

**ANALYSIS OF MtDNA AND Y CHROMOSOME
HAPLOGROUPS OF KHAN YOUNIS-GAZZA POPULATION IN
RELATION TO GENETIC DISEASES AND SOCIAL BEHAVIOR**

**A thesis submitted
by
Rana Farouk El Farra
B.S. (1994) & M.S. (1999)**

For

The Degree of Ph.D. in Biochemistry

**Department of Chemistry
Faculty of Science
Cairo University
June 2008**

Copyright © Rana F. El Farra
All Rights Reserved

CAIRO UNIVERSITY

APPROVAL SHEET FOR SUBMISSION

of a thesis submitted by

Rana Farouk El Farra

Title of thesis: ANALYSIS OF MtDNA AND Y CHROMOSOME
HAPLOGROUPS OF KHAN YOUNIS-GAZZA POPULATION IN
RELATION TO GENETIC DISEASES AND SOCIAL BEHAVIOR

This thesis has been approved by for submission by the supervisors:

.....
Date

.....
Prof. Dr. Hamdy Hassanain, Chair

.....
Date

.....
Prof. Dr. Sayed Goudah

.....
Date

.....
Dr. Abdalla Abed

Prof. Dr. Rifaat H. Hilal

.....
Chairman of Chemistry
Department, Faculty of
Science, Cairo University

ABSTARCT

ANALYSIS OF MtDNA AND Y CHROMOSOME HAPLOGROUPS OF KHAN YOUNIS-GAZZA POPULATION IN RELATION TO GENETIC DISEASES AND SOCIAL BEHAVIOR

Rana Farouk El Farra

Chemistry Department, Faculty of Science

Doctor of Philosophy

It is possible to identify certain mitochondrial DNA and Y chromosome haplogroups within a local population of Khan Younis in Palestine and correlate these haplogroups with the tribal affiliations, histories, genetic disorders, marriage and migration patterns. This study considers ten local tribal affiliations from Khan Younis. Other tribal affiliations from Gaza City, Yafa, Egypt and Jordan are included for paired comparisons. Endogamy, described as marriage within original tribal clan or tribes should enrich certain haplotypes within the tribal groups. Certain lineages and tribal affiliations can be correlated with mtDNA haplogroups determined by sequencing HVRI and with Y chromosome STRs haplogroups by multiplex primer systems. The single mutation cladistic analyses of both genetic markers as well as the monophylic origin likelihood ratio test suggest a common ancestor for this population. Comparing the Y chromosome STRs haplogroups to literature cited ME populations deciphered common migration patterns. Maternal age-

related genetic disorders could be correlated to mtDNA haplogroups. The level of mtDNA polymorphism and nucleotide diversity of this population is comparable to other limited and isolated populations.

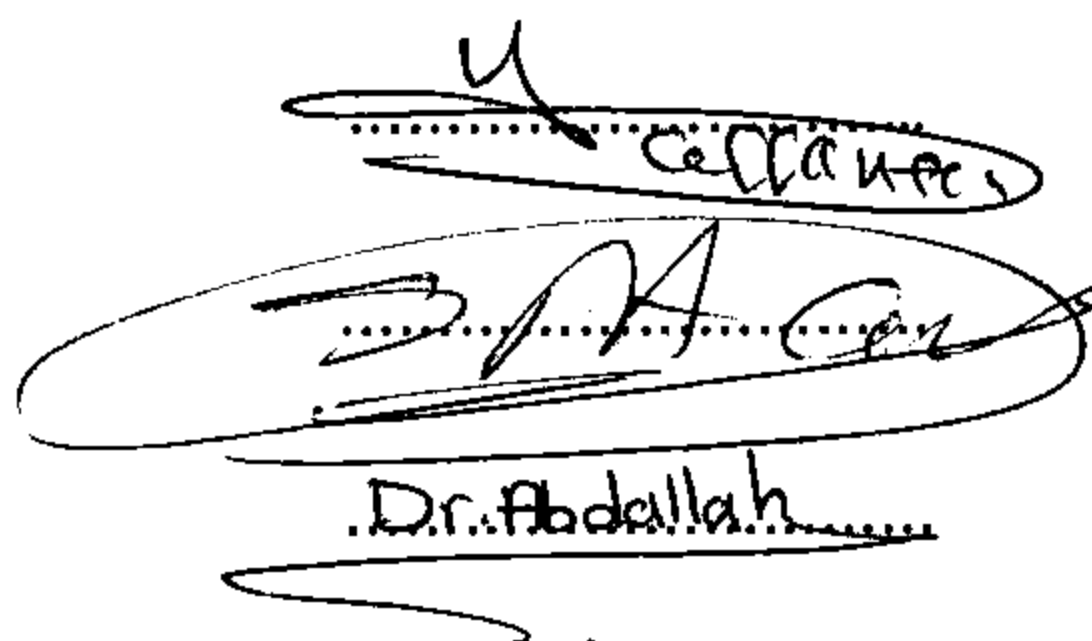
Key words: MtDNA, Haplogroups, Endogamy, HVRI, Y Chromosome STRs, Monophylic Origins, Polymorphism.

Supervisors:

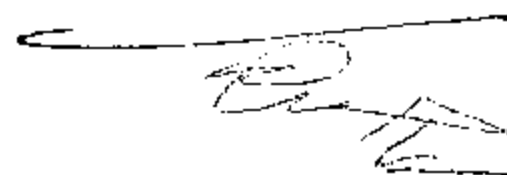
Prof. Dr. Hamdy M. Hassanin

Pro. Dr. Sayed Goudah

Dr. Abdalla Abed



Prof. Dr. Rifaat H. Hilal



Chairman of Chemistry Department,
Faculty of Science,
Cairo University

ACKNOWLEDGEMENTS

I would like to thank all of those who helped to make this dream come true. I would like to thank my mother for showing me the light at the end of a dark tunnel. I would like to thank my daughter Salma for her love and support during my Ph.D. journey. I want to thank my sisters Rola and Rihaf and my brothers Dr. Hamdi and Ahmed for their support. My deep gratitude goes to **Prof. Dr. Hamdy M. Hassanin**. Only with his help and support this work saw the light. I also would like to thank **Dr. Sayed Goudah** for his support. I appreciate Cairo University for allowing me this opportunity to learn and further expand my horizons.

I am grateful to **Dr. Manal El Hamshary** and **Dr. Nadim El Kharouf** for their endless support and guidance. They have been very sincere friends and colleagues. Many thanks are to **Dr. Abdalla Abed** and **Dr. Ahmed Salem** for their valuable support.

LIST OF ABBREVIATIONS

Mt DNA	Mitochondrial DNA
ME	Middle East
STRs	Short Tandem Repeats
D-loop	Displacement Loop
NCR	Non-Coding Region
bp	Base pair
RFLPs	Restriction Fragment Length Polymorphisms
CR	Control Region
Myr	Million Year
HVR-I	Hypervariable Region-I
NYR	Non-recombining Y Chromosome Region
PARs	Pseudoautosomal Regions
SRY	Sex-Determining Region on Y gene
AZF	Azoospermia Factors
DYS	DNA Y Chromosome Segment
SNPs	Single Nucleotide Polymorphisms
HLA	Human Leukocyte Antigen
UEP	Unique Event Polymorphism
CMH	Cohen Modal Haplotype
MH	Modal Haplotype
CE	Common Era
PAUP	Phylogenetic Analysis Using Parsimony
UPGMA	Unweighted Pair Group Method, Arithmetic Method
NJ	Neighbor Joining
JC model	Jukes-Cantor model
HKY+G	M. Hasegawa, H. Kishino, T.Yano plus Gamma
PCR	Polymerase Chain Reaction
DNA	Deoxy Ribonucleic Acid
RNA	Ribonucleic Acid
ds DNA	Double Stranded DNA
ss DNA	Single Stranded DNA
ddNTPs	Dideoxy Nucleoside Triphosphates
TMR	Carboxy Tertamethyl Rhodamine
JOE	6-carboxyl-4', 5'-dichloro-2',7'-dimethoxy Fluorescein
FL	Fluorescein

TABLE OF CONTENTS

ABSTRACT.....	IV
ACKNOWLEDGMENTS.....	VI
TABLE OF CONTENTS.....	VIII
SUMMARY.....	XV
INTRODUCTION.....	1
BACKGROUND & LITERATURE REVIEW.....	4
MIDDLE EAST: HISTORY AND GENEALOGIES.....	4
SOCIAL AND FAMILIAL AFFILIATION OF KHAN YOUNIS.....	5
MITOCHONDRIAL DNA.....	8
MTDNA GENETICS.....	8
D-LOOP STRUCTURAL ORGANIZATION.....	10
MITOCHONDRIAL DNA SOMATIC VS. GERM-LINE MUTATION... 	10
NEUTRAL MTDNA VARIATION AND HUMAN MTDNA EVOLUTION.....	12
HETEROPLASMY AND MITOCHONDRIAL DNA.....	13
MTDNA RECOMBINATION.....	14
MITOCHONRIAL DNA DIVERGENCE RATE.....	16
MITOCHONRIAL DNA DISORDERS.....	17
Y CHROMOSOME GENETICS.....	19
Y CHROMOSOME HAPLOGROUP DIVERSITY.....	20
Y CHROMOSOME MIDDLE EAST MARKERS AND HAPLOGROUPS.....	23
LEVANT POPULATION.....	23
NORTH AFRICA, OMAN, YEMEN AND ARABIAN PENINSULA.....	28
Y CHROMOSOME VS. MTDNA HAPLOGROUP DIVESITY.....	29
DATA ANALYSIS METHODS.....	35
POPULATION GENETICS.....	35
ALIGNMENT.....	36
CONSENSUS SEQUENCE AND ANCESTRAL STATE	

RECONSTRUCTION.....	37
OPTIMALITY CRITERIA AND MODEL OF EVOLUTION.....	37
ESTIMATION OF MODEL PARAMETERS.....	39
PHYLOGENETIC TREES.....	39
MTDNA SEQUENCE DATA AND PHYLOGENETICS INFERENCE	
CONFEDENCE.....	41
SEARCHING FOR OPTIMAL TREES.....	43
HYPOTHESIS TESTING IN A PHYLOGENETIC FRAMEWORK.....	43
COMPARING TWO TREES.....	44
TESTING FOR A MOLECULAR CLOCK.....	44
TREE ROOTING.....	45
TREE CALIBRATION.....	45
ESTIMATION OF COALESCENCE PARAMETERS.....	46
<i>Neutrality Theory</i>	46
<i>Effective Population Size</i>	46
<i>Coalescence Process</i>	47
<i>Coalescence Parameters</i>	47
<i>Detection of Gene Flow From Alleles Phylogeny</i>	48
<i>Cladogram Estimation</i>	49
TECHNIQUES.....	51
MITOCHONRIAL DNA ANALYSIS.....	51
POLYMERASE CHAIN REACTION.....	51
AUTOMATED DNA SEQUENCING.....	52
 MATERIALS AND	
METHODS.....	53
 SAMPLE DESCRIPTION.....	53
DNA	
EXTRACTION.....	56
<i>Buccal Cell Collection And Cell Lysis</i>	56
<i>Protein Precipitation</i>	57
<i>DNA Precipitation</i>	57
<i>DNA Hydration</i>	58
PCR AMPLIFICATION.....	58
<i>Primers</i>	58
<i>Primer Set Modification</i>	59
<i>Description of PCR Cycle</i>	59
<i>PCR Conditions</i>	59
<i>PCR Product Clean Up Protocol</i>	59
AUTOMATED SEQUENCING.....	60
<i>PCR For Sequencing</i>	60

<i>Purification Of the PCR Product</i>	60
<i>Sequencing PCR</i>	61
<i>Sephadex Spin Column Purification</i>	61
<i>Electrophoresis</i>	61
<i>Sequencher Program</i>	62
BIOINFORMATICS	62
<i>ClustalW/TCoffee Programs</i>	62
<i>PAUP</i>	62
<i>Absolute Distance Matrix</i>	63
<i>NJ/UPGMA Tree</i>	63
<i>Minimum Evolution Tree</i>	63
<i>Consensus Of Minimum Evolution Trees</i>	64
<i>Bootstrap Values</i>	64
<i>Obtaining Monophylic Trees</i>	64
<i>Monophylic Vs. Polyphylic Tribal Origins</i>	65
<i>Detection of Gene Flow (Number of Migrants)</i>	65
<i>Molecular Clock</i>	65
<i>Association Of Geographic Location With Haplogroups</i>	66
<i>Calculations of Theta θ Population Diversity</i>	66
Y CHROMOSOME STR TESTING	66
<i>STR Multiplex amplification Setup</i>	66
AMPLIFICATION THERMAL CYCLING	67
<i>9700 Thermocycler Amplification Parameters</i>	67
Y CHROMOSOME STR MULTIPLEX PRIMERS DESCRIPTION AND SEQUENCE	67
RESULTS	72
DATA SAMPLE DESCRIPTION	72
MITOCHONDRIAL DNA ANALYSIS	78
<i>Polymorphic Sites</i>	78
<i>Cladistic Analysis</i>	87
<i>Model Test Scores</i>	93
<i>Neighbor Joining Tree</i>	93
<i>Minimum Evolution Trees And Consensus</i>	95
<i>NJ Boot Sharp Values</i>	96
<i>Monophyly Vs. Polyphyly Likelihood Ratio Test</i>	96
<i>Geographic Association Test</i>	97
<i>Molecular Clock Test</i>	97
<i>THETA (θ) Values Or Population Diversity</i>	98
<i>Estimation Of Gene Flow By Nm</i>	98
<i>Maternal Genetic Disorders Correlations</i>	99
Y CHROMOSOME STRS ANALYSIS	101
<i>Y Chromosome STRs</i>	101

<i>Cladistic Analysis</i>	104
<i>Paternal Geographic Origin Correlations</i>	105
 DISCUSSION	 108
 MITOCHONDRIAL DNA	 108
Y CHROMOSOME	125
 CONCLUSIONS	 132
 LITERATURE CITED	 135
 APPENDIX I: ANCIENT HISTORY OF PALESTINE	 151
5500-2300 B.C.E.	151
2300-2000 B.C.E.	152
2000-1600 B.C.E.	154
1600-1100 B.C.E.	155
1100-626 B.C.E.	156
 APPENDIX II: KHAN YOUNIS: A GAZA STRIP HISTORICAL PERSPECTIVE	 160
 GAZA: TOWNS AND VILLAGES	 162
 APPENDIX III-a: THOSE WHO LIVED IN KHAN YOUNIS	 164
EL AGHA FAMILY	164
EL JUBUR FAMILY	164
EL ASTAI, FAMILY	165
EL AQQAD FAMILY	165
EL FARRA FAMILY	166
SHURRAB FAMILY	166
EL MEZAYYEN FAMILY	167
SHAAT FAMILY	167
FASFOUS FAMILY	168
EL SAQQA FAMILY	168
EL BALBEESY FAMILY	168

EL HELOU FAMILY.....	169
EL ASHI FAMILY.....	169
 APPENDIX III-b: FAMILY TREES OF SOME TRIBES THAT LIVE IN KHAN YOUNIS.....	 170
SHURRAB FAMILY TREE.....	170
EL JUBUR FAMILY TREE.....	171
EL SAQQA FAMILY TREE.....	172
SHIAÁT FAMILY TREE.....	173
EL FARRA FAMILY TREE.....	174
 APPENDIX IV: BIOLOGICAL FAMILY TREE.....	 175
 APPENDIX V: MTDNA HVRI OF 19 HAPLOGROUPS CLUSTAL W EDITED SEQUENCES.....	 176
 APPENDIX VI: HISTORICAL MIGRATION PATTERNS MAPS OF LEVANT.....	 181
THE LEVANT AND SURROUNDINGS.....	181
THE LATE ROMAN LEVANT, 200-600.....	182
THE BYZANTINE/FATIMID STANDOFF, 970-1070.....	183
MAMLUK SUPREMACY, 1260-1516.....	184
THE OTTOMAN LEVANT, EARLY SEVENTEENTH CENTURY.....	185
THE CONTEMPORARY LEVANT, 2000.....	186
 APPENDIX VII: KHAN YOUNIS Y CHROMOSOME STR DATA.....	 187
 APPENDIX IIX: KHAN YOUNIS SAMPLE PATERNAL AND MATERNAL GENEALOGIES.....	 188

LIST OF TABLES

Table 1: A summary of population under study tribal affiliations and geographic locations.....	76
Table 2: Summary of Paternal Tribal Affiliation, sample numbers, maternal tribal affiliation origins and geographic location.....	77-78
Table 3: Summary of sample numbers, haplogroups, paternal tribal affiliation and maternal tribal affiliations.....	83-84
Table 4: A summary of polymorphic sites on HVR-I of D-loop as compared to Anderson's reference sequence.....	85-86
Table 5: Position and description of polymorphic sites of haplogroups 18, 19 as compared to the highly polymorphic haplogroup 17.....	87
Table 6: A summary of haplogroups numbers and the age-related maternally inherited genetic disease.....	100
Table 7: Y chromosome STRs of a selected local Khan Younis population.....	103
Table 8: Y chromosome literature cited several populations in the Middle East.....	103
Table 9: A summary to correlate Y chromosome STRs haplogroups to the clan/tribal affiliation.....	104
Table 10: Khan Younis Y chromosome STRs data.....	187
Table 11: Khan Younis paternal and maternal four generations genealogies.....	188

LIST OF FIGURES

Fig. 1	Map of study locations.....	2
Fig. 2	Map of the Middle East.....	5
Fig. 3	Cladogram constructed using minimum number of mutations of the collapsed 19 haplogroups.....	90
Fig. 4	Cladogram constructed using the collapsed 19 haplogroups illustrating maternal tribal affiliations.....	91
Fig. 5	A cladogram constructed using nineteen haplogroups illustrating maternal origin geographic location.....	92
Fig. 6	NJ tree of 19 haplogroup.....	94
Fig. 7	50% majority rule consensus tree.....	96
Fig. 8	Y chromosome STR cladogram of twenty five individuals of Khan Younis local population.....	107
Fig. 9	Shurrah Family Tree.....	170
Fig. 10	El Jubur Family Tree.....	171
Fig. 11	El Saqqa Family Tree.....	172
Fig. 12	Shaât Family Tree.....	173
Fig. 13	El Farra Family Tree.....	174
Fig. 14	The Levant and Surroundings.....	181
Fig. 15	The Late Roman Levant.....	182
Fig. 16	The Byzantine/Fatimid Standoff.....	183
Fig. 17	Mamluk Supremacy.....	184
Fig. 18	The Ottoman Levant, Early Seventeenth Century.....	185
Fig. 19	The Contemporary Levant, 2000.....	186

SUMMARY

Name: Rana Farouk El Farra

Nationality: Palestine

Date of Birth: June 22nd, 1970

Degree: Doctor of Philosophy

Specialization: Biochemistry

Supervisors: 1. Prof. Handy Hassanin

2. Prof. Dr. Sayed Goudah

3. Dr. Abdallah Abed

Title of thesis: ANALYSIS OF mtDNA AND Y CHROMOSOME
HAPLOGROUPS OF KHAN YOUNIS-GAZZA POPULATION IN
RELATION TO GENETIC DISEASES AND SOCIAL BEHAVIOR

This thesis is a preliminary investigation to correlate marriage and migration patterns, history, tribal affiliations and genetic disorders with Y chromosome STRs patterns and mtDNA HVR-I markers of a local population of Khan Younis City in Gaza Strip-Palestine. A total number of 91 samples were collected and analyzed using the PCR and automated sequencing to obtain genetic data for mtDNA HVR-I. Twenty five male individuals were analyzed using Y chromosome STRs by applying PCR multiplex systems.

The results show that it is possible to identify both mtDNA haplogroups and Y chromosome STR haplogroups within the population under study. The study considers ten tribal affiliations including: El Farra, Shurrab, Pastous, El Mezayyen, El Agha, El Saqqa, El Aqqad, Abdel Bari and El Jubur.