



e-ISSN:2582 - 7219



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 4, Issue 8, August 2021



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 5.928



9710 583 466



9710 583 466



ijmrset@gmail.com



www.ijmrset.com



Bioinformatics Approaches to Understand the Role of African Genetic Diversity in Diseases

Sudheer Menon

Researcher, Department of Bioinformatics, Bharathiar University, Coimbatore, Tamil Nadu, India

ABSTRACT: Human adaptability and history sculpt the variation in genes and make them vulnerable to infectious diseases. By using a computational algorithm, African genomes and their immigration are being studied to reveal the spreading of different diseases in humans, animals, and plants. There are different genetic bases that make the differences in terms of disease vulnerability in different organisms. Development in genetics and bioinformatics have been providing us good tools that find the genes which are involved in the disease. Strains of Newcastle disease virus (NDV) are found in poultry in West and Central Africa. In Africa, Sweetpotato is a widely grown crop, and there are certain strains of viruses found in Sweetpotato. The viruses that are spreading disease in Sweetpotato are SPFMV (Sweetpotato Feathery Mottle Virus), SPMMV (Sweetpotato Mild Mottle Virus), SPCSV (Sweetpotato Chlorotic Stunt Virus).

KEYWORDS: Genetic diversity, African Population, Infectious Diseases, Bioinformatics Tools, Isozyme.

INTRODUCTION

Genetic diversity refers to the total number of genetic characteristics in the genetic makeup of any species. The adaptations in which populations adapt to environmental changes are due to genetic diversity. Multigene mechanisms are involved in the genetic vulnerability to diseases.

If the genetic vulnerability of any infectious disease in humans is considered, then the population level is more important than the individual level to study the disease. Pathogens diversity and their distribution over a region are also affected by the geographical and zonal distribution.

There are two main approaches to study the genetic diversity in diseases.

- Candidate gene approach
- Linkage approach

In the first approach that is the Candidate gene approach, the hypothesis is formed that a given gene or gene set is involved in the disease vulnerability.

In the linkage approach, with the help of the microsatellites, the specific region of the genome is discovered from the whole genome which causes disease vulnerability.

Lod score is used in the parametric linkages.

$$\text{Lod} = \log_{10} \left[\frac{\text{(probability of data if disease and marker are linked)}}{\text{(probability of data if disease and marker recombine freely)}} \right]$$

Newcastle disease in poultry is caused by a virus named avian paramyxovirus. Newcastle disease virus (NDV) is subcategorized into lentogenic, mesogenic, asymptomatic enteric on the basis of their severity in the chickens (OIE, 2012). In West Africa, virulent and avirulent strains of the Newcastle virus are found. In sub-Saharan Africa virulent strains of the Newcastle disease virus are found.

Sweetpotato Virus Disease is the most destroying syndrome in sweet potatoes. In Africa Sweetpotato production affecting the virus is revealed (Karyeija et al., 1998).



Human diseases are linked with protein misfolding. The diseases that are linked with Intrinsically disease proteins (IDPs) and Intrinsically disease regions (IDRs) are genuine and in huge numbers. New tools in bioinformatics showed the characterization of diseases associated with IDPs and IDRs unfoldome. Intrinsically Disease Proteins (IDPs) have some unique structural and functional characteristics that are involved in the pathogenesis of many diseases.

Genetic Susceptibility to infectious diseases

The following conditions are fulfilled in the gene identification methods (Cooke GS et.al., 2001).

- 1- Phenotypic traits are defined authentically
- 2- Genetic component of vulnerability to any disease is strong enough
- 3- A few number of genes are involved in the vulnerability to a disease
- 4- Availability of the involved Mendelian inheritance genes

It is discovered that infectious diseases depend upon the complex interactions between genes and their environmental parameters. It is revealed that neutral genes in the population or species have a direct relation with the time elapsed due to the reason they have common ancestors (Historical Markers).

Neutral polymorphism has limited predictive power on adaptive polymorphism by its very nature. On the other hand, it offers a basic population genetic framework for the species under investigation, as well as information on the degree of genetic similarity across populations, migration and genetic exchange rates, and a variety of other characteristics.

Isozyme discovery

Because a direct study of genetic polymorphism was not accessible in the late 1970s, population genetics had a theoretical foundation that was extremely speculative, but it has now been proven to be correct. The research was reported in Nature Biotechnology. In the 1970s and 1980s, isoenzymes were all the rage. A staggering amount of documents covering virtually the whole reign of the living were published. Indeed, Multilocus enzyme electrophoresis(MLEE) is an excellent example of a “generalist” genetic marker, one that may be used for any organism(Tibayrenc M, 1998).For finding the genetic variability in Human Multilocus enzyme electrophoresis is used (Nei M et.al., 1993).

Genetic Diversity and population in Africa

Demographic history (e.g., changes in population size, short- and long-range migratory events, and mixing) as well as locus-specific factors like natural selection, recombination, and mutation impact the pattern of genetic diversity in current African groups (Pilkington MM et.al., 2008). Despite the importance of Africa in understanding contemporary human origins and disease genetic risk factors, it has been underrepresented in human genetic studies. Much of what we know about genetic variation comes from a small number of Africa's 2000 ethnolinguistic groupings, and the majority of this information comes from mtDNA and Y chromosome research (Garrigan D et.al., 2007).

Additional substructure exists among African groups, according to analyses, notably between hunter-gatherer and agriculturalist communities (Jacobsson M et.al., 2008).

The CEPH diversity panel, on the other hand, only contains eight African groups, four of which are agricultural Bantu-speakers with recent shared ancestry.

Infectious diseases in Human Population of Africa

In humans, host genetic diversity influences vulnerability to a variety of infectious illnesses. A multitude of genetic modifications that offer infection resistance has developed as a result of repeated exposure to various pathogens. Despite the fact that the number of known candidate genes for infectious diseases has grown, progress in identifying genes that impact infectious disease susceptibility and/or resistance in various African groups has been modest. Malaria, AIDS (Autoimmune Deficiency Syndrome), and tuberculosis are some of the major diseases that spread in the huge population.

Malaria is the major population disease in Africa. About 90% population in Sub-Sahara Africa is affected by Malaria (Frodsham AJ et.al., 2004). Plasmodium falciparum parasite-infected more than 500 million population in Africa. Red blood cells play an important role in the immune response against Malaria.

Numerous studies have found that the genetic vulnerability to malaria infection differs among ethnically varied African groups. For example, a variation in the promoter region of the IL4 gene is linked to a reduction in P. falciparum infection in pastoralist Fulani from Mali, as demonstrated by reduced parasite burden, while no such genetic relationship is seen in the pastoralist Fulani from Nigeria (Dolo A et.al., 2005).



The second infectious disease is Autoimmune Deficiency Syndrome (AIDS) which is caused by Human Immunodeficiency Virus (HIV). 75-84% of deaths in Sub-Sahara Africa are caused by HIV(Winkler C 2004). Some individuals who were exposed to HIV, rapidly increase in the disease's advanced stages. Some individuals did not infect with AIDS despite the exposure to HIV(O'Brien SJ et.al., 2004).HIV enters the host cells through Chemokine receptors (O'Brien SJ et.al., 2004) .it is demonstrated that the CCR2 gene act as a resistance in HIV.

Infection with Mycobacterium tuberculosis, which leads to tuberculosis, is a leading cause of death worldwide, especially in resource-poor nations. Furthermore, in Sub-Saharan Africa, HIV infection is closely linked to an increased risk of tuberculosis (TB) (Corbett EL et.al., 2006). In Africa, TB rates range from 50-300 cases per 100,000 people (215). In African communities, genetic diversity has been found to impact susceptibility to tuberculosis.



Fig 1. Malaria causing agent



Fig 2. HIV causing AIDS, affecting the cells

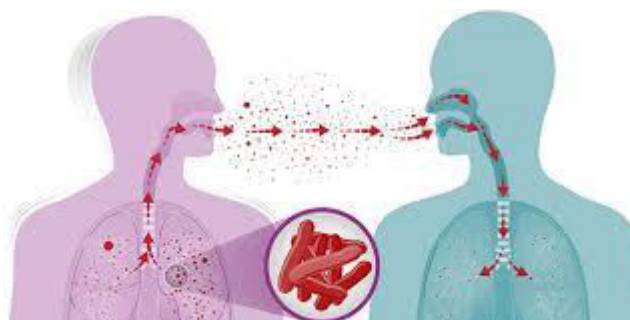


Fig 3. Spreading of Tuberculosis TB



Infectious disease	Known genes associated with disease susceptibility in Africans	Effect of genetic loci on infection or disease progression	References
HIV/AIDS	Chemokine receptor 2 (CCR2)	CCR-2 64I allele associated with ↓ disease progression	(146,235)
	Human leukocyte antigen (HLA) Locus	HLA A2/6802 supertype and HLA DRB1*01 associated with ↓ HIV infection in female sex workers from East Africa	(121)
	Killer immunoglobulin-like receptors (KIR)	KIR3DS1 in combination with HLA-B Bw4–80Ile associated with ↓ disease progression in African Americans and West African sex workers	(95,126)
	Interferon regulatory factor (IRF-1)	Three polymorphisms (located at 619, microsatellite region, and 6516 of genes) associated with ↓ susceptibility to infection in East African sex workers	(15)
	TRIM5α	136Q and 43Y alleles protect against HIV infection in African Americans	(108)
	APOBEC3G	186R allele associated with ↑ disease progression in African Americans	(8)
	CUL5	Haplotype10 associated with ↑ disease progression of HIV-1 in African Americans	(9)
Tuberculosis	Human homolog of natural-resistance-associated macrophage protein 1 (NRAMP1)	Several polymorphisms (5' microsatellite repeat, INT4, and a deletion in 3' UTR) associated with ↑ risk for tuberculosis in East and West Africans	(19,20a,188)
	UBE3A	7-bp deletion at 5' end of UBE3A gene associated with ↑ susceptibility in West African families	(27)
	Chromosome 15 and X chromosome	Unknown loci on these chromosomes associated with ↑ tuberculosis susceptibility	(20)
	Human leukocyte antigen (HLA) locus	DRB1*1302 allele and DQB1*0301—0304 alleles associated with ↑ tuberculosis susceptibility in Venda population from South Africa	(115)
	Vitamin D receptor (VDR) locus	F-b-A-T haplotype associated with ↓ susceptibility in a South African population	(115)
	Vitamin D receptor (VDR) locus	Polymorphism at codon 352 of VDR gene associated with ↓ susceptibility in Gambian population	(20)
	Vitamin D receptor (VDR) locus	Polymorphisms (FokI and ApaI) associated with ↑ susceptibility in West African families.	(149)
	CD209	Promoter –336G allele associated with ↓ susceptibility in West Africa and Malawi	(215)
		Promoter –336A and –871G alleles associated with ↓ susceptibility in admixed colored population from South Africa	(18)
		Intron 6 variant associated with ↑ susceptibility in West African population	(149)
	Pentraxin 3 (PTX3)	G-A-G haplotype associated with ↓ susceptibility in West Africa	(149)
Malaria	β-globin gene	HbS, HbC, and HbE alleles associated with ↓ malaria susceptibility in East and West Africans	(13a,216a)
	α-globin gene	α+-thalassemias associated with ↓ malaria susceptibility in East Africans	(58a)
	β-globin/α-globin genes	HbS and α+-thalassemia variants inherited together associated with ↑ susceptibility in East Africans	(232)
	Duffy gene	FY*O allele associated with ↓ malaria susceptibility in Africans	(83,86)

Table 1. Infectious diseases and the associated genes in the African Population

Bioinformatics tools used for disease identification

Infectious illness diagnosis and surveillance are becoming more common because of bioinformatics tools and procedures that analyze next-generation sequencing (NGS) data. Reviewing the use of bioinformatics tools, common databases, and NGS data in clinical microbiology, with an emphasis on molecular identification, genotyping, microbiome research, and antibiotic resistance studies, is of interest. Bioinformatics tools are widely utilized in pathogen detection, characterization, and typing. Following the growing use of genomic methods in the diagnosis and treatment of viral, bacteria, and fungal infections, this study was published.



Tool Names	URL
1. Lasergene	http://dnastar.com
2. CLCbio workbench	http://www.clcbio.com/products/clc-main-workbench/
3. Geneious	http://www.geneious.com/
4. ChimeraSlayer	http://microbiomeutil.sourceforge.net/#A_CS
5. AmpliconNoise	http://qiime.org/scripts/ampliconnoise.html
6. CATCH	http://science.sckcen.be/en/Institutes/EHS/MCB/MIC/Bioinformatics/CATCH
7. Mauve	http://gel.ahabs.wisc.edu/mauve
8. UCHIME algorithm	http://drive5.com/usearch/manual/uchime_algo.html
9. mothur	https://www.mothur.org/
10. DECIPHER	http://DECIPHER.cee.wisc.edu

Table 2. Bioinformatics tools

H3ABIONET is widely used network for analyzing and collecting the genomic data in Africa. Another network or approach was developed named, NETMAP that was used for the precise Data transfer between Africa and the rest of the world (<https://redcap.h3abionet.org/h3aprd/>).

For the analysis of Malaria, RV-approach was used. RV-approach consists of two steps.

- 1- Screening of antigen from the whole protein sequence by using tools such as, SignalP, VaxiJen, TMHMM etc.
- 2- Predicting the immunogenicity of T-cells epitope by using NetCTL program.

II. CONCLUSIONS

The study of ethnically varied human groups, particularly in Africa, is critical for recreating human evolutionary history and understanding the genetic foundation of complex diseases and phenotypic adaptability. Malaria, AIDS, and Tuberculosis are the major diseases found in the African population.

As mentioned in this review, several bioinformatics tools are available for evaluating data in the fight against and control infectious illnesses. However, bioinformatics tools for drug resistance testing, pathogen-host interaction, and treatment outcomes are available. The need to make bioinformatics tools and their application is more accessible and integrated with clinical microbiology and infectious diseases.

REFERENCES

- 1) OIE. 2012. Newcastle disease. In Manual of diagnostic tests and vaccines for terrestrial animals. OIE, Paris, France. http://www.oie.int/fifileadmin/Home/eng/Health_standards/tahm/2.03.14_NEWCASTLE_DIS.pdf.
- 2) Cooke GS, Hill AVS. Genetics of susceptibility to human infectious diseases. *Nat Rev Genet* 2001;2:967–77.



- 3) Tibayrenc M. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *Int J Parasitol* 1998;28:85–104.
- 4) Tibayrenc M. The impact of human genetic diversity in the transmission and severity of infectious diseases. In *Infectious Disease: Host–Pathogen Evolution* (ed. K. Dronamraju). Cambridge University Press, Cambridge, 2004, pp. 315–24.
- 5) Tishkoff SA, Kidd KK. Implications of biogeography of human populations for “race” and medicine. *Nat Genet* 2004;36(11):S21–7.
- 6) Valois HV. *Les races humaines*. Presses universitaires de France, Paris, 1971
- 7) Wright S. The genetical structure of populations. *Ann Eugen* 1951;15:323–54.
- 8) OIE. 2012. Newcastle disease. In *Manual of diagnostic tests and vaccines for terrestrial animals*. OIE, Paris, France. http://www.oie.int/fifileadmin/Home/eng/Health_standards/tahm/2.03.14_NEWCASTLE_DIS.pdf
- 9) Dolo A, Modiano D, Maiga B, Daou M, Dolo G, et al. Difference in susceptibility to malaria between two sympatric ethnic groups in Mali. *Am. J. Trop. Med. Hyg* 2005;72:243–248. [PubMed: 15772314]
- 10) Dong Y, Zhu H, Sagnella GA, Carter ND, Cook DG, Cappuccio FP. Association between the C825T polymorphism of the G protein β 3-subunit gene and hypertension in blacks. *Hypertension* 1999;34:1193–1196. [PubMed: 10601117]
- 11) Drayna D. Human taste genetics. *Annu. Rev. Genomics Hum. Genet* 2005;6:217–235. [PubMed: 16124860]
- 12) Sudheer Menon (2020) “Preparation and computational analysis of Bisulphite sequencing in Germfree Mice” *International Journal for Science and Advance Research In Technology*, 6(9) PP (557-565).
- 13) Sudheer Menon, ShanmughavelPiramanayakam and Gopal Agarwal (2021) “Computational identification of promoter regions in prokaryotes and Eukaryotes” *EPR International Journal of Agriculture and Rural Economic Research (ARER)*, Vol (9) Issue (7) July 2021, PP (21-28).
- 14) Sudheer Menon (2021) “Bioinformatics approaches to understand gene looping in human genome” *EPR International Journal of Research & Development (IJRD)*, Vol (6) Issue (7) July 2021, PP (170-173).
- 15) Sudheer Menon (2021) “Insilico analysis of terpenoids in *Saccharomyces Cerevisiae*” *international Journal of Engineering Applied Sciences and Technology*, 2021 Vol. 6, Issue1, ISSN No. 2455-2143, PP(43-52).
- 16) Sudheer Menon (2021) “Computational analysis of Histone modification and TFBs that mediates gene looping” *Bioinformatics, Pharmaceutical, and Chemical Sciences (RJLBPCS)*, June 2021, 7(3) PP (53-70).
- 17) Sudheer Menon Shanmughavelpiramanayakam, Gopal Prasad Agarwal (2021) “FPMD-Fungal promoter motif database: A database for the Promoter motifs regions in fungal genomes” *EPR International Journal of Multidisciplinary research*, 7(7) PP (620-623).
- 18) Sudheer Menon, ShanmughavelPiramanayakam and Gopal Agarwal (2021) *Computational Identification of promoter regions in fungal genomes*, *International Journal of Advance Research, Ideas and Innovations in Technology*, 7(4) PP (908-914).
- 19) Sudheer Menon, Vincent Chi Hang Lui and Paul Kwong Hang Tam (2021) *Bioinformatics methods for identifying hirschsprung disease genes*, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, Volume 9 Issue VII July, PP (2974-2978).
- 20) Drayna D, Coon H, Kim UK, Elsner T, Cromer K, et al. Genetic analysis of a complex trait in the Utah Genetic Reference Project: a major locus for PTC taste ability on chromosome 7q and a secondary locus on chromosome 16p. *Hum. Genet* 2003;112:567–572. [PubMed: 12624758]
- 21) Frodsham AJ, Hill AV. Genetics of infectious diseases. *Hum. Mol. Genet* 2004;13(Spec No 2):R187– R194. [PubMed: 15358724]
- 22) Fullerton SM, Bartoszewicz A, Ybazeta G, Horikawa Y, Bell GI, et al. Geographic and haplotype structure of candidate type 2 diabetes susceptibility variants at the calpain-10 locus. *Am. J. Hum. Genet* 2002;70:1096–1106. [PubMed: 11891618]
- 23) Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–2229. [PubMed: 12029063]
- 24) Garant MJ, Kao WH, Brancati F, Coresh J, Rami TM, et al. SNP43 of *CAPN10* and the risk of type 2 diabetes in African-Americans: the Atherosclerosis Risk in Communities Study. *Diabetes* 2002;51:231–237. [PubMed: 11756346]
- 25) Garrigan D, Hammer MF. Reconstructing human origins in the genomic era. *Nat. Rev. Genet* 2006;7:669–680. [PubMed: 16921345].



26. Garrigan D, Kingan SB, Pilkington MM, Wilder JA, Cox MP, et al. Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. *Genetics* 2007;177:2195–2207. [PubMed: 18073427]
27. Garrigan D, Mobasher Z, Severson T, Wilder JA, Hammer MF. Evidence for archaic Asian ancestry on the human X chromosome. *Mol. Biol. Evol* 2004;22:189–192. [PubMed: 15483323]
28. Gonder MK, Mortensen HM, Reed FA, de Sousa A, Tishkoff SA. Whole-mtDNA genome sequence analysis of ancient African lineages. *Mol. Biol. Evol* 2007;24:757–768. [PubMed: 17194802]
29. Graffelman J, Balding DJ, Gonzalez-Neira A, Bertranpetit J. Variation in estimated recombination rates across human populations. *Hum. Genet* 2007;122:301–310. [PubMed: 17609980]
30. Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, et al. Analysis of one million base pairs of Neanderthal DNA. *Nature* 2006;444:330–336. [PubMed: 17108958]
31. Guthery SL, Salisbury BA, Pungliya MS, Stephens JC, Bamshad M. The structure of common genetic variation in United States populations. *Am. J. Hum. Genet* 2007;81:1221–1231. [PubMed: 17999361]
32. Hamblin MT, Di Rienzo A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet* 2000;66:1669–1679. [PubMed: 10762551]
33. Hamblin MT, Thompson EE, Di Rienzo A. Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet* 2002;70:369–383. [PubMed: 11753822]
34. Hanchard N, Elzein A, Trafford C, Rockett K, Pinder M, et al. Classical sickle beta-globin haplotypes exhibit a high degree of long-range haplotype similarity in African and Afro-Caribbean populations. *BMC Genet* 2007;8:52. [PubMed: 17688704]
35. Harding RM, McVean G. A structured ancestral population for the evolution of modern humans. *Curr. Opin. Genet. Dev* 2004;14:667–674. [PubMed: 15531162]
36. Harris EE, Meyer D. The molecular signature of selection underlying human adaptations. *Am. J. Phys. Anthropol Suppl* 2006;43:89–130.
37. Hawks J, Wang ET, Cochran GM, Harpending HC, Moyzis RK. Recent acceleration of human adaptive evolution. *Proc. Natl. Acad. Sci. USA* 2007;104:20753–20758. [PubMed: 18087044]
38. Helgason A, Palsson S, Thorleifsson G, Grant SF, Emilsson V, et al. Refining the impact of *TCF7L2* gene variants on type 2 diabetes and adaptive evolution. *Nat. Genet* 2007;39:218–225.
39. Hill AV. Aspects of genetic susceptibility to human infectious diseases. *Annu. Rev. Genet* 2006;40:469–486. [PubMed: 17094741]
40. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet* 2005;6:95–108. [PubMed: 15716906]
41. Hollox EJ, Poulter M, Zvarik M, Ferak V, Krause A, et al. Lactase haplotype diversity in the Old World. *Am. J. Hum. Genet* 2001;68:160–172. [PubMed: 11095994]
42. O'Brien SJ, Nelson GW. Human genes that limit AIDS. *Nat. Genet* 2004;36:565–574. [PubMed: 15167933]
43. Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Brittenham G, et al. Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. *Am. J. Hum. Genet* 2004;74:1198–1208. [PubMed: 15114532]
44. Pilkington MM, Wilder JA, Mendez FL, Cox MP, Woerner A, et al. Contrasting signatures of population growth for mitochondrial DNA and Y chromosomes among human populations in Africa. *Mol. Biol. Evol* 2008;25:517–525. [PubMed: 18093995]
45. Plagnol V, Wall JD. Possible ancestral structure in human populations. *PLoS Genet* 2006;2:e105. [PubMed: 16895447]
46. Poulter M, Hollox E, Harvey CB, Mulcare C, Peuhkuri K, et al. The causal element for the lactase persistence/non-persistence polymorphism is located in a 1 Mb region of linkage disequilibrium in Europeans. *Ann. Hum. Genet* 2003;67:298–311. [PubMed: 12914565]
47. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet* 2001;69:124–137. [PubMed: 11404818]
48. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet* 2001;69:1–14. [PubMed: 11410837]



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor:
5.928

ISSN

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY



9710 583 466



9710 583 466



ijmrset@gmail.com

www.ijmrset.com