GENETIC COMPARISON OF DIFFERENT HUMAN POPULATION GROUPS USING INTERNAL TRANSCRIBED SPACER SEQUENCE

MUNA I. HADI G. ALATIYAH

A dissertation submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Biotechnology)

Faculty of Biosciences and Medical Engineering
Universiti Teknologi Malaysia

January 2013

To my beloved mother and father

ACKNOWLEDGEMENT

First , Alhamdulillah ,thanks to Allah for finishing the research and I would like to express my sincere thanks and appreciation to my supervisor Assoc.Prof Dr.Mohd .Shahir Shamsir Omar, who generously guided me with his knowledge and experience in every step of my thesis . My special thanks to Dr. Salehhuddin Bin Hamdan and I would like to take this opportunity to deliver my appreciation to him for letting me work in animal tissue culture laboratory and special thanks to all technical staff members in the bioinformatics and animal tissue culture laboratories especially Siti Nur Hafizah binti Soid at UTM for their assistance and collaboration during my laboratory work. Thanks to all my friends who showed their concern and support to me. Their views and advices are useful indeed.

ABSTRACT

Genetic identity testing for humans has been used to examine the variations in the polymorphic regions of the human DNA. These methods include the RFLP (Restriction fragment length polymorphisms), the STR (short tandem repeats) markers that are most commonly used loci for human identification as well as the mtDNA or Y chromosome in forensic medicine and paternity tests. Currently, the internal transcribed spacer (ITS) region is used for the evolutionary analysis of different species of animals, plants, fungi, yeast. However, the ITS region is yet to be used in genetic identity testing for humans. In this study, the genetic comparison of different human population groups using ITS sequence of mtDNA was performed. Two segments of the region of human mtDNA were selected and sequenced in order to determine if any sufficient single nucleotide polymorphisms (SNPs) exist and if it is suitable for genetic identity testing in human. Specific primers were designed to amplify the ITS regions using PCR from sample extracted from the blood samples of different nationalities of students from Faculty of Biosciences and Bioengineering, UTM (Iran, Nigeria, Kurdistan, Malaysia, Palestine and Luxembourg). The result from the bioinformatics analysis showed a significant number of SNPs in the ITS region. There are 40 SNPs found in the ITS sequences and 35 SNPs in the NADH1sequences of the mtDNA. The phylogenic analysis of the result revealed the phylogenetic relationship and a distinct genetic difference between the different nationalities that may be used as a marker for human genetic identification in the future.

ABSTRAK

Ujian identiti genetik untuk manusia telah digunakan untuk mengkaji variasi dalam kawasan polimorfik DNA manusia. Kaedah-kaedah ini termasuk penanda RFLP (Sekatan serpihan panjang polimorfisme), penanda lokus STR (pengulangan selaras pendek) yang paling biasa digunakan dalam pengenalpastian lokus manusia serta kromosom mtDNA atau Y dalam perubatan forensik dan ujian paterniti. Sekarang, rantau salinan peruang dalaman (ITS) digunakan untuk menjalankan analisis evolusi bagi pelbagai spesies haiwan, tumbuhan, kulat, yis. Walau bagaimanapun, rantau ITS belum lagi digunakan dalam ujian identiti genetik manusia. Dalam kajian ini, perbandingan genetik bagi populasi manusia yang berbeza dengan menggunakan jujukan ITS mtDNA telah dilakukan. Dua segmen rantau mtDNA manusia telah dipilih dan dijujukkan dalam usaha untuk menentukan kewujudan polimorfisme nukleotida tunggal (SNP) yang mencukupi dan kesesuaiannya untuk ujian identiti genetik dalam manusia. Primers khusus telah direka untuk mengamplifikasikan rantau ITS dengan menggunakan PCR untuk mengekstrak sampel darah pelajar dari Fakulti Biosains dan Biokejuruteraan, UTM yang terdiri daripada kewarganegaraan yang berbeza (Iran, Nigeria, Kurdistan, Malaysia, Palestin dan Luxembourg).Hasil daripada analisis bioinformatik menunjukkan terdapat jumlah tererti SNPs di rantau ITS. Terdapat 40 SNPs ditemui di jujukan ITS dan 35 SNPs dalam jujukan NADH1 mtDNA. Keputusan analisis filogenetik memperlihatkan hubungan filogeni dan perbezaan genetik yang ketara antara bangsa yang berbeza yang boleh digunakan sebagai penanda bagi pengenalan genetik manusia pada masa depan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

µl	-	Microliter
µM	-	Micromole
µg	-	Microgram
ng	-	Nanogram
bp	-	Base pair
◦C	-	Degree Centigrade Celsius
conc.	-	Concentration
D.W	-	Distilled water
DNA	-	Deoxyribonucleic acid
dNTPs	-	Deoxnucleotide triphosphate
EDTA	-	Ethylenediaminetetraacetic Acid
Fbb	-	Faculty of Biosciences and Bioengineering
For	-	Forward
g	-	Gram
hr	-	hour
ID	-	Identification Number
ITS1	-	Internal TRANSCRIBED spacer
ml	-	Milliliter
MW	-	Molecular Weight
mtDNA	-	Mitochondrial DNA
min	-	Minute
mg	-	Milligram
N.A.	-	Nothing at all
Nb	-	Number
nm	-	Monometer

NCBI	-	National Center for Biotechnology Information
N.C.	-	Negative Control
No.	-	Number
OD	-	Optical Density
OXPHOS	-	Oxidative Phosphorylation
PCR	-	Polymerase chain reaction
PID	-	Percentage Identity between the two sequences at each aligned position.
RNase	-	ribonucleases
Rev	-	Reveres
rpm	-	Revolution Per minute
Sec	-	Second
TAE	-	Tris-acetate EDTA
Tm	-	Melting temperature
tRNA	-	transfer RNA
UTM	-	Universiti Technology Malaysia
UV	-	Ultra Violet

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Genetic identity testing for humans has been used to examine the variations in the polymorphic regions of the human DNA. These methods use unique genetic markers such as the RFLP (Restriction fragment length polymorphisms) ,the STR (short tandem repeats) ,the mtDNA or Y chromosome for human identification, forensic medicine and paternity testing(Goodwin, Linacre, & Hadi, 2011).

The ribosomal DNA (rDNA) contains unique tandem repeats has been proposed as a new DNA fingerprinting source for human genetic identification. There are five chromosomes in the human genome that contain the rDNA chromosome: 13, 14, 15, 21and 22. The ribosomal DNA consists of NTS (non-transcript spacer), 18S, ITS1(internal transcribed spacer), 5.8S, ITS2 and 28S. The coding region with the gene is in 18S, 5.8S and 28S. The ITS(internal transcripted spacer)found between 18S, 5.8S regions and between 5.8S, 28S regions (Nei & Rooney, 2005).

Currently, the internal transcribed spacer (ITS) region has been used for studying the phylogenetic analyses of the evolution of tandemly arranged genes in different species (animals, plants, fungi and yeast) but has not yet been used in genetic identity testing in humans (Abrahams & Benjeddou, 2009).

1.2 Problem statement

Currently, the ITS region has not yet being used in genetic identity test in human. The tandemly repeated genes of ITS in rDNA are suitable candidate loci for molecular evolutionary marker studies because of their universal presence, high copy number and functional similarity. The presence of short repeated motif, high GC content and a higher change rate than other gene variable region make this an attractive option compared to existing identification markers (Blouin, 2002).

1.3 The aim of the study

This study investigates whether the variations of the ITS region of mtDNA can be used for human genetic identity testing, forensic medicine or paternity analysis. Different samples sourced from UTM FBB students were for this purpose. Students from different ethnic groups were used as examples to test this hypothesis. The sequencing and analysis of ITS of rDNA were used to construct the phylogenetic tree of the different UTM FBB

students to determine the relationship between groups of different populations. The aim of this study was to determine if the variation or polymorphic loci in ITS is sufficient to differentiate between different population groups and to detect possible sequence motifs associated with different ethnic groups.

1.4 Objectives

- To detect all the genome variations and the conserved regions of the mtDNA from existing genomic databases.
- To isolate and purify the maximum mtDNA from the whole blood sample from representations from different ethnic groups (Malay, Chinese, Nigerian, Palestine, Iranian and Luxemburg).
- To determine the most efficient and the best primers for successful amplification of region of interest in human mtDNA.
- To analyze the genetic variations using multiple sequence alignment and phylogenetic tree between the different samples of population group.

1.5 Significance of Study

The ITS region with the polymorphic loci or the variation in mtDNA in human can become a useful marker as the variability of the ITS region of rDNA that are due to

mutation, deletion or insertion in mtDNA(Gonzalez, Sylvester, Smith, Stambolian, & Schmickel, 1990) should be sufficient for the use as genetic identification. The ITS region has not been used before for human genetic identity and success in this project would create a potential novel application in human genetic identification and paternity analysis.

REFERENCES

Abrahams, Z., & Benjeddou, M. (2009). The value of Non-CODIS miniSTR genotyping systems in forensic casework in South Africa. *Afr. J. Biotechnol, 8*(11), 2432-2436.

Akane, A., Matsubara, K., Shiono, H., Yuasa, I., Yokota, S., Yamada, M., et al. (1990). Paternity testing: blood group systems and DNA analysis by variable number of tandem repeat markers. *Journal of forensic sciences, 35*(5), 1217.

Blouin, M. S. (2002). Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer. *International journal for parasitology, 32*(5), 527-531.

Brown, J. (2004). *dna fingerprinting.* worcester polytechnic institute.

Budowle, B., Allard, M. W., Wilson, M. R., & Chakraborty, R. (2003). Forensics and Mitochondrial DNA: Applications, Debates, and Foundations*. *Annual Review of Genomics and Human Genetics, 4*(1), 119-141.

Butler, J. M. (2005). *Forensic DNA typing: biology, technology, and genetics of STR markers*: Academic Press.

Caburet, S., Conti, C., Schurra, C., Lebofsky, R., Edelstein, S. J., & Bensimon, A. (2005). Human ribosomal RNA gene arrays display a broad range of palindromic structures. *Genome research, 15*(8), 1079-1085.

Cavelier, L., Jazin, E., Jalonen, P., & Gyllensten, U. (2000). MtDNA substitution rate and segregation of heteroplasmy in coding and noncoding regions. *Human genetics, 107*(1), 45-50.

Coble, M. D., Just, R. S., O'Callaghan, J. E., Letmanyi, I. H., Peterson, C. T., Irwin, J. A., et al. (2004). Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. *International Journal of Legal Medicine, 118*(3), 137-146.

Cui, C., & Tseng, H. (2004). Estimation of ribosomal RNA transcription rate in situ. *BioTechniques, 36*(1), 134-138.

Erickson, J., Rushford, C. L., Dorney, D., Wilson, G. N., & Schmickel, R. D. (1981). Structure and variation of human ribosomal DNA: molecular analysis of cloned fragments.

Fan, L., & Yao, Y. G. (2011). MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations. *Mitochondrion, 11*(2), 351-356.

Giles, R. E., Blanc, H., Cann, H. M., & Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. *Proceedings of the National academy of Sciences, 77*(11), 6715.

Gonzalez, I., Sylvester, J., Smith, T., Stambolian, D., & Schmickel, R. (1990). Ribosomal RNA gene sequences and hominoid phylogeny. *Molecular biology and evolution, 7*(3), 203-219.

Goodwin, W., Linacre, A., & Hadi, S. (2011). *An introduction to forensic genetics* (vol. 2): Wiley.

Grzybowski, T. (2000). Extremely high levels of human mitochondrial DNA heteroplasmy in single hair roots. *Electrophoresis, 21*(3), 548-553.

Hall, B. G. (2004). *Phylogenetic trees made easy: a how-to manual*: Sinauer Associates Sunderland, Massachusetts.

Hillis, D. M., & Dixon, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biology*, 411-453.

Howell, N., Smejkal, C. B., Mackey, D., Chinnery, P., Turnbull, D., & Herrnstadt, C. (2003). The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *The American Journal of Human Genetics, 72*(3), 659-670.

Ingman, M., & Gyllensten, U. (2001). Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. *Journal of Heredity, 92*(6), 454-461.

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Hypervariable'minisatellite' regions in human DNA. *Nature, 314*(6006), 67-73.

Jorde, L. B., Watkins, W., Bamshad, M., Dixon, M., Ricker, C., Seielstad, M., et al. (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *The American Journal of Human Genetics, 66*(3), 979-988.

Linch, C. A., Smith, S. L., & Prahlow, J. A. (1998). Evaluation of the human hair root for DNA typing subsequent to microscopic comparison. *J. Forensic Sci, 43*(2), 305-314.

McLennan, D. A. (2010). How to read a phylogenetic tree. *Evolution: Education and Outreach, 3*(4), 506-519.

Neckelmann, N., Li, K., Wade, R. P., Shuster, R., & Wallace, D. C. (1987). cDNA sequence of a human skeletal muscle ADP/ATP translocator: lack of a leader peptide, divergence from a fibroblast translocator cDNA, and coevolution with mitochondrial DNA genes. *Proceedings of the National academy of Sciences, 84*(21), 7580.

Nei, M., & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annual review of genetics, 39*, 121.

Otlowski, M., Barlow-Stewart, K., Taylor, S., Stranger, M., & Treloar, S. (2007). Investigating genetic discrimination in the Australian life insurance sector: the use of genetic test results in underwriting, 1999-2003. *Journal of law and medicine, 14*(3), 367.

Pakendorf, B., & Stoneking, M. (2005). Mitochondrial DNA and human evolution. *Annu. Rev. Genomics Hum. Genet., 6*, 165-183.

Pakstis, A. J., Speed, W. C., Kidd, J. R., & Kidd, K. K. (2007). Candidate SNPs for a universal individual identification panel. *Human genetics, 121*(3), 305-317.

Rosa, A., Fonseca, B. V., Krug, T., Manso, H., Gouveia, L., Albergaria, I., et al. (2008). Mitochondrial haplogroup H1 is protective for ischemic stroke in Portuguese patients. *BMC medical genetics, 9*(1), 57.

Shumaker, J. M., Metspalu, A., & Caskey, C. T. (1996). Mutation detection by solid phase primer extension. *Human mutation, 7*(4), 346-354.

Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S. T. (2003). Human gene mutation database (HGMD®): 2003 update. *Human mutation, 21*(6), 577-581.

Stoneking, M., Hedgecock, D., Higuchi, R. G., Vigilant, L., & Erlich, H. A. (1991). Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *American journal of human genetics, 48*(2), 370.

Sylvester, J., Whiteman, D., Podolsky, R., Pozsgay, J. M., Respess, J., & Schmickel, R. (1986). The human ribosomal RNA genes: structure and organization of the complete repeating unit. *Human genetics, 73*(3), 193-198.

Tritschler, H. J., & Medori, R. (1993). Mitochondrial DNA alterations as a source of human disorders. *Neurology, 43*(2), 280-280.

Tully, L. A., Parsons, T. J., Steighner, R. J., Holland, M. M., Marino, M. A., & Prenger, V. L. (2000). A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. *The American Journal of Human Genetics, 67*(2), 432-443.

Wallace, D. C. (1995). 1994 William Allan Award Address. Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *American journal of human genetics, 57*(2), 201.

Wallace, D. C. (2007). Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annu. Rev. Biochem., 76*, 781-821.

Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science, 280*(5366), 1077-1082.

Wang, P., Liu, Y., Yin, Y., Jin, H., Wang, S., Xu, F., et al. (2011). Diversity of microorganisms isolated from the soil sample surround Chroogomphus rutilus in the Beijing region. *International journal of biological sciences, 7*(2), 209.

Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., et al. (2001). The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proceedings of the National academy of Sciences, 98*(18), 10244-10249.

Wilson, G. N., Knoller, M., Szura, L. L., & Schmickel, R. D. (1984). Individual and evolutionary variation of primate ribosomal DNA transcription initiation regions. *Molecular biology and evolution, 1*(3), 221-237.

Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J., et al. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science, 292*(5518), 883-896.