LABORATORY OF MOLECULAR BIOLOGY UNIVERSITY OF WISCONSIN-MADISON

R.M. Bock Laboratories
1525 Linden Drive
Madison, WI
53706-1596

Office: 608-262-1586
CMB: 608-262-3203
FAX: 608-262-4570
www.molbio.wisc.edu

To the Systems Biology/Microbiology Search Committee,

Please accept the attached CV, statement of research interests, teaching philosophy, and representative publications in response to the **faculty position in systems biology/microbiology** job advertisement posted in the September 2nd issue of the journal Science.

I am extremely interested in both the department of biology and the biocomplexity institute at Indiana University and am confident that my interests in the evolution of gene function in yeast will complement the research programs of your current faculty in the department.

Thank you for your consideration. Please do not hesitate to contact me with any questions or concerns by phone or e-mail.

Sincerely,

Barry Williams
Laboratory of Molecular Biology and
Howard Hughes Medical Institute
University of Wisconsin
1525 Linden Drive
Madison, WI 53706
608-824-8690
bwilliams2@wisc.edu

Understanding the functional meaning of protein sequence evolution

Barry Williams

Despite the wealth of DNA sequence data and sophisticated computational methods of describing patterns of sequence variation, there remains a fundamental mystery as to the functional significance of the conservation and divergence of protein sequences. While sequence comparisons are routine in modern biology, there is little empirical data demonstrating the forces that shape variation, conservation, and divergence. What are the fitness effects of substitutions segregating within species and fixed between species? Do substitutions with fitness effects correspond to functional evolution? How pervasive is compensatory mutation and epistasis among substitutions? What is the structural and molecular/biochemical nature of mutations that alter function versus compensatory substitutions? What types of substitutions are tolerated at a given gene?

In order to pursue these questions I have developed a research platform using baker's yeast (*Saccharomyces cerevisiae*) that enables me to engineer any genetic alteration to the genome and to measure the most precise fitness estimates possible for these genetic changes. Precise replacement of orthologous genes from closely related species has revealed that genetic changes fixed between species, while not altering gene function, carry large fitness effects due to virtually ubiquitous epistasis. Unexpectedly, the fitness effects of the same replacement can be either positive or negative, depending on the environment. This indicates that genomes evolve adaptively, but not for functional diversification, instead protein evolution is largely a process of compensatory mutations that buffer organisms to maintain proper cellular processes and biochemical interactions. These findings provide the empirical evidence to support a new forefront in understanding the meaning of adaptive evolution on a genome-wide scale. To assess the degree to which the biology of *S. cerevisiae* is a product of evolutionary history versus domestication, and to guide the selection of genes with interesting evolutionary patterns for further gene replacements experiments, I have generated both phylogenomic and population genomic data sets from wild yeast populations. Phylogenomic studies resulted in the most robust phylogeny among a previously unknown group of organisms, and produced a data set that has rapidly become one the most thoroughly studied in all of molecular evolution. Population genomic analyses have shown natural populations of *S. cerevisiae* to be non-domesticated, moderately sized, and sexually out-crossing with evidence for genome-wide selection. Hence, the biology of this organism is still determined by its interactions in the wild and highly similar to that of most eukaryotes, so the evolutionary genetics of yeast can elucidate biological processes in other organisms and I now have the necessary evolutionary genetic framework to do so.

Postdoctoral research

Upon arriving at the Carroll lab, I transitioned from work primarily in population and conservation genetics (leading to protection of a rare butterfly species (Williams 2002, Williams et. al. 2003a)), to characterizing the molecular genetic basis of microevolutionary change in morphology. In collaboration with Trisha Wittkopp, I successfully elucidated genetic factors responsible for divergence in pigmentation between non-model species of flies within my first year as a post-doc (Wittkopp et al. 2003). I now continue this line of work by examining the developmental genetic basis of divergence in male genitalia among species of the Drosophila melanogaster subgroup. This exposure to new ideas has triggered my passion to pursue an experimental means of understanding functional aspects of adaptive genetic change. I decided that yeasts are the most experimentally versatile system that could best address these issues, so I attended the Cold Spring Harbor yeast genetics course to accelerate my mastery of the methodological toolkit. Using yeasts as a model system for functional evolutionary genetics I have addressed three areas of research: 1) the evolutionary relationships among yeasts as revealed by phylogenomics, 2) the historical demography and patterns of selection in natural populations of *S. cerevisiae*, and 3) the fitness effects of mutations fixed between closely related species of yeasts.

Yeast Phylogenomics

In collaboration with Antonis Rokas and Nicole King, we used then unpublished yeast genomic data to reconstruct the evolutionary history of eight key species (Rokas et al. 2003). Because those genomes were largely syntenic, we identified genes that were truly orthologous, and could be aligned over much of their length, which resulted in one of the largest, most reliable phylogenetic data sets to date. We took advantage of this genome scale data set to address one the greatest challenges facing molecular systematics, incongruence among multiple data sets. We showed that analyses of one or a few genes often resulted in strong support for conflicting phylogenetic trees, which suggests that analysis of a few genes is not sufficient to establish or refute phylogenetic hypothesis. In contrast, we were able to show that much larger data sets, while still representing less than one percent of total genome, provided maximal support for a given phylogenetic hypothesis. One indicator of the power of these analyses is that this data set has become one of the most thoroughly cited and analyzed data sets in molecular evolution. The robust phylogeny we obtained provides a rigorous comparative framework for evolutionary hypothesis testing, even though many of the species studied have only been discovered within the last decade.

Yeast Population Genomics

Saccharomyces cerevisiae is one of the most thoroughly studied organisms in biology, yet little is known concerning the evolutionary history and ecology of this species in the wild. Previous views were that of an entirely domesticated species, derived from a wild ancestor within the last few thousand years, which likely resulted in mutant strains fit only for industrial environments. To examine these predictions I generated a population genomic data set by sequencing over 100 loci from both wild and domesticated isolates collected across the species' full distribution. Surprisingly, *S. cerevisiae* did not exhibit characteristics of a strictly domesticated species nor did it behave as a typical microbial species. Natural populations exhibited weak genetic isolation and domesticated strains have been recently derived multiple times from an ancient ancestral species. The effective population size was only on the order of hundreds of thousands with both outcrossing and inbreeding individuals in those populations, and there was weak selection on almost exclusively deleterious mutations segregating within species. These results are consistent with population genetic data from a wide variety of eukaryotes. This underscores why yeasts are a valuable tool for genetic research, because understanding the evolutionary processes that produce phenotypes in yeasts provides a powerful model system for understanding those same processes in other species.

Testing predictions of functional genomics: orthologous gene replacement

Evidence suggests that most genes are central to cell maintenance and exhibit functional equivalence across eukaryotes despite sequence divergence; conversely, functional diversification must be generated through sequence divergence. Since virtually all genes exhibit sequence divergence between species, the proportion of changes contributing to functional diversification and the fitness effects of substitutions fixed over the vast majority of the genome are unclear. The "gold standard" test to identify functionally significant genetic differences between organisms is to transfer the genetic change of interest into a different genetic background; however, gene transfer is extremely laborious when possible and genes cannot be transgressed across species boundaries. The yeast experimental system I have developed provides a unique opportunity to precisely transfer any genetic difference between strains of the same species, between closely related sister species, or even between very distantly related eukaryotes. Transfer of genes among strains within species plus transformation of reconstructed ancestral genes tests whether novel substitutions in a given lineage of *S. cerevisiae* are adaptive or deleterious. Gene transfer from the sister species (*S. paradoxus)* can determine the fitness effects of substitutions fixed over time. These tests ascertain the fitness effects of adaptive substitutions among functionally diverged genes, but also address whether functionally conserved genes with sequence divergence exhibit genic incompatibility or neutrality. Transfer of genes from

incrementally more distantly related species can be used to estimate the rate at which such genic incompatibilities arise. These latter tests provide information concerning the evolutionary contingency of maintaining protein interactions or proper protein structure and stability.

The fitness assay I developed is performed with two marked isogenic strains of yeast grown simultaneously and replicated under a variety of environmental conditions. Since the two strains start the competition at equal frequencies, any change in their ratio is a direct measure of the difference in reproductive output, or differential fitness. The strains are marked with an insertion of the green fluorescent protein (GFP), except that one strain has a single nucleotide substitution altering its fluorescence to blue. The key result of this strategy for marking competing strains is that their precise ratio in co-culture can be measured using flow cytometry. Since this instrument can rapidly count 100,000 cells per sample, a difference in the growth rate of 0.002% can be measured after just 10 days of competitive growth.

Results from orthologous gene replacements

The initial phase of this work is nearly complete and I have determined the fitness effects of 8 precise gene replacements from *S. paradoxus* in the *S. cerevisiae* background under 3 different environmental conditions. Surprisingly, gene replacement nearly always had an effect on fitness even though all genes exhibited conserved function. The epistatic fitness effects of gene replacement were nearly always dependent on the environment so that the fitness effects of some gene replacements resulted in increased fitness, and this effect could be reversed to an equally strong decrease in fitness in a second environment. From an evolutionary point of view these effects were extremely large, ranging from 5% to 0.2%, and would lead to rapid loss or fixation of such a gene under these conditions. Alternatively, from a molecular biology perspective these evolutionarily significant substitutions would not be recognized as of consequence because none of the effects would have been detectable under standard growth or functional assays used in yeasts genetics. Since most gene replacements involve roughly one hundred substitutions, these data also indicate that individual substitutions must be largely nearly neutral in their fitness effects. Lastly, there was no relationship between fitness effect and the rate of evolution, number of substitutions, gene function, essentiality, or the presence of a gene duplicate. The dependence on genetic background was therefore an unpredictable yet ubiquitous effect. These findings signify that genomes steadily evolve by substitutions of small fitness effects that, while dependant on the environment, buffer phenotypes through maintenance of proper interactions among the underlying genetic components.

Future work

Compensatory protein evolution and tolerance to amino acid change

Proteins can evolve incompatibility between species without functional divergence through suppression of deleterious mutations by compensatory mutations at other sites, which would result in highly epistatic genomes as is consistent with my gene replacement data. Previous work indicates that compensatory mutations could be an unrecognized, yet fundamental aspect of all molecular evolution. If common, this hypothesis represents a revolutionary understanding of protein evolution. Rather than neutrality, molecular evolution is largely driven by selection. However, adaptation does not optimize or diversify protein function, instead proteins continually "run in place" as deleterious mutations each generate a constellation of potential new suppressors or compensatory mutations that are fixed before deleterious mutations can be swept out of a population or reverted. This new view of evolution could lead to a better understanding to such diverse topics as genetic suppression, genetic disease prevalence, epistasis, dominance, speciation, and the evolution of sex. However, even though it is predicted to be ubiquitous, especially among closely linked sites (within genes), there is virtually no empirical evidence demonstrating the prevalence of compensatory mutations in the wild. Given the fitness effects of whole gene substitutions from related species into *S. cerevisiae*, I have the unique opportunity to test the prevalence of intragenic compensatory mutations by generating chimeric gene

fusions between species to look for an increased fitness effect in chimeric genes relative to whole gene replacements and wild type genes. The relative simplicity of gene replacement also creates the opportunity to potentially map some of these compensatory mutations to single amino acid substitutions. Finally, molecular and biochemical assays, commonly used in yeast molecular biology, will determine how protein function, stability, regulation, known physical interactions and in some cases solved structure has been altered and solved by compensatory mutations accumulated between species.

To understand the link between protein evolution and fitness I will also carry out the systematic mutation of a single gene via replacement of every individual amino acid, at a small and well studied gene, with a radical substitution (e.g. alanine) and determine the fitness effects of each. Such alanine scans are common in molecular biology in order to map interaction domains, for example, but such a systematic study of mutation and fitness has never been completed. I can re-examine those amino acids that cause a change in fitness with more conservative amino acid substitutions or classes of substitutions that have been tolerated in other lineages of yeasts. This information will provide a missing link between molecular biology and evolution to correlate mutations known to alter function with the fitness effects of those mutations.

The functional evolution of gene duplicates in the HSP70 family of Ascomycetes

A central theme of molecular evolution is that gene duplication may be a major driving force in evolutionary diversification. In collaboration with Jaroslaw Marszalek (University of Gdansk, Poland) and Elizabeth Craig (University of Wisconsin), we are combining the wealth of functional genetic and biochemical knowledge of the *S. cerevisiae* heat shock protein (HSP) 70 family developed by my collaborators, with an evolutionary understanding of paralog maintenance. I have traced the complete evolutionary history of this gene family using genomic data from fungi, and found that the 14 HSP70 paralogs of *S. cerevisiae* were derived from 7 paralogs ancestral to all Ascomycetes, but were not generated by a known genome duplication event. Interestingly, some duplicates exhibited high rates of evolution and functional specialization, while other duplicate pairs exhibited concerted evolution via pervasive gene conversion. My collaborators have previously characterized the protein interactions, and their specific amino acid sites, necessary for a case of specialized paralog function. Future work will determine the efficiency of a single copy ancestral protein for each of the paralog's unique functions. Next, we will test whether mutations in sites critical for protein interactions for different functions lead to genetic trade-offs in fitness and how paralogs that have lost such trade-offs impart genetic robustness to these interactions.

Conclusion

The research platform I have developed for *S. cerevisiae* has the potential to test any number of hypotheses that have been difficult to address in evolutionary biology. I have focused on evolutionary genetics and the fitness effects of protein change, but this platform can be used in the future to address any question regarding genetic change and I have a particular interest in addressing the evolution of gene regulation. My previous work has provided the only experimental, empirical data to support pervasive compensatory evolution and future work will transition to obtaining a molecular and biochemical understanding of the process of compensatory evolution. I will also continue to collaborate with biochemists, cell biologists, and geneticists to add an evolutionary perspective to a comprehensive, systems approach to understanding gene function.

Rokas A.*, B.L. Williams*, N. King, and S.B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*. 425:798-804.

Wittkopp P.J.*, B.L. Williams*, J.E. Selegue, and S.B. Carroll. 2003. Drosophila pigmentation evolution: divergent genotypes underlying convergent phenotypes. *Proceedings of the National Academy of Sciences of the U.S.A.* 100(4):1808-1813.

Williams B.L., J. Brawn and K. Paige. 2003. Genetic effects of recent habitat fragmentation on the wide ranging, high gene flow butterfly *Speyeria idalia* (Nymphalidae). *Molecular Ecology* 12(1):11-20.

Williams B.L. 2002. The role of extinction in taxonomy and conservation: a case history of the regal fritillary. *Conservation Biology*. 16(1):148-157.

* Co-first Authors

Teaching Philosophy

Barry L. Williams

I consider education, both in the classroom and as a mentor, to be a primary responsibility of university faculty members. My goal as an educator is not only to ensure that students acquire knowledge and a fascination with the biological sciences, but more importantly, that they develop the problem solving skills necessary to think both independently and critically. To achieve this goal the curriculum must focus less on the quantity of course content, but instead address fundamental concepts through a number of active learning techniques. Educational research clearly shows that a successful learning environment includes a variety of active learning methods to address a variety of student learning styles. I plan to be part of a growing movement in undergraduate biology education to move away from standard lectures in front of a chalk board, towards the incorporation of active learning via participation in the scientific process.

For example, a standard genetics lecture on mating type switching in yeast might include a diagram of the transitions from haploid to diploid cells and to the production of four haploid gametes, listing each of the genes involved, and how they are regulated at each step. As the number of topics covered in the typical course is quite large, a subject such as this might receive as much as two hours of lecture time. However, mating type switching provides a wonderful opportunity to delve more deeply into the process of genetic reasoning. Incorporation of the active learning principles mentioned above would start by grabbing the students' attention, pointing out that the lowly yeast, used in making bread, bear, and wine, has managed to solve the age-old problem of finding a sexual partner by creating its own! The next step is to pique the curiosity of the students with a fascinating puzzle. How did yeast, in the last 200 million years, develop the ability switch from a single haploid cell, only capable of asexual reproduction, to a diploid cell capable of producing gametes? Instead of monologue, the course would involve a dialogue to develop a list of ideas and outline how we might engineer this genetic circuit, given the ancestral set of genes. This type of brainstorming tends to appeal to social learners and provides students with the opportunity to be active participants in the learning process. I can assemble these ideas in audio-visual presentations (slides, time lapse movies, animations), and incorporate not only student ideas, but pepper the examples with some of the tricks that yeast have actually evolved (for example, mating type switching is initiated with a double strand break at the mating type locus, yeasts have effectively solved this problem by "stealing," through horizontal gene transfer, and modifying a transposable element that now cuts only at the mating type locus and always on cue). This audio-visual approach appeals to students who are primarily visual learners. A more hands-on approach would involve student groups generating predictions of what would happen if a given component in the system were missing. We can discover how these very experiments actually unfolded in the past by reading the primary literature; additionally, such yeast gene knock-out strains are readily available and easily phenotyped in the lab. This last exercise requires that the students formulate falsifiable hypotheses, articulate their ideas clearly among there peers, and revise their ideas as they observe the results of their experiments.

Good teaching practices must also include regular evaluations of both the students and the instructor so that coursework can continually evolve to address portions of each lesson plan that have proven ineffective. The first, and most important, step is to identify common misconceptions and the key core concepts for a given course so that the specific aims of each session are focused on

overcoming misconceptions and comprehension of core concepts. Next, tests and student evaluations must be developed to effectively measure whether these specific aims have been adequately addressed and whether students have retained that content. I strive to be as up-front as possible with students at the beginning of every class session by stating the specific aims for that day, and end each session by presentation of an example evaluation question. Finally, the most effective way for students to address a challenging course is with enthusiasm and determination. I present ideas that are on the cutting edge of modern science and of general interest (HIV evolution, human-chimp evolution) to keep students of all backgrounds interested. I also focus on inquisitive based learning principles that let students make their own discoveries, develop essential critical thinking skills, and treat each topic as a fascinating puzzle to be solved.

Another important aspect of education at a university is mentorship. I personally developed my interest in biology much more through undergraduate research experience than through any course, so I understand how influential such experiences can be in the development of a scientific career. I have mentored six undergraduate and graduate students both as a graduate student and as a post-doc, and I have learned how important proper project scale and time management are to the successful completion of undergraduate research. It is also important that undergraduate research projects, while conceptually developed by the faculty, be carried out primarily by one student to provide a sense of pride in their accomplishments. Yeast provides a versatile and simple system to carry out an endless array of such projects. As pointed out by Nobel Laureate Lee Hartwell, discoveries in yeast genetics are possible because of innovations like toothpicks and velvet.

Having started teaching while still an advanced undergraduate, I have extensive experience in several areas of ecology, genetics, and evolutionary biology. I was fortunate to be hand selected by faculty, based on my previous enthusiasm in their courses, for many of the courses I taught. As a result I have been a teaching assistant in 8 different courses, some of them multiple times, and received university-wide teaching awards for my work in two of them. My duties have ranged from teaching introductory biology labs, to leading discussions and exam reviews, to writing and grading graduate course essay exams, to writing lectures that I have delivered in courses ranging from ten to over 600 students. My extensive experience has given me the opportunity to observe several different obstacles to teaching in the past, and these experiences have helped me devise approaches that minimize the students focus on grades alone, and draw out a curiosity for knowledge.

Regarding courses that I would be interested in teaching, I enjoy presenting issues of general biology to non-science majors. Non-scientists play a pivotal role (e.g. as judges or politicians) in deciding how many science related topics are addressed by society. This highlights the importance of their appreciation for the natural world and the scientific process. I am also eager to develop introductory and upper division courses and seminars in any aspect of evolutionary biology, genetics, and microbiology. These would include evolution, genetics, genomics, evo-devo, bioinformatics, molecular evolution, systematics, population and quantitative genetics, ecological genetics, microbial genetics, evolutionary ecology, and conservation biology. I am also interested in collaborating with colleagues to co-teach courses so that we can each teach to our strengths when one field presents complementary sub-disciplines.