MPE II Abstracts

Invited Keynote Talks

1. A. C. Wilson Keynote: Mechanisms of Protein Evolution

David Pollock, University of Colorado School of Medicine

2. Dobzhansky Keynote: The Eyes Have it - The Evolution of Visual Pigment Function

Belinda Chang, University of Toronto

3. Lotka Keynote: Evolution of Proteins

Richard Goldstein, National Institute for Medical Research, UK

4. Ohta Keynote: The Effect of Protein Structure on Sequence Evolution and of Evolutionary Process on the Lineage-Specific Distribution of Protein Folds Across Genomes

David Liberles, Johan A. Grahnen and Benjamin P. Oswald, University of Wyoming

Protein sequences evolve with constraints dictated by population level and mutational processes in addition to structural and functional constraints. Most common models for protein evolution ignore protein structure and function and assume that all sites in a protein have identical evolutionary processes acting upon them. We have developed a course-grained model of amino acid side chains and corresponding force field and use this to simulate the evolution of protein sequences that fold into a specific structure and bind to a specific partner. Like other models in the field, this model does not describe native sequences well and considerations of negative selective pressures to avoid sequences that do not fold or function properly will be discussed. In the second part of the talk, the distribution of protein folds across genomes will be examined. Since the late 1990s, it has been realized that the distribution of protein folds differs among genomes. The classic explanation for this is that different protein functions accommodating different lifestyles in different organisms necessitates these differences in lineage-specific fold distribution. Against a neutral null model, this functional hypothesis and two alternative hypotheses are tested in eukaryotic genomes. One hypothesis is that large effective population size organisms select for folds that are more robust to deleterious point mutation. Another hypothesis is that mutation-selection balance leads to the evolution of the most evolvable folds in small population size lineages.

5. Wright Keynote: To be Announced

Andy Clark, Cornell University

6. Waddington Keynote: Substitution patterns, life-history traits and the evolution of the population-genetic environment

Nicolas Lartillot, University of Montreal

Substitution processes operating on DNA and protein sequences are influenced by selection, but also by effective population size as well as key parameters of the genetic system (mutation and recombination rates, strength of biased gene conversion). These variables are in turn correlated with life-history traits, such as body size or generation time. All these connections suggest that information can be extracted from sequence alignments about ancestral traits and population genetic environments, which can then be mapped onto phylogenetic trees. In this direction, I will present a Bayesian probabilistic framework for modeling the macroevolutionary process in an integrative manner. The framework takes as an input protein multiple sequence alignments, data about life-history traits of extant species and fossil calibrations. Relying on mechanistically justified substitution models, it then jointly estimates divergence times, life-history evolution, correlations between substitution patterns and quantitative traits and, ultimately, ancestral population genetic environments. Application of the method to placental mammals reveals extensive correlations between life-history and molecular evolution, providing stimulating observations for testing macroevolutionary hypotheses.

7. Lewontin Keynote: Epistasis between mutant sites in mammalian hemoglobin: insights from protein engineering

Jay Storz, University of Nebraska

8. George C. Williams Keynote: Uncovering the molecular basis of phenotypic disunity in innate immune system genes across mammals

Andrew E. Webb[1], Claire C. Morgan[1], Noeleen B. Loughran[1,2], Thomas A. Walsh[1] and Mary J. O'Connell[1,3]*

[1] Bioinformatics and Molecular Evolution Group, School of Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland. [2]Current address:Toronto Medical Discovery Tower East (101), Hospital for Sick Children, Toronto, Ontario M5G 1L7, Canada. [3]Current address: Department of Organismic and Evolutionary Biology, Natural History Museum and Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138.

Knockouts of members of the toll-like receptor 4 (TLR4) pathway of the innate immune system in mouse do not produce the same immunological phenotype as in human, e.g.

IRAK-4 and MyD88. The discovery of a small molecule that activates the receptor for granulocyte-colony-stimulating factor (G-CSF), a cytokine that triggers the growth of white blood cells showed promise for chemotherapy patients (to boost their immune systems). However this only worked in mice and not in humans. These are instances of what we term phenotypic disunity (PD) and they mark functional diversification in the immune systems of human and mouse. Previously we have described the functional effects of mutating positively selected sites *in vitro* in the innate immune system gene Myeloperoxidase, and have shown a clear link between positive selection and protein functional shift. We have scaled up this approach and have analysed (*in silico*) all innate immune system genes across over 20 mammals. Our aim with this work is to determine the molecular events that lead to divergence of function (PD) in the innate immune system of mammals. Here we use signatures of adaptive evolution as a proxy for functional shift and together with previous publications on observed phenotypic disunity, we generate hypotheses and locate specific amino acids that are likely to have resulted in PD. We have applied complete genome data from a variety of mammals to determine disunity cases at the species level and have applied population level data from human and mouse to determine the patterns of conservation/variation in putatively positively selected sites. Through this work we will gain with a more complete understanding of (i) the evolution of the immune system, and (ii) the role of positive selection in PD, and we will potentially have a predictive tool in the future for the generation of improved models of human disease.

9. Müller Keynote: From humble beginnings (?); stories a viral RNA can tell us about evolution

Jeff Kieft, University of Colorado School of Medicine

An interest of my lab is the ability of structured RNAs to manipulate complex cellular machines, one example being the ribosome. Recently, we discovered that a relatively small viral RNA can drive a highly non-canonical mode of translation initiation in multiple domains of life. Thus, this one RNA functionally bridges billions of years of divergent evolution to exploit those parts of the ribosome that have not changed. Digging into the mechanism of this RNA, we see that even smaller versions, representing discrete structural elements, are enough to drive this process in bacterial ribosomes. As we examine the structure-function relationship of these RNAs, a plausible story emerges that explains how these RNAs could evolve to be more efficient and more specific, and how RNAs like these could expand the proteome.

10. Mayr Keynote: The origin of Eukaryotes: New data and new methods

James McInerney, National University of Ireland, Maynooth

Over the last few years, we have collected a wealth of information about eukaryotes - we have sequenced their genomes, measured the expression of their genes and their rates of evolution, investigated which genes were essential for life, which proteins interacted with one another, which genes were responsible for illness and so forth. At the same time, we

experienced stagnation in our efforts to ask the fundamental question - how did eukaryotes come to exist? Ever larger datasets of phylogenetic information were only partially helpful, in large part because of horizontal gene transfer among prokaryotes (against which eukaryotes were to be compared) and because of uncertainty in orthology and the very ancientness of the origins of eukaryotes. In this talk, I will show that progress can be made in evolutionary biology generally and in our efforts to understand eukaryote evolution specifically, if we take a broad, inclusive approach and we examine characters that are generally outside of the standard set we usually see in phylogenetic and phylogenomic analyses.

11. Carl Woese Keynote: Protein evolution in the Planctomycetes-Verrucomicrobia-Chlamydiae superphylum

Naomi L. Ward, Department of Molecular Biology, University of Wyoming, Laramie, WY, USA

The generation of a natural, phylogeny-based taxonomy for the Bacteria and Archaea was one of the most important legacies of the work of Carl Woese. This taxonomy, in addition to clarifying the relationships of well-known bacterial groups (principally those of medical or agricultural importance), also brought to light the phylogenetic distinctness of many poorly characterized groups. Within this latter category fall the planctomycetes and verrucomicrobia, two groups of bacteria originally studied for their unusual cellular morphology and biochemical properties. Early rRNA-based analyses, by Woese and co-workers, revealed that these organisms occupied distinct lineages within the tree of the Bacteria, very distant from model organisms such as *Escherichia coli* and *Bacillus subtilis*. Much more recent analyses indicate a superphylum relationship that embraces the phyla Planctomycetes, Verrucomicrobia, and Chlamydiae (PVC), as well as Lentisphaerae, Poribacteria, and OP3. The organisms of the superphylum are physiologically divergent. They include obligate human pathogens, free-living soil and aquatic microorganisms, and organisms found in close association with metazoan hosts. Certain lineages also feature the presence of complex endomembranes of unknown origin and function, as well as cellular processes that superficially resemble those seen in eukaryotes. Therefore, study of this group may provide information on the evolution of complex intracellular and extracellular structures in bacteria, and lineage-specific processes associated with the development of host association and acquisition of pathogenic potential. This has recently become possible through the availability of genome sequences for multiple organisms of the PVC superphylum (helping to fulfill Woese's 1998 call for "a phylogenetically representative genomic screen of the microbial world"). Several examples of comparative evolutionary analyses at both the whole-genome level (featuring analysis of indel substitutions and gene family dynamics), and single-genome level will be discussed.

12. Haldane Keynote: The evolution of a catalytic mechanism

Tony Dean, University of Minnesota

An outstanding question in molecular evolution is the origin of new catalytic chemistries. I will describe the evolution of two catalytic mechanisms. Isocitrate dehydrogenase contains an invariant active site lysine not found in related enzymes carrying out the same overall catalytic chemistry on homologous substrates. We show this lysine is essential to catalysis in isocitrate dehydrogenase and needed to overcome ground state stabilization of the substrate in the Michealis complex. The evolution of hydroxynitrile lyase, a plant defense enzyme, from a family of esterases represents a dramatic change in catalytic chemistry. In esterases catalysis begins with nucleophilic attack by a serine hydroxyl on the carbonyl carbon of the ester bond to form an acyl-enzyme intermediate with release of the alcohol. An activated water is then used to hydrolyze the carboxylic acid from the serine. In hydroxyl nitrile lyase the very same serine acts as a base, abstracting the proton from the hydroxyl to form a carbonyl with concomitant release of cyanide. The shift from nucleophyllic attack to acid-base chemistry is achieved by just three amino acid replacements while retaining the essential catalytic machinery. Our results provide fundamental insights into the evolution of biological novelty.

13. Fisher Keynote: The Drift-Barrier Hypothesis and the Evolution of Subcellular Featuers

Michael Lynch, University of Indiana

A general principle in evolutionary biology is that the efficiency of selection declines with reductions in population size - once the level of refinement of a molecular feature attains the point at which subsequent beneficial mutations have fitness effects less than the power of drift, further steps toward molecular perfection are no longer possible. Hence, unless all beneficial mutations have large effects (unlikely), the limits to the evolution of molecular perfection will be more extreme in small populations. This drift-barrier hypothesis has general implications for all aspects of evolution, including the performance of enzymes and the stability of proteins. It also implies that effective neutrality is the expected outcome of natural selection, an idea first suggested by Hartl et al. in 1985. In the context of these arguments, I will discuss several aspects of protein and cellular evolution that appear to be governed by the stochastic effects of drift: 1) the inverse relationship between mutation rates and both population size and proteome size; 2) the substantial investment that cells make in surveillance mechanisms for error-prone processes; and 3) the widespread tendency of proteins to evolve oligomeric structures.

14. Kimura Keynote: An attempt to estimate tiny fitness effects and an attempt to exploit huge fitness effects

Alexander R. Griffing, Liwen Zou, Eric A. Stone, and Jeffrey L. Thorne, North Carolina State University*

Two early-stage projects will be summarized in this talk. In the first, I will describe our efforts to quantify the recessivity/dominance of fitness effects of synonymous mutations. In the second, I will outline our approach to studying p53 gene evolution via the combination of interspecific sequence data and human disease data.

Accepted Talks

15. Non-contiguous structure-based protein recombination: a tool for engineering proteins and studying natural recombination

Devin L. Trudeau, Matthew A. Smith, and Frances H. Arnold.*

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA

Genetic recombination is found throughout the tree of life. Examples include horizontal gene transfer in bacteria, V(D)J recombination in the adaptive immune system, and meiotic recombination in sexual reproduction. There has been speculation about the selective advantages of sex and recombination, which include increasing the fitness variation of a population and allowing the purging of deleterious mutations. We are studying the costs and benefits of recombination at the protein level. We explore this experimentally by making chimeric proteins and observing how their biophysical properties are affected by such factors as recombination sites, mutation rate, structural disruption, and parent identity. One major difficulty with recombining distantly-related sequences is that most of the resulting 'chimeric' proteins are unfolded and non-functional. To increase the probability of creating folded chimeras, this laboratory has developed useful structure-guided approaches. The SCHEMA energy, for example, quantifies structural disruption in terms of amino acid contacts that are broken by recombination. We have found that chimeras with a high SCHEMA energy (highly disrupted) have a low probability of being folded and functional, while chimeras with a low SCHEMA energy have a higher probability. To study the behavior of protein chimeras arising from highly divergent parents (identities as low as 30%), we have used the SCHEMA energy to guide the design of chimeric proteins composed of non-contiguous blocks from up to four parents. With this method we have been able to create functional proteins with more than 100 mutations relative to the closest parent. The resulting chimeric protein libraries are diverse at the amino acid level; they also exhibit a range of properties such as thermostability, activity, pH tolerance, and expression. An important observation we have made is that sequence blocks from different parents contribute in a largely additive manner to the thermostability of a protein chimera. This allows us to make linear regression models for the thermostabilities of chimera libraries, which we use to predict the most stable chimeras. We are currently using this approach to make highly stable endo-β-1,4-glucanases, a class of cellulases which is important to the biofuels industry. Characterizing chimeric libraries of endo-β-1,4-glucanases, as well as other proteins, will help us to understand the structural basis for protein properties like thermostability and function at elevated temperatures. Moreover, we aim to improve the predictability of chimera expression, activity, and ability to refold after heat treatment, with a view towards developing structure-guided recombination as a general protein engineering tool.

16. Not different, just better: the adaptive evolution of a glycolytic enzyme

Katherine A. Donovan[1], Sarah A. Kessans[1], Fen Peng[2], Tong Zhu[1], Tim F. Cooper[2] & Renwick C.J. Dobson[1]*

[1]Biomolecular Interaction Centre, School of Biological Sciences, University of Canterbury, New Zealand; [2]Department of Biology and Biochemistry, University of Houston, USA.

Although we have a good understanding of adaptation at the organismal level, there is paucity of data addressing how organisms adapt at the molecular level. A major challenge in studying the adaptation of organisms to their environment is to demonstrate a causal relationship between the organism-level phenotype acted on by natural selection and the underlying molecular changes (e.g. changes in protein function). Our study builds on a laboratory experiment in which 12 replicate bacterial populations were evolved from a common ancestor in a glucose-limited environment for >50,000 generations. The fitness of each population increased relative to the ancestor, indicating adaptation to the experimental environment. Whole genome and candidate gene sequencing found that mutations occurring in the replicate populations exhibit a high degree of parallelism at the gene level. One gene that was mutated in all 12 populations is *pykF,* which encodes a glycolytic enzyme central to the regulation of energy metabolism, pyruvate kinase. Why is *pykF* a focal point for mutations? This question demands a molecular 'picture' of the evolved pyruvate kinase enzymes. I will present structural, biochemical and fitness data to assess the effect of the adaptive mutations. To examine whether the evolved *pykF* alleles confer a fitness benefit in and of themselves, we replaced the wild-type *pykF* gene in the ancestor with each of the evolved *pykF* genes. Competitive fitness assays (against the ancestor) demonstrate that, although each of the *pykF* genes conferred a fitness benefit, the magnitude of the fitness benefit was different in each case. Enzymatic analysis demonstrates that the evolved enzymes have surprisingly different characteristics with respect to activity, substrate binding and allosteric regulation. In addition, thermal stability assays demonstrate both increased and decreased stability when compared to the wild-type. Despite the variability in fitness and function, the crystal structures and small angle X-ray scattering profiles for the evolved pyruvate kinase enzymes are surprisingly similar to each other and the wild- type. Comparing the B-factor profiles of our structures, however, hint at differences in protein dynamics. In conclusion, although the long-term evolution experiment demonstrates a high degree of parallelism with respect to fitness and mutational patterns, our data suggest much less parallelism with respect to protein function. Moreover, our results point to protein dynamics as an important mode for adaptive evolution in proteins.

17.A phylogenetic model for progressive stages of cancer

Jing Zhao

University of Georgia

As biotechnology advances rapidly, a tremendous amount of cancer genetic data become available, providing an unprecedented opportunity for understanding the origin and

progression of cancer. Previous studies have demonstrated that cancer progression is an evolutionary process. However, the probabilistic model for understanding this evolutionary process remains limited. We develop a phylogenetic model in which progressive stages of cancer are described as the tips of a phylogenetic tree. It is assumed that duplication and deletion occur in accordance with a continuous time Markov Chain along the branches of the tree. The simulation study suggests that the Bayesian phylogenetic model can consistently estimate the duplication and deletion rates as cancer advances. The Bayesian phylogenetic model was applied to a real data set. The analysis found 60 genes that are significantly associated with the onset of cancer.

18. Latent effects of Hsp90 mutants revealed at reduced expression-levels

Dan Bolon

In natural systems, selection acts on both protein sequence and expression-level, but it is unclear how selection integrates over these two dimensions. We recently developed the EMPIRIC approach to systematically determine the fitness effects of all possible point-mutants for important regions of essential genes in yeast. Here, we systematically investigated the fitness effects of point-mutations in a putative substrate-binding loop of yeast Hsp90 over a broad range of expression-strengths. Negative epistasis between reduced expression-strength and amino-acid substitutions was common, and the endogenous expression-strength frequently obscured mutant defects. By analyzing fitness effects at varied expression-strengths, we were able to uncover all mutant activity effects. The majority of mutants caused partial defects, consistent with this region of Hsp90 contributing to a mutation-sensitive and critical function. These results demonstrate that critical regions of proteins can tolerate mutational defects without obvious impacts on fitness.

19. Differential expansion of select gene families in snake genomes

Drew R. Schield[1], Daren C. Card1, Qing Li[2], Mark Yandell[2], Todd A. Castoe[1]

1University of Texas at Arlington, 701 S. Nedderman Drive, Arlington, TX 76019; 2University of Utah, 6160 Eccles Institute of Human Genetics, Salt Lake City, UT 84112

We used information from domain-based analyses of three squamate reptile genomes (*Anolis, Python,* and *Ophiophagus*), and *a-priori* identified genes of interest, to compare patterns of expansion and contraction of gene families in snake genomes. Our results include identification of a diversity of patterns that have clear ties to the unique and extreme phenotypes that characterize snakes. These results include inferences of differential expansion of venom-related genes in the cobra, contraction of opsins in the ancestral snake lineage, and expansion of vomeronasal receptors in multiple lineages of snakes.

20. Snake venom three-finger toxins as a model of small protein structure/function evolution

Stephen P. Mackessy

School of Biological Sciences, University of Northern Colorado, 501 20th St., CB 92, Greeley, CO 80639-0017 USA

Snake venoms have evolved as an efficient replacement for mechanical means of predation, allowing for rapid incapacitation of prey with minimal contact by the predator. For the successful deployment of this chemical weaponry, a limited repertoire of proteins has been co-opted for overexpression in the venom glands. Among these, the three-finger toxins (3FTxs) appear to be most diverse functionally, with more than 12 distinct pharmacologies described. These small proteins consist of 60-74 amino acid residues, with 4-5 disulfide bridges. Somewhat enigmatically, virtually all members of this protein superfamily adopt a canonical, highly conserved molecular fold, defined by 3 adjacent loops consisting primarily of β pleated sheets and stabilized by 4 invariant disulfides in a globular core. In general, residues immediately flanking the Cys residues are conserved, but the non-structural residues show extreme variability. Proposed mechanisms of sequence evolution include point mutational changes and ASSET (accelerated exchange of exon segments), and exon rates of mutation far exceed introns. We have isolated, sequenced and characterized several non-conventional 3FTxs from rear-fanged snake venoms which share the three finger fold but which show taxon-specific toxicity, producing a lethal flaccid paralysis in lizards and birds but not affecting mammals at high doses. Sequence similarity analyses with over 100 other snake venom 3FTxs revealed only two short sequence motifs with potential to impart species selectivity. Within this hypervariable protein family, small primary structural changes produce significant functional shifts, perhaps via surface charge density changes, leading to the extreme specificity of toxin ligands for specific receptor subtypes.

21. Evolution of Pyruvate Specificity in Apicomplexan Lactate Dehydrogenases

Jeffrey I Boucher[1], Joseph R Jacobowitz[1], Scott Classen[2] & Douglas L Theobald[1]

[1]Brandeis University, Department of Biochemistry, Waltham MA 02454, USA.; [2]Physical Biosciences Division, Lawerence Berkeley National Laboratory, Berkeley, CA 94720, USA.

The evolution of novel enzyme functions is poorly understood. Outstanding questions involve: 1) the effect of epistasis on evolutionary irreversibility, 2) the role of promiscuous intermediates during evolution of function, and 3) the number of mutations required to produce a novel function. Lactate dehydrogenase (LDH) and Malate dehydrogenase (MDH) are homologous enzymes that share a fold yet possess strict specificity for their substrates, pyruvate for LDH and oxaloacetate for MDH. LDH has convergently evolved from MDH four times, including the evolution of Apicomplexan LDHs via the lateral gene transfer of a bacterial MDH. A striking difference between

modern Apicomplexan MDHs and LDHs is a six amino acid insertion within the "specificity loop" of the LDHs. Incorporating this insertion into a modern Apicomplexan MDH does not confer pyruvate activity and reduces oxaloacetate activity. However, using ancestral sequence reconstruction, we have shown that this insertion is responsible for the evolution of pyruvate specificity in Apicomplexan LDHs. Introducing the insertion within the ancestral MDH confers pyruvate activity with no loss of oxaloacetate activity. This ancestral MDH with the insertion is a promiscuous enzyme with comparable activity towards oxaloacetate and pyruvate. Subsequently mutating an Arg to Lys within the specificity loop commits the enzyme to LDH function, yielding pyruvate specificity only 10-fold less than the modern *P.falciparum* LDH. We have solved the crystal structures of the ancestral MDH, the ancestral LDH, and the ancestral MDH with the six amino acid insert. These structures confirm that the primary structural difference between the proteins is the conformation of the specificity loop. Our results show that the convergent evolution of pyruvate specificity in Apicomplexan LDHs arose through a neofunctionalization mechanism involving strong epistasis, a highly specific, promiscuous intermediate, and few mutations of large effect.

22. The evolution of RuBisCO stability at the thermal limit of photoautotrophy

Scott R. Miller, Michele A. McGuirl and Darla Carvey

Division of Biological Sciences, Missoula, MT, 59812

A longstanding question in evolutionary biology is how organisms adapt to novel environments. In North American hot springs, diversification of a clade of the cyanobacterium Synechococcus into hotter environments has resulted in the unique innovation of a light-driven ecosystem at temperatures up to 74 °C, and temperature adaptation of photosynthetic carbon fixation with the Calvin cycle contributed to this process. Here, we investigated the evolution of thermostability of the Calvin cycle enzyme ribulose-1, 5-bisphosphate carboxylase/oxygenase (RuBisCO) during Synechococcus divergence. Circular dichroism thermal scans revealed that the purified RuBisCO of the most thermotolerant Synechococcus lineage is more stable than those of other lineages or of resurrected ancestral enzymes. Using site-directed mutagenesis, we next identified four amino acid substitutions that together increased stability and activity of this enzyme at higher temperatures. These clustered near critical subunit interfaces distant from the active site. Each of the four amino acids is also observed in a less thermostable Synechococcus RuBisCO, and the impact on stability of three of these appears to be epistatic. Recombination analyses that allow for recurrent mutation as well as patterns of synonymous variation surrounding these sites suggest that the evolution of a more thermostable RuBisCO may have involved homologous recombination. Our results provide insights on the molecular evolutionary processes that shape niche differentiation and ecosystem function.

23. Are Type I and Type II rhodopsins are related through convergent or divergent evolution?

Kristine Mackin

There is a longstanding controversy as to whether Type I and Type II rhodopsins are related through convergent or divergent evolution. Using bacteriorhodopsin from Haloterrigena turkmenica, we created mutants in which the seven transmembrane helices are reordered in the protein's primary sequence to create unique folds. Seven of these mutants retain the ability to insert into the membrane of the E. coli cells used in expression; they also covalently bind and pack against the retinal chromophore in a native-like tertiary structure. These seven mutants are all capable of light-activated proton pumping, with some surpassing the activity of the wild-type. These results are evidence against convergent evolution; since multiple folds are functional, independently evolving proteins are unlikely to arrive at the conserved GPCR fold. We propose that Type I and Type II rhodopsins diverged from an ancient common ancestor, and that accumulated mutations have wiped away any detectable sequence similarity.

24. The Adaptive Evolution Database: An expanding database of phylogenetically-indexed gene families

Russell A. Hermansen, Benjamin Oswald, Stormy Knight, and David Liberles

One of the fundamental questions that molecular evolutionists attempt to answer is, "What makes species unique?" To facilitate answering this question, The Adaptive Evolution Database (TAED) has been built which is a database composed of phylogenetically-indexed gene families from the phylum Chordata. The database is a collection of gene families generated through multiple sequence alignments, maximum likelihood phylogenetic trees, branch specific measurements of positive selection using the ratio of nonsynonymous substitution rates to synonymous substitution rates (dN/dS), and gene tree/species tree reconciliation to categorize duplication events and reconcile each gene tree with respect to the reference NCBI species tree. Ratios of dN/dS are estimated using the free-ratios branch model in PAML, which are validated through a likelihood ratio test against the one-ratio branch model, and are evaluated for dS saturation. Gene tree/species tree reconciliation is applied in a parsimony framework based upon minimizing the number of duplication events implied by the reconciliation. An automated pipeline is being developed to assist in the automatic generation of the phylogenetic trees and ensure that they are maintained to reflect current genetic information. TAED is currently available to use online at www.wyomingbioinformatics.org/TAED. The database currently allows users to search by specific genes, gene functions, or species. The database may also be queried using a BLAST search. TAED also allows for full visualization of gene family data by employing three distinct ways of viewing the phylogenetic trees: normal phylograms, OneZoom TAED trees, and TreeThrasher (TT). TT is a 3D hyperbolic tree visualizer that also serves as a database front end. It was developed for TAED, but can function more generally for any database that is phylogenetically indexed. TT is capable of visualizing the Chordate species tree, which contains links to data underlying the branch in the species tree, such as gene family data.

25.The dynamic nature of protein structures

Jason Lai (1), Jing Jin (2,3, 4), Jan Kubelka (2), and David A. Liberles (1)

Department of Molecular Biology, University of Wyoming, Laramie, WY 82071; Department of Chemistry, University of Wyoming, Laramie, WY 82071; Program in Biochemistry and Molecular Biology, Goshen College, Goshen, IN 46526; Current Address: Department of Biochemistry and Cell Biology, Rice University, Houston, TX 77521

Since the dynamic nature of protein structures is essential for enzymatic function, it is expected that functional evolution can be inferred from the changes in protein dynamics. However, dynamics can also diverge neutrally with sequence substitution between enzymes without changes of function. In this study, a phylogenetic approach is implemented to explore the relationship between enzyme dynamics and function through evolutionary history. Protein dynamics are described by normal mode analysis based on a simplified harmonic potential force field applied to the reduced C(α) representation of the protein structure while enzymatic function is described by Enzyme Commission numbers. Similarity of the binding pocket dynamics at each branch of the protein family's phylogeny was analyzed in two ways: (1) explicitly by quantifying the normal mode overlap calculated for the reconstructed ancestral proteins at each end and (2) implicitly using a diffusion model to obtain the reconstructed lineage-specific changes in the normal modes. Both explicit and implicit ancestral reconstruction identified generally faster rates of change in dynamics compared with the expected change from neutral evolution at the branches of potential functional divergences for the α-amylase, D-isomer-specific 2-hydroxyacid dehydrogenase, and copper-containing amine oxidase protein families. Normal mode analysis added additional information over just comparing the RMSD of static structures. However, the branch-specific changes were not statistically significant compared to background function-independent neutral rates of change of dynamic properties and blind application of the analysis would not enable prediction of changes in enzyme specificity.

26. A two-bead coarse-grained model of protein evolution based on physical properties of amino acid residues

Dohyup Kim, Johan Grahnen, Jan Kubelka, and David A. Liberles

Proteins evolve according to the laws of physical chemistry, biochemistry and population genetics. Although classic site independent models are simple to use, it does not describe evolutionary processes for evolutionary analysis. To study the protein sequence evolution, a two-bead coarse-grained model based on physical properties of amino acid residues was developed by the Liberles group for evolutionary analysis. This model was used to evaluate the effect of both positive and negative selectional pressure in rate heterogeneity and rate heterotachy of the protein coding sequences. To further improve upon this model, two things were considered, the role of decoy structures and parameter optimization. First, in the protein folding simulation, selecting just for the native conformation does not guarantee that that sequence does not fold into structures that are not native transiently or permanently. By evaluating both intrinsic and extrinsic decoys

more carefully, the quality of the model can be improved. Second, an energy function that contains different terms of physical and thermodynamic properties of amino acid sequence is used in the model. These terms are adjusted for each protein to add specificity to the model. Appropriate weighting of contributing terms can be improved by using more sophisticated statistical parameter optimization methods. In this poster, the performance of an improved model is tested and compared by examining properties of proteins such as the distribution of dN/dS ratios across the structure and the distribution of DDG values for mutations. And I will present the importance of selecting appropriate decoy sets as well as weighting terms optimized for specific proteins.

27.Evolution of kinetic parameters within pathways

Alena Orlenko and David A. Liberles

Department of Molecular Biology, University of Wyoming, Laramie, WY 82071

It is clear that selective pressure acts on pathway components non-randomly. Biophysical properties of the enzymes catalyzing reactions within a pathway are expected to be one of the main evolutionary constraints. Research in this direction has been limited. In this project, we are going to fill this gap and explore the co-evolution of enzymes within pathways by analyzing the co-evolution of enzyme biophysical parameters (binding constants, catalytic constants). The main hypothesis that will be tested is that selective pressure along the pathway correlate with the sensitivity of enzymes to perturbations of kinetic parameters. We expect the conservative parts of a collection of proteins to have a central influence on the pathway output, and even a small perturbation of their kinetic constants will lead to dramatic changes in the output concentrations. We also hypothesize that proteins within a pathway should co-evolve according to a compensatory mechanism involving mutation-selection balance to keep the output product level approximately the same when negative selection is acting on the pathway. We expect that effective population sizes and mutation rates will influence the dynamics of this mechanism.

28.Evolution of Sec pathway proteins in PVC bacteria

Olga K. Kamneva, David A. Liberles, Naomi L. Ward

Molecular Biology Department, University of Wyoming, Laramie, WY

The Planctomycetes, Verrucomicrobia, Chlamydiae (PVC) super-phylum contains bacteria with either complex cellular organization or simple cell structures. Genome content evolution of this group has not been studied in a systematic fashion, which could reveal genes underlying the emergence of PVC-specific complex intracellular structure. Here we use a comparative genomics approach to identify gene families potentially associated with cellular compartmentalization in this bacterium. We analyze 26 PVC genomes and the genomes of several outgroup species. We identified a number of gene families with presence/absence profiles highly correlated with the presence/absence of intracellular membranes. Analyzing domain composition of those gene families, we find that an unexpectedly high fraction of those carry signal peptides which suggests the

involvement of the Sec-pathway with the compartmentalization of PVC bacteria and protein targeting to intracellular membranes. We further analyze the phylogeny of Sec proteins in PVC bacteria and find divergent duplicates of SecA ATPase genes in a number of PVC genomes and domain rearrangements within canonical SecA genes in Planctomycetes species.

29.Evolutionary dynamics of conformational flexibility in Flaviviruses

Juan Felipe Ortiz, Madolyn MacDonald, Patrick Masterson, and Jessica Siltberg-Liberles

Department of Molecular Biology, University of Wyoming, Laramie, WY 82071

Conformational flexibility allows proteins to modify their structure according to changes in their environment. This characteristic is observable in proteins across the tree of life. Here, we have investigated if regions that promote conformational flexibility are conserved over evolutionary time and we also address if changes in conformational flexibility could be related to functional plasticity. Using parsimony, the evolutionary dynamics of conformational flexibility, inferred from structural disorder prediction, was addressed in flaviviruses genome-wide. The results reveal fluctuations in structural disorder among lineages, but not in a temporal manner. Some protein regions were observed to undergo rapid transitions between an ordered and a disordered state in response to amino acid substitutions, while other regions were always ordered. No correlation between evolutionary rate and order-disorder transition rate could be established. Lineage-specific composition of regions with structural flexibility could alter the theromodynamics of the conformational ensembles, allowing these proteins to change promiscuous functions, such as transient interactions, without losing their predominant function. In flaviviruses, rapid evolutionary dynamics of structural disorder was observed and could be a potential driving force for phenotypic divergence.

30. A new model for estimating substitution processes using transposable elements

Aaron Wacholder and David D. Pollock*

Biochemistry and Molecular Genetics, University of Colorado School of Medicine

Accurate estimation of neutral (or nearly neutral) substitution processes is critical for understanding mechanisms of protein evolution, since the evolution of functional molecules is a compound process including context-dependent mutation rate as well as selection based on molecular function. Transposable elements, particularly SINEs (short interspersed nuclear elements), are excellent candidates for such study because they are highly active in many genomes, and many copies are generally thought to arise from single elements (and thus their ancestral sequence can be easily inferred). The standard view of transposable element evolution assumes a set of “master element” sequences, each actively replicating during a particular range of time, whose relations to each other

can be represented by a phylogenetic tree. However, recent empirical research suggests that many elements can be active concurrently, contradicting the master model. Moreover, we found implausibly high levels of variation at particular sites within subfamilies, indicating a faulty relationship between subfamily assignment and the ancestral sequence for individual elements, and possible error in the tree relating subfamilies. We therefore developed an MCMC method to infer the ancestral distribution of TEs given an observed extant distribution in a single genome. This method does not assume a particular phylogenetic tree relating these ancestors, and utilizes the transition probabilities from ancestors to descendents to calculate posterior probabilities for all ancestors. Our results indicate that elements sometimes have multiple possible ancestors with high posterior probabilities. Furthermore, our analysis suggest that ancestral TEs are related by a network rather than a tree-like structure, with bidirectional flow between classes of active elements.

31. Influence of the membrane environment on the molecular evolution of G protein-coupled receptors

Stephanie Spielman

University of Texas, Austin

G protein-coupled receptors (GPCRs) represent one of the largest families of membrane proteins in eukaryotes. In Metazoa, GPCRs play significant roles in a wide array of critical physiological functions, including perception, metabolism, and homeostatic regulation. Mutant GPCRs frequently induce severe diseases (such as cancer or hormonal disorders), and over 40% of current pharmaceuticals target GPCRs. Despite their clear functional importance, little is known about GPCR evolution. We have conducted a large-scale computational study of the evolutionary dynamics in mammalian GPCRs. Through deep species sampling, we were able to calculate site-specific evolutionary rates across over 350 GPCRs. From this analysis, we infer that GPCR regions embedded in the membrane (transmembrane domains) evolve more slowly, on average, than do other regions of the same protein. Despite this attenuated evolutionary rate, transmembrane domains exhibit more among-site evolutionary variability, and frequently contain a protein's most quickly evolving sites. Further, we note that chemosensory receptors, an important class of GPCRs that receive external chemical stimuli, evolve at a much faster rate than do GPCRs that receive endogenous signals. In particular, chemosensory GPCR display a significant rate increase in their transmembrane domains. While the plasma membrane does exert strong biophysical constraints on GPCR evolution, GPCR transmembrane domains display a remarkable ability to evolve and display much evidence of positive selection.

32. Convergent evolution of substrate specificity in a Parabasalid malate/lactate dehydrogenase family

Phillip Steindel, Emily Chen, Douglas Theobald*

During the evolutionary history of the lactate and malate dehydrogenase (L/MDH) superfamily, substrate specificity has switched from malate to lactate four independent

times, making these enzymes an ideal model system for exploring the biophysical basis of both substrate specificity and convergent evolution. In the trichomonads, an order of amitochondriate protists, an unusual LDH evolved from an MDH via a relatively recent gene duplication. Trichomonas vaginalis MDH has high malate activity and little lactate activity, while the LDH has only a slight preference for lactate, as well as activity towards other similar substrates. This promiscuity is conserved throughout the trichomonad LDHs. To better understand the mechanistic basis for these changes in specificity, we have reconstructed several ancestral trichomonad LDHs and MDHs. These reconstructions reveal two major transitions during LDH evolution, a switch from a highly specific MDH to a less specific MDH, then a switch to a promiscuous LDH. Two mutations during this first transition, R91L and G230W, are similar to mutations at identical positions during other MDH to LDH switches. Mutagenesis of ancestral enzymes, however, shows that these mutations are only partially responsible for this transition. We have solved crystal structures of the modern day Trichomonas vaginalis LDH and MDH, which reveal several other conserved mutations near the active site possibly responsible for substrate specificity changes.

33. The Hidden Markov P-Clouds (HiMP Clouds) project

Jaime Merlano and David D. Pollock

Biochemistry and Molecular Genetics, University of Colorado School of Medicine

To understand how functional molecules evolve, it is an important consideration to understand the underlying mutational and nearly neutral evolutionary process. To this end, we are working to improve methods to evaluate evolution based on transposable elements. The P-clouds method is a fast alignment-free approach (Koning et al 2011) to detect repetitive sequences in genomes. The objective of the current project is to take the P-clouds approach to the next level by increasing sensitivity/specificity and aiming to a more precise prediction of element-specific repeats or transposable elements (TE) such as MIR, Alu, and LAVA. In the creation and structure of element-specific P-clouds, some P-clouds tend to be adjacent to each other and therefore share similar properties. With these P-clouds characteristics in mind, an alignment-free hidden Markov model (HMM) is being constructed where the P-clouds are element-specific states and state transitions are calculated as probabilities between P-cloud states. The hidden states may represent clusters or families of P-clouds that share similar properties and homology (i.e. super P-clouds).

34. Evolutionary dynamics of protein domains across snake and lizard genomes

Daren C. Card[1], Kyle J. Shaney[1], Drew R. Schield[1], Qing Li[2], Mark Yandell[2], David Pollock[3], & Todd Castoe[1]

[1]Department of Biology, University of Texas at Arlington, Arlington, TX 76019; [2]Department of Human Genetics, University of Utah, Salt Lake City,

UT 84112; [3]Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045

Comparative genomics examining protein domains can be useful in determining large-scale differences in the content of the proteome among species, and inferring plausible causes of phenotypic variation between lineages. We used the MAKER2 annotation pipeline to identify genes and their corresponding protein domains in the Green anole lizard (Anolis carolinensis) genome and two snake genomes: the Burmese python (Python molurus bivittatus) and the King Cobra (Ophiophagus hannah). By analyzing the copy number of particular domains across these genomes, we identified several types of domains that have experienced large, or what we believe to be biologically relevant, changes among species. Our results identify instances of expansions in number of several protein domains in the ancestral snake lineage, and also in subsequent snake lineages that shed some light on the extreme phenotypes evident in certain snake clades (e.g. infrared sense, venom, and extreme physiological regulation).

35. (dN/dS<1) = Selection for Function – Myth or Reality?

Omer S. and Gogarten J.P.

Department of Molecular and Cell Biology, University of Connecticut, Storrs CT

In the evolutionary studies of protein coding DNA sequences, a small ratio in the rates for nonsynonymous to synonymous substitutions (dN/dS ratio) is interpreted as evidence for purifying selection and implicitly, selection for function. Recent studies have found low dN/dS ratios in many genes present in bacterial genomes such as GTA (Gene Transfer Agent) genes or ORFans. These ratios are interpreted to reflect purifying selection acting on the proteins, suggesting that these genes produce functional proteins that would benefit either the gene itself (selfish gene), the host organism, or the host population (group selection). Our preliminary results suggest this interpretation to be wrong. Prophage structural gene sequences within conserved gene neighborhoods in different Escherichia coli strains carry signatures of purifying selection indistinctive from that acting on functional genes. The syntenic position of the analyzed prophage genes in the host genome indicates that over the time span represented in our dataset these genes were only vertically inherited and never transferred in phage particles as part of the analyzed genealogy. Our findings, based on simulations under a neutral evolution scenario, reveal that purifying selection acts on these prophage gene sequences, and that the hypothesis of their neutral evolution can be rejected ($P << 0.000001$). We speculate that the selection against amino acid replacements does not reflect a selection for function, but a selection against protein toxicity of the mutants.

36. Bridging the gap between the molecular-structure effects of mutations and the population genetics in Arabidopsis

Carrie Olson-Manning[1], Thomas Mitchell-Olds[1], Jane Richardson[2]

[1]Department of Biology, Duke University, Durham, NC; [2]Department of Biochemistry, Duke University, Durham, NC

Rate heterogeneity varies greatly both within and between different proteins. However, the degree to which these observed evolutionary differences are explained by population genetic history, natural selection, and physical/biochemical constraints remains unclear. The role these effects play in the realized mutational spectra confuses inferences about the functional effects of natural protein polymorphism. Here, in a collaboration between evolutionary and structural biologists, we ask what evolutionary forces and constraints shape protein structure at the amino acid level and how those patterns relate back to natural variation. We examine polymorphism in 80 natural accessions of Arabidopsis thaliana and interspecific variation with its relative A. lyrata. For the many A. thaliana crystal structures in the protein databank, we evaluate compatibility of observed polymorphisms with local structure and possible effects on intermolecular interactions, using all-atom contact analysis of sterics and hydrogen bonding. This in-depth structural detail takes into account not only properties of the amino acid (i.e., size, charge and solvent accessibility) but also the specific 3D context of its surroundings. We then examine what evolutionary forces are acting on these protein structures using traditional sequence signatures of selection. And, finally, we seek to connect the two approaches both statistically and with specific illustrative examples. This strategy combines population genetics theory and structural biochemistry towards the goal of understanding the functional effect of segregating enzyme polymorphism.

37. A mixture model for bias and error in genomic data reduces false positive identification of heterozygotes

Rachel S. Schwartz and Reed A. Cartwright

Center for Evolutionary Medicine and Informatics The Biodesign Institute, Arizona State University

Accurate identification of heterozygous sites in genomes is essential to identification of many de novo mutations. Currently, 40–60% of de novo mutations identified by statistical methods using next generation sequencing data are false positives, which greatly increases the number of sites that must be validated. In this study, we examined heterozygous sites for one human with high next generation sequence coverage. We found that a mixture of Dirichlet-multinomial distributions provided an excellent fit to the distribution of reads at each site. The fit of the model suggests that it can be incorporated into SNP calling methods that currently assume a binomial distribution of reads with some error rate. Additionally, we examined the subset of heterozygous sites previously identified by the 1000 genomes project as variable in the human population. The majority of these sites fell into one of the components in the model for all heterozygous sites. This distribution had almost no read bias compared to the distributions for the remaining sites. This result suggests that a mixture model may provide a way of reducing false positive identification of heterozygous sites.

38. Estimating Indel Models via Simulation and Optimization

Akash Khare and Reed A. Cartwright

Center for Evolutionary Medicine and Informatics The Biodesign Institute, Arizona State University

Because of their importance in molecular evolution, many attempts have been made to infer indel (insertion and deletion) rates from comparative genomic data. Previous methods tend to be inadequate because they either do not control for alignment ambiguity or can only estimate biased models. In this study we have developed a simulation approach that corrects for estimation bias and can make accurate inference of insertion and deletion mutation patterns based on comparative genomic data from sister species. We first use EMDEL to calculate summary statistics for a dataset of real sequences. Between distantly related species, EMDEL's estimates will be biased because it does not account for overlapping indels. To compensate, we generate simulations via DAWG and search for a set of simulation parameters that reproduces the summary statistics of the real data. The resulting simulation parameters are an unbiased estimate of indel model parameters for the real data.

39. Llambda: Estimating Indel Rates and Length Distributions from a Multiple Sequence Alignment

Jessica Albanese and Reed A. Cartwright

Center for Evolutionary Medicine and Informatic,s The Biodesign Institute, Arizona State University

Indels (insertion and deletion) are a fundamental but understudied process in molecular evolution. In order to study the evolution of indel rates and length distributions across the tree of life, accurate and efficient methodologies are necessary. In this project, we update the LAMBDA.PL estimation procedure which is a component of the simulation software DAWG. Our new software, LLAMBDA, improves accuracy by better handling overlapping gaps and uses an outgroup to distinguish insertions and deletions. As input, LLAMBDA takes a multiple sequence alignment and a phylogenetic tree with branch-lengths. It outputs maximum-likelihood estimates of the per-substitution rates and length distributions of insertion and deletions. It outputs estimates for several different families of length distributions along with goodness-of-fit criteria that allows scientists to intelligently determine which models are the most biologically accurate.

40. EnRICH: Extraction and Ranking using Integration and Criteria Heuristics, a new software tool to identify high quality candidate genes from high-throughput datasets

*Xia Zhang, M Heather West Greenlee, Jeanne M Serb**

Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa, USA

High throughput screening technologies enable biologists to generate candidate genes at a rate that, due to time and cost constraints, cannot be studied by experimental approaches in the laboratory. Thus, it has become increasingly important to prioritize candidate genes for experiments. To accomplish this, researchers need to apply selection requirements based on their knowledge, which necessitates qualitative integration of heterogeneous data sources and filtration using multiple criteria. A similar approach can also be applied to putative candidate gene relationships. While automation can assist in this routine and imperative procedure, flexibility of data sources and criteria must not be sacrificed. A tool that can optimize the trade-off between automation and flexibility to simultaneously filter and qualitatively integrate data is needed to prioritize candidate genes and generate composite networks from heterogeneous data sources. We developed the java application, EnRICH (**E**xtractio**n** and **R**anking using **I**ntegration and **C**riteria **H**euristics), in order to alleviate this need. Here we present a case study in which we used EnRICH to integrate and filter multiple candidate gene lists in order to identify potential retinal disease genes. As a result of this procedure, a candidate pool of several hundred genes was narrowed down to five candidate genes, of which four are confirmed retinal disease genes and one is associated with a retinal disease state.

41. Selective pressure analysis of human and mouse innate immune genes at both the species and population level

Andrew E. Webb[1], Claire C. Morgan[1], Thomas A. Walsh[1] and Mary J. O'Connell[1,2]

[1] Bioinformatics and Molecular Evolution Group, School of Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland. [2] Current address: Department of Organismic and Evolutionary Biology, Natural History Museum and Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138.

In modern immunology, the mouse has become the predominant model organism - unsurprising considering their reported similarities to human in both physiology and genetics. However, an increasing number of instances have been reported in the scientific literature whereby mouse models of human disease do not mimic their human counterparts. This raises important questions about the widespread suitability of mouse in modeling innate immunity and about the molecular underpinnings of these variations. We wished to assess if there are signatures of positive selection in these genes that could potentially lead to the functional disparity observed at the phenotypic level. To this end we have formulated a bioinformatic pipeline to analyze the molecular evolution of these innate immune genes. We have generated a dataset containing 725 innate immune-related human genes from the highly curated database InnateDB. We then identified homologs across 21 high-coverage (>6X) vertebrate genomes. Utilizing these gene datasets we performed two stages of analysis. The first stage assayed for lineage-specific positive selection (LSPS) whereby homologs were grouped into protein families for alignment, model selection, Bayesian phylogenetic reconstruction, and selective pressure analysis where appropriate. The second phase of our analysis assessed the putatively positively

selected genes (and sites) in the context of SNP frequencies from population data for human (1000 Genomes and HapMap projects) and mouse (17 Genomes). Sites under LSPS were assessed for variation from background levels within the population data. Here we present our findings to date.

42. Adaptive fine-tuning of the key photosynthetic enzyme Rubisco to fit different climates

Maxim V. Kapralov[1], Takuya Nakazato[2] and Dmitry A. Filatov[1]

[1] Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK; [2] Department of Biological Sciences, The University of Memphis, Memphis, Tennessee 38152, USA

Ribulose-1,5-bisphospate carboxylase/oxygenase, called for short Rubisco, serves as a gateway for inorganic carbon to enter metabolic pathways in most ecosystems on Earth. Rubisco is literally the most abundant enzyme in the world and comprises up to 50% of all soluble protein in photosynthetic tissues, which is the price that plants have to pay for its large size and very slow turnover. As the performance of Rubisco can greatly affect crop yields, substantial efforts have been made to study its structure and function using directed mutagenesis, with the aim to artificially improve Rubisco performance. However, neither the practical problem of delivering better enzymes for crops nor the fundamental questions about Rubisco evolution in different groups of plants have been resolved so far. We analyse results of the Rubisco mutagenesis experiment performed by nature during evolution of flowering plants. Using Maximum likelihood and Bayesian analyses of DNA sequences, we found that genes encoding Rubisco are evolving under positive Darwinian selection in angiosperms. Further, we pinpointed residues under directional selection in c. 7000 flowering plant species and checked whether amino acid replacements at these sites correlate with climate parameters experienced by studied species. Three amino acid replacements under selection showed significant correlation with annual temperature and one of them also showed significant correlation with annual precipitation. All sites under selection are located in functionally important regions of the Rubisco enzyme. Our results suggest that Rubisco properties are being adjusted by natural selection to better fit the environmental conditions and that currently predominant "one size fits all" model for Rubisco kinetics is incorrect.

43. Functional characterization of natural variants at positively-selected sites in rhodopsin

D. David Yu, Amir Sabouhanian, and Belinda S.W. Chang, Departments of Cell & Systems Biology, Ecology & Evolutionary Biology, University of Toronto

Although in recent years much interest has centered on the development of likelihood-based codon models (dN/dS) to detect selection, the vast majority of these studies have been solely computational, with only a small fraction combining computational analyses with experimental investigations of positively-selected sites. Rhodopsin is the dim-light

sensor in vertebrates, the first step in the visual transduction cascade in the rod photoreceptor cells of the retina. Random sites codon models were used to test for positive selection in a dataset of vertebrate rhodopsins largely derived from AToL sequencing efforts (>400 taxa). Sites inferred to be undergoing positive selection in vertebrate rhodopsin were found to be clustered in portions of the transmembrane domains thought to be important in retinal uptake and release. Site-directed mutagenesis was used to generate all natural amino acid variants found at three of our positively-selected sites. In total, 24 mutant pigments were expressed *in vitro*, and spectroscopically assayed for three aspects of rhodopsin function, including λmax, and the rate of retinal release in the light-activated and dark-inactive states. Some of the variants at positively-selected sites were found to cause significant shifts in various aspects of rhodopsin function. This finding is in contrast to two previous studies of rhodopsin that found no correspondence between dN/dS-based site predictions, and functional assays of λmax. Finally, we are also interested in investigating associations between natural variants of rhodopsin and ecological characters that might affect aspects of rhodopsin function, such as temperature, and habitat depth, and present preliminary results here. Our investigations of natural variants in rhodopsin highlight the need for further studies that combine computational approaches with experimental studies in order to better understand the molecular evolution of protein structure and function.