Available online at www.sciencedirect.com
ScienceDirect
Gene 387 (2007) 1–6
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Genomic heterogeneity in the density of noncoding single-nucleotide and microsatellite polymorphisms in *Plasmodium falciparum*

Sarah K. Volkman [a], Elena Lozovsky [b], Alyssa E. Barry [c], Trevor Bedford [b], Lara Bethke [a], Alissa Myrick [a,1], Karen P. Day [c], Daniel L. Hartl [b], Dyann F. Wirth [a,*], Stanley A. Sawyer [d]

[a] *Department of Immunology and Infectious Disease, Harvard School of Public Health, Boston, MA, USA*
[b] *Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA*
[c] *Department of Medical Parasitology, New York University School of Medicine, New York, NY, USA*
[d] *Department of Mathematics, Washington University, St. Louis, MO, USA*

Received 28 February 2006; received in revised form 6 July 2006; accepted 24 July 2006
Available online 14 August 2006
Received by F.G. Alvarez-Valin

Abstract

The density and distribution of single-nucleotide polymorphisms (SNPs) across the genome has important implications for linkage disequilibrium mapping and association studies, and the level of simple-sequence microsatellite polymorphisms has important implications for the use of oligonucleotide hybridization methods to genotype SNPs. To assess the density of these types of polymorphisms in *P. falciparum*, we sampled introns and noncoding DNA upstream and downstream of coding regions among a variety of geographically diverse parasites. Across 36,229 base pairs of noncoding sequence representing 41 genetic loci, a total of 307 polymorphisms including 248 polymorphic microsatellites and 39 SNPs were identified. We found a significant excess of microsatellite polymorphisms having a repeat unit length of one or two, compared to those with longer repeat lengths, as well as a nonrandom distribution of SNP polymorphisms. Almost half of the SNPs localized to only three of the 41 genetic loci sampled. Furthermore, we find significant differences in the frequency of polymorphisms across the two chromosomes (2 and 3) examined most extensively, with an excess of SNPs and a surplus of polymorphic microsatellites on chromosome 3 as compared to chromosome 2 ($P=0.0001$). Furthermore, at some individual genetic loci we also find a nonrandom distribution of polymorphisms between coding and flanking noncoding sequences, where completely monomorphic regions may flank highly polymorphic genes. These data, combined with our previous findings of nonrandom distribution of SNPs across chromosome 2, suggest that the *Plasmodium falciparum* genome may be a mosaic with regard to genetic diversity, containing chromosomal regions that are highly polymorphic interspersed with regions that are much less polymorphic.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Genetic diversity; Coalescent analysis; Malaria

1. Introduction

Genetic variability of *Plasmodium falciparum* underlies its transmission success and thwarts efforts to control disease. Resistance to all antimalarial drugs arises rapidly and is mediated by mutations in key target (Peterson et al., 1990; Wang et al., 1997) or transport genes (Foote et al., 1990; Fidock et al., 2000). Antigenic variation in key proteins is the basis of immune evasion (Biggs et al., 1991) and may in part explain the lack of sterilizing immunity in humans (Bruce-Chwatt, 1963; Neva, 1977; Marsh, 1992). Thus, mutation is a key virulence determinant in the parasite. Yet, there appears to be nonrandomness in the frequency

Abbreviations: SNPs, single-nucleotide polymorphisms; MLEs, maximum likelihood estimates; N_e, effective population size; μ, nucleotide-site mutation rate; bp, base pair(s); θ, theta; kb, kilobase(s); cM, centimorgan; CI, confidence interval; MS, microsatellite; Tyr, tyrosine; Phe, phenylalanine; Cys, cysteine; Asp, asparagines.

* Corresponding author. 665 Huntington Avenue, I-703, Boston, MA 02115, USA. Tel.: +1 617 432 4629; fax: +1 617 432 4766.
E-mail address: dfwirth@hsph.harvard.edu (D.F. Wirth).
[1] Current address: Division of Infectious Diseases, San Francisco General Hospital, University of California, San Francisco, CA, USA.

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.07.026

at which polymorphic mutations occur in different genes in *P. falciparum*. Genetic variation in antigenic (Kemp et al., 1990), drug resistance (Foote et al., 1990; Su et al., 1997; Biswas et al., 2000; Fidock et al., 2000) and pathogenesis determinants is abundant, whereas SNPs in housekeeping genes (genes involved in general cellular or metabolic processes that are not obvious targets of rapid selective change) and introns is much less abundant (Rich et al., 1998; Volkman et al., 2001). This contrast has led to the hypothesis that the *P. falciparum* genome may be a mosaic with regions of significant diversity juxtaposed with regions that are practically monomorphic. Possible explanations for such a mosaic pattern include chance differences in the average coalescence time of genes in different chromosomes, or that the parasite population has undergone a series of selective sweeps, resulting in a patchwork genome with different evolutionary histories. Assessing the density of polymorphisms therefore has important implications, both in terms of the population structure and in terms of the identification of genes under diversifying selection, including potential drug and vaccine targets.

There is ample evidence for recent selective sweeps within the genome associated with the emergence and subsequent spread of drug resistance. Regions of the genome that have undergone recent selective sweeps include regions of chromosome 7 surrounding the *pfcrt* locus (Wootton et al., 2002) and regions of chromosome 4 surrounding the *dhfr* (Nair et al., 2003; Roper et al., 2003) and *dhps* (Roper et al., 2004) loci. These selective sweeps are recent in the evolutionary history of *P. falciparum*, presumably because they are a consequence of recently applied drug pressure. Other selective sweeps based on biological or immune selection may also have occurred as well as older selective sweeps, but it may require a greater depth of analysis to reveal these regions using a population genetics approach.

To investigate SNP density and microsatellite polymorphisms in more detail, we carried out an in-depth analysis of sequence diversity in the noncoding flanking and intronic sequences in *P. falciparum*. We focused our analysis on genes from chromosomes 2 and 3, primarily to reconcile previous work that yielded conflicting estimates of the SNP frequency between these two chromosomes (Volkman et al., 2001; Mu et al., 2002, 2005); however, we also included sequences from other chromosomes in our analysis to provide a more global view of the genome. We identified and compared polymorphisms both on a genome-wide basis and directly between chromosomes 2 and 3. We provide evidence for dramatic differences in the frequency of polymorphisms, both among genes in each chromosome and overall between chromosomes 2 and 3. These results may imply different evolutionary histories for these chromosomes, consistent with drift events or selective sweeps.

2. Materials and methods

2.1. Identification of polymorphisms

Genetic loci were amplified using the polymerase chain reaction, cloned and sequenced from various *P. falciparum* isolates (HB3, 7G8, D6, W2, Muz 12.4, 3D7, D10, Muz 37.4, Muz 51.1 and KF1776) as described (Volkman et al., 2001). These isolates represent both the K1-type and the MAD20-type alleles of *msp1*. Parasite DNA was genotyped using the *msp2* locus (Snounou et al., 1999) to ensure independent parasite isolates were used and sequences were aligned to generate consensus sequences for each of the isolates at each of the loci to determine the polymorphisms among isolates.

All sequences were scanned for microsatellite repeats using Tandem Repeats Finder (Benson, 1999). Parameters were chosen so that repeats identified by the program encompassed as many length polymorphic microsatellites as possible, while ignoring extraneous sequences. Regions of at least 12 repetitive base pairs were identified as repeats, regardless of the number of base pairs within the repeating unit. These parameters allowed identification of greater than 98% of the polymorphic regions in the dataset. The remaining polymorphic regions that the program did not identify were entered into the dataset manually. This method allowed for an objective, consistent, and unbiased determination of microsatellite sequence boundaries. For further information about this method, as well as precise parameters values please see: http://www.oeb.harvard.edu/hartl/lab/publications/PlasmoMS/index.html

2.2. Data analysis

The compiled dataset was analyzed by polymorphism type, polymorphisms within repetitive or nonrepetitive sequences, by chromosome, and by noncoding region (flanking or intronic sequence) to identify any significant differences regarding the frequency or distribution of polymorphisms within the dataset as described in the text. Polymorphism types included SNPs (both outside of and within microsatellites), microsatellite repeat-length polymorphisms, and other polymorphisms including small insertions and deletions. Tests of significance are noted in the text, but included both Fishers Exact Test and the Poisson Test using 2×2 contingency analysis.

2.3. Coalescent analysis

The infinite-sites model was used to estimate the product (θ) of the effective population size (N_e) and the nucleotide-site mutation rate (μ) for each of chromosomes 2 and 3. The model assumes that nucleotide sites within individual loci are tightly linked, but with free recombination between loci. Under these assumptions, likelihoods for θ for each chromosome can be written as a product of likelihoods over loci. Infinite-sites likelihoods for individual loci were derived using a recursion of Griffiths and Tavare (Griffiths and Tavare, 1994, 1995). Bayesian methods were used to estimateθ, so that the product likelihoods were also multiplied by a noninformative prior.

3. Results

3.1. Identification of polymorphisms

Forty-one loci from various chromosomes were sequenced from five geographically distinct parasite genomes including

Download English Version:

https://daneshyari.com/en/article/2819908

Download Persian Version:

https://daneshyari.com/article/2819908

Daneshyari.com