Visualizing evolution in real-time method for strain engineering

Luis H. Reyes, Texas A&M University
James D. Winkler, Texas A&M University
Katy C. Kao, Texas A&M University
The adaptive landscape for an industrially relevant phenotype is determined by the effects of the genetic determinants on the fitness of the microbial system. Identifying the underlying adaptive landscape for a particular phenotype of interest will greatly enhance our abilities to engineer more robust microbial strains. Visualizing evolution in real-time (VERT) is a recently developed method based on in vitro adaptive evolution that facilitates the identification of fitter mutants throughout the course of evolution. Combined with high-throughput genomic tools, VERT can greatly enhance the mapping of adaptive landscapes of industrially relevant phenotypes in microbial systems, thereby expanding our knowledge on the parameters that can be used for strain engineering.

Characterization of the adaptive landscape will significantly enhance our knowledge on the important parameters underlying complex phenotypes needed for the rational engineering of strains.

In adaptive evolution, clones are typically isolated from the evolving population after an arbitrarily elapsed time or at the end of the experiment. However, since the evolving population is heterogeneous, interclonal competition (clonal interference; Shaver et al., 2002; Kao and Sherlock, 2008) may lead to the extinction of beneficial mutants. Depending on the population structure during the course of evolution, the random isolation of adaptive mutants may fail to identify some adaptive mutations that arise during the course of the evolution. This review will (1) discuss factors that influence population structure and the impact of complex population dynamics on evolutionary engineering and (2) describe a novel evolutionary engineering method called visualizing evolution in real-time (VERT), that was recently developed to help address some of these limitations in traditional evolutionary engineering approach.

ADAPTIVE LANDSCAPE

The idea of an adaptive landscape was first introduced as “surfaces of selective value” by Wright in 1931 (Wright, 1931, 1982, 1989). The adaptive landscape is a multi-dimensional surface representation of the biological fitness of an organism in a particular environment. In an adaptive landscape map for a specific condition, each genotype is correlated with a fitness value (see Figure 1 for an illustration). The resulting landscape can be flat with a single optimum where the evolving population is required to acquire a specific set of mutations, or can be rugged where the accessible optima will depend on the starting point within the landscape. It has been demonstrated that bacteria encounter both types of landscapes in evolution experiments (Orr, 2005; Winnerich et al., 2006;
Gresham et al., 2008). Natural selection usually drives a population to the closest local optimum, but not necessarily the global optimum. Evolving populations tend to be trapped in suboptimal solutions (Lenski et al., 1998) in asexual systems. Thus to reach the global optimum, processes that allow for large “jumps” in the adaptive landscape, such as recombination and horizontal DNA transfer, are necessary to reach new regions of the adaptive landscape in a semi-rational manner. Recombination allows the combination of beneficial mutations with positive synergy and the removal of deleterious mutations acquired in the evolutionary process while horizontal gene transfer allows the acquisition of new functions.

THEORIES GOVERNING POPULATION STRUCTURE DURING ASEXUAL EVOLUTION

Numerous theories have been proposed for the population structure in in vitro adaptive evolution experiments. Several factors, including the selective pressure, size of the population, rate of mutations, frequency of beneficial mutations, and relative fitness of beneficial mutants, are involved in determining the population structure during evolution. In the simplest case, a well-adapted mutant rises in the population, and due to its increased fitness compared to background, the genotype will expand and eventually replace the parental population. This population structure is applicable to situations where the evolution is mutation-limited, the population size is small, and the time between the establishments of successive mutations is much larger than the time it takes for a beneficial mutant to fix in the population (strong positive selection). This theory, called clonal replacement (also called succession-fixation regime or strong-selection weak-mutation regime), implies that only one mutation can become fixed at a time, leading to successive complete selective sweeps (depicted in Figure 2A). The resulting population can be assumed to be homogeneous except during the periods when the beneficial mutant is sweeping through (Desai et al., 2007). However,

FIGURE 1 | Simplified adaptive landscape for two alleles (for one background genotype in one condition). The figure depicts fitness values for beneficial (positive relative fitness values) and deleterious (negative relative fitness values) combinations of alleles.

FIGURE 2 | Theories governing population structure during asexual evolution. The graphs represent the population structure as a function of time during asexual evolution. The capital letters represent different beneficial mutations in the population. The gridded boxes represent a snapshot of the frequency of different genotypes in the population at that one point in time. (A) Clonal replacement model, where successive sweeps and fixation of different beneficial mutations take place in a small population; snapshots of the genotypes at different elapsed times show that the population is homogeneous except when the beneficial mutant is sweeping through the population. (B) Clonal interference model, where different adaptive mutants compete until one with the largest fitness advantage sweeps through and becomes the founding genotype for subsequent evolution (e.g., mutations A, B, and C compete until C completely takes over the population). (C) Multiple mutations model, where multiple mutations occur in the same lineage before fixation. In the latter two population structures, some adaptive mutations are lost from the population, and depending on when adaptive mutants are isolated, some mutants (and thus the underlying molecular mechanisms for adaptation) may not be identified.
when the mutations are established at a faster rate than the rate of fixation, multiple mutant lineages can coexist and compete for resources until one with the largest fitness advantage outperforms all the other genotypes and becomes the next founding genotype for subsequent evolution. This theory, known as clonal interference (or one-by-one clonal interference), assumes that a single mutation can be fixed at a time, producing heterogeneous populations except immediately after the sweeping of the fittest mutant (depicted in Figure 2B). This theory focuses on the competition between different mutations with positive relative fitness (Gerrish and Lenski, 1998; Orr, 2000; Gerrish, 2001; Kim and Stephan, 2003; Campos and de Oliveira, 2004; Wilke, 2004). The two theories described above assume that only one beneficial mutation can be fixed at a time. However, if the population size is large enough or the rate of mutation is high enough, multiple mutations can occur in the same lineage before fixation, leading to the multiple-mutation model (Douki et al., 2007; depicted in Figure 2C). The importance of this third theory on population structure has been demonstrated in several theoretical and experimental studies (Teitel and Bell, 2001; Shaver et al., 2002; Backstrom and Gordo, 2004). In general, the population size in laboratory conditions is large enough that either one-by-one clonal interference or multiple mutations models shape the population structure.

FACTORS INFLUENCING POPULATION DYNAMICS

As mentioned above, factors such as mutation rate, relative fitness advantage, population size, and rate of beneficial mutations are important in shaping the population dynamics during evolution. We will briefly discuss each of these factors and how they impact adaptive evolution in different experimental systems. Since the evolution dynamics is dependent on the mutation rate, one would assume higher mutation rate to be advantageous for speeding up evolution by generating mutational diversity. However, an increase in mutation rates does not necessarily accelerate the pace of adaptation (Arjan et al., 1999). While a low mutation rate would result in a slow discovery of beneficial mutations, prolonged exposure to high mutation rate (such as the use of a mutator strain) increases the occurrence and accumulation of deleterious mutations as well as the hitchhiking of apparent "silent" mutations during the course of evolution, increasing the genetic load (Elena and Lenski, 2003; Gresham et al., 2000; Barrick et al., 2009). This is evidenced by the rarity of mutator strains in Lenski’s long-term adaptive evolution experiment with Escherichia coli, where mutators were found only after thousands of generations of evolution (Elena and Lenski, 1997; Stuegowski et al., 1997; Arjan et al., 1999; Vulic et al., 1999; Lenski et al., 2003) and the fitness advantage conferred by the mutator strains is most likely a result of overcoming a mutation-limited bottleneck during the evolution. Mutagens are often used to increase genetic diversity in evolution experiments. However, since it is not convenient to periodically mutagenize the evolving population, a controllable mutator system can be used, where the expression of mutator alleles can be induced only when needed (Selkonova et al., 2001).

The time it takes a beneficial mutation to become the majority in the population is called the fixation time and is an important factor in determining the population dynamics during evolution. This fixation time depends mainly on two factors, genetic drift and the fitness advantage of the beneficial mutation in comparison with the background, and is inversely proportional to the relative fitness advantage of the beneficial mutant (Lenski et al., 1991). A beneficial mutation with a 10% relative fitness advantage will become the majority of the population after approximately 250 generation in serial batch transfer experiments (Elena and Lenski, 2003) and 100 generations in continuous culture experiments (Gresham et al., 2008). Genetic drift is defined as the probability that a beneficial mutation survives extinction (Joanna, 2013). In vitro adaptive evolution experiments, the main source of drift is genetic bottleneck due to random sampling. This phenomenon takes place when a significant amount of the population suddenly vanishes, as occurs when a fresh batch culture is inoculated from an overnight culture. The survival probability of an allele carrying a beneficial mutation that arose in the culture will depend on its proportion in the culture at the time of transfer and the amount of inoculum transferred; therefore there is a chance that it could be completely lost due to the stochasticity of sampling. In evolution experiments using serial batch transfers, genetic bottlenecks between transfers affect heterogeneity by transferring a small fraction of the population. A reduction of the effect of genetic bottleneck could be achieved by using continuous culture systems such as chemostats or turbidostats (Conrad et al., 2013), where a much smaller genetic bottleneck is present.

VISUALIZING EVOLUTION IN REAL-TIME

As stated above, the population sizes in most in vitro evolution experiments are large enough to result in heterogeneous populations due to the effects of clonal interference and multiple mutations. Thus, adaptive evolution experiments can significantly benefit from a more systematic isolation of adaptive mutants and ramping-up schedules for selective pressures. The VERT system was developed (Kao and Sherlock, 2008; Huang et al., 2011) to address these limitations in traditional adaptive evolution experiments. The basis for VERT is the use of isogenic, but differentially labeled (typically with fluorescent proteins) strains to seed the initial evolving population. As a beneficial mutant arises and expands in the population, the colored subpopulation that it belongs is expected to increase in proportion. Using fluorescent activated cell sorting (FACS), the relative proportions of each of the colored subpopulations at each point in time can be measured. Each sustained expansion in the proportion of a colored subpopulation is called an “adaptive event.” Thus, the tracking of the different colored subpopulations can serve as a tool for determining when a fitter mutant arises in the population.

The relative subpopulation frequency data collected throughout the course of adaptive evolution represent the history of the population. The observed increase in the relative proportion of a colored subpopulation from consecutive data points is assumed to be the result of the expansion of an adaptive mutant. Therefore, adaptive mutants can be isolated from samples based on the observed expansions and contractions, by sorting out the colored subpopulation that is expected to contain the adaptive mutant of interest. Since experimental data can suffer from noise,
the identification of adaptive events may be challenging. Visual inspection of the data to identify adaptive events is a crude, but relatively effective method (Kao and Sherlock, 2008; Huang et al., 2011). However, since small changes in relative frequencies may be difficult to distinguish from noise, computational methods will provide less biased annotation of adaptive events; our group recently developed a supervised learning method for analysis of VERT data (Winkler and Kao, 2012).

The basic feature of the VERT system, the number of labeled subpopulations, is the aspect that can most readily be manipulated directly by the experimentalist, but is somewhat restricted by the available equipment and properties of the labels themselves. The number of fluorescent markers used represents distinct subpopulations that can be visualized during the course of an evolution experiment. VERT labels must have distinguishable emission spectra and preferably have no significant fitness effect in the condition of interest. Widely used fluorescent proteins such as GFP, YFP, and RFP can be detected on most FACS machines and usually have little effect on the physiology of their host strains. At a minimum, two labeled subpopulations are trivially required to observe population dynamics. Three subpopulations, employing RFP, GFP, and YFP labeled strains, have been used successfully (Kao and Sherlock, 2008; Huang et al., 2011) in fungal systems. Additional subpopulations can be included if suitable equipment is available. Simulated evolution may prove a useful tool for unraveling the connection between adaptive event discovery and initial population diversity.

Visualizing evolution in real-time-based in vitro adaptive evolution experiments can be used in either serial batch transfer or continuous culture systems. Provided that the different fluorescently marked strains show no significant fitness bias, then equal proportions of each strain might be used to seed the population for evolution. Samples are then withdrawn and quantified using FACS every few generations to track the population dynamics. It is typically assumed that the adaptive mutant will expand until a fitter mutant arises in another subpopulation and expands sufficiently to impede its’ expansion. It is further assumed that the generation at which the expanding subpopulation has reached a maximum proportion will contain the largest fraction of the adapted subpopulation, and is defined as an adaptive event. Under the assumption that the adaptive mutant responsible for the specific adaptive event is at its’ highest proportion at the end of the sustained expansion, the adaptive mutant is isolated from the expanding subpopulation from which adaptive mutants were isolated from are indicated. Detailed genotypic and transcriptomics analyses of the isolated adaptive mutants showed convergence in the perturbation of the protein kinase A regulatory network in independent lineages (Kao and Sherlock, 2008). Subsequent development and application of a two-colored VERT system in E. coli for n-butanol tolerance revealed previously undiscovered resistance mechanisms (Reyes et al., 2012).

**Figure 3** Example population dynamics from a three-colored VERT system (adapted from Kao and Sherlock, 2008). The colored bars represent the relative proportions of each colored subpopulation. An increase in the relative proportion of a colored subpopulation is indicative of the occurrence and expansion of an adaptive mutation in that subpopulation, and is defined as an adaptive event. Under the assumption that the adaptive mutant responsible for the specific adaptive event is at its’ highest proportion at the end of the sustained expansion, the adaptive mutants are isolated from the expanding subpopulation from which adaptive mutants were isolated are numbered 1-5.
CONCLUSION
Understanding the adaptive landscape for the phenotypes of interest in evolutionary engineering tools. The use of evolutionary engineering has been used extensively to improve strains for complex phenotypes where there is limited knowledge on the associated genetic determinants. Advances in molecular biology tools and methods have significantly improved our ability to obtain insights into the molecular mechanisms involved in the desire phenotypes in isolated adaptive mutants from in vitro evolution experiments. VERT is a recently developed tool for evolutionary engineering that can help to provide a rough population structure for the evolving population, allowing the systematic isolation of adaptive mutants and ramp-up of selective pressure. Combined with advanced computational tools, use of VERT in evolutionary engineering can help to gain additional insight regarding the adaptive landscape for complex phenotypes.

ACKNOWLEDGMENTS
The authors would like to thank NSF MCB-1054276 (Kay C. Kao), and the Graduate Research Fellowship program (James Winkler) for partial financial support.

REFERENCES


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.