International Research Journal of Biological Sciences ISSN 2278-3202
Vol. **10(1),** 14-18, February **(2021)** *Int. Res. J. Biological Sci.*

Bioinformatics tools: alternative approach for Poly Cystic Ovary Syndrome (PCOS) Detection

Riya and Parul Thapar*
Biotechnology, School of Pharmaceutical and Health care Sciences, CT University, Ludhiana, India
parul.thapar@yahoo.com

Available online at: www.isca.in, www.isca.me
Received 20th June 2020, revised 25th October 2020, accepted 19th November 2020

Abstract

Poly cystic ovary syndrome (PCOS) is a complex endocrine disorder prevailing among women of reproductive age, specifically high in teenage girls and has become the most accepted cause of menstrual irregularities and infertility among them. Therefore, it was thought to explore an alternative approach for disease detection, prevention or treatment for PCOS with the help of bioinformatics. Bioinformatics is an expanding field of science involving biology, computer science and mathematics. It is growing in every field of life science including molecular sciences, biotechnology, medicine, agriculture and more. Genetic information stored in the bioinformatics tools can be used to develop personal medicine. In another study, it is said that although the genetics and mechanism of PCOS are not yet understood, the computational tools may be helpful in finding the cause of this syndrome and this will also help in prevention of the disease. In the present study, the gene and genome sequences responsible for causing PCOS have been identified using bioinformatics tools like BLAST, PDB, NCBI. This will help to prevent the disease by genetic manipulations. Finally, the primers have been designed using primer designing tools in NCBI which can be further used for the treatment of the disease by manipulating the identified gene through polymerase chain reaction.

Keywords: Bioinformatics tools, BLAST, Poly cystic ovary syndrome (PCOS), Protein Data Bank (PDB), CLUSTAL W.

Introduction

Poly cystic ovary syndrome (PCOS) is a complex endocrine disorder prevailing among women of reproductive age, specifically high in teenage girls and has become the most prominent cause of menstrual irregularities and infertility among them. PCOS is primary cause of female infertility due to failure in oocyte- follicle maturation[1]. This disorder is associated with modernisation of living in middle- income and high-income population. Young adolescent girls experience full range of symptoms from irregular menses, amenorrhea, menorrhagia, hirsutism, acne, skin pigmentation, alopecia and ovarian cysts. Other symptoms like anxiety, depression and thyroid problems may exist. Obesity or propensity to weight gain is a common feature, though it is not uncommon in non- obese women. There are no specific diagnostic tests designed for PCOS, but certain other tests like pelvic exam, blood tests, ultrasound, checking for the signs of excess hair growth, insulin resistance and acne can be performed. In a study of adolescent girls with PCOS have a higher prevalence of metabolic syndrome than general adolescent girls with 308 times increased risk. In a study, Palvi et al. revealed that PCOS can linked with malignancies like cancerous conditions of the Endometrium, Breast, Uterine in early reproductive age. Due to the criteria to determine PCOS are still changing there is no exact value how many women are affected but estimated one of 10-15 women are affected with PCOS[2]. Diagnosis of PCOS in adults can follow three different guidelines, that are National Institutes of health criteria, Androgen Excess PCOS Society Criteria and Rotterdam Criteria[3]. The medications recommended by the doctors to regulate menstrual cycles are the combination of birth control pills and progestin therapy. Examples are Clomiphene, Letozole, Metformin and Gonadotropins. To reduce excessive hair growth, the medicines used are Oral Control Pills (OCPs), Spironolactone, Metformin (Mt) and Eflornithine. But these medications create certain side effects in young women[4-7]. Metformin is linked with fatal and nonfatal lactic acidosis and OCPs have certain side effects like weight gain, cardiovascular and thromboembolic events[8]. Therefore, it was thought to explore an alternative approach of disease detection, prevention or treatment for PCOS with the help of bioinformatics.

Bioinformatics is an expanding field of science involving biology, computer science and mathematics. It is growing in every field of life science including molecular sciences, biotechnology, medicine, agriculture and other areas. Various tools and databases used in the field of bioinformatics are National Centre for Biotechnology Information (NCBI), Gen bank, European Nucleotide Archive, Ensemble, Basic Logic Alignment Search Tool (BLAST) which mainly contain information of genome of different organisms present in the universe. Due to diversity and complexity of PCOS, its methodology and pathogenesis are unknown, therefore, understanding the molecular mechanism of occurrence and development of PCOS is crucial to develop more effective diagnostic and therapeutic strategies[9]. Personal medicine is the

customised way of healthcare for treating patients using genetic information. These bioinformatics tools and databases are used to demonstrate genetic information[10].

In a study, the pathogenesis by the candidate genes causing PCOS was conducted using bioinformatics analysis. In another study, it is said that although the genetics and mechanism of PCOS are not yet understood, the computational tools may be helpful in finding the cause of this syndrome and this will also help in prevention of the disease[4-7]. To identify PCOS genes, reverse genetics like microarray studies, Genome-wide association studies, computational methods and others are used but no efficient algorithm has been developed to predict PCOS genes. In fact, bioinformatics algorithms have been successfully developed to infer candidate genes in other fields and these could be introduced to PCOS research[11].

In the present study, genome sequences and the gene responsible for causing PCOS have been identified so that this particular disease can be prevented by genetic manipulation and there cannot be requirement of taking medicines to treat the disease. The objectives of the study are: i. Searching for genes causing PCOS. ii. Retrieving the sequence for the identified genes. iii. Designing the primers for detecting the disease.

Materials and methods

Searching of Genes Causing PCOS: From the human genome, two main genes were searched that code for the proteins within the ovary. The source used was, "The human protein atlas" by the method described in human proteome[10].

Retrieving Sequences for the Genes: The two different genes were searched and sequences were retrieved for both the genes using bioinformatics NCBI database. The NCBI database includes Basic Logic Alignment Search Tool (BLAST) which is available for retrieving the sequences. i. For the first gene, the home page of NCBI was opened, in "all databases" tab; chosen "nucleotide", and put the searched gene in the search bar and then enter. A new web page appeared, chosen the desired gene from all (*Homo sapiens*) and then i.e. "Homo sapiens kelch domain containing 8A, transcript variant 2, mRNA" sequence appeared. By clicking on it, the graphics and FASTA sequence for desired gene was obtained. ii. For the second gene, the home page of NCBI was opened, in all databases, chosen "gene", and put the searched gene in the search bar and then enter. A new web page appeared, chosen the desired sequence from "all" and the information of that particular gene sequence appeared. At the end of the webpage, from the related sequences of the genome, the desired nucleotide (*homosapiens*) was selected. Then, the complete sequence of the human DNA sequence from clone RP5-87J13 on chromosome Xq22.1-22.3 appeared. By clicking on it, the graphics and FASTA sequence for desired gene was obtained.

For both the retrieved sequences, the number of the sequences was obtained using "highlight sequence" feature and the number of base pairs was calculated by using CLUSTAL W tool.

Designing of Primers for detecting the disease: The designed primers for the sequence may help in the determination of the disease and it can be treated by manipulating the desired identified gene. i. From the first identified sequence, after the FASTA sequence was retrieved, again visited the NCBI database to design the primers. Then, clicked on "pick primers" in the NCBI software for primer designing. The length of the forward and reverse primer was filled as 100 to 400 and 500 to 800 respectively and then clicked on "get primers". Therefore, the primers were designed for the identified gene. ii. From the second identified sequence, after the FASTA sequence was retrieved, again visited the NCBI database to design the primers. Then, clicked on "pick primers" in the NCBI software for primer designing. The length of the forward and reverse primers was filled as100 to 400 and 500 to 900 respectively and then clicked on "get primers". Therefore, the primers were designed for the second identified gene.

Results and discussion

Searching of Genes Causing PCOS: The two different genes were searched from the human protein atlas. The first gene identified was KLHDC8A. This gene encodes a kelch domain sequence which can be the cause of cancer. Sometimes the gene encoding kelch domain sequence is regulated due to the formation of tumours which activates the tumour into Onco cells and causes cancer. Since this gene is also enriched in ovary tissue therefore it is the most prominent cause of PCOS[6]. The second gene identified was MUM1L1 i.e. melanoma associated antigen (mutated) 1-like 1. This gene encodes a protein which contains a mutated melanoma-associated antigen-1 domain on its membrane. The proteins which contain mutated antigens are expressed at high levels in certain types of cancers. The tissues enhanced are ovary and thyroid gland. It is not detected into immune cells and predicted location is intracellular. It targets the melanoma cells if present in the ovaries[7].

In a study conducted by Panda *et al*[12] similar results have been isolated from Uni Prot and PDB tools for follicle stimulating hormone receptor with accession number: AAB26480, Uni Prot id: P23945 and PDB id: 1XUN. Also, genes like Melanocortin 4 receptor with accession number: AAO92061, Uni Prot id: P32245, PDB id: 2IQP have been identified with the same bioinformatics tools.

In a study conducted by Jie-Xue Pan *et al*[13] a total 23675 annotated ENSEMBL genes were detected and subsequent analysis was included of which 12814 genes were upregulated and 10861 genes were down regulated in GCs of PCOS women.

Retrieving Sequences for the Identified Genes: The sequences were retrieved through NCBI database using

bioinformatics tools like BLAST for both the genes KLHDC8A and MUM1L1 respectively. The number of sequences retrieved for the gene KLHDC8A was1, 632 by using NCBI software using "highlight sequence" feature and the number of base pairs obtained using CLUSTAL W tool was 2,970. The obtained data is represented in the Table-1. The number of sequences retrieved are shown in the Figure-1 and the base pairs in the Figure-2(a). The number of sequences retrieved for the gene MUM1L1 was 61,921 by using NCBI software using "highlight sequence" feature and the number of base pairs was 61,957 by using CLUSTAL W tool. The obtained data is represented in the Table-1. The number of sequences retrieved are shown in the Figure 3 and the base pairs in the Figure-2(b).

Similar results have been found in a study by Panda *et al*[12]. The sequence for the follicle stimulating hormone receptor gene encoding the PCOS has been identified in the accession number: AAB26480.

Table-1: Features of the Retrieved Sequences.

Genes Identified	No. of sequences	No. of base pairs
KLHDC8A	1,632	2,970
MUM1L1	61,921	61,957

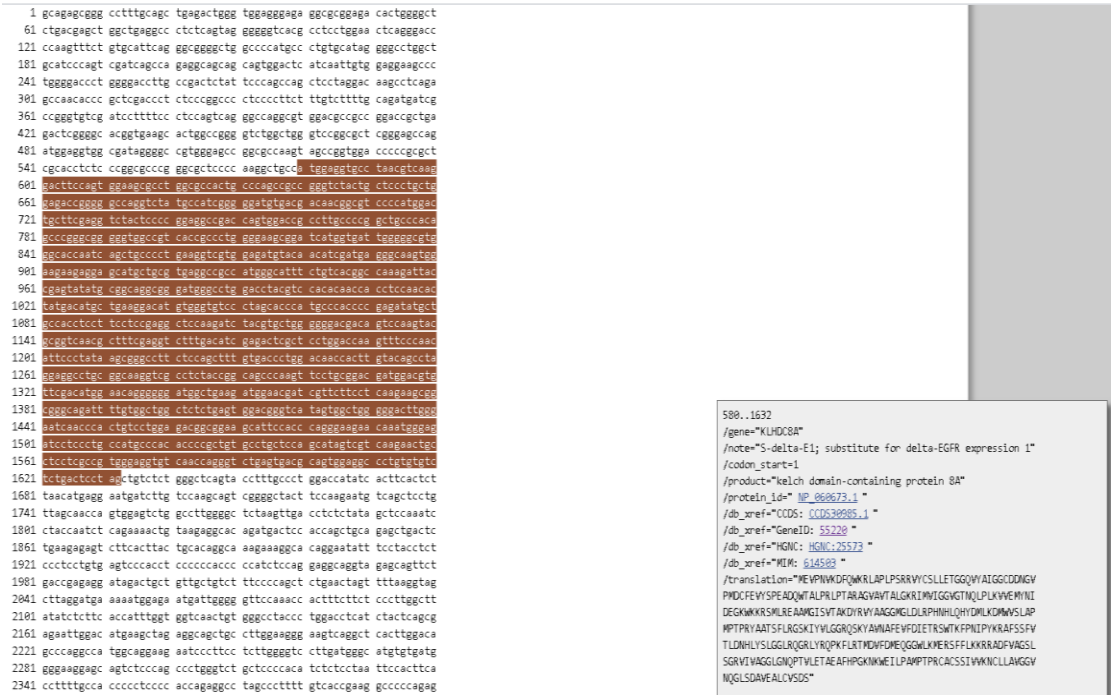
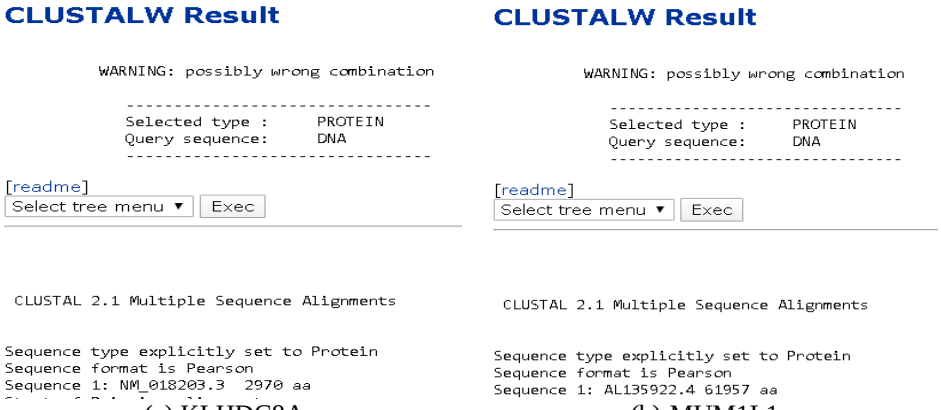


Figure-1: Number of sequences for gene KLHDC8A.



(a) KLHDC8A (b) MUM1L1

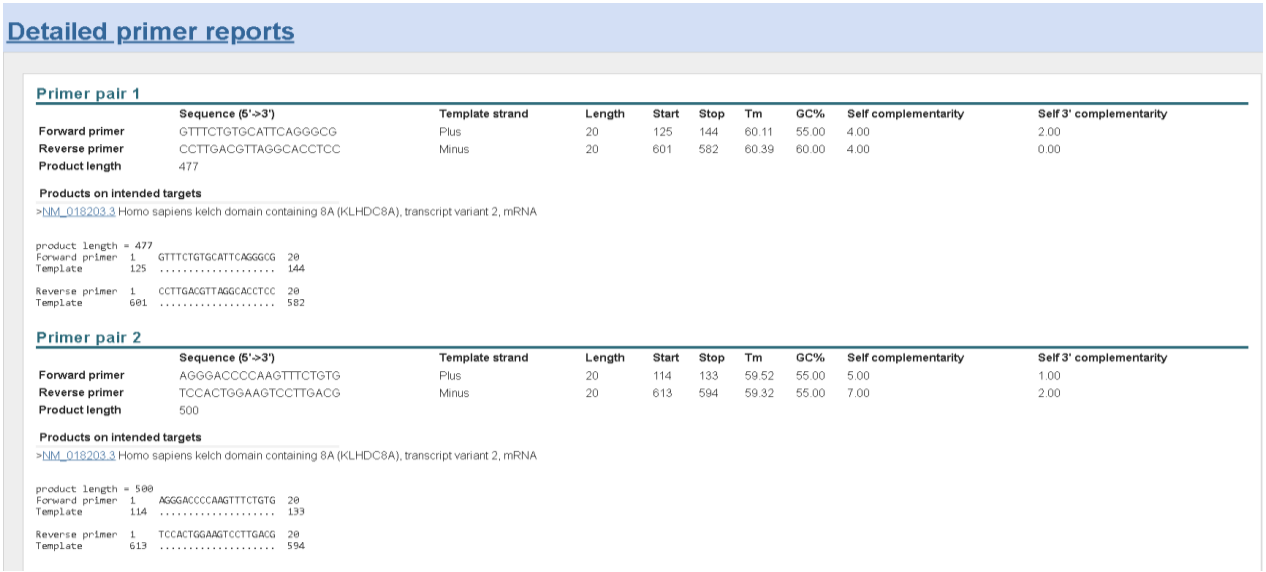
Figure-2: Number of base pairs of identified genes.

Figure-3: Number of sequences for gene MUM1L1.

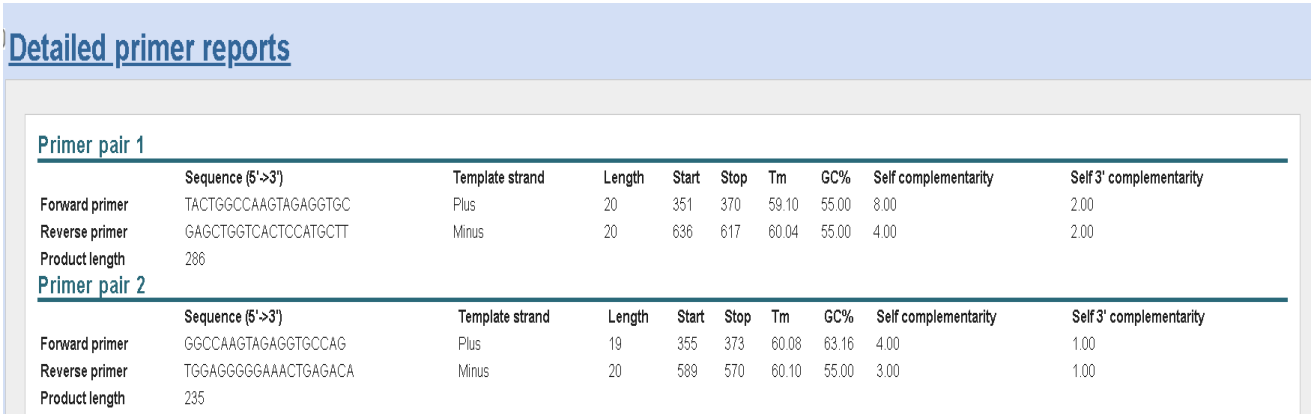
Designing of Primers for detecting the disease: The primers for the retrieved sequences were designed through NCBI database. For each gene, KLHDC8A and MUM1L1**,** 10 different primers were designed. The designed primers are shown in the Figure 4-(a), (b). These primers can be used in future for the treatment of the PCOS disease by manipulating the identified gene through polymerase chain reaction followed by genetic engineering.

In a study conducted by Artimani *et al*[14], similar results have been identified for primer designing of follicle stimulating hormone receptor gene using bioinformatics tool and software. Accession number for the gene is 000145 and forward and reverse primers were GAGAGCAAGGTGACAGAGATTC and TTGATGTAGAGCAGGTTGTTGG respectively.

In a study by Chen *et al*[15], similar sequences were obtained for the target genes of PCOS, i.e. forward ptimers – AGTCGATG ATGCTAGCTGA and CTAGCTAGATAGCTAGCTACG; reverse primers- CGTAGCTAGCTAGCTTACG and CGATGCATATTAGCTACGATGC.



(a) KLHDC8A



(b) MUM1L1

Figure-4: The designed Primers.

Conclusion

Poly Cystic ovary syndrome (PCOS) is a common cause of an ovulatory infertility in upto 5-10% of women of reproductive age. It is believed that both genetic and environmental factors play important roles in the occurrence and development of PCOS. It is an endocrine disorder. The medications recommended by the doctors to regulate menstrual cycles are the combination of birth control pills and progestin therapy, but these pills have certain side effects like weight gain, cardiovascular and thromboembolic events. Therefore, as an alternative, computational tools may be helpful in finding the cause of this syndrome and this will also help in prevention of the disease. The various tools and databases used in the field of bioinformatics are National Centre for Biotechnological Information (NCBI), Gen bank, European Nucleotide Archive, Ensemble, Basic Logic Alignment Search Tool (BLAST). The present study shows, genes responsible for causing the disease (PCOS)- KLHDC8A and MUM1L1 have been identified. The sequences for the identified genes have been retrieved using the computational software like CLUSTAL W. Through NCBI and BLAST, 10 pairs of the primers for each identified gene have been designed which can be further used for the detection of PCOS by polymerase chain reaction and treatment by genetic manipulation. This can help in the treatment of the disease without any side effects which are caused by taking medications.

References

1. Hou, Y.,Wang, Y., Xu, S., Qi, G., & Wu, X. (2019). Bioinformatics identification of microRNAs involved in polycystic ovary syndrome based on microarray data. *Mol. Med. Rep*., 20(1), 281-291.

2. Palvi, S., & Sheveta, V. (2019). Image Segmentation for detecting Polycystic Ovarian Disease using Deep Neural Network. *Int. J. Comp. Sci. Eng.*, 7(3), 534-537.

3. Samer, E. H., Lynn B., Layal H. H., Fadi, G. M., & Georges, D. (2016). Poly Cystic Ovarian Syndrome: An Updated Overview. *Front Physiol*., 7, 124.

4. Prathap, A., Subhalakshmi, T. P., & Varghese, P. J. (2018). A Cross-sectional Study on the Proportion of Anxiety and Depression and Determinants of Quality of Life in Polycystic Ovarian Disease. *Ind. J. Psychol. Med.*, 40(3), 257-262**.**

5. Choudhary, A., Jain, S., & Chaudhari, P. (2017). Prevalence and symptomatology of polycystic ovarian syndrome in Indian women: is there a rising incidence?. *Int. J. Reproduction Contracep. Obstetrics Gynecol*., 6(11), 4971-4975.

6. Bharathi, V., Swetha, S., Neerajaa, J., Madhavica, V., Janani, D. M., Rekha, S. N., Ramya, S., & Usha, B. (2017). An epidemiological survey: Effect of predisposing factors for PCOS in Indian urban and rural population. *J. Middle East Fertility Society*, 22, 313-316.

7. Radhakrishnan, R., & Verghese, A. A. (2018). Study on anxiety and depression among patients with polycystic ovary syndrome. *J. Drug Delivery Therapeutics*, 8(5), 338-340.

8. Domecq, J. P., Prutsky, G., Mullan, R. J., Sundaresh, V., Wang, A. T., Erwin, P. J, Welt, C., Ehrmann, D., Montori, V. M., & Murad, M. H. (2013). Adverse Effects of the Common Treatments for Polycystic Ovary Syndrome: A Systematic Review and Meta-Analysis. *J. Clinical Endocrinolol. Metabolism*, 98(12), 4646-4654.

9. Bi, X., Zhai, Z., & Wang, S. (2019). Identification of the key pathways and genes related to polycystic ovary syndrome using bioinformatics analysis. *Gen. Physiol. Biophys*., 38, 205-214.

10. Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... & Pontén, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220).

11. Zhang, X. Z., Pang, Y. L., & Li, Y. H. (2018). Computational characterization and identification of human polycystic ovary syndrome genes. *Sci Rep*., 8, 12949.

12. Panda, P. K., Rane, R., Chandran, R. R., Singh, S., & Panchal, H. (2016). Genetics of PCOS: A systematic bioinformatics approach to unveil the proteins responsible for PCOS. *Genomics Data*, 8, 52-60.

13. Jie, X. P., Ya, J. T., Fang, F. W., Ning, N.H., Yu, Q. X., Jun, Y. Z., Ye, L., Fan, Q., Qing, M., Jian, X., Jian, Z. S., & He, F. H. (2018). Aberrant expression and DNA methylation of lipid metabolism genes in PCOS: a new insight into its pathogenesis. *Clin. Epigenet.*, 10, 6.

14. Tayebe, A., Massoud, S., Reza, A., Mahnaz, A., Iraj, A., Mahnaz, Y., Sara, S. A., Nooshin, S., Jamshid, K., & Mehdi, M. (2015). Down regulation of adiponectin system in granulosa cells and low levels of HMW adiponectin in PCOS. *J. Assisted Reproduction Genetics*, 33(1), 101-110.

15. Chen, L., Ling, M. H., Wang, F. U., Yang, H. Y., Huang, X. Y., Zhou, W., & Sun, H. X. (2017). Genome wide association study for SNPs associated with PCOS in human patients. *Experimental Therapeutic Med*., 7(9), 4896-4900.

16. Thapar, P. (2018). Bioinformatics- tools and applications. Proceedings from 12th INDIA Com International Conference, New Delhi, India, 14th- 16th March. 5044-5047.