Some thoughts about designing experimental evolution (EE) experiments:

One obvious difference between EE and typical lab experiments is the much longer time scale involved. Depending on the system, environment and question, experiment lengths of >1 year are common. This length poses a challenge: you need to get the design right for both the immediate question and, ideally, for potential future questions. I think generally good advice is to (1) propagate replication and treatments at the high-end of what you envision using and (2) store population samples at -80 as often as practical (perhaps around every 100-250 generations). It's easy to drop treatments from subsequent analyses if things get too much, it's much less convenient to add them.

Some aspects of experimental design to be considered.

Replication. How many replicate populations should be propagated? If you want to test some question as to the effect of a selection environment on evolutionary response, then a level of replication will be needed depending on expected measurement error and intrinsic variability in response. (You can use a 'power analysis' to get some feeing for this). A typical number might be around 6 replicate populations per treatment. Higher replication makes it possible to make statistically reasonable inferences from even small differences in the average response between populations evolved in different treatments. Different types of questions can dictate different considerations. For example, lower replication might suffice to test for a 'main' effect – are two treatments different from one another – than for an 'interaction' effect – does some response differ depending on the combination of two factors.

Treatment number. There is obviously a temptation to assess as many different treatments as possible. It is often relatively easy to propagate large numbers of populations (e.g. in a 96 well plate), but be aware of the trade-off with lower replication within each treatment.

Environment. Some aspects of the environment will be dictated by the motivating question – e.g., the nature of the limiting resource. Other things to consider include, volume, transfer regime, base medium, culture container and shaking regime. One key thing to bear in mind is consistency. Any regime that suffers from high day-to-day variation will result in noisy estimates of evolutionary responses.

> *Container.* Think about the desirability of making high throughput measurements in a regime as similar as possible to the evolution environment. E.g. if growth curve measurements will form an integral part of the analysis, and you know that bacterial growth depends on the container material (e.g. polystyrene vs. polypropylene), you might choose containers for propagation made of the same material you will use for analysis.
>
> *Transfer volume.* This should be big enough to reproducible. It is easy to get a higher number of generations per day by transferring smaller numbers of

cells into fresh medium. Note, though, that the number of evolution generations does not directly correspond to the opportunity for evolution. The higher the dilution factor, the higher the chance that new (and therefore rare) variants in a population will be left behind. E.g., a paper by Wahl and Gerrish (Evolution 55: 2606-2610) that predicts that the best dilution factor for maximising the rate of bacterial adaptation is ~10. If it is possible to use a 96 well pin replicator to do daily transfers, that makes life easy. If the medium supports a consistent number of cells at saturation (this will usually be the case in defined liquid environments, maybe not on solid), and the length between transfers allows this number to be reached, the number of generations per transfer cycle is simply $\log_2$(dilution factor).

Base medium. Depending on the question you are asking, it will probably simplify subsequent analysis if the evolution environment is ecologically simple. Because most regimes will propagate enough individuals that multiple mutations will be present in a population at one time, it is never safe to consider any one individual to be representative of its source population. Complex environments exaggerate this complication. For example, inclusion of many resources might allow different sub-populations to arise and be maintained by specializing to different ecological niches. It is usually possible to screen for these kind of effects, but it does make it more difficult to think about the analysis. For one thing, the notion of relative fitness becomes complicated when it depends on the frequency of a type in the population– this will usually not be the case if types interact only through competition for a single resource. Choose a medium that you can be confident will be very similar between different batches. If your experiment has a treatment in a constant environment, you really want the environment to be constant.

Shaking. Anything acting to reduce environmental heterogeneity (like shaking) reduces the chance of multiple ecological types. Good or bad, depends on you question.

Transfer regime. Daily is convenient – but no reason not to go longer or shorter. It will usually be important (at least to keep track of number of generations) if time between transfers is long enough that the culture has become saturated and short enough the cell death is low. Decide on an interval around the ideal time of transfer that you will accept (e.g. +/- 1hour). We store the previous days cultures in the fridge until it is clear that the next block of cultures is OK. If it is not, we restart cultures from the refrigerated set. Any problems going back longer than one day and we restart from the last frozen set.

Storage. Frequently storing populations is a good idea in that it allows subsequent analysis of dynamics. We freeze every 100-250 generations, depending on the experiment. My recommendation would be to start out freezing often, then, reduce if it is too onerous. We used to store clones periodically, in retrospect, I think clones can be isolated from stored populations once it is clear that they will be useful. You

can store whole 96 well blocks, but, at least periodically, it is a good idea to store populations in individual tubes to reduce the risk of contamination.

Starting genotypes. Any subsequent analysis of reproducibility depends on mutational changes being independent. Start replicate populations with a small number of individuals to reduce the chance that rare mutations will be carried over into the beginning of multiple replicate populations. Think about how you are going to do subsequent assays. Make sure the ancestor is suitable – does it have a marked derivative? If you start the experiment with derivatives of the same genotype with different markers (e.g., for us, Ara+ and Ara-; GFP+ and GFP-), it is usually desirable that the markers used do not effect evolutionary trajectories. In practice, it is hard to be absolutely confident of this. Any fitness effect is probably bad, but even neutral markers may interact with subsequent mutations. (It is useful to have differently marked populations to enable screening for cross-contamination.)

Block design. It is crucial to be able to screen for contamination of populations. External contamination will probably be easy to identify through periodic plating of populations. Cross-contamination is harder to track. When we propagate long-term in 96 well blocks we follow a checker-board pattern to intersperse inoculated wells with blank wells that provide a means to detect splashing. Any time a blank well is contaminated, we start that days propagation over. We also arrange populations such that closest neighbours differ at a neutral marker to enable easy screening for cross-contamination. Populations can begin to look very different over time, it can be helpful to photograph screening plates to track changes and be more confident that weird looking populations are OK.

Fitness competitions:

#1. Ara marker competitions (protocol (and some text) based on http://myxo.css.msu.edu/ecoli/compet.html)

The basic idea is to mix a reference strain with a test strain or population where the reference is chosen to have the opposite Ara marker then the test. A sample of the mix is plated at the beginning and end of a competition period and the change in ratio is used to estimate the relative fitness difference. In general, the environment (including culture medium and vessel) should be identical to that used in the evolution of the test strains.

The following protocol is for the arabinose marker standard competition assays, in which for a given generation each of the 12 populations is competed against the reciprocally-marked ancestor with three-fold replication. In this experiment, populations were evolved in flasks containing 10 ml DM25 medium.

Strains:

REL606 (Ara^- ancestor), REL607 (Ara^+ ancestor), and 12 mixed-population samples.

Resources:

~80 TA plates, ~1.5 l of DM25, ~200 ml of LB.

Day -3

Label 14 flasks by strain. Add 9.9 ml of LB to each. Ancestral strains can be "looped" into appropriate flasks. Evolved populations must be allowed to thaw slightly, then use pipetter to remove 0.1 ml (i.e. usual transfer volume during the evolution experiment) of each into appropriate flasks.

Day -2 (preconditioning 1)

Label 72 flasks, 18 for each baseline strain and 3 for each evolved strain. Add 9.9 ml DM25 to each. Dilute each LB culture by 10^{-4} (via one dt) into 18 (baseline) or 3 (evolved) corresponding flasks. Each competitor measurement is propagated independently for each fitness competition.

Day -1 (preconditioning 2 – optional, but be consistent)

Label 72 flasks, 18 for each baseline strain and 3 for each evolved strain. Add 9.9 ml DM25 to each. Dilute each of the previous days DM25 cultures 10^{-2} into corresponding flask.

Day 0

Label 36 flasks, 3 for each evolved strain. Add 9.9 ml DM25 to each. Label 36 TA plates, also noting t=0. Pair-up yesterday's preconditioning flasks for the evolved populations and ancestral strains possessing the opposite Ara marker. Use each preconditioned strain only once. For each pair, mix 0.05 ml (i.e. 1:200 dilution) of each competitor into the corresponding fresh DM25 flask, then **immediately** sample at 5×10^{4} dilution (via one 100-fold dilution, then spread 0.05 ml) onto correctly labeled TA plate.

Day 1

Label 36 TA plates, also noting t=1. Sample each competition culture at 2×10^{5} (two times 100-fold dilutions, then spread 0.05 ml) onto correct TA plate. Count t=0 plates.

Next Day

Count t=1 plates.

Note: When competing strains that you anticipate will have small fitness differences, continuing the competition for 1 extra day can give better estimates. In this case, propagate a 100-fold dilution into fresh DM25 at the end of Day1. Plate a sample from this flask at the end of day 2.

Fitness estimates: We usually estimate the fitness of the test strain as its relative fitness (r) against the reference strain. Details at: http://myxo.css.msu.edu/ecoli/srvsrf.html

Competing 1 test strain to a reference strain – three replicates:

-80 freezer stocks
(red = ara- reference; white Ara+ test)

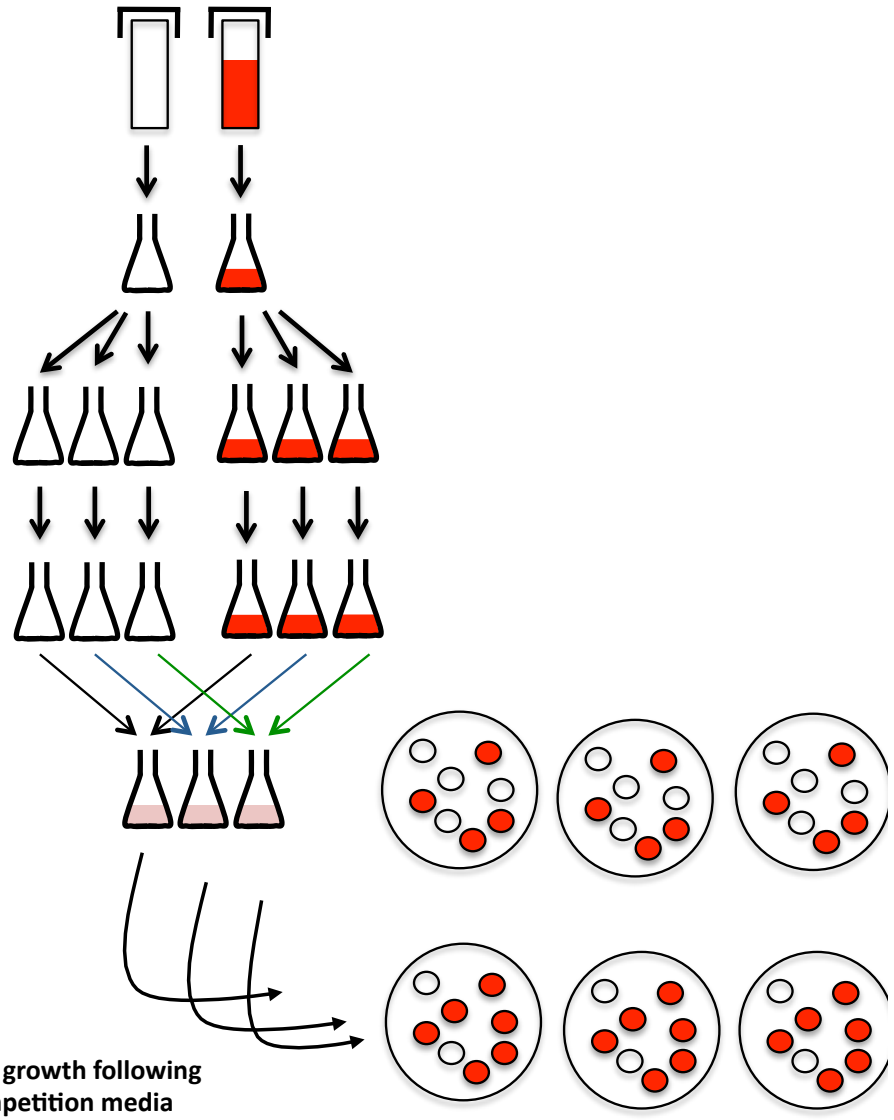
Day -3. LB inoculated with freezer stocks

Day -2. Preconditioning 1
(10^{-4} dilution* LB -> DM25)

Day -1. Preconditioning 2
(10^{-2} dilution DM25 -> DM25)

Day 0. Competition. Mix competitors
and plate $5x10^{-4}$ dilution** from this mix onto TA

Day 1. Competition. Plate $5x10^{-6}$ dilution** from
this mix onto TA



You will need to adjust dilution factors to *allow ~100-fold growth following transfer from LB (that supports ~$4x10^9$ cfu/ml) to your competition media and **to achieve ~200-400 colonies per plate if competition supports a density substantially different from DM25 (~$5x10^7$ cfu/ml)

#2 Fluorescent (e.g. GFP+) marker competitions.

Conditions are likely to be machine specific. Below is our protocol for discriminating between *E. coli* B REL606 (ancestor) cells and various evolved derivatives. We use an Accuri C6 flow cytometer.

We use P_L and P_{A1} (T7) promoters to drive expression of GFPmut3.1.

Dilute grown cells to around $5x10^6$ cfu/ml.

In our machine background noise is considerable at the size of bacterial cells. We can use a narrow gate to reduce noise but we find that the position of this gate often needs to be tweaked between runs (to maximise signal:noise). Considering also that the initial gate is usually made on cell size, and that evolved and ancestral cells can have quite different sizes so that the exact position of the gate will determine the frequency of the two types, we use an alternative approach where all cells are dyed to distinguish them from background.

Invitrogen has a Red dye sampler pack. Having gone through this we find that SYTO17 is the best dye for our needs. It readily stains live cells and produces a distinct peak in our 675 nm detector that allows cells to be distinguished from noise. A downside is that it (and all other red dyes we have tested) act to quench GFP and YFP signal. We therefore try and find a concentration of SYTO17 dye that allows cells to be distinguished while keeping (e.g.) GFP+ and GFP- cell peaks distinct in the appropriate detector channel. After a lot of trial and error, we find 150-200 nM of dye in water works best. It is possible the addition of DMSO (to increase cell permeability) or a pre-incubation of cells in a concentrated dye solution followed by suspension in water to lower the concentration, might produce better results in some situations. Use of a stronger/brighter GFP might also be useful, though changes in costs of expression would need to be calculated.