OPEN ACCESS Freely available online

PLoS one

Characterization of Within-Host *Plasmodium falciparum* Diversity Using Next-Generation Sequence Data

Sarah Auburn[1,2]*, Susana Campino[1☯], Olivo Miotto[3,4☯], Abdoulaye A. Djimde[5], Issaka Zongo[6], Magnus Manske[1], Gareth Maslen[1], Valentina Mangano[7], Daniel Alcock[1], Bronwyn MacInnis[1], Kirk A. Rockett[1,8], Taane G. Clark[1,9], Ogobara K. Doumbo[5], Jean Bosco Ouédraogo[6], Dominic P. Kwiatkowski[1,3,8]

1 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom, **2** Global Health Division, Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia, **3** Centre for Genomics and Global Health, University of Oxford, Oxford, United Kingdom, **4** Mahidol-Oxford Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, **5** Malaria Research and Training Centre, Faculty of Medicine, University of Bamako, Bamako, Mali, **6** Institut de Recherche en Sciences de la Santé, Direction Régionale de l'Ouést, Bobo-Dioulasso, Burkina Faso, **7** Department of Public Health Sciences, Section of Parasitology, University of Rome La Sapienza, Rome, Italy, **8** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **9** London School of Hygiene & Tropical Medicine, Keppel Street, London, United Kingdom

Abstract

Our understanding of the composition of multi-clonal malarial infections and the epidemiological factors which shape their diversity remain poorly understood. Traditionally within-host diversity has been defined in terms of the multiplicity of infection (MOI) derived by PCR-based genotyping. Massively parallel, single molecule sequencing technologies now enable individual read counts to be derived on genome-wide datasets facilitating the development of new statistical approaches to describe within-host diversity. In this class of measures the F_{WS} metric characterizes within-host diversity and its relationship to population level diversity. Utilizing *P. falciparum* field isolates from patients in West Africa we here explore the relationship between the traditional MOI and F_{WS} approaches. F_{WS} statistics were derived from read count data at 86,158 SNPs in 64 samples sequenced on the Illumina GA platform. MOI estimates were derived by PCR at the *msp-1* and *-2* loci. Significant correlations were observed between the two measures, particularly with the *msp-1* locus ($P = 5.92 \times 10^{-5}$). The F_{WS} metric should be more robust than the PCR-based approach owing to reduced sensitivity to potential locus-specific artifacts. Furthermore the F_{WS} metric captures information on a range of parameters which influence out-crossing risk including the number of clones (MOI), their relative proportions and genetic divergence. This approach should provide novel insights into the factors which correlate with, and shape within-host diversity.

Citation: Auburn S, Campino S, Miotto O, Djimde AA, Zongo I, et al. (2012) Characterization of Within-Host *Plasmodium falciparum* Diversity Using Next-Generation Sequence Data. PLoS ONE 7(2): e32891. doi:10.1371/journal.pone.0032891

Editor: Georges Snounou, Université Pierre et Marie Curie, France

Received January 11, 2012; **Accepted** February 7, 2012; **Published** February 29, 2012

Funding: The genetic component of this research was funded by the Wellcome Trust, the Bill and Melinda Gates Foundation, and the Medical Research Council. SA, SC, OM, MM, GM, DA, BM, KAA, TGC and DPK were funded by the Wellcome Trust, the Bill and Melinda Gates Foundation, and the Medical Research Council. AAD is supported by a European and Developing Countries Clinical Trial Partnership Senior Fellowship (grant 2004.2.C.f1) and a Howard Hughes Medical Institution International Scholarship (grant 55005502). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: sa3@sanger.ac.uk

☯ These authors contributed equally to this work.

Introduction

Particularly in high transmission regions, individuals may carry multiple genetically distinct parasite clones in a single malaria infection (multi-clonal infection). This dynamic poses numerous challenges to malaria control. Recombination between genetically distinct parasite clones (out-crossing) is a major risk factor for generating novel parasite variants with clinically important phenotypes such as virulence, drug resistance or immune evasion. Parasite diversity dynamics also have important implications for host immune acquisition, and influence the selection and spread of drug resistance [1,2,3,4,5]. Furthermore, the intense competition enabled between distinct clones within an infection appears to promote the evolution of highly virulent *Plasmodium* parasites [6,7,8,9]. Numerous studies have sought correlations between parasite multi-clonality and clinical phenotypes such as patient age, immune status, clinical outcome, drug sensitivity, gametocyte production and gametocyte infectivity [10,11,12,13,14,15,16,17]. However, these studies have been constrained by the limited availability of tools to effectively gauge the genetic composition of multi-clonal infections. Constrained in part by these limitations we still have only a poor understanding of how within-host diversity is shaped in different epidemiological settings.

Our current understanding is that multi-clonal infections may be generated by super-infection (the composition of multiple infectious bites with genetically distinct clones), and that this may be a frequent occurrence in high transmission areas. However, it has also been demonstrated that a single mosquito inoculation may carry a large amount of parasite diversity [18], and that this again may be dependent on the population level diversity. In addition, within-host dynamics may be shaped by factors such as host immunity and density-dependent control mechanisms [19]. In order to better understand the dynamics of multiple-clone infections and parasite out-crossing, effective tools to address

within-host diversity within the context of the population-level diversity are essential.

Traditionally, within-host parasite diversity has been described by the number of distinct clones within an infection, or, multiplicity of infection (MOI). A common approach is the use of PCR to derive the number of repeat length variants observed at a few highly diverse loci, most frequently variable number tandem repeats (VNTRs) in the genes encoding antigens such as the merozoite surface protein 1 (MSP1) and 2 (MSP2) [20,21]. The large repertoire of alleles at a single VNTR locus is practical for finger-printing malaria parasites, particularly for drug surveillance. However, limitations on the number and "nature" of the loci constrain effective characterization of within-host diversity. Furthermore, as the clone counting approach does not account for the population level of diversity, limited insight is gained into the risks of out-crossing in a given population.

Massively parallel sequencing technologies now enable characterization of parasite diversity at high depth and molecular resolution [22,23,24]. Individually, SNPs do not capture as much diversity as VNTRs, and so the clone counting MOI method is not effective for SNP data. Alternatively, the single molecule sequencing approach of platforms such as the Illumina Genome Analyzer enables individual allele counts to be derived at each SNP position, and the opportunity for new statistical approaches to describe within-host parasite diversity based on population genetics such as the F_{WS} metric (Manske, Miotto et al., in preparation). This metric characterizes not just within-host diversity, but also its relationship to local population diversity, essentially measuring the risk of out-crossing/inbreeding.

Here, using PCR at the *msp1* and *2* loci, we explore the relationship between the traditional VNTR-based MOI and genome-wide, SNP-based F_{WS} metric in clinical (non-cultured) *Plasmodium falciparum* samples from malaria-endemic regions in West Africa.

Results and Discussion

PCR-based estimation of MOI

The traditional method for characterizing within-host diversity in *P. falciparum* entails MOI estimation using genotype data from the *msp-1* and *-2* loci [20]. This approach utilizes family specific and repeat length variants in each gene to differentiate parasite clones (see Methods). In the West African samples, genotyping success rates at the *msp-1* and *-2* loci were 100% (64/64) and 98% (63/64), respectively. The distributions of MOI estimates for each assay/population are presented in Figure S1. A summary of these distributions is presented in Table 1. MOI estimates were slightly higher in Burkina Faso (mean = 2.95) than Mali (mean = 2.57), but the difference was not significant (t = 1.28, $P = 0.209$). These estimates are similar to previous observations in Mali and Burkina Faso [25,26]. In the combined population (West Africa), the mean MOI estimate was 2.83. The gene-specific MOI estimates demonstrated similar distributions to the maximum MOI (Figures S2, 3), although *msp-2* estimates (mean = 2.54) were moderately higher than *msp-1* (mean = 2.19) (t = −1.82; $P = 0.071$). Moderate gene-specific differences were also observed between populations. The *msp-1* MOI estimates were higher in Burkina Faso (mean = 2.33) than Mali (mean = 1.90), with borderline significance (t = 1.94, $P = 0.059$). The *msp-2* MOI estimates were only moderately higher in Burkina Faso (mean = 2.64) than Mali (mean = 2.33), and the difference was not significant (t = 0.91, $P = 0.368$). In summary, the MSP MOI data indicated high multi-clonality in West Africa, with slightly higher levels in Burkina Faso than Mali, and subtle differences in the estimates derived from the *msp-1* and *-2* loci.

Genome-wide characterization of within-host diversity (F_{WS})

The F_{WS} metric describes the relationship between the diversity observed within a patient to that of the population using estimations of heterozygosity (i.e. the probability that two randomly selected parasites carry different alleles at a given locus). In any given population, heterozygosity is dependent on the allele frequencies at that locus, as described by the Hardy-Weinberg principle, and is therefore influenced by the level of out-crossing in the population. By combining allele counts across all samples in the population, we can derive an estimate of population allele frequencies at each locus, and consequently estimate heterozygosity at the population level (H_S). Similarly, for each sample, using allele counts at each locus, we can estimate within-host heterozygosity (H_W), enabling computation of the F_{WS} measure. Essentially, the F_{WS} metric provides a measure of the risk of out-crossing between the parasites within an individual to generate new genotypes during recombination in the mosquito host. Thus, the metric captures information on all of the sample parameters which influence the risk of new parasite genotypes being generated at meiosis. These include not just the overall diversity within an individual but also the level of similarity between the parasites and their relative proportions. A low F_{WS} reflects a low risk of inbreeding/high risk of out-crossing and thus high within-host diversity. Whilst simple calculations of the heterozygosity at each locus within a sample provide some level of information on within-host diversity, the ability to describe this diversity in the context of the level of diversity in the population is critical to capturing information on the risk of out-crossing. Consequently, population level heterozygosity estimates are essential for effective computation and interpretation of the F_{WS} metric.

We previously described F_{WS} distributions in a range of *P. falciparum* populations, demonstrating global patterns of within-host diversity consistent with our knowledge of the epidemiology and human (and parasite) migration patterns in the regions studied (Manske, Miotto *et al.*, in preparation). In the West African samples addressed here, a mean F_{WS} of 0.73 was observed, with 28% samples exhibiting "high" F_{WS} estimates (i.e. $\geq$0.95: see Figure S2). In each of Mali and Burkina Faso, the mean F_{WS} scores were 0.81

Table 1. Summary of MOI and F_{WS} estimates in Burkina Faso, Mali and the combined (West Africa) population.

	Burkina Faso	Mali	All (West Africa)
Mean Maximum MOI (range)	2.95 (1–5)	2.57 (1–4)	2.82 (1–5)
Mean MSP1 MOI (range)	2.33 (1–4)	1.91 (1–4)	2.19 (1–4)
Mean MSP2 MOI (range)	2.64 (1–5)	2.33 (1–4)	2.54 (1–5)
Mean F_{WS} (range)	0.69 (0.19–1.00)	0.81 (0.32–1.00)	0.73 (0.19–1.00)

doi:10.1371/journal.pone.0032891.t001

(29% ≥0.95) and 0.69 (30% ≥0.95) respectively. The low proportion of samples with high F_{WS} scores is indicative of a large degree of panmixis (low sub-structure) amongst the parasites in the populations sampled here. Presumably, super-infection with genetically distinct parasites is a relatively frequent occurrence. It follows that the risks of parasite out-crossing, and consequent threat to malaria control efforts, are moderately high in these populations.

In accordance with the *msp*-based results, greater within-host diversity (lower F_{WS}) was observed in Burkina Faso (mean = 0.69) than Mali (mean = 0.81) (t = −2.08, P = 0.042). This observation may reflect subtle differences in the epidemiology of the sites studied here, such as rural (Mali) versus urban (Burkina Faso) influences on transmission dynamics [27,28]. Further exploration of the F_{WS} profile in different epidemiological settings is required to address this and other putative determinants of within-host parasite diversity. This should also enable more effective interpretation of the relative out-crossing risks associated with different F_{WS} scores.

>A wide range of F_{WS} scores were observed in the West African populations (Figure S2). Thus, even after gauging within-host diversity estimates in terms of the out-crossing risk and related parameters (number, relative proportion, and degree of divergence between clones), there appears to be heterogeneity in within-host diversity in the West African populations. It remains to be determined how well this heterogeneity reflects the heterogeneity in clinical outcomes such as disease severity, gametocyte prevalence and infectivity, patient age and immune status, in West Africa and other populations. To date, the parasite genetic basis of these outcomes has generally only been addressed with regard to clone count measures of within-host diversity (review in [29]). The extent to which the clone counting MOI approach reflected the additional diversity parameters captured by the F_{WS} metric remained uncertain. We therefore explored this relationship further.

Relationship between genome-wide F_{WS} metric and MSP1+MSP2 MOI estimates

Using the F_{WS} score as a proxy to genome-wide within-host SNP diversity, we assessed the correlation between this metric and the MSP-based MOI estimates. A significant negative correlation was observed between the MSP1 MOI and F_{WS} (rho = −0.480, $P = 5.92\times10^{-5}$), as illustrated in Figure 1. The F_{WS} correlations with each of the MSP2 and maximum MOI estimates were less strong but both remained significant (rho = −0.366, P = 0.003; rho = −0.384, P = 0.002, respectively) (Figure 1). These correlations demonstrated that the MOI and F_{WS} metric broadly agreed on their interpretations of within-host diversity. However, differences in two key areas, 1) the analytical approach, and 2) the number and nature of loci examined, may underlie the remaining deviations between the two measures.

A moderate proportion of samples (13/64 [20.3%]) demonstrated multiple clones by MSP genotyping but appeared to be largely "clonal" with respect to the F_{WS} (≥0.95). These results presumably reflect infections with multiple clones exhibiting limited divergence and thus limited within-host diversity. As detailed below, any of a number of differences, both technical and analytical, between the PCR-based MOI and F_{WS} approaches may underlie these discrepancies. Fewer samples (4/64 [6.25%]) exhibited a clonal MSP outcome concurrent with high diversity F_{WS} (≤0.70). The ability to capture the relatedness between the clones within an infection should facilitate our understanding of the impact of within-host regulatory effects such as inter-parasite competition and host immunity on parasite diversity dynamics.

In addition to the core analytical approach, the PCR-based MOI approach may be constrained in ability to capture the complexity of an individual infection owing to sensitivity to the number of loci examined, the level of polymorphism at each locus, family-specific versus intra-family variation and selective pressures at one or more loci. In these respects, the genome-wide, SNP-based, F_{WS} approach is more robust. In general, even with highly diverse loci, more markers ensure a more accurate representation of diversity [30]. Indeed, a larger number of multiple clone infections were detected by the maximum MOI (92% [59/64]) than with the *msp-1* (79% [50/63]) and *-2* (77% [49/64]) loci individually. However, practicalities limit the number of loci that can be examined by PCR.

Owing to differences in the level and nature of polymorphism, inter-locus variation in diversity is not unexpected within a multi-clonal malaria infection. Indeed, the *msp-1* and *-2* genes exhibited modest differences in polymorphism, reflected in their different strengths of correlation with the F_{WS} metric (Figure 1). Inter-locus differences may also result from locus-specific selective pressures which are not representative of the genome as a whole. This is a potential limitation of MSP1 and 2, which appear to be under selective pressure from the host immune system [31].

The distinction of alleles at VNTR loci by electrophoretic mobility is another potential limitation of the PCR-based approach, as differences in sequence composition between equally migrating fragments may be missed. Furthermore, it is generally accepted that VNTRs evolve at different rates and under different mechanisms from SNPs [32,33,34,35]. Thus, VNTRs may not provide an effective representation of the SNP diversity in the genome. Rather, as observed with the moderately high frequency (20.3%) of isolates demonstrating largely clonal dynamics with F_{WS} estimates but polyclonal dynamics at the MSP loci, VNTR-based estimates may overestimate the true genome-wide within-host diversity.

A potential shortfall of the F_{WS} approach is the integrity of the genome-wide dataset. A highly stringent SNP discovery and allele-calling process was employed here to ensure high confidence in the dataset (see Methods). In addition, only good quality samples with high sequence depth were included (Figure S3). The comparable sensitivity of the PCR-based and genome-sequencing approach has yet to be determined. However, in contrast to the potential ambiguity in detecting and distinguishing bands on gels, the ability to count individual reads in Illumina sequence data enables more objective allele-calling.

Juliano and colleagues achieved extensive sequence depth at the *msp-1* and *-2* loci by focusing 454 sequencing on amplicons of these two genes alone [36]. Using the clone-counting approach, facilitated by haplotype information, an average of ~5 more *msp-1* and *-2* variants were detected relative to the genotyping data. As with the PCR-based MOI method, this approach is powerful for distinguishing clones in clinical drug trials. For characterization of within-host diversity however, the use of just 2 VNTR loci under potential selective pressure renders this approach sensitive to the features discussed above. With continual improvements in read length and sequence yield on high-throughput, single-molecule sequencing platforms, haplotype-based clone counting approaches similar to those described by Juliano and colleagues should be feasible on a broader, genome-wide, scale.

Conclusion

The traditional approach to estimating MOI using VNTRs at a handful of loci such as *msp-1* and *-2* is a simple and moderately cheap method with demonstrated practicality for finger-printing clones in drug trials. However, genome-wide measures such as the F_{WS} statistic, by capturing information not just on the number of clones in a sample but also on their respective ratios and the degree of inter-clone variation at hundreds of thousands of polymorphic positions, provide a greater wealth of information on within-host

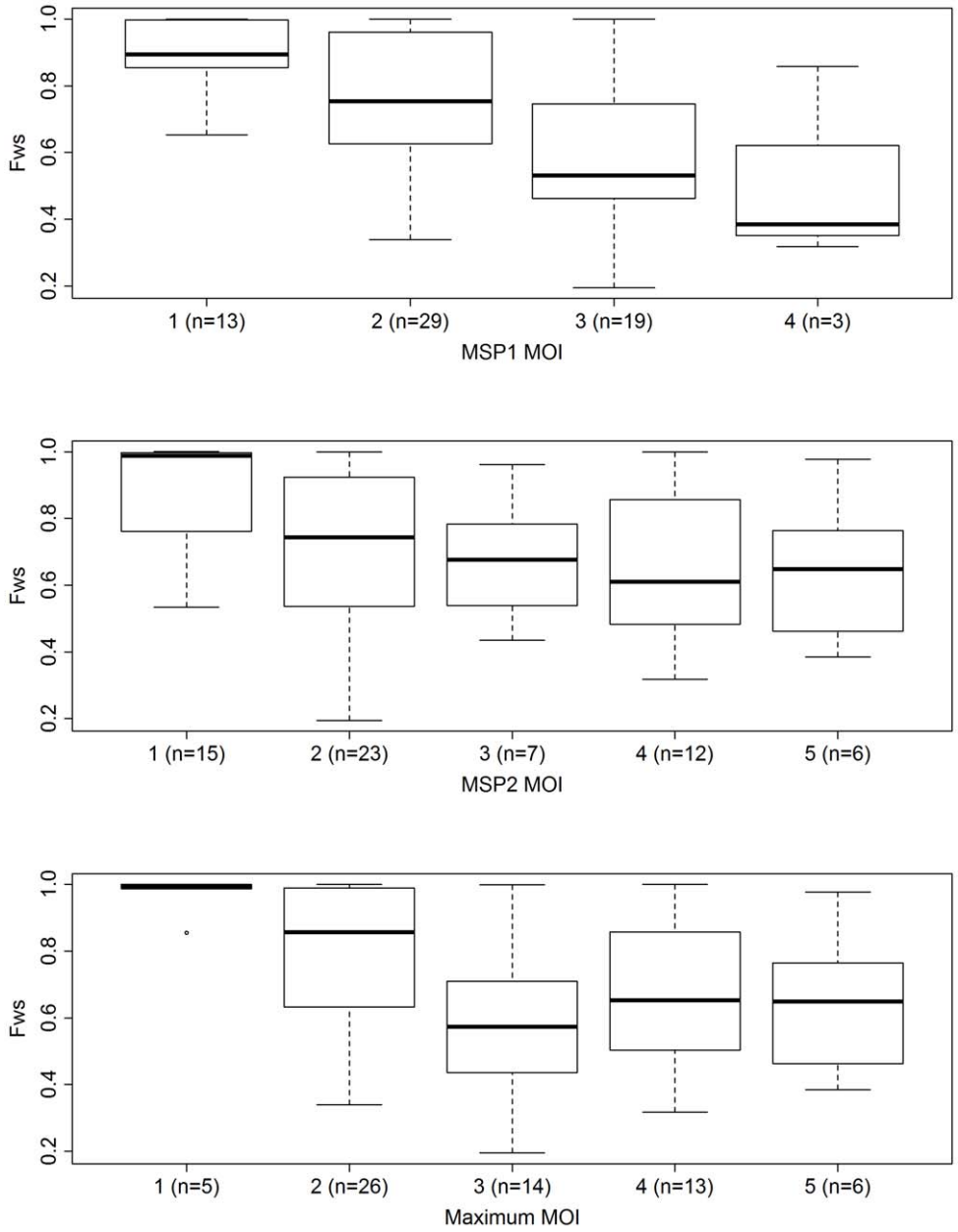

Figure 1. F_{WS} against MSP1 MOI, MSP2 MOI, and maximum MOI.
doi:10.1371/journal.pone.0032891.g001

diversity which is highly robust against potential locus-specific artifacts. With rapidly advancing progress in whole genome sequencing technologies, including ever increasing read lengths and continual reductions in cost, this approach is now feasible in hundreds of samples. The ability to assess within-host diversity in genome-wide datasets from non-cultured (clinical) samples sourced from low resource field settings, as demonstrated here, should revolutionize clinical and epidemiological studies of *Plasmodium*.

Methods

Samples

Samples were collected from field sites in Mali and Burkina Faso within the framework of a large, multi-center *P. falciparum* genome sequencing project to facilitate SNP discovery and population genetic characterization (Manske, Miotto *et al.*, in preparation). Consenting patients of all ages and ethnic groups presenting at the clinic with symptoms of uncomplicated malaria and *P. falciparum* parasitaemia were recruited to the study. In Mali, samples were collected from clinics in two rural villages (Kolle and Faladje). In Burkina Faso, samples were collected from permanent health dispensaries in three suburbs of Bobo-Dioulasso (Colsamma, Ouezzinville, and Sakaby). Details of the sample processing procedures in the laboratory are described elsewhere [37]. Briefly, venous blood (2–8 ml) was processed to deplete the human white blood cell fraction, and DNA extraction was undertaken using Qiagen QIAamp blood extraction kits as per the manufacturer's instructions.

Ethics

Informed, written consent was obtained from patients over 18 years of age and from a parent or guardian for younger patients. The study was approved by the Comite d'Ethique de la Faculté de

Médecine de Pharmacie et d'Odontostomatologie, Bamako, Mali, and Comite d'Ethique Institutionnel du Centre Muraz, Bobo-Dioulasso, Burkina Faso.

Sequencing and SNP Calling

Sixty-four samples (21 Mali, 43 Burkina Faso) with >500 ng total DNA and <60% human DNA contamination were sequenced on the Illumina Genome Analyzer platform. Details of the sequencing, SNP discovery and genotype calling process are described elsewhere (Manske, Miotto *et al.*, in preparation). Briefly, up to 6 lanes were sequenced per sample, with 37, 54 or 76 bp paired-end reads. Coverage distributions are presented in Figure S3. The "raw" sequence data for the samples can be accessed in the European Nucleotide Archive (www.ebi.ac.uk/ena/data/search/?query = plasmodium). For each sample, sequence data from multiple lanes was merged, and mapped to the *P. falciparum* reference genome (3D7 version 2.1.5; ftp://ftp.sanger.ac.uk/pub/pathogens/Plasmodium/falciaprum/3D7/3D7.version2.1.5) using the *bwa* program (available from http://bio-bwa.sourceforge.net) with the default parameters [38]. The resulting alignments were processed in *samtools* (http://samtools.sourceforge.net/cns0.shtml) using the default parameters to identify positions with one or more bases differing from the reference sequence (*putative* SNPs). Genotypes were called at 86,158 high confidence SNP positions (*typable* SNPs) identified by a stringent quality-filtering SNP discovery process detailed elsewhere (Manske, Miotto *et al.*, in preparation). This SNP list is referred to as version 1gamma and is available on the MapSeq database (www.mapseq.net/pf). For the purpose of assigning confident genotype calls to the dataset, all genotypes in the version 1 gamma dataset with less than 5 reads were assigned a status of missing data.

Calculation of F_{WS}

Details of the underlying basis for calculations used to derive the F_{ws} metric are described elsewhere (Manske, Miotto *et al.*, in preparation). The F_{WS} metric was calculated for each individual sample, using the formula $F_{WS} = 1-(Hw/Hs)$, where Hw is the within-individual heterozygosity and Hs is the within-population heterozygosity. At each biallelic SNP, heterozygosity was estimated using the formula $H = 1-(p^2+q^2)$, where p and q are the frequencies of the two alleles ($p = 1-q$). At each SNP, p and q were estimated for each individual as the proportions of sequencing reads that carried each allele in the individual sample. At population level, the allele frequencies at the SNP were estimated as the mean of the allele frequencies in the individuals comprising the population sample; the frequency of the least common allele at that SNP in the West African population was noted as the *minor allele frequency* (MAF). Since heterozygosity depends on allele frequencies, each of the 86,158 SNPs was assigned to one of ten equally-sized MAF intervals ([0.0–0.05], [0.05–0.1] … [0.45–0.5]). Estimates of Hw and Hs were calculated for each MAF interval, as the means of the Hw and Hs across all SNPs in that interval. For each individual, Hw estimates were plotted against the corresponding Hs estimates for all MAF intervals, and the slope of the resulting straight-line plot was used to evaluate the Hw/Hs ratio, and thus compute F_{WS} for the individual.

Genotype-based MOI Estimation

The traditional method of genotyping in the *P. falciparum* merozoite surface protein - 1 (*msp-1*) and -2 (*msp-2*) genes was undertaken on 64 of the typable samples using a nested PCR approach [20]. This method utilises family specific variation and variation in repeat length in each gene to distinguish parasite clones. In general, each parasite will exhibit one of three family-specific sequences, K1, R033 or MAD20, in the *msp-1* gene, and one of two family-specific sequences, FC27 or IC/3D7, in the *msp-2* gene. Within each family-specific sequence, regions comprising repeat-length polymorphisms enable further distinction of clones. The number of distinct bands observed in each sample was summed across the families in each of the *msp-1* and *-2* genes separately, providing gene-specific MOI estimates. The maximum MOI estimate was then derived as the maximum gene-specific MOI estimate for each sample. The standard Welch two-sample t-test was used to measure the significance of difference in MOI estimates between populations and assays using R software.

Supporting Information

Figure S1 MSP-based MOI Estimates.
(TIFF)

Figure S2 Distribution of F_{ws} scores in the West African samples. Dashed lines indicate thresholds for highly diverse ($F_{ws} \leq 0.7$) and moderately "clonal" ($F_{ws} \geq 0.95$) samples.
(TIFF)

Figure S3 Read depth distribution at the typable SNP positions in the West African samples. Distribution of SNP coverage at read depth (number of sequenced nucleotides covering a given locus) thresholds of 1, 5, 10, 15, 20 and 25.
(TIFF)

Acknowledgments

We wish to thank the patients who donated samples, the numerous individuals who facilitated sample collections in the field sites, and the Wellcome Trust Sanger Institute library preparation and sequencing teams.

Author Contributions

Conceived and designed the experiments: SA SC OM DPK. Performed the experiments: SA SC VM DA BM. Analyzed the data: SA OM MM GM TGC. Contributed reagents/materials/analysis tools: AAD IZ OKD JBO KAR DPK. Wrote the paper: SA OM. All authors read and approved the final version of the manuscript.

References

1. Hastings IM, D'Alessandro U (2000) Modelling a predictable disaster: the rise and spread of drug-resistantmalaria. Parasitol Today 16: 340–347.
2. Contamin H, Fandeur T, Rogier C, Bonnefoy S, Konate L, et al. (1996) Different genetic characteristics of Plasmodium falciparum isolates collected during successive clinical malaria episodes in Senegalese children. Am J Trop Med Hyg 54: 632–643.
3. Forsyth KP, Philip G, Smith T, Kum E, Southwell B, et al. (1989) Diversity of antigens expressed on the surface of erythrocytes infected with mature Plasmodium falciparum parasites in Papua New Guinea. Am J Trop Med Hyg 41: 259–265.
4. Marsh K, Otoo L, Hayes RJ, Carson DC, Greenwood BM (1989) Antibodies to blood stage antigens of Plasmodium falciparum in rural Gambians and their relation to protection against infection. Trans R Soc Trop Med Hyg 83: 293–303.
5. Newbold CI, Pinches R, Roberts DJ, Marsh K (1992) Plasmodium falciparum: the human agglutinating antibody response to the infected red cell surface is predominantly variant specific. Exp Parasitol 75: 281–292.
6. Bell AS, de Roode JC, Sim D, Read AF (2006) Within-host competition in genetically diverse malaria infections: parasite virulence and competitive success. Evolution 60: 1358–1371.
7. de Roode JC, Culleton R, Bell AS, Read AF (2004) Competitive release of drug resistance following drug treatment of mixed Plasmodium chabaudi infections. Malar J 3: 33.
8. de Roode JC, Pansini R, Cheesman SJ, Helinski ME, Huijben S, et al. (2005) Virulence and competitive ability in genetically diverse malaria infections. Proc Natl Acad Sci U S A 102: 7624–7628.

9. Mackinnon MJ, Read AF (2004) Virulence in malaria: an evolutionary viewpoint. Philos Trans R Soc Lond B Biol Sci 359: 965–986.
10. al-Yaman F, Genton B, Reeder JC, Anders RF, Smith T, et al. (1997) Reduced risk of clinical malaria in children infected with multiple clones of Plasmodium falciparum in a highly endemic area: a prospective community study. Trans R Soc Trop Med Hyg 91: 602–605.
11. Engelbrecht F, Togel E, Beck HP, Enwezor F, Oettli A, et al. (2000) Analysis of Plasmodium falciparum infections in a village community in Northern Nigeria: determination of msp2 genotypes and parasite-specific IgG responses. Acta Trop 74: 63–71.
12. Felger I, Smith T, Edoh D, Kitua A, Alonso P, et al. (1999) Multiple Plasmodium falciparum infections in Tanzanian infants. Trans R Soc Trop Med Hyg 93 Suppl 1: 29–34.
13. Mayor A, Saute F, Aponte JJ, Almeda J, Gomez-Olive FX, et al. (2003) Plasmodium falciparum multiple infections in Mozambique, its relation to other malariological indices and to prospective risk of malaria morbidity. Trop Med Int Health 8: 3–11.
14. Owusu-Agyei S, Smith T, Beck HP, Amenga-Etego L, Felger I (2002) Molecular epidemiology of Plasmodium falciparum infections among asymptomatic inhabitants of a holoendemic malarious area in northern Ghana. Trop Med Int Health 7: 421–428.
15. Smith T, Beck HP, Kitua A, Mwankusye S, Felger I, et al. (1999) Age dependence of the multiplicity of Plasmodium falciparum infections and of other malariological indices in an area of high endemicity. Trans R Soc Trop Med Hyg 93 Suppl 1: 15–20.
16. Taylor LH, Walliker D, Read AF (1997) Mixed-genotype infections of malaria parasites: within-host dynamics and transmission success of competing clones. Proc Biol Sci 264: 927–935.
17. Zwetyenga J, Rogier C, Tall A, Fontenille D, Snounou G, et al. (1998) No influence of age on infection complexity and allelic distribution in Plasmodium falciparum infections in Ndiop, a Senegalese village with seasonal, mesoendemic malaria. Am J Trop Med Hyg 59: 726–735.
18. Druilhe P, Daubersies P, Patarapotikul J, Gentil C, Chene L, et al. (1998) A primary malarial infection is composed of a very wide range of genetically diverse but related parasites. J Clin Invest 101: 2008–2016.
19. Portugal S, Carret C, Recker M, Armitage AE, Goncalves LA, et al. (2011) Host-mediated regulation of superinfection in malaria. Nat Med 17: 732–737.
20. Snounou G, Beck HP (1998) The use of PCR genotyping in the assessment of recrudescence or reinfection after antimalarial drug treatment. Parasitol Today 14: 462–467.
21. Viriyakosol S, Siripoon N, Petcharapirat C, Petcharapirat P, Jarra W, et al. (1995) Genotyping of Plasmodium falciparum isolates by the polymerase chain reaction and potential uses in epidemiological studies. Bull World Health Organ 73: 85–95.
22. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53–59.
23. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–380.
24. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, et al. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309: 1728–1732.
25. Bereczky S, Dolo A, Maiga B, Hayano M, Granath F, et al. (2006) Spleen enlargement and genetic diversity of Plasmodium falciparum infection in two ethnic groups with different malaria susceptibility in Mali, West Africa. Trans R Soc Trop Med Hyg 100: 248–257.
26. Soulama I, Nebie I, Ouedraogo A, Gansane A, Diarra A, et al. (2009) Plasmodium falciparum genotypes diversity in symptomatic malaria of children living in an urban and a rural setting in Burkina Faso. Malar J 8: 135.
27. Coene J (1993) Malaria in urban and rural Kinshasa: the entomological input. Med Vet Entomol 7: 127–137.
28. Omumbo JA, Guerra CA, Hay SI, Snow RW (2005) The influence of urbanisation on measures of Plasmodium falciparum infection prevalence in East Africa. Acta Trop 93: 11–21.
29. Kiwanuka GN (2009) Genetic diversity in Plasmodium falciparum merozoite surface protein 1 and 2 coding genes and its implications in malaria epidemiology: a review of published studies from 1997–2007. J Vector Borne Dis 46: 1–12.
30. Havryliuk T, Ferreira MU (2009) A closer look at multiple-clone Plasmodium vivax infections: detection methods, prevalence and consequences. Mem Inst Oswaldo Cruz 104: 67–73.
31. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, et al. (2000) Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. Mol Biol Evol 17: 1467–1482.
32. Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, et al. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368: 455–457.
33. Forbes SH, Hogg JT, Buchanan FC, Crawford AM, Allendorf FW (1995) Microsatellite evolution in congeneric mammals: domestic and bighorn sheep. Mol Biol Evol 12: 1106–1113.
34. Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4: 203–221.
35. Schlotterer C (1998) Genome evolution: are microsatellites really simple sequences? Curr Biol 8: R132–134.
36. Juliano JJ, Porter K, Mwapasa V, Sem R, Rogers WO, et al. (2010) Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. Proc Natl Acad Sci U S A 107: 20138–20143.
37. Auburn S, Campino S, Clark TG, Djimde AA, Zongo I, et al. (2011) An Effective Method to Purify Plasmodium falciparum DNA Directly from Clinical Blood Samples for Whole Genome High-Throughput Sequencing. PLoS One 6: e22213.
38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.