International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST218665

In-Silico Screening of Prime PCOS Biomarkers for the Identification of a Potential Model Organism

Shruti Shastry*[1], Soumyashree Ghosh[1], Ruqayya Manasawala[1]
*[1]Student, Department of Biotechnology, Jai Hind College Autonomous, Mumbai, Maharashtra, India
[1]Student, Department of Biotechnology, Jai Hind College Autonomous, Mumbai, Maharashtra, India
[1]Assistant Professor, Department of Biotechnology, Jai Hind College Autonomous, Mumbai, Maharashtra, India

Article Info
Volume 8, Issue 6
Page Number : 448-454

Publication Issue
November-December-2021

Article History
Accepted : 15 Dec 2021
Published : 24 Dec 2021

ABSTRACT

Polycystic ovarian syndrome (PCOS) is a multigenic endocrine disorder observed in women of reproductive age. Although the condition is characterized by the presence of polycystic ovaries and excess production of androgens, the exact aetiology has not been well deciphered due to the unavailability of a suitable model organism. Defects in the two prime biomarkers namely CYP11A and CYP19A1, have been found to play a role in disease progression. The objective of this study was to carry out an *in-silico* assessment of these two genes to identify a potential model organism for the efficacious study of PCOS. Bioinformatics tools such as BLAST and EMBOSS were used for local and global alignment respectively, to find sequence homology and thereby, establish a model organism. Sequence comparison was followed by phylogenetic analysis and secondary structure prediction of the enzymes encoded by the respective genes. Our *in-silico* study revealed *Gorilla gorilla* to be an ideal candidate for the study of PCOS owing to its high sequence and structural similarities with the human gene counterparts. Future prospects of the research include *in-vitro* analysis of the biomarkers on *Gorilla gorilla* ovarian theca cell line to pave the way for therapy.

Keywords: PCOS, model organism, in-silico, hyperandrogenism

I. INTRODUCTION

Polycystic Ovarian Syndrome (PCOS) is customarily characterized by long-standing anovulation, the presence of multiple ovarian cysts and excess production of androgens. However, the exact aetiology of this condition has not been well deciphered[1]. Clinically, this condition manifests itself in 6 to 20% of reproductive-aged women[2]. The condition is characterised by irregularities in the menstrual cycle, hirsutism, temporal alopecia, acne and elevated ratios of Luteinizing hormone (LH) and Follicle stimulating hormone (FSH)[3]. Neoteric research considers PCOS to be a condition affecting the adolescent age groups as well as adult women. As stated by the Rotterdam Diagnostic Criteria, prevalence rates lie in the range of 3 to 26%[4]. In PCOS, the maturation of the ovarian follicle is arrested at an early antral stage resulting in the accumulation of small antral follicles, giving a characteristic polycystic morphology[5]. Several studies have drawn an interrelation between the

development of this condition and a variety of environmental factors responsible for the gradual decline in insulin sensitivity. The expression of these factors is primarily modified by lifestyle changes such as excess body weight and increased intake of saturated fats present in processed and ready to eat fast foods. The development of this multifaceted disorder could be attributed to various factors stretching from unrestrained steroidogenesis, excessive oxidative stress and the susceptibility of the genetic compartments[4].

Various genetic and epigenetic variations influence its autosomal dominant inheritance pattern[6]. Two prime contenders playing an active role in the development of hyperandrogenism are the cytochrome P450 family 11 subfamily A member (CYP11A) and the cytochrome P450 family 19 subfamily A polypeptide 1 (CYP19A1) genes. CYP11A actively participates in the initial steps of androgenesis by encoding for the enzyme supervising the cleavage of the side chain of cholesterol[6]. A polymorphism in this gene is distinguished by repeats at the CYP11A promoter region. This genotype has been strongly correlated with PCOS in 20 pedigrees that express the disease. Another key enzyme involved in Oestradiol and Estrone synthesis is encoded by the CYP19A1 gene. Polymorphisms in this gene modifies the action of Aromatase enzyme related to steroidogenesis and follicular maturation[6]. This Aromatase coding gene, predominantly expressed in the ovarian cells, is regulated in a very tissue-specific manner through ten promoters that are alternatively used in different cell types[7]. Studies have reported that CYP19A1 gene expression and the consequent oestradiol production is significantly reduced in preovulatory PCOS follicles in relation to healthy controls[6]. Even though common model organisms such as *Ovis aries* or *Macaca mulatta* are available, not much has been understood regarding the disease progression or its potential therapeutics pointing towards the fact that these model organisms may not indeed be most suitable for the study of genes involved in PCOS[6].

The principal objective of this study was to carry out a comprehensive *in-silico* screening of the two prime biomarkers of PCOS, namely CYP11A and CYP19A1, in an effort to identify potential model organisms for study with maximum homology. A cohort of bioinformatics tools were employed for sequence retrieval and structure analysis.

II. METHODS AND MATERIAL

A. Nucleotide Sequence Analysis

1. Sequence Retrieval: Homo sapiens cytochrome P450 family 11 subfamily A member (CYP11A1), transcript variant 1, mRNA; nuclear gene for mitochondrial product (NM_000781.3) and Homo sapiens cytochrome P450 family 19 subfamily A polypeptide 1 (CYP19A1) mRNA, complete CDS (DQ118405) were retrieved from the National Centre of Biotechnology Information (NCBI) nucleotide database.
2. Local Pairwise Sequence Alignment: Basic Local Alignment Search Tool (BLAST) version 1.17 available on NCBI was used to compare the obtained nucleotide sequences with database sequences. Nucleotide sequences of five organisms with highest query coverage and percent identity were retrieved from the NCBI nucleotide database. Local pairwise alignment for CYP11A & CYP19A1 with each of the selected 5 organisms was also performed on EMBOSS with Water algorithm.
3. Global Pairwise Sequence Alignment: Global pairwise alignment was performed for CYP11A (NM_000781.3) and CYP19A1 (DQ118405) with *Pan paniscus, Gorilla gorilla, Ovis aries* and *Macaca mulatta* nucleotide sequences for the respective genes on EMBOSS with Stretcher algorithm. Global pairwise alignment was also

performed for CYP11A (NM_000781.3) and CYP19A1 (DQ118405) with *Pan paniscus, Gorilla gorilla, Ovis aries* and *Macaca mulatta nucleotide* sequences for the respective genes on BLAST with Needleman and Wunsch algorithm.

B. **Evolutionary Study**

A Phylogenetic tree was constructed for CYP11A (NM_000781.3) and CYP19A1 (DQ118405) nucleotide sequences on BLAST available on NCBI using the Neighbour-Joining method with a significant difference of 0.05.

C. **Protein Analysis**

Amino acid sequences for Homo sapiens CYP11A1 (P05108) and CYP19A1 (P11511) in FASTA format were retrieved from UniProt. Physical and chemical parameters of the protein encoded by *Homo sapiens*, *Gorilla gorilla, Ovis aries* and *Macaca mulatta* genes were obtained using Prot-Param. Secondary structure of the protein encoded by *Homo sapiens*, *Gorilla gorilla, Ovis aries* and *Macaca mulatta* genes were also predicted using the Chou-Fasman algorithm.

III.RESULTS AND DISCUSSION

A. **Nucleotide Sequence Analysis**

1. Local Pairwise Sequence Alignment: Mega-BLAST was performed to identify organisms showing highest similarity to human genes CYP11A and CYP19A1. BLAST performs local alignment, using the K-tuple method of alignment. The BLAST tool compares the query sequence with the database sequence in a pairwise manner. 5 organisms showing the highest percent identity of >90 % were selected for further *in-silico* analysis. The top five organisms showing highest query coverage and percent identity to the human gene CYP11A and CYP19A1 are shown in Table 1. EMBOSS by EMBL used to perform local pairwise alignment between the human gene sequence and gene sequences of the top 5 model organisms, corroborated with the results given by BLAST.

TABLE 1
RESULTS GIVEN BY BLAST AND EMBOSS FOR THE GENES OF INTEREST

CYP11A				CYP19A1			
Organisms	Accession No	BLAST (%)	EMBOSS (%)	Organisms	Accession No.	BLAST (%)	EMBOSS (%)
Pan paniscus	XM_003811110.4	99.24	99.2	*Pan paniscus*	XM_024925627.2	99.46	99.5
Gorilla gorilla	XM_004056502.2	99.24	99.2	*Pan troglodytes*	XM_009429125.3	99.69	96.2
Pongo abelii	XM_024232611.1	98.26	98.3	*Gorilla gorilla*	XM_031002144.1	99.56	97.8
Colobus angolensis	XM_011932917.1	97.17	97.2	*Pongo abelii*	XM_024232323.1	98.69	95.3
Macaca nemestrina	XM_011756963.1	96.95	97.0	*Hylobates moloch*	XM_032174255.1	98.26	94.5
Macaca mulatta	XM_015142701		96.4	*Macaca mulatta*	XM_015142241.1		97.5
Ovis aries	NM_001093789.1		78.6				

The nucleotide sequences of conventionally-used model organisms *Ovis aries* and *Macaca mulatta* were also subjected to sequence alignment. From local sequence alignment studies the organisms *Pan paniscus* and *Gorilla gorilla* were narrowed down for further analysis.

2. Global Pairwise Sequence Alignment: Global alignment enables an end-to-end alignment of the genes. EMBOSS Stretcher and BLAST Needleman were used to perform global alignment and the results are shown in Table 2.

TABLE 2. RESULTS OF EMBOSS (STRETCHER) AND BLAST (NEEDLEMAN) OF *HOMO SAPIEN* CYP 11A & CYP19A1 WITH *PAN PANISCUS, GORILLA GORILLA, OVIS ARIES* AND *MACACA MULATTA* GENES.

CYP11A	Organisms	*Pan paniscus*	*Gorilla gorilla*	*Macaca mulatta*	*Ovis aries*
	Percent Identity from EMBOSS Stretcher (%)	67.4	89.7	87.8	75.7
	Percent Identity from BLAST Needle (%)	67.0	90.0	88.0	76.0
CYP19A1	Organisms	*Pan paniscus*	*Gorilla gorilla*	*Macaca mulatta*	
	Percent Identity from EMBOSS Stretcher (%)	37.3	54.6	54.1	
	Percent Identity from BLAST Needle (%)	37.0	55.0	54	

The nucleotide sequences of conventionally-used model organisms *Ovis aries* and *Macaca mulatta* were also subjected to global alignment with the human genes of interest. The results given by EMBOSS Stretcher were verified using BLAST with Needleman algorithm to obtain concordant results. *Gorilla gorilla* was chosen to be the most suitable contender as model organism based on its higher percent identity to both the human genes. *Gorilla gorilla* also fared better than *Macaca mulatta* and *Ovis aries* in sequence alignment studies.

B. Evolutionary Study

A phylogenetic tree was constructed after carrying out Mega-BLAST with sequences highly similar to the query. The phylogenetic tree was constructed using the Neighbour-Joining method, with a significant difference of 0.05, displaying sequences with least evolutionary distance.

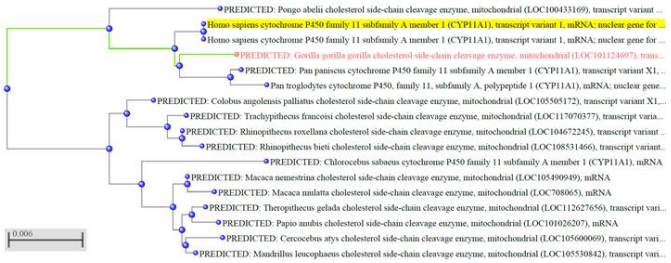

Figure 1: Phylogenetic tree showing evolutionary relationship between Homo sapien and Gorilla Gorilla CYP11A

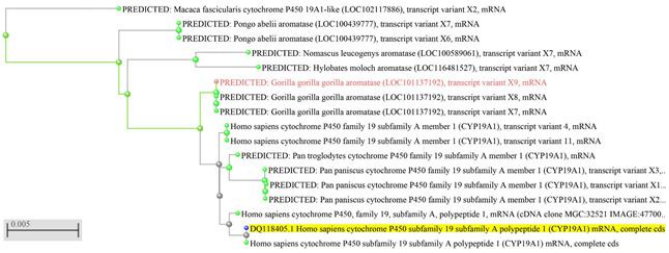

Figure 2: Phylogenetic tree showing evolutionary relationship between *Homo sapien* and *Gorilla Gorilla* CYP19A1

For both the genes, the prospective model organism *Gorilla gorilla* and *Homo sapiens* lie on nearby clads, ascertaining a close evolutionary relationship between the organisms for the genes of interest.

C. Protein Analysis

The retrieved amino acid sequences were submitted to the Prot-Param tool to compare the physical and chemical properties of the corresponding proteins as shown in Table 3.

TABLE 3

PHYSICAL AND CHEMICAL PROPERTIES OF THE PROTEINS ENCODED BY *HOMO SAPIEN*, *GORILLA GORILLA, OVIS ARIES* AND *MACACA MULATTA* CYP11A AND CYP19A1 GENES USING PROT-PARAM *MULATTA* CYP11A AND CYP19A1 GENES PREDICTED USING CHOU-FASMAN TOOL

Paramater	*Homo sapiens* CYP11A	*Gorilla Gorilla* CYP11A	*Macaca mulatta* CYP11A	*Ovis aries* CYP11A	*Homo sapiens* CYP19A1	*Gorilla Gorilla* CYP19A1	*Macaca mulatta* CYP19A1
Number of Amino Acids	521	521	521	520	503	503	503
Molecular Weight	60102.32	60048.33	60180.45	60356.94	57882.98	57868.95	57972.12
Theoretical pI	8.89	8.94	8.89	9.33	7.20	7.20	7.20
Total -vely Charged Residues	56	55	57	56	59	59	59
Total +vely Charged Residues	61	61	62	68	59	59	59
Instability Index	Stable	Stable	Stable	Stable	Stable	Stable	Stable
Aliphatic Index	89.06	89.06	88.12	88.67	98.81	98.81	94.73
Hydropathicity Index	-0.245	-0.232	-0.257	-0.316	- 0.003	-0.003	-0.034

The tool Prot-Param takes into consideration the amino acid sequences of the respective genes to deduce physical and chemical parameters of the protein so formed. The physical and chemical properties of the proteins coded by their respective genes in *Homo sapiens* and *Gorilla gorilla* were found to be highly similar *in-silico*.

Chou-Fasman tool predicts the secondary structure of the protein in the form of helices, sheets and turn propensities by scanning the sequence for amino acid residues that have high probabilities of forming any of these secondary structures. The results given by the Chou-Fasman tool are tabulated in Table 4.

TABLE 4

SECONDARY STRUCTURES OF THE OF THE PROTEINS ENCODED BY *HOMO SAPIEN, GORILLA GORILLA, OVIS ARIES* AND *MACACA MULATTA* CYP11A AND CYP19A1 GENES PREDICTED USING CHOU-FASMAN TOOL.

	Helix (%)	Sheets (%)	Turns (%)
Homo sapiens CYP11A	74.1	48.9	12.1
Gorilla gorilla CYP11A	74.1	49.3	11.9
Macaca mulatta CYP11A	74.9	47.6	12.3
Ovis aries CYP11A	75.8	42.1	12.5
Homo sapiens CYP19A1	76.7	52.1	10.7
Gorilla gorilla CYP19A1	76.1	52.1	10.7
Macaca mulatta CYP11A	77.1	52.3	10.5

The secondary structure of the protein encoded by the human CYP19A1 and CYP19A1 genes was found to be highly similar to the corresponding genes in *Gorilla gorilla in-silico*.

Thus, in all parameters, *Gorilla gorilla* was considered most suitable for the study of PCOS, even superior to the conventionally used model organisms – *Ovis aries* and *Macaca mulatta*.

IV. DISCUSSION

Research in the field of PCOS is inadequate due to the unavailability of suitable model organism, hence, in this study, a model system is being established through sequence retrieval and comparison. A model organism can be defined by several characteristics such as the availability of a well sequenced genome, ease of visualisation, convenient breeding and manipulation, large number of embryos, easy maintenance and defined ethical implications[9]. Common model organisms for the study of PCOS are *Ovis aries* and *Macaca mulatta*, however these research endeavours seldom translate into therapeutics for humans[6]. This study was focussed at the genetic level to find similarities in genes and proteins attributed to cause PCOS like symptoms, in order to find a potential model with maximum homology to their human gene counterparts.

Bioinformatics is the branch of science supervising the acquisition, storage, retrieval and analysis of crucial biological data. *In-silico* tools can be used to recover gene and protein sequences of several species that are stored in curated databases and deduce homology between organisms.

BLAST is an ideal tool for preliminary sequence analysis - in this study, it was employed to narrow down the search for a PCOS model from hundreds of database organisms to top 5 that could be further investigated. The shortlisted five organisms were found to be belonging to the order of primates. These results gcorroborated with the concept of evolutionary continuity, wherein the function and behaviour of biochemical processes were found to be resembling in evolutionarily related organisms[10]. BLAST tends to overlook gaps present within sequence stretches and begins aligning from a locus of similarity.

A natural extension of pairwise local alignment is global alignment. Global alignment involves end to end alignment indicating single nucleotide matches and mismatches[11]. Organisms showing highest percent similarity to the *Homo sapien* genes in BLAST were taken forward for global alignment. Thus, for global alignment *Gorilla gorilla* and *Pan paniscus* were chosen. The sequence alignment steps were parallelly conducted for the sequences of the conventionally used model organisms for PCOS, i.e., *Ovis aries* and *Macaca mulatta*.

Global alignment has several advantages over local such as an end-to-end alignment, consideration of gaps and a more practical representation of percent identity EMBOSS with Stretcher Algorithm and BLAST with Needleman algorithm were used to narrow down to the prospective model organism which was deduced to be *Gorilla gorilla.*

In order to study the evolutionary journey of the genes across species, a phylogenetic tree was created. *Gorilla* and *Homo sapien* gene sequences were found to be present on nearby clads, thereby establishing a close evolutionary and genetic relationship.

Although sequence comparison revealed Gorilla to be the closest match for study of the human biomarker genes, PCOS is generally defined by malfunctions at the protein-level[6]. Thus, it was vital to ensure that the protein characterisation studies could also be carried out using *Gorilla gorilla*. There are two aspects to characterising a protein coded by a gene - the secondary structure and the physicochemical properties exhibited by the protein. By confirming *in-silico* similarity in structure and physicochemical properties of proteins encoded by Gorilla genes, their behaviour *in-vivo* can be expected to be similar to their human gene counterparts. An in-vitro analysis of the genes and their corresponding proteins must be carried out to establish such an expectation.

The secondary structure is predicted in terms of helix propensities, sheet, turn and coil propensities using the Chou-Fasman tool. The secondary structure of the respective proteins in humans and *Gorilla* was found to be highly similar, down to the percentage of each of the secondary structural components. However, this only ascertains the structure of the proteins, and cannot verify if the behaviour of the proteins in-vitro would exactly mirror each other.

V. CONCLUSION

The study identified *Gorilla gorilla* to be the most suitable model organism for the efficacious study of PCOS. It was also confirmed meticulously that the genes of interest are only found in mammals, with primates showing highest similarity with the human genes.

Future prospects of the research include visualisation of the protein encoded by the respective genes and possible sites of mutations. An ideal model organism showing maximum homology to their gene counterparts in *Homo sapiens*, will make the studying of this disorder trouble-free. *In vitro* studies of the genes can be carried out using a *Gorilla gorilla* ovarian cell line. The model organism proposed by our project promises to supersede current systems, thereby facilitating better study and development of therapy. While *in-silico* tools can help in preliminary studies, a major limitation lies in their inability to guarantee the same results *in-vitro.* Bioinformatics tools aide in making calculated predictions and can drastically save time and monetary efforts involved in futile trial and error[14].

VI. ACKNOWLEDGMENTS

We would like to extend our humble gratitude to Jai Hind College Autonomous, Mumbai, the Department of Biotechnology, the Co-ordinators of the Biotechnology Department, Dr. Nissey Sunil and Dr. Kruti Pandya, as

well as the teaching and non-teaching faculty of the department for paving way for research to be conducted, and being pillars of counsel, confidence and motivation. The authors declare no conflict of interest.

VII. REFERENCES

[1]. Shaikh N, Dadachanji R & Mukherjee S. Genetic Markers of Polycystic Ovary Syndrome: Emphasis on Insulin Resistance. International Journal of Medical Genetics. 2014.

[2]. Pena AS, Witchel SF, Hoeger KM, Oberfield SE, Vogiatzi MG et al. Adolescent polycystic ovary syndrome according to the international evidence-based guideline. BMC Med. 2020.

[3]. Legro RS. Evaluation and Treatment of Polycystic Ovary Syndrome. Endotext. 2017.

[4]. Hayek SE, Bitar L, Hamdar LH, Mirza FD & Daoud G. Polycystic Ovarian Syndrome: An Updated Overview. Frontiers in Physiology. 2016; 7.

[5]. Khan MJ, Ullah A, & Basit S. Genetic Basis of Polycystic Ovary Syndrome (PCOS): Current Perspectives. Appl Clin Gene. 2019; 12:249-260.

[6]. Shareef E & Alwakeel M, New markers for the detection of polycystic ovary syndrome, Obstetrics & Gynecology International Journal, 10 (2019) 257-268.

[7]. Panghiyangani R, Soeharso P, Suryandari WA, Wiweko B, Kurniati M et al. CYP19A1 Gene Expression in Patients with Polycystic Ovarian Syndrome. J Hum Reprod Sci. 2020; 3:100–103.

[8]. Simmons D. The Use of Animal Models in Studying Genetic Disease: Transgenesis and Induced Mutation. Nature Education. 2008; 1.

[9]. Alberts B. Molecular Biology of the Cell (5th ed.). Garland Science, 2017.

[10]. Fu DL & Fu H. An Evolutionary Continuity Principle for Evolutionary System of Organism Divisions. American Journal of Agriculture and Forestry. 2018; 6:60-64.

[11]. Xiong J. Essential Bioinformatics. Cambridge University Press, 2006.

[12]. Needleman SB & Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology. 1970; 48:443–53.

[13]. Chou PY & Fasman GD. Prediction of protein conformation. Biochemistry. 1974; 13:222–245.

[14]. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins R, et al. Protein Identification and Analysis Tools on the ExPASy Server. John M. Walker, The Proteomics Protocols Handbook, Humana Press, 2005, 571-607.

[15]. Rhee SY, Bioinformatics. Current Limitations and Insights for the Future, Plant Physiology. 2005; 132:569-570.

[16]. Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990; 215:403-410.

[17]. Smith TF & Waterman MS. Identification of common molecular subsequences. Journal of Molecular Biology. 1981;147.

[18]. Myers EW & Miller W. Optimal alignments in linear space. Computer Applications in the Biosciences. 1988;4.

[19]. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suit. Trends in Genetics. 2000; 16:276–277.

Cite this article as :

Shruti Shastry, Soumyashree Ghosh, Ruqayya Manasawala, "In-Silico Screening of Prime PCOS Biomarkers for the Identification of a Potential Model Organism", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 6, pp. 448-454, November-December 2021. Available at doi : https://doi.org/10.32628/IJSRST218665

Journal URL : https://ijsrst.com/IJSRST218665