International Journal of Advanced Research in Biological Sciences
ISSN: 2348-8069
www.ijarbs.com
(A Peer Reviewed, Referred, Indexed and Open Access Journal)
DOI: 10.22192/ijarbs Coden: IJARQG (USA) Volume 9, Issue 5 -2022

Research Article

Special Issue on Potential Applications of Bioinformatics in Biological Sciences - PABBS 2022

DOI: http://dx.doi.org/10.22192/ijarbs.2022.09.05.01.001

In silico analysis of AMMECR1 protein-protein and its phylogenetic interaction

Jeyabaskar Suganya[1*], Nishandhini.M[2], Mark Antony. J [3], Rajesh Kumar.G[4], Mahendran Radha[5]

[1*]Assistant Professor, [2]Assistant Professor, [3]Student, [4]Assistant Professor, [5]Professor.
[1,2,3.5]Department of Bioinformatics, School of Life Sciences, VISTAS, Pallavaram, Chennai-600117, Tamil Nadu, India.
[4]Assistant Professor, Department of Pharmacology, Govt. Kilpauk Medical College, Chennai, Tamil Nadu, India.

Abstract

Protein-protein interactions are particularly essential among these relationships because of their diversity, specificity, and adaptability. Protein-protein interactions (PPIs) are important for understanding protein function, disease occurrence, and therapeutic development. Phylogenetics and Evolutionary Biology is a branch of biology that studies evolutionary relationships between groups of species as well as computational modelling approaches for studying biological, behavioural, and social systems.The AMMECR1 gene was found to be the main cause of Alport syndrome in humans in a recent study, but the actual function of the gene is still anonymous. As a result, determining the function of either genes or proteins was critical, as it would aid in the development of a novel medication for the protein AMMECR1. The protein-protein interaction and phylogenetic analysis of the AMMECR1 protein were used in this study to anticipate the activities of the protein that causes Alport syndrome. The function of the AMMECR1 protein was predicted using the results of the current *in silico* protein analysis, which will be useful for further investigation.

Keywords: AMMECR1 protein, Alport syndrome, Protein-protein Interaction, Phylogenetic analysis.

Introduction

The AMMECR1 gene is found on chromosome Xq22.3 and encodes the AMMECR1 protein [Zhou HM et al., 2015]. In humans, the symptoms of contiguous gene deletion syndrome include A: Alport disorder, M: Mental retardation, M: Midface impairment, and E: Elliptocytosis, as well as generalised hypoplasia and heart issues [Cai C et al., 2019].The syndrome is caused mainly byanomissionofXq22.3 Chromosome, which encompasses with several genes. For example: AMMECR1 encodes anunknown function for nuclear-localized protein. On analysing the C-terminal domain(residues 122 to 333) of protein AMMECR1 was found to have a extremely conserved region, with homologues in bacteria, archaea, and eukaryotes [Andreoletti G*et al.*, 2017].

The excessive tenacity of the AMMECR1 domain recommends an essential biological function i.e Central dogma of life. The LRGCIG - 6-amino acid motif has been revealed to be functionally significant during the evolutionary analysis in the AMMECR1 domain. The AMMECR1 domain is broken into two smaller subdomains. Five alpha-helices and five beta-strands make up the big subdomain, which encompasses both the N- and C-terminal sections. An antiparallel beta-sheet is formed by these five beta-strands [Moyses-Oliveira M*et al.*, 2018].

The small subdomain which also forms an antiparallel β sheet is made up of 4 α helices and 3βstrands. The conserved motif (LRGCIG) is positioned at βin N-terminal loop structure and itsside chains pointed towards the crossing point of 2 subdomains. Only 2 loops connected 2 subdomains and found that there is minute interaction between domains. As a result, these subdomains might migrate vigorously when substrate enters the domains or cleft. The cleft's enormous size indicates a large substrate such as a nucleic acid or protein. Negatively charged nucleic acids on the other hand are unable to travel across the gap because the inner side is devoid of positively charged residues [Basel-Vanagaite L*et al.*, 2017].

Interaction evidence from a variety of sources must be considered when building a functional association network for an organism's proteins; the applicability of these sources varies depending on the proteins in question, their biological roles, and the extent to which they have been studied experimentally [Andreoletti G *et al.*, 2017]. Data integration across different evidence sources has been shown to improve overall network quality and it is also thought to be required given the various modalities by which proteins can be connected. There are three types of sources of interaction evidence: I prior information gleaned from curated route databases or scientific papers, (ii) computational interaction predictions gleaned from a range of algorithms, and (iii) direct lab trials gleaned from a variety of assays in both low- and high-throughput [Damian Szklarczyk *et al.*, 2021].

One of several online sites dedicated to organism-wide protein association networks is the STRING database [Huang JK *et al.*, 2018]. The FunCoupGeneMANIA, HumanBase/GIANT, IMP, IID, ConsensusPathDB and HumanNetfields are all frequently used as resources in String database [Ogris C*et al.*, 2018]. These resources differ in the types of interaction evidence they include, the organisms they cover, and the online interface elements [Warde-Farley D*et al.*, 2010]. STRING attempts to prioritise coverage, completeness of evidence sources and usability [Greene CS *et al.*, 2015]. It allows users to log in and save their searches, as well as provide online viewers to aid in the examination of the supporting material for protein-protein interaction [Wong AK *et al.*, 2015].

STRING is a database of protein-protein interactions that is both known and predicted [Kotlyar M *et al.*, 2019]. The interactions arise from computer prediction, knowledge transfer across species, and interactions gathered from other (primary) databases [Kamburov A *et al.*, 2013], and they include both direct (physical) and

indirect (functional) correlations [Hwang S*et al.*, 2019]. Currently, the STRING database contains 24'584'628 proteins from 5'090 species [Doncheva NT *et al.*, 2019]. STRING 10 versions now supports5090 species, which is more than double the previous version [Szklarczyk D *et al.*,2015]. The most significant new feature is the ability to the upload complete datasets of genome as input in the database which the allow users to analyse subclasses as interaction complexes and perform gene-set enhancement investigation across the database [Franceschini A *et al.*, 2013]. STRING also uses the well-known taxonomy systems like GO and KEGG for enrichment analysis, but it also offers innovative classification systems based on high-throughput text mining [Von Mering C *et al.*, 2005].

MEGA (Molecular Evolutionary Genetics Analysis) 5.03 software has evolved throughout time to satisfy the growing demand for advanced evolutionary analysis to uncover organismal and genome evolutionary patterns and processes [Koichiro Tamura *et al.*, 2021]. It was first launched in 1993 as an interactive interface for statistical molecular evolution methods on the Microsoft Disk Operating System [Caspermeyer J 2018.]. MEGA's scope and utility have evolved over the years as additional methodologies, tools, and interfaces have been added, resulting in current integrated software for comparative sequence analysis [Claramunt S *et al.*, 2015]. For molecular phylogenetic analysis, MEGA initially included distance-based and maximum parsimony approaches [Hipsley CA *et al.*, 2014].

MEGA's scope was expanded by adding data gathering and integration of major methodologies for aligning sequences [Kumar S *et al.*, 2018].For molecular evolutionary analyses, maximum likelihood (ML) and Bayesian approaches were included later [Stecher G *et al.*, 2020]. MEGA now includes tools for finding the best-fit substitution model(s), calculating evolutionary distances and divergence periods, reconstructing phylogenies, predicting ancestral sequences, detecting selection, and diagnosing disease mutations [Patel R *et al.*, 2018].

Materials and Methods

Method for protein-protein interaction:

The AMMECR1 protein's interactions with other proteins were identified using the STRING database (http://string-db.org/). The AMMECR1 protein of Alport syndromewas employed as a query, and the results were analysed.The STRING database seeks to bring together all known and projected protein interactions, including both physical and functional interactions. STRING accomplishes this by gathering and scoring evidence from a variety of sources, including I automated text mining of scientific literature, (ii) databases of interaction experiments and annotated complexes/pathways, (iii) computational interaction predictions based on co-expression and conserved genomic context, and (iv) systematic transfers of interaction evidence from one organism to another. STRING aspires for broad coverage; the resource's next version 11.5 will include around 14 000 creatures. We also show how to utilise STRING to query genome-wide, experimental data, including how to find enriched functionality and potential biases in the user's query data automatically. STRING is a useful online resource available at https://string-db.org/.

Method for Phylogenetic Tree Reconstruction:

The amino acid sequences retrived from the string database were aligned and a phylogenetic tree was constructed using MEGA 5.03. We saved and analysed a phylogenetic tree. The Molecular Evolutionary Genetics Analysis (MEGA) software is a desktop tool for comparing homologous gene sequences from various species or multigene families, with a focus on inferring evolutionary relationships and patterns of DNA and protein evolution. This is the most recent stable release with bug fixes. This application includes several methods for measuring evolutionary distances from nucleotide and amino acid sequence data, three phylogenetic inference methods (UPGMA, neighborjoining, and maximum parsimony), and two statistical tests for

topological differences. Biologists can use this integrated workbench to mine data from the web, align sequences, perform phylogenetic analyses, test evolutionary hypotheses, and create publication-quality displays and descriptions.

Results and Discussion

The interaction of AMMECR1 protein with other proteins in *Homasapaiens* was found using STRING database. The AMMECR1 protein sequence of *Homasapaiens* was used as a query. The results showed the protein interaction network with 11 different proteins that includes (Figure 1):

- HR - Lysine-specific demethylase hairless with the score of 0.930
- THOC2 - THO complex subunit 2with the score of 0.823
- EMC2 - ER membrane protein complex subunit 2 with the score of 0.819
- TMEM164 - Transmembrane protein 164 with the score of 0.755
- RGAG1 - Retrotransposon gag domain containing 1 with the score of 0.747
- NXT2 - Nuclear transport factor 2 like export factor 2 with the score of 0.734
- COL4A5 - Collagen alpha-5(IV) chain with the score of 0.733
- ACSL4 - Long-chain-fatty-acid--CoA ligase 4 with the score of 0.689
- KCNE1L - Potassium voltage-gated channel isk-related subfamily e member 1-like with the score of 0.671
- GUCY2F - Guanylate Cyclase 2F with the score of 0.644

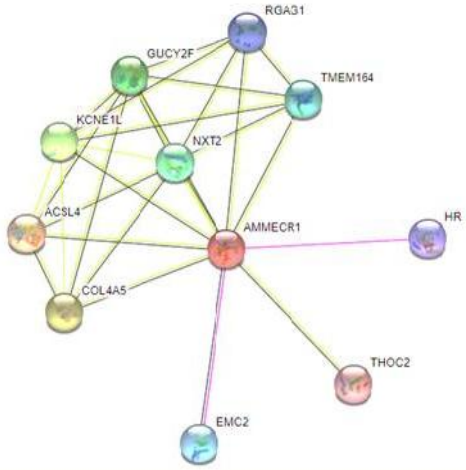
Table 1: Predicted proteins for AMMECR1 along with its functions using String database

S.No	Protein Name	Function
1.	AMMECR1	AMME syndrome applicant gene 1 protein; chromosomal region gene 1 for Alport syndrome, mental retardation, midface hypoplasia, and elliptocytosis
2.	HR	Histone demethylase that precisely demethylates both mono- and dimethylated 'Lys-9' of histone H3. Hair biology (through collagen targeting), brain activity, and cell cycle may all be controlled by this transcription regulator.
3.	THOC2	It is required for the efficient export of polyadenylated RNA and spliced mRNA. The THO subcomplex of the TREX complex is thought to link mRNA transcription, processing, and nuclear export by binding with spliced mRNA rather than unspliced pre-mRNA.
4.	EMC2	It contains a tetratricopeptide repeat domain
5	TMEM164	It is a member of the TMEM164 family.
6.	RGAG1	It contains Retrotransposon gag domain containing 1
7.	NXT2	Protein export regulator for NES-containing proteins and it's also involved in mRNA nuclear export.
8.	COL4A5	Type IV collagen is the most important structural component of glomerular basement membranes (GBM), which forms a 'chicken-wire' meshwork with laminins, proteoglycans, and entactin/nidogen.
9.	ACSL4	Long-chain fatty acids are activated for both cellular lipid production and breakdown via beta-oxidation. Arachidonate and eicosapentaenoate are the preferred substrates.

10.	KCNE1L	KCNQ1 is a repolarizing cardiac potassium ion channel that functions as an inhibitory beta-subunit.
11.	GUCY2F	It's likely that it has a specialised function in photoreceptor rods and/or cones. Guanylate cyclase receptors are enzymes involved in the resynthesis of cGMP, which is essential for the return of the dark state after phototransduction.

Figure 2: Sequences Alignment for 11 protein sequences using MEGA software

Figure 1: The interactions of AMMECR1 using String database



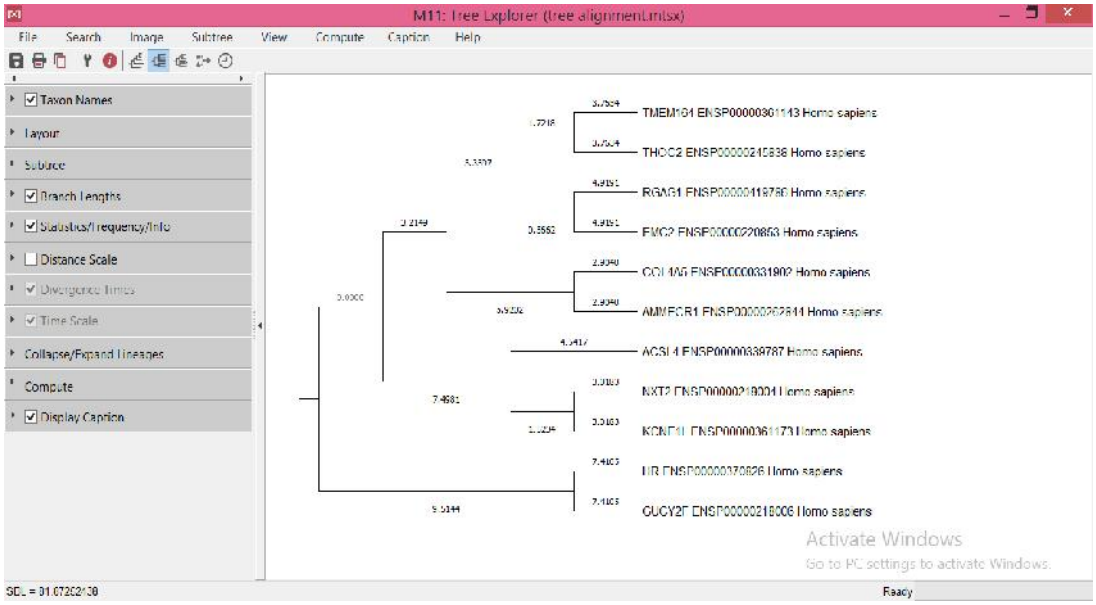
Your Input:

		Neighborhood	Gene Fusion	Cooccurence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
AMMECR1	AMME syndrome candidate gene 1 protein; Alport syndrome, mental retardation, midface hypoplasia and elliptocytosis chromosomal region gene 1 (333 aa)									

Predicted Functional Partners:

ACSL4	Long-chain-fatty-acid–CoA ligase 4; Activation of long-chain fatty acids for both synthesis of cellular lipids, and degradation ...							•		0.930
COL4A5	Collagen alpha-5(IV) chain; Type IV collagen is the major structural component of glomerular basement membranes (GBM), ...							•		0.823
KCNE1L	Potassium voltage-gated channel isk-related subfamily e member 1-like; Potassium voltage-gated channel subfamily E regul...							•		0.819
GUCY2F	Retinal guanylyl cyclase 2; Probably plays a specific functional role in the rods and/or cones of photoreceptors. It may be th...							•		0.755
NXT2	Nuclear transport factor 2 like export factor 2; NTF2-related export protein 2; Regulator of protein export for NES-containing ...							•		0.747
TMEM164	Transmembrane protein 164; Belongs to the TMEM164 family							•		0.734
EMC2	ER membrane protein complex subunit 2; Tetratricopeptide repeat domain containing					•				0.733
RGAG1	Retrotransposon Gag-like protein 9; Retrotransposon gag domain containing 1							•		0.689
HR	Hr, lysine demethylase and nuclear receptor corepressor; Lysine-specific demethylase hairless; Histone demethylase that sp...					•				0.671
THOC2	THO complex subunit 2; Required for efficient export of polyadenylated RNA and spliced mRNA. Acts as component of the ...			•						0.644

Figure 3: Phylogenetic treegenetrated by MEGA 5.03



The multiple sequence alignment file was used for the phylogenetic analysis by MEGA 5.03. From the multiple sequence alignment of AMMECR1, it is clear that the sequences analyzed in 11 protein sequences share the strong similarity among themselves. On analysing the phylogenetic tree based upon the protein sequences of *Homo sapiens* demonstrate the presence of 5 clusters among the 11 different proteins. The phylogenetic analysis was carried out by using distance method and clustering by UPGMA (Unweighted Pair Group Method with Arithmetic Mean), A tree of AMMECR1 drawn by using MEGA 5.03 reveals the formation of an outgroup comprising of HR, THOC2,EMC2,TMEM164, RGAG1, NXT2, COL4A5, ACSL4, KCNE1L,GUCY2F protein sequence. The sequences of AMMECR1and COL4A5 show the considerable similarity between each other with the distance of 2.9048. The remaining 9 sequences of *Homo sapiens* from different protein form 4 groups.

- First group comprises of THOC2 and TMEM164 with the distance of 3.70634.
- Second group comprise of RGAG1 and EMC2 with the distance of 4.9191.
- Third group comprise of ACSL4, NXT2 and KCNE1L with the distance of 4.5417 & 3.0183.
- Fourth group comprise of HR and GUCY2F with the distance of 7.4105

Conclusion

The study showed the protein-protein and evolutionary relationship of AMMECR1of *Homo sapiens* which plays a role in Alport syndrome. The string database showed that AMMECR1protein interacts with 11 different proteins of *Homo sapiens* that includes HR, THOC2,EMC2,TMEM164, RGAG1, NXT2, COL4A5, ACSL4, KCNE1L and GUCY2F. The maximum protein protein interaction was observed in COL4A5 protein when compared with other protein. The evolutionary relationship shows that AMMECR1of *Homo sapiens* is highly conserved across all 11 proteins of *Homo sapiens* and that the AMMECR1 is closest to COL4A5with good evolutionary distance of 2.9048. From the results of present *in silico* protein analysis predicted that the function of AMMECR1 protein can be more or less same as COL4A5.

Acknowledgments

We acknowledge Vels Institute of Science, Technology and Advanced Studies (VISTAS) for providing us with required infrastructure and support system needed.

Conflict of interest:

The authors declare they have no competing interests.

References

Andreoletti G, Seaby EG, Dewing JM., O Kelly I, Lachlan K, Gilbert RD and Ennis S. 2017. AMMECR1: a single point mutation causes developmental delay, midface hypoplasia and elliptocytosis. Journal of medical genetics. 54(4); 269-277.

Andreolett G,Seaby EG, Dewing JM, Kelly I, Lachlan K, Gilbert RD. 2017. AMMECR1: A Single Point Mutation Causes Developmental Delay, Midface Hypoplasia and Elliptocytosis. J Med Genet. 54: 269-277.

Basel-Vanagaite, L, Pillar N, Isakov O, Smirin-Yosef, P, Lagovsky, I, Orenstein N, 2017.X-Linked Elliptocytosis with Impaired Growth Is Related to Mutated AMMECR1. Gene, 606:47-52.

Cai C, Xu M, Su Y and Zhou H. 2019. Phosphorylation of AMMECR1 at Serine16 Is Not Essential for Its Nuclear Localization. Journal of Biosciences and Medicines. 7:146-153.

Caspermeyer J. 2018. MEGA software celebrates silver anniversary. MolBiolEvol. 35(6): 1558-1560.

Claramunt S andCracraft J.2015. A new time tree reveals Earth history's imprint on the evolution of modern birds. Sci Adv. 1(11):e1501005.

Damian Szklarczyk, Annika L Gable, Katerina C Nastou, David Lyon, Rebecca Kirsch, SampoPyysalo, Christian von Mering.2021.The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research. 49(D1): D605-D612.

Doncheva NT, Morris JH, Gorodkin J, Jensen LJ.2019. CytoscapeStringApp: network analysis and visualization of proteomics data. J Proteome Res. 18: 623-632.

Drysdale R, Cook CE, Petryszak R, Baillie-Gerritsen V, Barlow M.,Gasteiger E, Gruhl F, Haas J, Lanfear J, Lopez R.2020.The ELIXIR Core Data Resources: fundamental infrastructure for the life sciences. Bioinformatics. 36: 2636-2642.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, Von Mering C.2013.STRING v9.1: protein-protein interaction networks with increased coverage and integration. Nucleic Acids Res. 41:D808-D815.

Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, Zhang R, Hartmann BM, Zaslavsky E, Sealfon SC.2015. Understanding multicellular function and disease with human tissue-specific networks. Nat. Genet. 47: 569-576.

Hipsley CA and Muller J. 2014. Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. Front Genet. 5:138.

Huang JK, Carlin DE, Yu MK, Zhang W, Kreisberg JF,2018; Tamayo P, Ideker T. Systematic evaluation of molecular networks for discovery of disease genes. Cell Syst. 6: 484-495.

Hwang S, Kim CY, Yang S, Kim E, Hart T, Marcotte EM, Lee I. 2019.HumanNet v2: human gene networks for disease research. Nucleic Acids Res. 47: D573-D580.

Kamburov A, Stelzl U, Lehrach H, Herwig R.2013 The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. 41: D793-D800.

Koichiro Tamura, Glen Stecher, Sudhir Kumar, 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution. 38(7):3022-3027.

Kotlyar M, Pastrello C, Malik Z, Jurisica I. 2019. IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. Nucleic Acids Res. 47:D581-D589.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular Evolutionary Genetics Analysis across computing platforms. MolBiolEvol. 35(6): 1547-1549.

Moyses-Oliveira M, Giannuzzi G, Fish RJ, Rosenfeld JA, Petit F, Soares MF.2018 Inactivation of AMMECR1 Is Associated with Growth, Bone and Heart Alterations. Hum Mutat. 39. 281-291.

Ogris C, Guala D, Sonnhammer ELL.2018.FunCoup 4: new species, data and visualization. Nucleic Acids Res. 46: D601-D607.

Patel R and Kumar S. 2019. On estimating evolutionary probabilities of population variants. BMC Evol Biol. 19(1): 133.

Stecher G, Tamura K and Kumar S. 2020. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. MolBiolEvol. 37(4): 1237-1239.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP.2015. STRING v10: Protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43: D447-D452.

Von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen M.A, Bork P.2005. STRING: known and predicted protein-protein associations integrated and transferred across organisms. Nucleic Acids Res. 33: D433-D437.

Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT. 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 38: W214-W220.

Wong AK, Krishnan A, Yao V, Tadych A, Troyanskaya OG. 2015. IMP 2.0: A multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks. Nucleic Acids Res. 43: W128-W133.

Zhou HM, Cai C, Xu M and Li G. 2015. Research advances of AMMECR1. Biophysics. 3:1-6.

Access this Article in Online	
	Website: www.ijarbs.com
Quick Response Code	Subject: Proteomics
DOI: 10.22192/ijarbs.2022.09.05.01.001	

How to cite this article:
Jeyabaskar Suganya, Nishandhini.M, Mark Antony. J, Rajesh Kumar.G, Mahendran Radha. (2022). *In silico* analysis of AMMECR1 protein-protein and its phylogenetic interaction. Int. J. Adv. Res. Biol. Sci. 9(5): Special Issue 1: 1-11.
DOI: **DOI:** http://dx.doi.org/10.22192/ijarbs.2022.09.05.01.001