NO PICO

PAPER ID: PMC4943078

TITLE:
The population genetics of drug resistance evolution in natural populations of viral, bacterial and eukaryotic pathogens

ABSTRACT:
Drug resistance is a costly consequence of pathogen evolution and a major concern in public health. In this review, we show how population genetics can be used to study the evolution of drug resistance and also how drug resistance evolution is informative as an evolutionary model system. We highlight five examples from diverse organisms with particular focus on: (i) identifying drug resistance loci in the malaria parasite Plasmodium falciparum using the genomic signatures of selective sweeps, (ii) determining the role of epistasis in drug resistance evolution in influenza, (iii) quantifying the role of standing genetic variation in the evolution of drug resistance in HIV, (iv) using drug resistance mutations to study clonal interference dynamics in tuberculosis and (v) analysing the population structure of the core and accessory genome of Staphylococcus aureus to understand the spread of methicillin resistance. Throughout this review, we discuss the uses of sequence data and population genetic theory in studying the evolution of drug resistance.

Introduction:
Pathogen evolution is a major public health concern with enormous societal consequences around the world. Pathogens evolve quickly, allowing them to jump from multiple host species to humans (e.g. SARS coronavirus or Ebola virus; Woolhouse et al. 2005), to become more virulent and evade immune pressure within humans (e.g. influenza virus; Grenfell et al. 2004) and to become resistant to drugs used to combat them (WHO 2014d). In particular, drug resistance results in numerous deaths, increased hospitalizations and prolonged treatments (WHO 2014d). In addition to the costs borne by human health, the Centers for Disease Control and Prevention estimate the economic costs of drug resistance to be on the order of tens of billions of dollars a year in the United States alone (CDC 2011). While pharmaceutical intervention has played a critical role in our efforts to control epidemic pathogens, there is an urgent need to understand the evolution and spread of drug resistance.
Pathogens share common biological features that allow them to adapt rapidly under extreme selective pressures (imposed, for example, by drugs) on observable timescales. Pathogen populations with large sizes, high mutation rates and short generation times are likely to generate drug resistance mutations in a short amount of time. Pathogens also encompass a wide variety of organisms that include eukaryotes, prokaryotes and viruses. This means that our efforts to understand how pathogens evolve drug resistance need to account for factors such as: variable genome sizes [ranging from a few kilobases (Kb) in RNA viruses to many megabases (Mb) in Plasmodium falciparum], variable mutation rates (from 10−9 per base pair per generation in eukaryotes to 10−5 in some RNA viruses), alternative methods of genetic exchange (including viral re-assortment and horizontal gene transfer in bacteria) and clonal vs. recombinant forms of reproduction (see Box 1 for more on diverse modes of genetic exchange and recombination). In addition, the scale at which drug resistance evolves varies widely. For example, in HIV drug resistance mostly evolves within patients and is rarely transmitted (Wheeler et al. 2010), whereas in other pathogens (e.g. malaria), drug resistance is often transmitted between hosts, making certain drugs ineffective for a large fraction of new patients (Hyde 2005).
To prevent or at least manage drug resistance in pathogens, it is imperative to understand the evolutionary aspects of the problem. However, evolutionary biologists have traditionally left most of the study of drug resistance to the medicine and epidemiology communities (Read & Huijben 2009). This has changed in recent years with interesting population genetic work being done on natural populations of pathogens, some of which is featured in this review. This change is, at least in part, driven by the availability of genetic data for a variety of pathogens. A new cross-disciplinary field is emerging that uses large amounts of genetic data and new theoretical advances from the world of population genetics. Some of this work is directly aimed at finding ways to prevent drug resistance evolution or spread, whereas other work is more basic in nature and mostly uses the pathogen as a model system to learn about evolution in general. In the best cases, we learn lessons on how to prevent resistance and on general evolutionary principles simultaneously.

The aim of this review was to introduce researchers from different fields to new and exciting work on drug resistance evolution happening at the intersection of population genetics and medical fields such as virology and infectious disease. Perhaps more importantly this review is intended as an advertisement of the strengths of using natural pathogen populations to study evolutionary principles as well as the strengths of using evolutionary theory to better understand the dynamics of infectious diseases. To this end, we present five case studies at the intersection of evolutionary theory and pathogen drug resistance. While the five examples we have chosen are not meant to be exhaustive in any way, each example showcases an evolutionary question in the context of a different pathogen, each with a unique biology and differences in genome size, methods of genetic exchange, reproduction and recombination (see Box 1).
We will describe the following five examples:

Selective sweep mapping in the malaria parasite P. falciparum resistant to the antimalarial drug artemisinin. This is a great example of how population genetic methods originally developed for other sexually recombining eukaryotic organisms, such as humans, could be applied to P. falciparum, which experiences a very high rate of sexual recombination. The mapping of the drug resistance mutations in P. falciparum has allowed researchers to track the prevalence of artemisinin resistance worldwide.
The role of epistasis in oseltamivir (Tamiflu) resistance in influenza. The evolution and spread of oseltamivir resistance in influenza in 2007–2008 was unexpected. Even though the causal mutation for oseltamivir resistance was known, it was thought to be too costly to viral fitness to spread widely. Phylogenetic methods allowed researchers to discover mutations that interacted with the resistance mutation to ameliorate its detrimental effects, allowing the resistance mutation to spread. Influenza has become a model that helps researchers understand the importance of genetic interactions between sites during the course of evolution and the potential for predicting future evolution.
The importance of standing genetic variation (SGV) (also referred to as minority variants) to treatment failure for human immunodeficiency virus (HIV). Drug resistance mutations are well known in HIV and the evolution of drug resistance happens independently in different patients. This makes it possible to determine the importance of SGV for drug resistance in HIV. Several studies found that both SGV and de novo mutations contribute to drug resistance in HIV.
Clonal interference dynamics in Mycobacterium tuberculosis. Drug resistance mutations are also well known in M. tuberculosis. Through tracking these mutations, recent studies have revealed that clonal interference plays a large role in the evolutionary dynamics within patients. An examination of these dynamics offers insights into how multiply-resistant strains emerge and also provides a rare look into a natural population evolving without recombination.
Population structure analysis of the core and accessory genome of methicillin-resistant Staphylococcus aureus (MRSA). Staphylococcus aureus reproduces clonally, yet drug resistance is often acquired via the spread of mobile genetic elements (or horizontal gene transfer), such as the gene cassette SCCmec that confers methicillin resistance. While the genealogy of the core genome of MRSA reveals that relatively few lineages have spread across the globe, phylogenetic analysis of SCCmec reveals that methicillin resistance arises on a more local scale.

These five examples will be used to illustrate that the growing availability of sequence data facilitates the application of population genetic methods to important questions in the evolution of drug resistance.

Population genomic methods identify a region on chromosome 13 associated with artemisinin resistance ::: Linkage disequilibrium statistics for the detection of selective sweeps in the eukaryotic pathogen Plasmodium falciparum
:
Three recent studies by Cheeseman et al. (2012), Takala-Harrison et al. (2013), and Miotto et al. (2013) examined patterns of LD and haplotype homozygosity in the P. falciparum genome and identified a region on chromosome 13 associated with artemisinin resistance. This locus was later characterized as the kelch13 locus (Ariey et al. 2014). These three studies benefited greatly from deep samples of P. falciparum from multiple populations, some of which displayed the artemisinin drug resistance phenotype of slow parasite clearance rate while others did not (Noedl et al. 2008; Dondorp et al. 2009). This data allowed the authors to contrast genomic patterns in strains from geographical locations displaying the slow parasite clearance rate phenotype vs. those that did not.

In the first major study, Cheeseman et al. (2012) examined genomic data from samples from three populations: Laos, Thailand and Cambodia, where resistance is observed only in Thailand and Cambodia (Cheeseman et al. 2012). This data set offered the opportunity to contrast the Thai and Cambodia samples with Laos to discover putative loci conferring resistance to artemisinin. Cheeseman et al. (2012) utilized a two-step approach to identify genomic regions underlying resistance to artemisinin. First, the authors used the XP-EHH (Sabeti et al. 2007) and FST (Lewontin & Krakauer 1973) statistics and identified 33 regions as significant candidates under selection in at least one population. These statistics were particularly appropriate for their multipopulation data set because they are designed to contrast genomic signatures in two populations (Box 2). Cheeseman et al. (2012) found that 10 of the 33 regions discovered were associated with positive selection in studies examining different drugs, validating their approach. Second, Cheeseman et al. (2012) examined the associations of these 33 regions with the slow parasite clearance rate phenotype and identified two adjacent single-nucleotide polymorphisms (SNPs) 14 kb apart in their sample on chromosome 13 showing significant associations. Cheeseman et al. (2012) then performed fine mapping in this region and applied the EHH statistic (Box 2) to visually confirm the extended haplotype homozygosity pattern in a 35-kb region on chromosome 13 characteristic of a selective sweep.
Similar to the Cheeseman et al. (2012) study, Takala-Harrison et al. (2013) examined multiple populations from Cambodia, Thailand and Bangladesh for signatures of selection and association with the slow parasite clearance phenotype, with only Cambodian and Thai samples showing resistance phenotypes. The authors used efficient mixed-model association (EMMA, see Box 2) (Kang et al. 2008) to test the association of each SNP in their data set and the parasite clearance phenotype, thereby identifying four SNPs significantly associated, including two on chromosome 13. In conjunction with the association test, Takala-Harrison et al. (2013) applied XP-EHH (Sabeti et al. 2007) and FST (Lewontin & Krakauer 1973) to the Cambodia population, using the Thai and Bangladeshi populations as comparison populations. The authors found that only the polymorphisms on chromosome 13 previously identified with the EMMA test also had significant XP-EHH and FST values. Takala-Harrison et al. (2013) used haploview (Barrett et al. 2005) to visualize the extended haplotype homozygosity in the region.
Miotto et al. (2013) examined 10 locations in West Africa and South-East Asia and found that there was an exceptionally high amount of population substructure within Cambodia. Upon closer examination, the authors found that three of four clusters of P. falciparum found in western Cambodia showed slow parasite clearance rates, slow decay in LD, and loss of haplotype diversity, while the fourth cluster prevalent in northeastern Cambodia did not show any of these characteristics. This leads the authors to conclude that western Cambodia harbours at least three distinct populations of artemisinin-resistant P. falciparum. In particular, one of the subpopulations with resistant strains of P. falciparum showed a single haplotype extending across half of chromosome 13, corroborating the evidence from Cheeseman et al. (2012) and Takala-Harrison et al. (2013) that this locus is implicated in resistance to artemisinin.

Experimental confirmation of the kelch gene as a likely target of adaptation ::: Linkage disequilibrium statistics for the detection of selective sweeps in the eukaryotic pathogen Plasmodium falciparum
:
While Cheeseman et al. (2012), Takala-Harrison et al. (2013) and Miotto et al. (2013) were all able to localize a putative locus on chromosome 13 strongly associated with the slow parasite clearance phenotype, it was only recently that Ariey et al. (2014) were able to identify causative mutations with high confidence. Ariey et al. (2014) used an in vitro drug selection technique to subject a parasite line to high doses of artemisinin for 5 years. Comparing the sequenced data from the selected line with that of a clonal population not experiencing any selection, the authors identified eight mutations in seven genes that were present in the artemisinin treatment group but absent from the control group. Ariey et al. (2014) narrowed their list of candidates and concluded that only the mutation appearing in the kelch gene on chromosome 13 appeared at the same time as when artemisinin resistance developed in their treatment group. To determine whether there was concordance between the presence of this mutation in the kelch gene and artemisinin-resistant parasites from Cambodia, Ariey et al. (2014) sequenced the locus at which these mutations were present in parasite samples from patients showing the drug-resistant phenotype and from patients who did not in different geographical locations in Cambodia. They found that mutations in the kelch gene were strongly associated with the slow clearance phenotype observed in the locations where malaria is prevalent. In this extended analysis, Ariey et al. (2014)

identified 17 mutations in the kelch gene, and all were significantly associated with artemisinin resistance.
Several follow-up studies confirmed the findings of Ariey et al. (2014). Ashley et al. (2014) examined the geographical extent of resistance by tracking the prevalence of mutations in the kelch gene (see Fig. 1). The authors found several single point mutations in kelch significantly associated with slow parasite clearance rates. Recent work (Cheeseman et al. 2015; Miotto et al. 2015; Takala-Harrison et al. 2015; Tun et al. 2015) has also precisely mapped the origins and extent of mutations associated with artemisinin resistance (see Fig. 1). Ghorbal et al. (2014) used the CRISPR-Cas9 system to introduce a mutation (C580Y) implicated in artemisinin resistance (Ariey et al. 2014) into kelch, which produced the slow parasite clearance phenotype and demonstrated the first direct link between a mutant kelch and the characteristic phenotype. This work was expanded upon by Straimer et al. (2015) to confirm the role of multiple mutations and additional genetic factors in conferring artemisinin resistance.
Of the five examples reviewed in this study, this example of using classic statistical methods to find the association between the kelch gene and artimisinin resistance in P. falciparum is perhaps the most familiar to population geneticists. As a eukaryotic pathogen with a high rate of recombination, sequence analysis of the P. falciparum genome was amenable to traditional genome scan approaches. Clever stratification of the samples gave additional power to the genomic scans, and in combination with novel experimental methods, researchers were able to find a locus under recent strong selection that was responsible for the evolution of resistance to artimisinin. The identification of the kelch gene by population genetics methods offers a clear example where standard methodologies to identify selective sweeps are powerful in identifying a causal locus. Given that P. falciparum can be manipulated experimentally to confirm computational predictions, this organism is an attractive choice for future studies of drug resistance, especially since malaria continues to be a costly disease and new drug-resistant loci may be unknown.

Epistasis as a determining factor in the evolution of oseltamivir resistance ::: Evolution of oseltamivir resistance in influenza virus:
The interactions among sites in the genome can give rise to a phenomenon called epistasis where the effects of one mutation at a site are dependent on the presence or absence of mutations at other sites. Epistatic interactions, when they exist, ensure that the fitness effects of a drug resistance mutation are dependent on the genetic background on which it emerges. Epistasis in drug resistance evolution has been characterized by several in vitro studies, including the case of the five interacting mutations involved in cefotaxime resistance in Escherichia coli that produce predictable mutation orders (Weinreich et al. 2006) and the hundreds of interacting mutations that determine viral fitness in antiretroviral resistance in HIV (Hinkley et al. 2011).
In the case of oseltamivir resistance, the H274Y mutation was shown to have detrimental effects on viral fitness after it was first identified in the laboratory (Ives et al. 2002; Abed et al. 2004; Herlocher et al. 2004). This led researchers to believe that viruses carrying H274Y were ‘unlikely to be of clinical consequence' (Ives et al. 2002). Why then did the resistant H1N1 strains subsequently reach a global frequency of nearly 95% (WHO 2009) during the 2008–2009 flu season? Bloom and others showed that the H274Y mutation reduces the amount of folded neuramindase that reaches the host cell surface and thereby reduces the virus' fitness. They hypothesized that secondary mutations at other sites in the influenza genome may have acted to ameliorate deleterious effects of the H274Y mutation, permitting it to reach high global frequency (Bloom et al. 2010).
Bloom and colleagues took sequences from before and after 2007 and created a phylogenetic tree. They found five nonsynonymous mutations that separated the 2008 strains in which H274Y was common from the 1999 strains in which H274Y was found to have a strongly negative fitness effect. They then added the mutations one by one to the 1999 strain and found that two of the five mutations (R222Q, V234M) had a strong effect on the amount of neuraminidase at the cell surface. In viruses that had these two mutations, H274Y no longer reduced the fitness of the virus. In the absence of oseltamivir, the created triple-mutant virus (with H274Y, R222Q, V234M mutations) had a fitness comparable to wild type. In the presence of oseltamivir, the triple-mutant virus had much higher fitness than wild type (Bloom et al. 2010). A phylogeny with the status of these three mutations (H274Y, R222Q and V234M) is shown in Fig. 2.
Subsequent studies also identified potential epistatic interactions between mutations in hemagglutinin and neuraminidase that impacted viral fitness and may have influenced the spread of oseltamivir resistance (Hensley et al. 2011; Ginting et al. 2012; Behera et al. 2015). Taken

together these results suggest that epistasis had a profound impact on the evolution of resistance to oseltamivir and allowed the H274Y resistance mutation to spread through the global population. This work also provides cautionary evidence for the clinical assessment of drug resistance mutations. Secondary mutations that interact epistatically with drug resistance mutations are important factors that need to be incorporated when making predictions about the epidemiological consequences of drug resistance mutations. Knowing that even deleterious drug resistance mutations can spread when they emerge on permissive backgrounds tells us that we should pay close attention to identifying and monitoring potential epistatic interactions during virological surveillance.

Epistasis and the predictability of adaptive evolution ::: Evolution of oseltamivir resistance in influenza virus:
To what extent is epistasis a general feature of adaptive evolution? Both theoretical and empirical work predict that adaptation itself may enrich for epistatic interactions (Draghi et al. 2011; Draghi & Plotkin 2013; Gong et al. 2013; Rajon & Masel 2013; Szendro et al. 2013; Gong & Bloom 2014). This means that epistasis is likely to play a more general role in adaptive evolution and is not limited to specific case studies. It also means that researchers interested in adaptation need model systems in which to test theory regarding epistasis. Influenza resistance evolution has been a formative model for testing population genetic predictions regarding the role of epistasis in adaptive evolution.
Oseltamivir resistance evolution provided some of the first evaluations for methodologies that can potentially predict the sites within proteins that interact with one another based on sequence data and phylogenetic trees. Bloom and colleagues initially predicted the R194G mutation as a potential candidate for a positive epistatic interaction with H274Y using a phylogenetic method (Bloom & Glassman 2009) to infer stabilizing effects of mutations. They showed that R194G did indeed restore H274Y-mutant surface expression to wild type levels in the absence of oseltamivir (Bloom et al. 2010). Meanwhile Kryazhimskiy et al. (2011) examined hemagglutinin and neuraminidase sequences of multiple influenza strains and successfully predicted a known epistatic interaction (Collins et al. 2009) between H274Y and another mutation (D344N, see Fig. 3) in addition to the V234M and R222Q mutations identified by Bloom et al. (2010). Their methodology was based on a statistical ranking of the co-occurrence of specific amino acid substitutions within a phylogeny.
Predicting the evolution of influenza is an intriguing prospect for evolutionary geneticists, which we will briefly discuss in the Discussion. These studies have shown that phylogenetic methods can not only affirm theoretical predictions regarding the patterns of epistasis but can also predict epistatic interactions that may be of clinical importance. They also establish a connection between protein phylogenies and fitness that may prove to be useful for future work in both population genetics and infectious disease.

SGV contributes to treatment failure in HIV ::: The role of SGV for drug resistance evolution in HIV:
Before an HIV-infected person starts treatment, a blood sample is taken to sequence the virus (see Schutten 2006 for an overview of genotyping assays). Sanger sequencing of HIV before the start of treatment is standard in clinical practice, but next-generation sequencing is not (Simen et al. 2009; Codoñer et al. 2011). A sequence of the protease, reverse transcriptase, and integrase genes are used to determine whether any drug resistance mutations are present at high frequencies, and if so, these results help the clinician and patient choose a combination of drugs with which to start treatment (Hirsch et al. 2008). If the majority of the viral population in a patient carries a resistance mutation, then this information will be used to choose a drug regimen that will work for the specific virus. However, drug resistance mutations at low population frequencies (at 20% or less) are not detected by standard sequencing protocols (Simen et al. 2009). An important question therefore is whether this low-frequency SGV is present in most patients, and if so, whether it allows the viral population to adapt and evolve drug resistance.
An important study by Paredes et al. (2010) used asPCR to determine whether drug resistance mutations were already present as minority variants in the viral population of 183 patients. These patients originally took part in a clinical trial, so blood samples from before the start of treatment and at treatment failure (if applicable) were stored for them. Treatment failure is defined as having virus detectable in the blood at levels higher than should be expected given that the patient is on treatment. The authors focused on two important resistance mutations in the reverse transcriptase gene, K103N and Y181C, because the patients were treated with a combination of reverse transcriptase inhibitors. In 73 of the 183 patients, they found that either K103N or Y181C

was already present as a minority variant (mostly around 1% frequency), but they could not detect the mutation in the other 110 patients. Of the patients with the minority variant, treatment failed in 26 but was successful in 47. Out of the patients without the minority variants, treatment failed in only 16 but was successful in 94. Altogether, treatment failed in 23% of the 183 patients. When only considering the patients without detectable SGV, treatment failed in 15% of the patients. This means that in 8% of the patients, failure can be attributed to the presence of a minority variant (see Fig. 4).
The estimate from Paredes et al. (2010) that 8% of treatment failures come from SGV may be an underestimate because the authors only looked at the presence of two drug resistance mutations, and they may not have detected the variants in all cases. However, the two mutations are the most important for the treatment they looked at, and the result is fairly similar to a different estimate which we will describe below.
A less direct method to study the role of SGV for the evolution of HIV drug resistance was developed by one of us (Pennings 2012). We re-analysed data from a previous study (Margot et al. 2006) and looked for excess adaptation early during treatment. We found that in the study of interest, resistance evolution happened at a constant rate in the second and third year of treatment. In both of those years, the virus acquired resistance in 3.5% of the patients who previously had a virus without drug resistance. Such a constant rate of evolution of drug resistance was also observed in several other trials (e.g. UK Collaborative Group on HIV Drug Resistance and UK CHIC Study Group and Others 2010). In years 2 and 3 of treatment, SGV that was present before treatment probably plays no role, so the 3.5% likely reflects evolution of resistance from de novo mutations. In the first year of treatment, it is expected that both new mutations and SGV contribute to the evolution of resistance, so if resistance evolved in more than 3.5% of the patients' viral populations, this suggests that SGV played a role.
In Margot et al. (2006), there were 600 patients in total in the first year of treatment. The expected number of people with resistance after 1 year was 21 (this is 3.5%, the rate that is seen in year 2 and year 3). In reality, there were 57 people with resistance, so 36 more than expected. Thirty-six of 600 is 6% (see Fig. 4), so it was estimated that SGV leads to resistance in 6% of the patients. One may note that the total number of failing patients is much higher in Paredes et al. (2010). This may be because they have a stricter criterion for treatment success [<200 copies for Paredes et al. (2010) vs. <400 copies for Margot et al. (2006)]. Also, the patients were on similar but not the same treatments (3TC/AZT/EFV or ABC/3TC/AZT/EFV in the Paredes study and 3TC/TDF/EFV or 3TC/d4T/EFV in the Margot study).

Quantifying the importance of SGV in drug resistance evolution ::: The role of SGV for drug resistance evolution in HIV:
In the light of different overall rates of evolution in the two studies, it is perhaps surprising to see that the estimated rate of evolution from SGV is similar (8% and 6%). When Pennings used a third data set to estimate the rate of evolution of resistance due to SGV—based on a trial of long treatment interruptions with 435 patients (Danel et al. 2009)—her estimate was again 6%, which suggests that these estimates are fairly robust (Pennings 2012).
Adaptation from SGV has been discussed extensively in the literature; however, there are relatively few studies that have attempted to quantify its relative importance as a mode of adaptation. Altogether, the results from studying HIV suggest that SGV plays an important and quantitatively predictable role in treatment failures for this particular disease. SGV may also play a large role in treatment outcomes for other diseases, both those caused by pathogens and others such as cancer (Bozic et al. 2013). Next-generation sequencing can be used to determine whether drug-resistant SGV exists in patients prior to treatment, thus informing clinical decisions for individual patients. Apart from its clinical relevance, the particular example of HIV drug resistance evolution is also well-suited for validating evolutionary theory regarding adaptation from SGV (Orr & Betancourt 2001; Hermisson & Pennings 2005; Barrett & Schluter 2008).

The dynamics of clonal interference in TB ::: Clonal interference among nonrecombining populations of Mycobacterium tuberculosis
:
Recombination is thought to be evolutionarily adaptive because it allows different adaptive mutations that reside on different genetic backgrounds in a population to appear together (Fisher 1930; Muller 1932; Felsenstein 1974). In addition, it can unlink positively selected traits from deleterious passenger mutations and mitigate the effects of linked selection (Hill & Robertson 1966; Birky & Walsh 1988; Good & Desai 2014). In populations with no recombination, each

mutation remains on the genetic background on which it originally occurred. Therefore, if multiple positively selective traits enter the population simultaneously on different backgrounds, these traits cannot be recombined to augment each other. Instead, they must compete against each other as they both rise in frequency in the population (Muller 1932). This process is known as clonal interference (Gerrish & Lenski 1998).

While clonal interference has been extensively modelled (Rouzine et al. 2003; Desai & Fisher 2007; Park & Krug 2007; Martens & Hallatschek 2011; Neher 2013) and investigated in laboratory experiments (Miralles et al. 1999; Pepin & Wichman 2008; Kvitek & Sherlock 2013; Lang et al. 2013), few systems exist that allow us to test these predictions in natural populations. Globally, influenza seems to evolve clonally across years (Strelkowa & Lässig 2012), but intrahost populations of nonrecombining bacteria allow for many semi-independent evolutionary trajectories to be compared over much shorter time scales. Although there is evidence that HIV experiences clonal interference within patients, recombination complicates its study (Pandit & de Boer 2014). There is also a strong interest in clonal evolution in cancer, but sampling at different time points is often much harder for cancer than for M. tuberculosis or HIV, which makes cancer as a study system more difficult (Ding et al. 2012; Walter et al. 2012).

Intrahost evolution of M. tuberculosis provides a fairly straightforward way to examine clonal interference in vivo and to understand the extent to which it can slow the evolutionary process. Because there is no recombination, allele frequencies from pooled resequencing can, in some cases, be reconstructed into haplotypes (see below). Its large genome size (~4.5 Mb), well-documented drug resistance mutations, and long time course of infection mean that patterns of diversity can be tracked over time. Although previous studies have shown the presence of competing clonal lineages in M. tuberculosis populations in a patient using neutral markers (Al-Hajoj et al. 2010; Navarro et al. 2011), we can further use our understanding of drug resistance mutations to better understand the evolutionary dynamics of within-host populations. Among patients who are infected with drug-sensitive strains of M. tuberculosis, drug resistance can evolve within a patient during treatment. By tracking the frequencies of drug resistance mutations over time in a single patient, we can observe clonal interference between lineages that each carry beneficial mutations in real time.

Sun et al. (2012) conducted the first analysis of this kind by tracking drug resistance allele frequencies in three patients across at least two time points. The first patient was entirely free of drug resistance mutations at the onset of treatment, but by the second time point of sampling had four segregating drug resistance mutations. At the final time point, 94% of the sample had a single drug resistance mutation, and 6% was divided among other drug-resistant strains, including those containing drug resistance mutations not present at the second time point. This suggests that while M. tuberculosis rapidly acquires resistance mutations, different drug-resistant strains compete against each other due to lack of recombination. The second patient entered the study with an M. tuberculosis population fixed for a certain drug resistance mutation (rpoB L533P) but remained sensitive to the antibiotic rifampicin. Eighteen months later, this patient had a population whose genetic composition was dominated by a different drug resistance mutation (rpoB H526Y) and was now resistant to rifampicin, suggesting that successive sweeps of alternative drug resistance mutations can lead to multidrug-resistant strains.

Eldholm et al. (2014) performed a similar study, but with much greater sequencing depth, in a single patient who was followed over 42 months as extensive drug resistance was acquired by the pathogen. The patient, who was started on a standard antibiotic regimen (pyrazinamide, rifampicin and isoniazid), was given increasingly uncommon drugs as their tuberculosis population acquired more drug resistance mutations. The dynamics of clonal interference are shown in Fig. 5, based on data from Eldholm et al. (2014). Allele frequencies were measured at several time points, so alleles with similar frequency trajectories can be assumed to be on the same background or in the same clone. Therefore, the clonal frequency changes can be inferred across time.

Eldholm et al. (2014) identified 12 drug resistance mutations reaching frequency >25%, of which only seven ultimately fixed (shown in green), suggesting that clonal interference may purge over 40% of strongly selected drug resistance mutations even after reaching high frequencies (shown in blue).

The knowledge of these positively selected drug resistance mutations can also allow us to understand the dynamics of co-occurring hitchhiking mutations that were already on the background on which the drug resistance mutation arose. Eldholm et al. (2014) found that 23 mutations not directly associated with drug resistance also fixed in the population over the 42 weeks of sampling. While these mutations may have been neutral, compensatory, or positively

selected through adaptation unrelated to drugs, it is also possible that some or all of them are slightly deleterious but rose in frequency due to their linkage to beneficial mutations.
From Eldholm et al. (2014) it may be noted that, in most cases, the drug resistance mutations were added to haplotypes one by one, which suggests that each individual mutation leads to an increase in fitness. This is troubling, because part of the rationale behind using multiple drugs at the same time is that it should make it much harder for the pathogen to evolve resistance, as it needs to acquire multiple mutations at the same time. It has been suggested that imperfect drug penetration may explain the evolution of resistance in tuberculosis despite multidrug therapy (Lipsitch & Levin 1998; Moreno-Gamez et al. 2015).
The study of clonal interference, thus far largely investigated within experimental and theoretical models, has a unique application in the study of clonal dynamics of natural populations of the human pathogen M. tuberculosis. Using the presence of drug resistance mutations as a positive control for strong directional selection, we are able to better understand the evolutionary dynamics of clonally evolving pathogen populations such as M. tuberculosis. Although we currently only have detailed data for a few well-studied patients that include deep coverage of the entire genome across multiple time points, it appears to be the case that clonal interference in M. tuberculosis populations occurs in a way that is similar to what is seen in laboratory experiments. With the knowledge of hundreds of drug resistance mutations and the availability of affordable whole-genome sequencing, we expect to see many more studies of within-host dynamics of tuberculosis in the near future. We especially look forward to studies where population's genetics can be used to understand genetic patterns associated with diverse treatment outcomes.

Clinical background and historical analysis of the Staphylococcus aureus genome ::: Evolution of drug resistance within populations of Staphylococcus aureus bacteria via horizontal gene transfer:

Staphylococcus aureus is a human commensal that is known for causing dangerous skin, blood stream and other hospital-associated infections as well as having the ability to become resistant to the majority of antibiotics (Lowy 2003; Stryjewski & Corey 2014). In addition to methicillin resistance, which first emerged in the 1960s, there is also a growing number of cases of resistance to other antibiotics, including vancomycin, quinolones, aminoglycosides, streptogramins, oxazolidinones and rifamycins (Lowy 2003; Stryjewski & Corey 2014). MRSA can be particularly deadly. For example, mortality rates for patients with MRSA bloodstream infections has been reported to be ~30% (De Kraker et al. 2011). In the USA ~94 000 infections and ~19 000 deaths per year are caused by MRSA (Stockman 2009). Drug resistance can arise via new mutations within the genome of a single clone (Strahilevitz & Hooper 2005; Howden et al. 2008), via horizontal gene transfer between clones (Coombs et al. 2011) and even via horizontal gene transfer from another bacterial species (Hanssen & Ericson Sollid 2006; Bloemendaal et al. 2010; Malachowa & Deleo 2010; Smyth et al. 2012). Staphylococcus aureus strains acquire methicillin resistance via the mobile genetic element SCCmec [most likely via transduction (defined in Box 3) (Maslanova et al. 2013)], which integrates as a cassette of genes into the bacterium's chromosome.
The genome of S. aureus is ~2.8 Mbp total, where ~2.3 Mbp compose the core genome and ~0.5 Mbp compose the accessory genome. The most common method for analysing the genomes of S. aureus isolates has been multilocus sequence typing (MLST), in which a standardized set of housekeeping genes from the core genome are sequenced and categorized into allelic types, allowing placement of each isolate into a defined ‘clonal complex' (Enright & Day 2000; Enright et al. 2002; Feil et al. 2004; Maiden et al. 2014). MLST has historically been applied to the core genome, revealing that S. aureus is a highly clonal species with relatively few recombination events in the core genome (Robinson & Enright 2004). When MLST is applied to international collections of isolates, it appears that a small number of clonal lineages are responsible for most infections. For example, Oliveira et al. (2002) applied MLST to over 3000 MRSA isolates from hospitals across Europe, South America, and the USA, and found that just five clones caused 70% of the infections (Oliveira et al. 2002). As findings such as these emerged, it was thought that methicillin resistance was likely to have spread across the population of S. aureus when relatively few clones acquired SCCmec and then clonally expanded (i.e. genomewide sweeps) (Kreiswirth et al. 1993; Crisóstomo et al. 2001; Feil & Enright 2004). However, closer analysis of the population structure of the S. aureus accessory genome reveals a different story.

The mobile element SCCmec as a clinical indicator of hospital- and community-associated infections ::: Evolution of drug resistance within populations of Staphylococcus aureus bacteria via horizontal gene transfer:
In the first few decades of MRSA outbreaks, it was found that the strains causing hospital-associated (HA-MRSA) vs. community-associated (CA-MRSA) infections displayed distinct clinical and genetic differences (these differences may be eroding, as will be addressed later). HA-MRSA tends to be resistant to a wider array of antibiotics and often causes blood stream and other hospital-related infections in individuals with additional medical conditions. CA-MRSA tends to be more virulent (caused by virulence genes also transmitted via HGT) and often causes skin infections in otherwise healthy individuals (David & Daum 2010; Chua et al. 2014). These clinical differences are partly mediated by the mobile genetic element called SCCmec, a chromosomal gene cassette that confers resistance to methicillin and can also confer resistance to other antibiotics (Katayama et al. 2000; Hanssen & Ericson Sollid 2006). Interestingly, genetic changes within this gene cassette track successive epidemic waves of MRSA (Katayama et al. 2000; Chambers & DeLeo 2009), and extensive work has been done to investigate the source and spread of this mobile element, both as a phylogenetically informative sequence and as an adaptive factor in its own right. SCCmec is currently classified into 11 types (with more subtypes) using mutations and the orientation of segments within the cassette (Milheiriço et al. 2007; IWG-SCC 2015).
There have been a number of studies showing that the population structure of S. aureus is different between HA-MRSA and CA-MRSA strains (David & Daum 2010; Mediavilla et al. 2012; Stryjewski & Corey 2014). Both HA-MRSA and CA-MRSA strains are thought to arise when methicillin-sensitive clones newly acquire a SCCmec cassette from a source population, most likely the staphylococcus species S. epidermidis (Wu et al. 1996; Hanssen et al. 2004, 2005; Meric et al. 2015). It has been found that HA-MRSA strains tend to carry the longer SCCmec types I, II, and III elements in just a few clonal backgrounds, whereas CA-MRSA strains tends to carry the shorter SCCmec types IV and V elements in diverse clonal backgrounds (Enright et al. 2002; Robinson & Enright 2003; David & Daum 2010; Coombs et al. 2011). For example, the early HA-MRSA pandemic could be relatively neatly classified into just five clonal complexes (Musser & Kapur 1992; Fitzgerald et al. 2001; Enright et al. 2002; Robinson & Enright 2003). In contrast, as researchers began analysing the strains responsible for the emerging CA-MRSA health threat, they found these strains to display more diversity and a stronger association between clonal type and geographic location (Chua et al. 2010; David & Daum 2010; Coombs et al. 2011). As of 2006 the SCCmec type IV element alone was found to have entered at least nine different clonal complexes of S. aureus (Lina et al. 2006).
The clinical and genetic differences between hospital-associated (HA-) and community-associated (CA-) MRSA are eroding in the modern pandemic, mainly due to the emergence of CA-MRSA strains that seed an increasing number of hospital outbreaks (Chambers & DeLeo 2009; David & Daum 2010; Mediavilla et al. 2012; Hsu et al. 2015). A study of MRSA infections in 2004–2005 in the city of San Francisco, California, found that ~90% of MRSA infections were acquired in the community (Liu et al. 2008). In the United States, there is currently a predominant CA-MRSA clone, called USA300 (typically containing SCCmec Type IV), that is responsible for the vast majority of community-associated infections in addition to causing hospital outbreaks (Seybold et al. 2006; Chambers & DeLeo 2009). Mathematical models predict that these CA-MRSA strains will eventually replace the strains traditionally categorized as HA-MRSA, rendering the community-associated and hospital-associated categories far less useful (D'Agata et al. 2009). The success of these CA-MRSA strains is potentially mediated by the arrival of the SCCmec Type IV cassette, which may be particularly suited to recurrent horizontal transfer because it confers faster growth rates at little or no fitness cost (Okuma et al. 2002; Diep et al. 2008; Chambers & DeLeo 2009; Mediavilla et al. 2012).

High-resolution population structure analysis of Staphylococcus aureus using the mobile element SCCmec
::: Evolution of drug resistance within populations of Staphylococcus aureus bacteria via horizontal gene transfer:
As the clinical and genetic differences between CA- and HA-MRSA strains have eroded in the modern pandemic, so too has the emphasis on their differences within recent literature. Instead, attention is moving towards leveraging new sequencing technologies (Maiden et al. 2014) to achieve higher resolution into the global patterns of methicillin resistance acquisition. Overall, analyses of the core genome of MRSA strains suggest that relatively few methicillin-resistant

clones have dispersed widely across the globe (Robinson & Enright 2003), while analyses of the accessory genome of MRSA suggest that new strains emerge locally via frequent de novo transfers of the SCCmec element. In other words, it appears that there has been clonal expansion of relatively few core genomes, while the accessory genome frequently goes through gene-specific sweeps. To illustrate this, Lina et al. (2006) applied MLST to both the core genome and the SCCmec element and found no association between clonal complex background and SCCmec element sequence type. The authors therefore concluded that SCCmec appears to transfer both repeatedly across distinct clonal complexes (particularly in the case of SCCmec IV) and repeatedly within a clonal complex, which suggests that SCCmec spreads through gene-specific sweeps. Chua et al. showed in another study that MRSA from geographically and genetically distant isolates (i.e. distantly related core genomes) can nevertheless have relatively conserved accessory elements, including SCCmec type IV (Chua et al. 2011), which again suggests that transfer happens often and sweeps are gene-specific. Another example is a study by Nübel et al. (2008) where the authors used an improved sequence typing method of the core and accessory genome to investigate the genetic relationships within a single clonal lineage (ST5, belonging to clonal complex CC5) using 135 isolates from across 22 countries and six continents. In contrast to lower resolution studies that came before, their analyses of the core genome sequences found geographically-associated phylogenetic clades within the ST5 clonal lineage. Additionally, their analyses of the SCCmec sequence revealed that at least 23 independent transfers of SCCmec occurred into this clonal group alone, with acquisitions appearing to occur locally (rather than a single acquisition then disseminating globally; Fig. 6). We could refer to this case of multiple origins of the adaptive SCCmec element as a soft, gene-specific sweep (Pennings & Hermisson 2006). The authors estimate that previous calculations for the rate of SCCmec acquisitions were at least an order of magnitude too low.

Taken together, we now know that genetically diverse core genomes can nevertheless contain closely related resistance alleles and that frequent transfers can occur even within a clonal lineage. This makes understanding the evolutionary history of these alleles complicated, because alleles can be transmitted independently of the clonal expansion of core genome lineages. Nevertheless, these studies of drug resistance in S. aureus illustrate that researchers can achieve fine resolution into the demographic and adaptive history of S. aureus using improved sequencing technologies, in this case facilitated by the exploration of both the core and accessory genomes. As researchers begin to use whole-genome sequencing in S. aureus, an even finer resolution into the evolutionary history of MRSA outbreaks is possible. For example, Holden et al. (2013) were able to pinpoint both the population size changes and geographic origin of a specific clone (EMRSA-15/ST22), and Harris et al. (2013) were able to construct the transmission pathway of a hospital outbreak between staff and patients. Thus, the growing availability of sequence data and whole-genome sequencing technologies has provided the ability to determine that resistance may generally be unconstrained by global transmission dynamics and instead can be tracked on local spatial scales. This presents researchers with the opportunity to use sequencing as a tool to monitor transmission networks within hospitals to limit the numbers of hospital-acquired drug-resistant infections. Overall, the transmission and repeated evolution of drug resistance in bacterial populations capable of HGT presents important and interesting problems for reconstructing population structure and spatial distributions of resistance.

The advantages of pathogens as evolutionary model systems ::: Discussion:

The field of population genetics traditionally leans heavily on the use of fruit flies, yeast, Escherichia coli, and humans as model organisms. However, it is becoming clear that many pathogens can also be excellent model systems. Pathogen evolution as a model evolutionary system has many advantages. Pathogens receive a lot of attention from biomedical researchers, and as a consequence, they are well studied and extensively sampled. Some pathogens already function as model organisms in the lab [e.g., influenza (Foll et al. 2014), HIV (van Opijnen & Berkhout 2005), polio (Acevedo et al. 2014) and E. coli (Toprak et al. 2012)], which allows researchers to assess characteristics of population genetic and epidemiological relevance from laboratory populations, for example the cell surface expression assays that were performed in the previously mentioned study of oseltamivir resistance in influenza. In some pathogen systems, researchers can also apply known selective pressures with precision in a laboratory environment, thus allowing high-resolution measurements of an organism's response to selection. Additionally, because some pathogens evolve drug resistance quickly and repeatedly within patients, there are cases where many replicate evolutionary histories can be studied (HIV, HCV, TB, etc.).

Pathogens are also among the organisms for which the most genetic sequence data are available for population genetics analysis: In 2014, over 270 000 S. aureus sequences, 50 000 influenza sequences, 40 000 HIV sequences, and 5000 malaria sequences were released in GenBank alone (Benson et al. 2013; accessed 26 May 2015). Moreover, decreased costs of next-generation sequencing will make the amount of new data for pathogens increase exponentially. Throughout this review, we highlight uses of sequence data and population genetics theory in studying pathogens as evolutionary model systems. We believe that pathogen evolution will continue to be a fruitful area of research for evolutionary biologists where fundamental insights can lead to beneficial public health consequences.

Frontiers in the study of pathogen evolution and drug resistance ::: Discussion:
As more sequencing data emerge in the foreseeable future, there are some specific questions that we would like to see addressed at the intersection of population genetics and drug resistance in pathogens. First, we are expecting more work that bridges within- vs. between-host dynamics (e.g. Lythgoe & Fraser 2012). This is important because considerable genetic variation is created and selected for within the host; however, the within-host processes of short-lived infections have not yet received sufficient attention. We are expecting more within-host work, including time series and deep samples for pathogens that are traditionally studied at the between-host level. The M. tuberculosis work described in this review is a good example, but new within-patient work on influenza is also being done (Rogers et al. 2015). Second, with deeper samples and more time series, it should become increasingly possible to estimate selection pressures on pathogens and to estimate epistatic interactions between mutations. For some of this work, methods that are currently used for identifying mutations and evaluating selection in viruses grown in cell culture (Lou et al. 2013; Foll et al. 2014) will hopefully find use in clinical samples taken from infected patients. Extending these methods to patient data can potentially give researchers the ability to study the evolutionary dynamics of pathogens within patients over the entire course of an infection.
Another important goal for these evolutionary studies of pathogens was to increase our predictive power of pathogen evolution. Until very recently predicting evolution seemed like science fiction, but recent work is changing our perceptions of what is possible. In addition to the predictive phylogenetic methodologies mentioned in our discussion of epistasis in influenza, a method that uses the shapes of protein phylogenies to predict evolutionary trajectories was evaluated using influenza data (Neher et al. 2014). The work shows that predictability is low when evolution occurs by big steps, but much higher when evolution proceeds by small steps (Neher et al. 2014). Although we tend to fix our attention on large adaptive events, such as drug resistance or immune escape, evolution by small steps may be more common than previously expected (Bhatt et al. 2011; Strelkowa & Lässig 2012; Gong et al. 2013). In an organism like influenza where new vaccines are manufactured every year, the application of these and other new predictive methodologies that use population genetics theory (Luksza & Lässig 2014) are potentially groundbreaking.
We hope that the works on drug resistance reviewed here offer a glimpse into a growing field at the intersection of studies in pathogen evolution and evolutionary genetics. We also hope to have illustrated how these fields of study can benefit greatly from each other and that this review contributes to the exciting future of this field.

unknown_1:
All authors designed the study and contributed to the writing and editing of the manuscript. B.A.W. wrote the chapter on Influenza, N.R.G. wrote the chapter on Malaria, Z.J.A. wrote the chapter on MRSA, A.F.F. wrote the chapter on TB and P.S.P. wrote the chapter on HIV.