

Ahead of the curve: next generation estimators of drug resistance in malaria infections

Nicole Mideo¹, David A. Kennedy^{1,2}, Jane M. Carlton³, Jeffrey A. Bailey⁴, Jonathan J. Juliano⁵, and Andrew F. Read^{1,2,6}

¹ Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA

² Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA

³ Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10003, USA

⁴ Division of Transfusion Medicine and Program in Bioinformatics and Integrative Biology, University of Massachusetts School of Medicine, Worcester, MA 01655, USA

⁵ Division of Infectious Diseases, School of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA

⁶ Department of Entomology, The Pennsylvania State University, University Park, PA 16802, USA

Drug resistance is a major obstacle to controlling infectious diseases. A key challenge is detecting the early signs of drug resistance when little is known about its genetic basis. Focusing on malaria parasites, we propose a way to do this. Newly developing or low level resistance at low frequency in patients can be detected through a phenotypic signature: individual parasite variants clearing more slowly following drug treatment. Harnessing the abundance and resolution of deep sequencing data, our 'selection differential' approach addresses some limitations of extant methods of resistance detection, should allow for the earliest detection of resistance in malaria or other multi-clone infections, and has the power to uncover the true scale of the drug resistance problem.

Drug resistance and clearance curves

The evolution of drug-resistant pathogens is a major challenge in the fight to control infectious diseases. Malaria parasites are a prime example of this: resistance has evolved to nearly every antimalarial drug in use [1] and appears to be emerging against the current front-line artemisinin derivatives [2–7]. Ensuring the continued efficacy of these important drugs requires good surveillance, early and rapid detection of resistance, and containment of its spread [3,8–13]. Ideally, resistance would be detected when it is at a low level (drug 'tolerance') and low frequency in a patient, well before it has become a clinical problem. Once a new drug resistance mutation has generated sufficient treatment failure to arouse suspicion, its spread is probably well advanced. Current methods used to detect resistance struggle to detect rare or low level resistance. Here we propose a new approach that in principle has substantially greater sensitivity.

Corresponding author: Mideo, N. (nlm12@psu.edu)

Keywords: *Plasmodium*; deep sequencing; parasite clearance curves; mixed infections; selection differential; artemisinin.

1471-4922/\$ – see front matter

© 2013 Elsevier Ltd. All rights reserved. <http://dx.doi.org/10.1016/j.pt.2013.05.004>



Three main tools are used to detect drug-resistant malaria parasites. For simplicity, here we use the term 'resistance' to include any significant reduction in drug sensitivity below that found in wild type populations prior to drug use, including weak resistance or drug tolerance, although we note that there is disagreement over whether these phenotypes should be considered resistance [14,15]. These three tools are: (i) molecular markers of resistance, (ii) *in vitro* susceptibility tests, and (iii) therapeutic efficacy tests. Limitations of these approaches have been described in detail elsewhere (e.g., [12,16,17]). Briefly, searching for molecular markers requires knowledge of parasite candidate genes or pathways, yet little can be known in advance about the mechanisms underlying resistance when it is just starting to emerge. This is particularly so in malaria parasites where resistance mechanisms frequently involve loci other than those encoding drug targets [18]. Further, the list of known genetic markers of resistance will never be an exhaustive set of all pathways to resistance; new resistance mechanisms could regularly be evolving, and mechanisms need not be genetic. Testing the susceptibility of parasites to drugs *in vitro* is problematic because *in vivo* phenotypes do not necessarily correlate with *in vitro* performance (e.g., [4]), in addition to being technically challenging, laborious, and expensive. Although less technically demanding, therapeutic efficacy tests are logistically challenging and imprecise. The World Health Organization protocol [19] involves extensive follow up, monitoring patients for up to 1 month or more, during which time any estimate of resistance could be confounded by reinfections or relapses from liver stages. That protocol also sets out rigid criteria for parasitemia and other symptoms over the course of infections that indicate parasite resistance, thus failing to control for any possible sources of interindividual variation (e.g., host factors such as immunity, drug compliance, or coinfection) [8,11,16,20,21].

Recent studies of drug resistance in malaria parasites have addressed some of these problems by using a more

quantitative tool for assaying therapeutic efficacy, the parasite clearance curve [2,4,6,7,22,23]. The remarkable efficiency with which artemisinins act generates a pattern of log-linear decline in parasite densities on a very short timescale [24–26]. With sufficiently frequent sampling of parasite densities, the slope of their decline can be used to calculate a half-life (the time it takes for parasites to reach half of their initial density) [24,27,28]. Parasites that are cleared more slowly from a patient have longer half-lives and are deemed more ‘resistant’. Comparisons of the average half-lives of parasites in different host populations have demonstrated artemisinin resistance emerging in Cambodia [4–6], Thailand [2], and Myanmar [7] and failing to emerge, for now, in Laos [23] and Mali [29].

Parasite clearance curves are important tools for defining host population level distributions of parasite half-lives, and they have the advantage of being able to account for some of the variation that occurs between hosts. For example, some infections show a lag phase before parasite densities start to drop after treatment is initiated, or a tail phase where parasite densities remain above the detectable threshold after a period of log-linear decline [24,27]. The Parasite Clearance Estimator [27] controls for these nonlinearities when estimating a half-life. Although these patterns could be the result of host effects, they could also be measurement errors or, importantly, could indicate parasite subpopulations within a host that respond differently to treatment [24]. Parasite clearance curves are not able to give finer resolution on this sort of within-host variation because they are not designed for making inferences at this level. A parasite clearance curve offers a single estimate of the clearance rate of parasites: a weighted average of the clearance trajectories of all parasites within a patient [24]. Malaria infections in humans seldom consist of one parasite clone; many parasite genotypes share their hosts with other genotypes (e.g., [30–38]). A consequence of this is that resistant parasites will often share their host with sensitive parasites, particularly early in the evolutionary process as resistance is spreading. Attempts to interpret clearance curves for individual patients are affected by this within-host diversity.

Consider a patient that harbors two parasite clones: a fast-clearing, drug-sensitive clone and a slow-clearing, drug-resistant clone. When the clearance of these individual clones cannot be directly tracked, two opposing conclusions can be reached, depending on the relative frequency of the clones. If these two clones are at equal frequencies in a patient when drug treatment is initiated, then the clearance curve observed follows the trajectory of the resistant parasite, because it quickly becomes numerically dominant in a drug-treated infection (Figure 1A). If, however, the resistant clone is initially at low frequency in the patient, then our ability to ‘see’ those resistant parasites in the infection is vastly diminished. The clearance curve now follows the trajectory of the sensitive clone (Figure 1B). From clearance curves alone, it is impossible to tell that the patient in Figure 1B is harboring resistant parasites and potentially transmitting those parasites. Current approaches may therefore underestimate the scale of the drug resistance problem.

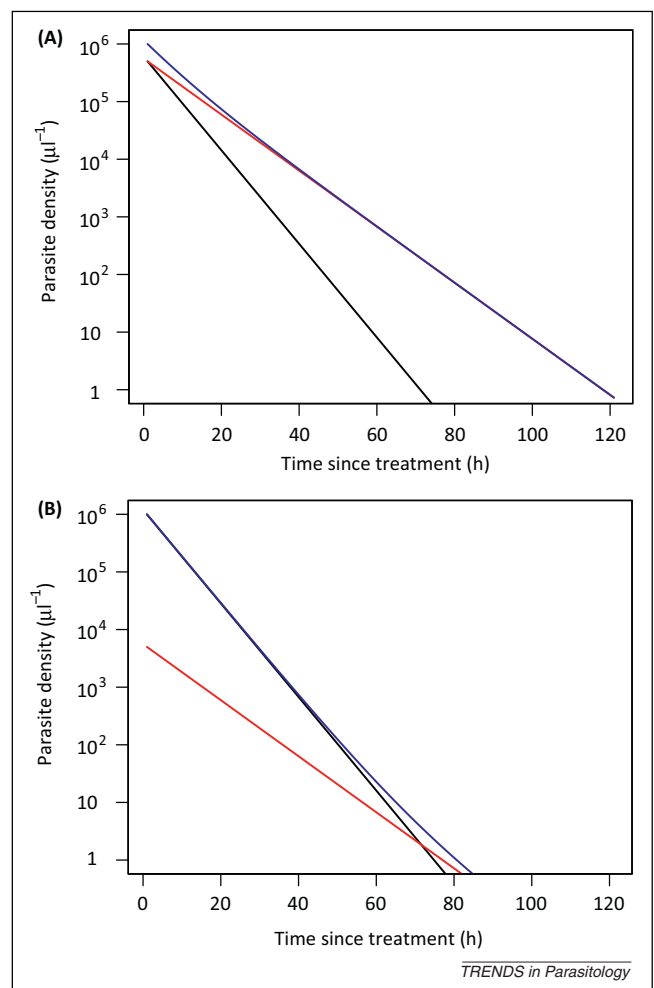


Figure 1. Clearance curves of diverse infections. Predicted clearance curves of sensitive (black) and resistant (red) parasite clones in hypothetical infections after drug treatment is initiated (ignoring any possible lag or tail phases [24,27]). The blue line shows the clearance curve that would actually be measured for an individual patient using standard approaches (i.e., it shows the change in the total parasite density over time). Initial frequencies of the resistant clone are (A) 50% and (B) 0.5%. Half-lives for sensitive and resistant clones are taken from [2] (3.7 h and 6.2 h, respectively). Despite harboring resistant parasites in both cases, the infection in (A) appears resistant whereas in (B) it appears sensitive.

Capitalizing on within-host diversity with next generation approaches

We propose that, rather than being an obstacle to identifying drug resistance in individual patients, the within-host diversity of infections can instead be exploited to detect it. When drug treatment is applied to mixed infections, clones with any degree of resistance will rise in frequency in the parasite population within that host. Even weakly drug tolerant clones will become more frequent because fully susceptible parasites will be cleared at faster rates. Thus, an increase in the frequency of any clone-specific genetic marker is indicative of up-selection by drugs of a tolerant or resistant clone in a sea of susceptible parasites. The critical requirement for identifying such a change in frequency is a detection technique that can quantify the relative abundance of different clones and be sufficiently sensitive to detect rare variants in a mixed population. Next generation sequencing (NGS) technologies, when they are used to deeply sequence a single highly

polymorphic region hundreds or thousands of times in a single infection, can do this. The incredible scale and high resolution of the data generated through these technologies have exposed the true (and vastly underestimated) within-host diversity of malaria infections [30,39–42] and prevalence of rare drug-resistant variants in HIV infections (e.g., [43–46]). Using these same deep sequencing tools to detect drug resistance in malaria infections, as well as quantify the extent and magnitude of that resistance, is the next logical step.

With longitudinal samples, including a pretreatment sample and perhaps as few as one subsequent sample at least one replicative cycle later, deep sequencing can generate a picture of how individual parasite variants respond to drug treatment. This approach, which we refer to as the ‘selection differential’ approach because that is what is being estimated [47], provides access to those clone-specific parasite clearance curves that are unobtainable with current tools. Revisiting the hypothetical patients described in Figure 1, if deep sequencing was used to quantify the relative abundance of different clones in place of tracking parasite densities, a clear signature of drug resistance would emerge regardless of whether the resistant variant was initially at a high or low frequency (Figure 2). Importantly, identifying drug resistance with deep sequencing does not require knowing (or sequencing) the molecular markers of resistance. Instead, the phenotype of resistance can be seen – slower clearance, increasing relative abundance – by sequencing any coding region that is highly polymorphic. Because parasite recombination does not occur within human hosts, such a gene region is in essence a clone-specific marker. The ability to detect all of the variants in an infection, and hence the power of the method, will depend on the polymorphism of the gene region targeted; a high degree of polymorphisms allows for a relatively rare resistant clone to be exclusively represented by one or a few alleles within an individual infection.

The selection differential approach offers several advantages as a tool to detect drug resistance. First, what we propose is a short-term test (perhaps just two or three replicative cycles) meaning that even if a parasite variant is below the limit of detection in a pretreatment sample (e.g., if sequencing is not sensitive enough to detect a variant at 0.05%, as depicted in the left panel of Figure 2B), it would be highly unlikely that its detection in a subsequent sample would be due to a reinfection or relapse. Second, what we propose should in principal be feasible with as few as two blood samples from patients, although certainly more power will be garnered by additional time points, and these may be necessary if lag phases are a parasite genetic effect. Finally, what we propose is inpatient, meaning that all patient-specific variables (such as immune status, coinfection, drug uptake, compliance, dosing, drug quality, etc.) are held constant while the clearance rates of individual parasite clones within that patient are investigated. Although recent studies have reported that a substantial amount of the variation in the clearance times is attributable to parasite genetics [2,48], some of it is not. Our approach therefore accounts for more of this variation.

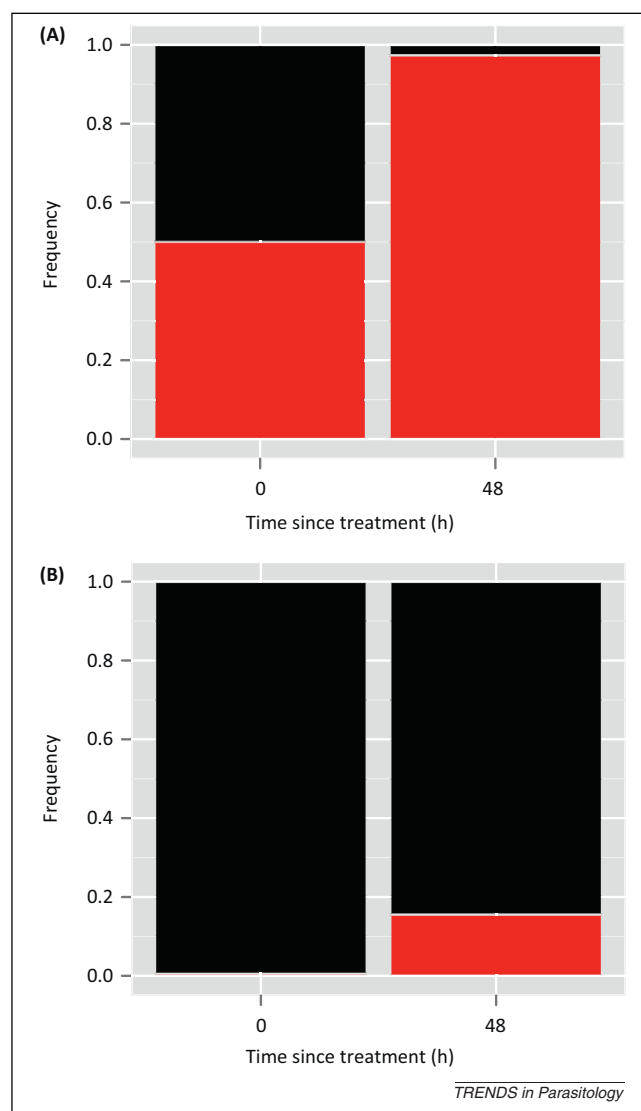


Figure 2. The phenotypic signature of drug resistance. Relative abundance of drug-sensitive (black) and -resistant (red) parasites before and after drug treatment of hypothetical infections in Figure 1. Half-lives are as given in Figure 1, and initial frequencies of the resistant clone are again (A) 50% and (B) 0.5%. In both cases, the relative abundance of the red parasite variant dramatically increases after drug treatment, indicating that it is being cleared more slowly than the black variant and providing a striking signal of resistance. Deep sequencing can provide this type of data on the relative abundance of parasite variants.

Next generation challenges

Using deep sequencing for early detection of drug resistance is not without its own challenges (Box 1). Chief among these is discriminating real changes in the relative abundance of different parasite variants over time from noise. Accurate characterization of variants within an infection and quantification of the relative abundance of those variants are key for inferring drug resistance from the changes that occur over the course of treatment. But deep sequencing is error-prone and observation error will always muddle attempts to characterize a large parasite population based on a small sample. In a forthcoming article, we develop statistical methods to analyze the type and quality of data involved and use these to provide estimates of the power of the approach to detect low frequency and low level resistance. In Box 2, we illustrate

Box 1. Technological challenges

Next generation sequencing of malaria parasites has been used to quantify and track the clonal diversity within individual infections [30,41,42], to examine diversity and population genetic patterns at the whole parasite population level (among hosts in a given region; [39–41]), and to correlate population genetic patterns with resistance phenotypes [58]. These studies represent a combination of whole genome and amplicon sequencing approaches, and each has its own challenges. For studies using whole genome sequencing [41,42,58], relatively large amounts of DNA are required. This necessitates collection of large volumes of whole blood, cold storage of samples between the field and the laboratory, filtering of samples to remove host white blood cells, and, sometimes, short-term *in vitro* culturing to increase parasite numbers (and DNA) [59,60]. The key challenges for whole genome sequencing studies, then, occur at the sampling and sample-processing end of the pipeline. In addition, when used for studying multi-clone infections, whole genome sequencing provides limited depth and short reads, both of which limit the ability to study variants occurring at low frequencies.

These challenges are less of a problem for amplicon sequencing, the approach that we are advocating. Less DNA is required for amplicon sequencing [30,39,40], and sufficient volumes can be

obtained directly from blood spots on filter paper, which require desiccation rather than refrigeration for storage. Because amplification of a target locus by PCR is performed, large blood samples and filtering of host white blood cells are unnecessary [61]. However, this PCR step also has the potential to introduce base pair substitutions due to polymerase errors and, more challenging, chimeric sequences due to incomplete primer extension on one haplotype followed by completion of extension on another (PCR-mediated recombination; e.g., [62]). Such chimeras are likely to be a significant contributor to the noise described in Box 2. These issues shift the key challenges with amplicon sequencing towards the bioinformatics and sequence-processing end of the pipeline. Computational tools have been developed to deal with chimeras and other PCR-introduced artifacts (e.g., [63–65]), but these have so far been restricted to making assignments in clonal populations or at higher taxonomic levels (i.e., genus level). For sexually recombining organisms such as malaria, chimeras and true recombinants lack *a priori* distinguishing features. For now, studies try to circumvent these problems by using stringent criteria for calling variants, for example, by requiring a variant to occur in multiple technical replicates, as in [39].

the variation that can arise across technical replicates from these sources of error and briefly describe how, by borrowing tools from community ecology, this noise can in principle be harnessed to develop a null expectation and define thresholds for inferring resistance.

Implicit in our selection differential approach is an assumption that infections are likely to be composed of a small number of drug-resistant parasites in a sea of sensitive ones. This may not be the case in regions of Southeast Asia where resistance to artemisinins is emerging and where the data suggest that the majority of infections harbor a single clone (e.g., [2]). In those cases, our approach is irrelevant because parasite clearance curves do an adequate job (recall Figure 1A). However, if and when resistance spreads to high transmission settings in Africa

(where vigilance and early detection are especially important [49]), resistant parasites are likely to occur in mixed infections. Indeed, several studies looking for known resistance mutations (i.e., resistance to non-artemisinin drugs) have found parasites harboring those mutations commonly occurring in mixed infections with wild type parasites (e.g., [34,50–52]). A further wrinkle comes from recent work that shows that mixed genotype *Plasmodium falciparum* infections tend to be composed of closely related parasites [53], and this could mean that resistant parasites often share an infection with other resistant parasites. If so, similarly resistant parasite variants in an infection would be cleared at the same rate, so the relative abundance of those variants would not change. Our proposed approach would fail to identify infections with resistance at high frequency,

Box 2. Quantifying resistance through noise

In a recent study [39], DNA was extracted from single blood spots from 100 individual malaria patients prior to receiving treatment, and then divided into two for PCR and sequencing. An approximately 250-bp region of the highly polymorphic *Plasmodium falciparum* *msp* gene was amplified, and the amplicon product was sequenced on the Roche 454 platform. For each replicate sample from individual patients, reads were clustered into parasite variants (haplotypes). Based on control sequencing, the authors required that a variant be at a frequency of at least 1% and occur in both replicates before being considered a true variant for their analysis. Here, we include all initially called variants in each replicate regardless of their frequency or reproducibility to provide a maximal estimate of noise. Figure 1A shows the relative abundance of variants, as estimated in the two technical replicates, for two patients (labeled 1111 and 1128) before treatment. Most of the 100 patients generate graphs that look like patient 1111, that is, the frequency of variants is relatively reproducible across replicates. For a minority of patients (fewer than 10), the pattern looks more like patient 1128, with large shifts in frequency across replicates. No systematic underlying cause of these shifts could be determined and all samples passed sequencing quality control cut-off values [39]. The dark blue variant in patient 1128 undergoes precisely the sort of change expected for a drug-resistant parasite during treatment. If such patterns can arise due to noise across a single time point, how can one be sure that the same pattern across time points indicates resistance?

With such technical replicates from pretreatment sampling, one can generate a null expectation against which to compare changes that

occur over the course of treatment. Useful metrics can be adapted from community ecology where measures of dissimilarity are used to compare species assemblages (e.g., the Bray–Curtis and Manhattan measures; see [66–69] for overviews). For example, for each variant within a patient, the distance of the point given by the two pretreatment frequency estimates of that variant to the expected 1:1 line can be calculated. We define d_0 as the sum of the distances of all variants within a patient:

$$d_0 = \frac{1}{\sqrt{2}} \sum_{i=1}^n \left| p_{0,i}^1 - p_{0,i}^2 \right|$$

for a patient harboring n variants, where the pretreatment frequency of the i th variant is estimated as $p_{0,i}$ (the superscript denotes the technical replicate). In Figure 1B, we illustrate this approach for the two representative patients. Calculating d_0 for each of the 100 patients in [39] gives rise to the distribution of d_0 values in Figure 1C; this represents our null expectation against which we could compare the distance of the points in a pre- versus post-treatment frequency plot to the 1:1 line. If drug treatment changes the relative abundance of parasite variants due to the presence of resistant variants, then the points in this plot will lie farther from the 1:1 line. Specifically, if

$$d_T = \frac{1}{\sqrt{2}} \sum_{i=1}^n \left| p_{0,i} - p_{T,i} \right|$$

where $p_{T,i}$ is the post-treatment frequency estimate of the i th variant, then we expect d_T to lie outside of the distribution of d_0 .

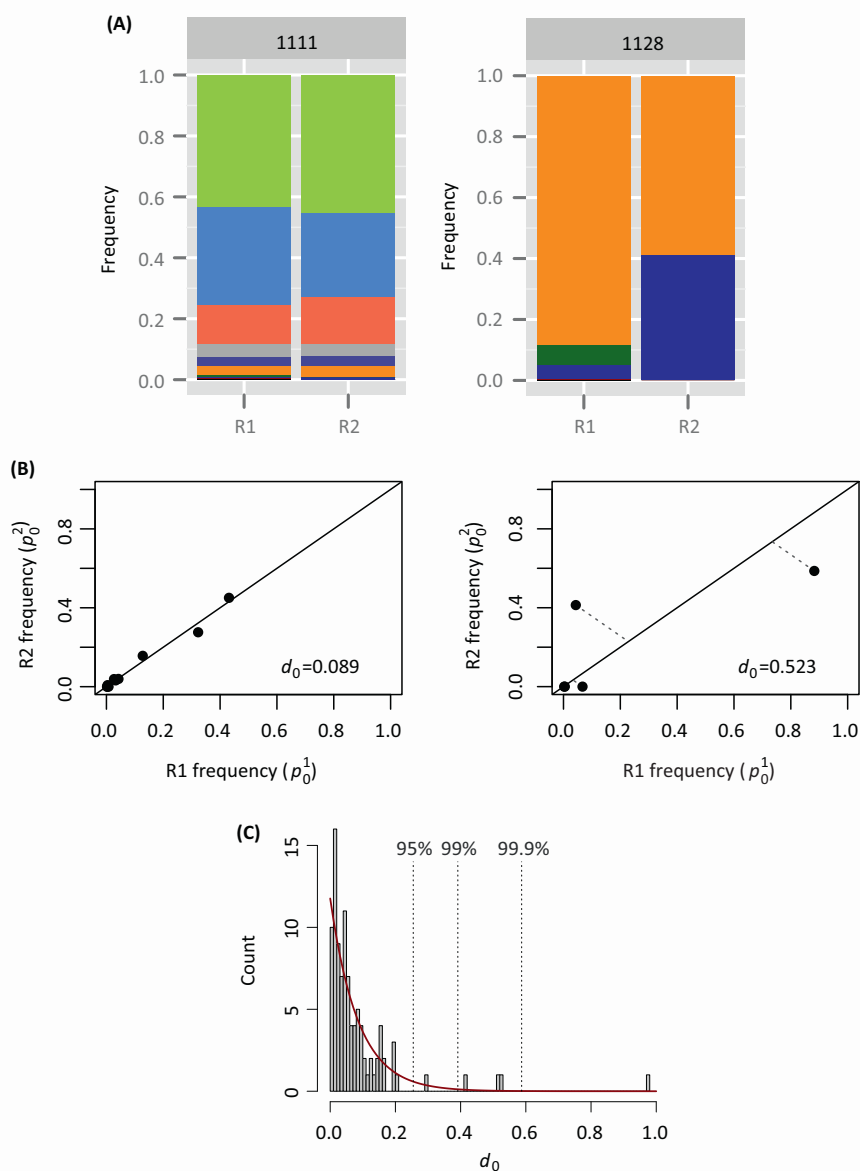


Figure 1. Noise and the summed distance metric. **(A)** Relative abundance of parasite variants, as determined by 454 sequencing of an approximately 250-bp region of the *P. falciparum* *csp* gene, from two representative patients [39]. For each patient, the relative abundances have been estimated twice (technical replicates, R1 and R2). Within patients, each color represents a unique parasite variant. **(B)** Plotting these frequencies in a different way reveals a metric for quantifying noise. Here, each point represents the two estimates of frequency for an individual parasite variant. The broken lines indicate the distance to the 1:1 line, and the d_0 value gives the sum of these distances for all variants within a patient. If technical replicates were perfectly repeatable, all points would lie along the 1:1 line and d_0 would be 0. **(C)** The distribution of d_0 values calculated from all 100 patients in [39], as well as a fitted exponential curve (red; $\lambda = 11.8$). The broken lines indicate the values of d_0 below which 95%, 99%, or 99.9% of the fitted distribution lies (equivalently, d_7 values above these cut-off values are expected to occur less than 5%, 1%, or 0.1% of the time by chance alone).

TRENDS in Parasitology

but this is not what our approach aims to do – again, clearance curves do an adequate job in this case. And, because all *de novo* mutations for resistance necessarily begin in a mixed infection, if *de novo* resistance is a significant source of treatment failure then our approach provides a method to detect these resistant variants as soon as they appear.

Our approach assumes that parasites in drug-treated infections follow the log-linear clearance curves discussed earlier, but a number of biological processes could qualitatively alter these patterns of decline. For example, what if there are multiple parasite cohorts with 48-h cycles that are

offset by a day? What happens if parasite variants within an infection vary in their cell cycle duration? In the first case, daily sampling to capture each cohort over a full replicative cycle should solve the multiple cohort problem. In the second case, it is not yet clear how variation in cell cycle duration would affect estimates of the relative abundance of variants over the course of drug treatment. Assuming it affects drug-sensitive and drug-resistant parasites equally, then this complexity would be captured in any null distribution (Box 2). But, if drug resistance is a function of this complexity, for example, if resistance is mediated by changes to cycle length (as has been suggested, e.g., [54]), then the situation

may be more complicated and mechanistic modeling of within-host processes could shed light on how this would affect measures of change. Our selection differential approach could also be constrained if drug treatment results in parasite densities below the threshold required for sequencing before a post-treatment sample is taken. With a better understanding of the noise and limits of sequencing data, it will be possible to quantify the parameter space (level and frequency of resistance) over which our approach is robust.

Finally, selection pressures beyond those imposed by drug treatment may be acting in infections. This means that the relative abundance of different parasite clones will change for reasons other than differential sensitivity to drugs (e.g., differential susceptibility to antigen-specific immune responses) and may obscure the selective effects of drugs. Understanding the scale of the changes in response to these other pressures could be achieved with longitudinal samples of untreated asymptomatic infections or with pre- and post-treatment samples from patients in areas where resistance is currently thought to be absent. This additional source of noise could also be built into the null expectation (Box 2). It is our speculation that selection in response to drugs will overwhelm other evolutionary forces such as selection via biased parasite killing by the immune system. For example, in experimental rodent malaria infections, which are more rapidly contained and cleared than human infections, all immune responses (specific and nonspecific) have been estimated to kill 85% of parasites per day at maximum efficiency and less than that throughout much of the infection [55]. It is unclear how much of that immune killing is strain-specific, but even the maximum of 85% equates to a parasite reduction ratio of approximately ten, two to four orders of magnitude lower than what has been estimated for the efficacy of artemisinins [25]. This suggests that drug selection has the potential to overwhelm strain-specific immune selection, but further data (or mechanistic modeling) is needed to

understand the magnitude of selection differentials that can be generated through immunity alone (Box 3).

Concluding remarks and future perspectives

The introduction of NGS technologies brought with it the promise of allowing researchers to 'follow the evolution of... resistance in real time' [56]. The data generated from these technologies should allow us to see the up-selection of drug-resistant parasites within individual infections. Indeed, deep sequencing has been used, for example, to follow the dynamics of known resistance mutations in HIV infections [57]. Yet, as we have argued, these technologies have a much broader utility, allowing for the dynamics of resistance to be tracked even when genetic resistance mechanisms have yet to be discovered.

The selection differential approach will in principle permit the most rapid diagnosis of resistance in malaria infections, with the fewest possible blood samples, and the best possible accuracy (bypassing issues of reinfection and relapse). Further, by allowing previously undetectable resistant parasites to be identified, these tools—in addition to contributing to surveillance and monitoring programs—could feed samples and data into studies seeking to identify genetic markers and pathways of resistance. Alternatively, with candidate mutations identified, the selection differential approach could be adapted to determine whether candidate resistance alleles are indeed being up-selected during drug treatment. Finally, should the costs and the turnaround time of the sequencing technology continue to decrease, it could in theory be possible to use this approach to identify individual patients who will not respond to treatment, giving clinicians an opportunity to change treatment regimens or partner drugs before clinical failure occurs. This would have the benefit of not only improving patient health but also removing the drug-specific selection pressure on those parasites.

Acknowledgments

We thank members of the Research and Policy in Infectious Disease Dynamics Program of the Science and Technology Directorate, Department of Homeland Security and the Fogarty International Center, National Institutes of Health (NIH) for discussion, and four anonymous reviewers for helping to substantially improve this manuscript. This study was funded by National Institute of Allergy and Infectious Diseases, NIH grants R01AI089819 to J.J.J. and U19AI089676 to J.M.C., and a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellowship to N.M.

References

- 1 World Health Organization (2010) *Global Report on Antimalarial Drug Efficacy and Drug Resistance: 2000–2010*. World Health Organization
- 2 Phyto, A.P. *et al.* (2012) Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. *Lancet* 379, 1960–1966
- 3 Dondorp, A.M. *et al.* (2010) Artemisinin resistance: current status and scenarios for containment. *Nat. Rev. Microbiol.* 8, 272–280
- 4 Dondorp, A.M. *et al.* (2009) Artemisinin resistance in *Plasmodium falciparum* malaria. *N. Engl. J. Med.* 361, 455–467
- 5 Noedl, H. *et al.* (2008) Evidence of artemisinin-resistant malaria in western Cambodia. *N. Engl. J. Med.* 359, 2619–2620
- 6 Amaratunga, C. *et al.* (2012) Artemisinin-resistant *Plasmodium falciparum* in Pursat province, western Cambodia: a parasite clearance rate study. *Lancet Infect. Dis.* 12, 851–858
- 7 Kyaw, M.P. *et al.* (2013) Reduced susceptibility of *Plasmodium falciparum* to artesunate in southern Myanmar. *PLoS ONE* 8, e57689

Box 3. Outstanding questions

- How well do different sequencing platforms estimate the relative abundances of known mixtures? Given this noise, how rare could a resistant clone be in an infection and still be detected using the selection differential approach?
- What magnitude of change in relative abundances can be generated through immune selection alone, that is, how effective at killing parasites are clone-specific immune responses? On what timescale are these changes observable?
- How does sampling and sequencing parasites at different stages in their replicative cycle affect estimates of relative abundance?
- How much variation in cell cycle duration is there among the parasites within an infection?
- If there is variation in cell cycle duration, and the stage of sampled parasites affects estimates, what sampling regime could provide a meaningful picture of the relative abundance of clones?
- Does residual DNA from dying or dead parasites contribute to estimates of relative abundance from deep sequencing?
- Are the lag phases in parasite clearance curve a host effect (e.g., drug metabolism) or a parasite effect (e.g., parasite subpopulations that are inaccessible to drugs) [24]? If lags are parasite effects that are unrelated to drug sensitivity, how many more samples would be needed to rule out changes in relative abundance due to different lengths of lag phases?

- 8 Sibley, C.H. and Price, R.N. (2012) Monitoring antimalarial drug resistance: applying lessons learned from the past in a fast-moving present. *Int. J. Parasitol. Drugs Drug Resist.* 2, 126–133
- 9 Guerin, P.J. *et al.* (2009) Global resistance surveillance: ensuring antimalarial efficacy in the future. *Curr. Opin. Infect. Dis.* 22, 593–600
- 10 Plowe, C.V. (2009) The evolution of drug-resistant malaria. *Trans. R. Soc. Trop. Med. Hyg.* 103, S11–S14
- 11 Chretien, J-P. *et al.* (2007) Improving surveillance for antimalarial drug resistance. *J. Am. Med. Assoc.* 297, 2278–2281
- 12 Vestergaard, L.S. and Ringwald, P. (2007) Responding to the challenge of antimalarial drug resistance by routine monitoring to update national malaria treatment policies. *Am. J. Trop. Med. Hyg.* 77, 153–159
- 13 Noedl, H. (2005) Artemisinin resistance: how can we find it? *Trends Parasitol.* 21, 404–405
- 14 Meshnick, S. (2012) Perspective: artemisinin-resistant malaria and the wolf. *Am. J. Trop. Med. Hyg.* 87, 783–784
- 15 White, N.J. (2012) Counter perspective: artemisinin resistance: facts, fears, and fables. *Am. J. Trop. Med. Hyg.* 87, 785
- 16 Fairhurst, R.M. *et al.* (2012) Artemisinin-resistant malaria: research challenges, opportunities, and public health implications. *Am. J. Trop. Med. Hyg.* 87, 231–241
- 17 Talisuna, A.O. *et al.* (2004) History, dynamics, and public health importance of malaria parasite resistance. *Clin. Microbiol. Rev.* 17, 235–254
- 18 Hyde, J.E. (2007) Drug-resistant malaria – an insight. *FEBS J.* 274, 4688–4698
- 19 World Health Organization (2009) *Methods for Surveillance of Antimalarial Drug Efficacy*. World Health Organization
- 20 Francis, D. *et al.* (2006) Geographic differences in antimalarial drug efficacy in Uganda are explained by differences in endemicity and not by known molecular markers of drug resistance. *J. Infect. Dis.* 193, 978–986
- 21 Wongsrichanalai, C. *et al.* (2002) Epidemiology of drug-resistant malaria. *Lancet Infect. Dis.* 2, 209–218
- 22 Cheeseman, I.H. *et al.* (2012) A major genome region underlying artemisinin resistance in malaria. *Science* 336, 79–82
- 23 Mayxay, M. *et al.* (2012) No evidence for spread of *Plasmodium falciparum* artemisinin resistance to Savannakhet Province, Southern Laos. *Am. J. Trop. Med. Hyg.* 86, 403–408
- 24 White, N.J. (2011) The parasite clearance curve. *Malar. J.* 10, 278
- 25 White, N.J. (1997) Assessment of the pharmacodynamic properties of antimalarial drugs *in vivo*. *Antimicrob. Agents Chemother.* 41, 1413–1422
- 26 White, N.J. (1994) Clinical pharmacokinetics and pharmacodynamics of artemisinin and derivatives. *Trans. R. Soc. Trop. Med. Hyg.* 88, S41–S43
- 27 Flegg, J.A. *et al.* (2011) Standardizing the measurement of parasite clearance in falciparum malaria: the parasite clearance estimator. *Malar. J.* 10, 339
- 28 Beshir, K.B. *et al.* (2010) Measuring the efficacy of anti-malarial drugs *in vivo*: quantitative PCR measurement of parasite clearance. *Malar. J.* 9, 312
- 29 Maiga, A.W. *et al.* (2012) No evidence of delayed parasite clearance after oral artesunate treatment of uncomplicated falciparum malaria in Mali. *Am. J. Trop. Med. Hyg.* 87, 23–28
- 30 Juliano, J.J. *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
- 31 Baruah, S. *et al.* (2009) Temporal and spatial variation in MSP1 clonal composition of *Plasmodium falciparum* in districts of Assam, Northeast India. *Infect. Genet. Evol.* 9, 853–859
- 32 Soulama, I. *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
- 33 Vafa, M. *et al.* (2008) Multiplicity of *Plasmodium falciparum* infection in asymptomatic children in Senegal: relation to transmission, age and erythrocyte variants. *Malar. J.* 7, 17
- 34 Jafari, S. *et al.* (2004) *Plasmodium falciparum* clonal population dynamics during malaria treatment. *J. Infect. Dis.* 189, 195–203
- 35 Magesa, S.M. *et al.* (2002) Diversity of *Plasmodium falciparum* clones infecting children living in a holoendemic area in north-eastern Tanzania. *Acta Trop.* 84, 83–92
- 36 Babiker, H.A. *et al.* (1999) The epidemiology of multiple *Plasmodium falciparum* infections. 3. Genetic structure and dynamics of *Plasmodium falciparum* infections in the Kilombero region of Tanzania. *Trans. R. Soc. Trop. Med. Hyg.* 93, S11–S14
- 37 Konaté, A.T. *et al.* (1999) The epidemiology of multiple *Plasmodium falciparum* infections. *Trans. R. Soc. Trop. Med. Hyg.* 93, S21–S28
- 38 Smith, T. *et al.* (1999) Epidemiology of multiple *Plasmodium falciparum* infections. 4. 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, S15–S20
- 39 Bailey, J.A. *et al.* (2012) Use of massively parallel pyrosequencing to evaluate the diversity of and selection on *Plasmodium falciparum* csp T-cell epitopes in Lilongwe, Malawi. *J. Infect. Dis.* 206, 580–587
- 40 Gandhi, K. *et al.* (2012) Next generation sequencing to detect variation in the *Plasmodium falciparum* circumsporozoite protein. *Am. J. Trop. Med. Hyg.* 86, 775–781
- 41 Manske, M. *et al.* (2012) Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* 487, 375–379
- 42 Auburn, S. *et al.* (2012) Characterization of within-host *Plasmodium falciparum* diversity using next-generation sequence data. *PLoS ONE* 7, e32891
- 43 Rozera, G. *et al.* (2009) Massively parallel pyrosequencing highlights minority variants in the HIV-1 env quasispecies deriving from lymphomonocyte sub-populations. *Retrovirology* 6, 15
- 44 Mitsuya, Y. *et al.* (2008) Minority human immunodeficiency virus type 1 variants in antiretroviral-naive persons with reverse transcriptase codon 215 revertant mutations. *J. Virol.* 82, 10747–10755
- 45 Hoffmann, C. *et al.* (2007) DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. *Nucleic Acids Res.* 35, e91
- 46 Wang, C. *et al.* (2007) Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Res.* 17, 1195–1201
- 47 Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*. (4th edn), Longman
- 48 Takala-Harrison, S. *et al.* (2013) Genetic loci associated with delayed clearance of *Plasmodium falciparum* following artemisinin treatment in Southeast Asia. *Proc. Natl. Acad. Sci. U.S.A.* 110, 240–245
- 49 Talisuna, A.O. *et al.* (2012) Mitigating the threat of artemisinin resistance in Africa: improvement of drug-resistance surveillance and response systems. *Lancet Infect. Dis.* 12, 888–896
- 50 Owusu-Agyei, S. *et al.* (2009) Epidemiology of malaria in the forest-savanna transitional zone of Ghana. *Malar. J.* 8, 220
- 51 Zhong, D. *et al.* (2008) Molecular epidemiology of drug-resistant malaria in western Kenya highlands. *BMC Infect. Dis.* 8, 105
- 52 Juliano, J.J. *et al.* (2007) Minority-variant pfcr1 K76T mutations and chloroquine resistance, Malawi. *Emerg. Infect. Dis.* 13, 872–877
- 53 Nkhoma, S.C. *et al.* (2012) Close kinship within multiple-genotype malaria parasite infections. *Proc. R. Soc. Lond. B* 279, 2589–2598
- 54 Dodd, A. *et al.* (2011) Artemisinin-induced parasite dormancy: a plausible mechanism for treatment failure. *Malar. J.* 10, 56
- 55 Metcalf, C.J.E. *et al.* (2011) Partitioning regulatory mechanisms of within-host malaria dynamics using the effective propagation number. *Science* 333, 984–988
- 56 Kahvejian, A. *et al.* (2008) What would you do if you could sequence everything? *Nat. Biotechnol.* 26, 1125–1133
- 57 Hedskog, C. *et al.* (2010) Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. *PLoS ONE* 5, e11345
- 58 Miotto, O. *et al.* (2013) Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat. Genet.* <http://dx.doi.org/10.1038/ng.2624>
- 59 Auburn, S. *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
- 60 Kozarewa, I. *et al.* (2009) Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat. Methods* 6, 291–295
- 61 Allouche, A. *et al.* (2000) High-throughput sequence typing of T-cell epitope polymorphisms in *Plasmodium falciparum* circumsporozoite protein. *Mol. Biochem. Parasitol.* 106, 273–282

- 62 Ashelford, K.E. *et al.* (2006) New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl. Environ. Microbiol.* 72, 5734–5741
- 63 Quince, C. *et al.* (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinform.* 12, 38
- 64 Haas, B.J. *et al.* (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-yrosequenced PCR amplicons. *Genome Res.* 21, 494–504
- 65 Edgar, R.C. *et al.* (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200
- 66 Cha, S.H. (2007) Comprehensive survey on distance/similarity measures between probability density functions. *Int. J. Math. Mod. Meth. Appl. Sci.* 1, 300–307
- 67 Clarke, K.R. *et al.* (2006) On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. *J. Exp. Mar. Biol. Ecol.* 330, 55–80
- 68 Magurran, A.E. (2004) *Measuring Biological Diversity*. Blackwell Science Ltd.
- 69 Faith, D.P. *et al.* (1987) Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69, 57–68