Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular & Biochemical Parasitology 143 (2005) 173–182
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Estimating SNP proportions in populations of malaria parasites by sequencing: Validation and applications☆

Paul Hunt*, Richard Fawcett, Richard Carter, David Walliker

Institute for Immunity and Infection Research, School of Biological Science, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, Scotland, UK

Received 22 December 2004; received in revised form 11 May 2005; accepted 12 May 2005
Available online 11 July 2005

Abstract

We have developed a rapid and simple method for determining accurately the proportions of alleles or individual malaria clones in a mixed infection. The technique uses a nested PCR reaction to amplify, from parasite mixtures, alleles of genes differing by single nucleotide polymorphisms, simultaneously, using common primers to non-polymorphic sequences. The mixed products are sequenced, and the heights of fluorescence peaks associated with different nucleotides at the polymorphic site used to quantitate the proportions of each allele in the mixture. We have confirmed the accuracy and precision of the method using a set of well-validated mixtures of genetically different malaria parasites. This technique can be used in the mapping of genetic loci underlying phenotypic traits and in the evaluation of the effects of different alleles upon the reproductive success (fitness) of parasites.

Keywords: *Plasmodium chabaudi*; Malaria; Alleles; Polymorphism; Single nucleotide polymorphism (SNP); DNA sequencing

1. Introduction

Mixed infections of more than one genetically distinct clone are common in natural infections of malaria parasites [1]. Quantitation of the proportions of each clone in such mixtures allows us to address important biological and epidemiological questions. For instance, the relative proportions of clones with different genotypes can be followed over time, allowing measurements of the overall fitness of specific clones, for example, those differing in drug-sensitivity.

Several methods have previously been used to determine the proportions of genetically different parasites in mixtures. These include (i) deriving a series of individual clones from the mixture and counting the number of genotypes among them [2,3]; (ii) using allele-specific antibodies which recognise a subset of parasites in the mixture to label the parasites on fixed blood-smears, enabling direct counts of these parasites to be made by fluorescence microscopy [4,5]; (iii) mutation-specific in situ PCR to label specific alleles in parasites on microscope slides, which can then be counted by microscopy [6]; and (iv) real-time quantitative PCR (RTQ-PCR) [7].

Each of these methods suffers certain drawbacks, such as being time-consuming, expensive, requiring the generation of reagents and assays for a specific genetic locus, or problems of repeatability. RTQ-PCR is probably the most accurate of the PCR-based methods, and is inexpensive after the initial capital outlay on development. The technique has been used recently to characterise simple and complex mixtures of clones of the rodent malaria species *Plasmodium chabaudi*, making use of the highly polymorphic gene *pcmsp1* (encod-

Abbreviations: AFLP, amplified fragment length polymorphism; bp, base pair; DNA, deoxyribonucleic acid; LGS, linkage group selection; PCR, polymerase chain reaction; RFLP, restriction fragment length polymorphism; RTQ-PCR, real-time quantitative PCR; SNP, single nucleotide polymorphism

☆ *Note*: Nucleotide sequence data reported in this paper for *pcpppk/dhps* have been deposited in GenBank™, accession numbers AY781173 and AY781174.

* Corresponding author. Tel.: +44 131 651 3601; fax: +44 131 650 6564.
E-mail address: Paul.Hunt@ed.ac.uk (P. Hunt).

0166-6851/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2005.05.014

ing merozoite surface protein 1), alleles of which vary at numerous sites [7]. So far, however, RTQ-PCR has not been applied successfully to alleles of genes which differ by just a single nucleotide.

Here we describe a technique for analysing parasite mixtures, which we call proportional sequencing. The method is rapid, and in principle can be applied to any single nucleotide polymorphism (SNP) at any locus, with only minimal requirements for optimisation and assay development. This contrasts with RTQ-PCR, in which rigorous development of each individual assay is needed. We use here a series of artificial mixtures of two clones of *P. chabaudi* used in an earlier study using RTQ-PCR of *pcmsp1* [7], but making use of *pcpppk-dhps* (encoding the bifunctional enzyme hydroxymethylpterin pyrophosphokinase/dihydropteroate synthase), alleles of which show several synonymous and non-synonymous SNPs.

2. Materials and methods

2.1. Parasite clones

Two cloned lines of the rodent malaria species *P. chabaudi* are used here: AS-PYR1 derived from cloned isolate AS after pyrimethamine selection, and AJ. These clones differ from one another by numerous amplified fragment length polymorphisms (AFLP) and other markers, the frequency of polymorphisms being between 1 in 1270 bp and 1 in 125 bp [8]. More specifically, the *dhps* nucleotide sequences of AJ and AS-PYR1 (GenBank accession numbers AY81173 and AY781174, respectively) include a number of synonymous and non-synonymous SNPs some of which are shown in Fig. 1.

2.2. Artificial mixtures of AS-PYR1 and AJ DNA

A series of artificial mixtures of AS-PYR1 and AJ, differing in their composition from 0% AS-PYR1 to 99.75% AS-PYR1, was produced by combining blood samples from mice containing AS-PYR1 and AJ at known parasite densities in different proportions (Table 1) [7]. DNA samples were then extracted as previously described [7]. The proportions of AS-PYR1 and AJ in these samples were previously validated by RTQ-PCR of *pcmsp1*, in the same study.

2.3. Proportional sequencing

Proportional sequencing estimates the proportions of alleles of a gene in a mixture by measuring the heights of the peaks in DNA sequence gel electropherograms which correspond with the nucleotides at the polymorphic sites under study.

The procedure used here was as follows. First, a fragment of *pcdhps* (Fig. 1) was amplified by a nested PCR reaction. Genomic DNA derived from mixtures of AS-PYR1 and AJ was amplified using outer primers pcdhps-11 (5′-GTACGCAGAATAT TTCAAATG-3′) and pcdhps-12 (5′-CCTTGAATACCCAATAAAAAG-3′). The product was diluted × 50 in sterile distilled water. One microliter of the diluted product was then amplified in a nested reaction using inner primers pcdhps-07 (5′-CTTTTGTTTCTCATAATCCAG-3′) and pcdhps-08 (5′-TTCTTTGCAAAACCTAAACC-3′). A negative control (no template) was included with each batch of reactions. If a negative control produced an amplified product, all samples within this batch were discarded. Five replicates of each genomic DNA sample were amplified.

Table 1
Artificial mixtures of AS-PYR1 and AJ

Artificially mixed sample (calculated proportions)		RTQ-PCR determinations of *pcmsp1* alleles	
%AS-PYR1	% AJ	%AS-PYR1	% AJ
0.0	100.0	0.0	100.0
2.0	98.0	2.0	98.0
6.0	94.0	2.9	97.1
20.0	80.0	18.7	81.3
32.5	67.5	33.3	66.7
50.0	50.0	46.2	53.8
67.5	32.5	67.7	32.4
80.0	20.0	77.7	22.3
94.0	6.0	93.3	6.7
98.0	2.0	98.1	1.9
99.0	1.0	98.8	1.2
99.8	0.3	99.6	0.4

Artificial mixtures were produced by mixing of volumes of blood samples of equal parasite density obtained from mice infected with pure AS-PYR1 or AJ. They were validated by RTQ-PCR at *pcmsp1*, as previously described [7].

The amplified products were purified with QIAquick PCR purification kit (Qiagen) and sequenced using an ABI Prism® BigDye™ Terminator (Applied Biosystems) cycle sequencing ready reaction kit (version 1) in the forward and reverse directions, using primers pcdhps-07 and pcdhps-08, respectively. Products were fractionated in an ABI 377 sequencer. The heights of the individual fluorescent peaks at polymorphic sites 1–4 (Fig. 1) were then measured using Seqed 1.0.3 software (Applied Biosystems) (Fig. 2). Non-polymorphic peaks were also measured in the forward and reverse directions (Fig. 1).

2.4. Data analysis

An index of the percentage of either the AS allele, f_{AS} (called the 'output %AS'), or the AJ allele, f_{AJ}, in a mixture of both alleles of *pcdhps* was calculated by three types of treatment of the data, as follows:

(i) A 'simple' treatment, in which the proportion of AS in the mixture, f_{AS} ('output %AS'), was expressed as the percentage of the AS peak height, h_{AS}, compared to the total of the two peak heights $h_{AS} + h_{AJ}$. Thus, $f_{AS} = h_{AS} \times 100/(h_{AS} + h_{AJ})$. Simi-