The Malaysian Human Genome Variation Project

Wan-Hatin WI, Nur-Shafawati AR, **Zilfalil BA**, 1Malaysia Human Genome Variation Consortium.

Introduction

Malaysia is a multi cultural, multi religious and multi ethnic country comprising of numerous ethnic groups. Malays form the major ethnic group and can be further subdivided into various sub-groups. The genome wide mapping of these ethnic groups for the creation of genetic database will provide the genetic map of the country migration history and uncover the extent of genetic similarity between the various populations along with the study of common diseases in the Malaysian populations.

Methodology

The SNP multilocus genotype data of the Malay subethnic groups were obtained through the use of the Affymtetrix 50K microarray chips and were analysed together with those of 12 other populations from Thailand, Indonesia, China, India, Africa and Orang Asli sub groups in Peninsula Malaysia obtained from the Pan-Asian SNP (PASNP) database. Phylogenetic analyses of all 54,794 autosomal SNPs were implemented by Neighbor Joining algorithm. Bayesian clustering procedure using STRUCTURE software was performed to examine the ancestry of all 472 individuals.

Results

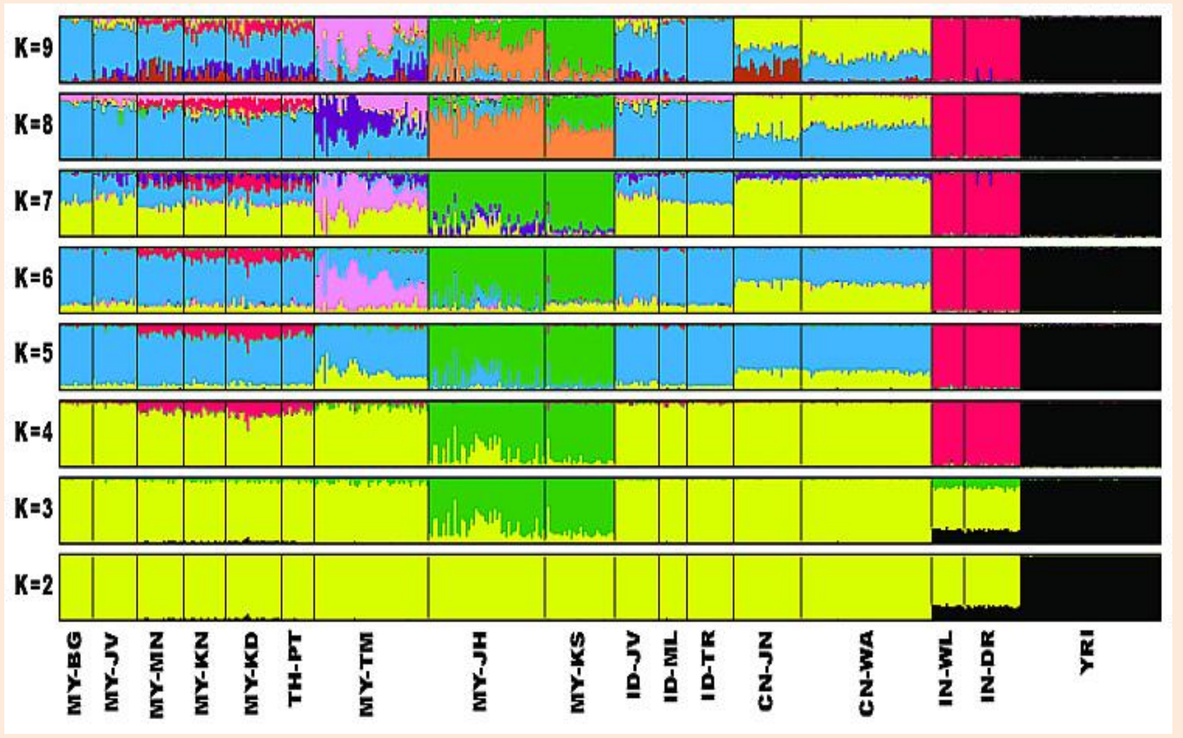
The phylogenetic analyses showed the Malay sub ethnic groups are separated into at least 3 different clusters. The Jawa, Bugis and Minang have a very close genetic relationship with Indonesian populations indicating a common ancestral origin, while the Kelantan and Kedah formed a distinct group on the tree indicating that they are genetically different from the other Malay sub ethnic groups. From our results, we could also conclude that while the Malays belong to the Southern Mongoloid group of races, there are evidence of gene flow from the Indians to Malays.

Conclusion

Our results have provided information on the genetic differentiation between the sub ethnic groups and valuable insight into the origin of the Malay sub ethnic groups in peninsula Malaysia. The created database on these populations will complement the HapMap database and contribute valuable information on the genetic background of the South-East Asian population.

Acknowledgements

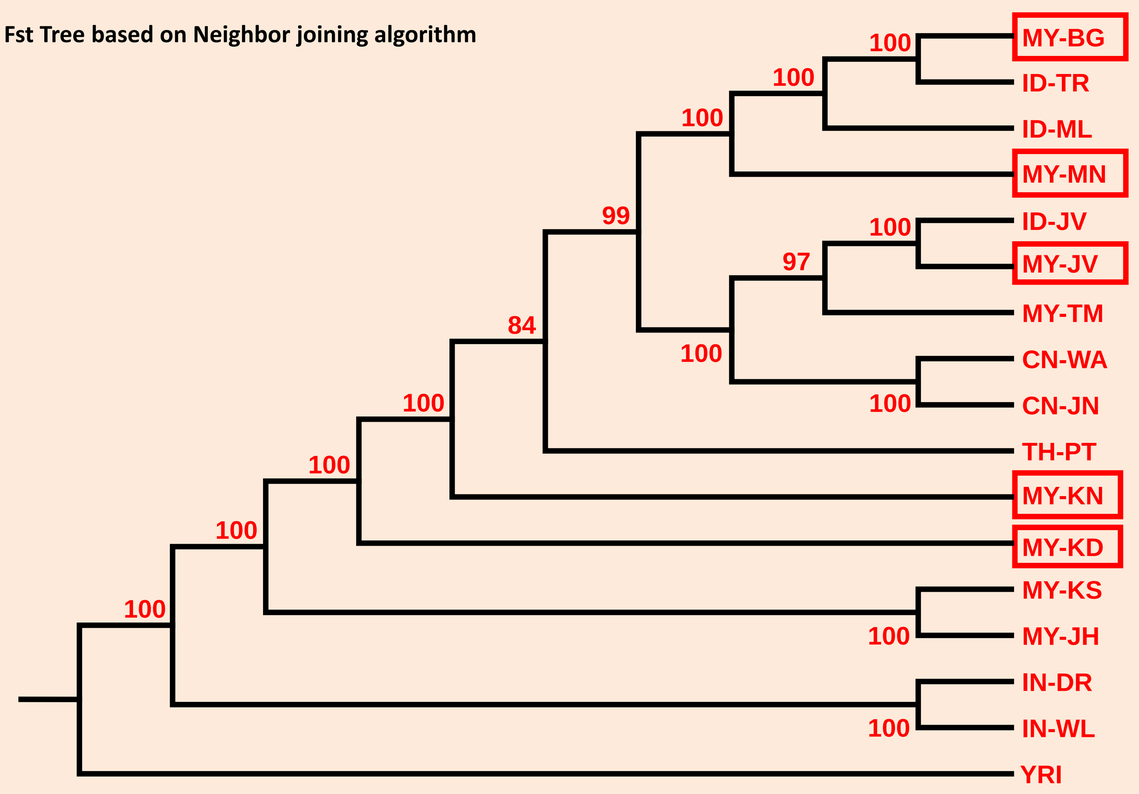
This research was supported by Fundamental Research Grants Scheme (FRGS) project 203/PPSP/6170025 and USM Fellowship Scheme. We acknowledge the contributions by the other members of this study group from the School of Health Sciences and the School of Dental Sciences, Universiti Sains Malaysia. We thank all the subjects who have participated in this research and those who have helped us in the data collection. Special thanks to the UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia and Matrix Analytical Sdn. Bhd., Malaysia for allowing us to use their laboratory facilities; CAS-MPG Partner Institute Computational Biology, Fudan University, Shanghai, China for helping us with the data analysis; PASNPI Consortium for allowing us to use the genotype data.



Proportion of membership coefficient (Q) of each pre-defined population in each of the 6 clusters from STRUCTURE.

COUNTRY & POPULATION (NO. OF SUBJECTS)	SAMPLE ID	CLUSTER (K)					
		MALAYS & INDONESIANS	INDIANS	CHINESE	AFRICAN	NEGRITOS	PROTO-MALAYS
MALAYS:							
JAWA (19)	MY-JV	**0.7954**	0.0094	0.1444	0.0006	0.0164	0.0338
BUGIS (14)	MY-BG	**0.8605**	0.0086	0.1137	0.001	0.0064	0.0098
MINANG (20)	MY-MN	**0.7558**	0.1202	0.0948	0.002	0.0114	0.0158
KELANTAN (18)	MY-KN	**0.6701**	0.1546	0.1133	0.002	0.0226	0.0374
KEDAH (24)	MY-KD	**0.6454**	0.2108	0.1033	0.0016	0.0142	0.0246
PROTO-MALAY*:							
TEMUAN (49)	MY-TM	0.3922	0.0086	0.0724	0.0008	0.0222	**0.5038**
NEGRITOS*:							
JAHAI (50)	MY-JH	0.1157	0.0028	0.0554	0.0006	**0.8094**	0.0162
KENSUI (30)	MY-KS	0.0312	0.013	0.1269	0.0006	**0.8181**	0.0102
THAILAND:							
TH-PT (14)		**0.7093**	0.1401	0.0991	0.0014	0.0188	0.0314
INDONESIA*:							
JAVA:							
JAWA (19)	ID-JV	**0.7845**	0.0074	0.1555	0.001	0.017	0.0346
SUMATRA:							
MELAYU (12)	ID-ML	**0.8567**	0.0198	0.1095	0.0022	0.006	0.0058
SULAWESI:							
TORAJA (20)	ID-TR	**0.8736**	0.0074	0.11	0.0008	0.005	0.0032
CHINA*:							
JINUO (29)	CN-JN	0.5173	0.003	**0.4699**	0.0006	0.0044	0.0048
WA (56)	CN-WA	0.5486	0.004	**0.4284**	0.001	0.0074	0.0106
INDIA*:							
LANGUAGE:							
MARATHI (14)	IN-WL	0.0058	**0.9772**	0.0042	0.0048	0.0044	0.0036
TELUGU (24)	IN-DR	0.0062	**0.9774**	0.0042	0.0036	0.0056	0.003
AFRICA*:							
YORUBA (60)	YRI	0.0016	0.0028	0.0018	**0.9902**	0.002	0.0016

Fst Tree based on Neighbor joining algorithm



Contact us:

Head, 1 Malaysia Human Genome Variation Consortium
Associate Professor Dr. Zilfalil Bin Alwi,
email: zilfalil@kb.usm.my

Secretariat,
Malaysia Human Genome Variation Consortium,
email: 1mhgvc.hvp.secretariat@gmail.com

Phone: 0060-9-767-6531

Visit us: http://1mhgvc.kk.usm.my/index.html