A single-step assay for the Gerbich-negative allele of glycophorin C

Brendan Scott ^{*}, Simon Easteal

John Curtin School of Medical Research, Australian National University, Canberra, ACT 0200, Australia

ARTICLE INFO

Article history:

Submitted 24 October 2007

Revised 15 February 2008

Available online 14 April 2008

(Communicated by Sir D. Weatherall, F.R.S., 25 February 2008)

Keywords:

Gerbich
Glycophorin
Malaria
Papua
Sialic

ABSTRACT

The Gerbich erythrocyte surface protein, glycophorin C (GYPC), can be used by *Plasmodium falciparum* to invade erythrocytes. The Melanesian Gerbich-negative antigenic condition (Ge⁻) is frequent in some populations where malaria is endemic, suggesting that it protects against malaria. We have determined as precisely as possible the breakpoint of the chromosomal deletion that causes the Ge⁻ condition by comparing the partial GYPC sequence of a Papuan Ge^{-/-} homozygous individual with known sequences of GYPC. This localisation has allowed us to develop a robust single-step PCR assay suitable for rapid screening of Ge⁻. This method is easier to implement than existing methods, can reliably identify heterozygous individuals, and will considerably aid efforts to study the distribution of Ge⁻ and its role in protection against malaria.

© 2008 Elsevier Inc. All rights reserved.

Introduction

The Gerbich-negative antigenic condition [1] results from the absence or alteration of sialic acids on the extracellular domain of the erythrocyte surface protein Glycophorin C (GYPC), caused by the absence of the third exon of the GYPC gene (Fig. 1). Ge⁻ is an apparently harmless recessive condition, and homozygotes are rare in most populations (1 in 45 000) [2], but in some parts of Papua New Guinea (PNG), the condition is common e.g. 39.3% of Atzeran-speakers in the Markham Valley are Ge⁻ homozygotes [3]. Ge⁻ is of interest because of its interaction with malaria. Ge⁻ erythrocytes are bound less efficiently than normal erythrocytes by *Plasmodium falciparum* [4–6] and by the *P. falciparum* surface protein EBA-140/BAEBL [7]. In some areas of PNG, Ge⁻ individuals have been reported to have lower levels of falciparum and/or vivax infection than Ge⁺ individuals [8], although this has not been borne out by more recent studies [9,10].

Detecting the Ge⁻ allele has been problematic. The GYPC gene contains two nearly identical copies of a 3.5 kb DNA segment. The first copy (Repeat 1) incorporates exon 2 and the second copy (Repeat 2) incorporates exon 3 of the gene (Fig. 1A). Ge⁻ alleles contain a deletion that results in the loss of exon 3 [11,12]. The large size of the deletion (~3.5 kb) and the sequence similarity of the repeats make the design of DNA-based typing methods difficult. Southern blotting has been used to differentiate Ge⁻ on the basis of the presence of one or two of the 3.5 kb repeats [12,13]. However, Southern blotting is labour-

intensive and time-consuming. A PCR-based assay was developed by Patel et al. [9], but its requirement for multiple PCRs per individual limits its application at a larger scale.

An improved assay for Ge⁻ would help shed light on its relationship to malaria and it would be useful for screening blood transfusion recipients from the Pacific region, since transfused Ge⁻ individuals may be at risk of haemolytic anaemia [2,14] and Ge⁻ newborns have required transfusion [15].

We have developed a simple single-step, PCR-based Ge⁻ assay and tested it on samples previously screened by Southern blotting. This was possible by first determining the breakpoint of the Ge⁻ deletion by sequencing the region of GYPC flanking the Ge⁻ deletion from a Papuan Ge⁻ homozygote and aligning the Papuan Ge⁻ sequence with Repeat 1 and Repeat 2 sequences from public databases.

Materials and methods

Genomic DNA was extracted from frozen peripheral blood lymphocytes from a homozygous Ge⁻ individual from Popondetta, Oro province, Papua New Guinea, using QIAGEN QIAamp DNA Mini Kits (QIAGEN Pty Ltd Australia, Doncaster, Victoria). Long PCR was used to amplify a primary PCR product with a forward primer (Cf, 5'-AGGTGCTGCTAGGCATGGAGAG-3') designed with a 3' nucleotide that is complementary to Repeat 1 but not Repeat 2, and a reverse primer (2Ar, 5'-CCTCATTGGTGTGGTACGTG-3') that is complementary to exon 4. Templates of 37.5 ng DNA were amplified in a 50 µl PCR mixture containing 5 µl PCR buffer (Expand Long Template PCR System buffer 3 (Boehringer Mannheim, Germany); 500 µM of each dNTP; 0.3 µM of each primer; 1.5 mM additional MgCl₂; and 3.75U Taq DNA polymerase. PCR consisted of: 92 °C, 2 min; 63 °C, 20 s;

^{*} Corresponding author.

E-mail address: Brendan.Scott@anu.edu.au (B. Scott).

Fig. 1. *Ge*(-) genotyping assay. Three primers (N493f, 198fA, 1582L) are included in the same PCR. In wildtype alleles (A), 198fA and 1582L prime the amplification of a 1805 bp fragment; in *Ge*(-) alleles (B), N493f and 1582L prime the amplification of a 1486 bp fragment. The two fragments can be separated by electrophoresis in agarose gels, enabling the identification of three genotypes: *Ge*(-/-): 1486 bp; *Ge*(-/+): 1486 bp + 1805 bp; *Ge*(+/+): 1805 bp. Examples of genotyping results (C) are shown for some of the individuals from Wosera, Papua New Guinea, previously genotyped by Southern blot, and used to validate the assay; and for the *Ge*(-) Papuan individual (-/-) and a *Ge*(+/+) Terena individual from Brazil (+/+), which were used as positive controls in developing the assay.

68 °C, 8 min; followed by 23 cycles of 92 °C, 10 s; 63 °C, 20 s; 68 °C for: 8 min (8 cycles), 9–13 min (3 cycles each); followed by 68 °C for 5 min. PCR products ~2.5 kb in length for the Papuan *Ge*(-) homozygote and ~6 kb in length for wildtype controls were visualised after electrophoresis on agarose gels. PCR products were purified using a QIAGEN QIAquick PCR Purification Kit.

Where necessary, primary products were used as templates in a second round of long PCR using nested primers complementary to exons 2 (80f, 5'-GCCTCTGCCTCCACCACAATGC-3') and 4 (984r, 5'-ATGACGAA-GAGGAGGGAGACTA-3'). PCRs were carried out with 1:250 final concentrations of the primary products, using the same cycling program, in 50 µl mixtures containing: 5 µl Expand Long Template PCR System buffer 3 (Boehringer, Mannheim, Germany); 500 µM of each dNTP; 0.2 µM of each primer; 1 mM additional MgCl₂; and 3.75U *Taq* DNA polymerase.

The products of this second, or nested, PCR for the Papuan *Ge*(-) individual were purified from bands cut from agarose gels and used as templates for dideoxy-sequencing. The nested primers, 80f and 984r, were used for sequencing in conjunction with: 237f (5'-GGTGAGAT-GAGGGTAACTGA-3'); 112r (5'-AAATGGTCAAAGTTCAAAT-3'); 1523U

(5'-AGTTTGAATGGCGTGAATG-3'); 347L (5'-AGGCAAAGGCAGGCT-CAAGT-3'); 579U (5'-CTCTGTGTAACCTGGCATTC-3'); 943U (5'-GGA-GTCCCAGCAGAAGTGAG-3'); and 896L (5'-TGGGGAGGGACTGCTGAAGA-3'). The sequences of overlapping fragments were assembled into a single 2131 bp contig, designated EF434170. Searches using the NCBI BLAST tool [16,17] confirmed that the sequence of the 5' section of EF434170 was identical to published sequences from the 3' sections of Repeat 1, Repeat 2, and the European *Ge*(-) allele (GENBANK accession numbers M24627, M24629, and M24628 respectively, <http://www.ncbi.nlm.nih.gov/>), and that the sequence of the 3' section of EF434170 is identical to all published exon 4 sequences.

EF434170 was aligned with both Repeat 1 and Repeat 2 sequences in a partial *GYPC* contig assembled from genome sequence data at the Washington University Genome Sequencing Center (WUGSC) [18], using MACAW [19] and ClustalW 1.7 [20] (Fig. 2). The deletion breakpoint was determined on the basis that upstream of the breakpoint, the sequence containing the deletion (EF434170) would most resemble Repeat 1 and downstream of the breakpoint it would most resemble Repeat 2.

Fig. 2. Localisation of the Ge(-) deletion breakpoint from alignment of the Papuan Ge(-) sequence with GYPC Repeats 1 and 2 from wildtype chromosome. The breakpoint occurs within the 113 bp boxed region of the aligned sequences that separates a region in which the Ge(-) sequence resembles the Repeat 1 sequence from a region in which it resembles the Repeat 2 sequence. *Sites corresponding to previously described single nucleotide polymorphisms rs2280240 and rs28387226. | marks the 3' boundary of Repeats 1 and 2.

The Ge(-) assay based on a single PCR reaction was designed with 2 forward primers: N493f (5'-ACATAGGTATGTCATATGATCTG-3'), which binds specifically to Repeat 1, and 198fA (5'-CAAGGTGCTGC-TAGGCATGGAGAA-3'), which binds specifically to Repeat 2; and reverse primer, 1582L (5'-CAGCCTGGGAAATGGTCAAA-3'), which binds to a site downstream of Repeat 2 (Fig. 1). The assay is predicted to produce 1486 bp (N493f/1582 L) and/or 1805 bp (198fA/1582 L) amplicons that distinguish the 3 Ge genotypes as follows: Ge(-/-): 1486 bp; Ge(+/-): 1805 bp; Ge(-/+) 1486 bp and 1805 bp. The corresponding binding site for N493f in Repeat 2 is mismatched at 8 nucleotides, including the 3' end nucleotide, while the corresponding site for 198fA in Repeat 1 is mismatched only at the 3' end nucleotide.

Genomic DNA from a Papuan Ge(-/-) individual, from a non-Melanesian Ge(+/-) individual (a Terena Amerindian from Brazil), and from a number of individuals from a population where Ge(-) is common (the Gogol river valley, Papua New Guinea [8,12]) were amplified using the three primers (N493f, 198fA and 1582L) in 50 µl PCR mixtures containing: 0.3 µM primer N493f; 0.4 µM primer 198fA; 0.7 µM primer 1582L; 0.4U Taq DNA polymerase; 50 ng template. PCRs were run for an initial cycle of: 95 °C, 2 min; 55 °C, 30 s; 72 °C, 1 min, followed by 30 cycles of: 95 °C 1 min; 55 °C 30 s; 72 °C, 1 min.

PCR products from both homozygous controls and from a heterozygous individual were purified, and dideoxy sequenced in both directions using a combination of primers appropriate to each allele i.e. 1582L+N493f for Ge(-), and 1582L+198fA Ge(+). The sequence data obtained were compared to public database sequences of Repeat 1 and Repeat 2 to confirm that the appropriate target sequences had been amplified.

DNA from 19 Papuan individuals from Wosera, PNG, who had previously been screened for Ge type by Southern blotting [12,13] was

quantified by spectrophotometry, diluted to 25 ng/µl, and typed using the three-primer assay, along with positive controls (Papuan Ge(-/-); Terena Ge(+/-); and negative (no DNA) controls. An estimated 50 ng of each template was used. Amplicons were separated on 1% agarose-TBE-ethidium bromide gels at 90 V for 50–60 min (Fig. 1C).

Results and discussion

We obtained a 2131 bp sequence contig (EF434170), extending from exon 2 within Repeat 1 to exon 4 from the genomic DNA of a Ge(-/-) Papuan individual. BLAST searches of EF434170 against known GYPC sequences confirmed that it is a partial sequence of GYPC. By comparing the sequence of this contig with publicly available Repeat 1 and Repeat 2 sequences, we were able to narrow the site of the Ge(-) deletion breakpoint to a 113 bp region near the 3' end of the repeats (Fig. 2). This location is consistent with the findings of Southern blot analysis [12] and previous speculation from the structure of the wild-type GYPC gene [11].

Within the 113 bp region, there are 3 apparent differences between Repeats 1 and 2 (Fig. 2), but these cannot be used reliably to locate the deletion breakpoint. There is a 3-base indel (CTT) missing from Repeat 1 and the Ge(-) sequence, but present in Repeat 2. However, the three bases are also present in the older Genbank sequences of a partial first repeat (M24627), a partial second repeat (M24629), and a corresponding partial sequence from a European Ge(-) homozygote (M24628), suggesting that it may be a cloning or sequencing artefact.

There are also two single nucleotide sites at which the Ge(-) sequence is identical to the Repeat 2 sequence. These sites correspond to two known single nucleotide polymorphisms (SNPs; rs2280240 and rs28387226) and the alternate nucleotides at these sites are the

same as those described for the corresponding SNPs. The observed difference between the repeats may, therefore, reflect polymorphism in one or both of them, which would make them unreliable indicators of the breakpoint location. Conversely, real differences between the repeats at these positions may previously have been misidentified as SNPs. Some misidentification of repeat differences as SNPs is not unexpected, since there is a high degree of similarity between the repeats; the public databases dbSNP and WUGSC report a higher density of intronic SNPs within the repeats than in adjacent parts of *GYPC* (data not shown). If the three-base indel is an artefact, and these SNPs are not true SNPs but misidentified differences between the repeats, then the region of the breakpoint could be narrowed to within 30 bp.

The identification of the breakpoint enabled us to design forward PCR primers, N493f and 198fA that bind specifically to sites in Repeat 1 and Repeat 2 respectively and which produce different length PCR products in combination with the same reverse primer 1582L (Fig. 1). We tested the specificity of the primers used in the assay by sequencing amplicons they produced and aligning these sequences with the reference sequences for Repeat 1 and Repeat 2, between which there are 4 nucleotide differences. Unprocessed sequence data were used to avoid bias, which meant that some sites could not be compared because of the incomplete nature of the data. N493f-primed sequence for the Papuan Ge(–) homozygous control was identical to Repeat 1 at all 4 repeat-specific sites. N493f-primed and 198fA-primed sequences from the heterozygous control were identical to Repeat 1 reference sequence at 2/2 sites, and Repeat 2 reference sequence at 3/3 sites, respectively.

Fifteen of the nineteen Wosera samples tested produced band patterns consistent with their previous genotyping by Southern blotting (see examples in Fig. 1C). The remaining four samples failed to produce amplicons. The failure of these assays may result from incorrect estimation of genomic DNA concentration in these samples. In the homozygous controls, PCR amplification failed when DNA template concentration was either raised or lowered, demonstrating that template DNA concentration is critical to the success of the assay.

However, both gel visualization and spectrophotometry indicated appropriate amounts of high quality template were used in the reactions and when template concentrations were increased or decreased, the samples continued to produce no results. Polymorphisms at sites complementary to the 3' end of the reverse primer could explain failure to detect both alleles, since this primer is used to prime both PCR reactions. There are no SNPs listed at these sites in public databases. However, the SNPs in these databases were ascertained in relatively small samples from Africa, Asia and Europe, and are unlikely to represent all the common variation in populations in New Guinea.

In a few cases, varying the template concentration for these samples produced faint bands, possibly resulting from non-specific amplification, that are inconsistent with results obtained by Southern blotting. This result further emphasizes the importance of appropriate template concentration for the successful use of this assay. It implies that when the template is resistant to amplification, incorrect concentration may cause samples to be mistyped.

This assay has important advantages over previous techniques for screening for the Ge(–) allele. It can distinguish heterozygotes, and can be widely applied without specialised technical knowledge or access to rare antibodies. All genotypes produce PCR bands, and thus genotyping does not rely on the absence of any amplicons to distinguish “null” alleles. Additionally, this screening method does not rely on the intensity of PCR bands on a gel, which may lead to mislabelling of heterozygotes

and homozygotes and make false negatives more likely. While the occurrence of Ge(–) deletions with multiple origins seems unlikely, the three-primer PCR-based assay would nonetheless be robust to most deletions that would remove exon 3 through unequal crossing over, including M24628, the European Ge(–) allele in Genbank. Finally, it only requires a single PCR for each sample, and is in this respect simpler than the PCR-heteroduplex assay of Patel et al. [9].

Acknowledgments

The authors wish to thank G. Koki of the Papua New Guinea Institute of Medical Research, from whose collection the Ge(–) sample was obtained. We also thank Professors S. Serjeantson, D. York, X. Gao, and P. Main, who obtained other samples used here. S. Pluskche, G. Wagner, and S. Serjeantson conducted the Southern blotting. We are most grateful to Professor S. Serjeantson for the advice on older sample material. BS was supported by an Australian National University PhD Scholarship.

References

- [1] R.R. Race, R. Sanger, Blood groups in man, 6th ed. Blackwell Scientific Publications; distributed by J. B. Lippincott, Oxford, 1975.
- [2] P.B. Booth, K. McLoughlin, The Gerbich blood group system, especially in Melanesians, *Vox Sang.* 22 (1972) 73–84.
- [3] P.B. Booth, R.W. Hornabrook, The inheritance of the Gerbich blood groups: an analysis of New Guinean families, *Hum. Biol. Oceania* 2 (1973) 72–78.
- [4] A.H. Chishti, J. Palek, D. Fisher, et al., Reduced invasion and growth of *Plasmodium falciparum* into elliptocytic red blood cells with a combined deficiency of protein 4.1, glycophorin C, and p55, *Blood* 87 (1996) 3462–3469.
- [5] D.C. Mayer, J.B. Mu, X. Feng, et al., Polymorphism in a *Plasmodium falciparum* erythrocyte-binding ligand changes its receptor specificity, *J. Exp. Med.* 196 (2002) 1523–1528.
- [6] G. Pasvol, Receptors on red cells for *Plasmodium falciparum* and their interaction with merozoites, *Philos. Trans. R. Soc. Lond., B Biol. Sci.* 307 (1984) 189–200.
- [7] D.C. Mayer, O. Kaneko, D.E. Hudson-Taylor, et al., Characterization of a *Plasmodium falciparum* erythrocyte-binding protein paralogous to EBA-175, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 5222–5227.
- [8] S.W. Serjeantson, A selective advantage for the Gerbich-negative phenotype in malarious areas of Papua New Guinea, *P. N. G. Med. J.* 32 (1989) 5–9.
- [9] S.S. Patel, R.K. Mehlotra, W. Kastens, et al., The association of the glycophorin C exon 3 deletion with ovalocytosis and malaria susceptibility in the Wosera, Papua New Guinea, *Blood* 98 (2001) 3489–3491.
- [10] S.S. Patel, C.L. King, C.S. Mgone, et al., Glycophorin C (Gerbich antigen blood group) and band 3 polymorphisms in two malaria holoendemic regions of Papua New Guinea, *Am. J. Hematol.* 75 (2004) 1–5.
- [11] C. Le Van Kim, Y. Colin, D. Blanchard, et al., Gerbich blood group deficiency of the Ge-1,-2,-3 and Ge-1,-2,3 types. Immunochemical study and genomic analysis with cDNA probes, *Eur. J. Biochem.* 165 (1987) 571–579.
- [12] S.W. Serjeantson, B.S. White, K. Bhatia, R.J. Trent, A 3.5 kb deletion in the glycophorin C gene accounts for the Gerbich-negative blood group in Melanesians, *Immunol. Cell Biol.* 72 (1994) 23–27.
- [13] G.E. Wagner, Molecular Genetic Analysis of Genes associated with inherited resistance to malarial parasitaemias, John Curtin School of Medical Research, Australian National University, Canberra, 1995.
- [14] R. Barnes, T.L.T. Lewis, A rare antibody (anti-Ge) causing haemolytic disease of Newborn, *Lancet* 278 (1961) 1285–1286.
- [15] P.A. Arndt, G. Garratty, G. Daniels, et al., Late onset neonatal anaemia due to maternal anti-Ge: possible association with destruction of erythroid progenitors, *Transfus. Med.* 15 (2005) 125–132.
- [16] S.F. Altschul, W. Gish, W. Miller, et al., Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403–410.
- [17] T.L. Madden, R.L. Tatusov, J. Zhang, Applications of network BLAST server, *Methods Enzymol.* 266 (1996) 131–141.
- [18] E.S. Lander, L.M. Linton, B. Birren, et al., Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860–921.
- [19] G.D. Schuler, S.F. Altschul, D.J. Lipman, A workbench for multiple alignment construction and analysis, *Proteins* 9 (1991) 180–190.
- [20] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.