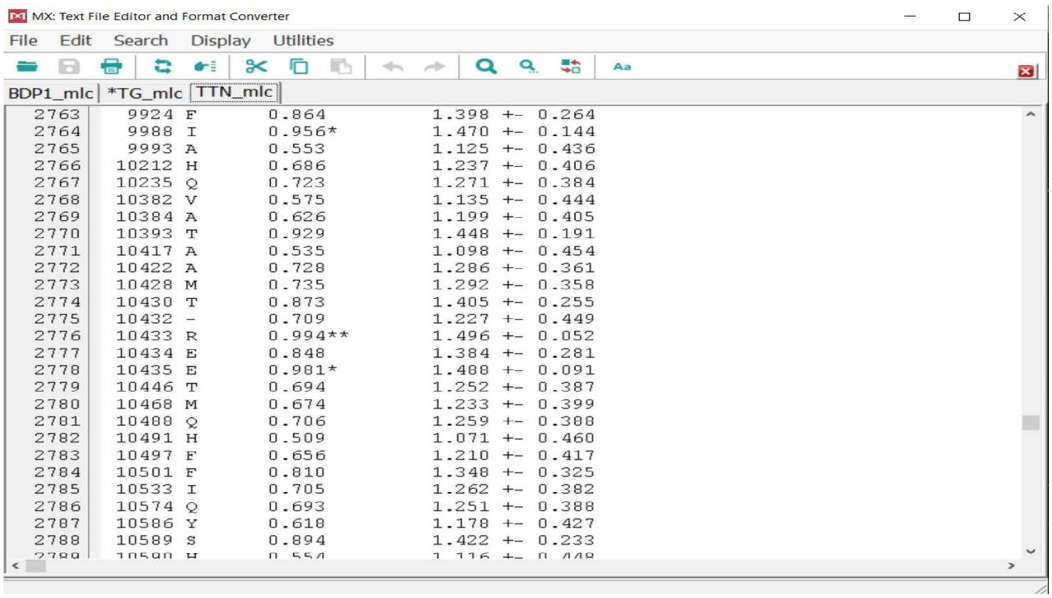


Figure 9: The Process Of Selection Of The Three Genes Was Identified To Be Under Positive Selection.

3. Results and Discussion

The proteins are molecules that play many critical roles in our body. In our study by used a comparative genomics approach to identify genes that are under positive selection in seven genes of thyroid cancer. We find that positive selection targets a wide range of different functions in the thyroid cancer genes, including cell surface proteins such as the TG gene. Three genes were identified to be under positive selection. According to the Bio Cyc database, the three Genes were assigned to Location and functional categories. Three genes were identified to be under positive selection. According to the Bio Cyc database, the three Genes were assigned to Location and functional categories (Table 1). As in previous studies [10], they find that the proteins that typically undergo positive selection are exposed on the cell surface. Many of these proteins are directly involved in interactions with the host, phages, or other bacteria. In our results may suggest that the finding of TG and TTN in intracellular space and membrane respectively may indicate that a substantive of this positive selection may be related to other adaptive processes than immune/defense evasion and interactions of the cell with phages and other microorganisms, possibly metabolic adaptations in response to a changing environment. Thyroglobulin has been shown to interact with Binding immunoglobulin protein. For that, we strongly suggest that the positive selection acting on the TG gene and TTN gene is caused by selection to avoid phage binding and/or binding of defense-related molecules from the host immune/ defense system. In these genes, the Positive selection looks alike to be related to selection to avoid recognition by a host immune system and/or binding of phages and colicins [23,24].

3.1 Location of Positively Selected Sites and Sequence Variation

We plotted the genic location of positively selected sites for the 10 groups that had sites detected from whole-sequence analysis Positively selected sites were not homogeneously distributed among regions; 69% (83 of 116) of sites were located in LRRs. The heterogeneous distribution of positively selected sites was clear from the comparison of the proportion of sites under selection within the NBS and LRR regions, the two domains that occur in all proteins. Two-by-two contingency tests revealed that sites under positive selection occur significantly more frequently in the LRR domains ($P \ll 0.001$; $\chi^2 = 56.13$). Nonetheless, 33 positively selected sites were located in non-LRR regions.

We also studied the distribution of indels across regions in the groups in which positive selection was detected. In a two-by-two contingency test, LRRs had a significantly larger proportion of indels than non-LRR domains ($P \ll 0.001$; $\chi^2 = 145.14$). However, the high incidence of indels in the LRR was not unique to proteins under positive selection. Positively selected sites were not detected in groups 1, 5, 7, and 16, but indels were also more frequent in the LRR than the NBS for these sequence groups ($P \ll 0.001$, $\chi^2 = 91.36$). These observations are important for two reasons. First, the high incidence of gaps in the LRR regions provides additional evidence that LRRs are more labile than other domains. Second, gaps alone do not account for the high incidence of positively selected sites in LRRs.

Table 1 : Positively Selected Positions In Genes.

Gene	P-value M7:M8	P-value M8:M8a	Positively Selected Positions
BDP1	1.8739E-24	8.49829E-18	443s,702L,778D,1204F,2185R,2528L
BRAF	0.347221283	0.101829993	
HMCN1	7.26898E-14	#NUM!	
HRAS	#NUM!	#NUM!	
NRAS	#NUM!	0.316215894	
TG	5.4209E-07	0.00248918	9T,1866Q,1492V
TTN	3.8448E-166	1.75045E-87	4086D,4128T,4376L,4393M,4732V,4751S,4774L,5337M, 10433R,10435E,10848R,10862L,11158Q,11371L,12291T, 12292I,12296R,12434T,12499L,12500R,12509H,12510P, 12513R,12520K,12522H,12538P,12555E,12556V,12568K, 12590A,12844A,12901I,12903P,12919K,12932A,12939V, 12963P,12967R,12974A,12985K,13207V,13236T,13258P, 13263P,13351A,13356P,133557E,13358A,13363V,13470V, 13512L,13514P,13524K,13568L,13654P,13825L,13888E, 13893I,13895I,14181T,20378L,22476L,23082R,23785V, 28292I,32516T.

Table2 : Genes That Show Evidence Of Positive Selection.

Gene name	Gene	Likelihood ratio	dN/dS	Function	Location
thyroglobulin	TG	9.14853	0.29289	making thyroglobulin protein, one of the largest proteins in the body. which found only in the thyroid gland [2].	extracellular space
B double prime 1	BDP1	73.833708	0.56531		Nucleus
Titin	TTN	393.099362	0.11445	making a very large protein named titin which plays an important role in skeletal muscles and in cardiac muscle [3].	cytoskeleton, nucleus, membrane

Table 3 : Genes Models Test.

Gene	Model 0 (lnL)	dN/dS	Model 7 (lnL) null model
BDP1	-48574.23142	0.56531	-47934.763
BRAF	-7712.401333	0.05714	-7655.0213
HMCN1	-86441.95578	0.16802	-85331.935
HRAS	-2561.463252	0.01492	-2538.7533
NRAS	-1673.179197	0.0075	-1671.2698
TG	-57284.60195	0.29289	-56128.97
TTN	-512311.6603	0.11445	-501928.08

Table 4 : Genes Models Parameters.

Gene	Model 7 (Parameters)	Model 8 (lnL) alt model	Model 8 (Parameters)	Model 8a (lnL) null model
BDP1	p = 0.46434 q = 0.33840	-47880.129	p0 = 0.92401 p = 0.54738 q = 0.45765 (p1 = 0.07599) w = 2.25355	-47917.04626
BRAF	p = 0.08412 q = 1.15138	-7653.9635	p0 = 0.99778 p = 0.09249 q = 1.36260 (p1 = 0.00222) w = 1.90081=	-7655.301804
HMCN1	p = 0.33623 q = 1.50670	-85301.682	p0 = 0.95879 p = 0.46693 q = 2.76439 (p1 = 0.04121) w = 1.14128	-85301.34069
HRAS	p = 0.09687 q = 4.78581	-2538.7551	p0 = 0.99999 p = 0.09687 q = 4.78581(p1 = 0.00001) w = 7.89021	-2537.479488
NRAS	p = 0.01271 q = 0.47851	-1671.2704	p0 = 0.99999 p = 0.01277 q = 0.48175(p1 = 0.00001) w = 1.00000	-1671.772675
TG	p = 0.37421 q = 0.78368	-56114.542	p0 = 0.95215 p = 0.43836 q = 1.10917(p1 = 0.04785) w = 1.41412	-56119.11664
TTN	p = 0.12951 q = 0.77598	-501547.2	p0 = 0.98656 p = 0.14922 q = 1.04079(p1 = 0.01344) w = 3.79429	-501743.7497

Table 5 : Lrt (M8 Vs M8a).

Gene	LRT (M7 vs M8)	p-value	LRT (M8 vs M8a)	p-value	Positively Selected Positions
BDP1	109.268044	1.8739E-24	73.833708	8.49829E-18	
BRAF	2.115586	0.347221283	2.676646	0.101829993	
HMCN1	60.50515	7.26898E-14	-0.683108	#NUM!	
HRAS	-0.0036	#NUM!	-2.551154	#NUM!	
NRAS	-0.001256	#NUM!	1.004534	0.316215894	
TG	28.855668	5.4209E-07	9.14853	0.00248918	
TTN	761.764802	3.8448E-166	393.099362	1.75045E-87	

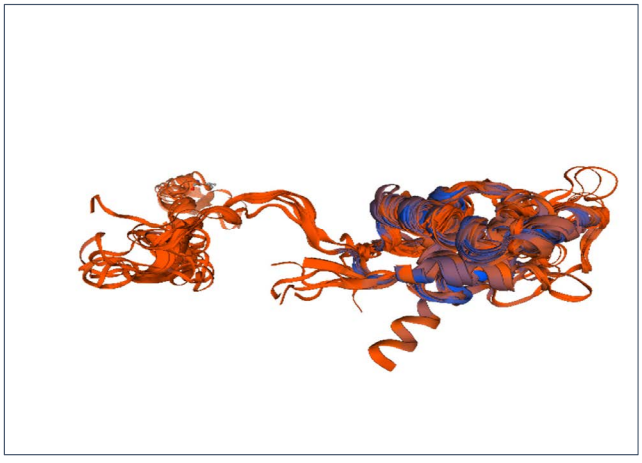


Figure10: Bdp1 Protein Structure

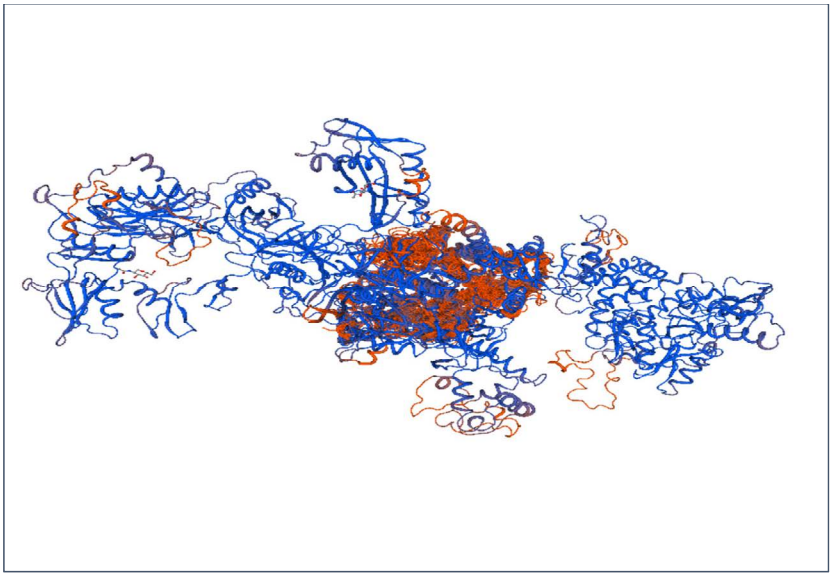


Figure 11:Tg Protein Structure

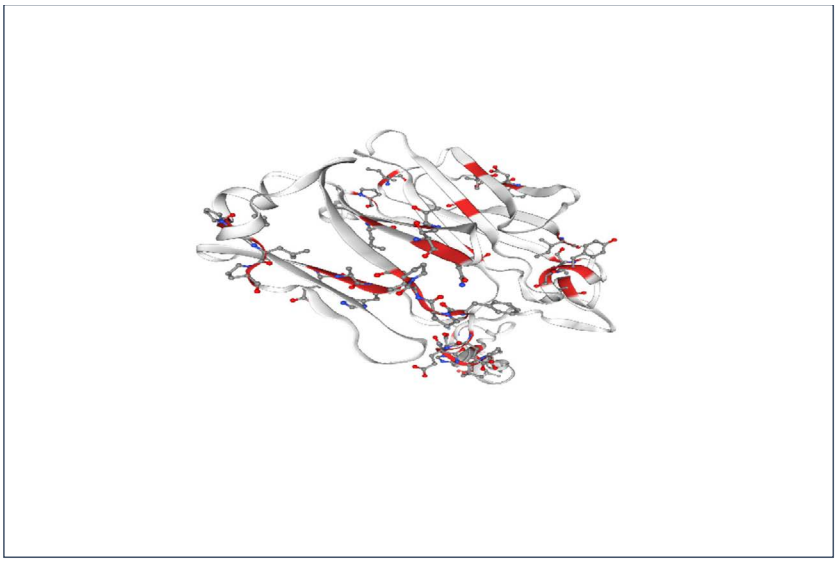


Figure 12: Ttn Protein Structure

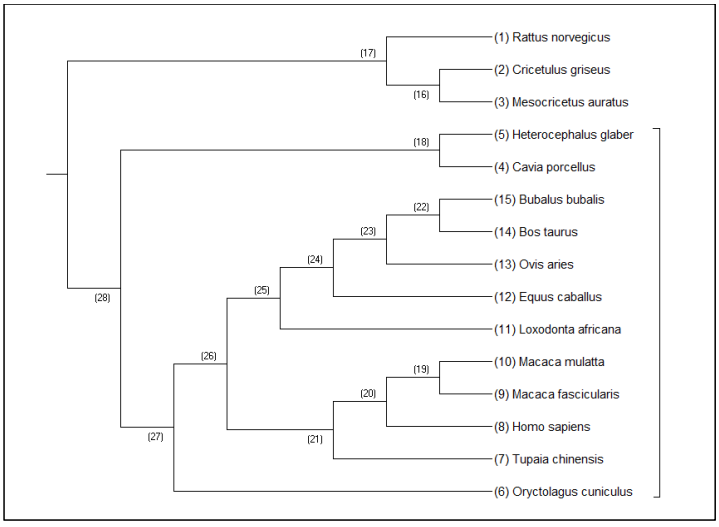


Figure 13:Bdp1 Phylogenetic Tree

Citation: Madani, S., Gotsiridze, I., (2024). Explore Selection Patterns in Cancer-Associated Genes in Various Species: The Comparative Analysis. CME Journal of Clinical Case Report,1(1), 1-12.

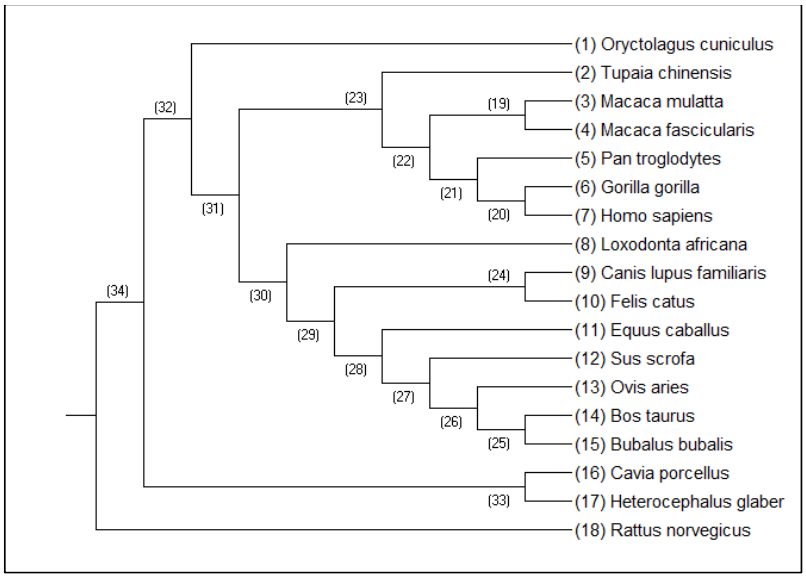


Figure 14:Tg Phylogenetic Tree

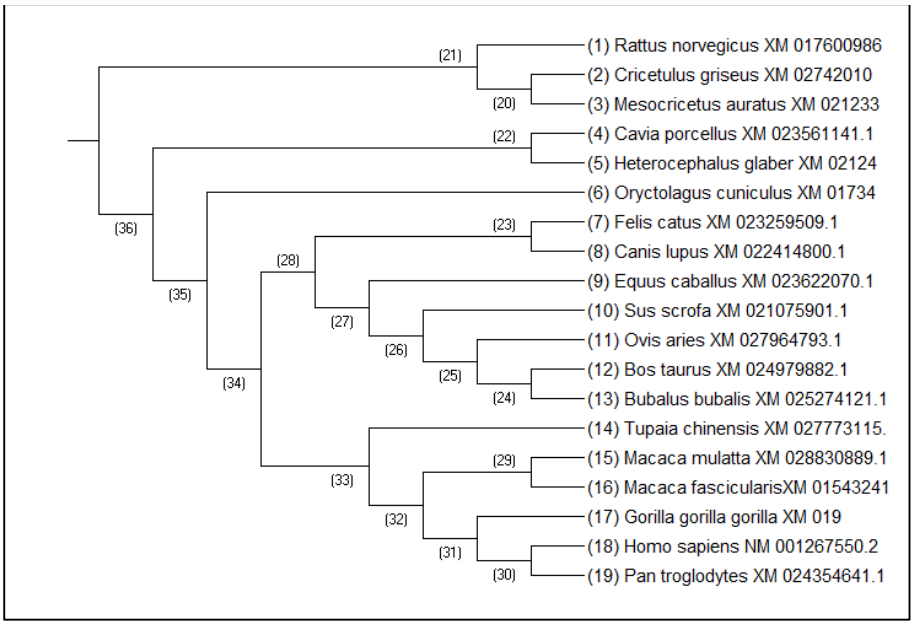


Figure 15: TN Phylogenetic Tree

References

1. Yang, Z., Nielsen, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. Journal of molecular evolution, 46, 409-418.
2. Kimura, M. (1983). The neutral theory of molecular evolution. Cambridge University Press.
3. Laura, M. (2008). Contrasts in Japanese Society and Culture in Advertisements: 1960s and 1990s (Doctoral dissertation, Gakushuin University).
4. W.H.Li, Molecular Evolution, Sinauer Associates, Sunderland,Mass, USA, 1997.
5. Koonin, E. V., Wolf, Y. I. (2010). Constraints and plasticity in genome and molecular-phenome evolution. Nature Reviews Genetics, 11(7), 487-498.
6. Lynch, M., Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. science, 290(5494), 1151-1155.
7. Hughes, A. L., Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature, 335(6186), 167-170.
8. Endo, T., Ikeo, K., Gojobori, T. (1996). Large-scale search for genes on which positive selection may operate. Molecular biology and evolution, 13(5), 685-690.
9. Tsunoyama, K., Gojobori, T. (1998). Evolution of nicotinic acetylcholine receptor subunits. Molecular biology and evolution, 15(5), 518-527.
10. Crow, J. F. (2017). An introduction to population genetics theory. Scientific Publishers.
11. Onodera, W., Asahi, T., Sawamura, N. (2019). Data for

positive selection test and co-evolutionary analysis on mammalian cereblon. Data in brief, 26, 104499.
12. Ban Y, Greenberg DA, Concepcion E, Skrabanek L, Villanueva R, Tomer Y. Amino acid substitutions in the thyroglobulin gene are associated with susceptibility to human
13. and murine autoimmune thyroid disease. Proc Natl Acad Sci U S A. 2003 Dec 9;100(25):15119-24. Epub 2003 Dec 1.
14. Targovnik, H. M., Citterio, C. E., Rivolta, C. M. (2011). Thyroglobulin gene mutations in congenital hypothyroidism. Hormone research in paediatrics, 75(5), 311-321.
15. Caron, P., Moya, C. M., Malet, D., Gutnisky, V. J., Chabardes, B.,et al. (2003). Compound heterozygous mutations in the thyroglobulin gene (1143delC and 6725G→ A [R2223H]) resulting in fetal goitrous hypothyroidism. The Journal of Clinical Endocrinology & Metabolism, 88(8), 3546-3553.
16. González-Sarmiento, R., Corral, J., Mories, M. T., Corrales, J. J., Miguel-Velado, E., et al. (2001). Monoallelic deletion in the 5'region of the thyroglobulin gene as a cause of sporadic nonendemic simple goiter. Thyroid: Official Journal of the American Thyroid Association, 11(8), 789-793.
17. Prpić, M., Franceschi, M., Romić, M., Jukić, T., Kusić, Z. (2018). Thyroglobulin as a tumor marker in differentiated thyroid cancer–clinical considerations. Acta Clinica Croatica, 57(3), 518.
18. Yang, Z., Nielsen, R. (1998). Synonymous and nonsynonymous rate variation in nuclear genes of mammals. Journal of molecular evolution, 46, 409-418.
19. Kimura, M. (1983). The neutral theory of molecular evolution. Cambridge University Press.
20. Kimura in the 1960s and 70s
21. W.H.Li, MolecularEvolution, SinauerAssociates, Sunderland, Mass, USA, 1997.
22. Koonin, E. V., Wolf, Y. I. (2010). Constraints and plasticity in genome and molecular-phenome evolution. Nature Reviews Genetics, 11(7), 487-498.
23. Lynch, M., Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. science, 290(5494), 1151-1155.
24. Hughes, A. L., Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature, 335(6186), 167-170.
25. Endo, T., Ikeo, K., & Gojobori, T. (1996). Large-scale search for genes on which positive selection may operate. Molecular biology and evolution, 13(5), 685-690.
26. Tsunoyama, K., & Gojobori, T. (1998). Evolution of nicotinic acetylcholine receptor subunits. Molecular biology and evolution, 15(5), 518-527.
27. Crow, J. F. (2017). An introduction to population genetics theory. Scientific Publishers.