BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 8(25)

Analysis of G6PD enzyme deficiency in Saudi population

Khalid K Alharbi, Alaa Salem Abed, Rabbani Syed*, Imran Ali khan & Sabeena Muhammed Mustafa

Clinical Laboratory Sciences Department, College of Applied Medical Sciences, King Saud University, P.O. Box 10219, Riyadh 11433, Kingdom of Saudi Arabia; Rabbani Syed – Email: rabbanisyd@gmail.com; *Corresponding author

Received September 19, 2012; Accepted October 17, 2012; Published December 19, 2012

Abstract:
The evolutionary conservation of a housekeeping gene such as G6PD is greater than that of tissue-specific genes, presumably because the latter may require more specific adaptation to the physiology of individual organisms. The abundance of distinct mutation sites and their clinical manifestations make G6PD ideal for structure-function analysis. Therefore, it is of interest to screen of G6PD deficiency in the blood donors in Kingdom of Saudi Arabia. We report the mean and variation of enzyme activity in a huge set of Suadi to non-Saudi population with reference to the entire population. The sequence level conservation of G6PD among distant species is demonstrated using phylogenetic trees. These observations have implications in the sequence-structure-function understanding of G6PD with reference to its association to several human diseases.

Keywords: G6PD enzyme, Phylogenetic tree, Gene –Gene Distance, Mutations.

Background:
Glucose-6-phosphate dehydrogenase (G6PD) is an enzyme in the pentose phosphate pathway (PPP) that plays an important role in protecting cells from oxidative damage by producing NADPH and reduced glutathione. In the erythrocyte, which lacks a nucleus, mitochondria and other organelles, PPP is the only biochemical pathway for generating reducing capacity **[1]** .G6PD deficiency is common in malaria endemic regions and is estimated to affect more than 400 million people worldwide. It is a hereditary genetic defect, which is one of the most prevalent polymorphisms and enzymopathies in humans, particularly in males **[2]**. It is postulated that the high frequency of G6PD deficiency has arisen because G6PD deficient variants confer some protection or resistance against malaria caused by Plasmodium falciparum and Plasmodium vivax **[3].** Erythrocyte exposure to oxidative stress causes haemoglobin denaturation, ultimately resulting in haemolysis. Haemolytic anaemia in G6PD-deficient individuals can be triggered by a range of oxidative agents, such as infections and certain foods and drugs, including anti-malarials **[4]**. The G6PD gene is located at the telomeric region of the X chromosome (band Xq28), consisting of 13 exons and 12 introns. It encodes 515 aminoacids and a GC-rich (more than 70%) promoter region. G6PD-deficient erythrocytes are more susceptible to destruction by oxidative stress than normal erythrocytes due to the lower NADPH levels **[4].** Individuals with this genetic defect may exhibit non-immune haemolytic anaemia in response to a number of stimuli, most commonly infections or exposure to certain medications or chemicals **[5].**

G6PD is X-linked, and so deficient variants are expressed more commonly in males than in females. Worldwide, an estimated 400 million people are G6PD deficient with the distribution corresponding to areas in which malaria is, or has been, prevalent. The evolutionary conservation of a housekeeping gene such as G6PD is greater than that of tissue-specific genes, presumably because the latter may require more specific adaptation to the physiology of individual organisms. Total lack of G6PD is almost certainly lethal in mammals, but partial G6PD deficiency confers biological advantage with respect to

malaria. Most of the aa replacements causing G6PD deficiency take place in positions that are conserved, but less than fully conserved. Thus, the range of permissible G6PD mutations in the microevolution of the human species bears a clear imprint of the macro evolutionary history of this gene. With increasing numbers of full genome sequences becoming available, the approach we have used for the analysis of G6PD mutations is likely to be generally applicable to genes (particularly housekeeping genes) underlying other human diseases **[6].** The abundance of distinct mutation sites and their clinical manifestations make this enzyme ideal for structure-function analysis studies. The main goal of the present study is to screen of G6PD deficiency in the blood donors in Kingdom of Saudi Arabia and also in this analysis, we also summarise the occurrence or degree of mutations by residue conservation of G6PD sequence in various species.

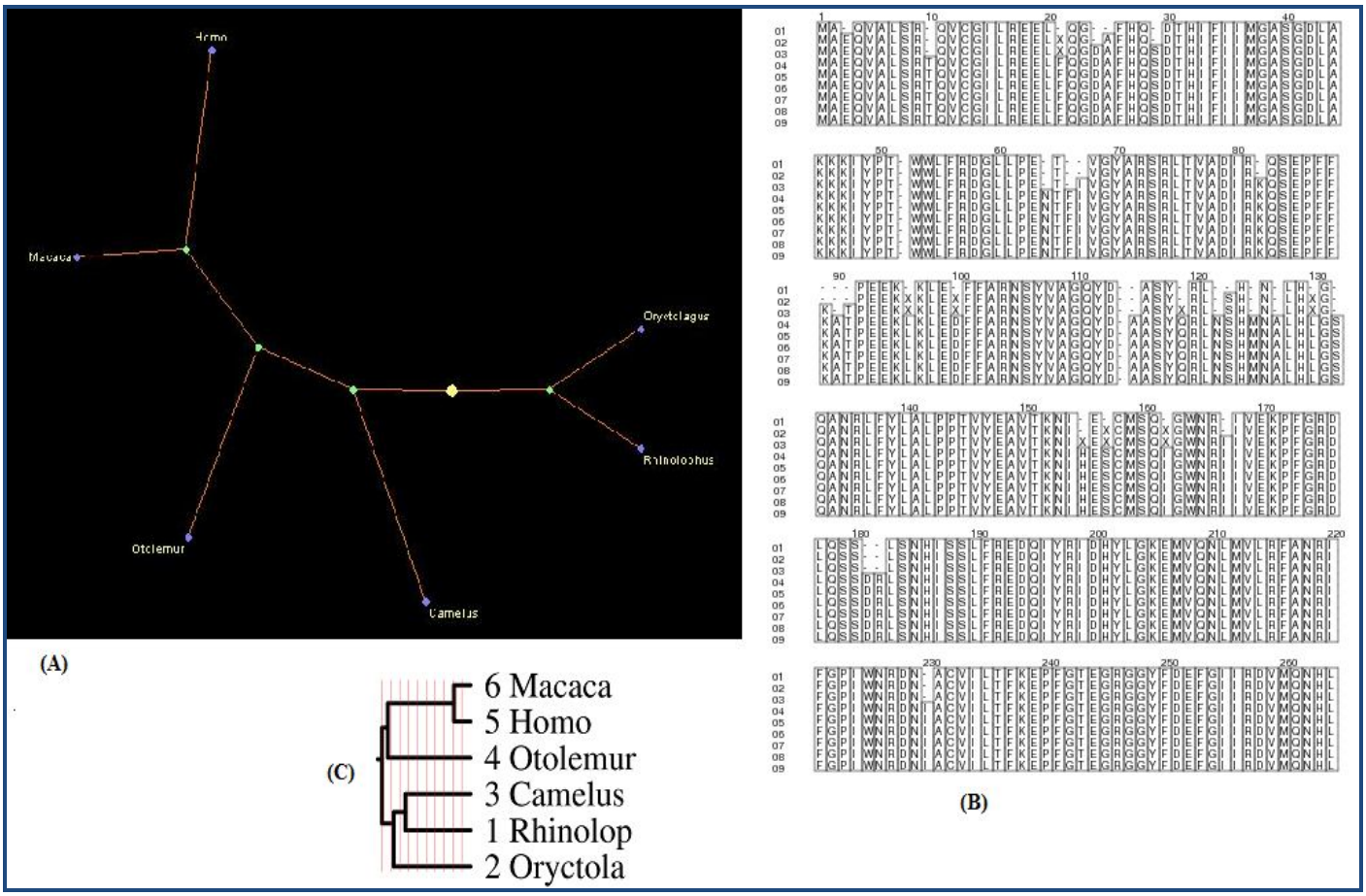

Figure 1: (A) Shows the Phylogenetic tree for the sequences; **(B)** MSA for G6PD from different species; **(C)** Phylogeny (rooted) for G6PD from different species.

Methodology:

Samples Collection

1540 donor samples were obtained from Saudi male donors which presented to Central blood bank, driver school blood bank (DALLA), King Abdul-Aziz specialist Hospital in Taif city in West region of Kingdom of Saudi Arabia during the period (3 July 2011 to 12 Aug 2011). The Age of these donors is between 17 to 50 years with a mean age of 33.5, and tested for the activity of G6PD. Whole blood sample obtain from whole blood bag after blood donation, and we used 6ml potassium EDTA tube, and kept in refrigerators 2-8^{0}C, and tested within 12 hours.

G6P-DH Fluorescence Screening Test

This test used as screening test for G6PDD, by detect the activity of G6PD enzyme in blood sample to reduction of NADP and formation of reduced form NADPH. (This method is that of Beutler & Mithell **[7]** modified on the recommendation of the ICSH (70)). Add 5 µl of Anti-coagulated blood (which was collected in EDTA tube) directly into 100 µl of the working reagent. Mix and incubate the sample with working reagent at temperature (25^0) for 10 minutes. Then take 10 µl of the resulting solution and place on the paper provided and leave to dry. When the filter paper is completely dry view under along wave UV-lamp in a darkened room. It will fluorescence activity when the sample have a normal or slightly deficient. In case of no fluorescence activity after 10-minute incubation, the sample has marked G6PD deficiency of enzyme.

G6PD determination test

This test is different from florescence test, it is used to detect the quantity of NADP that reduced to NADPH by G6PD enzyme. The amount of formation of NADPH is related to the activity of G6PD enzyme. Pipette 400 µl of lysing reagent to a test tube. Add 100 µl of Anti-coagulated blood. Mix well and wait for 5 minutes. In another test tube add 3ml of buffer and 100 µl of NADP reagent. To this mixture add 50 µl of the lysate prepared in step 1. Wait for 5 minutes. To the above mixture add 50 µl of

the substrate and put it in a sample covette, and interred in spectrophotometer. Take a result from spectrophotometer as (mU/ml); milli Units per ml and divided by the RBC value (millions) to get the G6PD milli Units/ RBC (millions).

Genome sequence data

Based on the similarity of the query sequence obtained from Uniprot DB, we could find the closely related sequences. High-throughput NCBI blast (E-value cutoff e-5) was used to search homologous sequences, which yielded 6 sequences. Basic Local Alignment Search Tool (BLAST) **[8]** is a sequence similarity search and alignment program that which can be used by a web interface or as a stand-alone tool to compare the query sequences to the database sequences.

Multiple Sequence Alignment

Clustal W was the program **[9]** used for Multiple Sequence Alignment Program (MSA) to determine the conserved sequences among the templates. The obtained alignment files were stored as text file inorder to study the degree of conservation by tree prediction method.

Evolutionary Studies

Phylogenetic tree were predicted by Phylodraw software **[10]** in order to determine the relatedness of G6PD to other related among other species. PhyloDraw is a unified viewing tool for phylogenetic trees. PhyloDraw supports various kinds of multi-alignment formats (Dialign2, Clustal-W, Phylip format, NEXUS, MEGA, and pairwise distance matrix) and visualizes various kinds of tree diagrams, e.g. rectangular cladogram, slanted cladogram, phylogram, unrooted tree, and radial tree. By using several control parameters, users can easily and interactively manipulate the shape of phylogenetic trees. The tree was calculated from distance matrices determined from percent identity or aggregate. In this method, pairwise distances used to cluster the sequences are represented as the percentage of mismatches between two sequences. The branch lengths are the percentage mismatch between two nodes

Evolutionary Trace Server (TraceSuite II) was the software used to trace the g6pd consensus sequences among species. The concept behind this algorithm is to trace the evolutionary significance among individual sequences based on the occurrence of the amino acid residues in a protein sequence by their relative evolutionary importance, and when a structure is available for that protein, it also can display a structural map of where top-ranked residues fall.

Discussion:

Frequency of G6PD deficiency in KSA is very high when compared to other study population **[11]**, and still not be screening routinely for blood donor **[12].** We estimate the activity of G6PD enzyme in blood donors sample by using fluorescence screening test to follow up the enzyme activity. The normal range for G6PD enzyme in normal patient is (100-200 mU/10^9), in own study (mean ± SD) of G6PD enzyme in 39 deficient donor is (10.8 ±13.1) and the range is (57.5 - 0.206), but the (mean ± SD) of G6PD enzyme in Saudi deficient donor is (8.87 ± 10.2) and the range is (46.16 - 0.206). Finally the (mean ± SD) of G6PD enzyme in non Saudi deficient donor is (14.25 ± 17) and the range is (57.5 - 0.871), **Table 1 (see supplementary material).** The phylogenetic tree **(Figure 1)** we obtained for G6PD clearly supports a common evolutionary origin throughout all the organisms considered for the study and is consonant with taxonomy. The degree of homology results shows critical regions as consensus sequences in all the species. These conserved sites/specific pattern suggests that within the selected species in G6PD are likely to be involved in the development of functional specificity. These may play a crucial role in binding to a specific substrate or can be part of active site of protein. Besides the specific pattern sites determined are regarded as functionally important and the properties of the motifs may also suggest specific functional roles for the protein. The tree shows the average gene-gene distance as 0.048 and root-gene distance as 0.027.

More than a hundred naturally occurring mutations of human glucose-6-phosphate dehydrogenase (G6PD) have been identified at the amino acid level. The abundance of distinct mutation sites and their clinical manifestations make this enzyme ideal for structure-function analysis studies **[13].** These results of computer-assisted analyses contribute to a further understanding of the structure-function relationships of human G6PD deficiency. The results from Evolutionary Trace Server (TraceSuite II) include six clusters which are more or less conserved among species and as shown in figure s 4 and 5. This web server provided an improved representation of results adds useful input and output options and integrates the graphical representation of the clusters which shows the evolutionary distance between the selected sequences among the species. The cluster shows human and macaca have more or less similar kind of evolutionary significance in the case of G6PD mutation and they lay on the same phase in evolution **[14].**

Recognition of sequence variations especially by mutation that lead to functional diversification within a sequence family is not a trivial task. Our study focused on the fact that, functional specificity signals must be separated from strong background signals resulting from the phylogenetic differences between the protein subfamilies or subgroups. The mutation in the sequences can be analyzed through pattern and evolutionary studies which can be helpful to distinguish the specific distribution of amino acids within and across the subfamilies. Hence, it is very important to determine the degree of conservation of mutation by analyzing the occurrence of the residues that can provide significance information for evolutionary studies **[15].**

Conclusion:

The application of sequence similarity study approach and phylogenic studies as shown by this work may significantly advance the understanding of protein sequence-structure-function relationships and guide experimental characterization of protein function. Overall, our analysis reveals the presence of G6PD mutated sequences among species and gene clusters provide significant differences which helps to identify the evolutionary path that might have included. This analysis highlights the conserved presence as well as functional importance of G6PD revealed by mutational studies.In the present case, gene-gene and gene-root distances highlighted relatively well the feature of evolutionary data for homologous between species.

Conflict of Interest: No

Acknowledgement:
We are thankful to Mentor and Staff of College of Applied Medical Sciences (CLS Department), King Saud University, Riyadh for helping us in compiling this study by providing facilities from the university Grand.

References:

[1] Carter N *et al. Malar J*. 2011 **10:** 241 [PMID: 21849081]
[2] Phompradit P *et al. Malar J*. 2011 **10:** 368 [PMID: 22171972]
[3] Louicharoen C *et al. Science*. 2009 **326**: 1546 [PMID: 20007901]
[4] Beutler E & Duparc S, *Am J Trop Med Hyg.* 2007 **77:** 779 [PMID: 17978087]
[5] Frank JE, *Am Fam Physician*. 2005 **72**: 1277 [PMID: 16225031]
[6] Eosario notaro *et al. FASEB J.* 2000 **14:** 485 [PMID: 10698963]
[7] Beutler E & Mitchell M, *Blood.* 1968 **32**: 816 [PMID: 4386875]
[8] Altschul SF *et al. J Mol Biol*. 1990 **215**: 403 [PMID: 2231712]
[9] Julie D *et al. Nucl Acids Res*. 1994 **22**: 4673 doi: 10.1093/nar/22.22.4673
[10] Jeong-Hyeon Choi *et al. Bioinformatics*. 2000 **16:** 1056 doi:10.1093/bioinformatics/16.11.1056
[11] Alhamdan NA *et al. Genetics Medicine*. 2007 **9:** 372 [PMID: 17575503]
[12] Alabdulaali MK *et al. Asian J Transfus Sci.* 2010 **4**: 31 [PMCID: PMC2847342]
[13] Yi-Sheng Cheng *et al. Journal of Biomedical Science*. 1999 **6**: 106
[14] Alexix Dereeper *et al. BMC Evolutionary Biology*. 2010 **10:** 8
[15] Pei J *et al*. *Bioinformatics*. 2006 **22**: 164

Edited by P Kangueane
Citation: Alharbi *et al.* Bioinformation 8(25): 1260-1264 (2012)

BIOINFORMATION

open access

Supplementary material:

Table 1: Enzyme activity ((IU / g) in deficient donor's sample

Variable	Saudi	Non Saudi	Total
MEAN	8.871	14.25	10.80
SD	10.17	17.09	13.12
MAX	46.16	57.5	57.5
MIN	0.206	0.871	0.206