Nig. J. Anim. Prod. 2018, 45(1): 9 - 17
©Nigerian Society for Animal Production

Computational Genomics of RELA gene of Pig and its protein prescription

Oluwole, O. O.
Institute of Agricultural Research and Training,
Obafemi Awolowo University, Nigeria
oluwafunmike@yahoo.co.uk +234(0)7031567164

Abstract

RELA gene is a pleiotropic transcription that is present in almost all living cells where several signal transduction events end points were initiated by many arrays of stimuli that were related to many biological processes such as inflammation, immunity, differentiation, cell growth, tumorigenesis and apoptosis in the cells. The nucleotide and amino acid sequence (AAS) of RELA gene of pig and other mammalian species were downloaded from the National Center for Biotechnology information (NCBI) data base, United State of America and Universal protein resources (Uniprot) database, United Kingdom. Data generated were analyzed by bioinformatics tools. The results showed a high variation of AAS among the mammals where only four mammals have the same AAS length (551) with pig while others were not. The RELA gene of pig shared very high percent identity with all mammals ranging from 95.5% (warthog) to 83.5% (guinea pig) and similarity percentage ranged from 85.1 to 99.5% where the warthog had the highest percentage value while the least mammal was Guinea pig. The tree showed RELA gene with two clades; the first for mammals and the second for non-mammals. The RELA gene of pig was in the same cluster with the other Suidae family, warthog and babirusa. For the protein structure obtained, only 43% can be meaningfully predicted from the 61% residues modelled at more than 90%. The domain architecture of RELA protein structure contained mainly RNA Recognition motif and region of low complexity. The domainstarting position of all the organisms ranged from 191 to 210 while the termination of all the mammals ranged from 102 to 306 amino acids.The region of low complexity starting point ranged from position 173 to 462while termination positions ranged from 184 to 474.

The evolution relationship of RELA gene of pig and other mammalian species were highly related and conserved among the mammalian species; they have high comparability and they evolved from common ancestors.

Keywords: Pig, RELA gene, protein structure, phylogenetic tree.

Introduction

The RELA gene is a gene that encodes a protein called Transcription factor P65 known as Nuclear Factor NF-Kappa-B (NF-$_k$B) P65. It is involved in formation, nuclear translocation and activation of NF-$_k$B heterodimer (Nolan *et al.,* 1991). It is essential because it is involved in all types of cellular processes such as cellular metabolism, chemotaxis etc. The major function of RELA is the phosphorylation and acetylation that are crucial for post-translational modifications that are required for activation of NF-κB. It also modulate immune responses and its activation is positively associated with multiple types of cancer in humans.

The NF-kappa-B complexes are held in the cytoplasm in an inactive state complexed with members of the NF-kappa-B inhibitor (I-kappa-B) family. In a conventional activation pathway, I-kappa-B is phosphorylated by I-kappa-B kinases (IKKs) in response to different activators, subsequently degraded thus liberating the active NF-kappa-B complex which

translocated to the nucleus. The NF-kappa-B p65-p65 complex appears to be involved in invasin-mediated activation of IL-8 expression (http:// www. uniprot .org/ uniprot/Q04206). The inhibitory effect of I-kappa-B upon NF-kappa-B the cytoplasm is exerted primarily through the interaction with p65. The p65 shows a weak DNA-binding site which could contribute directly to DNA binding in the NF-kappa-B complex. It associates with chromatin at the NF-kappa-B promoter region via association with DDX1 (http://www.genecards.org/cgi-bin/carddisp.pl?gene=RELA). It is also essential for cytokine gene expression in T-cells (Bettelli *et al*., 2005).

As a complex compound with other protein, it contains a N-terminal REL-homology domain (RHD), and also a C-terminal transactivation domain (TAD), where RHD is involved in DNA binding, dimerization and NF-κB/REL inhibitor interaction. On the other hand, TAD is responsible for interacting with the basal transcription complex (Chen and Greene, 2004). RELA and p50 is the mostly commonly found heterodimer complex among NF-κB homodimers and heterodimers, and is the functional component participating in nuclear translocation and activation of NF-κB ((http://www.genecards.org/cgi-bin/carddisp.pl?gene=RELA).).

The RELA participates in adaptive immunity and responses to invading pathogens via NF-κB activation. Mice without individual NF-κB proteins are deficient in B- and T-cell activation and proliferation, cytoline production and isotype switching (Li and Verma, 2002). He also reported that mutations in RELA are found responsible for inflammatory bowel disease as well. Therefore, this study was undertaken to evaluate the phylogenetic relationship between RELA gene of pig and other mammalian species and to predict its protein structure and its domain architecture.

Materials and methods

Retrieval of amino acids sequences

The nucleotide and amino acid sequences (AAS) of RELA gene of pig were downloaded from the National Center for Biotechnology information (NCBI) data base, United State of America and Universal protein resources (Uniprot) database, United Kingdom.

Determination of Percentage Identity and Similarity

The Identity and Similarity percentage of AAS of RELA gene of pig and other mammals were identified by conducting pairwise comparison of their AAS using two or more sequences of Basic local alignment search tool (BLAST).

Phylogenetic analysis

The phylogenetic tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Saitout and Nei, 1987). The tree with the highest log likelihood (-1182.1917) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 25 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of

171 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar *et al*, 2016).

Prediction of Protein Structure and Domain Architecture

The AAS of RELA protein was submitted into phyre2 online (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi) for protein structure prediction and analysis.The protein parameters (physical and chemical properties) of pig AAS length of 551 were obtained from Expasy Bioinformatics resource portals where Protfam tool was used to obtain the protein parameters.

Domain architecture of RELA gene was predicted using SMART software online (Tables 3). The amino acids sequence obtained from NCBI database was used to determine the domain architecture.

Results and discussions

The organisms and the accession numbers of their amino acid sequence (AAS) with their sequence length that were retrieved from NCBI and Uniprot were shown in Table 1. There was high variation of AAS among the mammals where only four mammals have the same AAS length (551) with pig while others were not as shown in Table 1. The animals that have the same length of AAS with pigs were Warthog, Cattle, Wild yak and Gibbon.AAS length is the number of bases or amino acid (AA) in the canonical sequences. The mammals with different AAS from that of pigs were not bolded. The sequence variations observed were probably by deletion or insertion of some AA sequences due to convergent evolution. These changes that occur in DNA and proteins were products of evolution (Jin, 2006). The building blocks of these biological macromolecules, nucleotide bases, and amino acids form linear sequences that determine the primary structure of the molecules (Kedasher *et al*., 1999).

Table 1: The percent identity and si milarity between the AAS of RELA gene of Pig and of other mammals

Animal	Accession number	Class of animals	Amino acid Sequence length
Pig	F2QA75	Mammalian	**551**
Warthog	F2QA76	Mammalian	**551**
Bibirus	F2QA77	Mammalian	549
Cattle	A1XG22	Mammalian	**551**
Camel	TOMEL2	Mammalian	560
Horse	F7B9X5	Mammalian	550
Cat	A0A0D6CBM2	Mammalian	555
Giant Panda	G1MG73	Mammalian	528
Wild Yak	L8IMF1	Mammalian	**551**
Human	Q2TAM5	Mammalian	337
Gorilla	G3SHV6	Mammalian	508
Gibbon	G1R1D5	Mammalian	**551**
Sheep	W5Q7L5	Mammalian	555
Dog	F1PCU1	Mammalian	542
Green monkey	A0A0D9R6K4	Mammalian	548
Giant Panda	G1MG73	Mammalian	528
Rat	Q7TQN4	Mammalian	550
Mouse	Q04207	Mammalian	549
MustelaFuro	M3YJN8	Mammalian	511
Chinese hamster	G3IDM9	Mammalian	546
Human	Q2TAM5	Mammalian	377
Chimpanzee	A2T6W4	Mammalian	364
Guinea pig	H0VZS0	Mammalian	496
Orangutan	A2T7F9	Mammalian	364
Chicken	A0A1D5P4Q4	Aves	558

The gene identity and similarity between the AAS of RELA gene Pig and of other mammals is shown in Table 2.Identity can be defined as the amount of characters which match exactly between two different sequences.The RELA gene of pig shared very high percent identity with all mammals ranging from 95.5% (warthog) to 83.5% (guinea pig) with the exception pf chicken as an outgroup that had the least (50.3%) value as shown in Table 2. The identity percentage obtained from the alignments reveals the amount of characters that match exactly between different sequences. The percentage of Identity observed from RELA gene of pig and the other mammals was due to high pattern of evolution and differentiation. This result also show that the RELA gene of all these mammals were homologous, perform the same function and have high conservation. According to Joshi and Xu (2007), the two sequences that have more than 70% of Identity have probability of more than 90% to perform the same function. This result is in corroboration with the findings of Oluwole (2017) (S100A8 gene) and that of Durosaro *et al.,* 2016 (Bovine Reprimogene).

Table 2: The percent identity and similarity between the AAS of RELA gene of Pig and other species

Animal	Identity (%)	Similarity (%)	e-value
Pig	100	100	0.00
Warthog	99.5	99.5	0.00
Babirus	98.2	98.9	0.00
Cattle	93.2	94.6	0.00
Camel	95.7	96.7	0.00
Horse	94.9	96.4	0.00
Cat	93.5	95.1	0.00
Giant Panda	90.9	92.4	0.00
Wild Yak	93.0	94.4	0.00
Human	97.7	99.1	0.00
Gibbon	92.9	94.9	0.00
Sheep	91.1	92.7	0.00
Dog	91.5	93.8	0.00
Green monkey	93.0	94.7	0.00
Giant panda	90.9	92.4	0.00
Rat	89.5	92.0	0.00
Mouse	88.9	91.7	0.00
Mustelafuro	91.8	93.5	0.00
Chinese hamster	89.5	91.1	0.00
Gorilla	86.2	88.4	0.00
Chimpanzee	91.0	92.9	0.00
Guinea pig	83.5	85.1	0.00
Bornean Orangutan	89.9	91.3	0.00
Chicken	50.3	60.5	0.00

The similarity percentage ranges from 85.1 to 99.5% where the Warthog had the highest percentage value while the least mammal was Guinea pig and the chicken as an out group had 60.5% as shown in Table 2. Similarity percentage can be defined as the degree of resemblance between two aligned sequences. It also reveals the extent to which residues is aligned. The higher Similarity percentage obtained between the

RELA gene of pig and other species implied that they all have common ancestors and homologous since the expected-value (e-value) were low for all the species, therefore they were significant (0.05) and homologous (Lesk, 2002). They can also be used to assign protein into family.

The phylogenetic result as shown in Figure 2, where phylogenetic relationship between RELA gene of pig and that of the other organisms was shown where an optimal tree with the sum of branch length of 0.742 was obtained. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) was shown next to the branches (Felsenstein, 1985). The tree showed that the RELA gene has two clades where the first clade contained RELA gene for the mammals while the second clade contained RELA gene for the non-mammals-Chicken (Aves) which was an out group.The RELA gene of pig was in the same cluster with the other Suidae family, Warthog and Babirusa. The close relationship observed between the RELA gene of pigs and RELA gene of other mammals implied high comparability and evolution from a most recent common ancestor. It also showed that they shared eleven common ancestors from the terminal nodes to the root of the tree. The dendrogram obtained revealed that the RELA genes of different animal were clearly separated from each other according to their classes as shown in the Figure 2.

The protein physical and chemical properties of mammals with the same AAS are shown in Table 3. Despite having the same AAS length of 551, their protein structure physical and chemical properties were not the same except only their stability (unstable) and hydropathicity (negative). The N–terminal of both the pig and warthog were both Asparagine, the cattle and gibbons were having Methionine while the Wild yak was Leucine. Pig and Warthog had almost the same physical and chemical properties as shown in the Table 3. From the 61% residues modelled at more than 90% only 43% can be meaningfully predicted. Therefore the overall confidence was considered to be low (<70%). The result as shown below (Figure 1) can be viewed http://www.sbg.bio.ic.ac.uk/phyre2/phyre2_output/

From this study, the domain architecture revealed that the RELA protein contained mainly RNA Recognition motif and region of low complexity as shown in Table 4.The domainstarting positions of all the organisms range from 191 to 210 with the exception of Gorilla and orangutan that started from 6. The termination (end) of all the proteins range from 102 to 306 amino acids. Jiang *et al.* (1997) reported that domain is evolutionarily more conserved than other regions of a protein and tends to evolve as units, which are gained, lost, or shuffled as one module. He also added that the identification of motifs and domains in proteins is an important aspect of the classification and functional annotation of protein sequences.

The region of low complexity start point ranged from position 173 to 462while termination positions ranged from 184 positions to 474 (Table 4).The proteins also contained region of little diversity in the sequence of amino acid composition called low complexity region. Low complexity region evolved rapidly through mitotic replication slippage and meiotic recombination event.

Table 3: RELA protein parameters of pig and other mammals that have the same AAS length

Protein Parameters	Pig	Warthog	Cattle	Wild yak	Gibbon
Carbon	2654	2657	5322	2662	2667
Hydrogen	4175	4177	8376	4193	4190
Nitrogen	745	745	1498	749	744
Oxygen	813	810	1637	819	812
Sulphur	19	19	38	18	19
Total atom	8406	8408	16871	8041	8432
Instability Index	50.21	50.29	43.62	43.96	53.95
Stability	Unstable	Unstable	Unstable	unstable	unstable
Aliphatic index	72.99	73.18	72.61	73.14	73.68
Hydropathicity	-0.451	-0.450	-0485	-0.482	-0.463
N-terminal	Asparagine	Asparagine	Methionine	Leucine	Methionine

AAS =Amino acid sequence, c=carbon, H =hydrogen, N=nitrogen, O=oxygen, S= Sulphur

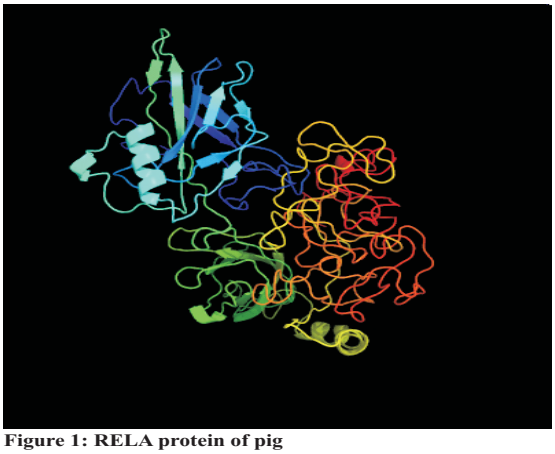

Figure 1: RELA protein of pig

Table 4: Domain architecture of RELA using SMART

Domain name	RNA Recognition motif			Low complexity	
Organism Name	start	End	e-value	start	End
Pig	191	287	2.81e-22	377	425
Warthog	191	287	2.81e-22	377	425
Babirusa	191	287	1.77e-21	276	423
Camels	200	196	2.81e-22	386	434
Horse	193	289	2.81e-22	324	346
Cat	193	289	2.81e-22	378	421
Cattle	193	289	1.73e-22	395	402
Wild yak	193	289	2.81e-22	395	402
Gabbon	193	289	2.81e-22	375	431
Sheep	193	289	2.81e-22	392	406
Dog	193	289	1.22e-22	378	405
Green monkey	210	306	2.81e-22	392	442
Giant Panda	210	306	2.81e-22	396	424
Rat	193	289	2.81e-22	377	389
Chinese hamster	194	290	2.81e-22	407	421
Guinea pig	193	289	2.81e-22	462	474
Human	193	289	2.81e-22	-	-
Gorilla	6	102	2.81e-22	187	244
Orangutan	6	102	1.7e-22	187	244
Chicken	198	294	6.55e-23	173	184

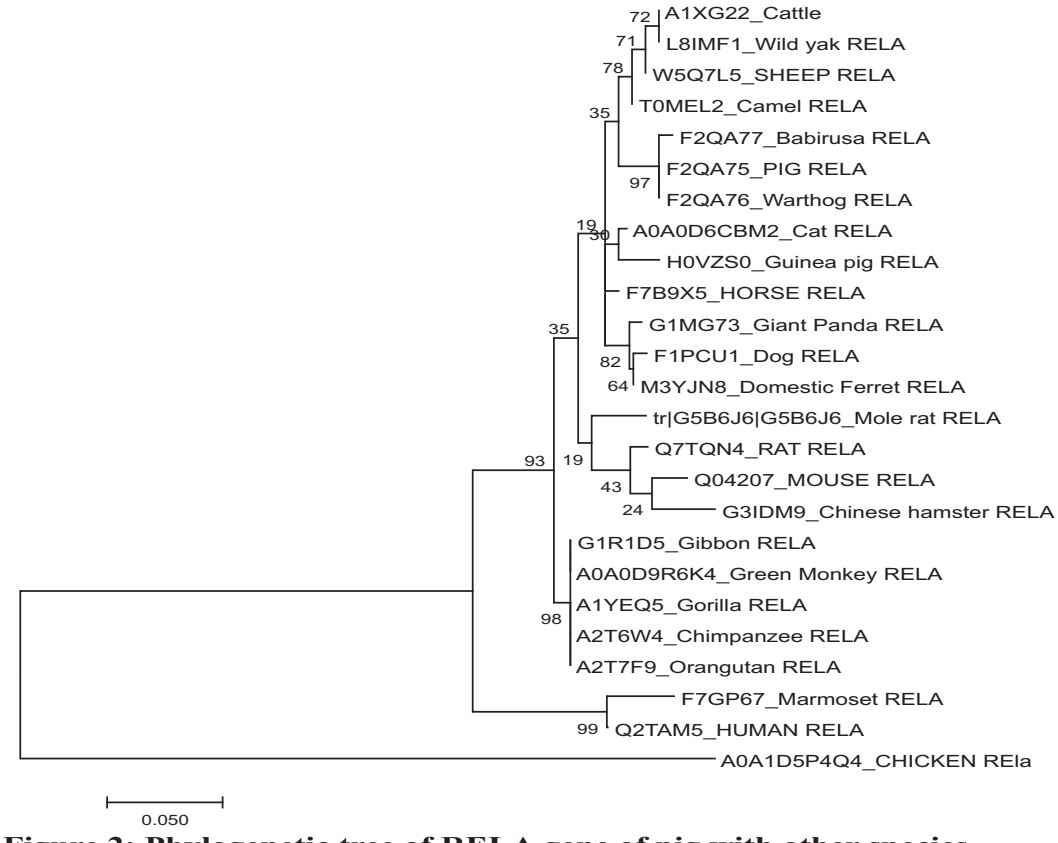

Figure 2: Phylogenetic tree of RELA gene of pig with other species

Conclusion

In conclusion, this study shows that RELA gene of pig and that of other mammalian species was highly comparable and evolved from common ancestors i.e it has high relatedness among the mammalian species studied. The high similarity and identity was observed among the RELA gene of pig and other mammalian species. The protein predicted was only 43% meaningful. The domain architecture of RELA protein structure contained mainly RNA Recognition motif and region of low complexity. The domain starting position of all the organisms ranged from 191 to 210 while the termination of all the mammals ranged from 102 to 306 amino acids. The region of low complexity starting point ranged from position 173 to 462 while termination positions ranged from 184 to 474. This study gives basic information that may be useful in the development of diseases resistance because RELA participates in adaptive immunity and responses to invading pathogens via NF-κB activation.

References

Baxevanis, A. D. and Quellete, B. F. F. 2001. Bioinformatics: A practical guide to Analysis of gene and protein. 2nd edition John Wiley and sons, Inc USA. 470pp.

Bettelli, E., Dastrange M. and Oukka, M. 20005. Foxp3 interacts with nuclear factor of activated T cells

and NF-kappa B to repress cytokine gene expression and effector functions of T helper cells. Proc. Natl. Acad. Sci. U.S.A. 102:5138-5143 [PubMed] [Europe PMC]

Chen, L. F. and Greene, W. C. 2004. "Shaping the nuclear action of NF-kappaB".Nature Reviews Molecular Cell Biology.**5** (5): 392–401. doi:10.1038/nrm1368. PMID 15122352

Durosaro, S. O., Ilori, B. M., Sanda, A. J. Onagbesan, O. M., Peters, S. O. and Ozoje, M. O. **2016.** Mutational and evolutionary Analysis of Bovine Reprimo Gene. Nigerian Journal of animal production. *The Nigerian society for animal production*. Vol 43(1), page 10-19

Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.

Fonseca, V. G., Fonseca, V. G., Rosa, J., Laizé, V., Gavaia, P. J. and Cancela, M. L. 2011. Identification of a new cartilage-specific S100-like protein up-regulated during endo/perichondral mineralization in gilthead seabream. *Gene Expression Patterns:* GEP [2011, 11(7):448-455]

GeneCards human gene database 2018. Rela gene, protein coding and Rela oncogene. Weizmann Institute of Science. All Rights Reserved. v4.6.1 Build 19

Jiang, W., Hou Y. and Inouye, M. 1997. CspA, the major cold-shock protein of Escherichia coli, is an RNA chaperone. *Journal of Biology and Chemistry*. 272:196–202.

Jin, X. 2006. Essential Bioinformatics.Cambridge University Press. 1:3-27.

Josh, T. and Xu D. 2007. Quantitative assessment of relationship between sequence similarity and function similarity. Biomed Central Genomics. 8;222-234.

Kedersha, N. L., Gupta, M., Li, W., Miller, I. and Anderson, P. 1999. RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2α to the assembly of mammalian stress granules. *Journal of Cellular Biology*.;147: 1431–1442.

Kumar, S., Stecher, G. and Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.Molecular Biology and Evolution 33:1870-1874.

Lehninger, A. L., Nelson, L. D. and Cox, M. M. 1993. Principe of Biochemistry 2nd edition. Worth publisher Inc. USA. 432pp

Lesk, A. M. 2002. Introduction to Bioinformatics 2nd Edition. Oxford University press, Oxford, 255pp.

Li, Q. and Verma, I. M. 2002."NF-kappaB regulation in the immune system".Nature Reviews. Immunology.**2** (10): 725–34. doi:10.1038/nri910. PMID 12360211

Nolan, G. P., Ghosh, S., Liou, H. C., Tempst, P., Baltimore, D. 1991. "DNA binding and I kappa B inhibition of the cloned p65 subunit of NF-kappa B, a rel-related polypeptide".Cell.**64** (5): 961–9. doi:10.1016/0092-

8674(91)90320-X. PMID 2001591

Saitou, N. and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425.*Uniprot 2017. Protein database*. http://www.uniprot.org/uniprot/C3S7K5

Zuckerkandl E. and Pauling L. 1965.Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.

Uniprot knowledgebase (UniprotKB). http://www.uniprot.org/uniprot/Q04206

Received: 14th November, 2017
Accepted: 18th February, 2018