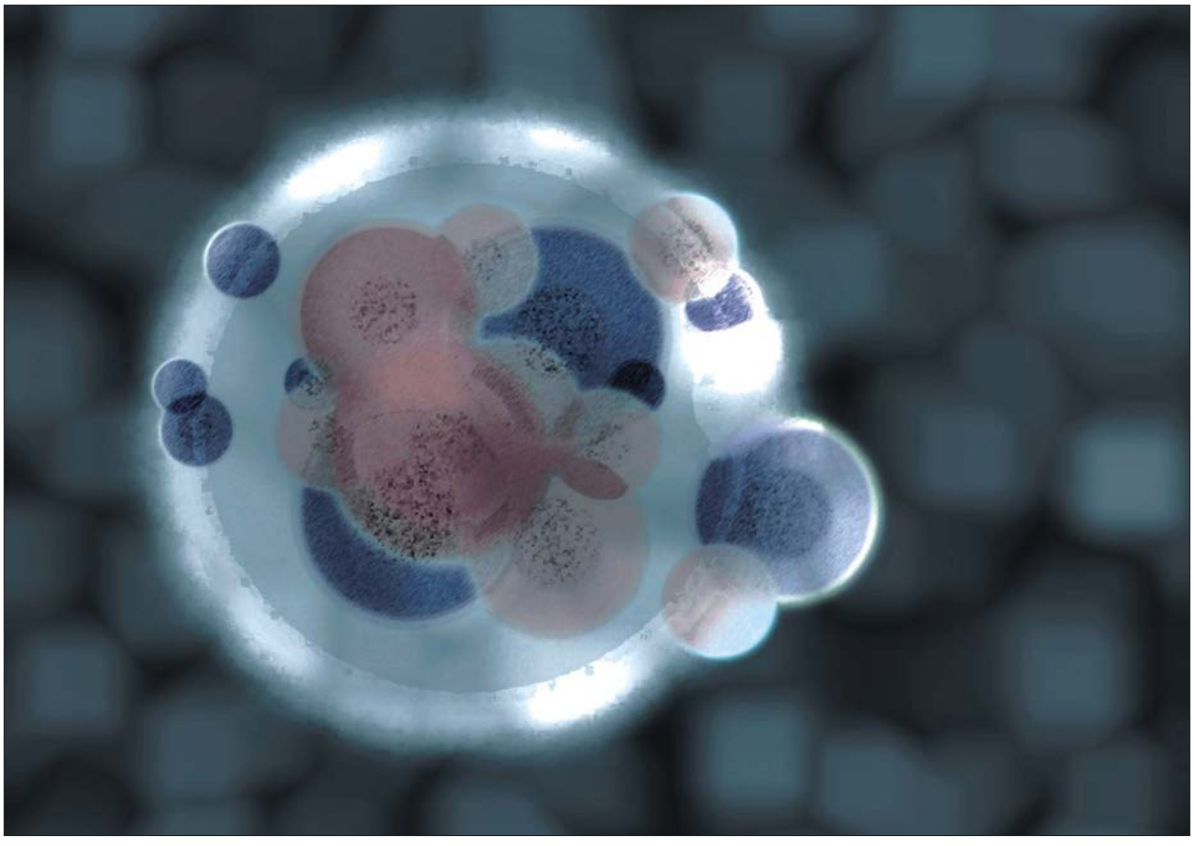

Journal of Biological Sciences

ISSN 1727-3048

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Biological Sciences 5 (6): 790-794, 2005
ISSN 1727-3048

Allelic Representation and its Effect on Genetic Variation: A Jordanian Population-based Study

Mawieh Hamad and Salem R. Yasin
Department of Biological Sciences, Faculty of Sciences and Arts, Hashemite University, Al-Zarqa, Jordan

Abstract: Here, we report on the relationship between allelic number and percentage genotype representation at highly polymorphic human genetic loci. To achieve this goal, blood samples were collected from 450 unrelated Jordanians and analyzed using PCR-based DNA typing kits. PGR ranged from 1.0 at loci with low allelic number to <0.05 at loci with high allelic number suggesting that there is some sort of an inverse relationship between allelic number and PGR. The lowest locus-based heterozygosity was observed at the TPOX locus (0.45) while the highest was observed at the D18S51 (0.88). Heterozygosity as a measure of genetic variation among Jordanians was calculated at 0.72. No clear relationship between the number of genotypes and heterozygosity was observed at any locus; this although, heterozygosity tends to generally increase as allelic number increase. Although the PGR for highly polymorphic loci precipitously decreases as the number of alleles increase, the level of genetic variation based on PCR-based polymorphic loci remains very high.

Key words: Allele, genotype, genetic variation, polymorphism, Jordan

INTRODUCTION

The importance of genetic variation within natural populations is paramount as it allows for the appearance of variable phenotypic profiles and hence permits natural selection to operate[1, 2]. The use of heterozygosity as a measure of genetic variation has been the procedure of choice by population geneticists for several decades[2-4]. Historically, genetic variation studies relied upon polymorphic protein typing using electrophoretic methods[5-8]. Using polymorphic protein typing techniques, genetic variation in humans was estimated at around 0.067 to 0.072[7], this was viewed as significant enough to allow for various microevolutionary processes to operate within different human populations[2,4]. However, as coding sequences in most species represent a minor proportion of the total DNA, polymorphic protein typing is, by definition, confined to a minor and distinct fraction of the DNA. Additionally, protein polymorphism is limited to two or few phenotypes in the majority of cases; consequently, the degree of genetic variation to be estimated using such data is likely to be limited as well. As such, estimating genetic variation merely on the basis of protein polymorphism is biased and limited as it does not necessarily reflect the actual level of variation. This coupled with the observation that non-coding DNA sequences can and do participate in the making of the phenotypic profile of the organism justifies, in our opinion, the use of PCR-based polymorphic DNA loci to measure genetic variation.

With the advent of available molecular methodologies that can be employed in DNA typing and analysis[9-12], it has become possible to carry out DNA typing for purposes ranging from forensic to evolutionary studies. In recent years, numerous reports on the polymorphism of a large number of genetic marker systems in different populations were published. The level of heterozygosity at most of such genetic marker systems is far greater than that obtained from classical protein polymorphism data[13-18]. However, in many of these genetic marker systems, absence of one or more alleles was shown to be of common occurrence with the consequent decrease in the number of genotypes[18]. For any polymorphic locus, not all alleles making up the locus are necessarily present in all human populations. The ratio of alleles detected in the population to the total number of alleles is referred to here as Percentage Allelic Representation (PAR). Additionally, the ratio of genotypes that are detected in the population to the total number of genotypes is referred to here as Percentage Genotype Representation (PGR).

The influence of this phenomenon on the actual degree of polymorphism has yet to be addressed. In this study, the relationship between allelic number and PGR

Corresponding Author: Dr. Mawieh Hamad, Biological Sciences Department, Faculty of Sciences and Arts, Hashemite University, P.O. Box 330077, Al-Zarqa 13133, Jordan
Tel: 00962 5 3903333 Fax: 962-5-3903349

Table 1: PAR, PGR and average heterozygosity in Jordanians at 23 different PCR-based genetic markers

Genetic locus	Total No. of alleles	No. of absent alleles	PAR[a]	No. of absent genotypes[b]/ total No. of genotypes	PGR[c]	% Heter[d]
LDLR	2	0	100.0	0 / 3	100.0	53.2
GYPA	2	0	100.0	0 / 3	100.0	50.4
D7S8	2	0	100.0	0 / 3	100.0	48.2
HBGG	3	0	100.0	0 / 6	100.0	61.2
Gc	3	0	100.0	0 / 6	100.0	66.9
HLA-DQA1	8	0	100.0	8 / 36	77.8	71.8
THO1	8	1	87.5	11 / 36	69.4	76.1
FES/FPS	8	2	75.0	20 / 36	44.4	74.5
TPOX	9	3	66.7	31 / 45	31.1	45.0
vWA31	11	1	91.0	43 / 66	34.8	84.9
D5S818	11	5	54.5	51 / 66	22.7	72.0
D3S1358	12	4	66.7	61 / 78	21.8	84.0
D8S1179	12	2	83.3	51 / 78	34.6	76.0
D13S317	12	5	58.3	59 / 78	24.4	79.0
D16S539	13	6	53.8	72 / 91	20.9	82.0
Penta D	14	2	85.7	72 / 105	31.4	87.0
F13A1	15	5	66.7	96 / 120	20.0	77.9
CSF1PO	16	8	50.0	121 / 136	11.0	67.0
D7S820	19	12	36.4	171 / 190	10.0	80.0
Penta E	22	7	68.2	195 / 253	22.9	73.0
D18S51	29	16	44.8	396 / 435	09.0	88.0
D21S11	35	21	40.0	594 / 630	05.7	80.0
FGA	39	28	28.2	748 / 780	04.1	84.0
Average heterozygosity:			72.0		43.3	72.3

[a]PAR (No. of present alleles/No. of all alleles); [b]Absent genotypes = number of absent but possible genotypes + number of impossible genotypes; [c]PGR (No. of present genotypes/No. of all possible genotypes based on total number of alleles): [d]Percentage heterozygosity

was assessed through evaluating the genotype frequency distribution of 23 different polymorphic loci in a sample population of unrelated Jordanians. Genetic variation among Jordanians was also calculated based on average heterozygosity at all 23 loci included in this study.

MATERIALS AND METHODS

Sample collection: Five milliliter blood samples were collected into plain tubes from a sample population consisting of a total of 450 unrelated Jordanians over the period from 1998-2002. The sample population was selected at random with respect to color, age, sex and area of residence within Jordan. Donors were informed of the goal of the study; they were asked to sign a form of consent on the understanding that names will be kept confidential. DNA was extracted utilizing ready to use Chelex 100 extraction kit (Sigma, St. Louis, USA).

DNA amplification and analysis: DNA samples were typed using commercially available typing kits. Amplification and typing of samples for the PM loci (LDLR, GYPA, HBGG, D7S8 and Gc) was carried out using the AmpliType® PCR amplification and typing kit (Perkin Elmer corporation, Norwalk, CT). Amplification and typing of samples for the HLA-DQA1 locus was carried out using the AmpliType HLA-DQAα forensic DNA Amplification kit (Perkin Elmer Corporation). For both of the above marker typing systems, a Perkin Elmer thermal cycler 480 was used (Perkin Elmer Corporation).

Typing of the STR loci HUMvWA31, HUMTH01, HUMF13A1 and HUMFES/FPS was carried out using the ABI Prism™ Monoplex Primer STR set (Applied Biosystems, Foster City, CA, USA). Multiplex PCR amplification of the D3S1358, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, D 1 6S539, Penta E, Penta D, CSF1PO and TPOX loci was carried out using the PowerPlex®16 System (Promega Cooperation, USA). Capillary electrophoresis for all loci was performed according to the protocols supplied in the manufacturers' Technical Manual. PCR products were denatured with formamide and electrophoresis was carried out on an ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, USA). Length of the amplified fragments was determined using the GeneScan Analysis 2.0.2 software based on Tamra 350 for the Monoplex sets and internal lane size standard (ILS 600) for the Multiplex set. Allele designations were determined using the Genotyper 2.0 and PowerTyper™ 16 Macro softwares by comparison of the sample fragments with those of the allelic ladder.

Genotype frequencies were directly calculated from the number of individuals expressing the specific genotype in the sample set; allelic frequencies were calculated based on the Hardy Weinberg Equilibrium. PAR was calculated as the number of present alleles divided by the total number of alleles for the specific locus; likewise, PGR was calculated as the number of genotypes present divided by the total number of genotypes for the locus in question. Number of absent genotypes was considered as that resulting from

subtracting the number of present genotypes from the total number of genotypes at the specific locus. Average heterozygosity was calculated as the sum heterozygosity at all loci divided by the number of loci.

RESULTS AND DISCUSSION

The number of the known alleles for the loci used in this study ranged from 2 to 39. All loci were considered as polymorphic based on the 95% criterion for polymorphism[4]. At loci with allelic number greater than 8, one or more alleles were missing from the Jordanian population (Table 1); consequently, at each of these loci, a considerable number of genotypes were impossible to occur. At most loci evaluated, certain genotypes were unexpectedly absent although the allelic pairs required to produce them were present; hence the term absent but possible genotypes. Absence of such possible genotypes is understandably due to one or more of a whole host of random variables influencing the pattern of allelic dispersion within the Jordanian population. Absence of impossible genotypes as well as some of the possible ones resulted in the degree of locus-based polymorphism to be lower than expected. Both PAR and PGR were <100% for 18 out of the 23 loci included in the study (Table 1). PGR decreased in a precipitous fashion as the total number of alleles making up the locus increased in that, where PGR remained at 100% for loci consisting of 2-3 alleles (LDLR, GYPA, D7S8, HBGG and Gc), it dropped to <10% for loci with allelic number greater than 28 (D18S51, D21S12 and FGA) (Table 1 and Figure 1). Even though PGR dramatically dropped as the total number of alleles increased, heterozygosity values remained higher than that obtained from polymorphic PCR-based loci with limited allelic number (Table 1). It was also much higher than that obtained from polymorphic protein typing data[19]. It can be argued that population-based PGR values may be more reflective of locus-based polymorphism than that obtained from the theoretical calculation of locus-based polymorphism on the basis of total allelic number. This is mainly because PGR takes into account the special set of circumstances influencing gene flow within a population by discounting absent genotypes; that is both possible and impossible genotypes alike.

The least heterozygous locus was the TPOX (0.45) while the most heterozygous locus was the D18S51 (0.88). No clear relationship between the number of alleles and the degree of heterozygosity was observed; for instance, the TPOX locus (9 alleles) was 45% heterozygous while the THO1 locus (8 alleles) was 76% heterozygous. Based on the genotypic frequency distribution at the 23 loci, average heterozygosity among Jordanians was calculated

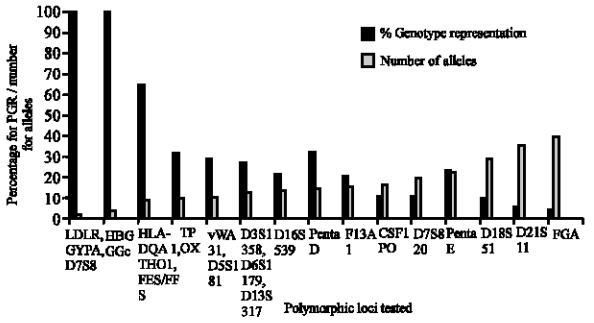

Fig. 1: The relationship between allelic number and PGR. The value for the total number of alleles at each of the 23 loci included in the study was plotted parallel to the corresponding PGR of the locus. All 23 loci were considered in this analysis (loci names were listed on the X axis); in cases were multiple loci shared the same number of alleles, an average PGR was calculated and used instead of single PGR values.

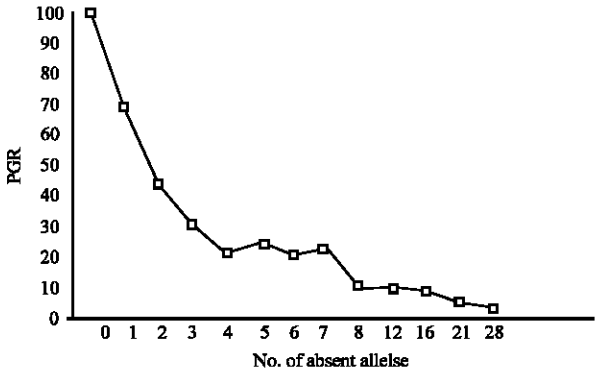

Fig. 2 a: The relationship between the number of absent alleles and PGR. PGR was plotted against the number of absent alleles for representative loci of different allelic number. In cases where multiple loci shared the same number of alleles, one locus was chosen at random as a representative locus

at 72.3%. This is significantly higher than that calculated based on data obtained from polymorphic protein typing procedures[4,7,8]. Should this finding hold true through testing of genetic variation using additional loci for the Jordanian population and through testing of genetic variation in other populations, a reassessment of classical data pertinent to genetic variation in humans as well as other species may become necessary. This although, previous studies have shown that typing of about 20 loci is sufficient to accurately measure genetic variation[4], the degree to which heterozygosity may change should greater number of loci be typed is very low.

Fig. 2b: The relationship between PAR and PGR. Both PAR and PGR for loci of different allelic number were plotted against allelic number. In cases where multiple loci shared the same number of alleles, one locus was chosen at random as a representative locus

The relationship between the number of absent alleles and PGR (Fig. 2a) and that between PAR and PGR (Fig. 2b) was further investigated. It is clear that, with, only few exceptions of the tested loci (loci D5S818, D13S317, F13A31 and Penta E), as the total number of alleles increase, the number of present alleles decrease and so does PAR (Fig. 2a). Similarly, the lower the PAR value for a locus, the lower the PGR value for the same locus (Fig. 2b). It must be noted that such relationships are not exactly linear; this is because the relationship between the number of alleles and PAR is not exactly linear. Nonetheless, as these two values increase, PGR decreases in a precipitous manner. Therefore, actual or real polymorphism at loci with multiple alleles may be reflected more clearly using PAR and PGR rather than total number of alleles. Mathematical treatment of these relationships should help in furthering our understanding of the intricacies involved in allelic dispersion patterns in different populations of sexually reproducing organisms. Why the variations in allelic frequencies in different populations, what are the forces behind these variations and how strong is the effect of such variations on microevolution are but few of the pertinent questions that may be further clarified following this approach.

Since genetic variation is essential for various microevolutionary processes to operate[2,4], being present at high levels may further ease the progression of various microevolutionary processes. This, notwithstanding the fact that variation within the coding portion of DNA is of Great0 significance regarding the progression of microevolution as it directly influences the phenotypic profile. Nonetheless, the contribution of non-coding DNA sequences to the phenotypic profile of the individual cannot be ruled out. Numerous reports have shown that variation at polymorphic non-coding DNA sequences differentially predispose to various disease states[20-22]. Therefore, considering heterozygosity or any other measure of genetic variation at the non-coding DNA level should not be discounted.

ACKNOWLEDGMENT

This study was supported by research grant SY-MH/0203, College of Graduate Studies and Scientific Research, Hashemite University, Jordan.

REFERENCES

1. Lewontin, R.C., 1974. The Genetic Basis of Evolutionary Change. Columbia University Press, New York, pp: 1-24.
2. Ridley, M., 1993. Evolution. Blackwell Scientific Plublications, Boston, Massachusetts, USA., pp: 72-92.
3. Ayala, F.J., J.W. Valentine, L.G. Barr and G.S. Zumwalt, 1974. Genetic variability in a temperate intertidal phoronid, *Phoronopsis viridis*. Biochem. Genet., 11: 413-427.
4. Ayala, F.J., 1982. Population and Evolutionary Genetics: A primer. The Benjamin Cummings Publishing Company In: Menlo Park, Californian, USA., pp: 29-55.
5. Ayala, F.J., M.L. Tracey, L.G. Barr, J.F. McDonald and S. Pereze-Salas, 1974. Genetic variation in natural populations of five *Drosophila* sp. and the hypothesis of the selective neutrality of protein polymorphisms. Genetics, 77: 343-384.
6. Gottlieb, L.D., 1977. Electrophoretic evidence and plant systematics. Ann. Missouri Bot. Gard., 64: 161-180.
7. Harris, H. and D.A. Hopkinson, 1972. Average heterozygosity in man. J. Hum. Genet., 36: 9-20.
8. Johnson, G., 1979. Genetic Polymorphism among Enzyme Loci. Genetic Physiology, Academic Press, New York, USA., pp: 239-273.
9. Budowle, B. and F.S. Baechtel, 1990. Modifications to improve the effectiveness of restriction fragment length polymorphism typing. Applied Theor. Electrophor., 1: 181-187.
10. Gill, P., A.J. Jeffreys and D.J. Werrett, 1985. Forensic application of DNA fingerprints. Nature, 318: 577-579.
11. Kimpton, C.P., P. Gill, A. Walton, A. Urquhart, E.S. Millican and M. Adams, 1993. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Methods Applied, 3: 13-22.

12. Waye , J.S., L. Presley, B. Budowle, G.G. Shutler and R.M. Fourney, 1989. A simple method for quantifying human genomic DNA in forensic specimens extracts. Biotechniques, 7: 852-855.
13. Hamad, M., S. Yasin and A. Elkarmi, 2001. On the polymorphism of HUMvWA31, HUMTH01, HUMF13A1 and HUMFES/FPS STR genetic loci in Jordanians. Kor. J. Genetics, 23: 157-161.
14. Hayes, J.M., B. Budowle and M. Freund, 1995. Arab population data on the PCR-based loci: HLA-DQA1, LDLR, GYPA, HBGG, D7S8, Gc and D1S80. J. Forensic Sci., 40: 888-892.
15. Tahir, M.A., A.Q. Al Khayat, F. Al Shamali, B. Budowle and G.E. Novick, 1997. Distribution of HLA-DQA1 alleles in Arab and Pakistani individuals from Dubai, United Arab Emirates. Forensic Sci. Intl., 85: 219-223.
16. Woo , K.M., M.S. Bruse and B. Budowle, 1995. Korean population data on the PCR-based loci LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1 and D1S80. J. Forensic Sci., 40: 645-648.
17. Yasin, S., M. Hamad and A. Elkarmi, 1999. Jordanian population data on the PCR-based loci: LDLR, GYPA, D7S8 and GC. Forensic Sci. Intl., 104: 17-21.
18. Yasin, S.R., 2002. Allele frequencies at nine PCR-based STR loci in Jordanians. Kor. J. Genet., 24: 327-334.
19. Koehn, R.K. and W.F. Eanes, 1978. Molecular Structure and protein Variation within and among Populations. In: Evolutionary Biology Volume II (Eds.), Hecht, M.K., W.C. Steere and B. Wallace. Plenum Publishing Corp, New York, pp: 39-100
20. Meloni, R., C. Laurent and D. Campion, 1995. A rare allele of a microsatellite located in the tyrosine hydrophosphate gene found in Schizophrenic patients. CR Acad. Sci. III, 318: 803-809.
21. Undlien, D.E., S.T. Bennett, J.A Todd, H.E. Akselsen, I. Ikäheimo, H. Reijonen, M. Knip, E. Thorsby and K.S. Ronningen, 1995. Insulin gene region-encoded susceptibility to IDDM maps upstream of the Insulin gene. Diabetes, 44: 620-625.
22. Weaver, J.U., P.G. Kopelman and G.A. Hitman, 1992. Central obesity and hyperinsulinemia in women are associated with polymorphism in the 5' flanking region of human insulin gene. Eur. J. Clin. Invest., 22: 265-270.