Directed Evolution and Genomic Analysis of Novel Yeast Species for More Efficient Biomass Conversion

Investigators

Gavin Sherlock, Assistant Professor, Genetics; R. Frank Rosenzweig, Associate Professor, Microbial Genetics; R. P. Levine, Professor (Emeritus), Biological Sciences; Barbara L. Dunn, Senior Research Scientist; Katja Schwartz, Research Associate.

Introduction

We propose to develop novel, adaptively evolved, hybrid yeast strains having phenotypes that enhance the efficiency of industrial-scale biomass conversion. Specifically, we propose to develop heat- and ethanol-tolerant yeast that convert both xylose and glucose to ethanol, studying key aspects of their functional and systems biology as they evolve. Initially, we will increase the amount of standing genetic variation upon which natural selection can act, by hybridizing species in the *Saccharomyces sensu stricto* group. Thereafter we will direct the evolution of rare, viable F2 hybrid progeny in chemostats, selecting for industrially desirable traits. Through successive rounds of crossbreeding and selection, we will be able to recombine these traits, without recourse to genetic modifications other than those wrought by natural selection. We will use DNA microarrays to assay both genome architectural and gene expression changes that occur during adaptive evolution. These studies will reveal the contribution of specific changes in DNA content and gene expression patterns to xylose fermentation and enhanced ethanol- and thermo-tolerance. These insights can be used to direct future efforts aimed at achieving still greater conversion efficiencies under industrially relevant conditions. Achieving this aim is essential to the national strategic goal of transforming a geology-based economy, supported by non-renewable resources, to a biology-based economy, supported by renewable resources that include forest and agricultural residuals. Life cycle analyses indicate that such an economy, centered on the concept of the integrated biorefinery, would have the added advantage of zero net greenhouse gas emission. The specific goals of this project are:

1. To create libraries of hybrid yeast strains which present large pools of standing genetic variation.

2. To select, through the application of evolutionary pressures, hybrid yeasts exhibiting phenotypes useful for the production of ethanol from plant biomass.

3. To design and test 60-mer oligonucleotide microarrays that contain probes for *S. cerevisiae* and one of the other *Saccharomyces sensu stricto* strains for use in array comparative genomic hybridization (a-CGH) and expression studies.

4. To perform a detailed comprehensive characterization of chromosomal rearrangements and gene expression changes that occur as a result of *de novo* speciation and adaptive evolution of these strains, and to characterize their stability.

Background

Introduction and Rationale

The US Department of Energy has identified the need to supplement transportation fuel supplies with sustainable, domestic resources as one of the nation's highest energy priorities (http://www.eere.energy.gov/biomass). Among the most promising strategies for attaining this goal is the integrated biorefinery, an industrial process based on the use of a flexible biomass feedstock to produce fuel, power, materials and chemical precursors for industrial syntheses. Implementation of this process would have the added benefits of reducing our reliance on foreign oil, invigorating rural economies through growth of the bioenergy industry and dramatically decreasing atmospheric accumulation of greenhouse gases. Life cycle analysis demonstrates that a renewable bioenergy economy would not only result in 100% carbon closure, but also bring about a dramatic reduction in the hydrocarbon precursors that lead to ozone formation [1]. Progress towards converting from a geology-based to a biology-based economy depends critically on improving the efficiency and flexibility of biomass conversion through improvements in both thermochemical and biochemical processes. Improvements in the latter require development of a microbial system capable of aggressively fermenting at elevated temperatures and ethanol concentrations the mixture of five- and six-carbon sugars that result from thermochemical and enzymatic pre-treatment.

Starch derived from corn is currently the preferred substrate for producing fuel-grade ethanol. However, to reduce cost and to meet the projected demand for this commodity, bioethanol producers will need to turn to cheaper lignocellulosic biomass feedstocks such as corn stover, corn fiber, rice straw, cereal straw, bagasse, forestry waste products, and possibly switch grass (National Renewable Energy Laboratory Sugar Platform Colloquies, NRL/SR-510-31970, May 2002). In order to utilize all of the substrate found in these cellulosic and hemicellulosic feedstocks it will be necessary to develop a microbial system can ferment both hexose and pentose sugars (NREL Colloquies on the Yeast Platform Project, NREL/SR-510-31690, February 2002).

Saccharomyces cerevisiae, because of its GRAS ("Generally Recognized As Safe") and model organism status, is the organism of choice for the development of such a system. While *S. cerevisiae* is able to ferment hexose sugars to ethanol, it normally lacks the ability to produce ethanol by the fermentation of xylose and L-arabinose, the pentose sugars present in hemicellulose. On the other hand, although it cannot utilize xylose as a sole carbon source, *S. cerevisiae* can catabolize this substrate, albeit slowly [2], as it possesses the genes that encode xylose reductase, xylitol dehydrogenase, and xylulokinase, the key steps in xylose metabolism [3-6]. Thus far, strains of *S. cerevisiae* capable of converting xylose to ethanol are transgenic, typically containing the XDH gene (encoding D-xylulose reductase/xylitol dehyrogenase) and the XR gene (encoding NAD(P)H-dependent xylose reductase) from the yeast *Pichia stipitis*. Strain TMB3001 [7] is one example of such a strain. Ethanol yield of TMB3001 has also been improved using evolutionary engineering, by growing mutagenized TMB3001 in chemostats for an extended period of time while slowly adapting from aerobic, to microaerobic, to anaerobic conditions [8]. Since genetically engineered strains (i.e., GMOs) may not be acceptable in certain markets and industrial settings (National Renewable Energy Laboratory Colloquies on the Yeast Platform Project, NREL/SR-510-31690, February

2002), we propose herein an alternative approach that makes use of classical genetics and is based on well-established evolutionary principles and techniques.

Xylose Metabolism

Xylose reductase (XR) catalyzes the first step of xylose utilization, the conversion of xylose to xylitol, which is then converted to D-xylulose by xylulose dehydrogenase (XDH). Significantly, *S. cerevisiae* has endogenous genes encoding XR (*GRE3*, *YPR1*, *YJR096W*) and XDH (*XYL2*) activities [5, 9, 10], and while their activities are too low to allow growth on xylose [10], overexpression of *GRE3* in the recombinant TMB3001 strain containing the *P. stipitis* XR and XDH leads to an increase in xylose consumption and ethanol production [10]. *S. cerevisiae* also has an endogenous xylulose kinase encoded by *XKS1* [3, 6], which converts D-xylulose to D-xylulose-5-P. This compound can then enter the pentose phosphate pathway eventually allowing for energy production and growth, and the evolution of CO_2 and ethanol. Endogenous transporter genes that are known to be able to import xylose into the cell are also present in *S. cerevisiae*. Thus wild-type *S. cerevisiae* contains all the necessary enzymes to allow growth on xylose as a sole carbon source; it appears that what prevents this from occurring is both the low activities of the initial enzymes as mentioned above, plus the accumulation of an unfavorable cell redox balance during xylose metabolism [11].

The Saccharomyces "sensu stricto" species: An ideal system for generating and characterizing genomic diversity.

The complete genomic sequences for four additional members of the *Saccharomyces 'sensu stricto'* species complex have recently become available [12, 13]. Reproductive isolation in these species is postzygotic, and haploid gametes from these species can mate with each other and form viable, but sterile zygotes [14-16]. Because a high number of spores can be screened after an inter-species mating, rare viable hybrid zygotes can be isolated and used for the real-time speciation of hybrid yeasts in the laboratory [17]. These rare viable spores are themselves postzygotically separated from each other, and span an enormous range of diversity. Furthermore, because they are created by the hybridization of two related, but distinct species, they are likely to be heterozygous for every locus for which there is a copy contributed by both species, and likely to exhibit heterosis, or hybrid vigor, and thus be good raw material for the selection of novel or enhanced traits.

Initial Results

Funding for this project became available in late Feb 2005, and for the subcontract in Montana in late April 2005, so there has been little progress thus far. The bulk of this report is instead directed towards our proposed Research Plan, as set out below. However, even in the short time that funding has been available to us, we have already begun to generate some data.

a) Establishment of standards for various sugars.

As part of our research, it will be important to be able to quantify the rate at which xylose is being catabolized, so that we will be able to determine the efficiency of its fermentation. We have been able to establish a protocol for generating xylose standard curves using GC-MS (Gas chromatography – mass spectrometry) from complex media.

In addition, we have done likewise for an intermediate in the pathway, xylitol, as well as some other sugars for comparison, glucose and L-arabinose.

b) Initial characterization of yeast species sugar utilization.

As detailed in the Research Plan below, there are several different yeast species within the *Saccharomyces sensu stricto* complex. Prior to generating and evolving hybrids, we have begun to characterize these strains with respect to their ability to utilize various 5-carbon sugars, by both plate and liquid assays, such that we can choose the appropriate partners for the hybridization that is most likely to maximize our liklihood of evolving strains that can efficiently ferment xylose. Preliminary studies indicate that *S. bayanus*, the most distantly related to *S. cerevisiae* within the *sensu stricto* complex, has a modest ability to grow on solid medium containing xylose as the carbon source. We are now characterizing this further using liquid media, with both *S. bayanus* and *S. bayanus – S. cerevisiae* hybrids. We view this result as extremely promising with respect to our proposed research.

c) Initial design and characterization of two-species microarrays.

As detailed in the research plan below, we have used the ArrayOligoSelector software to design custom microarrays capable of detecting DNA from yeast hybrids. Initially, we have designed an array capable of detecting both *S. cerevisiae* and *S. bayanus* DNA, containing just over 5000 probes for each genome. Our initial characterization of the array (see Figure 1 below) suggests that our design strategy will allow us to discriminate well between the DNA in the hybrid contributed from each parent.
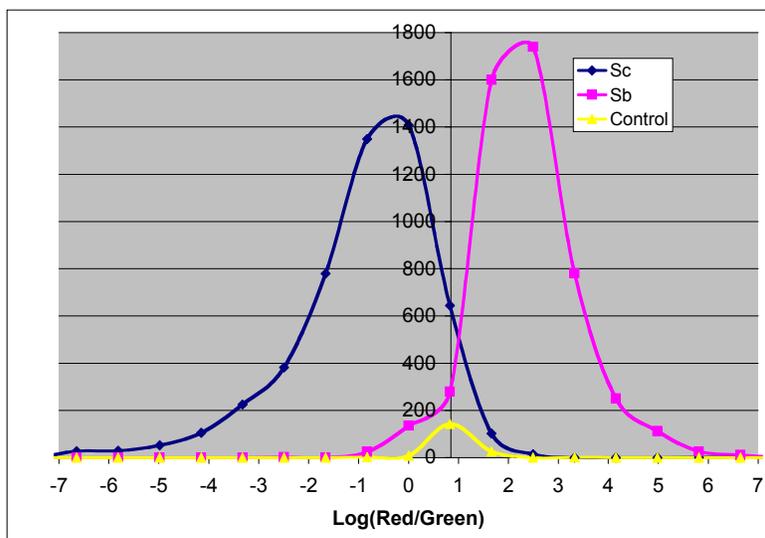
Figure 1 Dual species arrays allow us to discriminate between DNA from the two genomes. *S. cerevisiae* DNA was labeled with green (blue curve) while *S.bayanus* DNA was labeled with red (magenta curve). We see a clear discrimination between the two sets of probes.

Research Plan

This proposal seeks to use natural selection to generate novel yeast species that can efficiently convert xylose from plant biomass to ethanol at elevated temperatures and ethanol concentrations, as described below. We recognize that thermochemical pretreatment of hemicellulose results in a mixture of xylose and arabinose. However, because xylose is by far the dominant constituent and because *Saccharomyces spp.* already contain within their genomes the evolutionary raw material for xylose fermentation, we will concentrate our efforts on directing the evolution of an efficient xylose catabolic pathway.

Goal 1 - *To create genetically diverse pools of hybrid yeast strains.*

Through ongoing collaborations with the laboratory of Ed Louis we have access to parental allodiploid hybrid yeasts that they have generated between the sequenced laboratory strain of *S. cerevisiae* (S288c) and sequenced strains of the other *Saccharomyces sensu stricto* species. We will assay each hybrid for the phenotypes we intend to evolve – thermotolerance, ability to grow at high concentrations of ethanol, and the ability to grow on xylose. While we do not expect any of these initial hybrids to grow well on xylose (though see *Preliminary Results*) it is an obvious and easy experiment that we should carry out. We will select the hybrid combination that has the most favorable characteristics, based on these phenotypes, to be used as the basis for all of our further experiments.

We will isolate rare viable spores from the hybrid, which rapidly autofertilize to form diploid F2 hybrids. We will generate as large a collection of independent yeast hybrids as is practical, and pool them. These will serve as our initial library stock for subsequent experiments, and will be frozen in glycerol. For our first round of evolutionary selections (see below), we will either use cells directly from this stock, or cells from this stock which have been mutagenized to increase genetic variability, using EMS mutagenesis. Thus as the starting point for all our initial experiments, we will use a genetically diverse pool of yeast hybrids, which may or may not have been subject to mutagenesis (see below for when these will be used).

Goal 2 - *To select, through the application of evolutionary pressures, hybrid yeasts exhibiting desirable phenotypes that can be applied to the production of ethanol from plant biomass*

The power of combining experimental evolution with genomics. By combining experimental evolutionary studies with genomic analyses we will be able: *(a)* to select novel strains with desirable phenotypes from an extremely diverse and heterozygous population of yeast hybrids, *(b)* establish the genomic and transcriptomic reprogramming required to evolve these phenotypes, and *(c)* determine the stability of both the karyotypes and phenotypes of these novel strains.

We will use chemostats to select for desirable traits with respect to energy production using novel hybrid species as the raw material. The general experimental design is: A large pool of genetically variable, hybrid F2 yeast will be used to inoculate chemostat cultures. These hybrid pools will be propagated vegetatively in continuous chemostat culture for ca. 320 generations, initially under selection for one phenotype (growth at

high temperature, growth on xylose, or growth in high ethanol). Having evolved phenotypes individually, we will mate the evolved strains with each other in all pairwise combinations (e.g., xylose fermenters with high-temperature glucose fermenters). From each of these new hybrids, we will again select for rare viable spores, and use these to inoculate a second round of chemostat cultures, simultaneously selecting for two of the desired traits. Finally, we will mate each hybrid possessing two desired traits with a hybrid containing the remaining trait, and from each of these new strains we will again generate large populations of spores. These will be subjected to a third round of directed evolution, selecting for all three traits. This will require at least 48 chemostat experiments, with three experiments per selection regime (*glucose/glucose+sorbitol, aerobic/anaerobic conditions*], double and triple phenotype). Since one 320 generation chemostat experiment requires ~80 days, 48 such experiments would require more than 10 years! However, because we can run 6 experiments in parallel all selections can be easily carried out within the time frame of the project. Of course, there will be downtime between rounds of selection, and we may have to repeat selections using mutagenized cells if we are unable to evolve the desired traits from non-mutagenized strains. Nevertheless, we are confident that we will be able to run all of the chemostat experiments within the first two years of this project, after which we will be able to characterize the results of the selections in great detail.

Operation and monitoring of nutrient limited-chemostats. Selection experiments will be conducted using ATR SixFors (http://www.atrbiotech.com), a fully integrated, computer-controlled culture system consisting of six independently controlled bioreactors that can be run in batch, fed-batch or continuous mode at constant temperature and pH. We will operate a working volume of 300mL at dilution rate of 0.17 h^{-1}. Aerobic cultures will be sparged with sterile, humidified house air, anaerobic cultures with sterile, humidified high-purity N_2 gas.

On a daily basis (approx. every 4 Generations) we will estimate: *(a)* Generations elapsed (deduced from $Vol_{effluent}$ and $Vol_{working}$); *(b)* Culture purity (by light microscopy and plating); *(c)* Culture density (by $O.D._{600nm}$, and haemocytometry); *(d)* Frequency of a neutral marker, *e.g.,* cyclohexamide resistance, by plating cells of known dilution on selective and non-selective plates. Every day we will archive cells from three 15mL population samples, two as pellets at -80^0C, and one as a glycerol stock at -80^0C, which will enable us to conduct "molecular archaeology," and, in principle, identify the genetic alteration underlying a specific adaptive change. Also, should a given culture become contaminated, archiving makes it possible "re-wind the evolutionary tape" to a point prior to the onset of contamination and resume the experiment. The pelleted samples can be used for DNA isolation. By monitoring generation time and the frequency of neutral markers, we will be able to observe when adaptive sweeps have occurred.

On a weekly basis (approx. every 30 Gen^s), we will evaluate key features of the evolving populations' physiology that we expect to change as cells adapt to their respective environments. Specifically, we will measure the residual substrate concentration of key metabolites in cell-free filtrate taken directly from the chemostats. Fermentation substrates (e.g., glucose and xylose) and products (e.g., ethanol, glycerol, and acetate) will be assayed either enzymatically, or by GC-MS as appropriate.

On a monthly basis (approx. every 120 Gen[s]) we will take clones isolated from each experiment as CFUs on non-selective agar, prepare them for Pulsed-field gel electrophoresis (PFGE), and establish karyotypes for evolutionary intermediates in the adaptive process. The results of these analyses will determine our choice of strains to assay for array-CGH, as described in Goal 4 below. For experiments described in **IV.B3,** we will estimate specific activities of xylose reductase, xylitol dehydrogenase and xylulose kinase in yeast cell extracts using established spectrophotometric methods [7, 21]; for experiments described in **IV.B4** and **IV.B5**, respectively, we will assay sensitivity of the evolved strains to heat- and ethanol-shock using standard techniques.

In summary, **for each experiment, we will establish a record of: (1) adaptive change** inferred from fluctuation in the frequency of neutral markers, **(2) phenotypic change** deduced from: *(i)* residual metabolite concentrations, *(ii)* specific activity of xylose catabolic enzymes, *(iii)* heat and ethanol sensitivities, and **(3) genomic change** evident in PFGE-derived karyotypes and array-CGH profiles. In so doing, we will be able to produce strong correlative evidence for the adaptive value of specific chromosome rearrangements in yeast evolving under specific selective pressures.

Phenotype 1: **Selection of hybrids capable of metabolizing xylose to ethanol.**
Wild-type *S. cerevisiae* is unable to grow on the pentose sugar xylose as a sole carbon source. Thus we face a major challenge in our goal to direct the evolution of hybrid yeast that can ferment xylose to ethanol without recourse to transgenics. We should therefore ask: 'Is there sufficient raw material upon which evolution can act to produce this desirable trait?' It appears that the answer is a qualified "yes." Not only does *S. cerevisiae* contain the 3 endogenous XR genes and the one XDH gene, but in addition, the *S. cerevisiae* genes *SOR1* and *SOR2* are both 54% identical (66% similar) to XDH over their entire lengths. *SOR1* is induced by sorbitol, and encodes sorbitol dehydrogenase, which converts sorbitol to sorbose. The *P. stipitis* XR gene is 57% identical, and 72% similar to the *GRE3* gene product, which is induced by starvation stress [22], and therefore likely to be induced under the near-starvation conditions of the chemostat. The fact that at least 3 genes exist in *S. cerevisiae* with high homology to the *P. stipitis* XR genes, and another 3 exist with high homology to the *P. stipitis* XDH genes, raises the possibility that they may be mutated to acquire more *Pichia*-like properties (e.g., NAD(P)H-dependent xylose reductase activity in *P. stipitis* appears to confer a more favorable cell redox balance), or to have higher activity, either of which would allow selection in a chemostat for growth on xylose as a sole carbon source.

Of all the traits that we wish to select for within hybrid *Saccharomyces* species, conversion of xylose to ethanol is potentially the most difficult to derive. While *S. cerevisiae* can utilize D-xylulose (poorly), it is unable to grow on D-xylose [23], although it is reported to metabolize this substrate with slow kinetics [2]. These observations explain why xylose-fermenting *S. cerevisiae* have been engineered by integrating the XDH and XR genes from *Pichia stipitis*, as described above. We propose an alternative strategy based on directed evolution of large, genetically diverse input populations. Chemostat experiments selecting for yeast that ferment xylose will be initiated using libraries of hybrids, derived from either non-mutagenized or EMS-mutagenized pools. Evolutionary experiments will be conducted under both aerobic and anaerobic conditions.

The use of different pools and culture conditions should maximize our chances of creating strains having the desired phenotype. For each of the combinations we will adopt two approaches:

(a) *Gradual replacement of carbon source* - We will run chemostat experiments wherein we gradually apply a selective pressure to favor hybrid yeasts capable of exploiting xylose as a carbon source by gradually replacing a substrate that they can use (such as glucose) with xylose. At the onset of such experiments, glucose will be provided at concentrations just below limiting, so that the addition of xylose will select for clones capable of using xylose for growth. Over the course of the experiment, we will incrementally replace glucose with xylose, thus increasing the selective pressure on the population, until xylose is the only sugar left in the medium. We will adjust the relative concentrations of these sugars only after we know that an adaptive sweep has occurred. Should this strategy prove unproductive, we will explore the use of media containing the non-metabolizable sugar sorbitol. Sorbitol acts as an inducer of the *SOR1* and *SOR2* genes, and because of their homology to *XDH*, they may be co-opted to function as *bona fide* xylulose reductases.

(b) *Immediate selection for growth on xylose* - The gradual replacement of glucose (or some other sugar) with xylose, should select for strains that are able to grow on xylose, but may also select for strains simply better able to grow on the primary substrate. Indeed, this selective pressure may be stronger than the selection for xylose metabolism, and lead us into an evolutionary dead end. Therefore, we will also attempt to select for clones that are able to grow on xylose as sole carbon source. While this strategy may be unlikely to succeed (though see Preliminary Results), it will be immediately obvious if it has failed, and so will be a low-risk, high-payoff activity.

In the event that we isolate a strain or strains capable of using xylose, but only from aerobic conditions, we will then apply a second selection, to select for a variant capable of using xylose under anaerobic conditions, which should lead to the production of ethanol. We will use the approach described by Sonderegger and Sauer [8] who subjected yeast cultures successively to aerobic, microaerobic and anaerobic conditions, to select for strains that produce ethanol from xylose. Again, we will only switch conditions after we know that one or more adaptive sweeps has occurred.

Phenotype 2: Selection of hybrids capable of growing at high temperature. We will select for strains able to grow at higher temperatures than that at which yeast normally grows, as such a strain could be expected to increase the efficiency of ethanol production and other industrial processes. We will culture the starting hybrid pool in anaerobic, glucose-limited chemostats, and incrementally increase culture temperature over the course of each experiment. Experiments will commence at 30°C and continue to a 45°C endpoint, or for 320 generations, whichever comes sooner. Because we will have taken frequent samples, there is no danger of us losing strains evolved during this process should there be a catastrophic decrease in fitness when the temperature is incremented. Additionally, should a strain growing at, for instance 36°C, grow much faster at that temperature than does a strain evolved after an adaptive sweep at 38°C, we will still have access to the earlier strain, which probably has a more desirable phenotype.

Phenotype 3: Selection of hybrids capable of growth in high ethanol. While ethanol is well known to inhibit microbial growth, some strains of *S. cerevisiae* have adapted to high concentrations of ethanol [24, 25]. Yeasts that could ferment sugars under these conditions offer obvious advantages for industrial biofuel production, as they would enable more ethanol to be harvested from large batch cultures. We will select for ethanol-tolerant yeast by placing hybrid pools in anaerobic, glucose-limited chemostats, then incrementally adjusting influent media ethanol concentration. We will increase ethanol concentration up to 15% ethanol, 320 generations, or until no more adaptive sweeps are observed, whichever comes first.

Selection of Multiple Phenotypes. Once we have selected strains that individually possess one of the three desired phenotypes, we will use these strains to generate new hybrids. The auto-fertilized diploid hybrids that we initially place under selection readily produce viable spores [17], amenable to another round of mating. These new hybrids will then be sporulated, and as large a pool of viable spores, from the rare ones that exist, will be used in the next round of chemostat experiments to select for two phenotypes simultaneously. In this case, we know that the genetic material that has gone into these crosses can support growth under each of the two conditions separately. Unless the two traits are somehow antagonistic, we expect some spores to possess genotypes that confer both phenotypes. Hence, we will begin using as selective conditions the endpoints from which the individual strains were selected, thereby selecting first for those desired spores, and then further evolving them. It will be readily apparent if this has failed, as cells will wash out. Should this occur we will back off from these conditions, e.g., by lowering the temperature, and/or the ethanol concentration, and then slowly bring them back up to select for strains having both desired phenotypes. Once we have generated all possible double phenotypes, we will use a similar strategy to obtain strains possessing all three phenotypes, using all three possible pairwise combinations of strains developed during the second round of selection.

Goal 3 - *To design and test 60-mer oligonucleotide microarrays that contain probes for Saccharomyces cerevisiae and one of the other Saccharomyces sensu stricto strains for array CGH and expression studies.*

Array design. We will design 60-mer DNA oligomer arrays to study genome rearrangements and gene expression changes in evolved hybrid yeasts. These arrays will be purchased from Agilent Technologies (Palo Alto, CA, USA) with approximately 11,000 custom designed oligonucleotides on them, 5,500 from each species. Having two species on a single chip will enable us to assay the DNA content and the transcriptome of the hybrid contributed by both species simultaneously, conserving reagents and minimizing experimental variation. Using ArrayOligoSelector (http://arrayoligosel.sourceforge.net/), we will design the oligos to avoid cross hybridization, both within and between the two species whose genomes are represented on the array. We will randomize the placement of the oligonucleotides across the array such that probes from adjacent chromosomal locations will not be adjacent on the array, to avoid/detect spatial hybridization biases on the microarray. Likewise, oligonucleotides from the two species will be randomly interspersed across the array. As a filter for the oligonucleotide design process, we will use the combined sequence of both genomes, which the software uses to assess cross hybridization potential. We expect to design

about 6000 oligonucleotides per organism, though some weeding out of oligos, due to space considerations, will have to be done. This will not greatly affect our ability to comprehensively carry out array-CGH, as we will still have coverage of the entire genome, but we will lose the ability to assay some transcripts. While less than ideal, this does halve the cost of the microarrays for a small loss in coverage. When deciding which genes to omit, we will ensure that all genes encoding enzymes are represented, so we will be able to get an accurate picture of changes in metabolism during our evolutionary experiments.

Initial Array Characterization. To test our arrays for cross hybridization problems we will label and separately hybridize DNA from each species, and determine whether probes that were designed to assay DNA from the other species show significant signal. Probes showing significant cross hybridization will be eliminated from further analyses. These experiments will also indicate which probes show good signal for the species for which they were designed. Because we are able to order custom oligonucleotides arrays in small batches, we can refine our designs on subsequent iterations, to eliminate and/or replace probes that do not have the characteristics that we desire. For array CGH, it is most natural to compare an experimental sample to a sample with a known genomic complement. In the case of hybrid yeasts, it is thus obvious to isolate genomic DNA from both strains that were fused to form the hybrid, and mix equal genomic amounts of their DNA for use as a reference sample. Since the size of each contributing yeast's genome is known, we can simply calculate a ratio between the size of *S. cerevisiae*'s genome and the size of the genome of the *Saccharomyces* species to which it has been fused, and use this ratio such that we can combine genomic DNA from the two yeasts to give a resulting mixture that is 1:1 in terms of genome copies, rather than by mass. This mixture will then be used as the reference sample.

Experimental Design for expression microarray experiments. We will wish to assay transcription in the hybrid strain with respect to that of the original parental strain, to understand the changes that have occurred to the levels of various enzymes involved in metabolism. Such comparisons are non-trivial, because it is likely that we will be growing the evolved strains in radically different conditions than can be used to grow the parental strain. For instance, assuming that we have isolated strains capable of growing with xylose as the sole sugar, we will not be able to grow the parent in xylose for a straight comparison to the parent's transcriptome. Furthermore, there may be many genes expressed in the derived strains that are not expressed in the parent strain – therefore using RNA derived from the parent in the control channel for the expression arrays, which would be used to derive the denominator for expression ratios would be a poor strategy. An alternative strategy would be to use a common reference for the control sample, to which all samples could be compared. We could derive a common reference from all of our conditions (parental and derived strains grown in their various conditions) so that all genes expressed in any of the conditions would be represented in the reference. However, this would mean that we would be unable to perform our array hybridizations until the end of all experiments, when all samples would be in hand to create the reference. Since it would be much more effective to have an assessment of our evolved strains as we evolve them, we will instead use genomic DNA of the parental strain as a common reference for our expression studies. This means that we will

compare all expression samples to the same robust reference, which is guaranteed to have sequences capable of hybridizing to every spot.

Data Normalization and Controls. To normalize the signals between reference and experimental samples, we will use the methods of van Peppel *et al* [26], whereby several different control oligonucleotides are placed in duplicate throughout the array. From these control spots, a lowess curve will be derived, and this curve will then be used to normalize the data, as described [26].

For array CGH experiments, we will have to control for differing amounts of DNA in our strains. We will use FACS analysis to determine the ratio of DNA in a hybrid cell compared to the amount of DNA in each of the yeast cells from which the hybrid was derived. In the labeling reactions, we will use the same number of genomes worth of DNA from the hybrid as is used from the individual yeasts making up the reference sample. For example, if we consider the contributing strains as having 2 genomes worth of DNA combined, and the hybrid has 1.9 genomes worth of DNA in a G1 cell (e.g. there has been some non-reciprocal translocations) then we would use a 2:1.9 ratio of input DNA into the reference and experimental labeling reactions respectively. This will work irrespective of whether the cells from one culture of cells are predominantly in G1, and the cells from the other culture have a large number of M phase cells, as the ratio of DNA will be that of G1 cells. Thus our spiked-in controls would be used only to compensate for biases introduced in the labeling reaction. We will easily test this method of data normalization, using control strains containing known aneuploidies, as well as tetraploid vs. diploid, and diploid vs. haploid strains of *S. cerevisiae*, where we know every ratio should be two.

Array CGH Data Analysis. While browsing of the data graphically allows us to identify regions of aneuploidy, given the number of microarrays from which we expect to collect data, simple browsing will be impractical as a systematic analysis method. Thus, we will use the recently developed CLAC software [27] that implements a statistically robust algorithm for automated calling of gains and losses in array-CGH data. This software will allow us to rapidly process hundreds of microarrays worth of data to determine with statistical confidence (by setting an acceptable False Discovery Rate) the regions within each genome represented on a given microarray for copy number variation. An additional feature of the CLAC software, that is invaluable for our studies, is the ability to determine, based on data from multiple hybridizations, whether there are recurrent regions of loss or gain. We will analyze the data for a given species' genome for recurrent alterations both for different spores from a given hybrid, as well as across all hybrids. Identification of recurrent changes occurring while under selective pressure will be strongly indicative of restrictions to the possible evolutionary trajectories, as has already been seen in glucose limitation for non-hybrid strains . An additional layer of complexity that we can assay is whether syntenic regions of the two source genomes that make the hybrid are involved in reciprocal events. Thus we will determine, for both hybrid or ancestral strains, whether certain rearrangements predominate under certain selective conditions

Expression Data Analysis. The expression data will allow us to determine, how various 'modules of metabolism' are transcriptionally remodeled during our evolutionary selection experiments. We have already shown that various elements of metabolism are altered at the transcriptional level, such as glycolysis and the TCA cycle [19], and recently it has been shown that the majority of the metabolic network shows transcriptional regulation of some kind [28]. Thus, it is our expectation that some of the changes in metabolism that will occur during our evolutionary selection will either result from, or result in (due to feedback mechanisms) transcriptional changes in various metabolic genes. To understand how metabolism might be reprogrammed, we will 'paint' the expression data onto metabolic pathways using the freely available GenMapp software [29]. In addition, we will use hierarchical clustering (e.g. [30]) to elucidate clusters of genes that have shown similar alterations in their expression, and then analyze these groups of genes in the context of their Gene Ontology annotations [31], using GO::TermFinder [32]. GO::TermFinder is able to determine for a list of genes whether there is an overrepresentation of Gene Ontology terms compared to what would be expected by chance, and will enable us to easily determine biological themes from our clusters of genes, using the extensive GO annotations that are available for yeast genes from the *Saccharomyces* Genome Database [33]. These approaches are complementary, because GO::TermFinder may well help us pinpoint the processes that are affected, for subsequent visualization with GenMapp.

Goal 4 - *To perform a detailed characterization, on a whole-genome scale, of chromosomal rearrangements and expression changes that occur as a result of de novo speciation and selection of hybrid yeasts formed between Saccharomyces sensu stricto strains, and to characterize their stability. These studies will yield fundamental insight into the changes in genome architecture and global transcription that underlie evolutionary adaptation to xylose fermentation, and growth under conditions of elevated ethanol concentration and high temperature.*

For each evolutionary selection, we will have frozen daily samples at -80°C. Because we will have monitored both the generation time of the cells and the frequency of neutral markers at each point in the experiment, we will know with a great degree of accuracy when each adaptive sweep has occurred. Thus, we will be able to retroactively examine both the DNA and RNA content of a cell population immediately prior to and following an adaptive sweep. We expect to observe between 3 and 6 adaptive sweeps during the course of any ~320 generation individual experiment. For each successful chemostat experiment, we will therefore have between 12 and 28 samples (3 to 7 adaptive sweeps x 2 (for preceding and following samples) x 2 (for RNA and DNA)). We anticipate running approximately 3 selection regimes per every 6 – 9 months, with 3 rounds of selection (single, double and triple phenotypes), which means for comprehensive microarray characterization of all the evolved strains, we will need between 108 and 252 microarrays over the 3 years.

After we have selected strains with all three desirable traits, we will use microarrays to characterize their phenotypic (*i.e.,* transcriptional) and karyotypic stability. There are at least two circumstances in an industrial setting where it is important for these strains to be stable. First, strains will need to be propagated under standard laboratory conditions in order to produce large populations for inocula. Typically this would be carried out as a

short-term batch culture. Second, strains must be stable over the course of prolonged industrial fermentations. Thus, we will propagate evolved strains in batch cultures for a short time (10 generations) under standard laboratory culture conditions (YPD medium at 30°C), as well as in extended (50 generations) fed-batch cultures that mimic industrial fermentations. In both cases, samples will be taken at 5 generation intervals and assayed for their DNA content and RNA expression profiles. We will analyze these data to determine the genomic and transcriptomic stability of our evolved strains.

References
1. Sheehan, J., A. Aden, K. Paustian, K. Killian, J. Brenner, M. Walsh, and R. Nelson, *Energy and Environmental Aspects of Using Corn Stover for Fuel Ethanol.* Journal of Industrial Ecology, 2003. **7**: p. 117-146.
2. van Zyl, C., B.A. Prior, S.G. Kilian, and J.L. Kock, *D-xylose utilization by Saccharomyces cerevisiae.* J Gen Microbiol, 1989. **135 (Pt 11)**: p. 2791-8.
3. Ho, N.W.Y. and S.F. Chang, *Cloning of yeast xylulokinase by complementation of Escherichia and yeast mutations.* Enzyme and Microbial Technology, 1989. **11**(417-421).
4. Richard, P., M.H. Toivari, and M. Penttila, *The role of xylulokinase in Saccharomyces cerevisiae xylulose catabolism.* FEMS Microbiol Lett, 2000. **190**(1): p. 39-43.
5. Richard, P., M.H. Toivari, and M. Penttila, *Evidence that the gene YLR070c of Saccharomyces cerevisiae encodes a xylitol dehydrogenase.* FEBS Lett, 1999. **457**(1): p. 135-8.
6. Batt, C.A., S. Carvallo, D.D. Easson, M. Akedo, and A.J. Sinkskey, *Direct evidence for a xylose metabolic pathway in Saccharomyces cerevisiae.* Biotechnology and Bioengineering, 1986. **28**(549-553).
7. Eliasson, A., C. Christensson, C.F. Wahlbom, and B. Hahn-Hagerdal, *Anaerobic xylose fermentation by recombinant Saccharomyces cerevisiae carrying XYL1, XYL2, and XKS1 in mineral medium chemostat cultures.* Appl Environ Microbiol, 2000. **66**(8): p. 3381-6.
8. Sonderegger, M. and U. Sauer, *Evolutionary engineering of Saccharomyces cerevisiae for anaerobic growth on xylose.* Appl Environ Microbiol, 2003. **69**(4): p. 1990-8.
9. Traff, K.L., L.J. Jonsson, and B. Hahn-Hagerdal, *Putative xylose and arabinose reductases in Saccharomyces cerevisiae.* Yeast, 2002. **19**(14): p. 1233-41.
10. Traff-Bjerre, K.L., M. Jeppsson, B. Hahn-Hagerdal, and M.F. Gorwa-Grauslund, *Endogenous NADPH-dependent aldose reductase activity influences product formation during xylose consumption in recombinant Saccharomyces cerevisiae.* Yeast, 2004. **21**(2): p. 141-50.
11. Verho, R., J. Londesborough, M. Penttila, and P. Richard, *Engineering redox cofactor regeneration for improved pentose fermentation in Saccharomyces cerevisiae.* Appl Environ Microbiol, 2003. **69**(10): p. 5892-7.
12. Cliften, P., P. Sudarsanam, A. Desikan, L. Fulton, B. Fulton, J. Majors, R. Waterston, B.A. Cohen, and M. Johnston, *Finding functional features in Saccharomyces genomes by phylogenetic footprinting.* Science, 2003. **301**(5629): p. 71-6.
13. Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E.S. Lander, *Sequencing and comparison of yeast species to identify genes and regulatory elements.* Nature, 2003. **423**(6937): p. 241-54.
14. de Barros Lopes, M., J.R. Bellon, N.J. Shirley, and P.F. Ganter, *Evidence for multiple interspecific hybridization in Saccharomyces sensu stricto species.* FEM Yeast Res, 2002. **1**(4): p. 323-31.
15. Greig, D., R.H. Borts, E.J. Louis, and M. Travisano, *Epistasis and hybrid sterility in Saccharomyces.* Proc R Soc Lond B Biol Sci, 2002. **269**(1496): p. 1167-71.
16. Masneuf, I., J. Hansen, C. Groth, J. Piskur, and D. Dubourdieu, *New hybrids between Saccharomyces sensu stricto yeast species found among wine and cider production strains.* Appl Environ Microbiol, 1998. **64**(10): p. 3887-92.
17. Greig, D., E.J. Louis, R.H. Borts, and M. Travisano, *Hybrid speciation in experimental populations of yeast.* Science, 2002. **298**(5599): p. 1773-5.
18. Brown, C.J., K.M. Todd, and R.F. Rosenzweig, *Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment.* Mol Biol Evol, 1998. **15**(8): p. 931-42.

19. Ferea, T.L., D. Botstein, P.O. Brown, and R.F. Rosenzweig, *Systematic changes in gene expression patterns following adaptive evolution in yeast.* Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9721-6.

20. Dunham, M.J., H. Badrane, T. Ferea, J. Adams, P.O. Brown, F. Rosenzweig, and D. Botstein, *Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2002. **99**(25): p. 16144-9.

21. Bergmeyer, H.U., *Methods of Enzymatic Analysis.* 1983, Deerfield Beach, Florida: Verlag Chemie.

22. Garay-Arroyo, A. and A.A. Covarrubias, *Three genes whose expression is induced by stress in Saccharomyces cerevisiae.* Yeast, 1999. **15**(10A): p. 879-92.

23. Hsiao, H.Y., L.C. Chiang, P.P. Ueng, and G.T. Tsao, *Sequential utilization of mixed monosaccharides by yeasts.* Appl Environ Microbiol, 1982. **43**(4): p. 840-5.

24. Ghareib, M., K.A. Youssef, and A.A. Khalil, *Ethanol tolerance of Saccharomyces cerevisiae and its relationship to lipid content and composition.* Folia Microbiol., 1988. **33**: p. 447-452.

25. Alexandre, H., I. Rousseaux, and C. Charpentier, *Ethanol adaptation mechanisms in Saccharomyces cerevisiae.* Biotechnol Appl Biochem, 1994. **20 (Pt 2)**: p. 173-83.

26. van de Peppel, J., P. Kemmeren, H. van Bakel, M. Radonjic, D. van Leenen, and F.C. Holstege, *Monitoring global messenger RNA changes in externally controlled microarray experiments.* EMBO Rep, 2003. **4**(4): p. 387-93.

27. Wang, P., Y. Kim, J. Pollack, B. Narasimhan, and R. Tibshirani, *A method for calling gains and losses in array CGH data.* Biostatistics, 2004. **in press**.

28. Ihmels, J., R. Levy, and N. Barkai, *Principles of transcriptional control in the metabolic network of Saccharomyces cerevisiae.* Nat Biotechnol, 2004. **22**(1): p. 86-92.

29. Dahlquist, K.D., N. Salomonis, K. Vranizan, S.C. Lawlor, and B.R. Conklin, *GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways.* Nat Genet, 2002. **31**(1): p. 19-20.

30. Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein, *Cluster analysis and display of genome-wide expression patterns.* Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14863-8.

31. Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock, *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

32. Boyle, E.I., S. Weng, J. Gollub, H. Jin, D. Botstein, J.M. Cherry, and G. Sherlock, *GO::TermFinder - open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes.* Bioinformatics, 2004.

33. Dwight, S.S., M.A. Harris, K. Dolinski, C.A. Ball, G. Binkley, K.R. Christie, D.G. Fisk, L. Issel-Tarver, M. Schroeder, G. Sherlock, A. Sethuraman, S. Weng, D. Botstein, and J.M. Cherry, *Saccharomyces Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO).* Nucleic Acids Res, 2002. **30**(1): p. 69-72.

Contacts

Gavin Sherlock: sherlock@genome.stanford.edu
Frank Rosenzweig: frank.rosenzweig@mso.umt.edu