ABSTRACTS

Genomes and Evolution 2004
The Pennsylvania State University
University Park, Pennsylvania
June 17-20, 2004

Annual meeting, Society for Molecular Biology and Evolution
Annual meeting, American Genetic Association

GENOMES AND EVOLUTION 2004

Organizing Committee: Hiroshi Akashi, Douglas R. Cavener, Claude W. DePamphilis, S. Blair Hedges (Chairman), Wojciech Makalowski, Kateryna Makova, Masatoshi Nei, Stephen W. Schaeffer, and Shozo Yokoyama. Conference planning staff: Carolyn Andersen and Tammy Golden. Production staff: Brian Heckman and Stella Uyeno. Logo artwork: Jennifer Hines and Blair Hedges.

SOCIETY FOR MOLECULAR BIOLOGY AND EVOLUTION

Council (2004): John C. Avise (President), Naoyuki Takahata (Past-President), Jeffrey R. Powell (President-Elect), Sudhir Kumar (Secretary), Marta L. Wayne (Treasurer), G. Brian Golding (Councilor), Jody Hey (Councilor), Laura A. Katz (Councilor), and William Martin (Editor).

AMERICAN GENETIC ASSOCIATION

Council (2004): Shozo Yokoyama (President), Kent E. Holsinger (Past President), John Doebley (President Elect), Oliver A. Ryder (Executive Vice President), Norman F. Weeden (Treasurer), Linda Strausbaugh (Secretary), and Karen Cone, Sally Otto, Brad Shaffer, Caro-Beth Stewart, Sara Via, Robert Wayne, Adalgisa Caccone, Andrew Clark, and Scott V. Edwards (Council Members).

INVITED TALKS

Plenary Talks

Avise J, Hood L, Gehring W

Early Evolution of Life

Baldauf S, Koonin E, Doolittle F, Ochman H

Genome Evolution

Andersson S, Green E, Palmer J, Wessler S, Wolfe K

Molecular Phylogeny and Molecular Clocks

Cracraft J, Kumar S, O'Brien S, Vilgalys R

Development and Evolution

Carroll S, Levine M, Sinha N, Theissen G

Genome Evolution in Primates

Eichler E, Paabo S, Saitou N, Varki A

Origins and Evolution of Genetic Systems

Firestein S, Gojobori T, Klein J, Parham P

Molecular Polymorphisms

Batzer M, Clark A, Hartl D, Nachman M

Molecules and Biodiversity

Cheng C, Mathews S, Real L, Rieseberg L

Adaptive Evolution

Jeffery W, Kocher T, Okada N, O'Neill R, Zhang J

ORAL PRESENTATIONS

Walter M. Fitch Symposium

Bofkin L, Cai J, Engelhardt B, Raes J, Roy S, Torgerson D, Turner E, Zufall R

Early Evolution of Life

Andersson J, Archibald J, Brenner S, Brooks D, Caetano-Anolles G, Calteau A, Friedman B, Gaunt M, Koonin E, Makarova K, Man O, Nakamura Y, Penny D, Silva J, Snoeyenbos-West O, Stajich J, Yoon H

Genome Evolution

Achaz G, Adams K, Anxolabehere D, Boore J, Crease T, Denver D, Dermitzakis E, Doniger S, Emerson J, Feschotte C, Goldman N, Gotea V, Gu X, Hartling J, Iwashita S, Kroll E, Liang H, Makova K, Man O, Mararis G, Meyer A, Moran N, Neafsey D, Panopoulou G, Parsch J, Quesneville H, Ricker J, Rifkin S, Rispe C, Sémon M, Shiu S, Simillion C, Singh N, Tillich M, Wadhawan S, Wolf Y, Zhang P

Molecular Phylogeny and Molecular Clocks

Barker K, Blair J, Fain M, Gibas C, Hay J, Hoberman R, Kitazoe Y, Leebens-Mack J, Madsen O, Martins L, Moore J, Ray D, Roelants K, Rokas A, Sampedro-Jimenez J, Seo T, Sorenson M, Steen T, Takezaki N, Tamura K, van Tuinen M, Woolfit M

Development and Evolution

Becker A, Cutter A, Ikeo K, Kuraku S, Pollard D, Popadic A, Santini S, Wittkopp P

Genome Evolution in Primates

Abi-Rached L, Boissinot S, Gilad Y, Go Y, Hedges D, Kaessmann H, Meunier J, Murphy W, Ptak S, Rand D, Satta Y, Su B, Wang X, Wildman D

Origins and Evolution of Genetic Systems

Belov K, Esteves P, Grassot J, Kawasaki K, Miller H, Niimura Y, Piontkivska H, Wang W

Molecular Polymorphisms

Andolfatto P, Araki H, Bachtrog D, Briscoe A, Burk-Herrick A, Catania F, Cho S, Crawford D, Dyer K, Fay J, Ganko E, Jagadeeshan S, Pascual M, Pawlowska T, Schlenke T, Ting C, Uyenoyama M, Vekemans X, Webster M, Williams B, Williamson S, Won Y

Molecules and Biodiversity

Aubin-Horth N, Browning T, Lambert D, Ota T, Rissler H, Roca A

Adaptive Evolution

Greenberg A, Guillet-Claude C, Irwin D, Nunney L, Perez-Gonzalez C, Salzburger W, Salzman Y, Saunders M, Suzuki Y, Thorne J, Yan L, Zupunski V

POSTER PRESENTATIONS

Early Evolution of Life

Ge F, Gu X, Okamura H, Zhaxybayeva O

Genome Evolution

Baillie G, Bajaj M, Bergthorsson U, Biedler J, Bliss B, Braasch I, Brunner S, Bulazel K, Carlini D, Cavalcanti A, Chalkia D, Chambers S, Chung W, Crawford A, Cusack B, Dagan T, Dean J, Dehal P, Delaney T, Denawa M, Ezawa K, Ferreri G, Feschotte C, Francino P, Fraser H, Ganesan S, Gibas C, Gibas C, Gibson A, Goertzen L, Goetting-Minesky P, Gong Y, Gonzalez J, Griffin A, Grus W, Gu J, Hackett J, Hao W, Huang J, Huttley G, Itoh T, Jordan K, Kaneko S, Kumar S, Lapierre P, Lin C, Lu Y, Luo J, Mack J, Marzelli M, McGrath C, Meisel R, Merritt T, Moore R, Morris G, Mower J, Negre B, Nekrutenko A, Nozawa M, Ong H, Opiyo S, Perez C, Peterson S, Ponger L, Portnoy M, Prachumwat A, Pritham E, Raghupathy N, Ramos E, Raquel T, Richardson A, Rogozin I, Rosenberg M, Roy S, Schuenzel E, Shedlock A, Simon D, Sinclair C, Sirviö A, Song N, Stefanovic S, Subramanian S, Swigonova Z, Tanaka H, Teeling E, Tetzlaff K, Thornburg B, Urrutia A, Wall K, Wilson A, Wu G, Zhang X, Zufall R

Molecular Phylogeny and Molecular Clocks

Barr N, Battistuzzi F, Beckmann K, Blair J, Cui L, Druzhinina I, Hedges B, Hill E, Kim J, Kong H, Kosiol C, McNeal J, Melo-Ferreira J, Moses A, Pascual M, Raterman D, Scally M, Snoeyenbos-West O, Veerappan C, Vernot B, Wang G, Xing J

Development and Evolution

dePamphilis C, Ellis R, Good J, Goodisman M, Hersh B, Lin Z, Mannaert A, Maughan H, Meyer C, Nam J, Negre B, Zahn L

Genome Evolution in Primates

Cordaux R, Hayakawa T, Iwase M, Kitano T, Kouprina N, Mayer J, Taylor J, van de Lagemaat L

Origins and Evolution of Genetic Systems

Cho S, Erickson B, Gasch A, Hao L, Hazkani-Covo E, Landweber L, Lucas O, Nikolaidis N, Sanchez M, Sawai H, Tanaka T

Molecular Polymorphisms

Achaz G, Bauchet M, Bustamante C, Carboni A, Chaix R, Doherty M, Frankham R, Geraldes A, Gonder M, Haddrill P, Halder I, Hernandez R, Ishiyama H, Jensen J, Jensen-Seaman M, Kamau E, Ko W, Lazzaro B, Lu G, Mao X, Mena-Ali J, Norton H, Ohniwa R, Payseur B, Phinchongsakuldit J, Podlaha O, Rafalski A, Ramakrishnan U, Rothenburg S, Salcedo T, Stoebel D, Takahashi A, Whitehead A, Zhao Z, Zhu L

Molecules and Biodiversity

Carney S, Das J, Hajibabaei M, Huntley M, Kahila Bar-Gal G, Kato Y, Landry C, Pirog K, Pittman K, Raftis F, Reynolds N, Schwarz D, Simonelic K

Adaptive Evolution

Aris-Brosou S, Baines J, Barluenga M, Crouch J, Ferdig M, Goto H, Gu S, Johannesson H, Jørgensen F, Kasuga T, Ketcham K, Lawton B, McGraw L, Mishra P, Montano A, Montooth K, Opazo J, Pecon-Slattery J, Rooney A, Rossiter W, Schmidt D, Terry P, Uinuk-ool T, Watabe T, Weinreich D, Wlasiuk G, Wong A

ABSTRACTS

NOTE: Names of presenting authors listed in **bold**.

ABI-RACHED, Laurent
Emergence of the *KIR* genes and haplotypes in primates
Abi-Rached L. [1], Rajalingam R. [2], Guethlein L. A. [1], Parham P. [1]. (1) Stanford University, (2) UCLA Immunogenetics Center
The *killer cell immunoglobulin-like receptor* (*KIR*) genes encode cell-surface receptors mainly expressed on natural killer cells. This family, which contains 14 genes in humans is characterized by a common genomic localization within 150-200kb, a high level of allelic polymorphism and a variability in haplotype gene content. Species comparison further illustrated the diversity and rapid evolution of this family in primates, with few real orthologs and many recombinants. This extent of recombination, coupled with recent and rapid expansions of some of the *KIR* lineages makes it difficult to fully resolve the gene relationships with only the coding sequences.
To address this question, but also to characterize the underlying genomic events associated with the emergence of the different lineages, and to understand how the haplotypes were generated, we characterized a full chimpanzee *KIR* haplotype. Subsequently, we used several phylogenetic methods to study the relationships of the different parts of all the available *KIR* gene sequences. The results of these analyses will be presented, and together with the comparison of the gene localizations, used to reconstruct the history of the emergence of the *KIR* genes/*KIR* haplotypes in primates.

ACHAZ, Guillaume
Slow turnover of HIV-1 populations in chronically infected individuals.
Achaz G. [1], Palmer S. [2], Kearney M. [2], Maldarelli F. [2], Mellors J. W. [3], Coffin J. M. [2], Wakeley J. [1]. (1) OEB, Harvard University, (2) NCI, NIH, (3) University of Pittsburgh
A simple, non-parameteric test for population structure was applied to temporally spaced samples of HIV-1 sequences from the *gag-pol* region within two chronically infected individuals. The results show that temporal structure can be detected for samples separated by about 22 months or more. The performance of the method, which was originally proposed by Hudson *et al.* (1992) to detect geographic structure, was tested for temporally spaced samples using neutral coalescent simulations. By comparing levels of temporal structures in simulations to the levels observed in real data, we estimate the effective intra-individual population size of HIV-1 to be between 10**3 and 10**4 virus, which is in agreement with some previous estimates. The effective population size is extremely small compared to the experimentally observed one (about 10**10 virus). The definition and interpretation of estimates of such effective population parameters are discussed.

ACHAZ, Guillaume
***cis*-regulatory and protein evolution in orthologous and duplicate genes**
Castillo-Davis C. I. [1], Hartl D. L. [1], **Achaz G.** [1]. (1) OEB, Harvard University
The relationship between protein and regulatory sequence evolution is a central question in molecular evolution. It is currently not known to what extent changes in gene expression are coupled with the evolution of protein coding sequences or whether these changes differ among orthologs (species homologs) and paralogs (duplicate genes). Here, we develop a method to measure the extent of functionally relevant *cis*-regulatory sequence change in homologous genes and validate it using microarray data and experimentally verified regulatory elements in different eukaryotes species. We find that protein and regulatory evolution is functionally weakly coupled in orthologs but not paralogs, suggesting that selective pressure on gene expression and protein evolution is quite similar and persists for significant amount of time following speciation but not gene duplication. Additionally, duplicates

exhibit a dramatic acceleration of both regulatory and protein evolution in comparison with orthologs suggesting increased directional selection and/or relaxed selection on both gene expression patterns and protein function in duplicate genes.

ADAMS, Keith

Genome evolution and gene silencing in polyploids

Adams K. L. [1], Wendel J. F. [1]. (1) EEOB Department, Iowa State University

Most eukaryotes have genomes that exhibit high levels of gene redundancy, much of which appears to have arisen from one or more cycles of genome duplication (polyploidy), a process that is especially common in flowering plants. Polyploidization has played a major role in the evolution of plant genomes. Gene pairs duplicated by polyploidy may be co-expressed or one copy may be silenced or lost. Molecular techniques were used to screen the expression of over 450 genes and to assay in detail expression levels of 50 genes duplicated by polyploidy, using multiple organs and genotypes of polyploid cotton. Silencing of one gene copy was documented for several genes. Some genes show partitioning of expression patterns between duplicates, with one duplicate copy being silenced in some organs and the other copy being silenced in other organs, suggesting subfunctionalization. Newly created polyploids are being used to understand the immediate consequences of polyploidization on gene expression. Gene silencing is probably epigenetically mediated in the newly created polyploids.

ANDERSSON, Jan O.

Gene transfers from Nanoarchaeota to an ancestor of diplomonads and parabasalids

Andersson J. O. [1], Sarchfield S. W. [2], Roger A. J. [2]. (1) Uppsala University, (2) Dalhousie University

Rare evolutionary events, such as lateral gene transfers, may be useful to pinpoint, and correlate the timing of, key branches across the tree of life. Here we present phylogenetic analyses of two aminoacyl-tRNA synthetases that both indicate gene transfer events from the newly discovered hyperthermophilic archaeal phylum Nanoarchaeota to an ancestor of two unicellular eukaryotic groups: the diplomonads and parabasalids. Thus, these two protistan groups, often regarded as the two deepest independent offshoots of the eukaryotic lineage, share a common ancestor to the exclusion of the eukaryotic root. Our results also show that Nanoarchaeota is older than this common ancestor and that the split genes in the Nanoarchaeum equitans genome is a derived feature. This unexplored phylum likely had, and may still have, representatives living in close proximity to mesophilic eukaryotes.

ANDERSSON, Siv

Evolution of the mitochondrial genome and proteome

Andersson S. G. [1], Frank A. C. [1], Karlberg O.. [1], Bossau B. [2]. (1) Uppsala University, (2) Lyon University

The alpha-proteobacteria, from which mitochondria are thought to have originated, displays a 10-fold genome size variation. The recent sequencing of a dozen alpha-proteobacterial genomes, including our own completed genomes of *Rickettsia prowazekii*, the typhus pathogen, *Bartonella quintana*, the agent of trench fever and *Bartonella henselae*, the agent of cat-scratch disease, enables a global genomic comparison of alpha-proteobacteria and mitochondria. Here, we discuss the use of computational approaches to infer genomes at ancestral nodes and to quantify the flux of genes along the branches of the alpha-proteobacterial tree. Our analysis suggests that the alpha-proteobacterial ancestor contained 3,000 to 5,000 genes and was biochemically highly versatile, with a complete system for aerobic respiration and a broad biosynthetic capability. We follow the fate of these pathways in the various descendents, including the mitochondrion, and estimate the extent of gene loss, gene transfer and gene genesis that have contributed to shape the mitochondria. Our comparative analyses of eubacterial and mitochondrial proteomes have shown that modern mitochondria are the result of both reductive and expansive processes. The results are summarized in a model on the origin and evolution of the mitochondrial genome and proteome.

ANDOLFATTO, Peter

Distinguishing between selection and demography in genome wide scans of variability.

Andolfatto P. [1], Becquet C. [2]. (1) Dept. of Zoology, University of Toronto, (2) Dept. of Ecology and Evolution, Brown University

Considerable attention has focused on using patterns of nucleotide variation within species to identify targets of positive selection. Various statistical tests that have been developed generally rely on the assumption that the population has been of constant size. Yet species may experience frequent changes in population size. In particular, humans are thought to have experienced a population size reduction (or bottleneck) when emigrating from sub-Saharan Africa. Similar out-of-Africa bottlenecks have been proposed for *Drosophila melanogaster* and its sibling species, *D. simulans*. We use coalescent-based approaches to show how population bottlenecks can lead to patterns of variation that closely resemble those expected under positive selection. In particular, we show that bottlenecks can lead to considerable heterogeneity in patterns of variability across a recombining chromosome, and that these patterns can also be chromosome-specific. Our results have important implications for reliably distinguishing between selection and demography in genome-wide scans of nucleotide variability.

ANXOLABEHERE, Dominique

Transposable element domestication: recurrent recruitments of the DNA binding THAP domain.

Anxolabehere D. [1], Nouaud D. [2], Quesneville H . [1]. (1) Instiut Jacques Monod Univ. Paris 6, (2) Instiut Jacques Monod CNRS

A novel evolutionary conserved protein motif, the THAP domain defines a new family of cellular factors, the THAP proteins. It harbours strictly similarities with the DNA-binding domain (DBD) of the *P* element transposase of *D. melanogaster* (1). We show that the THAP domain is present in the proteins encoded by the *Drosophilidae P* neogenes originated from recurrent molecular domestications of different *P* elements families. Some of them have undergone exon shuffling involving other *P* transposons and resulting in duplication of THAP domain (2). These data represent repeated host recruitment of the DNA-binding function of the *P* transposon. In human the THAP-9 protein exhibits a significant identity with the P transposase, so it could correspond to be a P neogene. We discuss the evolutionary relationships between the THAP domain and the DBD of the P element transposase, which are largely phylogeneticaly distributed, at least from dipterans to the human. Ref. 1) Roussigne et al., 2003. *TBS* 28 (2) ; 2) Nouaud D. et al. 2003. *Mol. Biol. Evol*. 20.

ARCHIBALD, John

Genome reduction in eukaryotes: nucleomorph genomes as a case study

Archibald J. M. [1], Theophilou S. [1], Fong A. [1], MacKinnon M. [1]. (1) Department of Biochemistry & Molecular Biology, Dalhousie University

The cryptomonads and chlorarachniophytes are two highly unusual groups of microalgae that have acquired photosynthesis through a process called *secondary endosymbiosis*. This occurs when a phagotrophic eukaryote engulfs a photosynthetic eukaryote and retains its plastid (chloroplast). The molecular details of secondary endosymbiosis are not well understood, but it appears to be a complex process involving the large-scale movement of DNA from the nucleus of the engulfed eukaryotic cell to that of the host. In most algae of this sort, the gene transfer process has gone to completion and the endosymbiont nucleus has completely disappeared. However, in cryptomonads and chlorarachniophytes, the nucleus of the endosymbiont persists in a highly reduced form called a *nucleomorph*. As evolutionary intermediates in the process of intracellular gene transfer and genome reduction, nucleomorph-containing algae have the potential to provide an important glimpse into the nature of secondary endosymbiosis and the process of eukaryotic genome evolution. Toward this end, we are using a comparative genomics approach to examine the diversity of nucleomorphs in a wide range of chlorarachniophyte and cryptomonad species. Here we present preliminary data on the genomic regions surrounding the nucleomorph heat shock protein 70 (hsp70) and hsp90 loci in members of both groups.

ARIS-BROSOU, Stephane

Parameterization of codon models, bias and implications for translational selection

Aris-Brosou S. [1], Bielawski J. P. [2]. (1) North Carolina State University, (2) Dalhousie University

We compare the influence of two parameterizations of the codon model developed by Goldman and Yang on the estimation of synonymous rates of molecular evolution. The first parameterization strictly follows the original authors' as the model uses codon frequencies in the rate matrix (GY94). Another approach, inspired by Muse and Gaut, is to incorporate nucleotide frequencies in the rate matrix (MG94). We show by simulation that estimation of synonymous rates is biased under both parameterizations when codon frequencies are uneven. However, MG94 appears more robust than GY94 at small nonsynonymous to synonymous ratios, while the opposite is true at ratios greater than one. An application to real data shows that parameterization can in some cases affect both the direction and the significance of the correlation between synonymous rates and codon usage. This demonstrates that one should be very careful when estimating synonymous rates.

AUBIN-HORTH, Nadia

Neurogenomics of a Short-Circuited Life in Wild Salmon

Aubin-Horth N. [1], Letcher B. H. [2], Hofmann H. A. [1]. (1) Bauer Center for Genomics Research, Harvard University, (2) Natural Resources Conservation, University of Massachusetts

Atlantic salmon are best known for spectacular marine migrations before homing to spawn in natal rivers. However, many males do not migrate and instead become sexually mature at sizes ten times smaller than anadromous males. The ultimate evolutionary causes of this developmental plasticity as well as the morphological and life-history differences between these two reproductive tactics have been studied extensively. However, the nature and extent of coordinated changes that must occur at the physiological and molecular level to organize such an extreme change in phenotype is unknown. We show that individual brain gene expression profiles of early maturing and immature males from a wild population differ significantly, revealing important molecular components implicated in constructing these divergent phenotypes. The inclusion of several individuals in our analysis allows making robust inference on the molecular basis of complex traits variation resulting from regulation of the genome.

BACHTROG, Doris

Positive selection drives Y-chromosome degeneration in Drosophila

Bachtrog D. [1]. (1) Cornell University

Why does the Y chromosome harbour so few functional loci? Evolutionary theory predicts that Y-chromosomes degenerate because they lack genetic recombination. Both positive and negative selection models have been invoked to explain this degeneration, since both can result in the recurrent fixation of linked deleterious mutations on a non-recombining Y chromosome. To distinguish between these models, I investigate patterns of nucleotide variability along 37 kilobases of the recently formed neo-Y chromosome in *Drosophila miranda*. Levels of nucleotide variability on this chromosome are 30-fold lower than in highly recombining portions of the genome. While both positive and negative selection models can result in reduced variability levels, their effects on the frequency spectrum of mutations differ. Using coalescent simulations, I show that patterns of nucleotide variability on the neo-Y chromosome are unlikely under deleterious mutation models (including background selection and Muller's ratchet) but are expected under recent positive selection. These results implicate positive selection as an important force driving the degeneration of Y-chromosomes; adaptation at a few loci - possibly increasing male fitness - occurs at the cost of the majority of other genes on this chromosome.

BAILLIE, Gregory

Evolution and retrotransposition of betaretroviruses in mice, rats, and other mammals

Baillie G. J. [1], van de Lagemaat L. N. [1], Gagnier L. [1], Baust C. [1], Mager D. L. [1]. (1) Terry Fox Laboratory, BC Cancer Agency

We have discovered and characterised several groups of previously unknown endogenous betaretroviruses in the genomes of *Mus musculus* and *Rattus norvegicus*. Each group contains both mouse and rat elements, and several of the groups are more closely related to previously-known betaretroviruses from non-murine hosts. Some of the groups also include members from hosts which were not previously known to harbour betaretroviruses, such as the gray mouse lemur (*Microcebus murinus*) and Seba's short-tailed bat (*Carollia perspicillata*). Some of the mouse and rat elements possess intact open reading frames for *gag*, *pro*, *pol*, and/or *env* genes, and display characteristics of having retrotransposed recently.
The mouse type D endogenous retrovirus MusD belongs to one of these groups (β7), and is closely related to a group of active LTR retrotransposons, the ETns. We have developed a retrotransposition assay to test whether ETns retrotranspose using proteins encoded by MusD.

BAINES, John
Analysis of candidates for positive selection identified by a microsatellite screen of natural populations of the house mouse
Baines J. F. [1], Ihle S. [1], Ravaroarimanana B. [1], Tautz D. [1]. (1) Institut für Genetik der Universität zu Köln
A major goal of evolutionary biology is to describe the frequency of adaptive changes within populations and characterize the nature of the underlying genes and mutations responsible. The recent evolutionary history of the house mouse and the availability of its genome sequence make it an excellent candidate to begin addressing this goal. The genus *Mus* is believed to have originated on the Indian subcontinent and only recently colonized the rest of the world, providing good *a priori* expectation that current populations have adapted to new aspects of their environment. Out of 200 microsatellite loci screened for populations of both *M. musculus musculus* and *M. musculus domesticus*, approximately 10 loci display a strong signature of a selective sweep by the lnRV and/or lnRH statistics. Here, these candidates are investigated in more detail by PCR and direct sequencing of flanking regions and comparing the pattern of polymorphism to that of divergence using *M. caroli* as an outgroup.

BAJAJ, Mamta
Predicting amphipathic helices through statistical analysis
Bajaj M. [1], Moriyama H. [2], Moriyama E. [3]. (1) University of Nebraska Lincoln, (2) Chemistry, University of Nebraska Lincoln, (3) Biological Sc., University of Nebraska Lincoln
Many secondary structure prediction methods have been developed. However, very few methods are available for predicting amphipathic helices. Amphipathic alpha helices are very important for protein structure and functions. These alpha helices have hydrophobic and hydrophilic faces, which are corresponding to the protein side and the other side. Locating this helix helps in predicting the function of a protein such as DNA-binding proteins. We are developing a method that predicts such alpha helices based on a set of new conformational statistics. Training sets consisting amphipathic alpha helices are prepared from the Protein Data Bank based on torsion angle and surface accessibility calculations. We will discuss the performance of this new method comparing to other methods.

BARKER, Keith
A molecular time scale for avian diversification: the largest avian order (Passeriformes)
Barker F. K. [1], Cracraft J. [2]. (1) University of Minnesota, (2) American Museum of Natural History
Reconstructing the history of avian diversification is plagued by the absence of well-sampled fossil data for most groups and time periods. This is particularly true of the small, terrestrial passerine birds. Sequence data from passerines are providing a well resolved estimate of their evolutionary relationships, which yields important insights into their biogeography and evolutionary ecology. In concert with increasingly sophisticated molecular clock methods, these data also provide a new timescale for passerine diversification. Current estimates of the age of passerine lineages imply that passerines existed 30 Ma prior to their first appearance in the fossil record, and 60 Ma prior to their appearance in Eurasia. This

pattern is likely a function of passerine biogeography, which suggests the group arose in Gondwana prior to the Tertiary.

BARLUENGA, Marta

Sympatric speciation in a crater lake cichlid in the Neotropics: the Midas Cichlid species complex

Barluenga M. [1], Stoelting K. N. [1], Meyer A. [1]. (1) Department of Biology, University Konstanz

The existence of sympatric speciation - i.e., the divergence of populations in the absence of physical barriers - is controversial, although an increasing number of theoretical and empirical evidence suggests that under some conditions speciation in sympatry is possible. The polymorphic Midas Cichlid species complex (Amphilophus spp.) from several crater lakes in Nicaragua fits several of the key characteristics of this model with assortative mating on the basis of a color polymorphism coupled with ecological differentiation based on trophic apparatus and body shape, under fully sympatric conditions. Here we present an integrated study of several molecular markers - AFLP fragments, 10 microsatellite loci and mitochondrial DNA - combined with morphometric and ecological analysis, to identify a unequivocal case of sympatric speciation in one lineage of the Midas Cichlid species complex in the crater lake Apoyo in Nicaragua, Central America. Ecological specialization to alternative life styles - benthic and limnetic - resulting in marked body shape adaptations appears to be the main force of speciation in this system.

BARR, Norman

Molecular systematics of the genus *Ceratitis* (Diptera: Tephritidae)using mitochondrial and nuclear genes

Barr N. B. [1], McPheron B. A. [1]. (1) The Pennsylvania State University

The fruit fly genus *Ceratitis* (Tephritidae), native to the Afrotropics, consists of 88 described species in six recognized subgenera. Like other tephritids, *Ceratitis* species feed on and damage the fruit of host plants. Several species within the genus are polyphagous and pose a threat to agriculture; for example, the Medfly (*C. capitata*) attacks over 300 species in almost 70 plant families. Despite its importance to agriculture, the systematic positions of *Ceratitis* relative to closely related genera, subgenera within *Ceratitis*, and species within the *Ceratitis* subgenera are yet uncertain. I have produced *Ceratitis* phylogenies using mitochondrial (*cytochrome oxidase 1* and *NADH dehydrogenase 6*) and nuclear (*period*) genes, to help understand the systematics of this important group.

BATTISTUZZI, Fabia Ursula

A genomic timescale of prokaryote evolution and early Earth history

Battistuzzi F. U. [1], Feijao A. [2], Hedges S. B. [1]. (1) The Pennsylvania State University, (2) EMBL Heidelberg

A clear understanding of the history of our biosphere may reveal general principles applicable to other worlds. For example, it is widely believed that the metabolic activities of prokaryotes shaped the early evolution of the atmosphere on Earth. However, the important details of this, including the relative abundance of atmospheric gases through time and the evolutionary history of ecologically important organisms, have not yet been resolved. In particular, the timing of the evolution of methanogens is important for models that postulate the presence of methane in the atmosphere. Fossils are of little help in timing the origin of methanogens, but the availability of complete genomes of many species of prokaryotes provides abundant data for molecular clock analyses.

In this study we estimated divergence times of the major groups of prokaryotes with shared proteins from the complete genomes and proteomes of 73 species. We conducted protein sequence analyses using a supergene approach and with local-clock time estimation methods and carefully selected calibration points. The results support the existence of deep divergences in the tree of life. Consideration of current metabolisms of species and groups of prokaryotes permitted to draw some conclusions regarding the timing of origin of some ecologically important metabolisms.

BATZER, Mark

Mobile elements and primate genomic diversity

Ray D. A. [1], Callinan P. A. [1], Salem A. H. [1], Xing J. [1], Hedges D. J. [1], Jorde L. B. [2], **Batzer M. A.** [1]. (1) Dept. of Biological Sciences, Louisiana State University, (2) Department of Human Genetics, University of Utah

An analysis of several recently integrated Alu subfamilies was undertaken to assess Alu element associated genomic diversity. Using this approach we recovered of a number of Alu elements with different distributions throughout the primate lineage. Many of the Alu elements recovered from the human genome were restricted to the human lineage, with some that were polymorphic for insertion presence in diverse humans. Some of the other elements recovered from the human lineage also resided at orthologous positions in non-human primate genomes. Sequence analysis demonstrated that these Alu elements were the products of gene conversion events of older pre-existing elements, independent parallel forward insertions of older elements in the same genomic region, or authentic shared phylogenetic characters. The level of gene conversion between Alu elements suggests that it has had an impact on the nucleotide diversity within Alu elements. We have also identified genomic deletions associated with the retroposition and insertion of Alu elements. Retroposition mediated genomic deletion is a novel source of primate genomic variation. The distribution of Alu elements throughout various primate genomes makes them useful tools for resolving non-human primate systematic relationships.

BAUCHET, Marc

Underdominant selection in cow growth hormone genes

Bauchet M. P. [1], Meade K. [2], Shriver M. D. [1], McHugh D. E. [2]. (1) Penn State University, Dept of Anthropology, (2) Trinity College, Dept of Genetics

This analysis focuses on a set of 24 markers in two cow genes, growth hormone (GH) and growth hormone receptor (GHR) that were typed in 12 cow breeds - including a zebu group. Levels of linkage disequilibrium (LD) between each marker of the two genes were investigated after phasing using two methods (EM and maximum likelihood, with respectively Phase and SnpHap programs) and two different measures of LD (D' and r2). We also measured the LD as correlation between zygotic frequencies (RXC program). Very low levels of LD were observed in two these genes. However, numerous significant deviations from Hardy-Weinberg expectations caused by excess of homozygosity has been measured for several markers across breeds in gene GH. One hypothesis is that underdominant selection has been taking place, i.e. selection against heterozygotes.

BECKER, Annette

Evolutionary genetics of carpels: Using California poppy (*Eschscholzia californica* Cham.) as a basal eudicot model system

Becker A. [1], Smyth D. R. [1]. (1) Monash University, School of Biological Sciences

All flowering plants have carpels, female reproductive structures that enclose the eggs and develop into seed pods and fruits. Carpel development genes are being defined in *Arabidopsis*, a higher eudicot. But the evolutionary origin and the molecular basis of the enormous variety of carpel morphologies among flowering plants is not clear. Here, we describe the characterisation of genes that control carpel development in a more primitive plant, the California poppy (*Eschscholzia californica*). This new model species is a basal eudicot that can be manipulated transgenically allowing for functional analysis of gene functions.

We have been able to identify homologs of the *Arabidopsis* carpel development genes *AGAMOUS* and *CRABS CLAW* in the Californian poppy. Comparison of carpel genes in California poppy and *Arabidopsis* will help reveal core genes that underlie carpel development in all dicots.

BECKMANN, Kevin

Identification of Single Copy Genes for Phylogenetic and Functional Analysis

Beckmann K. G. [1], Leebens-Mack J. [1], Wall P. K. [1], Cui L. [1], Ma H. [1], dePamphilis C. [1]. (1) The

Pennsylvania State University
Despite frequent gene and genome duplication events, some genes are retained in plants as "single copy" through long periods of evolutionary history. To identify conserved single copy genes in flowering plants, we clustered the Arabidopsis and rice proteomes into putative gene families using Tribe MCL. 1300 tribes were identified that contained exactly one member in each species. tBLASTx queries against the TIGR unigene database were used to populate each of these 1/1 clusters with homologous sequences from other plant species. These genes were usually found as single copy when detected in a given plant species, with recent polyploids sometimes having an extra copy. We have developed a high throughput pipeline for alignment, postalignment processing, and phylogenetic analysis to test the prediction that conserved 1/1 families will have a history congruent with the known organismal history. The forces that could result in long-term conservation of single copy genes in plants will be discussed.

BELOV, Katherine

Evolution of the mammalian MHC class II region and genetic diversity of the platypus DZB gene

Belov K. [1], Colgan D. J. [1], Woodward R. [1], Eldridge M. D. [1]. (1) Australian Museum

There are three extant lineages of mammals * eutherians, marsupials and monotremes. Among eutherian orders, orthologous relationships are observed between MHC class II gene families. Such orthologies are not seen between the classical class II gene families of eutherians and non-mammals. MHC genes thus far isolated from marsupials and monotremes do not share orthologs between the two lineages. Nor are their genes orthologous to eutherian or avian gene families. Gene turnover in this complex is dynamic.

cDNAs from class II beta genes were isolated from the two extant monotreme lineages. Methods used to determine nomenclature of these genes and the assignation "DZB" will be discussed. High levels of DZB diversity were seen in platypuses from mainland Australia and Tasmania. The King Island population appears to be monomorphic. Further characterization of the marsupial and monotreme MHC is currently being conducted via cDNA and BAC library screening, BAC contiging and sequencing.

BERGTHORSSON, Ulfar

Transcription and RNA editing of horizontally transferred mitochondrial genes in flowering plants

Bergthorsson U. [1], Palmer J. D. [1]. (1) Department of Biology, Indiana University

Horizontal gene transfer (HGT) of mitochondrial genes between distantly related species of seed plants has recently been discovered. These HGT events have included both genes for ribosomal proteins (*rps2* and *rps11*) and genes coding for proteins involved in respiration (*atp1*, *atp8*, *nad1*). A central question regarding gene acquisition by HGT is whether horizontally transferred genes are functional. Moreover, some mitochondrial HGTs have occurred in organelles where the vertically transferred gene copy had already been lost following intracellular gene transfer to the nucleus. This raises the question whether the plants still retained the ability to correctly recognize RNA editing sites in these genes. We analyzed transcription and RNA editing status in several mitochondrial genes that were acquired by HGT and compared the expression of genes that were regained by HGT with genes of chimeric origin (part vertical, part horizontal) and genes that are present in two copies where one copy is of horizontal origin.

BIEDLER, Jim

Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity

Biedler J. [1], Tu Z. [2]. (1) Virginia Polytechnic Institute and State University, (2) Virginia Polytechnic Institute and State University

Over a hundred families of non-LTR retrotransposons (non-LTRs) were found in the *Anopheles gambiae* genome assembly by a reiterative and comprehensive search using the conserved reverse transcriptase (RT) domains of known non-LTRs as the starting queries. These families, which are defined by at least 20% amino acid sequence divergence in their RT domains, range from a few to approximately 2000 copies

and occupy at least 3% of the genome. 8 of the 15 previously defined clades are represented, plus two novel clades Loner and Outcast, which is more than what has been reported for any genome. The first invertebrate representatives of the L1 clade were also found. 21 families from 8 clades have multiple full-length copies with over 99% nucleotide identity and other sequence characteristics suggesting recent activity. Interesting genome distribution bias was found for two of these families. The maintenance of multiple recently active diverse lineages indicates a complex evolutionary scenario. These non-LTRs have the potential to be developed as tools for population studies and genetic manipulation of the primary vector of the devastating disease malaria.

BLAIR, Jaime
Molecular clock methodology and impact on time estimation
Hedges S. B. [1], **Blair J. E**. [1]. (1) The Pennsylvania State University
Despite the abundance of sequence data from genomes of organisms, molecular time estimates rely on accurate calibrations, usually from the fossil record. All such calibrations are minimum times of divergence. Nonetheless, if the calibration is robust, it may be close to the true time of divergence. Only one calibration is required, although additional robust calibrations are desirable if sequences are available. Here we look critically at some recent molecular clock studies that have yielded conflicting results. One proposing a young date (425-490 Ma) for the origin of land plants is contradicted by the discovery of land plant fossils from the Early Cambrian (520 Ma). Another finding relatively young dates for divergences among animal phyla (~580 Ma) contains estimates that violate the fossil record as well. In contrast, studies finding earlier (Precambrian) divergences for plants and animals are robust to different methods and remain broadly consistent with the fossil record.

BLAIR, Jaime
Phylogenetic Sequence Analysis of Eukaryotic Genomes
Blair J. E. [1], Hedges S. B. [1]. (1) The Pennsylvania State University
Complete genome sequences are now available for several vertebrate (human, mouse, fish) and invertebrate (fruit fly, mosquito, worm, tunicate) animals, as well as from fungi and plants. The number of genes in each genome ranges from 6,000 to more than 20,000, with each gene coding for a protein with typical length of 300-400 amino acids. This represents several orders of magnitude more data for evolutionary research than previously available. Potentially, these valuable new data will accelerate our understanding of early life. However, at the same time they create analytical challenges and the need for more sophisticated methods. Distinguishing gene duplication from speciation events and accommodating rate variation are two of the most difficult issues and they are addressed here. With some new bioinformatic approaches we have conducted complete genome sequence analyses and our results bear on the early evolution of animals and the rise in complex life on Early Earth.

BLISS, Barbara
Whole genome evidence for multiple transfers among plant organelles
Bliss B. J. [1]. (1) Pennsylvania State University
Classic phylogenetic analysis of sequence information is used to infer species relationships. Greater similarity of sequence is presumed to reflect conserved function or more recent common ancestor. Horizontal gene transfer (HGT) results in a topology that differs from the accepted organismal topology. Historically, HGTs have been uncovered one at a time, using careful hybridization experiments. We used a bioinformatics approach to develop a whole genome view of multiple organelles. Previously reported interorganellar transfers are clearly illustrated along with numerous new putative horizontal transfers. Simulations modeling the frequency of occurrence of high identity segments, along with consideration of sequence function were used to select several high identity segments. Phylogenetic analyses support the hypothesis that organelle genomes have been more extensively invaded by horizontal gene transfer than previously shown. Increasing availability of whole, well-annotated genomes and tools to manipulate

them will allow rapid identification and analysis of evolutionary events involving multiple genomes, including HGT.

BOFKIN, Lee

Measuring heterogeneity in evolutionary processes within and between genomes: novel applications of nucleotide substitution models to large multiple-alignments.

Bofkin L. N. [1], Whelan S. [1], Goldman N. [1]. (1) EMBL - European Bioinformatics Institute

The identification of functional elements within genomes is of great importance. Many recent studies have examined rate variation in large numbers of small windows of genomic sequence and inferred that those evolving at the lowest rates are functional. We have developed a novel statistical modelling technique that detects whether two adjacent windows are evolving to significantly different processes, using a maximum likelihood framework. The tests perform well because they take into account more factors than previous studies; we demonstrate the effectiveness of the technique using yeast alignments and identify previously uncharacterised putative functional features on *S. cerevisiae* chromosome 3. We also demonstrate changes in "average" evolutionary processes across regions, including the mammalian Cystic Fibrosis Transporter Gene Region. Furthermore, the method has been adapted to test for replication origins, revealing a signal at the primate β-globin locus that previous methods had not detected.

BOISSINOT, Stephane

Molecular Evolution of LINE-1 retrotransposons since the origin of primates

Khan H. [1], **Boissinot S.** [1]. (1) Department of Biology, Queens College, CUNY

LINE-1 (L1) retrotransposons constitute the most abundant family of autonomously-replicating retrotransposons in the human genome. Although L1 retrotransposons seem to have been continuously active since the origin of mammals, the rate of L1 amplification, and presumably the impact of L1 activity on genomes, has changed over evolutionary times. To understand which factors might affect the replicative success of L1 families, we analyzed the molecular evolution of the L1 families that amplified in the human lineage during the last 70MY. Three clearly distinct lineages of L1 families emerged before the origin of primates. These three lineages were simultaneously active until 43 MY ago when two of them became extinct. Their extinction coincided with the acquisition of a novel type of regulatory sequence in the lineage that persisted. We propose that the frequent replacement of regulatory regions during L1 evolution can explain changes in the rate of L1 amplification.

BOORE, Jeffrey

Bizarre Observations in Mitochondrial Genomics

Boore J. L. [1], Fong J. [2], Helfenbein K. [2], Lindberg D. [3], Macey J. R. [2], Masta S. [2], Medina M. [2], Mueller R. [3], Papenfuss T. [3], Simison B. [3]. (1) DOE Joint Genome Institute and UC Berkeley, (2) DOE Joint Genome Institute, (3) UC Berkeley

We've determined complete mitochondrial genome sequences for many organisms. Current observations contrast with early-established dogma, including: (1) mtDNAs missing multiple genes in chaetognaths and limpets; (2) double mtDNAs in nemertines; (3) duplications in amphisbaenians, salamanders, and others that can evolve in concert; (4) gene rearrangements that vary greatly in rate among lineages; (5) a homoplasious rearrangement shared between birds and some reptiles; (6) tRNA genes that are heavily edited in centipedes and spiders; (7) many gene junctions without a tRNA gene or secondary structure as is thought to be necessary for transcript processing; (8) split rRNA genes, so co-regulation by the vertebrate model is impossible. We can deduce modes of gene rearrangement to estimate their utility as phylogenetic characters. Much remains unknown about the evolution of these diminutive genomes, but developing software and databases promise to facilitate comparisons of mtDNA features.

BRAASCH, Ingo

Genome evolution of cichlid fishes - insights from receptor tyrosine kinases

Braasch I. [1], Salzburger W. [1], Meyer A. [1]. (1) Department of Biology, University Konstanz

The East African cichlid fishes are outstanding examples of adaptive radiation and explosive speciation. The evolutionary success of this group is mainly based on its diversity of trophic morphology and color patterns. Craniofacial cartilage, which builds the trophic apparatus, as well as pigment cells are of neural crest cell origin. We chose a candidate gene approach to investigate the genetic basis of cichlid coloration. The receptor tyrosine kinases kit and fms play crucial roles in the development of zebrafish coloration. We isolated and shotgun sequenced a 150 kb fragment from a cichlid BAC library that spans the fms gene locus. Like in pufferfish, the fms gene is linked to PDGFR-beta, another receptor tyrosine kinase. Additionally, we show that the cichlid fms gene is duplicated, most likely due to the postulated fish specific genome duplication.

BRENNER, Steven

Phylogeny of Ancient Proteins Ywis Reconstructed Using Structure (PAPYRUS)

Hill E. E. [1], **Brenner S. E.** [1]. (1) University of California, Berkeley

Traditional molecular phylogenies use nucleotide or protein sequences, but most homologous proteins share no significant sequence similarity. We are therefore reconstructing phylogenies of very distantly related proteins using features from protein structure, which is more conserved over time. Structural data have been growing exponentially for decades, with 54,745 domains known today.

We analyzed structures of homologous proteins lacking sequence similarity to identify informative characters. For example, a study of 4-helical cytokines elicited helix size, linker topology, and receptor as key features. The CAPER database contains these ancient relationships, and those found by collaborators. We are formalizing and constructing evolutionary models for dozens of knowledge-based characters, as well as semi-random features identified by feature-selection algorithms. These characters and their evolutionary models govern selection of appropriate phylogenetic reconstruction algorithms.

BRISCOE, Adriana

The spectrum of human rhodopsin disease mutations through the lens of interspecific variation

Briscoe A. D. [1], Gaur C. [2], Kumar S. [2]. (1) University of California, Irvine, (2) Arizona State University

Rhodopsin mutations account for a large fraction of genetic changes underlying the human eye diseases, Retinitis Pigmentosa (RP). The availability of rhodopsin sequences from a large number of vertebrates has allowed us to investigate factors important in the development of RP by contrasting interspecific differences with RP disease mutation data. We find that rhodopsin disease mutations are overabundant in highly conserved sites and that amino acids with any potential of variability among vertebrates are likely to harbour disease mutations less frequently. At any site in rhodopsin, the set of disease-associated amino acids does not show any commonality with the set of interspecific amino acids. Disease mutations are biochemically four times more radical than the interspecific variation. This pattern is also observed when disease mutations are categorized based on biochemical, physiological and psychophysical traits such as protein folding, electroretinogram amplitude, and equivalent field diameter. We introduce the concept of the expected chemical severity based on the normal human codon at a position. Results reveal that the analysis of disease mutations in the context of the original codon is very important for the practical application of evolutionary principles when comparing original and disease amino acid mutations.

BROOKS, Dawn

Inferred Thermophilic Amino Acid Compostion of Proteins in the Last Universal Ancestor of Life

Brooks D. J. [1]. (1) Washington Unversity School of Medicine

The Last Universal Ancestor (LUA) of all living organisms existed relatively close in time to the origin of

life itself, so its characteristics might provide some clues to the process by which life arose. For this reason, evidence regarding whether the LUA lived in a moderate or hot environment is of significant interest. The amino acid composition of proteins is informative, since this feature differs in thermophiles and mesophiles. We used expectation maximization to estimate the amino acid composition of a set of thirty-five proteins in the LUA, based on alignments of their modern day descendants, a phylogenetic tree relating those descendants, and a model of evolution. Irrespective of whether sequences of mesophiles or thermophiles were used in the alignments of modern day proteins, estimates of amino acid composition in the LUA were more similar to that of modern day thermophiles. These results support the proposal that the LUA was thermophilic.

BROWNING, Teena

Life or death - is it in the genes? Infectious disease, the MHC and genetic diversity

Browning T. L. [1], Zenger K. R. [2], Eldridge M. D. [1]. (1) Macquarie University / Australian Museum, (2) University of Sydney

Overall genetic diversity and/or possession of particular alleles is theorised to play a major role in selection and fitness, for example infectious disease susceptibility. Such hypotheses are difficult to evaluate quantitatively with direct measures of disease-related mortality.

The tammar wallaby (*Macropus eugenii*) is a medium sized macropod found in southwestern Australia. Tammar sudden death syndrome (TSD) is an orbivirus known to attack tammar populations with an ~50% morbidity rate. Disease onset is undetectable until the animal collapses and death rapidly follows. In the spring of 1998 an outbreak of TSD caused an epidemic in a captive colony of tammars. Of 80 animals, mortality was ~50%. Tammar wallaby MHC class II genes (DAB and DBB) have been characterised. Characterisation of class I genes is in progress. I will discuss the correlation between microsatellite diversity, MHC diversity and TSD epidemic victims and survivors. Microsatellite loci data suggest a correlation between genetic diversity and disease susceptibility.

BRUNNER, Stephan

Comparative Sequencing of Maize: Structure and Origin of Allelic Non-Colinearities

Morgante M. [1], Jung M. [2], Jung M. [2], Tingey S. V. [2], **Rafalski J. A.** [2]. (1) University of Udine, Udine, Italy, (2) DuPont Co. Crop Genetics

Two maize inbred lines, McC and B73, have been recently compared over more than 100 kb at the bronze1 genomic region (Fu H. and Dooner H.K., Proc Natl Acad Sci USA 2002, 99: 9573-9578). Contrary to expectations, the DNA sequences of the two inbreds not only differ extensively in the repetitive DNA segments, but also in some cases genes present in one allele are found missing in the other allele. Similar results were reported at another locus (Song R. and Messing J., Proc Natl Acad Sci USA 2003,100: 9055-9060). We set out to extend these observations to other genetic loci and survey the occurrence of genic and intergenic non-colinearities in maize gene pool. Four genomic segments of ca 250 kb each have been sequenced from two inbred lines, B73 and Mo17. Intergenic non-homologies were common and frequently consisted or relatively recent (<1my) insertions of retrotransposon-derived sequences. The differences between the two alleles ranged from 20% to nearly 50%, depending on the locus. Genic non-homologies were relatively rare and were usually pseudogenes. At one locus, Adh1, both alleles were nearly identical, except for SNPs and indels, which were present at normal frequency.

BULAZEL, Kira

Expanded centromere composition of the sex chromosomes of the marsupial *Macropus rufogriseus* (red-necked wallaby)

Bulazel K. [1], Ferreri G. C. [1], Metcalfe C. J. [1], Eldridge M. D. B. [2], O'Neill R. J. [1]. (1) University of Connecticut, (2) Macquarie University, Sydney Australia

The family Macropodidae (kangaroos and wallabies), one of the largest in the Marsupial clade, has 45 species ranging in diploid chromosome number from 2n=10/11 to 22. In this family the centromeres are

known sites of chromosome breaks and recombination, participant in the evolution of the karyotypic diversity across the clade. *Macropus rufogriseus* (red-necked wallaby) have unusually lengthened centromeres that are up to half the length of the chromosome. These centromeric regions of the X chromosome are heterochromatic and satellite and repeat rich. Though the centromeric regions are enlarged, CENP-E immunofluorescence restricts the functional centromere to a point localization within the larger region. Microdissection and microcloning the *M. rufogriseus* X centromere yielded three types of sequences present in the region. The resultant sequences were analyzed by Southern hybridization to assess presence and quantity, and by fluorescence *in situ* hybridization (FISH) to localized sequences to their specific chromosomal locations.

BURK-HERRICK, Angela
Molecular Evolution of BRCA2 exon 11 in Mammalia: A comparative phylogenetic approach
Burk-Herrick A. [1], Springer M. S. [1]. (1) University of California, Riverside
A large proportion of inherited breast cancers are due to mutations in the breast cancer susceptibility gene, BRCA2. In addition, the BRCA2 pathway recently has been linked to sporadic breast and ovarian cancer through a BRCA2 transcriptional repressor protein. The role of BRCA2 in both familial and sporadic breast and ovarian cancer provides a big incentive to understanding the gene's structure and function. BRCA2 encodes a very large protein that is involved in homologous recombination repair of DNA damage. BRCA2 has been shown to interact with RAD51 via its BRC repeats of exon 11 and a C-terminal domain. To define sites and regions of sequence conservation and elucidate functionally important sites, a portion of BRCA2 exon 11 was sequenced for a diverse array of mammalian taxa. Homologous alignments were used to generate phylogenetic trees, to identify conserved sites and sites undergoing positive selection, and to determine which missense mutations catalogued in the Breast Cancer Information Core (BIC) are most likely to increase oncogenic risk.

BUSTAMANTE, Carlos D.
Optimal Tests for Inferring Natural Selection from Comparative Population Genomic Data
Bustamante C. D. [1]. (1) Cornell University
The McDonal-Kreitman test (1991) provides an elegant and robust approach for identifying genes and genomic regions that may be the subject of adaptive, balancing, or weak negative selection. Recently, a family of related approaches have emerged for combinining genomic McDonald-Kreitman data for a pair of species (e.g., Bustamante et. al, 2002; Fay, Wycoff, and Wu 2002; Smith and Eyre-Walker, 2002; Sawyer et. al 2003; Eyre-Walker, 2004). There is general interest in the community in understanding the similarities and differences among the approaches. Using novel theoretical results as well as Coalescent simulations and forward simulations, we investigate the power, robustness, and accuracy of these approaches under a variety of models for the distribution of selective effects among new mutations as well as assumptions regarding linkage. We demonstrate that the reduction in effective population cased by natural selection at linked sites (e.g., Hill and Roberston, 1966; Birky and Walsh, 1988; Charlesworth, 1995) depends greatly on the mean and variance of the distribution of selective effects *among newly arising mutations* , and that this distribution can have profound and unintuitive effects on the expected McDonal-Kreitman cell entries and MK test results. Likewise, we demonstrate that the estimate of the selection coefficient based on the Sawyer and Hartl (1992) parameterization of the MK cell entries is close to the mean selection coefficient for *fixed differences*. We address how to estimate the mean and variance of selective effects from MK table data and, thus, how to estimate the proportion of fixed differences between a pair of species that are adaptive (and confidence intervals on this quantity). We also address various issues related to the estimated direction and magnitude of selection from MK table data when the population has experienced a recent change in population size.

CAETANO-ANOLLES, Gustavo
Universal sharing patterns in the evolution of proteins

Caetano-Anolles G. [1], Caetano-Anolles D. [2]. (1) University of Illinois, Urbana, Illinois, (2) VitalNRG, Knoxville, Tennessee
Protein evolution is imprinted in both the sequence and structure of evolutionary building blocks known as protein domains. These domains share a common ancestry and can be unified into a comparatively small set of folding architectures, the protein folds. We have traced the distribution of protein folds between and within proteomes belonging to Eucarya, Archaea and Bacteria along the branches of a universal tree of protein architecture reconstructed from fold-usage data. We found that folds shared by the three organismal domains were placed almost exclusively at the base of the rooted phylogeny and that there were clear evolutionary patterns related to protein architecture and organismal diversification. These include a relative timing for the rise of prokaryotes, congruent episodes of architectural loss and diversification in Archaea and Bacteria, and a late and quite massive rise of architectural novelties in Eucarya perhaps linked to multicellularity. These patterns provide a unique insight on deep evolutionary phenomena that relate to the origins of life.

CAI, Jing

Dependence of protein evolutionary rate on gene lineage specificity in ascomycetes

Cai J. J. [1], Woo P. CY. [1], Lau S. K. P. [1], Yuen KY. [1]. (1) Department of Microbiology, University of Hong Kong
We report a new highly statistically significant predictor of protein divergence rate, lineage specificity of gene, which decrypts the phyletic distribution of orthologs of a gene in genomes studied. Highly lineage-specific genes are narrowly distributed among less species along the lineage. Using the complete sets of protein-coding genes from six Ascomycetes, we predicted *A. nidulans-A. fumigatus* orthologs and *S. cerevisiae-C. albicans* orthologs and classified them into several lineage specificity classes, then compared the average nonsynonymous substitutions rate, *Ka*, among different classes. We found that highly lineage-specific genes evolve faster than do those more broadly phyletically distributed genes (Mann-Whitney *U* test, $P<0.001$). The correlation between the protein's divergence rate and the lineage specificity is stronger than that for other factors, such as, gene expression level and gene dispensability (Backwards stepwise regression standardized coefficient BETA(Lineage specificity)>=0.403; whereas BETA(Other factors)<=0.209). In conclusion, lineage specificity is a more direct reflection of gene's evolutionary importance.

CALTEAU, Alexandra

Horizontal transfer of two operons coding for hydrogenases between bacteria and archaea

Calteau A. [1], Gouy M. [1], Perrière G. [1]. (1) BBE - UMR CNRS 5558
The existence of massive gene transfers between hyperthermophilic bacteria and archaea is very controversed. To question the existence of these transfers, we decided to buid phylogenies using the homologous gene families of the database HOBACGEN-CG (86 complete prokaryotic genomes : 70 bacteria and 16 archaea). Preliminary results revealed three horizontal gene transfers between bacterial and archaeal species involving large clusters of genes. All of these transfers affected operons coding for multisubunit membrane-bound [NiFe]-hydrogenases involved in the energy metabolism of the donor genomes. One transfer is related to a 13-genes operon, called *mbx*, which probably arose in the genome of *Thermotoga maritima* from a species belonging to the *Pyrococcus* genus. The two others implied an operon of six genes, called *ech*, transferred independently to the genome of *Thermoanerobacter tengcongensis* and *Desulfovibrio gigas*, from a species belonging to the *Methanosarcina* genus.

CARBONI, Andrea

Preliminary study of *Phaseolus vulgaris* L. evolution through non-TIR NBS LRR Resistance Gene Homologues analysis

Carboni A. [1], Del Bianco F. [1], Ranalli P. [1]. (1) Istituto Sperimentale per le Colture Industriali
Six different genotypes of *Phaseolus vulgaris* L. were examined with degenerate PCR primers designed to

amplify the Nucleotide Binding Site domain of the non-TIR NBS Leuicine Rich Repeat subfamily: two cultivars (BAT93 and Jalo EEP558), parentals of the most saturated genetic map of this crop and representative of the two majors domestication centers (respectively Mesoamerica and the southern Andes), and four accessions representing geographical regions from Mexico to Central America and Peru showing different rates of resistance to root knot nematodes (*Meloidogyne* spp.). More than 300 clones were sequenced and some interesting and previously never identified cluster of RGHs within non-TIR subfamily were recognized. The presence of these new clusters of highly diverged non-TIR NBS sequences suggests that multiple ways of sequence dispersion are occurring in common bean. Our preliminary results tried to relate the specific evolutionary dynamics of this gene family with gene pools evolution.

CARNEY, Susan L.

Expression and molecular population genetic analysis of hemoglobin subunits in the deep-sea hydrothermal vent tubeworm, *Ridgeia piscesae*

Carney S. L. [1], Fisher C. R. [1], Schaeffer S. W. [1]. (1) The Pennsylvania State University

Ridgeia piscesae is a morphologically diverse species of tubeworm. The varied phenotypes are genetically indistinguishable and inhabit environments that differ in parameters such as temperature, oxygen and sulfide. We chose hemoglobin, a carrier of oxygen and sulfide, as a candidate molecule for studies of gene expression and population genetics in two of the most extreme *R. piscesae* morphs. Hemoglobin appears to directly link the animal to its environment because there are phenotype-specific differences in its structure and function, suggesting that this molecule may play a role in generating phenotypic plasticity in this species. Current studies will examine DNA sequences of the genes encoding the four different globin subunits to detect morph-specific variation. Quantitative RT-PCR studies are in progress to determine if morph-specific variation exists in the levels of expression of each globin subunit.

CATANIA, Francesco

Evidence for an ongoing local selective sweep in a European *D. melanogaster* population

Catania F. [1], Schlötterer C. [1]. (1) Institut für Tierzucht und Genetik

Recent microsatellite studies of European *D. melanogaster* reveal low, though significant, levels of population genetic differentiation. This pattern could be either due to shared ancestry or a result of high levels of gene flow. Here, we describe a pronounced differentiation between European populations for one microsatellite locus; one allele was present at a frequency of 0.16 in a population from The Netherlands, but a frequency of 0.0031 was observed in 13 additional European populations (320 individuals) further examined. A combined DNA sequence and microsatellite analysis indicated that this population specific allele was contained in a genomic region that was identical by descent over at least 620kb. These data suggest (ongoing) positive local selection driving a linked neutral allele to a higher frequency.

CAVALCANTI, Andre

Life without non-coding DNA: Analysis of a pilot genome of *Oxytricha trifallax*'s macronucleus

Cavalcanti A. R. [1], Weitz J. S. [1], Landweber L. F. [1]. (1) Princeton University

Ciliates are an unusual group of microbial eukaryotes. They all possess two types of nuclei - a somatic macronucleus (MAC), responsible for most RNA synthesis, and a germline micronucleus (MIC), responsible for genetic exchange in sexual reproduction. Massive DNA cleavage, rearrangement and amplification are required to form the new MAC genome from the MIC after sexual exchange. In *Oxytricha trifallax*, after all DNA processing, the newly formed MAC is almost devoid of non-coding DNA, and is composed of about ~24,000 kinds of gene-sized molecules (individual coding regions packaged with their regulatory sequences and telomeres), each ~2kb long and present in ~1000 copies. Scrambled genes, assembled from disperse gene-fragments in the MIC, are an added layer of complexity to spirotrich genomes.

We will present results obtained from a pilot genome sequencing project, and discuss a possible mechanism for the evolution of such a streamlined somatic genome.

CHAIX, Raphaëlle
Lineages, clans, tribes: mythological or genetic entities?
Chaix R. [1], Quintana-Murci L. [2], Khegay T. [3], Jacquesson S. [4], Austerlitz F. [5], Heyer E. [1]. (1) Musée de l'Homme, Paris, France, (2) Institut Pasteur, Paris, France, (3) Academy of sciences, Uzbekistan, (4) IFEAC, Uzbekistan, (5) CNRS, Université Paris-Sud, France
Traditional societies are often organized into descent groups called lineages, clans and tribes. To test the hypothesis of common ancestry within these descent groups, we compared ethnological and genetic data in five patrilineal populations from Central Asia. We show that people from the same lineage and/or clan do share a recent common ancestor, whereas no such common ancestry is observed at the tribe level. In contrast to traditional beliefs, a tribe might be a conglomerate of clans from diverse origins who subsequently reinvented a hypothetical ancestor to strengthen group unification.

CHALKIA, Dimitra
Evolution of Formins in Eukaryotes
Chalkia D. [1], Nikolaidis N. [1], Makalowski W. [1]. (1) IMEG, Biology Dep. Pennsylvania State University
Formins are multi-domain proteins defined by strongly conserved formin homology 2 (FH2) domains. In eukaryotes formins play key roles in actin-related cellular processes. Although there are many experimental studies on this protein family, the evolutionary relationships among its members are largely unknown. Here, we investigated the evolution of formins from several eukaryotic genomes.
Our analyses suggested that: a) There are seven phylogenetic groups of formins in vertebrates, six in invertebrates, two in fungi, two in plants, at least two in slime mold and one in lower eukaryotes (Entamoeba, Giardia). b) The diaphanous (DIA) group and the disheveled associated activator of morphogenesis group (DAM) are clustered together and also are the capuccino (CAPU) and formin homology 2 domain containing 1 (FHOD1) groups. c) In both Arabidopsis and rice two evolutionary distinct groups exist suggesting that they emerged before the separation of monocots/eudicots. d) Two slime mold formins cluster with the metazoan CAPU group, suggesting that this group has been diverged very early in the evolution of eukaryotes.

CHENG, Chris
Diversity and molecular evolution of antifreeze proteins
Cheng C. [1]. (1) University of Illinois
Antifreeze (AF) proteins are a group of structurally disparate but functionally identical novel molecules, with the unique ability to bind to ice crystals and inhibit ice growth. They have evolved in some polar marine fishes and land insects, allowing them to avoid freezing in extreme cold. Other AF-like proteins are found in freeze tolerant organisms where they limit ice expansion and tissue damage. The diverse AF structures bespeak multiple evolutionary origins (inferred precursors include lectins, enzymes and blood complement), and processes of parallel and convergent evolution. Additionally, the molecular mechanisms that generated AF genes encompass more than the customary duplication of a pre-existing gene. De novo duplications of a 9-nt element in the ancestral gene created the new coding region for the Antarctic fish AF glycoprotein. The AF glycoprotein that evolved independently in the northern codfish appeared to have arisen from entirely non-coding DNA, dispensing with the need for a pre-existing protein coding gene.

CHO, Soochin
Testing the balancing selection hypothesis on the complementary sex-determination gene of honeybees
Cho S. [1], Huang Z. Y. [2], Zhang J. [1]. (1) Dept of EEB, University of Michigan, (2) Dept of Entomology,

Michigan State University
Honeybees are haplodiploid organisms, with haploid males and diploid females. In *Apis mellifera*, the western honeybee, heterozygotes at the complementary sex determination (*csd*) locus are females, hemizygotes are males, and homozygotes become sterile males. This sex determination mechanism generates strong selection against homozygotes at *csd*, which should result in balancing selection on the gene. To test this hypothesis, we characterized the genetic diversity of csd in *A. mellifera* and its sister species *A. cerana*. Our results show (1) high levels of nucleotide polymorphism within species, (2) long persistence of polymorphic alleles, and (3) elevated rate of nonsynonymous substitutions, all supporting the balancing selection hypothesis.

CHO, Soochin

Frequent Loss of Introns During Nematode Evolution

Cho S. [1], Ellis R. E. [2]. (1) Dept of EEB, University of Michigan, (2) Dept Mol Biol/UMDNJ School of Osteopathic Medicine
Since introns were discovered 26 years ago, people have wondered how changes in intron/exon structure occur, and what role these changes play in evolution. To answer these questions, we are studying five sets of genes in nematodes related to *C. elegans*. Our data indicate that nematode introns are lost at a very high rate during evolution, almost 400-fold higher than in mammals. These losses do not occur randomly, but instead favor some introns and do not affect others. By contrast, intron gains are far less common than losses in these genes. Based on the sequences at each intron site, we suggest that several distinct mechanisms can cause introns to be lost, and that the total rate of loss during evolution is the sum of their individual contributions. In theory, the small size of *C. elegans* introns should increase the rate at which each of these types of loss can occur, and might account for the dramatic difference in loss rate between nematodes and mammals.

CHUNG, Wen-Yu

A Perl Module for Molecular Evolution

Chung W. [1], Nekrutenko A. [2]. (1) Center for Comparative Genomics and Bioinformatics, (2) Department of Biochemistry and Molecular Biology
Algorithms for the detection of natural selection from sequence data are routinely used in comparative genomics. Despite a large number of elegant and powerful implementations in the framework of molecular evolution, few evolutionary tools work well in concert and even fewer are suitable for large-scale analysis. Our goal is to change this situation by providing a set of PERL extension modules providing straightforward programming interface for major molecular evolutionary methods. The implementation is self-contained - one will only need to download and install our extension modules. This gives user a greater flexibility in developing truly portable software without losing the speed on critical computational tasks.

CLARK, Andrew

Single Nucleotide Polymorphisms in Human Populations

Clark A. G. [1]. (1) Cornell University
Single nucleotide polymorphisms provide an easily scorable set of markers that distinguish the genomes of individuals. In humans, SNPs allow inference about many features of our past evolutionary history. Most human SNP studies identify SNPs from a heterogeneous set of previous studies, so the first problem that needs to be tackled is to model and correct for the biases introduced by this ascertainment scheme. Methods for doing this ascertainment correction are now well established, and require only a careful specification of the scheme for initial discovery of the SNPs used in each study. We will then review efforts to infer past demographic history, past rounds of natural selection, and the role of variation across the genome in local recombination rate on the frequency spectrum and linkage disequilibrium patterns of

human SNPs. As simple as the problem of disease association may seem at first, all of these factors may impact the type I and type II error rates of tests of association.

CORDAUX, Richard

Amplification dynamics of Alu elements in the human genome

Cordaux R. [1], Hedges D. J. [1], Xing J. [1], Batzer M. A. [1]. (1) Louisiana State University

Alu elements make up more than 10% of the human genome. Their impact on the genome has been substantial during human evolution and it is therefore important to understand how these mobile elements spread within their host genomes. We analyzed the sequence variation of human Alu elements to investigate the relationships and expansion patterns of Alu subfamilies. Hence, we demonstrate, using a network phylogenetic approach, that Alu subfamilies do not follow a single "master gene" model of expansion. Rather, we find that Alu subfamilies typically contain 10-15% of secondary source elements that contributed as much as 20-30% of all subfamily members. In addition, mismatch distributions indicate that Alu subfamilies did not follow a uniform dynamics of expansion; we used the properties of these distributions to estimate when Alu subfamily expansions took place during human evolution. Our results contribute to a better understanding of how Alu elements could expand to more than one million copies in the human genome.

CRACRAFT, Joel

A Molecular Time-Scale for Avian Diversification

Cracraft J. [1], Barker F. K. [2]. (1) American Museum of Natural History, (2) James Ford Bell Museum of Natural History

Debate rages on the temporal pattern of diversification of modern birds (Neornithes). The fossil record has been interpreted as supporting the hypothesis that nearly all major lineages arose after the Cretaceous-Tertiary (KT) extinction event, whereas sequence data and molecular clock analyses have generally supported a pre-KT origin of these lineages. Empirically the debate has been unconstrained by lack of robust phylogenetic understanding, poor calibrations, or by inappropriate sequence data and time-estimation methods. We will examine new advances in avian relationships, calibrate the temporal history of clades using paleontological and biogeographic data, and then estimate nodal divergences using nuclear gene sequences and methods not dependent on clocklike behavior. Our findings suggest that multiple avian lineages began diversification prior to the KT boundary.

CRAWFORD, Douglas

Microarray Studies on the Functional Importance of Quantitative Variation: most etabolic genes are different and correlated with function

Oleksiak M. F. [1], Roach J. L. [2], **Crawford D. L.** [2]. (1) North Carolina State University, (2) University of Miami, RSMAS

Within and Among Populations there are considerable differences in gene expression (Oleksiak, et. al Nat. Gen. 32:261; 2002). To begin to understand the functional importance of this variation, microarray studies were used to investigate how the variation in gene expression covaried with phenotypic measures of metabolism. Microarray analyses of metabolic genes printed with a high-level of replication indicated that most genes have statistically significant difference in expression among individuals within a population. Most of this significant variation is correlated with one of four phenotypic measures and thus apparently is not spurious. Although much larger, the patterns of variation in gene expression is similar to patterns of protein polymorphisms, suggesting that the differences among individual is partially heritable. These data suggest that much of phenotypic variation is due to changes in gene expression.

CRAWFORD, Andrew J.

Phylogeny, Cenozoic biogeography, and multiple independent mitochondrial gene order rearrangements in Central American frogs of the genus *Eleutherodactylus*

Crawford A. J. [1], Smith E. N. [2]. (1) Smithsonian Tropical Research Institute, (2) University of Texas at Arlington

The genus *Eleutherodactylus* (Leptodactylidae) is currently the largest vertebrate genus, with over 860 species. This study represented the first phylogenetic investigation of the mainland component of this genus based on DNA sequence data. We sought to infer relationships among subgenera and among species groups within the Central American subgenus, *Craugastor*. We collected data from the nuclear gene, c-*myc*, and from the mitochondrial genome we sequenced the ND2-WANCY region including the origin of light strand replication. Phylogenetic results were broadly concordant between data sets and among inference methods. We used a Bayesian MCMC approach to disentangle evolutionary rates from divergence times. Our topological and temporal results support a previous model of three independent invasions by Eleutherodactylus into Central America. Our investigation also revealed at least six independent gene order rearrangements involving the ND2-WANCY region within this genus. Most involved duplication events and were autapomorphic among our sample of species. However, one rearrangement appeared to characterize a clade of species, and one duplication event was found only in one of four conspecific populations. Some rearrangements did not appeared to fit the model of tandem duplication and differential loss.

CREASE, Teresa

Insertion site variation in the DNA transposon *Pokey* in *Daphnia pulicaria*.

Crease T. J. [1]. (1) University of Guelph

Pokey is an autonomous DNA transposon of the *piggyBac* family that inserts into a specific TTAA site in the rDNA of *Daphnia*. All other characterized rDNA elements are non-LTR retrotransposons. *Pokey* also differs from other rDNA elements because it inserts into many other genomic locations. We subcloned and sequenced the 3' end and flanking region of *Pokey* elements that were screened from a cosmid library of *Daphnia pulicaria*. Phylogenetic analysis of the sequences suggests that *Pokey* transposes between rDNA and other genomic locations. Moreover, *Pokey* elements have diverged into multiple subfamilies whose sequence divergence from one another varies from 2% to 6%. Sequence divergence between elements within subfamilies is less than 1%. Although *Pokey* inserts into a single TTAA site in rDNA, the sequences flanking genomic TTAA insertion sites are very diverse. Ongoing work aims to establish what determines insertion site specificity in this element.

CROUCH, Jo Anne

Phylogenetic estimation of adaptation and speciation in populations of the phytopathogenic fungus *Colletotrichum graminicola*

Crouch J. A. [1], Clarke B. B. [1], Hillman B. I. [1]. (1) Rutgers University

Colletotrichum graminicola is a haploid, filamentous, primarily clonally-reproducing fungus responsible for anthracnose disease in economically important grasses and cereals of the Poaceae family. To investigate the impact of sympatric population distribution and the potential for host-plant specialization in lineages of this fungus, an evolutionary hypothesis was constructed through maximum likelihood analysis of three concordant unlinked loci. At least two distinct sibling species corresponding to host-plant origin were revealed by the phylogenetic analysis. Nucleotide substitution rates consistent with positive adaptive selection at the fungal mating-type gene provided evidence of reinforcement acting to reproductively isolate these species as they emerged in host-range restricted ecological niches. This hypothesis of adaptive evolution suggests that sexual reproduction may have played an important role in ancestral *C. graminicola* populations.

CUI, Liying

Rapid Evolution of Yabby Family Transcription Factors in Angiosperms

Cui L. [1], Leebens-Mack J. [1], Ma H. [1], dePamphilis C. W. [1]. (1) The Pennsylvania State University

The YABBY family transcription factors contain a Zn finger domain and a YABBY domain. Members of

the YABBY family, including the Arabidopsis CRC gene, specify the abaxial-adaxial polarity of lateral organs. We surveyed YABBY genes in the complete genomes of *Arabidopsis* and rice, plant ESTs from dbEST, loblolly pine, fern, moss, green alga and ESTs generated by the Flower Genome Project. The distribution of YABBY genes is limited to angiosperms and gymnosperms. Phylogenetic analysis shows that this gene family underwent a radiation at the origin of flowering plants leading to four distinct lineages in monocots and eudicots. Whereas the phylogeny supports the orthology of *CRC* and the rice *DL* gene, functional divergence has been reported. The Nymphaeace ortholog to *INO* also suggests conservation in this lineage. Additional rounds of duplication are evident in the *YAB3/FIL* lineage. The implication of gene duplication and functional divergence will be discussed.

CUSACK, Brian

Alternative splicing and evolution of duplicated genes

Cusack B. P. [1], Wolfe K. H. [1]. (1) Trinity College Dublin

Gene duplication and alternative splicing (AS) are both important sources of protein sequence variation, but the evolutionary relationship between these two processes has not been explored. Studying AS requires extensive sampling of the transcriptome but new datasets such as the ASAP database of AS patterns in mammalian ESTs make it possible to study AS on a genome-wide scale. We examined the relationship between AS and gene duplication in human and mouse and find a significant negative correlation between gene family size and the frequency of AS (measured as the fraction of splice junctions that show AS), which is consistent with subfunctionalization of duplicated genes. In orthologous genes compared between human and mouse, the frequency of AS is lower in genes that show strong constraints on sequence (low Ka/Ks ratios). The same effect is seen in comparisons of paralogs within human. In addition, genes showing species-specific AS have higher Ka/Ks ratios than genes that show conserved AS patterns. These results indicate that AS can accelerate protein sequence evolution.

CUTTER, Asher

Sexual and temporal dynamics of molecular evolution in C. elegans development

Cutter A. D. [1], Ward S. [1]. (1) University of Arizona

In this study, we investigate the dynamics of gene evolution across Caenorhabditis elegans ontogeny and among genes expressed differentially between each sex and gamete type. Using comparative sequence analysis with C. briggsae, we measure protein sequence evolution for different gene classes identified by genome-wide gene expression developmental time series. We demonstrate that genes expressed predominantly after reproductive maturity evolve more rapidly than genes expressed earlier in development, consistent with the mutation accumulation model of aging where relaxed selection occurs post-reproductively. We also find that genes involved in spermatogenesis evolve faster than other phenotypic classes of genes, whereas male-related genes evolve slower. The slower rate of male-related gene evolution indicates that selection acts to maintain males in these androdioecious species, despite their rarity, and the rapid evolution of sperm genes suggests that sexual selection acts on sperm.

DAGAN, Tal

Unitary Pseudogenes in Bacterial Genomes: Lifestyle and Gene Loss

Dagan T. [15], Graur D. [1]. (1) Department of Biology and Biochemistry, University of Houston

Genome size in bacteria is affected by lifestyle. For example, parasitism invariably entails the loss of genetic function and a consequent reduction in genome size. Patterns of gene loss can be inferred from studies of unitary pseudogenes, i.e., pseudogenes devoid of functional paralogs. A unitary pseudogene indicates that a cellular function has been irreplaceably eliminated. A high frequency of unitary pseudogenes is expected in genomes of organisms that have become dependent on "the kindness of strangers." In this study, we examine the effects of lifestyle on the frequency of unitary pseudogenes in 140 fully sequenced bacterial genomes. The bacteria were classified according to lifestyle into free-living, facultative symbionts and obligate symbionts. The obligate symbionts were, in turn, classified into

commensals, mutualists, and antagonists (parasites). As expected, parasitic bacteria were found to have lost a considerable number of gene functions. In particular, genes encoding metabolic functions were found to be prone to pseudogenization. Surprisingly, the ratio of unitary pseudogenes to functional genes did not differ significantly between free-living and facultative symbiotic bacteria.

DAS, Jayatri

Structural Conservation of the Nuclear-Encoded Subunits of Cytochrome C Oxidase Underlies Evolving Functional Sites

Das J. [1], Miller S. T. [2], Stern D. L. [1]. (1) Dept. of Ecology & Evolutionary Biology, Princeton, (2) Dept. of Molecular Biology, Princeton

Interspecific sequence comparison can reveal regions of evolutionary conservation that are under purifying selection due to both functional and structural constraints. We have combined comparative sequence analysis with structural, biochemical and physiological data to investigate the evolution and function of the nuclear-encoded subunits of cytochrome c oxidase (COX). This enzyme is physiologically important as the primary determinant of cellular oxygen uptake in oxidative phosphorylation. By mapping positions of amino acids that are conserved across a wide range of taxa onto the structure of bovine COX, we show that conserved residues are spatially grouped into functional domains. These domains include some known functional sites, such as a vertebrate binding site for thyroid hormone that is surprisingly conserved in insects and yeasts, as well as other uncharacterized regions. Analysis of the biochemical nature and spatial distribution of conserved amino acids suggests that selection acts to maintain the structural foundation for active sites across taxa. To define the functions of conserved sites identified by our structural analysis, we are conducting experimental assays in *Drosophila* cell culture. Currently, we are studying interactions of juvenile hormone and the insulin receptor pathway with COX to understand physiological regulation of oxidative phosphorylation.

DEAN, Jed

Expanding the role of Reciprocal Hemizygosity Scanning (RHS)

Dean E. J. [1], McCusker J. H. [2], Davis R. W. [1], Steinmetz L. M. [3]. (1) Stanford University, (2) Duke University, (3) EMBL

Several genetic factors contributing to the high-temperature growth (Htg) phenotype observed in clinical isolates of pathogenic S. cerevisiae have been determined (Steinmetz et al). The study identified two intervals on chromosomes XIV and XVI that co-segregate with Htg+ and used a unique RHS approach to investigate the contribution of each ORF within the interval to Htg+. RHS uses hybrid strains built by crossing the clinical isolate (Htg+) with a laboratory strain (Htg-) and compares pair-wise the growth ability of a hybrid strain containing a deletion in the clinically derived allele to a hybrid strain containing a deletion in the reciprocal laboratory derived allele. This analysis has been applied to quantitative trait locus intervals underlying Htg, and we are now in the process of expanding RHS to the entire genome. Each of these RHS strains are marked with a unique 20-mer found within the deletion construct, so that after a pool of the RHS strains is subjected to a specific condition, the growth ability of each strain can be determined using the relative hybridizations of the barcodes to a oligonucleotide array. This novel approach will shed new light onto the structure of complex traits in other organisms.

DEHAL, Paramvir

Insights into the Evolution of Gene Structure using Whole Genome Sequence

Dehal P. S. [1], Boore J. L. [1]. (1) Joint Genome Institute

There has been much discussion on the role of gene structural changes -- the insertion, deletion, or rearrangement of exons and introns -- in the evolution of genes and gene function. Now that whole genome sequences are available from multiple vertebrates, we have undertaken a systematic analysis of all such changes across the complete sets of orthologous genes. This data set allows us to examine the mechanisms and estimate rates of change in intron/exon boundaries and to address the origins of novel

intron/exon sequences. In this manner, we can evaluate the relative role of exon shuffling and other mechanisms in the evolution of gene function. Most gene studies focus on the patterns of sequence change and do not address how structural changes influence the genes evolution. Gene structure evolution has potentially broad implications for the creation of novel gene function and the evolution of vertebrate genomes.

DELANEY, Theresa
Investigating the Influence of Genetic Code Evolution over Different Genetic Distances
Delaney T. K. [1], Freeland S. J. [1]. (1) University of Maryland Baltimore County
Previous research has shown that over short evolutionary distances, the genetic code restricts patterns of protein evolution, but over longer distances this restriction relaxes leaving the biochemical properties of amino acids to dominate substitution patterns. But this remains largely a qualitative (not quantitative) observation.

The vast amount of protein sequence data now available in public databases allows for a further quantification of this effect. By analyzing a larger set of protein sequences at a higher resolution of evolutionary distances, we will investigate how mutation biases combine with the genetic code and the biochemical similarity of amino acids to predict interconversion rates of amino acids. From this analysis, we aim to develop a predictive quantitative model of protein evolution as a further resource to improve the accuracy of major bioinformatics tools used for biological sequence analysis, such as sequence alignment and homology detection.

DENAWA, Masatsugu
Intron evolution in the P-type ATPase superfamily: A new property of conserved introns in the conserved domains
Denawa M. [1], Ohniwa R. L. [1], Okamura H. [2], Takeyasu K. [1]. (1) Graduate School of Biostudies, Kyoto University, (2) Osaka Dental University
The P-type ATPase is a multi-membrane-spanning protein that occurs through prokaryote to eukaryote, forming a superfamily with different substrate specificities. This superfamily possesses 8 conserved regions and can be a good model system to study the evolution of introns. The database search identified 1495 introns in 126 P-type ATPase genes. The degree of positional conservation throughout the P-type ATPase evolution enabled the introns to be classified into 6 distinct categories; Class 1 as the oldest to Class 6 as the newest. The numbers of Class1 and Class 2 introns are smaller and occur more frequently in the conserved regions. In contrast, the numbers of Class 4, 5 and 6 introns are larger and appear more frequently in the non-conserved regions. These results suggest that new introns are hard to be inserted into the conserved regions, but once inserted, they tend to remain more stably than those in non-conserved regions.

DENVER, Dee
The transcriptional consequences of mutation accumulation
Denver D. R. [1], Streelman J. T. [2], Kim S. [3], Morris K. [2], Lynch M. [1], Thomas W. K. [2]. (1) Indiana University, (2) University of New Hampshire, (3) Stanford University
To investigate the genome-wide transcriptional consequences of spontaneous mutations, we carried out a series of microarray experiments involving a set of long-term *Caenorhabditis elegans* mutation-accumulation (MA) lines and their common N2 (lab strain) ancestor. The MA lines were propagated across hundreds of generations as single, randomly selected hermaphrodites; this treatment severely reduced the efficiency of natural selection and ensured that all but the most deleterious mutations accumulated over time in an effectively neutral fashion. Coupled with the microarray data, direct estimates of the rate and molecular spectrum of spontaneous mutation in the *C. elegans* MA lines allowed us to estimate the fraction of spontaneous mutations that have transcriptional consequences. Microarray

experiments involving five *C. elegans* natural geographic isolates were also done to compare patterns of transcriptional variation in wild *C. elegans* populations to that observed in the MA lines.

DEPAMPHILIS, Claude

The Floral Genome Project: Origin and Diversification of Floral Regulatory Genes

dePamphilis C. W. [1]. (1) The Pennsylvania State University

The Floral Genome Project (FGP) was initiated to help understand the origin and diversification of floral regulatory genes and their functions. We have identified homologs of floral regulatory genes through large scale EST sequencing and positive screening of genes expressed during early flower development in key basal angiosperm, monocot, and eudicot lineages. Microarray and in situ hybridization are being used to determine their expression patterns. The results are being used to derive a consensus set of floral regulators, and to identify gene duplications that may have led to important innovations during floral evolution. Current results, based on > 50,000 EST sequences, suggest the ancestral angiosperm had a diverse toolkit of floral regulators similar to modern derived models. Studies of contrasting gene families (MADS, SKP1, and Yabby) as well as phylogenomic analyses of and expression patterns from model species will illustrate the major themes of conservation and divergence identified to date.

DERMITZAKIS, Emmanouil

Polymorphism, Divergence and Genomic Distribution of Conserved Non-Genic Sequences (CNGs) in Humans

Dermitzakis E. T. [1], Reymond A. [2], Kirkness E. [3], Ucla C. [2], Gagnebin M. [2], Excoffier L. [4], Rossier C. [2], Antonarakis S. E. [2]. (1) Univ. of Geneva and Wellcome Trust Sanger Institute, (2) Univ. of Geneva, (3) The Institute for Genomis Research, (4) Univ. of Bern

Analysis of the human and mouse genomes revealed a large number of conserved non-genic sequences (CNGs). Although their conservation is high enough to support their functionality, their role remains unknown. We have compared 1628 CNGs and 950 conserved exons of Hsa21 between human, mouse and dog. We also did a whole genome analysis of CNGs and exons. Finally, we have explored levels of nucleotide variation of 95 CNGs in 10 Europeans and 10 West Africans and compared with exons, introns, regulatory regions and random intergenic sequences. Our results suggest that CNGs have selective constraints independent of their genic environment, their distribution in the genome is negatively correlated with the exon distribution and their levels of nucleotide variation are extremely low, with an excess of rare alleles, indicative of strong negative selection. CNGs are elements that may participate in known and unknown functions of the genome and may harbor substantial functional variation.

DOHERTY, Mary

Molecular Markers for Assessing Community Diverstiy of Coastal Ciliates

Doherty M. [1], Costas B. [2], Snoeyenbos-West O. L. [1], McManus G. [2], Katz L. A. [1]. (1) Smith College, (2) University of Connecticut

Biogeographic distrbutions of microbial eukaryotes is widely debated. While some claim that microbial organisms have a cosmopolitan distribution, other data suggests evidence of endemism. To assess these hypotheses for the distribution of Choreotrich and Oligotrich ciliates, we analyzed coastal community samples taken from Long Island Sound. We characterized sequences for SSU rDNA, ITS DNA, and mitochondrial rDNA genes to assess gene flow across time and space. These data will serve as a basis for continued study of phylogeographic distribution of Choreotrich and Oligotrich ciliates.

DONIGER, Scott

Identification of functionally constrained and unconstrained cis-regulatory sequences in Saccharomyces species

Doniger S. W. [1], Fay J. C. [1]. (1) Washington University

The degree of functional constraints imposed on cis-regulatory elements and the rate of gain/loss of cis-regulatory elements between species is important to understanding their evolution. To estimate these quantities, we examined 161 known or predicted transcription factor binding motifs in 4188 orthologous intergenic alignments from the *S. cerevisiae*, *S. paradoxus* and *S. mikatae* genomes. We found ~63k motifs conserved at orthologous positions in all three species. The expected number of conserved motifs is only ~8k, calculated from the genome average synonymous substitution rate: 0.83 substitutions per site across the three species. Furthermore, we found that ~20% of conserved motifs were lost on either the *S. cerevisiae* or *S. paradoxus* lineage. We validated our results using experimentally verified motifs, expression data and function annotations of flanking genes.

DOOLITTLE, W. Ford
Tree of Life, Web of Life
Doolittle W. F. [1]. (1) Dalhousie University
There is now ample support in evidence for the notion that gene exchange (recombination and lateral gene transfer) within and between species is a major and possibly the dominant force in prokaryotic genome evolution. Rather than make this general case once again, I will address three relevant specific issues, using examples from our own research.
(1) Within-species gene content variation and recombination, and their implications for prokaryotic species concepts.
(2) The size and stability of the core of unexchangeable or rarely exchanged genes shared among species, and the use of this core to trace organismal phylogeny.
(3) Evolving definitions of "phylogeny" and the reality of the universal Tree of Life.

DRUZHININA, Irina S.
Resolving the molecular evolution of recently diverged fungal taxa using Bayesian phylogenetic inferences: application to the "harzianum/tomentosum" clade in *Trichoderma*
Druzhinina I. S. [1], Bissett J. [2], Kubicek C. P. [1]. (1) TU Vienna, Institute of Chemical Engineering, (2) Agriculture and Agri-Food Canada ECORC
Investigating the phylogeny of a so far unresolved clade of the fungal genus *Trichoderma* (the *T. harzianum*/*T. tomentosum* clade), we have characterized the evolutionary properties of three genomic regions (ITS1 and 2, large intron of *tef1alpha* and last large exon of *ech42*), selected the best model of evolution and estimated the usability of the combined three fragments data matrix for the phylogenetic analysis. Subsequently, based on well-sampled matrix (124 sequences of ITS1 and 2 and the large *tef1alpha* intron, and 64 *ech42* gene sequences) we have applied Bayesian inferences to resolve the phylogeny of the clade.
The resulting tree clearly confirms the phylogenetic position of *T. velutinum, T. cerinum* and *T. helicum* as well as *T. tomentosum* and *H. tawa*. Moreover we demonstrate the phylogenetic position of three potentially new undescribed species of *Trichoderma*. We also conclude that the small spored species without sterile conidiophore elongations (*T. harzianum* and *T. aggressivum*) evolved after those with typical â€œpachybasiumâ€□-like morphology. *T. harzianum*, the major taxon of this clade, formed a proliferated mega-clade whose monophyletic unity was not significantly confirmed while some morphologically indistinguishable subclades (=lineages) have already undergone explicit sympatric speciation, as was also supported by biochemical profiles using BIOLOG phenotype microarrays.

DYER, Kelly
Molecular evolution of X-chromosome drive in *Drosophila recens*
Dyer K. A. [1], Jaenike J. [1]. (1) University of Rochester
The loci underlying X-chromosome drive are not known in any species; however, studying regions tightly-linked to the driving gene(s) may allow inferences about the evolutionary history of the trait. In the mycophagous fly *Drosophila recens*, ~5% of males harbor driving X's and thus sire mostly daughters. In

addition, females homozygous for the driving chromosome are sterile, presumably due to deleterious factors closely linked to the driving gene(s). In a species-wide sample of standard and driving X chromosomes, we show that standard chromosomes exhibit high levels of polymorphism, whereas drive is associated with a single whole-chromosome haplotype that harbors essentially no polymorphism. We experimentally corroborate this apparent chromosome-wide suppression of recombination in driving X's, and use the polymorphism data to infer the evolutionary history of X-drive in *D. recens*. Our results demonstrate the effect of intra-genomic conflict on both current patterns of polymorphism as well as long-term genome evolution.

ELLIS, Ronald
Females, Hermaphrodites and Developmental Bias
Cho S. [1], Baldi C. [2], **Ellis R. E.** [2]. (1) Department of EEB, University of Michigan, (2) Dept. of Molecular Biology/UMDNJ School of Osteopathic Medicine
Although self-fertilizing hermaphrodites are ideally suited for colonization, species that contain hermaphrodites are common in nematodes, but absent from insects or vertebrates. To understand this developmental bias, we are studying *C. elegans* and its close relatives. Our phylogenetic data show that mating systems have changed multiple times during the evolution of this genus. What traits favor these changes? Hermaphrodites are essentially females that produce sperm as larvae. We show that *fog-1* and *fog-3* control spermatogenesis in both female and hermaphroditic species. However, *fog-3* transcripts are expressed during larval development in all *XX* hermaphrodites, but never in *XX* females. This expression is controlled by sex-determination genes, which appear to act on conserved binding sites in the *fog-3* promoter. We show that several aspects of this sex-determination process allow rapid changes in mating systems, and that simple genetic manipulations can replicate these changes.

EMERSON, J.J.
Extensive Gene Traffic on the Mammalian X Chromosome
Emerson J. J. [1], Kaessmann H. [2], Betrán E. [3], Long M. [1]. (1) Dept. of Ecology and Evolution, University of Chicago, (2) Center for Integrative Genomics,Univ. of Lusanne, (3) Dept. of Biology, University of Texas, Arlington
Mammalian sex chromosomes have undergone profound changes since evolving from ancestral autosomes. By examining retroposed genes in the human and mouse genomes, we demonstrate that, during evolution, the mammalian X chromosome has generated and recruited a disproportionately high number of functional retroposed genes, whereas the autosomes experienced lower gene turnover. Most autosomal copies originating from X-linked genes exhibited testis-biased expression. Such export is incompatible with mutational bias and is likely driven by natural selection to attain male germline function. However, the excess recruitment is consistent with a combination of both natural selection and mutational bias.

ENGELHARDT, Barbara
Protein function prediction using a Bayesian model of molecular function evolution
Engelhardt B. E. [1], Jordan M. I. [1], Brenner S. E. [1]. (1) U.C. Berkeley
We have constructed a statistical graphical model of molecular function evolution based on phylogenomic principles. Our model enables inference of molecular function for un-annotated protein sequences. By taking into account the full repertoire of reconciled phylogenetic relationships within a protein family, the model infers accurate annotations even with sparse and noisy data, validated through experimental evidence available in the literature.

We have assessed this method on several Pfam families, including deaminases and laminin B, where it provides results superior to widely used methods of function annotation. We have found that

incorporating Ohno's hypothesis that gene duplications promote neofunctionalization increases the accuracy of the annotation inference.

ERICKSON, Brian

Genomic Analysis of Transporter Proteins in the Malaria Parasite *Plasmodium falciparum*

Erickson B. K. [1], Wang X. [2], Sanchez M. [1], Wang Y. [1]. (1) University of Texas at San Antonio, (2) Astrazeneca Pharmaceuticals

Malaria is one of the most devastating infectious diseases in the world. Drug resistance in the pathogenic parasite Plasmodium falciparum is widespread and there is an urgent need for the development of new drug targets. Transporters are potential drug targets because they play important roles in the parasite life cycle, especially in the uptake of solutes and nutrients from the extracellular medium and in the disposal of metabolic wastes. However, the first-pass annotation of the P. falciparum genome only predicted 30 transporters. We used the specialized transporter databases TCDB and TransportDB as query databases for a mutual blast of the P. falciparum genome. The initial search returned over 100 hits covering a broad spectrum of transporter types. Some may perform crucial yet uncharacterized functions. Various mechanisms including gene duplication and lateral gene transfer are proposed to have contributed to the evolution of transport machinery in the parasite genome.

ESTEVES, Pedro

Allelic variation at the VHa locus in natural populations of rabbit (*Oryctolagus cuniculus*, L) from Iberian Peninsula

Esteves P. J. [1], Lanning D. [2], Zhai S. K. [3], Ferrand N. [4], Knight K. L. [3], van der Loo W. [1]. (1) CIBIO/UP, FCUP & I.M.B.B./VUB, (2) Department of Microbiology and Immunology, LUC , (3) Department of Microbiology and Immunology, LUC, (4) CIBIO/UP, FCUP

The large interallelic distances between the three rabbit Ig *VHa* lineages, *a1*, *a2*, and *a3*, have suggested that the persistence time of the *VHa* polymorphism could amount to 50 million years. Rabbit originated in the Iberian Peninsula where two subspecies coexist, one of which is confined to Southwestern Iberia (*O. c. algirus*). We studied the *VH* loci in the original species range. Serological surveys revealed that a majority of sera of the subspecies *algirus*, when tested with *VHa*-locus specific alloantisera, showed no reaction ("a-blank"). Using RT-PCR, we determined sequences of rearranged *VH* genes expressed in seven algirus rabbits that were typed as either "a-positive" or "a-blank". The data show that the *VH* genes transcribed in "a-positive" rabbits are closely related to one or the other of the *VH1* alleles of domestic rabbits. In contrast, "a-blank" rabbits were found to preferentially use *VH* genes that, although clearly related to the known *VHa* genes, define a new major allotypic lineage, designated *a4*. We conclude that at least four distantly related lineages of the rabbit *VHa*-locus exist, one of which seems to be endemic in the Iberian range.

EZAWA, Kiyoshi

Statistical analysis of gene conversion in mouse and rat genomes

Ezawa K. [1], Oota S. [2], Saitou N. [1]. (1) National Institute of Genetics, (2) RIKEN Bioresource Center

Gene conversion, also known as non-reciprocal recombination, is considered to play an important role in concerted evolution of multigene families. In order to figure out a genome-wide picture of gene conversion events, we examined mouse and rat genome data in the EnsEMBL database(http://www.ensembl.org). We sampled about 1,000 'quartet's each of which consisting of 2 mouse genes and 2 rat genes suspected to have duplicated before the mouse-rat speciation.
Two kinds of statistical tests indicate that approximately 15 percent of those quartets have undergone gene conversion. Statistical analysis uncovered that (i) in almost every quartet having expecienced gene conversion, the mouse gene pair resides within a chromosome and so does the rat gene pair; and (ii) the frequency of gene conversion correlates negatively with the physical distance of genes. We also analyzed how the frequency correlates with the relative orientation and the evolutionary distance of a gene pair.

FAIN, Matthew

Parallel radiations in the primary clades of birds, as inferred from intron 7 of beta-fibrinogen

Fain M. G. [1], Houde P. [1]. (1) New Mexico State University

Knowledge of avian phylogeny is prerequisite to understanding the circumstances and timing of the diversification of birds and the evolution of morphological, behavioural, and life history traits. Recent molecular data sets have helped to elucidate the three most basal clades of the tree living birds, but relationships among neoavian orders (the vast majority of birds) remains frustratingly vexing. Here, we examine intron 7 of the beta-fibrinogen gene from the majority of nonpasserine bird families and all orders. These data reveal that Neoaves consist of two sister clades with ecological parallelisms that may be analogous to those found between marsupial and placental mammals. Some members of the respective clades have long been recognized as examples of convergent evolution, but it was not appreciated that they were parts of diverse parallel radiations. In contrast, some traditional orders of birds are found to be polyphyletic, with representative families in both radiations.

FAY, Justin

Segregation of gene expression differences among recombinant strains of *S. cerevisiae*

Fay J. C. [1], McCullough H. L. [2], Sniegowski P. D. [3], Eisen M. B. [2]. (1) Washington University, (2) Lawrence Berkeley National Laboratory, (3) University of Pennsylvania

Gene expression differences may make a substantial contribution to phenotypic variation found in nature. An important question is whether most gene expression differences themselves have a complex or simple genetic basis and whether the expression differences are due to cis or trans-regulatory changes. To address these issues we examined gene expression differences among 12 F2 recombinant strains from a cross between a copper sulfate-sensitive, rust-colored strain and a copper sulfate-resistant, white-colored strain. From three replicate experiments we found segregation patterns consistent with both monogenic and polygenic inheritance. Expression differences which cosegregated with resistance to copper sulfate segregated as a monogenic trait. However, expression differences which cosegregated with rust coloration segregated as a polygenic trait.

FERDIG, Michael

Biological processes underlying complex multiple-drug resistance in the malaria parasite, *Plasmodium falciparum*

Ferdig M. T. [1], Patel J. [1]. (1) University of Notre Dame

The malaria parasite, *Plasmodium falciparum*, has thwarted all efforts to eliminate it with drugs. In fact, widespread, increasing levels of resistances to diverse pharmacologically active compounds is yielding a generalized multiple-drug resistant (MDR) phenotype in populations of *P. falciparum* from several regions of the world. The observed steady march towards MDR is clearly not the result of a simple, single-gene mechanism (e.g. an efflux pump). Past studies have emphasized the role of single-gene determinants of resistance; however, this focus necessarily has ignored the genomic context in which these genes function. Consequently, rather than searching for resistance genes, we are exploring the complex, multi-genic physiological processes that underpin decreasing sensitivity to various drugs in the progeny of a genetic cross. Alignment of QTL scans of response phenotypes to different drugs reveals shared loci that may account for observed correlations (cross-resistance patterns) between drug responses in the progeny and in natural populations. Using an integrative genomic approach that combines QTL with SNPs, microarrays, and population genetics, we have identified candidate genes in these loci that could highlight the biological processes that can contribute gene variants that influence resistance evolution.

FERRERI, Gianni

The Paradoxical Conservation of an Active Marsupial Retroelement

Ferreri G. C. [1], Marzelli M. [1], Eldridge M. D. B. [2], O'Neill R. J. [1]. (1) University of Connecticut, (2)

Macquarie University
We have investigated the impact, origin and distribution of a retroelement within marsupials affiliated with centromeric expansions. Analysis has shown that this retroelement exhibits a wide phylogenetic distribution, encompassing the macropodine superfamily. Presented is preliminary evidence that this retroelement is still actively expressed within the majority of macropodines, encompassing 24 million years of divergence. Results indicate that this retroelement sequence is highly conserved within many of the species examined and is most likely part of a tandemly repeated array. Sequence features of this retroelement indicate that it is derived from a retrovirus and expression analyses strongly suggest a recent infection of a common macropodine ancestor. Furthermore, preliminary results indicate this retroelement carries a conserved localization within host chromosomes across this group and may have contributed to the extensive karyotypic divergence observed in this group of mammals.

FESCHOTTE, Cedric
***Merlin*: a new superfamily of DNA transposons present in diverse animal genomes and related to bacterial IS1016 insertion sequences**
Feschotte C. [1]. (1) University of Georgia
Several new families of DNA transposons were identified by computer-assisted searches in a wide range of animal species, including nematodes, flat worms, mosquitoes, sea squirt, zebrafish and humans. Many of these elements have coding capacity for transposases, which are related to each other and to those encoded by the IS1016 group of bacterial insertion sequences. Although these transposases display a motif similar to the DDE motif found in many transposases and integrases, they cannot be directly allied to any of the previously described eukaryotic transposases. Other common features of the new eukaryotic and bacterial transposons include similarities in their terminal inverted-repeats and 8-bp target site duplications. Together, these data indicate that these elements belong to a new superfamily of DNA transposons, called *Merlin*/IS1016, which is common in many eubacterial and animal genomes. We also present evidence that some of these transposons have been recently active in several animal species.

FESCHOTTE, Cedric
Cross-mobilization as a mechanism contributing to the amplification of miniature inverted-repeat transposable elements (MITEs)
Feschotte C. [1], Osterlund M. [1], Peeler R. [1], Wessler S. R. [1]. (1) University of Georgia
Miniature inverted-repeat transposable elements (MITEs) are the predominant type of interspersed repeat in the non-coding regions of plant genes in addition to being abundant in many animal genomes. Although MITEs are structurally reminiscent of DNA-mediated nonautonomous transposons, their origin and mode of amplification are still poorly understood. *Stowaway* is a superfamily of MITEs that is widespread and prolific in plant genomes. Genome-wide sequence analysis in rice suggests a model whereby most of the ~25,000 *Stowaway* MITEs present in this species were trans-mobilized by transposases encoded by a small number (<50) of distantly related *mariner*-like elements (MLEs). To test this model, we have reconstructed two distinct, potentially active MLE transposases from rice and undertaken their functional characterization. Results gathered from in vivo and in vitro DNA-binding assays provide support for the idea that rice MLE transposases can indeed interact with multiple, distantly related rice transposons, including several *Stowaway* MITE families.

FIRESTEIN, Stuart
Vertebrate Olfactory Receptors
Firestein S. [1], Zhang X. [2], **Zhang X.** [1]. (1) Columbia University, (2) Salk Institute
The vertebrate olfactory system is arguably the best chemical detector on the planet. It is capable of detecting thousands of low molecular weight compounds and discriminating between chemicals of often very similar structure. Underlying this capability are three large families of G-protein coupled receptors (GPCRs) expressed by specialized sensory neurons lining an epithelium in the nose. The olfactory

receptors represent the largest family of GPCRs by more than an order of magnitude. In mouse this family consists of 1200 functional genes and a 250 pseudogenes, while in man there are 1100 genes, but only 350 of them appear to encode functional proteins. In most mammals there are two distinct olfactory systems, one devoted to the detection of environmental odors important in feeding and predation, and another concerned primarily with chemical signals between conspecifics that govern mating and social behaviors. One system may be seen as critical of the survival of the individual, the other for survival of the species. Each system has its own set of receptors. The phylogenetic organization of the gene families as well as their patterns of expression shed light on evolutionary strategies in chemosensory systems.

FRANCINO, Pilar

Applying the comparative method to bacterial genome evolution

Francino M. P. [1], Lee L. C. [2], Kobayashi A. [3], Boore J. L. [1]. (1) DOE Joint Genome Institute & Berkeley Natl. Lab., (2) Genome Sequence Center, BC Cancer Agency Vancouver, (3) DOE Joint Genome Institute & Livermore Natl. Lab.

By taking advantage of the availability of numerous and diverse bacterial genome sequences, we have obtained a well-resolved global bacterial phylogeny on which we are tracing the evolution of genome level characters and molecular evolution parameters. In order to estimate absolute rates of evolution, we are employing maximum likelihood to estimate the times of divergence among species in the absence of a general molecular clock, and we are employing the comparative method to detect correlations between characters independent of phylogeny. Our initial analyses show that the size and GC content of a bacterial genome are negatively correlated with the rate of evolution of its encoded proteins. The implications of these correlations for genome evolution are discussed, in a context of selection, mutational load and cell economics.

FRANKHAM, Richard

Comparative rates of loss of quantitative and molecular genetic variation in finite populations

Frankham R. [1], Gilligan D. M. [2], Briscoe D. A. [2]. (1) Macquarie, James Cook & Harvard Universities, (2) Macquarie University, NSW 2109, Australia

Quantitative genetic variation, the primary determinant of the ability to evolve, is expected to be lost in small populations, but there is little data about the effect, and controversy as to whether it is similar to that for near neutral molecular variation. Genetic variation for abdominal and sternopleural bristle numbers and allozyme heterozygosity were estimated in 23 populations of *Drosophila melanogaster* maintained at effective population sizes of 25, 50, 100, 250 or 500 for 50 generations, as well as in 19 highly inbred populations and the wild outbred base population. Highly significant negative regressions of genetic variation as a proportion of initial values on inbreeding due to finite population size were observed for both quantitative characters and for allozyme heterozygosity, and the regression coefficients did not differ significantly. Arguments that quantitative genetic variation is not lost in small populations are unwarranted.

FRASER, Hunter

Detecting selection using a single genome sequence

Plotkin J. B. [1], Dushoff J. [2], **Fraser H. B.** [3]. (1) Harvard University, (2) Princeton University, (3) UC Berkeley

Selective pressures on proteins are usually measured by comparing homologous nucleotide sequences. Here we introduce a method to estimate selection on the basis of a single genome sequence. We catalogue the relative strength of selection on each gene in the entire genomes of *S. cerevisiae*, *M. tuberculosis*, and *P. falciparum*. Our analysis confirms that our method is consistent with estimates of selective pressures found by traditional comparative methods. Using this method to study genome-wide selection pressures in these organisms, we confirm that most antigens are under strong selection for amino acid substitutions. We also identify many uncharacterized proteins, including "ORFans" with no known orthologs in other organisms, that are under strong selection. Our method of estimating selective

pressures requires far fewer data than comparative sequence analysis, and it measures selection across an entire genome; the method can readily be applied to a large range of sequenced organisms.

FRIEDMAN, Brad

Evolutionary Patterns of Intron Gain and Loss

Nielsen C. B. [1], **Friedman B.** [2], Birren B. [3], Galagan J. E. [3], Burge C. B. [4]. (1) MIT Biology, Broad Institute, (2) MIT Biology, MIT Mathematics, (3) Broad Insitute, (4) MIT Biology

A correlation between intron density and positional bias has long been recognized. Introns are evenly distributed within the coding sequences of genes in intron-rich genomes, but are biased towards the 5' ends of genes in intron-poor genomes. To investigate the dynamics of intron evolution, orthologous proteins from four fungi were aligned and the patterns and positions of intron conservation were analyzed. Approximately 200 gains and 200 losses were inferred in each lineage. Reconstruction of the ancestral positional intron distribution revealed that the fungal ancestor exhibited an even stronger 5' bias than is observed today, a bias that has partially
eroded over the past 200 million years. The forces that shape intron distribution are therefore likely to vary substantially over long
evolutionary time scales and to involve a complex interplay between
gain and loss.

GANESAN, Skanth

Comparative Analysis of Gene Prediction Methods and Development of a Fungal Genome Database System

Ganesan S. [1], Harris S. [2], Moriyama E. N. [3]. (1) Dept. of Computer Science, University of Nebraska Lincoln, (2) Plant Pathology & Plant Science Initiative, UNL, (3) School of Biological Sciences & Plant Science Init.

The complete genome sequences of several filamentous fungi have recently become available as part of the Fungal Genome Initiative. Multiple gene prediction programs are being used to identify coding regions within these genomes. Despite many limitations, existing gene prediction methods and gene structure models are often applied to new sequences for which no model or method has yet been tuned. Our objective is to analyze the available gene mining methods by assessing their prediction performance as well as their use of varied genomic information. We are developing an integrated genome database system that will facilitate the genome annotation of three filamentous fungi; Neurospora crassa, Aspergillus nidulans and Fusarium graminearum. Our preliminary analysis of the above genomes has already revealed that each possesses a surprisingly large number of predicted genes with no apparent homologue in any other organism, thereby highlighting the need for accurate gene prediction programs.

GANKO, Eric

A role for retrotransposons in D.melanogaster gene evolution

Ganko E. W. [1], Franchini L. F. [1], McDonald J. F. [1]. (1) University of Georgia, Genetics Dept.

LTR retrotransposons (LTEs) are common mobile repeat sequences that may contribute to gene evolution through built-in regulatory and coding signals. To better understand the contribution of LTEs to *D.melanogaster* gene evolution we searched upstream and downstream of each LTE for the presence of neighboring genes. In all, 33% of LTE sequences (228/682) are within 1 kb of a gene or within gene boundaries. Of these 228 LTEs, 146 are within genes, predominantly in introns. Several truncated LTEs associated with genes are present in *D.melanogaster* populations around the world. Full length LTEs are more often associated with genes but have limited population distributions, most likely due to recent insertion events. Functional analysis indicates LTEs preferentially associate with genes having signal transduction, behavior, and development function, and are underrepresented in physiological process genes. Our results indicate that some LTEs may contribute to gene evolution in *D.melanogaster*.

GASCH, Audrey

Coevolution of *cis*-regulatory elements and their DNA binding proteins in fungi

Gasch A. P. [1], Moses A. M. [2], Chiang D. Y. [3], Fraser H. B. [3], Berardini M. [4], Eisen M. B. [5]. (1) Laboratory of Genetics/Genome Center UW-Madison, (2) Program in Biophysics, UC-Berkeley, (3) MCB Department, UC-Berkeley, (4) Life Science Division, Lawrence Berkeley Lab, (5) MCB Department, UC-Berkeley; Lawrence Berkeley Lab

Relatively little is known about the mechanisms through which gene expression regulation evolves. We have systematically explored the evolution of regulatory sequences in fungi, using a method that is not restricted to genomes whose sequences can be aligned. Reasoning that genes that are coregulated in *S. cerevisiae* are likely to be coregulated in related species, we identified orthologs of coregulated *S. cerevisiae* genes in 13 other fungi and identified sequences that are enriched in the noncoding regions of the presumptively coregulated genes in each species. Our results indicate that many of the known *S. cerevisiae* regulatory elements are conserved in other species, in some cases in distant relatives of this yeast. Since our initial presentation of this work, we have characterized one set of sequences from proteasome genes in *S. cerevisiae*, *C. albicans*, and *N. crassa* by measuring the interactions between these sequences and the putative transcription factor cloned from each species. Our results indicate that the DNA-binding specificities of these transcription factors have diverged in concert with the sequences found upstream of the proteasome genes in *S. cerevisiae* and *C. albicans* and suggest that the *N. crassa* factor has a different cellular function.

GAUNT, Michael

The evolutionary implications of genetic exchange via genome fusion in Trypanosoma cruzi

Gaunt M. W. [1], Yeo M. [1], Frame I. [2], Stothard R. [3], Carrasco H. [4], Taylor M. [1], Solis Mena S. [1], Veazey P. [1], Miles G. A. [1], Acosta N. [5], Rojas de Arias A. [5], Mile M. A. [1]. (1) London School of Hygiene and Tropical Medicine, (2) The Wellcome Trust, (3) Imperial College, London, (4) Universidad Central Venezuela, (5) Universidad Nacional de Asuncion, Paraguay

Trypanosomatids (Kinetoplastida) are basal eukaryotes and with the exception of *Trypanosoma brucei* genetic exchange has been considered absent or cryptic. The mechanism of genetic exchange of *Trypanosoma cruzi* has recently been determined by observing extant genetic exchange between biological clones that were transfected with different drug-resistant markers. Analysis of the resulting double-drug resistant progeny showed the genetic mechanism is very distinct from the mendelian-based mechanism proposed for *T. brucei* and involves fusion of parental genotypes, allele loss and uni-parental inheritance of kinetoplast maxicircle DNA. Moreover, there are strong genetic parallels between the laboratory cross and the genotypes of natural isolates. Genome fusion has enabled genetic exchange between *T. cruzi* lineages which span millions of years of temporal divergence and evidence suggests it is not restricted to this species. This concept provides a new dimension to genome duplication, particularly as it operates close to the crown of life.

GERALDES, Armando

Contrasting patterns of introgression at X-linked loci across the European rabbit hybrid zone

Geraldes A. [1], Nachman M. W. [2], Ferrand N. [1]. (1) CIBIO - Universidade do Porto, (2) EEB - University of Arizona

Patterns of introgression at different genes following secondary contact between diverged lineages can provide insight into the genetic architecture of species differences. The European rabbit (*Oryctolagus cuniculus*) is composed of two distinct evolutionary lineages: *O. c. algirus* and *O. c. cuniculus*. These two groups show highly divergent mtDNA and Y chromosome haplotypes that diverged about two MYA, and these two groups overlap along a narrow geographical zone in the central Iberian Peninsula. Here, we surveyed nucleotide variability at three X-linked loci (*Smcx*, *Phka2*, and *Hprt*) representing a total of 7917bp. Forty three individuals were sampled from three geographic regions: north-east Iberia and France, south-west Iberia, and central Iberia. In addition two outgroup species (*Lepus spp.*) were

sequenced. *Smcx* lies near the centromere and had low levels of nucleotide variability, high levels of linkage disequilibrium and high levels of population differentiation, while *Phka2* and *Hprt* are telomeric and had higher levels of variability, less linkage disequilibrium, and little population differentiation. These observations are consistent with the hypothesis that genes in regions of little recombination (*Smcx*, Y chromosome, mtDNA) introgress less following secondary contact, possibly as a consequence of linkage to sites under selection.

GIBAS, Cynthia
GenoMosaic: On-Demand Multiple Genome Comparison and Comparative Annotation
Gibas C. J. [1], Sturgill D. [1]. (1) Department of Biology, Virginia Tech
The goal of the GenoMosaic project is to support new method development for on-demand multiple genome comparison. Each genome to be compared can be modeled as a string of generic features of any type that can be computationally defined, related by adjacency information within and among genomes. The generic feature abstraction makes it possible to study the arrangement of features in the genome at a level of detail which includes RNA genes, putative regulatory regions, SNPs, overlapping transcripts, intron splice junctions, alternative polyadenylation signals — in short, to incorporate significant sequence details which are not necessarily within protein-coding regions. This abstraction is amenable to functional implementation as a relational data model upon which novel query capabilities can be built, and provides objects that can be analyzed using algorithms for comparison of strings and lists. As an initial effort, we have implemented a prototype using a representative set of comparative and content-based annotation methods to reduce a collection of prokaryotic genomes to a feature mosaic representation. Entity-Relationship modeling was then used to develop a data model capable of storing detailed results, including complete parameters for each instance of analysis.

GIBAS, Cynthia
An Interactive Web Resource for Comparative Analysis of Chloroplast Genomes
Gibas C. J. [1], Kaluszka A. [1], Hilu K. W. [1]. (1) Department of Biology, Virginia Tech
Among the three plant genomes (nuclear, mitochondrial, and chloroplast), the chloroplast genome has been instrumental in advancing our understanding of plant evolution and systematics. The chloroplast genome became a focus in such studies because of its relatively small size (around 150 kb), highly conserved nature, and the presence of genes in single copies, as opposed to the multigene families that predominate in the nuclear genome. Several small and large-scale sequencing projects are on pace to deposit dozens of new chloroplast genome sequences in GenBank over the next few years, and with this influx of data comes a shift to whole-genome methods of analysis. Genome Organization Analysis Tool (GOAT) is a program that performs comparative sequence analysis on ordered gene lists from annotated genomes, provides visual and tabular output, and provides means of accessing and analyzing related gene sequence data, for the purpose of comparing genome organization at the gene order level. GOAT can be used to compare any two or more genomes or chromosomes on demand, or configured to provide access to precomputed comparisons of a specific group of genome sequences. We use this software to present an interactive resource for analysis of chloroplast genomes.

GIBAS, Cynthia
A Phylogeny of Land Plants Based on Whole-Genome Analysis of Chloroplast Using Correlated Peptide Motifs
Gibas C. J. [1], Hilu K. W. [1], Stuart G. [2], Berry M. [3]. (1) Department of Biology, Virginia Tech, (2) Department of Life Sciences, Indiana State University, (3) Department of Computer Science, U. of Tennessee
We have applied a new computational method to estimate pairwise species distances for a collection of sequenced chloroplast genomes primarily from land plants. A singular value decomposition (SVD) method is used to define multidimensional protein vectors based on a data matrix in which each protein

is represented as a vector of overlapping tetrapeptide motifs. Protein vectors are summed to produce species vectors, from which interspecies distances can be determined. We have computed distances based on tetrapeptide frequencies in the entire protein content of the chloroplast as well as from subsets of proteins classified by traditional phylogenetic analysis as slow or fast evolving. These trees generally capture many accepted relationships among land plants, although a few unexpected branchings are obtained. Analysis of the entire protein content of the chloroplast is more successful in confirming accepted relationships than analysis of slow or fast evolving genes alone.

GILAD, Yoav

A comparison of the olfactory receptor gene repertoires of human and chimpanzee

Gilad Y. [1], Man O. [2], Paabo S. [3], Glusman G. [4]. (1) Yale University School of Medicine, (2) Weizmann Institue of Science, (3) MPI for evolutionary Anthropology, (4) Institute for Systems Biology

Olfactory receptor (OR) genes constitute the basis of the sense of smell and are encoded by the largest mammalian gene superfamily of more then 1,000 genes. In humans, but not in mice or dogs, the majority of OR genes became pseudogenes. Here, we take advantage of the recently sequenced genome of the chimpanzee and compare the OR gene repertoire of human to its closest living relative. In agreement with previous reports, we found that the chimpanzee have a significantly lower proportion of OR pseudogenes than human. By contrasting shared vs. species-specific coding region disruptions, we estimate that humans have started to rapidly accumulate OR pseudogenes roughly 3.2 MYA. We further estimate that 112 human intact OR genes are evolving under no evolutionary constraint and may become pseudogenes. We found one chimpanzee-specific OR subfamily expansion, four expansions in human, and support for the action of positive selection on certain human OR genes. These observations may suggest that while humans have an overall reduced need for the sense of smell compared with chimpanzee, specific sensory requirements may have shaped the functional human OR gene repertoire

GO, Yasuhiro

Lineage-dependent loss of function and diversification of bitter taste receptor genes in primates.

Go, Y. [1], Satta Y. [1], Takenaka, O. [2], Takahata, N. [1] (1) Graduate University for Advanced Studies, Sokendai, (2) Primate Research Institute, Kyoto University

We characterize bitter taste receptor (*T2R*) genes in 13 primates to gain insight into their tempo and mode of evolution. In the human genome, there are 11 pseudogenes among 36 genes. Of these three are human-specific pseudogenes, and in non-human primates there are18 species-specific pseudogenes. This contrasts with the finding that the mouse genome contains only six pseudogenes among 41 genes. The extensive loss of function in primates is likely attributed to relaxed functional constraints. There is a trend that frequently duplicated genes have more pseudogenes. In functional genes, however, positive selection has been operating on amino acids in extracellular domains particularly in genes under frequent duplication. Two likely fates of duplicated genes, loss of function and adaptive diversification, are thus manifest in the primate *T2R* genes. It is argued that the repertoire of *T2R* genes in primates is lineage-dependent and reflects different responses to changes in the environment.

GOERTZEN, Leslie

Multiple copies of *atp8* in the mitochondrial genome of *Amborella trichopoda* support a horizontal gene transfer hypothesis.

Goertzen L. R. [1], Bergthorsson U. [1], Palmer J. D. [1]. (1) Indiana University

The *Amborella trichopoda* mitochondrial genome contains two distinct copies of *atp8* (*orfB*), the gene encoding subunit 8 of atp synthase. Copy number and organellar location are supported by Southern blots and evidence of RNA editing, a characteristic of plant mitochondrial genes. *Amborella* belongs to a basal lineage of angiosperms and phylogenetic analyses confirm that one copy of *atp8* also belongs to a relatively basal angiosperm lineage and is the likely product of vertical inheritance in the lineage leading to *Amborella*. The other copy of *atp8* appears to have been acquired by horizontal gene transfer from an

unidentified eudicot donor. cDNA clones of both *atp8* copies contain evidence for C > U RNA editing, suggesting that both are transcribed, edited and probably functional. These results are discussed in the context of evidence for two copies of the mitochondrial *atp1* gene in *Amborella trichopoda*, one of vertical and the other horizontal origin.

GOETTING-MINESKY, Paula

Male mutation bias in Perissodactyla and Cetacea

Goetting-Minesky P. [1], **Kvikstad E.** [1], Makova K. [1]. (1) The Pennsylvania State University

The number of germline cell divisions is higher in mammalian males than in females. This leads to male mutation bias and suggests that errors during DNA replication are a significant source of mutation. Here we investigate nucleotide substitution rates at introns of genes homologous between chromosomes X and Y in Perissodactyla (horses, rhinos) and Cetacea (whales, dolphins). We examine how ubiquitous male mutation bias is in mammals and attempt to identify factors contributing to the variation in its magnitude. Approximately 2 kb of intron sequences were amplified and sequenced from ZFX and ZFY for 4 species of Perissodactyla and 3 species of Cetacea. For Perissodactyla, the substitution rate appears to be 2.21 times higher on Y than on X, leading to the male-to-female mutation rate ratio (alpha) of ~5.59. The comparative data for Cetacea and Perissodactyla will allow us to assess whether differences in metabolic rate contribute to the variation in the magnitude of male mutation bias.

GOJOBORI, Takashi

Origins and Evolution of the Central Nervous System in Animals: Gene expression profiles in hydra neural cells and planarian brain.

Gojobori T. [1], Mineta K. [1], Hwang J. S. [1], Agata K. [2], Ikeo K. [1]. (1) National Institute for Genetics, (2) RIKEN Center for Developmental Biology, Japan

To address the question of whether the central nervous system (CNS) was derived from a common ancestor of deuterostomes and protostomes, we took a comparative approach using different species, particularly focusing on one of the lower bilateral animals, the planarian (Platyhelminthes, Tricladida), which is known to possess one of the most primitive brains. Determining the nucleotide sequences of about 10,000 ESTs from the head portion of planarians, we obtained 116 clones that may be related to the nervous system. Comparing these 116 planarian EST clones with all ORFs of the complete genome sequences of the human, fruit fly, and nematode, we found that the origin of nervous system-related genes greatly predated the emergence of the nervous system, and that these genes might have been recruited toward the nervous system. We also sequenced the ESTs of hydra (Cnidaria) that does not have a CNS. Using the DNA chips, we have conducted competitive hybridization experiments between a normal hydra and a mutant called *epi-hydra* that completely lacks neural cells. Since we successfully obtained ESTs that were specifically expressed in hydra neural cells, we would present the results of comparisons between those hydra ESTs and the planarian brain ESTs.

GOLDMAN, Nick

Estimating the frequency of events that cause multiple nucleotide changes

Whelan S. [1], **Goldman N.** [1]. (1) EMBL-European Bioinformatics Institute

Mathematical models of sequence evolution assume that all substitutions derive from point mutations. There is, however, increasing evidence that larger scale events, involving two or more consecutive sites, may also be important. We describe a model ('SDT') that allows for single nucleotide, doublet and triplet mutations and, when applied to protein-coding DNA, allows doublet and triplet mutations to overlap codon boundaries. We have implemented the SDT model for maximum likelihood phylogenetic inference, and have applied it to over 250 protein-coding sequence alignments from the Pandit database. We find the SDT model's inclusion of doublet and triplet mutations to be overwhelmingly successful in

giving significant improvements in fit of model to data, indicating that larger scale mutation events do occur. Distributions of inferred parameter values over all alignments analyzed suggest that these events are far more prevalent than previously thought. Detailed consideration of our results and the absence of any known mechanism causing three adjacent nucleotides to be substituted simultaneously leads us to suggest that the actual evolutionary events occurring may include still-larger scale events such as gene conversion, inversion or recombination, or series of rapid compensatory changes.

GONDER, Mary Katherine

Demographic History of Khoisan Speakers of Tanzania Inferred from mtDNA Control Region Sequences

Gonder M. K. [1], Mortensen H. [1], Hirbo J. B. [1], Mountain J. [2], Tishkoff S. A. [1]. (1) University of Maryland, (2) Stanford University

The Hadza and Sandawe are unique among Tanzanians because they speak "*click*" languages and subsist by hunting and gathering. However, their origins and relationships with their neighbors and with southern African "*click*"; speakers remains a genetic and linguistic puzzle. We present mtDNA control region sequences of the Hadza, the Sandawe and other Tanzanians (n>600) compared to mtDNA sequences from a global panel (n>1000). Several analyses were applied to these data to 1) characterize genetic variation among the Hadza and Sandawe compared to neighboring populations and 2) reconstruct past relationships of East African populations to one another and to other African populations. The oldest mtDNA lineages are present in Tanzania and occur most frequently among the Sandawe who share a unique relationship with southern African "*click*" speakers, but are absent among the Hadza. These results support archeological and fossil evidence indicating an east African origin of modern humans.

GONG, Yi

Estimation of Deleterious Genomic Mutation Rate of Drosophila melanogaster by the Binscy Assay

Gong ZY. [1], Woodruff R. C. [1], Thompson J. N. [2]. (1) Bowling Green State University, (2) University of Oklahoma

Mutation rate is an important concept in evolution. It has been pointed out that most new mutations with any effect on the phenotype must reduce fitness, and are deleterious. Present-day organisms have been subject to a long history of natural selection that has fixed nearly all mutations that are beneficial under the prevailing environmental conditions. Deleterious mutations are important in the evolution of sexual reproduction and genetic recombination, diploidy, mate choice, senescence, inbreeding, degeneration Y chromosomes and dosage compensation.

The objective of the current experiment is to indirectly estimate the deleterious genomic mutation rate of Drosophila melanogaster. We use a marked crossover-suppressing chromosome (Binscy) to detect mutations on the X chromosome that have been sequestered for several generations. We measure the decrement of viability (dM) by testing the ratio of Binscy males over total males including C(1;YS)oc ptg males, and the genetic variance between replicates (dV). Then formula of Bateman(1959) and Mukai(1964) (u= (dM)^2/ dV) was used to estimate the polygenetic mutation rate (u) for slightly deleterious mutations.

GONZALEZ, Josefa

A BAC-based physical map of the *Drosophila buzzatii* genome

Gonzalez J. [1], Nefedov M. [2], Casals F. [3], Delprat A. [3], Calvete O. [3], Bosdet I. [4], Schein J. [4], de Jong P. [2], Ruiz A. [3]. (1) Universitat Autonoma de Barcelona, Spain , (2) Children's Hospital Oakland Research Institute,USA, (3) Universitat Autonoma de Barcelona, Spain, (4) Genome Sciences Centre, Canada

Few genomic resources are available in the genus Drosophila besides the reference species *D. melanogaster* (Sophophora subgenus). *D. buzzatii* (*repleta* group) is a representative species of the Drosophila subgenus which is being widely used as a model in studies of genome evolution, adaptation and speciation. A genomic library of *D. buzzatii* has been constructed using the shuttle vector pTARBAC2.1. This library

comprises ~18.000 clones with an average insert size of ~150 kb (12-18x redundancy). A sample of ~10.000 clones has been fingerprinted by digestion with *EcoRI* and agarose gel electrophoresis, and assembled using FPC into 634 contigs with 2 or more clones (mean 15 clones) and 516 singletons. Furthermore, 385 clones representing 376 contigs have been unambiguously mapped by hybridization to single sites in the salivary gland chromosomes. This high-coverage BAC-based physical map of the *D. buzzatii* genome will be a valuable addition to the resources available to the Drosophila research community.

GOOD, Jeffrey

Molecular evolution and genomic location of genes expressed during spermatogenesis in the mouse

Good J. M. [1], Nachman M. W. [1]. (1) University of Arizona

In mammals, many genes involved in male reproduction are rapidly evolving and differentially sex-linked, consistent with the action of sexual selection. However, the importance of developmental timing and tissue specificity in shaping the evolution of male reproductive genes is unclear. We examined the genomic location and molecular evolution of ~1000 genes expressed in the germ line during spermatogenesis in mice. Some of these genes are expressed only in the germ line while others are also expressed in other tissues. However, only the subset of genes specific to the germ line had significantly elevated rates of evolution. This pattern was strongest in germ line-specific genes expressed primarily at later developmental stages. Overall, we found a paucity of genes expressed during spermatogenesis on the X chromosome, consistent with selection against X-linkage due to X-inactivation during meiosis. However, genes expressed only in the male germ line were not underrepresented on the X, suggesting proportionately more germ line-specific genes escape X-inactivation.

GOODISMAN, Michael

Genomic analysis of social insect development: A microarray-based study of gene expression in larval and adult ants

Goodisman M. D. [1], Isoe J. [2], Wheeler D. E. [2], Wells M. A. [2]. (1) The Georgia Institute of Technology, (2) The University of Arizona

The complex nature of animal societies affects many aspects of organismal biology. One significant example of this effect is the influence that social structure has on individual development in the highly social insects. To initiate an understanding of how ontogeny differs between social and nonsocial insects, we used cDNA microarray technology to examine differences in gene expression between larval and adult *Camponotus festinatus* ants. We then compared expression patterns obtained from our study to those observed in related studies of *D. melanogaster*. We found that many genes showed distinct patterns of expression between the larval and adult ants, a result that was confirmed through QRT-PCR. Genes involved in protein metabolism and structural activity tended to be more highly expressed in larval than adult ants. Surprisingly, the patterns of gene expression observed for homologous genes in *D. melanogaster* differed substantially from those observed in *C. festinatus*. This study represents one of the first genome-level analyses of gene expression in ants, and suggests that the molecular mechanisms involved in development will differ between social and nonsocial insect taxa.

GOTEA, Valer

Transposable Elements Contributed to our Proteome

Gotea V. [1], Makalowski W. [1]. (1) Department of Biology, Penn State University

The high content in repetitive elements of eukaryotic genomes has intrigued many, and for many years, transposable elements (TE) were regarded as unnecessary, genomic parasites, or selfish DNA. Different recent studies have shown that TE fragments can be found in coding regions of many genes at mRNA level. This, however, does not prove that TE-encoded protein fragments exist, and in disease cases, it rather suggests a non-functional protein. Determination of the three dimensional structure of the protein can be considered as definitive proof for translation of TE-fragment-containing mRNA and protein existence, but as no studies presented this kind of data, the contribution of TEs to our proteome remained

an open question. We searched the PDB repository for proteins that are likely to contain TE fragments and found a few positive examples. Our findings also suggest that some TEs (LINEs, LTRs) can contribute to our proteome, while others (Alus) are rather unsuitable for this purpose.

GOTO, Hiroki

Comparative study of the regulation and catalytic efficiency of alpha-amylase in *Drosophila melanogaster*

Goto H. [1], Szmidt A. E. [1], Yamazaki T. [2], Inomata N. [1]. (1) Kyushu University, (2) The Research Institute of Evolution

To clarify the causes of the amylase activity differences among fly strains with different AMY isozymes, we examined amylase activity of two types of transgenic flies with chimeric *Amy* genes (*Amyc161* and *Amyfc661*) on two different food media at two developmental stages. The two types had different 5□f flanking regions of the *Amy1* and *Amy6* genes. We found highly significant difference in amylase activity between the *Amyc161* and *Amyfc661* flies. This indicates that the sequence difference, 16 nucleotide differences, between the 5□f flanking regions of the *Amy1* and *Amy6* genes causes difference in amylase activity. These nucleotide differences were not located in the region harboring putative regulatory elements. We also measured two fitness components: developmental time and productivity. We could not detect any significant difference between the *Amyc161* and *Amyfc661* in the fitness components.

GRASSOT, Julien

Molecular Evolution of Receptor Tyrosine Kinases with Immunoglobulin-like Modules

Grassot J. [1], Gouy M. [2], Perrière G. [2], Mouchiroud G. [1]. (1) CGMC UMR 5534, (2) BBE UMR 5558

Receptor Tyrosine Kinases (RTK) are involved in fundamental control and regulation of cellular process in metazoan. They represent an important and actively studied protein family. RTK are grouped in classes that have a complex phylogenetic story (e. g. complete genome duplication, local duplication,). RTK with Immunoglobulin-like (IG-like) modules represents a major class among the RTK and are present from early metazoans, including sponges and choanoflagellates. Because they are representative of all the RTK, we have chosen to perform a comprehensive study of their monophyletic origin. We started by gathering all data available on IG-like RTK, and continued by performing phylogenetic trees on their conserved catalytic tyrosine kinase domain sequences. Our results give a comprehensive framework of evolution events occurring in the IG-like RTK classes divergence - and in a broader view to all the RTK-, with their key dates related to the metazoan evolution. It is a basis to better understand the phylogenetic story of all RTK. It will serve as a model to contribute to better understand the evolution of metazoan.

GREEN, Eric

Decoding the Human Genome by Multi-Species Sequence Comparisons

Green E. D. [1], NISC Comparative Sequencing Program [1]. (1) NHGRI, NIH

The comparison of sequences generated from different extant species has emerged as a powerful strategy for identifying functionally important regions of the human genome. As a complement to ongoing efforts to sequence entire genomes, the NISC Comparative Sequencing Program is pursuing multi-species genome explorations by sequencing and analyzing the same orthologous regions in many different vertebrates. To date, we have generated over 400 Mb of comparative sequence data from more than 30 different species. Comparative analyses of these data are revealing important insights about the patterns of sequence conservation among species. In particular, we are developing approaches for utilizing multi-species sequence comparisons to identify highly conserved non-coding sequences, which likely correspond to regulatory and other functionally important elements. In addition, we are examining the relative ?informativeness? of different species? sequences for identifying such highly conserved regions, which should help guide decisions about future sequencing efforts. Our studies should advance the

utility of comparative sequence analyses, and make important contributions towards unraveling the functional and evolutionary complexities of the human genome.

GREENBERG, Anthony
Ecological Adaptation During Incipient Speciation Revealed by Precise Gene Replacement
Greenberg A. J. [1], Moran J. R. [1], Coyne J. A. [1], Wu C. I. [1]. (1) University of Chicago
To understand the role of adaptation in speciation, one must characterize the ecologically relevant phenotypic effects of naturally occurring alleles at loci potentially causing reproductive isolation. The desaturase2 (ds2) gene of D. melanogaster is such a locus. Two geographically differentiated ds2 alleles underlie a pheromonal difference between the Zimbabwe and Cosmopolitan races. We used a site-directed gene replacement technique to introduce an allele of ds2 from the Zimbabwe population into Cosmopolitan flies. We show that the cosmopolitan allele confers resistance to cold as well as susceptibility to starvation when the entire genetic background is otherwise identical. We conclude that ecological adaptation likely accompanies sexual isolation between the two behavioral races of D. melanogaster. We are currently extending these results by testing the involvement of ds2 in behavioral isolation as well as looking for a signature of positive selection in the ds2 sequence.

GRIFFIN, Autumn
Understanding the Phylogeography of Ciliates Using Molecular Markers: *Halteria grandinella* and *Meseres corlissi*
Griffin A. J. [1], Baron A. [1], Snoeyenbos-West O. L. [1], Foissner W. [1], Katz L. [1]. (1) Smith College
We are using molecular markers to explore the biogeography of ciliates. There are currently two schools of thoughts on ciliate biogeography:
1. Cosmopolitan: most ciliate species have been discovered and all ciliates are cosmopolitan
2. Endemism/species rich: fewer than half of the existing species of ciliates have been described and there is a considerable amount of endemism.
To test these hypotheses, we sequenced the small subunit rDNA (SSU rDNA), the internally transcribed spacer regions (ITS) plus the 5.8 rDNA gene from populations of two ciliate species: the closely-related freshwater oligotrichs *Halteria grandinella* and *Meseres corlissi*. Our data reveal considerable genetic diversity underlying these similar morphological entities. Moreover, we find evidence of restricted gene flow due to geographic isolation in many of our populations. The one exception is populations from Massachusetts and San Francisco, which share identical haplotypes. We discuss the potential sources of gene flow among populations, and the implications of our data on the competing hypotheses.

GRUS, Wendy
Lineage-specific expansions of vomeronasal pheromone receptor gene families in mice and rats
Grus W. E. [1], Zhang J. [1]. (1) University of Michigan
Pheromones are used by individuals of the same species to elicit behavioral or physiological changes such as male-male aggression, puberty, estrus, and induction of mating, and they are perceived primarily by the vomeronasal organ (VNO). VNO pheromone receptors are encoded by the V1r and V2r gene superfamilies in mammals. 137 V1r genes of 12 families were previously identified from the mouse genome. We here report the identification from the draft rat genome sequence 93 V1r genes, grouped into 12 families. Interestingly, we found two mouse-specific and two rat-specific families. The numbers of genes in many families also differ substantially between the two species. These observations suggest that V1r genes are subject to rapid turnover by gene duplication and pseudogenization, consistent with their roles in species-specific recognitions.

GU, Jianying
Are isochores vanishing in mammals?
Gu J. Y. [1], Li W. H. [1]. (1) University of Chicago

Isochore structure, a large scale homogeneity of GC content, has been identified only in mammalian and bird genomes. Several independent lines of evidence suggested that GC-rich isochores seemed to have disappeared from mammalian genomes. We recently investigated the evolution pattern of GC content in mammalian lineage, and inferred the ancestral state of GC and the ancestral GC content using a maximum likelihood method. Our stringent searching criteria on the HOVERGEN database have resulted in two complementary datasets: 101 genes with a marsupial outgroup; 301 genes without a marsupial outgroup. We found declines in GC3 in the GC-rich and the GC-medium genes, while no evidence was found for GC-low genes in the major lineage including primates, rodents and carnivora. A similar pattern was identified by concatenating the genes in major eutheria orders. It is noticeable that the GC3 in lagomorpha lineage is higher than its closest neighbor - rodentia in GC-low and GC-medium genes.

GU, Sheng

Simulation studies on the role of premeiotic clusters of mutations in single-gene speciation using Genetic Simulation Library

Gu S. [1], Rajaei H. [2], Woodruff R. C. [1]. (1) Dept. of Biological Sciences, Bowling Green State University, (2) Dept. of Computer Sciences, Bowling Green State University

Single gene speciation is one type of speciation processes. It involves the occurrence of a dominant mutation at one locus, which could bring either phenotypic or behavioral changes that might lead to pre-mating isolation. It is usually believed that single-gene speciation is not possible, because a complete reproductively isolated mutant has no compatible mate. However, mutations do not always occur as single events. The existence of identical copies derived from one mutation event shed the light on the scenarios when single-gene speciation is possible. These identical mutants are called premeiotic clusters. To identify the possibility and importance of the role of premeiotic clusters in the single-gene speciation, we conducted a computational simulation study by using Genetic Simulation Library (GSL). The simulation result suggests that the effect of clusters on the single-gene speciation is significant only in very small populations if mutants are considered to random mate in the whole population. However, for species that tend to interact inside demes, clusters could play much more important role. One special scenario shows that the role of clusters is maximized when there exists time differential between the birth of mutants and that of the normal.

GU, Xun

Genome Phylogenetic Analysis Based on Extended Gene Contents, and Tree of Life

Zhang H. M. [1], **Gu X.** [1]. (1) Iowa State University

With the rapid growth of entire genome data, whole-genome approaches such as gene content become popular for genome phylogeny inference, including the tree of life. However, the underlying model for genome evolution is unclear, and the proposed ({\em ad hoc}) genome distance measure may violate the additivity. We formulate a stochastic framework for genome evolution, which provides a basis for defining an additive genome distance. However, we show that it is difficult to utilize the typical gene content data, i.e., the presence or absence of gene families across genomes, to estimate the genome distance. We solve this problem by introducing the concept of extended gene content, that is, the status of a gene family in a given genome could be either absence, presence as single-copy, or presence as duplicates, which can be used to estimate the genome distance and phylogenetic inference. Computer simulation shows that the new tree-making method is efficient, consistent, and fairly robust. The example of 35 microbial complete genomes demonstrates that it is useful not only to study the universal tree of life, but also to explore the evolutionary pattern of genomes.

GU, Xun

Phylogenomic Microarray Analysis Shows Rapid Expression Divergence in the Early Stage after Gene Duplication

Gu X. [1], Zhang Z. Q. [1]. (1) Iowa State University
Using the stochastic model recently developed for expression evolution after gene duplications, we conducted a genome-wide phylogenetic analysis of yeast gene families. We first inferred each gene family tree with the homologous protein sequences, and estimated roughly the duplication time(s) relative to the E.coli/yeast speciation. For each gene family we estimated the additive expression distances between any duplicate genes. The asymmetric evolution of gene expression after duplication was examined by the relative-rate test. Our major findings are as follows: (1) the rate of expression evolution in the early stage is at least as 10-fold as that in the late stage; and (2) the rapid evolution of expression pattern shortly after gene duplication occurs in one duplicate gene, while the other copy remains a slow evolution.

GUILLET-CLAUDE, Carine

Duplication and adaptive evolution of *knox-I* genes in conifers

Guillet-Claude C. [1], Isabel N. [2], Pelgas B. [1], Bousquet J. [1]. (1) Université Laval, (2) Canadian Forest Service
Class I *knox* genes are central regulators of meristem cell identity. We identified a diversified group of *knox-I* genes in conifers. Phylogenetic analyses indicated they all formed a monophyletic group. Four subgroups named *SKN1* to *SKN4* were well delineated, each regrouping pine and spruce sequences. Rates of substitution per year showed an evolution in two steps: faster rates after gene duplications, followed by lower rates. Positive directional selection was detected for branches harbouring an accelerated rate of evolution. Sites with highly significant amino acid rate shift indicative of functional divergence were identified for *SKN3* and *SKN4*, which were assigned to different linkage groups. *SKN1* and *SKN2* did not diverge as much from each other and they were located close to each other on the same linkage group. All these data indicate that chromosomal rearrangements, which can facilitate gene divergence, participated in the adaptive evolution of the *knox-I* gene family in conifers.

HACKETT, Jeremiah

Migration of a Plastid Genome to the Nucleus in a Peridinin Dinoflagellate

Hackett J. D. [1], Yoon H. S. [1], Bhattacharya D. [1]. (1) University of Iowa
Dinoflagellate algae are important primary producers and of significant ecological and economic impact because of their ability to form "red tides". They are also models for evolutionary research because of an unparalleled ability to capture plastids through endosymbiosis. The location and extent of the plastid genome in the dominant perdinin-containin dinoflagellates remain two of the most intriguing issues in plastid evolution. The plastid genome in these taxa is reduced to single-gene minicircles encoding an incomplete set of plastid proteins. The location of the remaining photosynthetic genes is unknown. We generated a data set of 6,480 unique expressed sequence tags (ESTs) from the toxic dinoflagellate *Alexandrium tamarense* to find the missing plastid genes and to understand the impact of endosymbiosis on genome evolution. Here we identify 48 of the non-minicircle-encoded photosynthetic genes in the nuclear genome of *A. tamarense*, accounting for the majority of the photosystem. Fifteen genes that are always found on the plastid genome of other algae and plants have been transferred to the nucleus in A. tamarense. The plastid-targeted genes have red and green algal origins. These results highlight the unique position of dinoflagellates as the champions of plastid gene transfer among photosynthetic eukaryotes.

HADDRILL, Penelope

Inferring the selective and demographic history of *Drosophila melanogaster* populations using multi-locus data.

Haddrill P. [1], Thornton K. [2], Charlesworth B. [1], Andolfatto P. [3]. (1) ICAPB, University of Edinburgh, (2) Dept. of Mol. Biol. & Genetics, Cornell University, (3) Dept. of Zoology, University of Toronto
It has been suggested that *Drosophila melanogaster* has a tropical African origin and has recently colonized temperate habitats. This colonization of temperate habitats has presumably been accompanied by adaptation and/or changes in population size, though the relative importance of these two factors is

debated. We present an analysis of nucleotide variability data for 10 X-linked loci from five populations of *D. melanogaster*. African populations harbor higher levels of variability and drastically lower levels of linkage disequilibrium than non-African populations. We evaluate these population differences in the context of simple models of selection and population size changes in the history of non-African populations. A recent drastic bottleneck in the history of both non-African populations appears to capture most features of the data. Shared haplotypes and similar levels of variability in non-African populations argue against local adaptation. African populations also show a departure from the expectations of a neutral population of constant size by several measures. This may reflect either population size changes in the distant past or recurrent positive selection.

HAJIBABAEI, Mehrdad

DNA barcoding: identification of animals using a molecular evolutionary approach

Hajibabaei M. [1], Janzen D. H. [2], Hebert P. DN. [1]. (1) University of Guelph, (2) Univeristy of Pennsylvania

A fast and accurate identification system is critical for exploring biodiversity. We used a molecular evolutionary approach to examine the efficacy of DNA based identification system for several groups of animals. This approach was tested on four taxonomically well characterized assemblages involving 1500 individuals of 500 (skipper butterflies, hawk moths, bees and birds). Analysis of sequence diversity in a 650 bp fragment of cytochrome oxidase subunit I (coI), as a DNA barcode, allowed successful species level identification. This was possible because of very low genetic divergence among conspecific individuals but deep genetic divergences among congeneric taxa. We also evaluated the effectiveness of DNA barcoding in different geographical regions and compared results from COI with those from another gene, cytochrome b (cytb), for a subset of samples 100 species.

HALDER, Indrani

Estimation of BioGeographical Ancestry in Admixed Populations

Halder I. [1], Mei R. [2], Jones K. [2], Akey J. M. [3], Shriver M. D. [1]. (1) The Pennsylvania State University, (2) Affymetrix Inc., Santa Clara, CA, (3) FHCRC, Seattle, WA

Understanding genetic variation in populations and how this variation contributes to differential disease risk among populations is an important area of research. We have typed a panel of 11,500 SNP markers in three resident US populations, (African Americans, Puerto Ricans and European Americans) in efforts to analyze patterns of genetic variability. The first two are considered admixed populations with varying levels of BioGeographical ancestry (BGA), while European Americans are assumed more homogeneous. Using a maximum likelihood based method and a separate Bayesian method we show that BGA estimates vary within all groups. To check for stratification within populations we have used the Individual Ancestry Correlation (IAC) test, where markers are split into two subsets and ancestry is estimated separately with both sets. Correlation between estimates obtained with the different marker sets indicates presence of genetic structure. While the admixed populations show stratification as expected, we also detected evidence for stratification within the European-American sample. The implications and source of this structure has been studied and is presented here. We have done additional simulation studies to check the accuracy of our estimates and to test if a subset of informative markers provides more precise estimates compared to estimates with all markers.

HAO, Li

Genomic Organization and Evolutionary Analysis of Ly49 Genes Encoding Rodent Natural Killer Cell Receptors: Rapid Evolution by Repeated Gene Duplication

Hao L. [1], Nei M. [1]. (1) Dept. of Biology, Pennsylvania State University

Ly49 genes regulate the cytotoxic activity of natural killer (NK) cells in rodents and provide important protection against virus-infected or tumor cells. About 15 Ly49 genes have been identified in mice, but only a few genes have been reported to date in rats. Here we studied all Ly49 genes in the entire rat genome sequence and identified 17 functional and 16 nonfunctional genes in a 1.8 Mb genomic region of

chromosome 4. Phylogenetic analysis of these genes indicated that the Ly49 gene family expanded rapidly in recent years, and this expansion was mediated by both tandem and genomic block duplication. The joint phylogenetic analysis of mouse and rat genes suggested that the most recent common ancestor of the two species had at least several Ly49 genes but that the majority of current duplicate genes were generated after divergence of the two species. In both species Ly49 genes are apparently subject to birth-and-death evolution, but the birth and death rates of Ly49 genes are higher in rats than in mice. The rate of gene expansion in the Ly49 gene family in rats is one of the highest among all mammalian multigene families so far studied.

HAO, Weilong
Patterns of bacterial gene movement
Hao W. [1], Golding G. B. [1]. (1) McMaster University
Lateral gene transfer has emerged as an important force in bacterial evolution. A substantial number of genes can be inserted into or deleted from genomes through the process of lateral transfer. Here we looked for atypical occurrence of genes among related organisms to detect laterally transferred genes. We have analyzed 50 bacterial complete genomes from nine groups.
For each group we use a 16s rRNA phylogeny and a comparison of protein similarity to map gene insertion/deletions onto their species phylogeny. The results reveal that there is poor correlation of genes inserted, deleted and duplicated with evolutionary branch length. In addition, the number of genes inserted, deleted, or duplicated within the same branch are not always correlated with each other. Nor is there any similarity within groups. Most strikingly, the number of insertions of foreign genes is much larger in the external branches of the trees. These insertions also greatly outnumber the occurrence of deletions and yet the genome sizes of these bacteria remain roughly constant. This indicates that many of insertions are specific to each organism and are lost before related species can evolve. Simulations of the process of insertion and deletion, tailored to each phylogeny, support this conclusion.

HARTL, Daniel
Molecular Evolution of Plasmodium Genomes
Hartl D. L. [1], Castillo-Davis C. I. [1], Bedford T. B. [1]. (1) Harvard University
The 30-Mb haploid genome of the malaria parasite *Plasmodium falciparum* consists of 14 chromosomes ranging in size from 0.65 to 3.4 Mb, and its 82% AT content ranks it among the most AT-rich in eukaryotes. To examine rates and patterns of genome evolution, we have compared nucleotide substitutions and intron gains or losses in pairs of orthologous and paralogous (duplicate) genes in *P. falciparum* and the rodent parasite *P. yoelii.* We find that duplicate genes in both *P. falciparum* and *P. yoelii* exhibit a dramatic acceleration of both nonsynonymous substitution and intron gain or loss when compared with orthologs, suggesting either relaxed selection or accelerated directional selection in duplicate genes. Among duplicate genes, rates of synonymous substitutions do not differ significantly from those in introns, suggesting that a large fraction of intron sites in *P. falciparum* evolve under little or no selective constraint.

HARTLING, Julia
Relationship between local packing and sequence diversity in protein structures
Hartling J. [1], Kim J. [2]. (1) Yale University, University of Pennsylvania, (2) University of Pennsylvania
The amount of sequence diversity (mutational robustness) underlying different protein structures varies dramatically, small number of structures are native for several sequence families (superfolds), whereas the majority of structures are native for a single sequence family. The diversity of the sequences underlying each protein structure is determined by functional constraints along with various qualities of the 3D structure. Topology or local packing of amino acids in proteins is one of the important factors determining the underlying sequence diversity. Different aspects of local packing - number of local neighbors within 6.5A of each amino acid, the regularity of packing architecture, and contact order

influence the size of sequence space in non-redundant set of protein structures. Moreover, the regularity of local packing is associated with other patterns in phenotype space of proteins, such as the overall symmetry of structures.

HAY, Jennifer

Fossil molecular outgroup for a phylogenetically isolated taxon, *Sphenodon*

Hay J. M. [1], Sarre S. D. [2], Lambert D. M. [1]. (1) Allan Wilson Centre, IMBS, Massey University, (2) Applied Ecology Research Group, U of Canberra

Outgroups form a requisite part of molecular phylogenetic studies by providing a root for phylogenetic trees to determine the polarity of evolution. However, phylogenetically isolated taxa have no close relatives that can be used as outgroups, so studies result in unrooted trees. This is especially a problem for so-called "living fossil" taxa that are virtually monotypic. *Sphenodon*, or tuatara, are an extreme case of this, whose extant members comprise a monogeneric order that separated from snakes and lizards 230 million years ago. We used a novel approach to solve this problem, by isolating and sequencing a partial nuclear DNA copy of the mitochondrial cytochrome b gene, and using the nuclear pseudogene as an outgroup to root a mitochondrial tree for all tuatara populations. The pseudogene represents a fossil mitochondrial gene as it changes little once inserted into the nuclear genome. Similarly, we are now looking at fossil mtDNA genes from ancient DNA of extinct populations.

HAYAKAWA, Toshiyuki

Gene conversion of Sialic binding domains by Pseudogenes in the CD33-related Siglec Gene Cluster

Hayakawa T. [1], Angata T. [2], Margulies E. H. [3], Mikkelsen T. [4], Green E. D. [3], Varki A. [1]. (1) UCSD, La Jolla, CA, (2) AIST, Tsukuba, Japan, (3) NIH, Bethesda, MD, (4) The Broad Institute, Cambridge, MA

Siglecs (sialic acid-binding immunoglobulin superfamily lectins) are a family of cell surface receptors involved in regulating the immune response. Nine of the 13 known primate Siglec genes the along with 14 Siglec pseudogenes comprise the CD33-related Siglec gene cluster on human chromosome 19. Gene conversion is a mechanism for copying part of a genomic sequence into another, contributing to genetic diversity. We analyzed genomic sequences of the CD33-related Siglec gene cluster in three primates (human, chimpanzee and baboon) and found three partial gene conversions between Siglec genes and pseudogenes, involving the extracellular domains that mediate sialic acid recognition. Two of these involved a pseudogene converting a Siglec gene in the human or chimpanzee lineage. ELISA assays using recombinant proteins show differences in sialic acid-binding properties between the converted Siglec and its non-converted ortholog. These findings indicate that gene conversion with pseudogenes has contributed to the functional evolution of the Siglecs.

HAZKANI-COVO, Einat

Rarity of conservation in highly regulated operons

Hazkani-Covo E. [1], Graur D. [2]. (1) Tel Aviv University, (2) Department of Biology and Biochemistry, University of Houston

The forces that drive the evolution of single genes have been studied extensively. For example, it was suggested that proteins involved in numerous protein-protein interactions are more "essential" to the organism than proteins involved in fewer interactions and, hence, evolve slower. However, very little is known about the forces that drive the evolution of complex genetic elements, such as bacterial operons. Operon architecture is seldom conserved during evolution. In this study we attempted to extrapolate from the gene to the operon. By analogy to the situation in proteins, we hypothesized that operons that are controlled by many transcription factors will be evolutionarily more conserved than operons under the control of a single transcription factor. The degree of operon conservation was inferred by ascertaining for each operon (whose number of transcription-factor interactions is known) in *Escherichia coli* whether or not it is completely or partially intact in 44 completely-sequenced bacterial genomes. No correlation was found between the number of transcription factors and operon conservation.

HEDGES, Blair

A molecular timescale for eukaryote evolution and the rise of complex life

Hedges S. B. [1], Blair J. E. [1]. (1) Penn State University

The pattern and timing of the rise in complex multicellular life during Earth's history has not been established. Great disparity persists between the pattern suggested by the fossil record and that estimated by molecular clocks. Here, we used all available protein sequence data and molecular clock methods to place constraints on the increase in complexity through time. Estimates of the maximum number of cell types of common ancestors showed an increase from two cell types at 2500 Ma to ~10 types at 1500 Ma and 50 cell types at ~1000 Ma. The results suggest that oxygen levels in the environment, and the ability of eukaryotes to extract energy from oxygen, as well as produce oxygen, were key factors in the rise of complex multicellular life. Mitochondria and organisms with more than 2-3 cell types appeared soon after the initial increase in oxygen levels at 2300 Ma. The addition of plastids at 1500 Ma, allowing eukaryotes to produce oxygen, preceded the major rise in complexity.

HEDGES, Dale

Differential *Alu* Mobilization and Polymorphism Among the Human and Chimpanzee Lineages

Hedges D. J. [1], Callinan P. A. [1], Cordaux R. [1], Xing J. [1], Barnes E. [1]. (1) Louisiana State University

Alu elements are primate-specific members of the SINE (Short INterspersed Element) retroposon family which comprise approximately 10% of the human genome. Here we report the first chromosomal-level comparison examining the Alu retroposition dynamics following the divergence of humans and chimpanzees. We find a two-fold increase in Alu insertions in comparison to the common chimpanzee (Pan troglodytes). The genomic diversity (polymorphism for presence or absence of the Alu insertion) associated with these inserts indicates that, analogous to recent nucleotide diversity studies, the level of chimpanzee Alu diversity is approximately 1.7 times higher than that of humans. Evolutionarily recent Alu subfamily structure differs markedly between the human and chimpanzee lineages, with the major human subfamilies remaining largely inactive in the chimpanzee lineage. We propose a population-based model to account for the observed fluctuation in Alu retroposition rates across primate taxa.

HERNANDEZ, Ryan

Selecting Forward: A Simulation Approach to Modeling Evolution

Hernandez R. D. [1]. (1) Cornell University

Recent studies have shown that nucleotide mutation rates are heavily influenced by mutation biases (e.g. transition/transversion and CpG biases). By incorporating these mutation biases into a finite-site forward simulation program, sequence data can be generated to determine their effects on amino acid composition and assess the robustness of common statistical tests of neutrality under mutation-specific selection schemes. The selective effect of each mutation is either drawn from a probability distribution or computed via various fitness functions based on physiochemical distances (determined *a priori*). Finally, the effects of demographic changes and recombination rates on the site frequency spectrum will be addressed.

HERSH, Bradley

Cis-regulatory evolution of the Ultrabithorax-target gene *knot* in the *Drosophila* haltere

Hersh B. M. [1], Carroll S. B. [1]. (1) University of Wisconsin

To understand how morphology evolves under the control of Hox genes, we hope to identify Hox target genes and explore the rules that govern the function and evolution of their cis-regulatory sequences. As the *knot* gene is expressed in the *Drosophila* wing but not haltere, it is a potential target of the *Ubx* Hox gene. We isolated a wing-specific cis-regulatory element of knot that contains a single activator (Ci) binding site and a cluster of UBX binding sites necessary for repression. The negative and positive control regions are physically separable, so UBX does not repress by competing for activator binding sites. The Ci binding site is conserved in five *Drosophila* species but the UBX binding sites are not. We identified a

second cluster of UBX sites that is conserved and can mediate *knot* repression in the haltere. Thus, the activator site is conserved during the evolution of this element, whereas different functional clusters of UBX sites sufficient to confer repression have arisen.

HILL, Emma
Characterizing Ancient Protein Evolutionary Relationships via Structure.
Hill E. E. [1], Brenner S. E. [1]. (1) University of California, Berkeley
Protein structure is more conserved over time than sequence; many proteins of common evolutionary descent share fold topology but not sequence similarity. Detailed analyses allow us to understand protein evolution in terms of three-dimensional structure and sequence, rather than sequence alone.

We used structure features (e.g., physical attributes, interactions, insertions, duplications) to determine protein relationships and to reconstruct phylogenies for protein families. We found that chain and helix length, crossovers, and receptors can determine relations of the four-helical cytokines; comparable features play key roles in other families we studied. We are currently quantifying the structural features that can be used to elucidate phylogenetic relationships for proteins of known structure and to differentiate homology from analogy.

HOBERMAN, Rose
Using physical-chemical properties of amino acids to model site-specific substitution propensities.
Hoberman R. A. [1], Lu Y. [1], Klein-Seetharaman J. [2], Rosenfeld R. [1]. (1) Carnegie Mellon University, (2) University of Pittsburgh Medical School
Existing models of molecular evolution capture much of the variability in mutation rates across sites. More biologically realistic models also seek to explain site-specific differences in substitution propensities between residue pairs, leading to more accurate and informative models of evolutionary dynamics. Toward this end, we describe a procedure for systematically characterizing the conservation of each position in a multiple sequence alignment in terms of specific physical and chemical properties. We use a Monte-Carlo method to ascertain the statistical significance of the findings and to control the False Discovery Rate. We use our method to annotate the diverse GPCRA family with a selection pressure profile. We demonstrate the computational and statistical significance of the properties we have identified, and discuss the biological significance of our findings. The latter include confirmation of experimentally determined properties as well as novel testable hypotheses.

HOOD, Leroy
Systems Biology and Evolution
Hood L. E. [1]. (1) Institute for Systems Biology
The advent of the Human Genome Project has transformed biology by providing a genetics parts list of all genes and proteins, by fueling the contention biology is an informational science, and by catalyzing the emergence of biological information (e.g., rapid DNA sequencing or DNA chips). From this has emerged a new approach to biology termed systems biology—*centered on the idea one can study biological systems by delineating the relationships of all of their component elements*—and, hence, come to understand the resulting systems properties. I will discuss these important points, give several examples of systems approaches, and conclude by discussing the emerging view of evolution as a means for generating organisms that can more effectively carry out biological computation.

HUANG, Jinling
Phylogenomic evidence of organelle loss and gene transfer in *Cryptosporidium parvum*
Huang J. [1], Mullapudi N. [1], Lancto C. [2], Scott M. [1], Abrahamsen M. [2], Kissinger J. C. [1]. (1) University of Georgia, (2) University of Minnesota
The apicomplexan parasite *Cryptosporidium parvum* differs from other apicomplexan organisms by

lacking a plastid organelle. We have screened the entire genome using a phylogenomic approach and identified genes of plastid/endosymbiont and prokaryotic origin. Our data suggest that *Cryptosporidum* evolved from a plastid-containing lineage and subsequently lost the organelle during its evolution. Expression analyses of several transferred genes show that these genes are expressed and their expression is regulated throughout the life cycle of *C. parvum*. The successful integration of the transferred genes has changed the genetic and metabolic repertoire of the parasite.

HUNTLEY, Melanie

Comparative Proteomics: Repetitive Simple Sequences and Neurological Proteins

Huntley M. A. [1], Golding G. B. [1]. (1) McMaster University

The most commonly shared peptide segment among the proteins in any given completed eukaryotic proteome is a simple sequence composed of one or a few amino acid residues. These repeats tend to characterize only eukaryotes, and are largely absent from archaea and eubacteria. Attempts to characterize the three dimensional structure of repeats have generally failed, and their function remains elusive. Proteins associated with disease and development of the nervous system are thought to contain repetitive, simple sequences. However, repeats are abundant within the proteins of species such as *Saccharomyces cerevisiae* which has few to no specialized developmental and neurological proteins. It is therefore of interest to determine if these specialized proteins have an excess of simple sequences when compared to other sets of compositionally similar proteins. We have determined the relative abundance of simple sequences within neurological proteins, and find no excess of repetitive simple sequence within this class. In fact, poly-glutamine repeats which are associated with many neurodegenerative diseases are no more abundant within neurological specialized proteins than within non-neurological collections of proteins. We also examined the codon composition of serine homopolymers to determine what forces may play a role in the evolution of extended homopolymers.

HUTTLEY, Gavin

Modeling the impact of DNA methylation on the evolution of *BRCA1*

Huttley G. A. [1]. (1) CBiS / ANU

The base 5-methylcytosin (^{m}C) plays an important functional role in the biology of mammals as an epigenetic modification, and appears to exert a striking impact on the molecular evolution of mammal genomes. The collective functions of ^{m}C revolve around its' effect on gene transcription while its influence on the evolution of mammal genomes derives from a greatly elevated spontaneous mutation rate of ^{m}C to T. ^{m}C is now known to occur at the dinucleotides CpG, CpA and CpT. As a step towards a comprehensive statistical examination of the role of ^{m}C in mammal molecular evolution we have developed novel Markov models of codon substitution that incorporate dinucleotide level terms relevant to ^{m}C mutation. Models were implemented using version 0.8 of the PyEvolve(http://cbis.anu.edu.au/software) and applied to a 2880 nucleotide long alignment of *BRCA1* exon 11 sequences from primates. We conclude that mutation of ^{m}C is a probable factor affecting *BRCA1* codons containing the dinucleotide CpG, a possible factor for CpA containing codons, and an unlikely factor affecting CpT containing codons. The confounding of estimated terms with the effects of natural selection indicate these effects must be addressed for comparisons between different coding and non-coding regions.

IKEO, Kazuho

Molecular Evolution of Nervous system from gene expression profile

Ikeo K. [1], Hwang J. S. [1], Ogura A. [1], Gojobori T. [1]. (1) National Institute of Genetics

Because of the effort of full genome sequencings, now, full genome sequence data become available. When we are thinking about human or some biological complexity phenomena on the molecular level, such full genome information seems really useful. Even though their usefulness for the understandings of biological phenomena, there is a big gap between the know ledges by using model organisms like mice or

fly. To confirm our knowledge or hypothesis on human, there is a difficulty to do experimentally. Comparative genomics with evolutionary view is really helpful to fill this gap.
We are now conducting the research projects from this view point using both the evolutionary and bioinformatical approaches focused on the evolution of nervous system. I will show several study for the evolution of nervous system using comparative approach of gene expression profile and also show more general over view of this field.

IRWIN, David
Evolution of new hormone function — loss and gain of a receptor
Irwin D. M. [1]. (1) University of Toronto
The vertebrate proglucagon gene encodes three glucagon-like sequences (glucagon, glucagon-like peptide-1 (GLP-1), and glucagon–like peptide 2 (GLP-2)) that have distinct functions in regulating metabolism in mammals. In contrast, glucagon and GLP-1 have similar physiological actions in fish, that of glucagon. Characterization of receptors for proglucagon-derived peptides from fish should provide insight into changes associated with the evolution of new physiological roles for glucagon-like peptides. We identified sequences similar to receptors for proglucagon-derived peptides from the pufferfish and zebrafish genomes. Phylogenetic analysis of these and other characterized vertebrate receptors for proglucagon-derived peptides demonstrate that receptors for glucagon, GLP-1 and GLP-2 have an origin before the divergence of fish and mammals; however, fish have lost the gene encoding the GLP-1 receptor. Functional GLP-1 receptors have been characterized in goldfish and zebrafish, but are most closely related to glucagon receptors. Both pufferfish and zebrafish have a second closely related glucagon receptor gene. We suggest that the binding activity of the duplicated receptors has diverged, yielding receptors specific for glucagon and GLP-1, but that ancestral downstream signaling has been maintained, resulting in both receptors retaining glucagon stimulated downstream effects.

ISHIYAMA, Hiroko
Nucleotide polymorphism and divergence in the *PgiC* gene between four closely related *Shorea* species (Dipterocarpaceae)
Ishiyama H. [1], Shukor N. A. [2], Inomata N. [1], Szmidt A. E. [1], Yamazaki T. [1]. (1) Dept. of Biology, Fac. of Sci.,Kyushu University, (2) Fac. of Forestry, Putra Malaysia University
Very little is known about the levels and patterns of DNA polymorphism and divergence in tropical trees. In this study, we investigated nucleotide polymorphism and nucleotide divergence in the following four *Shorea* species: *S. acuminata, S. curtisii, S. leprosula* and *S. parvifolia* from peninsular Malaysia using partial 1.2kb sequences of the *PgiC* gene. The nucleotide polymorphism for silent sites varied among species ranging from *Pi* = 0.0017 in *S. leprosula* to 0.0113 in *S. acuminata*. Significant negative value of Tajima□fs *D* for *S. leprosula* suggests that this species expanded recently. Dimorphic sequences were found in two species (*S. acuminata, S. parvifolia*), and a significant value of *Zns* was obtained for *S. parvifolia*. The level of nucleotide divergence between species was very low (*Dxy* = 0.0088-0.0190) suggesting they diverged recently. Furthermore, putative hybrids, which had haplotypes from different species were found suggesting the possibility of inter-specific hybridization.

ITOH, Takeshi
Intensity of Natural Selection after Horizontal Gene Transfer in Bacteria
Itoh T. [1], **Nakamura Y.** [2], **Gojobori T.** [2]. (1) National Institute of Agrobioloical Sciences, (2) National Institute of Genetics
Although it is widely accepted that the microbes have been exchanging a significant amount of genes beyond species during evolution, little is known about its evolutionary importance in terms of natural selection. We compared between closely related bacteria the horizontally transferred (HT) genes that were detected by a Bayesian method, and thereby estimated the synonymous and nonsynonymous distances by the Nei-Gojobori method. As a result it turned out that the ratios of the nonsynonymous to

synonymous distances were greater in HT genes than in non-HT genes. For example, between *E. coli* K-12 and O-157, the ratios were on an average 0.36 and 0.10 for HT and non-HT genes, respectively. These results suggest that purifying selection was extensively relaxed after the transfer events. Thus, in many cases functions of the HT genes might be deteriorated rapidly due to their neutral nature.

IWASHITA, Shintaro
Process of Retrotransposable Element-1-Involved Gene Duplication in the Creation of Protein Divergence: A Ruminant-Specific p97bcnt gene
Iwashita S. [1], Ueno S. [2], Hon-Nami K. [1], Osada N. [3]. (1) Mitsubishi Kagaku Institute of Life Scien, (2) Mitsubishi Kagaku Institute of Life Sciences, (3) University of Chicago
p97Bcnt (*bcntp97*) is a ruminant-specific paralogous gene created by retrotransposable element-1 (RTE-1)-involved gene duplication that includes an endonuclease domain of RTE-1 in its coding sequence. The gene is located 6 kb upstream of orthologous *bcnt* in the bovine genome (MBE 20, 1556, 2003). To access the process by which *bcntp97* was created, we compared *bcnt* and *bcntp97* organization between bovine and chevrotain, the earliest diverged ruminant taxon. Bovine and chevrotain gene organization is essentially the same except in the region upstream of the inserted RTE-1 domain (RTE-1 exon) of *bcntp97*, where the location of a reverse transcriptase domain (RTD), probably split from the inserted RET-1, differs. In the chevrotain gene, traces of the RTD are found in a 5' intron of the RTE-1 exon as an inverted form, while in the bovine gene, the RTD is located in the intergenic region of *bcntp97* and *bcnt*. These results showing drastic shuffling will help to trace the process of paralogous gene divergence during genome evolution.

JAGADEESHAN, Santosh
Characterization of rapidly evolving genes in Drosophila: sex genes evolve faster.
Jagadeeshan S. [1], Singh R. S. [1]. (1) McMaster University
Rapidly evolving genes (REGs) have important consequences on rates and patterns of morphological diversification, adaptive evolution and species formation. We have undertaken a large scale study to characterize rapidly evolving genes in Drosophila to systematically identify and characterize the proportion of REGs in testis, ovary and head tissues of Drosophila and study the forces responsible for their rapid divergence. We used high stringency southern hybridizations to screen a subset of testis, ovary and head cDNA libraries between related species of Drosophila. Signal ratio distribution from these assays reveal that testis has the highest proportion of REGs followed by ovary and then head. This trends indicates the existence of selection on reproductive genes. Sequence analysis of a subset of REGs from each tissue substantiates this trend. Moreover, ovary and testis cDNAs show higher Ka/Ks ratios among members of the *D. melanogaster* complex indicating that these proteins are subject to directional selection. The higher divergence of testis transcripts over both ovary and head provides a mechanism for faster male evolution and faster evolution of male hybrid sterility over inviability. These data suggest inherent differences in the pattern of gene evolution in reproductive and non-reproductive tissue, which may be driven by sexual selection.

JEFFERY, William
Evolution of Eye Degeneration in Cavefish
Jeffery W. R. [1]. (1) University of Maryland
We study the evolution of eye degeneration in the teleost *Astyanax mexicanus*, a single species with eyed (surface fish) and eyeless (cavefish) forms. In addition to eye and pigment regression, cavefish exhibit increased numbers of taste buds and other feeding structures. A candidate gene survey showed that about150 key eye regulatory and structural genes are expressed normally in cavefish. However, a small number of these genes show subtle changes in their embryonic expression patterns. The midline signaling gene *Sonic hedgehog* (*Shh*) and its downstream target genes are expanded in cavefish embryos. *Shh* overexpression in surface fish embryos resulted in a blind cavefish phenocopy. We have also shown that

Shh expression is sufficient and necessary for increasing taste bud development in cavefish. The results support a trade-off hypothesis in which adapative and regressive characters evolve in concert based on the pleiotropic effects of *Shh*.

JENSEN, Jeffrey

The Effects of Demography on Tests of Neutrality: A goodness-of-fit test to distinguish rejections due to demography from selection

Jensen J. D. [1], Kim Y. [2], Bauer DuMont V. [1], Bustamante C. D. [1], Aquadro C. F. [1]. (1) Cornell University, (2) University of Rochester

A number of statistical tests have been proposed to detect directional selection, based on DNA polymorphism data. The theory of genetic hitchhiking predicts that sufficiently strong, recent directional selection will greatly reduce variation both at the target of selection as well as at linked sites. One recent approach that both incorporates various aspects of the data as well as predicts the putative target of selection is the composite likelihood method proposed by Kim and Stephan (2002). However, through simulations, we demonstrate that this test is not robust to a number of demographic and population scenarios, with some realistic population parameters for Drosophila resulting in a high rate of false-positives. We explore a goodness-of-fit test that may allow for the discrimination of rejections based on selection from those based on population and demographic forces.

JENSEN-SEAMAN, Michael

Genetic Structure of the Laboratory Rat (*Rattus norvegicus*)

Jensen-Seaman M. I. [1], Orlebeke K. [1], Nihart J. [1], Nie J. [1], Thomas M. A. [2], Jacob H. J. [1]. (1) Medical College of Wisconsin, (2) Idaho State University

The genome sequence of the rat will facilitate its use in evolutionary studies, ranging from genome evolution to identifying the molecular basis for phenotyping traits, in addition to its long history as a model for human diseases. Like the mouse, rat strains were derived from wild ancestors, although it is believed that all rat strains are derived from a single subspecies. To understand the patterns of genetic diversity in the autosomal, mitochondrial (mtDNA), and Y-chr genomes of inbred rat strains, we sequenced the mtDNA D-loop and mtDNA Cox1 from 48 commonly used strains and three wild rats. We analyzed patterns of allele sharing at ~4800 nuclear microsatellites in these same strains. Finally, we are presently developing and genotyping Y-chr markers in the 48 inbred strains. Results from mtDNA show three major clades of inbred rats (maximum divergence of 1.24% in the D-loop), one corresponding completely with a group of Japanese strains identified from the autosomal data. These results, and ongoing analyses, indicate that inbred strains of rat posses a complex mix of genetic diversity in the autosomal, mtDNA, and Y genomes, a result of a complex breeding history.

JOHANNESSON, Hanna

Evolution of surface antigens of the human pathogenic fungi Coccidioides immitis and C. posadasii

Johannesson H. [1], Taylor J. W. [1]. (1) UC Berkeley, Dept. of Plant and Microbial Biology

We have investigated the evolution of two antigens (Proline-rich antigen, *PRA*, and Spherule outer wall glycoprotein, *SOWgp*) in the human pathogens *Coccidioides immitis* and *C. posadasii*. By using likelihood-based methods to compare models of selective pressure among codons we verified that *PRA* evolves under positive selection. No evidence of diversifying selection acting on *PRA* was found, thus the increased rate of evolution is not a result of avoidance of the host's immune system. The analyses suggest that selection was not stronger on the branch separating pathogenic and non-pathogenic species in the phylogeny of *Coccidioides* spp and their sister taxa, and we suggest that positive selection act on *PRA* as a consequence of spore cell-wall morphogenesis unique to each species. We found that the central part of *SOWgp* consists of repetitive units of 41-47 aa, evolving under concerted evolution by unequal crossing over. The species share repeat number polymorphism, indicating that *SOWgp* is under balancing selection.

JORDAN, King

Sequence evolution and the human gene expression network

Jordan I. K. [1], Mariño-Ramírez L. [1], Wolf Y. I. [1], Koonin E. V. [1]. (1) National Center for Biotechnology Information

I will describe an integrated analysis of genome-scale sequence and expression data that was used to examine the interplay between two sources of order, natural selection and physical self-organization, in the evolution of human gene expression. The topology of a human gene expression network, derived from tissue-specific expression profiles, shows scale-free properties that imply evolutionary self-organization via preferential node attachment. Genes with numerous co-expressed partners evolve more slowly on average than genes with fewer co-expressed partners, and genes that are co-expressed show similar rates of evolution. Surprisingly, however, we found no connection between the rate of gene sequence divergence and the extent of gene expression profile divergence between human and mouse. This suggests that distinct modes of natural selection may govern these processes, and I will propose a model of how convergence may influence the evolution of gene expression patterns.

JØRGENSEN, Frank G.

Comparative analysis of protein coding sequences from Human, Mouse and the domesticated Pig

Jørgensen F. G. [1], Hoboth A. [2], Hornshøj H. [3], Bendixen C. [3], Fredholm M. [4], Schierup M. H. [1]. (1) Department of Genetics and Ecology, Aarhus, (2) Bioinformatics Research Center (BiRC), Aarhus, (3) Danish Institute of Agricultural Sciences, (4) Department of Animal Science and Animal Health

We performed an evolutionary analysis on a collection of 1120 full length cDNAs from the domesticated pig. These were aligned to orthologues from human and mouse. The Japanese puffer fish *Fugu ruprices* was used as outgroup in the phylogenetic analysis. This analysis shows that rodents and artiodactyls split from the primate lineage within a very short period of time, but with rodents as outgroup to primates and artiodactyls. Since diversification, the pig and mouse lineage have on average experienced 1.44 and 2.86 times as many synonymous substitutions as humans, respectively, whereas the rates of non-synonymous substitutions are more similar. The analysis shows a higher average dN/dS ratio in the human lineage compared to the pig and mouse lineage. Codon based maximum likelihood models were used to investigate which genes that may have undergone positive Darwinian selection in each of the three lineages. Evidence of positive selection was detected in approximately 12.8 percent of the genes in human and 8.5 percent in pig.

KAHILA BAR-GAL, Gila

Genetic and morphometric differences between two populations of gray wolves in Israel: hybrids or not?

Kahila Bar-Gal G. [1], Tchernov E. [2], Greenblatt C. [3], O'Brien S. J. [1]. (1) NCI, (2) Evolution, Systematics and Ecology, The Hebrew University, (3) Kuvin Center for the Study of Infectious and Tropical Diseases

From the archaeological records, the gray wolf (*Canis lupus*) is known to have inhibited Israel since the late Pliocene period. During the Neolithic revolution wolves were domesticated in the southern Levant and since then, domestic dogs and wolves have lived together in the same region. The close contact between domestic dogs and wolves may have allowed for hybridization to occur between both species. Moreover, rapid human population growth in the last 50 years, has forced the wolves into increasingly smaller areas, fragmenting the population. This process has created two main populations of wolves, one in the Golan Heights and the other in the Negev and Arava. Therefore, the threat of cross breeding between domestic dogs and wild wolves has increased. We used both molecular and morphometric methods to characterize the wolf populations of Israel for evolutionary and conservation studies. Both nuclear and mitochondrial markers were amplified from modern and museum specimens. These results indicate that the current populations are significantly different from each other and should be considered as distinct managements units. In addition, hybridization appears to have occurred at higher frequencies in the northern population.

KAMAU, Esther

Linkage disequilibrium and diversity around the self-incompatibility locus.

Kamau E. [1], Charlesworth D. [1]. (1) University of Edinburgh

In *A. lyrata* the self-incompatibility (S-) locus region contains the genes encoding the pistil (female) and pollen recognition proteins, respectively SRK, SCR (or SP11 in Brassica). Tight linkage between the two genes is expected, because recombinant haplotypes would be self-compatible, while incompatibility requires co-adaptation of the two genes' sequences. The SRK gene has high diversity in natural populations of *A. lyrata*, as predicted since balancing selecion should act on S-alleles. If recombination occurs in the S-locus region, diversity should decrease with distance from the selected genes. We estimated polymorphism in four genes flanking the SRK gene, and used two-full sib families to test whether variants in these sequences co-segregate with the S-locus. Given that the S-loci are probably located in the middle of a long chromosome arm, which should recombine freely, our results suggest local suppression of recombination, but possibly only over a small region.

KANEKO, Satoko

Processed pseudogenes for regulating functional Makorin genes in mammals

Kaneko S. [1], Takahata N. [1], Satta Y. [1]. (1) Dept. of Biosystems Science, Graduate University for Advanced Studies

Makorin is a highly transcribed ribonucleoprotein with zinc-finger motifs and is encoded by multiple loci. A recent study by Hirotsune et al. (2003 Nature 423:91-96) demonstrated that in mice, the gene expression of functional Makorin1 is controlled by an expressed processed pseudogene Makorin1-p1. However, the high sequence similarity with Makorin1 immediately suggests that Makorin1-p1 arose relatively recently by retrotransposition, and the high Ka/Ks ratio implies that Makorin1-p1 has evolved in a neutral fashion. We therefore study the evolutionary history of Makorin1-p1 in genus Mus and argue that the regulation mechanism of Makorin1 gene expression is genus-specific. For this reason, we examine if Makorin1 gene expression in other mammals is similarly regulated by their own processed pseudogenes. A particular attention is paid to Makorin4 specific to the primate lineage. We will discuss the generality of processed pseudogenes for regulating functional Makorin gene expression.

KASUGA, Takao

Comparative genomics within the genus Neurospora

Gilbert L. B. [1], **Kasuga T.** [1], Townsend J. [1], Glass L. [1], Taylor J. W. [1]. (1) Plant & Microbial Biology, UC Berkeley

Little is known about the process of microbial adaptation. We are using microarray technology to help identify genes responsible for environmental adaptation in a model filamentous fungus *Neurospora*. *Neurospora* offers a unique opportunity to characterize the variability of global gene regulation within and between species. The genus *Neurospora* consists of eight closely related conidiating species, as well as several non-conidiating species. We have constructed a 70mer oligomer array for *Neurospora crassa* representing 3,366 genes, approximately one third of the genome. To assess the effectiveness of our array for other members of the *Neurospora* genus we have analyzed comparative genomic hybridizations for all five conidiating species of *Neurospora* as well as a few non-conidiating isolates. This technique uses genomic DNA as a substrate for labeling and hybridization to an oligo array. We can now estimate genome divergence among *Neurospora* species by comparing the ratio of fluorescences between samples.

KATO, Yumiko

Recent independent origins of annual *Zostera marina* from the perennial

Kato Y. [1], Aioi K. [2], Takahata N. [1], Satta Y. [1]. (1) The Graduate University for Advanced Studies (Sokendai), (2) Aoyama Gakuin Women's Junior College

Seagrasses are angiosperms and consist of five families, which are thought to become adaptive to aquatic life independently. Seagrass beds support the high biodiversity and productivity of coastal waters. Rapid

loss of seagrass beds, due mainly to artificial disturbances, endangers seagrass species worldwide. From a viewpoint of conservation biology, it is imperative to understand the origin and evolution of seagrass species. Zosteraceae is one of the five seagrass families. High species diversity of Zosteraceae is observed around the Japan Archipelago, suggesting that the family has evolved therefrom (Kato et al. 2003). *Z. marina* has two forms, perennials and annuals, around the Japan Archipelago. Although annuals might be derived from perennials, the genetic relationships have not been clarified yet. The analysis of nuclear *phyA* gene sequences shows recent origins of annuals from perennials independently in different geographic areas. It is argued that both environmental and genetic factors are involved in this transition. Kato, Y. et al. 2003. Genes Genet. Syst. 78. 329-342.

KAWASAKI, Kazuhiko

Genetic basis for the evolution of vertebrate mineralized tissue.

Kawasaki K. [1], Weiss K. M. [1]. (1) Dept. of Anthropology, Pennsylvania State University

Mammalian teeth form on extracellular matrix proteins; enamel on enamel-specific proteins, dentin and bone on collagen, with acidic proteins regulating their growth. We found that the enamel-specific proteins and dentin/bone acidic proteins are related to an avian eggshell matrix protein, mammalian milk caseins, and salivary proteins. The genes for these proteins arose from SPARC that codes a non-collagenous bone matrix protein, and form the secretory Ca-binding phosphoprotein (SCPP) gene family. The duplication of SPARC generated SPARCL1 in ancient bony fish, and tandem duplications of SPARCL1 created the many SCPP genes. Bony fish developed a dentin/bone SCPP; tetrapods enamel SCPPs; birds an eggshell SCPP; and mammals milk/saliva SCPPs. This suggests that the mechanism of mammalian tissue mineralization was developed in bony fish or tetrapods, and parallel functional specialization of duplicated genes facilitated adaptive evolution in vertebrates.

KETCHAM, Kelly

MC1R variation is not linked to small-scale melanic plumage patterns in two passerine birds.

Ketcham K. D. [1], Gibbs H. L. [1], Fleischer R. C. [2], Greenberg R. S. [2], Vickery P. D. [3]. (1) The Ohio State University, (2) Smithsonian Institute, (3) Center for Ecological Research

Studies that link molecular variation in the melanocortin-1 receptor (MC1R) gene with total body melanism in vertebrates have recently become models for understanding the molecular basis of adaptive variation in natural populations. We examined whether MC1R variation can be correlated with small-scale melanic plumage patterns by comparing MC1R sequences in melanic and non-melanic subspecies of two passerine birds: Swamp Sparrows (*Melospiza georgiana*) and Grasshopper Sparrows (*Ammodramus savannarum*). The levels of variation found in each species were typical of passerine birds but no fixed nonsysnonymous substitutions were found that correlated with plumage differences in either species. We conclude that variation in the coding region of the MC1R gene is not responsible for these small-scale pigment pattern variations and that variation in this gene may only determine rare instances of large-scale, whole body melanism in birds and not more common small-scale patterns.

KIM, Junhyong

CIPRes: Building the compuational infrastructure for estimating Tree of Life and benchmarking phylogeny algorithms

Kim J. [1], Hillis D. M. [2], Turner P. E. [3], Ancel-Meyers L. [2], Muse S. [4]. (1) University of Pennsylvania, (2) University of Texas, Austin, (3) Yale University, (4) North Carolina State University

The Cyber Infrastructure for Phylogenetic Research (CIPRes) project is a NSF funded consortium of 44 researchers from 14 institutions whose goal is to develop the computational infrastructure supporting the Tree-Of-Life initiative. One major component of the CIPRes project is the development of an extensive set of simulated and real datasets that will provide a uniform evaluation of phylogeny algorithms. In this talk, we present the broad outline of the CIPRes project and our approach for developing large-scale evolutionary simulations with high degree of complexity. We present preliminary results on a one-

million taxa simulation, a database for storing and querying trees and data at millions of taxa scale, and an example high dimensional complex simulated data and its effects on phylogenetic reconstruction. The ultimate goal of the CIPRes benchmark effort is to develop a broadly accepted set of evaluation methods. Therefore, part of the goal of this talk is to solicit community input for specifications for algorithm evaluations, evolutionary models, and database queries.

KITANO, Takashi
Genome Sequence Analyses of Human, Chimpanzee, and Gorilla HOXA Clusters
Kitano T. [1], Kim C. G. [1], Sumiyama K. [1], Dewar K. [2], Kohara Y. [1], Saitou N. [1]. (1) National Institute of Genetics, (2) McGill University
HOXA cluster regions of chimpanzee and gorilla were sequenced over 200kb and were compared with the human HOXA cluster region. Long repetitive elements were almost absent in the HOXA cluster regions. This observation supports the idea that one of the selective forces keeping the genes in the HOXA cluster tightly arranged stems from the fact that adjacent genes share common cis-regulatory elements. Nucleotide difference of HOXA cluster region between human and chimpanzee (0.75%) was lower than that of the genome-wide value (1.23%). Average dN (0.13%) and dS (0.69%) for human branch were slightly lower than those of average dN (0.28 %) and dS (0.78 %) for 103 genes (Kitano et al. in press). In the HOXA cluster region, the proportion of GC -> AT substitutions of the human lineage was significantly higher than those of the chimpanzee and the gorilla lineages, and the proportion of AT -> GC substitutions of the human lineage was lower than those of the chimpanzee and the gorilla lineages.

KITAZOE, Yasuhiro
A unified index for estimation of homoplasy and molecular phylogeny with application to mammalian genomes
Kitazoe Y. [1], Kishino H. [2], Okabayashi T. [1], Watabe T. [1], Okuhara Y. [1], Kurihara Y. [1]. (1) Kochi University, (2) University of Tokyo
With large sized genomic databse, any biases in the models of molecular evolution may lead to significant confusion. In particular, existing procedures of phylogenetic inference assume absence of correlated evolutions among lineages. Here, we propose a unified index for simultaneous estimation of homoplasy and phylogenetic trees. By way of multidimensional vector space representation of taxa based on estimated pairwise distances, the index is defined as the deviations from the expected pattern of tree structure. The biases in the distances are corrected so as to minimize the index. We detected strong biases of the estimated distances in three representative large datasets from placental mammals. While previous results showed incongruence between the datasets, our bias correction yielded a consistent topology, which implies paraphyly of insectivores. This is a sharp contrast to the hypothesis of multiple origins and is consistent with the palaeontological view.

KLEIN, Jan
Origins of the adaptive immune system in vertebrates
Klein J. [1]. (1) The Pennsylvania State University
An adaptive immune system (AIS) based on the use of major histocompability complex (Mhc) molecules, T-cell receptors (Tcrs), B-cell receptors (Bcrs = immunoglobulins), and proteins encoded in recombination activating genes 1 and 2 (RAG1 and RAG2), as well as bona fide lymphocytes and the thymus is apparently restricted to jawed vertebrates (gnathostomes). The status of jawless vertebrates (agnathans) with respect to these principal components of the AIS will be summarized and the possible origin of these components discussed.

KO, Wen-Ya
Lineage-specific molecular evolution in the Drosophila melanogaster species subgroup
Ko W-Y. [1], David R. M. [1], John A. [1], Lin C-F. [1], Piao S-F. [1], Akashi H. [1]. (1) The Pennsylvania State

University
We studied lineage-specific patterns of codon usage and protein evolution in six D. melanogaster subgroup species (D, melanogaster, D. simulans, D, teissieri, D. yakuba, D. erecta, and D. orena) and two ancestral lineages (D. teissieri-D. yakuba and D. erecta-D. orena). Analyses of 16 nuclear genes (8042 codons) revealed strong departures from equilibrium at silent sites in five of the 8 lineages examined. Four of the lineages show excesses of unpreferred silent substitutions and one of the lineages shows a recent increase in codon bias. These trends are consistent among genes. Rates of protein evolution also show strong lineage effects, but associations between changes in codon bias and patterns of protein evolution are unclear. The magnitude of parameters
governing both silent and protein evolution appear to vary frequently, and in a lineage-specific manner, among these species. Mutational and selective mechanisms underlying these patterns will be discussed.

KOCHER, Thomas D.
Rapid Speciation of Cichlid Fishes in African Lakes
Kocher T. D. [1]. (1) University of New Hampshire
East African cichlids have radiated into ~2000 species in the last 5MY, including more than 500 ecologically diverse species in Lake Malawi. Phylogenetic studies suggest a pattern of radiation in stages. The earliest divergence resulted in clades specialized for different physical habitats (e.g. rock vs. sand). This was followed by specialization of the oral jaws to exploit multiple feeding strategies within each habitat. The final stage is the diversification of species with unique male color patterns. We are working to identify the genes underlying these traits to ground theoretical models of speciation. Differences in jaw morphology are controlled by a relatively small number of QTL which show evidence of consistent directional selection. Using genetic and physical linkage maps developed for a related cichlid, we are cloning genes controlling color pattern and sex determination, with the ultimate goal of placing these sequence polymorphisms in historical context.

KONG, Hongzhi
Rapid Birth-and-Death Evolution in the *SKP1* Gene Family
Kong H. [1], Leebens-Mack J. [1], Albert V. A. [2], Ma H. [1], dePamphilis C. W. [1]. (1) The Pennsylvania State University, (2) University of Oslo
As a core component of the SCF ubiquitin ligases, the Skp1 protein can regulate many fundamental processes in eukaryotes. Previous studies indicate that multiple Skp1 homologs from the same plant or animal species evolved at highly heterogeneous rates; the rate heterogeneity is so severe that long-branch attraction always obscures the true relationships in the phylogenetic analysis of the entire gene family. Here we show that, within flowering plants, the most slowly evolving Skp1 proteins follow the organismal phylogeny, but the moderately and rapidly evolving genes appear to be members of large, species-specific clades. Significant differences in the gene families were observed even when SKP1 genes from the two rice genomes were compared, suggesting a high rate of gene turnover. *dN*/*dS* analysis further indicated that the vast majority of the moderately and rapidly evolving *SKP1* genes have evolved under relaxed or altered constraint; some may have even become pseudogenes. On the basis of these observations, a modified birth-and-death model was proposed to explain the asymmetric evolution of *SKP1* genes.

KOONIN, Eugene V.
A universal trend of amino acid gain and loss in protein evolution: The modern echo of code origin
Koonin E.V.[1], Jordan I.K.[1], Kondrashov F. A.[2], Adzhubei I.A.[3] Wolf Y.I.[1], Kondrashov A. S.[1], Sunyaev S.[3.] (1) National Center for Biotechnology Information, NIH, (2) Section of Evolution and Ecology, University of California at Davis, (3) Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School.

Amino-acid compositions of proteins differ substantially between taxa and, thus, can evolve. For example, proteins from organisms with GC-rich (AT-rich) genomes contain more (less) amino acids encoded by GC-rich codons. However, no universal trends in ongoing changes of amino acid frequencies have been reported. We compared sets of orthologous proteins encoded by triplets of closely related genomes from 15 taxa representing all three domains of life, bacteria, archaea, and eukaryota. Five amino acids (Cys, Met, His, Ser, and Phe) are accumulated in at least 14 taxa, whereas four others (Pro, Ala, Glu, and Gly) are consistently lost. The same nine amino acids are also currently accumulated (lost) in human proteins, as revealed by human polymorphism analysis. All amino acids with declining frequencies are thought to be among the first incorporated into the genetic code; conversely, all amino acids with increasing frequencies, except Ser, were probably recruited late. Thus, expansion of initially underrepresented amino acids, which begun over 3 billion years ago, continues to this day.

KOSIOL, Carolin

Protein Sequence Evolution: Markov or non-Markov?

Kosiol C. [1], Goldman N. [1]. (1) EMBL - European Bioinformatics Institute

In 1992, Henikoff and Henikoff derived the series of BLOSUM matrices whose elements are probabilities of amino acid substitutions, but are not based on a Markov model. BLOSUM matrices often perform better than evolutionary models for the purpose of comparing protein sequence alignments or database searches. It is unclear why this should be, but it may be because protein sequences behave in a non-Markovian manner. We show that some of the non-Markovian behaviour observed in the literature can be explained by an aggregated Markov process (AMP) which incorporates rate heterogeneity among different codon sites of the protein and the properties of the amino acids encoded by the sequence.

KOUPRINA, Natalia

The SPANX gene family of cancer-testis specific antigens: rapid evolution and amplification in African great apes and hominids

Kouprina N. Y. [1]. (1) NCI

Human SPANX genes comprise a gene family with five members (SPANX-A1, -A2, -B, -C, -D) encoding cancer-testis-specific antigens. These highly similar, paralogous genes cluster at Xq27. We isolated primate genomic clones homologous to human SPANX. Their analysis revealed an uncharacterized group of genes, SPANX-N, present in all primates and in mouse and rat. In humans, four SPANX-N genes comprise a series of tandem duplicates at Xq27; a fifth member is located at Xp11. SPANX-N genes are expressed in normal testis and melanoma cell lines; testis-specific expression of SPANX is conserved in mouse. Analysis of the taxonomic distribution of the long and short forms of intron indicates that SPANX-N is the ancestral form, from which the SPANX-A/D subfamily evolved in the common ancestor of the hominoid lineage. Strikingly, the coding sequences of the SPANX genes evolved much faster than the intron and 5'-untranslated regions. There is a strong correlation between the rates of evolution of synonymous and non-synonymous codon positions, both of which are accelerated twofold or more compared to the non-coding sequences. Thus, evolution of the SPANX family appears to have involved positive selection that affected not only the protein sequence but also the synonymous sites in the coding sequence.

KROLL, Evgueny

Starvation-associated genomic rearrangements and speciation in yeast

Kroll E. [1], Dunn B. [2], Coyle S. [3]. (1) Molecular Sciences Institut, (2) Stanford University, (3) University of California, Berkeley

It is thought that genomes do not change at high rates. We found that complete starvation dramatically increased the frequency of gross genomic rearrangements and chromosome duplications in the yeast Saccharomyces cerevisiae. Additionally, certain loci that are involved in metabolism, nutrient uptake, protein biosynthesis and oxidative stress were deleted or amplified in the majority of cells that survived

starvation. Consequently, we found that a subset of starved cells formed low-fertility hybrids in backcrosses, having become reproductively isolated from their ancestral population. The reproductively isolated clones were self-fertile, potentially becoming the founders of new species. Since starvation is a common and general condition in the wild, our results imply that evolutionary processes in populations under environmental stress may occur at rates much higher than previously thought. Also, stress might provide a means for sympatric speciation within the starving population.

KUMAR, Sudhir

Precision and Robustness of the genomic timescales

Kumar S. [1], Filipski A. [2], Hedges S. B. [3]. (1) Center for Evolutionary Functional Genomics, ASU, (2) Center for Evolutionary Functional Genomics, (3) Department of Biology, The Pennsylvania State University

Molecular data are emerging as an additional and indispensable means to determine the time of species divergence. Least squares, maximum likelihood, and Bayesian approaches are available to generate divergence time estimates from molecular data under local and global molecular clocks. All of these methods require *a priori* knowledge of constraints (in form of divergence time) upon one or more nodes in the evolutionary tree to calibrate the molecular clock, which is then employed to convert the observed sequence divergences at other nodes to time using an appropriate model of sequence evolution. Precision of the estimated molecular times is affected by a number of different attributes of the data, including the fossil (or calibration) divergence time uncertainty, gene and site sampling variances, rate variation among lineages, and the choice of the method used to combine information from multiple genes. We have systematically examined the affect of these sources of error on molecular time estimates when using the least squares and Bayesian methods. We present our assessment of the relative and absolute contributions of these factors in determining standard errors and confidence intervals of the molecular time estimates.

KUMAR, Sudhir

***MEGA3*: An Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment**

Kumar S. [1], Tamura K. [2], Nei M. [3]. (1) AzBio, Arizona State University, (2) Tokyo Metropolitan University, (3) The Pennsylvania State University

We announce the release of the Molecular Evolutionary Genetics Analysis software version 3.0 (*MEGA3*) for exploring and analyzing DNA or protein sequences from an evolutionary perspective. *MEGA3* provides an intuitive solution for creating, editing, and visualizing sequence alignments using a newly included multiple sequence alignment engine (based on CLUSTALW source code). It has an extensive repertoire of methods for estimating evolutionary distance estimation from nucleotide and amino acid sequence data, three different methods of phylogenetic inference and two statistical tests of topological differences. Tests for molecular clock as well as substitution/mutation pattern homogeneity among lineages are included. In addition, *MEGA3* computes statistical quantities such as nucleotide and amino acid frequencies, transition/transversion biases, codon frequencies (codon usage tables), and the number of variable sites in specified segments in nucleotide and amino acid sequences. *MEGA3* also facilitates analyses of large and multigene datasets by providing services for creating groups of sequences and specifying domains and genes. Advanced on-screen sequence data and phylogenetic-tree editors facilitate publication-quality outputs. Integrated and interactive designs, on-line context-sensitive helps, and a text- file editor make *MEGA* easy to use on the Microsoft Windows platform. *MEGA3* can be downloaded free of charge from *http://www.megasoftware.net*.

KURAKU, Shigehiro

Identification of gene co-option involved in turtle shell evolution

Kuraku S. [1], Usuda R. [1], Kuratani S. [1]. (1) Evolutionary Morphology, RIKEN CDB

The shell is one of the most distinct synapomorphies that define turtles. Unlike in other amniotes, turtle

ribs grow laterally to be entrapped in dorsal dermis to form the carapace. In early stages of development, turtle embryos are characterized by the carapacial anlage called carapacial ridge (CR). In order to reveal molecular mechanisms that permitted the acquisition of this evolutionary novelty, we performed a systematic screening of genes expressed specifically in the CR of soft-shelled turtle, *Pelodiscus sinensis*, by means of microbeads-based competitive hybridization. As a result of the validation of CR-specific expression by real-time PCR and *in situ* hybridization, orthologues of mammalian *Sp5*, *CRABP-I*, *APCDD1* and *LEF1* were identified. Phylogenetic tree inference and comparative expression analyses indicated that these four genes were recruited for expression in CR in the turtle lineage without undergoing any gene duplication. Functional analyses are now underway to understand how these genes are cooperatively acting to realize turtle-specific body patterning.

LANDRY, Christian

Genetic variation for phenotypic plasticity in budding yeast: effects on the regulatory network

Landry C. [1], Oh J. [2], Hartl D. L. [1], Cavalieri D. [2]. (1) Organismic and Evolutionary Biology, (2) Bauer Center for Genomics Research

Phenotypic plasticity is defined as the ability of a genotype to display different phenotypes. Gene regulation is a perfect example of plastic trait as it is modulated by environmental conditions. In the presence of Genotype X Environment interaction, or genetic variation for phenotypic plasticity, a genotypic value is not adequately defined by focusing on a phenotype (expression) in one environment but is rather better described by a reaction norm, e.g. variation as a function of the environment. The presence of G X E has important impacts on the maintenance of genetic variability, the estimation of quantitative genetic parameters and on the outcome of evolution in variable environments. It can also significantly change the genetic architecture of quantitative traits by changing the signs of genetic correlations across environments. In the present study, we use cDNA arrays to examine gene expression variation in multiple strains of *Saccharomyces cerevisiae* grown in different environments. We investigate the relative contribution of the genotypes and the environments in shaping phenotypic variation. We focus mainly on the importance of G X E and its impact on the patterns of genetic correlation among traits and the make-up of the underlying genetic regulatory network.

LANDWEBER, Laura

Patterns and Process of Gene Scrambling and Unscrambling in Ciliates

Kuo S. [1], Chang W. J. [1], Mollenbeck M. [2], Wong L. C. [1], Cavalcanti A. [1], Lipps H. J. [2], **Landweber L. F.** [1]. (1) Princeton University, (2) University of Witten

Spirotrichous ciliates undergo massive DNA elimination and rearrangement of their ~2.5 Gb germline micronuclear genome to construct the set of ~2 kb "nano-chromosomes" that comprise their ~50 Mb gene-rich somatic macronuclear genome. In many species, we estimate that 20-30% of all genes are "scrambled," i.e. both fragmented and permuted into several small unordered segments in the germline. These macronuclear-destined segments can be present on either strand within a locus, or even dispersed over unlinked loci in the germline. Experiments in our laboratory have surveyed the origin, evolution, and processing of scrambled genes in stichotrichous ciliates. I will describe new complex patterns of scrambled genes. For example, in one case in *Uroleptus* the macronuclear-destined segments for two independent transcripts are intertwined on two separate micronuclear loci. I will also describe our studies examining molecular intermediates in the developmentally regulated process of gene unscrambling in *Stylonychia* and *Oxytricha*.

LAPIERRE, Pascal

Whole genome analyses on the class levels: is there a consensus phylogeny?

Lapierre P. [1], Olga Zhaxybayeva O. [1], Gogarten J. P. [1]. (1) University of Connecticut

With availability of many completely sequenced genomes, it has been shown that every single genome has undergone many dynamic changes throughout evolution. Helped by horizontal gene transfers

(HGT), gene duplications, deletion pressure, genome inversions and other genome rearrangements, even two strains of the same species can have very different gene content. Individual genes often have different evolutionary histories, casting doubts that phylogenies calculated from single genes reflect the organismal phylogeny.

We performed comparative genome analyses for two classes of Bacteria: the cyanobacteria where HGT is inferred to have been a major contributor in the evolution of the clade and the gamma proteobacteria, a more heterogeneous group, where HGT appears to have been less of a factor. The latter finding might be due to the close association that some of the gamma proteobacteria have with multicellular eukaryotes (parasitic, symbiotic, or commensal), or due to the reduced genome size of some of the organisms included in the analysis. Two different approaches were applied to assess a genome mosaicism using whole genome analysis for those two classes of bacteria. The obtained results show that the extent of genome mosaicism is different for different classes of bacteria.

LAWTON, Betty R.

The Search for Genetic Footprints of Intergenomic Conflict: Imprinting and Rapid Evolution of IGF2 in the Poeciliid Fishes

Lawton B. R. [1], Mateos M. [2], O'Neill R. J. [1], Oshiro W. [1], Reznick D. [3], O'Neill M. J. [1]. (1) University of Connecticut, (2) University of Arizona, (3) University of California-Riverside

The Parent-Offspring Conflict model for the evolution of imprinting in mammals elegantly explains the correlation between the growth regulatory interests of mammalian parents, and the growth regulatory function and parent-specific expression profiles of imprinted genes. Comparative studies involving two oppositely imprinted genes, IGF2 and its growth antagonist IGF2r seem to demonstrate a link between the evolution of placentation and imprinting. Technical limitations, however, in studies of monotremes and birds and the likely common origin of placentation in marsupials and eutherians diminish the authority of these studies. The developmental expression profile and sequence evolution of IGF2 was examined in a family of live-bearing fishes, the Poeciliidae, in which placentation has evolved independently several times. While no evidence of imprinting was found in these fishes, extraordinarily high dn/ds ratios for IGF2, within certain lineages of this family, is suggestive of directed evolution driven by genetic conflict.

LAZZARO, Brian

Evolutionary and Quantitative Genetics of *D. melanogaster* innate immunity.

Lazzaro B. P. [1], Sackton T. B. [2], Clark A. G. [3]. (1) Entomology, Cornell University, (2) Ecology and Evolutionary Biology, Cornell, (3) Molecular Biology and Genetics, Cornell

Wild *D. melanogaster* populations harbor substantial genetic variability in the capacity to suppress bacterial infection. We have statistically tested associations of phenotypic variation in resistance to infection by mutiple bacteria with molecular variation in candidate immune response genes. These associations reveal genes carrying allelic variation with general effects on immune function (i.e., the effect is the same across pathogens), with pathogen-specific effects (e.g., significant effect on the response to Gram-positive bacteria, but no effect on anti-Gram-negative responses), and, intriguingly, with antagonistic effects (effects in opposite direction depending on the pathogen). We compare the quantitative genetic effects we observe with published and unpublished molecular population genetic data to draw general conclusions about the origin, mechanism and intensity of selection on specific *D. melanogaster* immune response genes.

LEEBENS-MACK, Jim

The utility of whole chloroplast genome sequencing for reconstructing deep nodes in plant phylogenies with an example from basal angiosperms

Leebens-Mack J. H. [1], dePamphilis C. W. [1], Raubeson L. A. [2], Cui L. [1], Zhang Y. [1], McNeal J. [1], Boore J. [3],

Fourcade M. [3], Kuehl J. V. [3], Wyman S. K. [4], Jansen R. K. [5]. (1) Department of Biology, Penn State University, (2) Dept. of Biol. Sci., Central Washington University, (3) DOE, Joint Genome Institute, (4) Dept. of Computer Science, University of Texas, (5) Section of Integrative Biology,University of Texas
Most recent molecular phylogenies based on one to several genes have suggested that the monotypic *Amborella*, or *Amborella* plus waterlilies, represents the basal, extant angiosperm lineage. Last year the chloroplast genome of *Amborella* was sequenced and all 61 protein coding genes shared with 12 other available land plant chloroplast genomes were used to assess its phylogenetic position. Analyses of both nucleotide and amino acid sequence data placed the monocot lineage, rather than *Amborella*, at the base of the angiosperms. We have sequenced five new chloroplast genomes to determine if limited taxon sampling could explain this surprising result. These include two monocots (*Acorus* and *Yucca*), a waterlily (*Nuphar*), a basal eudicot (*Ranunculus*), and a gymnosperm (*Ginkgo*). Phylogenetic analyses of both amino acid and nucleotide sequences provide strong support for the basal position of *Amborella*, and parametric bootstrap analyses implicate long-branch attraction in the original study.

LIANG, Han
Tandem Stop Codon Analysis in Yeasts
Liang H. [1], Cavalcanti A. R. [1], Landweber L. R. [1]. (1) Princeton University
It has been suspected for a long time that there is another stop codon following the real stop codon, as a backup if the real one were read through by a near cognate tRNA. However, the concept of a "tandem stop codon" still remains elusive. Here we found that a statistical excess of stop codons has evolved at the third codon position following the real stop codon UAA in yeasts. Comparative analysis indicated that the stop codons at this position are much more conserved than sense codons, suggesting the tandem stop codons are maintained by selection. We then further evaluated the influence of expression level and biological importance of genes on the distribution of tandem stop codons. Our results suggest that expression level of genes is the primary factor influencing tandem stop codons. The possible biological reasons underlying tandem stop codons will be discussed.

LIN, Zhenguo
Plant Homologs of Yeast DNA Repair Genes
Lin Z. [1], Chalkia D. [1], Makalowski W. [1], Ma H. [1]. (1) Pennsylvania State University
Multiple genes related to DNA repair and recombination have been studied in Saccharomyces cerevisiae and found to play important roles during meiosis. Although, protein homology does not guarantee functional identity, there is a good chance that two homologous proteins perform similar functions. To identify plant homologs of these genes, we selected twenty yeast protein sequences to search different proteome databases and performed further phylogenetic analysis and architecture analysis on potential homologs. Nine of these yeast proteins were found to have clear plant homologs (DMC1, HOP1, MRE11, MSH3, MUM2, RAD10, RAD27, RAD50, and SPO11), and two others (REC104, REC107) had potential homologs that require more careful analysis. We will discuss evolutionary history of these proteins in functional context. Further analysis of these homologs can provide valuable clues to study the functions of meiosis-related genes in plants and both the conservation and divergence of regulations of meiosis.

LIN, Chiao-Feng
Phylogenetic Analysis of U12-Dependent Introns in *Arabidopsis thaliana* and Rice Genomes
Lin C. F. [1], Makalowski W. [1]. (1) The Pennsylvania State University
U2-dependent and U12-dependent introns coexist in many higher eukaryotic genomes. The presence of U12-dependent introns both in animals and plants suggests that their origin might predate the separation of animals and plants. Genomic studies revealed a biased phylogenetic distribution and frequency of U12-dependent introns within a variety of genomes. While scarce in the human and the *Arabidopsis* genomes, U12-dependent introns seem to be absent from the nematode genome. The estimated frequency of U12-dependent introns in the *Arabidopsis* genome is only half of that in the human genome. To

understand evolution of U12-dependent introns in plants, we scanned another plant (rice) genome and analyzed homology groups involving genes that contain U12-dependent introns. The size of groups varied widely. Only few groups contain U12-dependent genes from both *Arabidopsis* and rice. This finding suggests that U12-dependent introns are not very stable and easily transform into U2-dependent introns.

LU, Guoqing

Mining Mitochondrial Single Nucleotide Polymorphisms (SNPs) associated with human population evolution and genetics diseases

Wang C. G. [1], **Lu G.** [1]. (1) University of Nebraska Lincoln

The studies on mitochondrial genetic diseases and mitochondrial DNA (mtDNA) intraspecies diversity are key topics in population genetics and medicine. Most mtDNA variations within and among populations are single base variants, known as single nucleotide polymorphisms (SNPs). SNPs as an abundant form of mitochondrial genome variation, however, have not been systematically studied in the field of human molecular evolution and genetic diseases. The goal of this research is to use mitochondrial genome as a model to study molecular evolution and disease-associated SNPs in humans. For this purpose, a bioinformatics tool consolidating mt SNP information in various public repositories and literature is developed. We will present here the preliminary findings of mitochondrial SNPs potentially associated with human population evolution and genetic diseases.

LU, Yi

The Error Minimization of Escherichia coli Protein-coding Genes

Lu Y. [1], Freeland S. J. [1]. (1) University of Maryland Baltimore County

Because the standard genetic code is not completely symmetrical, even codons that specify the same amino acid can vary in terms of the consequences of mistranslation. Mistranslation is an unavoidable "tax" on the efficiency of gene expression (and thus, ultimately, the fitness of a genome). I therefore hypothesize that natural selection might exert a preference for codons that, even when mistranslated, produce an amino acid with similar biochemical properties to those of the "intended" amino acid, maximizing the efficiency of gene expression.

To test this hypothesis, I have created a local database of Escherichia coli proteins that each share less than 50% sequence identity to the others. I have also written software that analyzes the codon preferences of the gene sequences corresponding to these proteins in terms of secondary structure. If successful, this software will demonstrate a qualitatively new link between patterns of synonymous codon usage and the protein structures that they encode.

LUCAS, Olivier

Evolution of a protein-protein interaction in vertebrates

Lucas O. J. [1], Prigge J. R. [1], Schmidt E. E. [1]. (1) Montana State University

The core of the TATA Binding Protein (TBP) is conserved from archea to humans. Nevertheless, a new N-terminus of TBP (Nt-TBP) arose with vertebrates and is conserved among all vertebrate lineages. Most of the mice lacking the vertebrate-specific Nt-TBP die at mid-gestation from a placental defect. Since Nt-TBP arose with the earliest vertebrates, we suspect it has a more primordial function. We are using the yeast two-hybrid method to investigate the evolution of protein-protein interactions involving the Nt-TBP. Initial screens will use amphioxus and hagfish. Data will be compared to Nt-TBP interactions being found in the lab using mouse yeast two-hybrid screens. We built a cDNA library from hagfish. We cloned hagfish full length TBP and Nt-TBP and are using these as "baits". From our first screen, we sequenced 13 putative interacting clones. Five of them represent cDNAs of the same gene. Further confirmation, identification and characterization are underway.

LUO, Jing

Evolution of Hox gene clusters in young polyploidy fishes

Luo J. [1], Meyer A. [1]. (1) University of Konstanz, Germany

HOX gene clusters are both functionally and evolutionarily conserved. The comparison of the number of HOX gene clusters across animal lineages provided the first evidence for genome duplications in their evolutionary history. An evolutionary investigation of HOX clusters following relatively recent genome duplication, such as newly formed polyploid species, should yield insight into the evolution of the gene clusters and the long-term evolutionary consequences of genome duplication. Here we report a study that compares the evolution of HOX clusters between recently and more anciently duplicated genomes. We constructed a BAC library from goldfish (*Carassius auratus*), a putative young allopolyploid species, and screened for HOX clusters. Goldfish HOX clusters will be sequenced and compared to those of other fish species.

MACK, Jennifer

Centromeric Drive in Mixed Karyotypes of P. Maniculatus and P. Polionotus F1, N1 AND F2 Hybrids

Mack J. A. [1], O'Neill R. J. [1]. (1) University of Connecticut

Delineation of the centromeric structure has proven difficult due to species specific variation, while its function remains highly conserved. This complexity could be due to genetic conflict with centromeres acting as selfish elements to ensure that a particular centromeric structure will be paramount in a population. The model system challenging this hypothesis is the mouse species Peromyscus maniculatus and Peromyscus polionotus. These species share a conserved diploid number of 48 although a difference in the number of chromosome arms, or fundamental number (FN), was observed. These species are able to interbreed making this system invaluable for studying chromosomal effects. In F1 hybrids, equal segregation of the parental chromosomes exists, however a significant shift from a lower FN to a higher is observed in backcrosses and F2s, indicative of centromeric drive. Centromeric sequences and their associated proteins are currently being evaluated to determine the nature of this effect.

MADSEN, Ole

Timing and pattern of the mammalian colonization of Madagascar as assessed by nuclear gene phylogenies

Poux C. [1], **Madsen O.** [1], Marquard L. [2], Vieites D. R. [2], de Jong W. W. [1], Vences M. [2]. (1) University of Nijmegen, (2) University of Amsterdam

Four orders of extant terrestrial placental mammals (i.e., afrosoricidans, rodents, primates and carnivores) can be found in Madagascar. These four groups of Malagasy mammals all display considerable evolutionary diversity in morphology and ecological adaptations, and for each group several hypotheses have been proposed for their dispersal to Madagascar. To elucidate their biogeographic history, we determined sequences from three functionally diverse nuclear genes (Androgen receptor, von Willebrand Factor and Alpha 2B adrenergic receptor) from a broad range of Malagasy and putative outgroup mammals. Maximum likelihood and Bayesian phylogenetic analyses strongly support the monophyly of each group, and their closest living relative was identified among African taxa, suggesting single dispersal from Africa to Madagascar for each group. Estimation of divergence times of these dispersals, with Bayesian methodology, will be presented. The results will be discussed in relation to the number of possible colonization events and to the hypothesis of a putative African-Malagasy landbridge in the Cenozoic.

MAKAROVA, Kira S.

Reconstruction of duplication events in early evolution of eukaryotes

Makarova K. S. [1], Wolf Y. I. [1], Mirkin B. G. [2], Koonin E. V. [1]. (1) National Center for Biotechnology Information, (2) School of Information Systems and Computer Science

In this work we reconstructed using a parsimony approach and analyzed ancient duplication events that

occurred along the branch separating eukaryotes from prokaryotes. We showed that 3836 orthologous gene sets at the end of this branch correspond to 2073 orthologous sets at its base (duplication ratio of 1.85). Similar reconstructions show the duplication ratio of 1.19 for archaea and 1.25 for bacteria, which emphasizes the burst of gene duplication upon the advent of the eukaryotes. Among the functional classes of eukaryotic proteins, the only one that showed a significant excess of large paralogous clusters over the mean is "Posttranslational modification, protein turnover, chaperones". Almost all genes in this functional class underwent multiple duplications during early eukaryotic evolution, which seems to reflect the increasing requirement for fine tuning of protein targeting, regulation of protein degradation, and formation of multisubunit complexes in the eukaryotic cell. Many of the large paralogous clusters consist of proteins containing repetitive domains involved in protein-protein interactions, such as WD40 and TPR repeats. It appears that proliferation of these "glue" proteins had been a decisive force in the emergence of the complex organization of the eukaryotic cell.

MAKOVA, Kateryna

Insertions and Deletions Are Male Biased Too: A Whole-Genome Analysis in Rodents

Makova K. D. [1], Yang S. [2], Chiaromonte F. [3]. (1) Department of Biology, The Pennsylvania State University, (2) Department of Biochemistry and Molecular Biology, , (3) Department of Statistics, The Pennsylvania State University

In mammals, due to the greater number of cell divisions in the male germline than in the female germline, nucleotide substitutions occur more frequently in males. The data on mutation bias in insertions and deletions (indels) are contradictory, with some studies suggesting no sex bias and others indicating either female or male bias. The sequenced rat and mouse genomes provide a unique opportunity to accurately estimate mutation rates from a large number of orthologous loci in organisms similar in generation time and in the number of germline cell divisions. Here we compare the mutation rates between chromosome X and autosomes for neutral sites in eutherian ancestral interspersed repetitive elements present at orthologous locations in the rat and mouse genomes. We find that small indels are male biased: the male-to-female mutation rate ratio (alpha) for indels in rodents is ~2. This is consistent with nucleotide substitutions and small indels occurring primarily during DNA replication.

MANNAERT, An

Posterior Hox gene variation in amphibians

Mannaert A. [1], Leyns L. [1], Bossuyt F. [1]. (1) Vrije Universiteit Brussel

Changes in the number and regulation of Hox genes have strongly influenced the diversification of metazoan body plans. Tetrapods typically have four Hox gene clusters, each containing up to 13 paralogous genes. The loss of some of these genes in several vertebrate lineages may have been accompanied by the secondary loss or simplification of appendages. In amphibians, each of the three orders has a very specific, conserved body plan, with different types and/or reduction of limbs (frogs and salamanders) or no limbs at all (caecilians). We conducted a PCR survey in several members of each amphibian order to screen for the presence of conserved homeobox sequences in paralogs 9-13. We subsequently performed genome walking to obtain adjacent sequences outside these partial homeobox fragments. Identification was completed by phylogenetic analysis of a data set that combined our fragments with other vertebrate Hox sequences from GenBank.

MAO, Xianyun

Identifying loci under positive natural selection indicated by high branch link in population comparison

Mao X. [1], Bigham A. [2], Moore L. G. [3], Shriver M. D. [2]. (1) IGDP in genetics, The Pennsylvania State, (2) Department of Anthropology, The Pennsylvania State University, (3) Department of Anthropology, UC Denver

The relationship between high Fst and natural selection was investigated based on several human SNP

databases. Local natural selection especially positive selection may increase linked SNP allele frequency differences measured as high Fst. Although demographic factors as well as genetic drift also affected differences between populations, identifying high locus-specific branch length and Fst on genome-wide SNP datasets helped with the selection of candidate genes or loci based on particular population-specific traits. For instance, comparison between indigenous American population adaptive to high altitude (Quechua) and another indigenous American population (Nahua) with these statistics resulted in a list of 451 SNPs (threshold Fst>0.20) and 636 genes within 40kb region of those markers were selected. Based on this dataset, candidate genes that caused the adaption on high altitude were partially revealed.

MARTINS, Leonardo
Distribution of Rate and Rate Variability in Yeast Genomes
Martins L. O. [1], Thorne J. L. [2], Kishino H. [1]. (1) University of Tokyo, (2) North Carolina State University
Level of evolutionary rate and its variability reflect variable selection pressure or functional constraints and change in generation time or mutation rate. By introducing a stochastic model for rate variability, Bayesian hierarchical procedures robustly estimate rates and times. Based on the MCMC sample, we propose an empirical Bayes procedure to estimate their distribution in genome, taking account of uncertainty in the Bayesian estimates. The procedure was applied to 106 orthologous genes from seven *Saccharomyces* species and one *Candida albicans* as outgroup. The mean and s.d. of the evolutionary rate were estimated to be 0.00691 [0.00661, 0.00723] and 0.00139 [0.00112, 0.00168] (/site/MYR) respectively. The squared CV of rates per MYR had mean and s.d. of 0.00322 [0.00278, 0.00408] and 0.00106 [0.00039, 0.00181]. It will become possible to estimate the correlation between the rate and rate variability, when a larger number of genomes becomes available for analysis in the near future.

MARZELLI, Meghan
The Role of Segmental Duplications in Genome Evolution in Marsupials
Marzelli M. E. [1], Ferreri G. C. C. [1], O'Neill R. J. [1]. (1) University of Connecticut
Segmental duplications, paralogous sequences with high identity found as tandem arrays or interspersed in a nonrandom fashion, are associated with gene rearrangement and disease susceptibility. Segmental duplications are correlated with breaks of synteny between the mouse and human genomes, suggesting a role in gene rearrangement and evolution in eutherian mammals. Our work focuses on the identification of a segmental duplication in the marsupial *Macropus eugenii*. A *M. eugenii* BAC library was screened with a Macropodine centromeric sequence, KERV-1. Two positive clones were mapped by fluoresence *in situ* hybridization to *M. eugenii* chromosomes. One clone hybridized at the junction of two syntenic blocks (two separate chromosomes in the marsupial *Aepyprymnus rufescens*). Flanking sequences of the BACs will be subcloned to identify the extent of paralogy. Such segmental duplications may play an important role in genome evolution in marsupials.

MATHEWS, Sarah
Photoreceptor evolution in plants: patterns of phytochrome divergence in green and non-green species
Mathews S. [1]. (1) Harvard University
Light is a critical resource for plants and at least three families of photoreceptors mediate photomorphogenesis, the control of plant form by ambient light conditions. Phytochromes are the receptors of red and far-red light, encoded in eukaryotes by a small family of genes that has arisen by serial duplication and functional divergence. Selective constraints on phytochromes have been altered during evolutionary transitions and/or after gene duplication. Early in the history of angiosperms, episodic positive selection influenced phytochrome A, suggesting that its function might have been important as angiosperms colonized the existing understory. In parasitic plants, where some aspects of development are uncoupled from light signals, phytochromes are evolving under relaxed constraints and harbor novel introns. Within *Arabidopsis*, both members of a closely related and partially redundant gene pair are evolving under purifying selection, but with some sites in the locus of minor function evolving

neutrally or under weak selection. While these sites are in regions thought to be determinants of differential function between the loci, there is evidence that function of the minor locus is not important in all ecotypes or genetic backgrounds. Together these observations suggest specific roles for innovation in photoreceptor molecules during diversification events.

MAYER, Jens

Characterization of human endogenous retrovirus families HERV-K11, HERV-K14 and HERV-K22

Lavie L. [1], Flockerzi A. [1], Medstrand P. [2], Meese E. U. [1], **Mayer J.** [1]. (1) Human Genetics, University of Saarland, Germany, (2) Cell & Molecular Biology, Lund Unversity, Sweden

A substantial amount of the human genome comprises human endogenous retroviruses (HERVs). Most HERV families await detailed characterization regarding status in the human genome, evolutionary age, evolution, etc. Those information are crucial to assess evolutionary behaviour of endogenous retroviruses after germ line fixation, as well as the various HERV families' influence on the host genome. We set out to characterize other thus far little described HERV-K families, namely HERV-K11, HERV-K14 and HERV-K22, also termed HERV-K(HML-8), HERV-K(HML-1) and HERV-K(HML-5), respectively. We characterized corresponding proviral sequences in the human genome sequence regarding structure, variants, evolutionary age and evolutionary behaviour. We generated translatable consensus sequences that supposedly represent primate-targeting exogenous retroviruses having been present at that time in evolution. Details for each family will be presented.

Supported by DFG research grants Ma2298/2-1 and Me917/16-1.

MCGRATH, Casey

Molecular appropraches to studying genome evolution in ciliates

McGrath C. L. [1], Zufall R. A. [1], Katz L. A. [1]. (1) Smith College

Ciliates are characterized by dimorphic nuclei: a germline micronucleus and a somatic macronucleus. This feature makes the study of gene and genome evolution in ciliates particularly interesting and particularly difficult. Some ciliates are known as extensive fragmenters: in creating a new macronucleus, the micronucleus is fragmented into gene-sized chromosomes, which are then differentially edited and replicated. Extensive fragmentation can be found in ciliates from diverse lineages, implying that this process may have evolved more than once. We are developing techniques for sequencing genes from both the micronucleus and macronucleus of extensive fragmenters. Analyses of resulting data will provide insights into the processes underlying gene and genome evolution among ciliates.

MCGRAW, Lisa

Genes regulated by mating, sperm or seminal proteins in mated female Drosophila melanogaster

McGraw L. A. [1], Gibson G. [2], Clark A. G. [1], Wolfner M. F. [1]. (1) Cornell University, (2) North Carolina State University

In *Drosophila melanogaster*, the act of mating initiates important biological processes in the female required for fertilization and for efficient production of progeny. Many of the physiological and behavioral changes that occur post-mating in females are triggered by male-derived molecules or cells that pass to the female from the male during the course of mating. Despite our growing understanding of the functions and roles of the male-derived components of mating, the genetic changes that occur in the female in response to these substances is virtually unknown. We used Affymetrix GeneChips to identify genes within the female whose transcript levels change in response to mating. Among the ~1750 genes that show mating-dependent changes, we identified 550 genes that are regulated by receipt of sperm, 160 genes that are regulated by accessory gland proteins (or Acps- a major component of the seminal fluid) and the remainder of which are regulated by the act of mating itself. The mating-dependent genes that we have identified contribute to many biological processes including protection, protein modification and metabolism.

MCNEAL, Joel

Mitochondrial DNA suggests 12 origins of parasitism in angiosperms and implicates parasitic plants as vectors of horizontal gene transfer.

Barkman T. J. [1], **McNeal J. R.** [2], Lim S. [1], Coat G. [2], Croom H. B. [3], Young N. D. [4], dePamphilis C. W. [2]. (1) Western Michigan University, (2) The Pennsylvania State University, (3) University of the South, (4) Duquesne University

Analyses designed to infer the phylogenetic positions of all haustorial parasitic angiosperm lineages were performed on a data matrix of 103 seed plant species using 3 mitochondrial genes: *atp1*, *coxI*, and *matR* (4,019 aligned base pairs). Overall, the mitochondrial DNA phylogenetic tree agrees with independent analyses in terms of non-parasitic plant relationships and reveals 12 independent origins of parasitism in flowering plants. In addition, results suggest multiple putative transfers of *atp1* from host to parasite lineages, revealing for the first time a vector for horizontal gene transfer in angiosperms. Finally, the phylogenetic relationships inferred for parasites indicate that the origins of parasitism in angiosperms are very strongly correlated with gains of the mobile *coxI* group I intron. Evolutionary implications of the phylogeny and horizontal gene transfer events will be discussed.

MEISEL, Richard

Conserved motifs are found at *D. pseudoobscura* rearrangement breakpoints.

Meisel R. P. [1], Richards S. [2], Gibbs R. A. [2], Bettencourt B. R. [3], Gelbart W. M. [3], Hradecky P. [3], Letovsky S. [3], Schaeffer S. W. [4]. (1) IGDP Genetics, The Pennsylvania State University, (2) HGSC, Baylor College of Medicine, Houston, TX, (3) FlyBase, Harvard University, Cambridge, MA, (4) Department of Biology, Penn State University

The third chromosome of *D. pseudoobscura* is polymorphic for >30 different gene arrangements generated by overlapping paracentric inversions. Comparative bioinformatic analyses of the *D. melanogaster* and *D. pseudoobscura* genomes and PCR approaches allowed us to identify inter- and intraspecific rearrangement breakpoints. The intraspecific breakpoint sequences in derived and ancestral arrangements revealed a conserved motif found at high frequencies in interspecific breakpoint regions and virtually absent in coding regions. The motif is found in an indirect orientation in pairs of intraspecific breakpoints suggesting that genetic exchange between pairs of the motif could generate paracentric inversions. If this element is responsible for genome rearrangement, purifying selection appears to remove this sequence from coding regions. Interbreakpoint matches to the conserved motif average 148 bp suggesting that selection favors accelerated divergence and degradation of the motif sequence.

MELO-FERREIRA, José

Fine-scale analysis of ancient mtDNA introgression of *Lepus timidus* into *L. granatensis* and *L. europaeus* from the Iberian Peninsula

Melo-Ferreira J. F. [1], Ferrand N. [2], Alves P. C. [2]. (1) CIBIO/UP, Sciences Faculty of Porto & IREC/UCLM, (2) CIBIO/UP & Sciences Faculty of Porto

Hybridization and introgression may have important consequences in the assessment of phylogenetic relationships and phylogeographic patterns among hare species. Recently, ancient mtDNA introgression of *L. timidus* was detected in *L. granatensis* and *L. europaeus* from the Iberian Peninsula, an area where *L. timidus* is not present since the last glacial period. In this work we developed a PCR-RFLP that discriminates the mtDNA of the hares from the Iberian Peninsula, as well as *L. timidus*, and sequenced the cytochrome b of introgressed individuals. High levels of introgression (90 to 100%) were detected in Northern Iberia, in both *L. granatensis* and *L. europaeus*, and a clear southward gradient was found across the *L. granatensis* distribution range. MtDNA from *L. timidus* was also identified in a third Iberian hare species, *L. castroviejoi*. The observation of two introgressed mtDNA sub-lineages from *L. timidus* showing a marked geographical structure suggests two separate waves of hybridization during the last glaciations.

MENA-ALI, Jorge

S*-allele diversity and sequence divergence in natural populations of *Solanum carolinense

Mena-Ali J. I. [1], Travers S. [1], Stephenson A. G. [1]. (1) The Pennsylvania State University

We obtained and sequenced partial cDNAs for the S-alleles of 28 plants from two natural populations of horsenettle (Cumberland, MD and State College, PA). We found 16 S-alleles: 11 in the Cumberland MD population and only 5 in the State College PA population. There were five new alleles—3 in the MD population and 2 in the PA population. The other 3 alleles in the PA population were also shared with the MD population. Unlike other studies looking at the diversity of S-alleles, the allele frequencies in each of our two populations departed significantly from equal. Eleven of the 30 S-alleles sampled from the MD population and 13 of the 26 S-alleles sampled from the PA population were identified as the S8 allele (previously found in both NC and TN populations). Such deviation from equal S-allele frequencies may reflect biological differences associated with specific S-alleles. We compared the similarity of individual sequences for different copies of the S8 allele to determine the level of intra-allelic sequence variation and to identify variable regions. We expect a higher level of sequence divergence in allelic copies from different populations relative to the variation from within the same population.

MERRITT, Thomas

The structure and population genetics of the breakpoints associated with cosmopolitan inversion *In(3R)Payne* in *Drosophila melanogaster*

Merritt T. J. [1], Matzkin L. M. [2], Zhu C. T. [1], Eanes W. F. [1]. (1) Stony Brook University, (2) University of Arizona

We have identified both the distal and proximal breakpoints of the cosmopolitan chromosomal inversion *In(3R)P* in *Drosophila melanogaster*. The breakpoints are complex with small duplicated segments of the genome present at both breaks. The distal breakpoint breaks the tolkin (*tok*) gene, which is duplicated at the proximal breakpoint. We sequenced a representative sample of standard and inverted karyotypes for a 2kb portion of the functional tok gene, as well as the truncated copy of *tok* found at the distal breakpoint. The copies of tok found in the standard arrangements possess levels of polymorphism typical of *D. melanogaster* genes. The *In(3R)P* truncated copy of *tok* shows numerous single base changes as well as small deletions and duplications. The functional *to*k gene associated with *In(3R)P* shows very little polymorphism in either introns or exons. The pattern of polymorphism is consistent with a recent origin of *In(3R)P*, on the order of *Ne* generations.

MEUNIER, Julien

Recombination Drives GC-content in the Human Genome

Meunier J. [1], Duret L. [2]. (1) University Lyon 1, (2) CNRS

Unraveling the evolutionary forces responsible for variations of neutral substitution patterns among taxa or along genomes is a major issue to allow the identification of functional sequence features. Mammalian genomes show large scale regional variations of GC-content (the isochores), but the substitution processes at the origin of this structure are poorly
understood. Here, we analyzed the pattern of neutral substitutions in 14.3 Mb of primate non-coding regions. We show that the GC-content toward which sequences are evolving is strongly correlated (r2 = 0.61) with the rate of crossovers (notably in females). This demonstrates that
recombination drives the evolution of base composition in human (probably via the process of biased gene conversion). The present substitution patterns are very different from what they had been in the past, resulting in a major modification of the isochore structure of our genome. This non-equilibrium situation suggests that changes of recombination rates occur relatively frequently during evolution, possibly as a consequence of karyotype rearrangements. These results have important implications for understanding the spatial and temporal variations of substitution processes in a broad range of sexual organisms, and for detecting the hallmarks of natural selection in DNA sequences.

MEYER, Axel

Evidence for the fish-specific genome duplication

Meyer A. [1], Van de Peer Y. [2], Taylor J. [3]. (1) University of Konstanz, (2) University of Ghent, (3) Victoria University

An evolutionary analysis of both structure and regulatory elements of vertebrate Hox genes and clusters, with emphasis on the differences between the Hox clusters of fish (Actinopterygia) and tetrapod (Sarcopterygia) lineages will be presented. We find higher rates of gene loss and gene sequence evolution in the Hox genes of fishes compared to those of land vertebrates. In contrast to the general conservation of genomic architecture and Hox gene sequences observed in sarcopterygians, the evolutionary history of actinopterygian Hox clusters likely includes an additional (third) fish-specific genome duplication that initially increased the number of clusters from four to eight. Comparative genomic analyses of gene families and of the entire Fugu genome (compared to other genomes) with emphasis on the analysis of duplicated protein coding genes (paranome) will be presented. Both types of data sets provide evidence for a fish-specific genome duplication about 350 million years ago. The phylogenetic timing and evolutionary implications of the fish-specific genome duplication in fish will be discussed.

MEYER, Christiane I.

Different genes cause similar melanophore patterns in two *Xiphophorus* species

Meyer C. I. [1], Stern D. L. [1]. (1) Princeton University

In the fish genus *Xiphophorus*, at least 10 out of 22 species are polymorphic for various macromelanophore patterns. The polymorphism is due to simple Mendelian segregation of a diallelic locus and the pattern phenotype is encoded by the dominant allele. In *X. maculatus*, the macromelanophore determining locus (*Mdl*) has been mapped to the sex chromosome. We tested whether the gene underlying a similar polymorphism in the sister species, *X. helleri*, maps to the same linkage group. The *X. helleri* locus, known as *Dabbed*, is not linked to the marker *egfrb* (epidermal growth factor receptor b), which is present in all *Xiphophorus* species and closely linked to *Mdl* in *X. maculatus*. This indicates that genes at two different loci are responsible for very similar color patterns in these two closely related fish. This result contrasts with the scenario recently described for artic skuas and snow geese, in which independent mutations have occurred in a single gene to produce convergent melanin-based color patterns in these two distantly related species.

MILLER, Hilary

The Major Histocompatibility Complex (MHC) of an Ancient Reptile Lineage, *Sphenodon* (Tuatara)

Miller H. C. [1], Belov K. [2], Daugherty C. H. [1]. (1) Victoria University of Wellington, (2) Australian Museum

The evolution and organisation of MHC genes differs among diverse vertebrate taxa. For instance, MHC class II gene families are conserved among different eutherian lineages, whereas avian MHC genes appear to undergo higher rates of gene duplication and/or concerted evolution making orthologous relationships difficult to discern. Both mammals and birds are derived from ancient reptiles, so analysis of the reptilian MHC may hold the key to understanding the differences in MHC evolution between birds and mammals. We are using genomic PCR and cDNA library screening to isolate class I and II MHC genes from the tuatara, the sole survivor of an ancient reptile lineage, *Sphenodontia*, which diverged from other reptiles about 230 million years ago. Preliminary results show that a tuatara class I gene has highest amino acid similarity (44-52%) to class I sequences of snake, axolotl and shark, but only 28-40 % similarity to mammalian and bird sequences, reflecting the antiquity of the *Sphenodon* lineage. This locus appears to retain moderate levels of MHC diversity, even in populations with low microsatellite DNA variation.

MISHRA, Paras Kuamr

Study of Causes of Hybrid Sterility in the *Drosophila Bipectinata* Complex and Its Impact on Testis and Seminal Vesicle Morphology

Mishra P. K. [1], Singh B. N. [2]. (1) Genetics Lab., Dept. of Zoology,Varanasi-221005, (2) Genetics lab.Dept.Of

Zoology, BHU, Varanasi-221005
Hybrid sterility is one of the most important isolating mechanisms that plays a crucial role in speciation by acting as a reproductive barrier. The *Drosophila bipectinata* complex comprises of four species namely *bipectinata, parabipectinata, malerkotliana* and *pseudoananassae*. Their male hybrids were sterile. In attempt to investigate the putative causes of hybrid sterility, we have analyzed the sperm motility and the presence of individualized sperm in hybrids. Individualized sperm were altogether absent in hybrids but immotile sperm were present in all hybrids except those of*pseudoananassae* where testes were aspermic. The above observations suggest that there were appreciable compatibilities among *bipectinata, parabipectinata and malerkotliana* but maximum incompatibilities of these species with *pseudoananassae*. Furhter,there was reduction in the size of seminal vesicles in all hybrids but their testis size remained larg except in the hybrids of *pseudoananassae* where the testis was atrophied. These results infer that *bipectinata, parabipectinata* and *malerkotliana* are phylogenetically closer while *pseudoananassae* is remotely related with these species.

MONTANO, Adriana Maria

Convergent and Parallel Evolution of Peptidoglycan Recognition Proteins in Mammalian Innate Immune System

Montano A. M. [1], Satta Y. [1], Takahata N. [1]. (1) Biosystems Science, The Graduate University for Advanced Studies (Sokendai)

Insects and mammals have developed similar mechanisms and molecular pathways to recognize and eliminate invading pathogens. Components of the innate immune system discriminate between hosts and microorganisms by recognizing repeating patterns of sugar residues found in pathogens. Peptidoglycan recognition proteins (PGRPs) are a family of pattern recognition molecules that bind to peptidoglycan, which is an essential cell wall component of virtually all bacteria. In humans there are four *PGRP* genes: *PGRP-S*, *PGRP-L*, *PGRP-Ia* and *PGRP-Ib*, in contrast to 13 homologs in Drosophila. Up to now there is no clear phylogenetic relationship of *PGRP* genes, partly due to their great extent of divergences in amino acid sequences. The purpose of this study is to clarify the origin and evolutionary relationship of *PGRPs* in mammals. Although four types of *PGRPs* have likely emerged prior to the mammalian divergence, subsequent evolution is lineage specific. Also, *in-silico* analysis of genomic and EST databases of various vertebrates suggests parallel and convergent evolution in *PGRP* genes at the amino acid sequence level. These results suggest that natural selection plays important roles in the diversification of *PGRP* genes.

MONTOOTH, Kristi

Modeling metabolic performance in Drosophila

Montooth K. L. [1], Clark A. G. [1]. (1) Cornell University

Several genes flanking the glucose-6-phosphate branchpoint in glycolysis evolve non-neutrally in Drosophila. We found that a QTL underlying the covariance between three enzymes at this branchpoint is also associated with metabolic rate and flight velocity. To understand how biochemical variation at this branchpoint impacts performance, we are modeling flight as a function of maximal activities for enzymes flanking glucose-6-phosphate. We are fitting a model based on flux through a trifurcating branch and estimating parameters using MCMC sampling. We are using a similar approach to model ethanol detoxification as a function of ADH, ALDH and AcCoAS activities. Flight performance and the detoxification of environmental toxins, likely impact fitness in natural populations. Understanding how variation within biochemical pathways affects physiological performance is a key step in understanding how natural selection, acting on performance traits, manifests itself within metabolic pathways.

MOORE, Jonathan

The Number and Distribution of Unrooted Binary Tree Shapes

Moore J. E. [1], Lake J. A. [2]. (1) Pomona College, (2) UCLA

Though the relationship between the number of taxa and the number of binary (fully resolved) unrooted

trees has been described, the exact relationship between the number of taxa and the number of binary unrooted tree *shapes* remains unknown. This relationship is both of theoretical and practical interest, since the distribution of tree shapes is also important for the calculations of prior probabilities for those engaged in Bayesian inference of phylogeny. For *n* taxa, we exhaustively calculated three invariants for trees that (a) yield the same results for trees of the same tree shape and (b) are highly likely to yield different results for trees of different tree shapes. Calculation of these invariants allows efficient grouping of trees into tree shapes, resulting in highly accurate estimates of the number of tree shapes for up to 24 taxa. Trends in these estimates are analyzed, as are the relationships between a tree shape's stringiness and the number of trees of that shape.

MOORE, Richard
The signature of selective sweeps in *Arabidopsis thaliana*
Moore R. C. [1], **Purugganan M. D.** [1]. (1) North Carolina State University
Genes having undergone a selective sweep exhibit greatly reduced nucleotide variation that increases as the recombination distance from the site under selection increases. Although a number of genes in the model plant *Arabidopsis thaliana* exhibit patterns of nucleotide variation indicative of a sweep, the extent to which this signature extends into the flanking chromosomal region around the locus is uncharacterized. In order to address this issue, we performed a molecular population genetics analysis of nucleotide polymorphism and divergence in chromosomal regions surrounding loci which have undergone a selective sweep in Arabidopsis. We predict that the size of selective sweeps in Arabidopsis will be influenced by the low rates of recombination in this selfing species.

MORAN, Nancy
Evolution of gene expression in the reduced genome of a bacterial symbiont (*Buchnera*)
Moran N. A. [1], Plague G. R. [1], Wilcox J. L. [1], Dunbar H. E. [1]. (1) University of Arizona
The endosymbiotic bacterium *Buchnera* aphidicola is related to *Escherichia coli* but has a highly reduced genome, retaining 15% of ancestral genes. Buchnera retains genes for biosynthesis of limiting essential amino acids, which are provisioned to the aphid hosts. Almost all regulatory genes are eliminated. We addressed the extent to which Buchnera shows ancestral or novel means of regulating pathways for amino acid biosynthesis as well as heat shock genes. Our approach was to manipulate growth conditions (temperature or diet) for the aphid and to examine the resulting transcript pools, using full genome microarrays and quantitative RT-PCR. *Buchnera* of different host species differ in presence of *metR*, which regulates methionine biosynthesis; we used this contrast to determine the role of *metR* in regulation of transcription genes. *Buchnera* shows constitutively elevated production of heat shock genes. Although its responses to environmental change are modest, there are some significant responses to both changes in amino acid supply and to temperature stress. Some of these changes are the same as those in related bacteria, and some appear to be novel.

MORRIS, Geoffrey
Transcription factor binding divergence of duplicate genes
Morris G. P. [1], Li W. H. [1]. (1) Dept. of Ecology & Evolution, University of Chicago
Divergence in the binding of transcription factors to promoters is thought to underlie much of gene expression evolution. Using published genome-wide location data for 106 transcription factors in S. cerevisiae we investigated divergence in the sets of transcription factors bound to duplicate genes. If promoter regions are retained during duplication then divergence in the set of bound transcription factors should increase over time, concomitant with coding sequence divergence. Accordingly, we found that transcription factor binding divergence in duplicate genes is weakly correlated with per-site synonymous (Ks) and non-synonymous (Ka) divergence. However, even duplicate pairs with low Ks (<0.1) have considerable variation in transcription factor binding divergence. This may be due to incomplete duplication of promoters, or deletions or rearrangements which cause the abrupt loss of

binding sites. We therefore investigate whether duplicate pairs found within syntenic regions have less transcription factor binding divergence and expression divergence than genes that are found at the edges or outside of syntenic regions.

MOSES, Alan

Maximum likelihood estimation and hypothesis testing for phylogenetic analysis of transcription factor binding sites

Moses A. M. [1], Eisen M. B. [2]. (1) Biophysics and Integrative Genomics, UC Berkeley, (2) Division of Genome Sciences, Lawrence Berkeley Lab

The binding sites of sequence specific transcription factors are an important class of functional non-coding DNA. Using a probabilistic framework for their evolution we have implemented software ("EMnEM") that can be used for maximum likelihood estimation and hypothesis testing. We confirm previous findings (based on parsimony and other methods) that rates of evolution in transcription factor binding sites are slower than surrounding sequence, and that there is variation in the rate of evolution across the different positions in the binding sites. With the availability of powerful, inexpensive computers, maximum likelihood methods are becoming increasingly popular for phylogenetic analysis of protein coding sequence; this may prove true for non-coding sequence as well.

MOWER, Jeffrey

The Genus *Plantago* Provides a Unique Opportunity to Identify Horizontal Gene Transfer in Plant Mitochondria

Mower J. P. [1], Young G. [1], Stefanovic S. [1], Palmer J. D. [1]. (1) Indiana University

Horizontal transfer is the exchange of genetic material between non-mating species. In plant mitochondria, horizontal transfer has been previously observed for genes and mobile introns. Unfortunately, many cases of horizontal transfer in plant mitochondria may go undetected because of the extremely low mitochondrial substitution rate, often leading to phylogenetic trees with limited resolution and low support values. In the genus *Plantago*, however, mitochondrial genes have experienced a dramatic increase in the rate of synonymous substitution relative to other angiosperms, providing a context to easily distinguish between genes of *Plantago* and non-*Plantago* origin. Phylogenetic analysis of the *atp1* gene from members of the genus *Plantago* revealed several cases of horizontal gene transfer. In at least one case, the donor appears to be a parasitic plant. Analyses using the mitochondrial SSU rDNA and *cox1* genes do not show evidence for horizontal transfer.

MURPHY, William

A rhesus macaque genome map reveals recent rearrangements in human genomic evolution

Murphy W. J. [1], Agarwala R. [2], Schaffer A. [2], Stephens R. [3], Smith C. J. [4], Crumpler N. J. [1], David V. A. [5], O'Brien S. J. [5]. (1) SAIC-Frederick, Inc. LGD, NCI, (2) NCBI, National Institutes of Health, (3) ABCC, NCI-Frederick, (4) Mt. Sinai School of Medicine, (5) Laboratory of Genomic Diversity, NCI-Frederick

The genomes of non-human primates have become highly visible candidates for comparative analysis, as they provide powerful models of human disease and a better understanding of the recent evolution of the human genome. Here we compare the order of 815 genomic markers mapped in a rhesus macaque radiation hybrid panel to the human genome, allowing for nearly complete syntenic cross reference at an average resolution of 3.5 Mb. Analysis of ordered conserved segments identified a modest number of chromosomal rearrangements between human and macaque. Multispecies comparisons with cat, dog, cattle, mouse and rat genomes allowed us to phylogenetically polarize rearrangement events and distinguish those that occurred in the human branch from those that occurred earlier in primate evolution. Analysis of primate specific rearrangement breakpoints, refined using mouse and human genome sequences, provide insight into the mechanisms promoting chromosomal rearrangement in recent primate evolution.

NACHMAN, Michael

The genetic basis of adaptive melanism in pocket mice

Nachman M. W. [1]. (1) University of Arizona

Identifying the genes underlying adaptation is a major challenge in evolutionary biology. Here I describe the molecular changes underlying adaptive coat color variation in a natural population of rock pocket mice, *Chaetodipus intermedius*. Rock pocket mice are generally light-colored and live on light-colored rocks. However, populations of melanic mice are found on dark lava, and this concealing coloration provides protection from avian and mammalian predators. Association studies using markers in candidate pigmentation genes revealed four mutations in the melanocortin-1-receptor gene, Mc1r, that appear to be responsible for adaptive melanism in one population of lava-dwelling pocket mice. Linkage disequilibrium decays rapidly both upstream and downstream of Mc1r, suggesting that a genomic scan for selection would have missed this gene despite its clear importance in adaptive evolution. Interestingly, other melanic populations of these mice on different lava flows show no association with Mc1r mutations, indicating that adaptive dark color has evolved recently in this species through changes at different genes. Estimates of migration rates between light and dark populations indicate that selection on coat color is strong. Over longer evolutionary timescales, variation at Mc1r suggests that this gene may underlie color variation in a variety of different vertebrates.

NAM, Jongmin

Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms.

Nam J. [1], Kim J. [2], Lee S. [3], An G. [3], Ma H. [1], Nei M. [1]. (1) Pennsylvania State University, (2) MSU-DOE PRL, (3) POSTECH

Plant MADS-box genes form a large gene family for transcription factors and are involved in various aspects of developmental processes including flower development. To have a deeper insight into the evolutionary pattern of this gene family, we enumerated all available functional and nonfunctional (pseudogenes) MADS-box genes from the Arabidopsis and rice genomes. Plant MADS-box genes can be classified into type I and type II genes on the basis of phylogenetic analysis. Conducting extensive homology search and phylogenetic analysis, we found 64 functional and 37 nonfunctional type I genes and 43 functional and 4 nonfunctional type II genes in Arabidopsis. We also found 24 functional and 6 nonfunctional type I genes and 47 functional and 1 nonfunctional type II genes in rice. Our phylogenetic analysis indicated that there were at least ~4 - 8 type I genes and ~15 - 20 type II genes in the most recent common ancestor of Arabidopsis and rice. We have also shown that type I genes have experienced a higher rate of birth-and-death evolution than type II genes in angiosperms. The implications of these findings will be discussed.

NEAFSEY, Daniel

Different regulatory mechanisms underlie similar transposable element profiles in the *Drosophila* and pufferfish genomes

Neafsey D. E. [1], Hartl D. L. [1]. (1) Harvard University

Comparative analysis of recently sequenced eukaryotic genomes has uncovered extensive variation in transposable element (TE) abundance, diversity, and distribution. The TE profile in the sequenced pufferfish genomes is more similar to that of *Drosophila melanogaster* than other sequenced vertebrates, in that pufferfish TEs exhibit low overall abundance, high family diversity, and spatial concentration in the heterochromatin. It has been suggested that selection against the deleterious effects of ectopic recombination between TEs has structured the TE profile in *Drosophila* and pufferfish, but not humans. We tested this hypothesis by measuring the population frequency of 48 euchromatic TE insertions in the genome of the green spotted pufferfish (*Tetraodon nigroviridis*). We estimated the selection coefficient acting on recent insertions using a sojourn time density function based on a diffusion approximation. We show that in contrast to *Drosophila*, euchromatic TE insertions in *Tetraodon* are selectively neutral, and that

the low copy number and compartmentalized distribution of TEs in the *Tetraodon* genome must be regulated by other means. Inference of regulatory processes governing TE profiles should take into account effective population size, incidence of inbreeding/outcrossing, and other species-specific traits.

NEGRE, Bárbara
Rearrangement of the *Drosophila* Hox gene complex: Expression of *labial*, *proboscipedia* and *abdominal-A* in species with different gene arrangements
Negre B. [1], Sánchez-Herrero E. [2], Suzanne M. [2], Akam M. [3], Ruiz A. [1]. (1) Universitat Autònoma de Barcelona, (2) Centro de Biología Molecular , (3) University of Cambridge
The *Drosophila* Hox gene complex has split at least three times in different lineages. One of the splits, fixed in the species of the *repleta* group, took place between *labial* (*lab*) and *proboscipedia* (*pb*), separating *lab* from the other anterior Hox genes and relocating it close to *abdominal-A* (*abd-A*). In order to test for an effect of the *lab-pb* split on their expression patterns, we have analyzed the expression of three Hox genes close to this split (*lab*, *pb* and *abd-A*) by *in situ* hybridization to embryos and imaginal discs in four *Drosophila* species: two of them (*D. buzzatii* and *D. repleta*) were fixed for the split whereas the other two (*D. melanogaster* and *D. virilis*) possess the ancestral arrangement. The results show that the expression of the three genes seem to be conserved, with no appreciable differences between the species with and without the split. The only observed change affects the expression of *pb* in the *D. virilis* embryo and seemingly is not related with the *lab-pb* split.

NEGRE, Bárbara
Rearrangement of the *Drosophila* Hox gene complex: Genomic characterization of the *labial-proboscipedia* split in *Drosophila buzzatii*
Negre B. [1], Casillas S. [2], Barbadilla A. [2], Nefedov M. [3], de Jong P. [3], Ruiz A. [2]. (1) Universitat Autònoma de Barcelona (Spain), (2) Universitat Autònoma de Barcelona. Spain., (3) Children's Hospital Oakland Research Institute.USA
In the genome of metazoan, Hox genes are usually clustered and arranged in the same order as they are expressed along the anteroposterior body axis. However, three different natural rearrangements of the Hox gene complex have been recorded in the genus *Drosophila*. We have characterized a split fixed in *D. buzzatii* that separated the gene *labial* (*lab*) from the other anterior genes of the complex and relocated this gene to a distant site near the posterior genes, *abdominal-A* (*abd-A*) and *Abdominal-B* (*Abd-B*). Two clones from a *D. buzzatii* BAC library, one containing the *lab* region and the other containing the *proboscipedia* (*pb*) region, were isolated and completely sequenced. The resulting sequences were analyzed and the arrangement of genes compared with that found in *D. melanogaster*. The split was caused by an inversion with one breakpoint located ~40 kb upstream of *lab* and <3 kb downstream of *pb*, between *pb* and the cluster of *Ccp* genes, and the other breakpoint located between the genes *CG31363* and *CG17836*.

NEKRUTENKO, Anton
Reconciling the Numbers: ESTs vs. Protein-Coding Genes
Nekrutenko A. [1]. (1) The Pennsylvania State University
The number of expressed sequences greatly surpasses the estimated number of protein-coding genes in mammalian genomes. An evolutionary approach reveals that only 9-14% of human and mouse expressed sequences are able to code for proteins. Clustering of these sequences using cross-species relationships suggests that millions of expressed sequences may correspond to only ~20,000 distinct protein-coding transcripts.

NIIMURA, Yoshihito
Evolutionary Changes of the Olfactory Receptor Gene Family in the Human and Mouse Lineages
Niimura Y. [1], Nei M. [1]. (1) The Pennsylvania State University
Olfactory receptor (OR) genes form the largest multigene family in mammalian genomes. Humans have

~800 OR genes, but >50% of them are pseudogenes. By contrast, mice have ~1,400 OR genes and pseudogenes are ~25%. To understand the evolutionary mechanisms that shaped the difference in the number of OR genes between humans and mice, we studied the genomic locations of all human and mouse OR genes and conducted a detailed phylogenetic analysis of the genes. We found that the organization of OR genomic clusters is well conserved between humans and mice in many chromosomal locations, suggesting that the difference in the number of OR genes have been generated mainly by gene duplications within each genomic cluster. We estimated that the most recent common ancestor (MRCA) between humans and mice had ~750 functional OR genes and mice acquired ~350 new OR genes after the human-mouse divergence. We also estimated that ~430 OR genes in the MRCA have become pseudogenes or eliminated from the genome in the human lineage. Therefore, both gene expansion in the mouse lineage and gene loss in the human lineage are responsible for the difference in the number of OR genes between the two species.

NIKOLAIDIS, Nikolas

Origins and evolution of immunoglobulin (Ig-like) domain containing receptors of the innate immune system

Nikolaidis N. [1], Nei M. [1]. (1) The Pennsylvania State University

The innate immune system in vertebrates uses several gene families encoding Ig-like receptors. The extended leukocyte receptor cluster (LRC) gene complex in mammals contains four such gene families (CD66, SIGLEC, LILR, KIR), some of which (e.g. the KIR gene family in primates) are known to have evolved very rapidly by gene duplication. To understand the origins and evolution of these gene families, we studied the evolutionary relationships of Ig-like domain encoding genes from human, mouse, chicken, frog and fish. Phylogenetic analysis of these genes showed that most LRC gene families have homologs in all species and generate intraspecific clades suggesting recent species-specific gene expansion. However, the analysis of the SIGLEC genes suggests that some member genes have diverged rather early in vertebrate evolution. Furthermore, many phylogenetic clades contain genes with divergent functions (e.g. KIR genes are closely related to the FCAR and GP6 genes). In addition, the KIR, LILR, and other LRC gene families from mammals are closely related to several genes from chicken, frog, and zebrafish. These results suggest that the gene families evolved following the model of divergent evolution and share a common ancestor that existed before the separation of fish and tetrapods.

NORTON, Heather

Pigmentation candiate gene variation and possible evidence for natural selection in human populations

Norton H. L. [1], Friedlaender J. S. [2], Merriwether D. A. [3], Koki G. [4], Mgone C. S. [4], Shriver M. D. [1]. (1) Penn State University, Department of Anthropology, (2) Temple University Anthropology Department, (3) SUNY Binghamton, Department of Anthropology, (4) Papua New Guinea Institute for Medical Research

Pigmentation of the skin is an easily visible example of human phenotypic variation. Natural selection may have shaped global variation in human skin pigmentation by acting on genes controlling the production and distribution of melanin, the primary pigment of the skin. Melanin acts as a natural sunscreen, making large amounts of melanin beneficial in environments where ultraviolet radiation is high. Recent studies suggest that SNPs in some pigmentation candidate genes may have an effect on normal variation in human pigmentation. This work examines locus-specific FST values (lsFST) for SNPs in 5-6 pigmentation candidate genes in pairwise comparisons of six human populations. These values are then compared to an empirical lsFST distribution based on 11,078 SNPs. High pairwise lsFST values between populations known to differ in skin pigmentation suggest that the SNP may have a functional effect. High pairwise lsFST values between populations with similar pigmentation may suggest convergent evolution for pigmentary phenotype. Where possible, levels of LD surrounding putative selected SNPs will be examined to help determine if high lsFST levels are due to drift or to natural selection.

NOZAWA, Masafumi

Origin of *sirene*: a paradigm of exon shuffling and gene duplication in *Drosophila*

Nozawa M. [1], Aotsuka T. [1], Tamura K. [1]. (1) Tokyo Metropolitan University

Exon shuffling and gene duplication are essential mechanisms for producing new genes, which result in diversity of genomes. In *Drosophila*, it has been reported that exon shuffling through retropositions of expressed genes produced several chimerical genes. In this study, we also discovered a new chimerical gene consisting of the head from a part of *CG11779* (annotated in the *D. melanogaster* genome) and the body from *Adh* in the species of the *D. bipectinata* species complex, and named it *sirene* after Greek myths. Sequence analyses of the genomic and complementary DNA of *sirene* showed that the exon shuffling was induced by reverse transcription of expressed *nos* gene, which is overlapping with *CG11779*, along with tandem duplication of *Adh*. Interestingly, the regulatory element of *sirene* was originated from the retroposed fragment of *CG11779*, suggesting possibility for retroposons to duplicate regulatory regions as well as coding regions.

NUNNEY, Leonard

Detecting natural selection at the molecular level: re-examining some "classic" examples.

Nunney L. [1]. (1) University of California, Riverside

The primary statistics used to detect the action of natural selection in gene sequence data is the ratio w = Ka/Ks, where Ka and Ks are the rates of nonsynonymous and synonymous substitution. Strong evidence for adaptive evolution is the finding of w > 1. However, the methodology used for detecting unusually high values of w and testing them against 1 is inconsistent, sometimes increasing the type 1 error rate by converting random extremes into apparently significant effects. To avoid this problem, we have proposed two related methods for detecting the action of natural selection, still focusing on the criterion w > 1 but also emphasizing the importance of detecting shifts in natural selection between related lineages. Applying these methods to published examples of adaptive evolution shows that while some, such as abalone sperm lysin, still showed strong evidence of adaptive evolution, others failed to show this expected pattern.

OHNIWA, Ryosuke

Molecular evolution of the SNARE families: Q-SNARE preceded R-SNARE.

Ohniwa R. L. [1], Uemura T. [1], Takeyasu K. [1]. (1) Graduate School of Biostudies, Kyoto University

The SNARE molecules strictly govern the intracellular vesicle transport in eukaryotes, which relies on a specific combination of 4 distinct SNARE motifs (Qa, Qb, Qc and R) to form a specific SNARE complex. In this report, we extended our homology search to the 16 Archaea and 89 Eubacteria genomes to identify a possible evolutionary process of the SNARE molecules. First the evolutionary and hydropathy profiles of the Qa, Qb, Qc and R SNARE motifs were collected from a complete set of the SNARE molecules in *Caenorhabditis*, *Arabidopsis*, *Mus*, *Saccharomyces* and *Drosophila*. Then, from the Mahalanobis distances between the profiles, the collected SNARE molecules were classified into 4 distinct families. On the basis of this classification, 1055 and 5480 genes were identified in Archaea and Eubacteria, respectively, and found to encode Q-SNAREs, but not R-SNARE. These results suggest that the Q-SNARE families are older than the R-SNARE family in evolution.

OKADA, Norihiro

Looking for the molecular basis of adaptive evolution of cichlid fishes

Okada N. [1]. (1) Tokyo Institute of Technology

The cichlid fishes in East Great Lakes exhibit great diversity of color patterns and jaw morphology. It is possible that mutations responsible for the changes of absorption spectra of visual pigments, and also those for color patterns and jaw morphology, might have been fixed due to selection during evolution. We are now trying to identify these mutations and characterize them.

Some deep-living species in the lakes changes aa sequences in RH1 genes to alter absorption spectra

toward short in order to adapt to their ddep light environment. Some species in Lake Victoria also change aa sequences of *lws* possibly to adapt to its turbid environment.

OKAMURA, Hideyuki
Kingdom specific subfamilies of the P-type ATPase evolved from methanogens and suggest the possibility of polyphyletic eukaryogenesis
Okamura H. O. [1], Denawa M. D. [2], Ohniwa R. O. [2], Kawai S. K. [1], Takeyasu K. T. [2]. (1) Department of Biology, Osaka Dental University, (2) Graduate School of Biostudies, Kyoto University
We performed PHI- and PSI-BLAST search for the P-type ATPase in 16 archaeal genomes, and found 41 P-type ATPases ORFs. Phylogenetic analyses showed that 3 of 41 ORFsÃ£â,¬â,¬belong to the type 2C subfamily (Na+/K+ and H+/K+ ATPase of animal) and 4 of 41ORFs to the type 3A subfamily (plasma membrane H+ ATPase of plant and fungi), respectively. Namely, 5 of 16 archaea have these eukaryote specific ion pumps and 4 of them are methanogens which branched
recently based on the 16s rRNA tree. In contrast, the deep-branching hyperthermophiles in the 16s rRNA tree had only type 1B (heavy metal transporter) or no ion pumps. Phylogenetic tree of type 2C and type 3A rejectedÃ£â,¬â,¬ monophyly of archaea. Especially, genus *Methanosarcina* was the sister group of eukaryote with a high bootstrap value in the type 2C tree. These results together with the preceding reports provide phylogenetic evidence that eukaryote evolved from methanogenes through polyphyletic eukaryogenesis.

O'NEILL, Rachel
Centromere Dynamics, Karyotypic Diversification and Speciation in Mammals
O'Neill R. J. [1]. (1) University of Connecticut
Species-specificity in karyotypes frequently involves centromere-associated rearrangements, including centric shifts, translocations, fusions, and inversions. Little is known about how the centromere, either as a functioning unit of chromatin or as a specific block of repetitive DNA sequences, acts in the creation of these types of chromosome rearrangements in an evolutionary context. Macropodine marsupials offer a unique system in which to study karyotypic diversification and speciation. Characterization of the composition and distribution of centromeric sequences within this group of mammals indicates these sequences have been involved in amplifications, segmental duplications, fissions and fusions. The centromere becomes unstable within interspecific hybrids and is associated with an increased frequency of chromosome aberrations, indicating that hybridization-induced genomic instability may play an important role in karyotypic diversification and speciation within this group.

ONG, Han Chuan
Repeated independent transfer of a mitochondrial gene to the nucleus in grasses
Ong H. C. [1], Brown C. W. [1], Palmer J. D. [1]. (1) Indiana University
Mitochondrial gene loss and transfer to the nucleus has been shown to be ongoing, widespread and frequent among angiosperms. Particularly, most ribosomal protein genes have experienced scattered loss via transfer to the nucleus within grass (Poaceae) mitochondria. By surveying the evolution of the mitochondrial *rpl5* gene in grasses, we observed up to four independent transfers with subsequent functional integration into the nuclear genome. For most grass taxa, mitochondrial *rpl5* is maintained as a pseudogene while the intact nuclear copy is expressed and targeted to the mitochondria by a 5' targeting peptide. Surprisingly, we also observed transitory states in rpl5 *gene transfer whereby some grasses continue to retain a copy of a full-coding* rpl5 gene in both the mitochondria and nucleus. Our results suggest that mitochondrial gene transfers to the nucleus in flowering plants occur at a very high rate with a prolonged intermediate state and are more prevalent than previously reported.

OPAZO, Juan C.
Evolution of insulin gene in caviomorph rodents

Opazo J. C. [1], Palma R. E. [2], Melo F. [3], Lessa E. P. [4]. (1) Center for Molecular Medicine & Genetics, WSU, (2) CASEB, depto. Ecologia, PUC, (3) Depto. Genética Molecular y Microbiología, PUC, (4) Lab. Evolución, Fac. Ciencias, Montevideo, Uruguay

Among mammals insulin is a conservative molecule that maintains its structure and function. Rodents that belong to the Suborder Hystricognathi represent an exception, having a very divergent molecule with unusual physiological properties. Accordingly, in this work we analyzed the evolutionary pattern of insulin gene in caviomorph rodents. We found that these rodents have higher omega (dN/dS) value than non hystricomorph mammals, and inside the group values are heterogeneous. We estimated some codons under positive selection, the second binding site (A13 and B17) and another related with hexamerization (B18, B20, B22 and B27). In the monomer structure all selected sites form a single patch around the second binding site, and in the hexamer patches are together forming three major ones. In the hexamer structure in contacts between chains B all selected sites are involved (except B18) and between faces in the center of the molecule all contacts are among selected sites. There is no clear hypothesis about this drastic change, however, experimental evidence show that this group of rodents has some peculiarities in growth function and coincidental or not these changes appeared together with important changes in life history traits.

OPIYO, Stephen

Usage of multivariate methods in protein family classification

Opiyo S. O. [1], Moriyama E. [1]. (1) University of Nebraska, Lincoln

The amount of amino acid sequences are increasing in databases. Various methods are being used to extract information from this molecular data. In this study, we examine two multivariate analysis methods: principal component analysis (PCA) and cluster analysis (CA). Protein families analyzed are membrane and soluble diiron enzymes. The objectives of this study are to use PCA to extract information from physico-chemical properties of the 20 amino acids, to use auto and cross covariance (ACC) to transform amino acid sequences into quantitative measures, and to use PCA and CA to classify the transformed protein sequences. ACC data transformation makes it possible to translate amino acids of different length into same number of variables. This enables us to use multivariate analysis methods on protein sequences without relying on multiple alignments but still including positional information. We will examine the performance of these multivariate methods in protein family classification.

OTA, Tatsuya

Characterization of major histocompatibility complex (MHC) class II beta genes in the Antarctic toothfish, *Dissostichus mawsoni*

Ota T. [1], Cheng C. C. [2], Miyake T. [3], Amemiya C. T. [3]. (1) The Graduate University for Advanced Studies, (2) University of Illinois, (3) Benaroya Research Institute at Virginia Mason

Notothenioid fishes are known for their dominant presence in the frigid waters of the Southern Ocean and for their unprecedented adaptations, many of which remain to be fully elucidated. As surveys of genetic polymorphisms have been instructive in our understanding of vertebrate evolution, we have characterized the MHC class II genes of the Antarctic toothfish, a large notothenioid fish that has recently come under scrutiny due to overt overfishing. MHC class II beta genes obtained by RT-PCR have been sequenced and numerous nucleotide substitutions have been found at potential antigen recognition sites. Our observations are reminiscent of those seen in the polymorphic MHC genes of other vertebrates and raise the possibility of using the loci for population assessment studies. Screening of a toothfish genomic BAC library of 5x coverage yielded 32 positive clones, some of which may contain multiple MHC class II beta loci. The evolutionary implications of our findings will be discussed.

PALMER, Jeffrey

Plants as a Model System for Studying Horizontal Gene Transfer

Palmer J. D. [1]. (1) Indiana University

Horizontal gene transfer (HGT) is widely recognized as a major force in bacterial evolution, but its prevalence and importance in eukaryotes, especially multicellular eukaryotes, is quite unclear. We have found that plant mitochondrial genes undergo surprisingly frequent and recent HGT, whereas chloroplast genes do not. Why this is so, and what some of the mechanisms and consequences of plant-to-plant mitochondrial HGT might be, will be discussed. The limited evidence bearing on HGT in nuclear genomes of plants, animals, and fungi will be discussed in the context of approaches to tackling this challenging problem.

PANOPOULOU, Georgia

How often duplicates adopt a novel role? Views from a WMISH screen of amphioxus genes and their duplicated zebrafish orthologs

Panopoulou G. [1], Weise V. [1], Groth D. [1], Hennig S. [1], Poustka A. J. [1], Lehrach H. [1]. (1) Max-Planck Institute for Molecular Genetics

Gene duplication is a major force driving the evolution of genomes. Gene duplicates account for 8-20% of the genes in eukaryotic genomes. We are carrying out a WMISH screen of amphioxus genes and their duplicated zebrafish orthologs. Teleost fish have undergone additional to the rest of vertebrates gene duplication, thus being an ideal system to study the functional diversification that can be reached after duplication. Amphioxus as the closest living invertebrate chordate to vertebrates can serve as a reference point for deducing ancestral gene function. It is hypothesised that functional diversification via either aminoacid substitutions in coding regions or changes in the number and type of shared regulatory motifs of duplicates can secure their retention in genomes. The above screen intends to answer: How often are duplicates adopted to non-overlapping expression domains? How often are these domains novel? Do novel expression patterns correlate with the rate that genes evolve?

PARHAM, Peter

Co-evolution of MHC class I and NK-cell receptors in primates

Parham Peter. [1]. (1) Stanford University

Natural killer cells are lymphocytes of the innate immune response that secrete cytokines and kill infected cells at early times of infection. In mammals, they also regulate implantation of the embryo during the first trimester of pregnancy. NK cells thus influence the reproduction and survival of organisms. In performing their functions mammalian NK cells use a variety of activating and inhibitory cell-surface receptors, many of which engage MHC class I or class I-like ligands. The MHC class I gene family is characterized by polymorphism within species and changes in gene organization between species. Genes encoding some of the NK-cell receptors have similar characteristics. Thus, the elaborate Ly49 gene family which murine NK cells use as lectin-like receptors for polymorphic MHC class I determinants, is represented by a single non-functional gene in humans. Conversely, the elaborate killer-cell immunoglobulin-like receptors (KIR) that human NK cells use to engage polymorphic MHC class I, is represented in mice by two genes with apparently different functions. To gain understanding of how KIR evolve and maintain functional interactions with MHC class I we have been comparing the two gene families in several hominoid species.

PARSCH, John

Molecular evolution of sex-biased genes in Drosophila

Parsch J. [1]. (1) University of Munich

Data from cDNA microarray experiments and comparative genomic studies of Drosophila are used to test the "*faster male evolution*" hypothesis. Orthologs of genes showing either male-, female-, or no sex-bias in their expression were compared over short (within the D. melanogaster subgroup) and long (within the subgenus Sophophora) timescales. Comparisons of nonsynonymous/synonymous substitution rates (dN/dS) between species of the D. melanogaster subgroup revealed that male-biased genes had significantly faster rates of evolution than both female-biased and unbiased genes. This difference was

due primarily to a higher dN in the male-biased genes. In comparisons between D. melanogaster and D. pseudoobscura, genes with highly-biased male expression were significantly more divergent than genes with highly-biased female expression. These results support faster male evolution and suggest that male and female genes are subject to contrasting selective forces.

PASCUAL, Marta
Polymorphism in autosomal and X-linked neutral loci. What does a recent colonization process tell us?
Pascual M. [1], Mestres F. [1], Balanya J. [1], Serra L. [1]. (1) Dept Genetics, Universitat de Barcelona
Recent colonizations allow to investigate the genetic consequences of founder effects and infer the processes affecting variability in natural populations. Highly polymorphic markers such as microsatellites can better detect the impact of a founder effect and were used to analyze the colonization of America by *Drosophila subobscura* occurred within the last 25 years. Two populations from each North and South America were compared to five European populations, within the ancestral range of the species. New World areas were not independently colonized, 78% of the alleles were common to both colonized areas while the rest (19%) were found in only one locality and always at low frequency (<0.06). The significantly higher reduction in variability in X-chromosomal compared to autosomal loci detected in American populations disappeared when a correction was applied considering an equal number of males and females among the colonizers and hence, different contribution from X chromosomes and autosomes to founder populations. Hence the strong bottleneck (4-9 effective founders) suffered by these populations during their colonization explains our findings without having to invoke to other selective forces. On the other hand a significantly larger variability in X-chromosomal loci was found in ancestral populations consistent with the background selection model.

PASCUAL, Marta
Selection and speciation of colour morphs in *Pseudodistoma crucigaster* (Ascidiacea: Polyclinidae)
Pascual M. [1], Tarjuelo I. [2], Posada D. [3], Crandall K. A. [4], Turon X. [2]. (1) Dept Genetics, Universitat de Biologia, (2) Dept Animal Biology, Universitat de Barcelona, (3) Dept Biochem, Genet & Immuno, Universidad de Vigo, (4) Dept Integrative Biology, Brigham Young University
Variation in pigmentation is common in benthic invertebrates, although few studies have shown the existence of genetic differentiation of chromatic forms in colonial organisms. We studied the genetic structure of a colonial ascidian with populations of different colour morphs in the northwestern Mediterranean. A fragment of the COI mitochondrial gene was sequenced in seven populations of *Pseudodistoma crucigaster* belonging to three different colour morphs (orange, yellow and grey). Maximum likelihood analyses showed two well-supported clades separating the orange morph from the yellow-grey morphotypes. Genetic divergence between these clades was 2.12%, and ST values between populations of the two clades were high (average 0.936), pointing to genetic isolation. Coalescence analyses indicate that the separation of the two clades took place right after the Messinian crisis of the Mediterranean Sea (5-6 MYA). Non-neutral mtDNA evolution is observed in our data when comparing the two clades, showing a significant excess of nonsynonymous polymorphism within the yellow-grey morphotype using the McDonald-Kreitman test, which is interpreted as a further support of reproductive isolation. We conclude that the two clades might represent separate species.

PAWLOWSKA, Teresa
Organization of individual genetic variation in arbuscular mycorrhizal fungi
Pawlowska T. E. [1], Taylor J. W. [1]. (1) University of California Berkeley
Arbuscular mycorrhizal (AM) fungi (Glomeromycota) are thought to be the oldest asexual multicellular organisms. Cells of AM fungi contain hundreds of nuclei. Unusual polymorphism of rDNA observed in individuals of AM fungi inspired a hypothesis that heterokaryosis, i.e. the coexistence of many dissimilar nuclei in cells, occurs throughout the AM fungal life history. Our study of the transmission of polymorphic genetic markers in natural isolates of *Glomus etunicatum,* and direct PCR amplification of

rDNA from microdissected nuclei, support the alternative hypothesis of homokaryosis. Presence of polymorphic rDNA copies in each nucleus signals a relaxation of concerted evolution, a recombination driven process responsible for homogenization of rDNA repeats. To understand causes of relaxation of concerted evolution in AM fungi we are investigating their genome structure by assessing genomic DNA renaturation kinetics and sequencing randomly selected genomic clones.

PAYSEUR, Bret

Testing Genotype-Phenotype Association in Natural Populations

Payseur B. A. [1], Clark A. G. [1], Boerwinkle E. [2], Sing C. F. [3]. (1) Cornell University, (2) University of Texas, (3) University of Michigan

Understanding the connection between phenotypic and genotypic variation for complex traits presents a formidable challenge to evolutionary biologists. Association testing in natural populations is widely used and has been successful in a number of cases. However, this method is not ideal because it requires explicitly identifying large numbers of genotypic classes when data from many sites are considered. Alternatively, bulk segregant analysis tests for differences in allele frequency between the upper and lower tails of the phenotype distribution. We generalize this approach to ask whether the distribution of multi-site genotypes differs between the tails. Statistical significance of observed differences is assessed by randomly permuting membership in phenotypic classes. We use this method to identify associations between single nucleotide polymorphisms (SNPs) and risk factors in cardiovascular disease using genotypes in 3,817 individuals at 80 sites from the APOA1/C3/A4/A5 gene cluster in humans. Significant differences between phenotypic classes in per-site standardized allele frequency and several other genotypic metrics are detected.

PECON-SLATTERY, Jill

Evolution of genes located on the Y-chromosome in the cat family Felidae

Pecon-Slattery J. [1], King V. [2], Pearks Wilkerson A. J J. [3], Murphy W. J. [3], Johnson W. E. [1], O'Brien S. J. [1]. (1) Laboratory of Genomic Diversity NCI-Frederick MD, (2) TCAG Rockville MD, (3) Laboratory of Genomic Diversity SAIC-Frederick MD

The recent radiation of the cat family Felidae offers a unique opportunity to examine patterns of evolution within Y-chromosome genes. Using phylogenetic methods, sequences from four single copy genes located in the non-recombining region of the Y chromosome are compared across 36 species of cat. Introns from *SMCY*, *UBE1Y* and *ZFY* (3604 bp) exhibit exceptional support for each of the eight recognized felid lineages and diagnostic substitutions identifying nearly all species. Analysis of the male-determining gene *SRY* and its adjacent genomic flanks (2957 bp) suggests this reproductively important gene to be tightly linked with speciation. Adaptive evolution of *SRY* is assessed by codon analyses using PAML. No specific sites are under selection, but an abundance of w>1 (w = Dn:Ds) among all branches within the *SRY* tree suggests most speciation events are marked by nonsynonymous changes. Overall, genes within the NRY are powerful markers of evolution across felid species.

PENNY, David

RNA processing in the Eukaryotic Ancestor

Collins L. J. [1], Matheson J. W. [1], **Penny D.** [1]. (1) Massey University

The nature of RNA processing in the last common eukaryote ancestor is enigmatic, but now involves complexes of ribozymes and associated proteins. In particular, we have used new bioinformatic techniques to search for spliceosomal and RNase-P components in genomes of basal lineages of eukaryotes. For proteins, an Ancestral Sequence Reconstruction technique has been developed that finds a large number of proteins in basal eukaryotes that occur in spliceosomes of Crown eukaryotes. A new version of RNAmotif adapted for parallel computers finds the matching RNA molecules. These have shown to be expressed by RT-PCR. From the distributions, it appears that in the eukaryotic ancestor, splicing was already a complex procedure and that there was more than one form of splicing present. We

now need to understand the alternative life-cycle strategies that may have diverged very early, and these have been studied by simulation. The results are incompatible with any simple fusion hypothesis for the first eukaryote.

PEREZ, Concepcion
Independent and Repeat Coat Color Mutations Found Amongst Australian Populations of Domestic Norway Rats
Perez C. [1]. (1) University of California, Berkeley
In December, 2002, a small breeding nuclei of domestic Norway rats was shipped from Brisbane, Australia to San Francisco, California. Included were representatives of a novel autosomal dominant spotting gene known as "downunder", which is characterized by enlarged and varied ventral spotting. Three color genes were also included: Australian blue, Australian mink, and Australian red-eyed dilution. To date, Australian blue has been confirmed as a recessive allelic independent repeat mutation of the blue gene previously described by Roy Robinson and symbolized as "d". The Australian mink gene is an autosomal recessive. Upon test breeding to American domestic stock, it is known that there is an American counterpart gene, and the Australian allele is dominant over the American allele. To date, all testcrosses of the Australian red-eyed dilution gene to American red-eyed stock have given negative results. Future testcrosses will include testing Australian red-eyed dilution to American pink-eyed dilution.

PEREZ-GONZALEZ, Cesar
Studies of Line-1 Element Interactions with the Human Genome
Perez-Gonzalez C. E. [1], Furano A. V. [1]. (1) National Institutes of Health
The Long Interspersed Nuclear Element 1 (LINE-1, L1) is a non-LTR retrotransposon present in all mammalian genomes. It contains two open reading frames (ORFs), the first encoding a protein capable of binding nucleic acids and protein, while the second contains the reverse transcriptase and endonuclease necessary for its mobility. L1 retrotransposition is responsible not only for it's own mobilization but also that of SINEs and processed pseudogenes, generating between 30-50% of the DNA in the human genome. Thus its activity has had a major impact on mammalian genomes. Sequence analysis of different LINE-1 families in primates has shown that a coiled-coil domain located in ORF 1 has undergone adaptive evolution, which suggests that this region may be involved in an interaction with the host. This could involve an interaction with a host protein required for retrotransposition or one that suppresses L1 activity. We are doing yeast two-hybrid screens using either a modern or ancient L1 family as bait to screen human cDNA libraries. Proteins shown to interact with L1 ORF1 via this screen will then be knocked out using siRNA or subjected to over expression assays to test their biological relevance.

PETERSON, Stephen
EF-1 alpha intron position and number changes among *Penicillium* species
Peterson S. [1]. (1) U.S. Department of Agriculture
The 5 prime half of the EF-1 alpha gene was sequenced from numerous Penicillium species for phylogenetic studies. AA sequences predicted from the DNA sequences revealed 0-3 introns in each species and these introns were inserted in eight different locations in the gene. MP Trees from the AA coding regions of EF-1 alpha and calmodulin, and 28S rDNA were generally congruent, but deep branches were unresolved. Molecular clock calculations were made using the GTR+I+G model. Using those calculations, some introns appear to have arisen in the past 4-10 My, and are unique to their lineages. Comparison to outgroup species that diverged 45 Mya reveals no consistent intron position shared by all species. In contrast, location and number of calmodulin introns was consistent for ingroup and outgroup species. Congruence of the gene trees suggests that the EF-1 alpha genes are not paralogous, and it is likely that the observed distribution of introns is evidence of recent intron loss and gain.

PHINCHONGSAKULDIT, Jaros

Genomic analysis of promoter: Long-PCR and Pseudo-recombination

Phinchongsakuldit J. [1], Crawford D. L. [1]. (1) University of Miami

We are developing *in vivo* methods to analyze the effect of nucleotide variation on gene expression. This approach requires the precise amplification of large genomic DNA. However, *in vitro* recombination during PCR can create chimera amplicons between two (or more) templates, and this occurs at a high frequency in long PCR. The DNA sequences derived from this pseudo-recombination can therefore be misleading. The pseudo-recombination rates in the amplifications of 5kb *Ldh-B* locus are from 2.38% to 36.4%. The attempt to minimize pseudo-recombination during PCR is conducted by optimizing the annealing time and primer concentration. The results show that decrease in annealing time up to 10 seconds can minimize the pseudo-recombination rate. However, further reduction beyond 10 seconds no longer decreases the pseudo-recombination rate. The optimization of the primer concentration does not show a consistent pattern in minimizing pseudo-recombination rates. Thus, the optimized annealing time and primer concentration can be different for different sets of primers.

PIONTKIVSKA, Helen

Between-Host Evolution of CTL Epitopes in Human Immunodeficiency Virus Type 1 (HIV-1): an Approach Based on Phylogenetically Independent Comparisons

Piontkivska H. [1], Hughes A. L. [1]. (1) University of South Carolina

In human immunodeficiency virus type 1 (HIV-1), mutations that escape from cytotoxic T-lymphocyte (CTL) recognition have been documented, and sequence analyses have provided indirect support for the hypothesis that natural selection has favored CTL-escape mutants within an infected host. In spite of such evidence for within-host selection by CTL, it has been more difficult to determine how natural selection by host CTL has influenced long-term evolution of HIV-1. We used statistical analysis of published HIV-1 genomic sequences to examine the role of natural selection in between-host evolution of CTL epitopes. Based on a phylogenetic analysis, we identified 21 pairs of closely related genomes isolated from different hosts and examined the pattern of nucleotide substitution in genomic regions encoding well-characterized CTL epitopes. The results revealed that certain CTL epitopes have been subject to repeated positive selection across the population, while others are generally conserved. Furthermore, evidence of positive selection was associated with divergence from the canonical epitope sequence and with an enhanced frequency of convergent amino acid sequence changes in CTL epitopes. The results support the hypothesis that CTL-driven selection has been a major factor in the long-term evolution of HIV-1.

PIROG, Katarzyna

Cryptic diversity in the tide-pool ciliate *Strombidium oculatum*.

Pirog K. A. [1], Griffin A. J. [1], Snoeyenbos-West O. L. [1], Costas B. [2], McManus G. B. [2], Katz L. A. [1]. (1) Department of Biological Sciences, Smith College, (2) Department of Marine Sciences, UConn

There are two contrasting views on ciliate phylogeography, but few molecular data to distinguish between hypothesis. The first maintains that ciliate populations are cosmopolitan, and most species have been described. The second posits a greater degree of endemism and argues for greater numbers of ciliates species. To distinguish between these hypotheses, I am analyzing multiple molecular markers (ITS regions, 5.8S rDNA, histone H4, mitochondrial SSU and a-tubulin) from twenty-three populations of two choreotrich ciliates: *S. oculatum* and *S. stylifer*. My preliminary analyses reveal that *S. oculatum*, previously described only from Europe, also exists in North America and that there is considerable genetic divergence within populations. In contrast, we find no genetic variation among North American populations of *S. stylifer*. These data are consistent with higher species number for the morphological species *S. oculatum*, and cosmopolitanism in *S. stylifer*. We discuss interpretations given the geological history of these regions.

PITTMAN, K.J.

The Concept of Subspecies in Evolutionary Biology

Pittman K. J. [1]. (1) University of Kansas

The concept of species is of central importance to biology. While evolutionary biologists have not been able to agree on one common species concept, evolutionary biologists do agree that species are real, true entities in nature, not just constructs of the human mind. If species are accepted as real entities, then evolutionary biologists should discuss the nature of subspecies, the taxonomic rank below subspecies. My research shows that subspecies are not real entities in nature but human constructs. As such, this greatly impacts how evolutionary biologists approach their research programs. Use of the subspecies concept has obscured where true biodiversity lies. If we as scientists cannot properly identify biodiversity, we cannot protect it. I will present evidence that shows that the rank of subspecies in taxonomy should be discarded. I will also show how evolutionary research will greatly benefit from an in-depth review and discussion of the subspecies concept.

PODLAHA, Ondrej

Non-neutral evolution of the transcribed pseudogene Makorin1-p1 in mice

Podlaha O. [1], Zhang J. [1]. (1) University of Michigan

Pseudogenes are commonly recognized by their lack of open reading frames in spite of sequence similarity to functional genes, and are generally thought to evolve neutrally without functional constraints. Makorin1-p1 is a transcribed pseudogene first identified in *Mus musculus*. Earlier studies revealed that the Makorin1-p1 transcript regulates the expression of its paralogous functional gene Makorin1. Here we show that Makorin1-p1 was formed after the separation of *Mus* from *Rattus* but before the divergence of *M. musculus* and *M. caroli*. The transcribed 5' half of Makorin1-p1 exhibits rates of point and indel substitutions that are 2 to 3 times lower than those in the untranscribed 3' half of the pseodogene, strongly suggesting that the 5' half is under functional constraints and is not neutrally evolving. Although the transcript of Makorin1-p1 likely functions by its sequence similarity to Makorin1, we find no evidence of gene conversion between them, indicating that functional conservation alone is sufficient to maintain their coordinated evolution.

POLLARD, Daniel

Evolutionary Properties of Early-Embryonic Enhancers in *Drosophila*

Pollard D. A. [1], Moses A. M. [1], Eisen M. B. [1]. (1) UC Berkeley/Lawrence Berkeley National Lab

The widely accepted paradigm that comparative sequence analysis will aid in the functional annotation of *cis*-regulatory regions in eukaryotic genomes can now be tested in *Drosophila* with the completed genome of *D. pseudoobscura*. Here we report a characterization of the evolutionary properties of functional enhancers and non-enhancer sequences. In addition to nineteen previously annotated early-embryonic enhancers, we analyzed twenty-three sequences previously identified in a computational screen for high-density clusters of early-embryonic transcription factors, five of which drive early-embryonic expression patterns and eighteen of which do not. As expected, functional enhancers tend to maintain a higher density of binding sites between the two species compared to non-functional sequences. Surprisingly, substitution rates in binding sites in both functional and non-functional sequences appear to be close to background noncoding substitution rates, however, insertion/deletion rates in binding sites appear to be much lower in functional enhancers than in non-functional sequences. Various models of binding site and enhancer evolution will be discussed.

PONGER, Loic

Evolution of the DNA methylation in the eukaryotes

Ponger L. [1], Li W. H. [1]. (1) University of Chicago

The DNA methylation is a natural modification of the DNA corresponding
to the transfer of a methyl group on the nucleotides performed by some specific

enzymes, the DNA methyltransferases (MTases). In eukaryotes, the methylation is specific to the cytosines but exhibits a great variety of patterns (ie, level of methylation, methylated sequences) associated with two main functions, the control of the repeated elements (in fungi) and the regulation of the gene expression (in vertebrates and plants). The similarities observed between the methylation in plants and vertebrates could be explained by a convergence in response to an increasing number of genes or could correspond to an ancestral trait modified in the fungi. To understand the evolution of the methylation, we searched all the putative DNA MTases described in all the complete eukayotic proteomes and we compared their phylogeny with their biochemical properties and with the methylation pattern of these organisms.

POPADIC, Aleksandar

Developmental basis of hind leg evolution in insects

Mahfooz N. S. [1], Li H. [1], **Popadic A.** [1]. (1) Wayne State University

Diversification of leg appendages is one of the hallmarks of morphological evolution in insects. In particular, insect hind (T3) legs exhibit a whole spectrum of morphological diversification, ranging from uniform to extremely modified. In order to elucidate the developmental basis of T3 leg evolution, we have examined the expression patterns of the homeotic gene, *Ultrabithorax* (*Ubx*), in a broad range of species. Our results show that *Ubx* expression in hemimetabolous insects is localized only in specific T3 leg segments undergoing differential growth (compared to their fore leg counterparts). Furthermore, the degree of enlargement is associated with the developmental timing of *Ubx* expression. Functional studies show that inhibition of the Ubx expression results in the decrease in size of specific T3 leg segments. These results suggest that the diversification of insect hind legs was influenced by changes in both the spatial and temporal regulation of the *Ubx* gene.

PORTNOY, Matthew

Multi-Species Comparative Sequence Analysis of the *GDF6* Growth Differentiation Gene

Portnoy M. E. [1], NISC Comparative Sequencing Program [2], Mortlock D. P. [3], Green E. D. [1]. (1) Genome Technology Branch, NHGRI, NIH, Bethesda, MD, (2) NIH Intramural Sequencing Center, NIH, Bethesda,MD, (3) Vanderbilt University Medical Center, Nashville,TN

Comparative sequence analysis is a powerful tool to study gene structure and regulation. We have taken such an approach to identify cis-acting regulatory sequences associated with *GDF6*, which is required for proper skeletal limb joint formation. Following cross-species BAC mapping, we generated genomic sequence orthologous to a ~200-kb region encompassing *GDF6* (in the human genome) from chimpanzee, baboon, pig, cow, cat, dog, rat, platypus, and zebrafish. The generated sequences were analyzed in parallel with available human, mouse, chicken, xenopus, and fugu sequences. All sequences were aligned using the program Multi-PipMaker and, utilizing two different algorithms, analyzed to identify multi-species conserved sequences (MCSs). The first method (WebMCS) uses a 25-base sliding window and takes into account the neutral substitution rate of each species. The second (Exactplus) requires exact matches of 6 to 25 bases among varying numbers of species, and then extends alignments from some number of remaining species. WebMCS identified between 58 and 153 MCSs and Exactplus identified between 114 and 2,196 MCSs across the region (depending on the parameters used with each program). Experiments are now underway to determine if these MCSs serve a functional role, such as regulating the expression of *GDF6*.

PRACHUMWAT, Anuphap

Higher Gene Duplicabilities for Metabolic Proteins than for Non-Metabolic Proteins in Yeast and *E. coli*

Marland E. [1], **Prachumwat A.** [2], Maltsev N. [1], Gu Z. [3], Li W. H. [3]. (1) Mathematics & Computer Science Division, Argonne National Laboratory, (2) Committee on Genetics, University of Chicago, (3) Ecology and Evolution, University of Chicago

Although the evolutionary significance of gene duplication has long been appreciated, it remains unclear what factors determine gene duplicability. In this study we investigated whether metabolism is an important determinant of gene duplicability because cellular metabolism is crucial for the survival and reproduction of an organism. Using genomic data and metabolic pathway data from the yeast (*Saccharomyces cerevisiae*) and *Escherichia coli*, we found that metabolic proteins indeed tend to have higher gene duplicability than non-metabolic proteins. Moreover, a detailed analysis of metabolic pathways in these two organisms revealed that genes in the central metabolic pathways and the catabolic pathways have, on average, higher gene duplicability than do other genes and that most genes in anabolic pathways are single-copy genes.

PRITHAM, Ellen

Unexpected diversity and differential amplification of transposable elements in four species of Entamoeba protozoans

Pritham E. J. [1], Feschotte C. [1], Wessler S. R. [1]. (1) The University of Georgia

We report the first comprehensive analysis of transposable element (TE) content in the compact genomes (~20 Mb) of four species of Entamoeba unicellular protozoans for which draft sequences are now available. Entamoeba histolytica and E. dispar, two human parasites, have many LINE retrotransposons, but few DNA transposons. In sharp contrast, the reptile parasite E. invadens and the free living E. moshkovskii contain few LINEs, but harbor a vast, diverse and recent population of DNA transposons. The recently amplified element encoded transposases belonging to 4 superfamlies including a variety of Tc1/mariner elements, which are widespread in animal, plant and fungi but were previously described only from ciliates among protozoans. Representatives of three other superfamilies (hAT, mutator, and piggyBac) are identified for the first time in any protozoan. The dramatic diversity of transposons and their differential amplification among closely related species with similarly compact genomes are discussed in the context of the biology of Entamoeba protozoans.

PTAK, Susan

Absence of the TAP2 recombination hotspot in chimpanzees

Ptak S. E. [1], Roeder A. D. [1], Stephens M. [2], Gilad Y. [1], Paabo S. [1], Przeworski M. [1]. (1) Max Planck Institute for Evolutionary Anthropology, (2) University of Washington

Recent experiments using sperm typing have demonstrated that, in several regions of the human genome, recombination does not occur uniformly but instead is concentrated in "hotspots" of 1-2 kb. Moreover, the crossover asymmetry observed in a subset of these has led to the suggestion that hotspots may be short-lived on an evolutionary time-scale. To test this possibility, we focused on a region known to contain a recombination hotspot in humans, TAP2, and asked whether chimpanzees, the closest living relative of humans, harbor a hotspot in a similar location. Specifically, we used a new statistical approach to estimate recombination rate variation from patterns of linkage disequilibrium in a sample of 24 western chimpanzees. This method has been shown to produce reliable results on simulated data and on human data from the TAP2 region. Strikingly, however, it finds very little support for recombination rate variation at TAP2 in the western chimpanzee data. Moreover, simulations suggest that there should be stronger support if there were a hotspot similar to the one characterized in humans. Thus, it appears that the human TAP2 recombination hotspot is not shared by western chimpanzees. This finding indicates that fine-scale recombination rates can change between very closely related species.

QUESNEVILLE, Hadi

***In silico* detection of new transposable element families in genomic sequences**

Quesneville H. [1], Autard D. [1], Andrieu O. [1], Nouaud D. [1], Anxolabéhère D. [1]. (1) Institut Jacques Monod

The procedures generally used to detect transposable elements (TEs) in nucleic acid sequences rely on sequence similarity search with previously characterized elements. However, these methods are likely to miss many elements. We tested a two steps strategy: (i) First we search for TE sequences by both

comparing the six-frame translations of the nucleic acid sequences of known TEs with these of the genomic sequence. The flanking sequences of these hits are aligned all together, and similar flanking sequences are used to build TE consensus sequences. (ii) Second, we search among the genome regions that have a base composition close to that of TEs using Hidden Markov Models (HMM). The HMM is trained on consensus sequences obtained in the first phase. We detect more than 300 new TE in *D. melanogaster*. This approach is very interesting for the detection and the identification of new TEs in organisms for which only very few TEs are known.

RAES, Jeroen

The role of frame shift mutations in the evolution of new gene functions after gene duplication

Raes J. [1], Van de Peer Y. [1]. (1) Ghent University - VIB

Due to their drastic nature, frame shift mutations are generally considered to yield defective proteins and are therefore quickly removed from a population. However, after gene duplication, a reduction of selection pressure on one of the duplicates might allow the persistence of these mutations and could lead to neofunctionalisation, as was shown recently (Vandenbussche et al., 2003 Nucl Acids Res).
In this study, we designed an automated approach to detect conserved frame shift mutations and used it to investigate the prevalence of this phenomenon in vertebrate gene families. Several instances of this process could be found, and we could show that frame shifts that occur after gene duplication events can remain conserved over large evolutionary timescales (up to 400 MY) in multiple species, indicating the probable functional importance of the newly created protein domains. Finally, our results revealed that alternative splicing, in which one of the splicing forms has a frame shift, also constitutes a mechanism for functional divergence and protein evolution.

RAFALSKI, Antoni

Comparative Sequencing of Maize: Structure and Origin of Allelic Non-Colinearities

Brunner S. [1], Morgante M. [2], Jung M. [1], Tingey S. V. [1], **Rafalski A.** [1]. (1) DuPont Crop Genetics, (2) Universita' di Udine

Analysis of genetic diversity among individuals of the same species is usually based on the assumption that that an allelic counterpart is found for each gene. Two maize inbred lines, McC and B73, have been recently compared at the *bronze1* genomic region (Fu H. and Dooner H.K., *Proc Natl Acad Sci* USA 2002, 99: 9573-9578). Contrary to expectations, the DNA sequences of the two inbreds not only differ extensively in the repetitive DNA segments, but also some genes present in one allele are found missing in the other allele. We extended these observations to other genetic loci and surveyed the occurrence of genic and intergenic non-colinearities in maize gene pool. Four genomic segments of ca 250 kb each have been sequenced from two inbred lines, B73 and Mo17. Intergenic non-homologies were common and frequently consisted or relatively recent insertions of retrotransposon-derived sequences. The differences between the two alleles ranged from 20% to nearly 50%, depending on the locus. Genic non-homologies were relatively rare and were usually pseudogenes. At one locus, Adh1, both alleles were nearly identical, except for SNPs and indels, which were present at normal frequency. The consequences of these findings for evolutionary history of maize will be discussed.

RAFTIS, Frances

Patterns of Replacements at Changing Protein Interfaces

Raftis F. M. [1], Golding G. B. [1]. (1) McMaster University

Functional constraints in proteins generally correspond to a change in the replacement rates of the residues involved. The change is typically a decrease, resulting in conservation of residues involved in protein-protein contacts, interaction with substrate, or maintaining globular structure. Though the interfaces of obligate homooligomers are generally conserved, there are some cases where this may not be so. A number of homologous proteins have exhibited a change in their quaternary structure in different species, resulting in extant versions of different subunit assemblages. We studied the patterns of

replacements at the interfaces of these proteins to see if and how their conservation differs from proteins with invariant quaternary structures.

RAGHUPATHY, Narayanan

The role of duplication in the evolution of vertebrate insulin signalling

Raghupathy N. [1], Durand D. [1]. (1) Carnegie Mellon University

We present results from a comparative genomics study on the role of duplication in verterbrate insulin signalling. In invertebrates, metabolism and growth are mediated through a single insulin receptor. In vertebrates, metabolism is mainly controlled by the insulin receptor, while growth and differentiation are primarily mediated through the IGF1 receptor. As a first step in investigating the hypothesis that the specialization of these functions arose through duplication of an entire ancestral signalling pathway, we investigated the human chromosomal regions surrounding the insulin ligand and receptor families for evidence of large-scale duplication. We present conserved gene clusters in these regions, which we have confirmed statistically using a new probabilistic method to reject the null hypothesis that paralogs in these gene clusters were duplicated independently.

RAMAKRISHNAN, Uma

Precision and Accuracy of Divergence Time Estimates from STR and SNPSTR Variation

Ramakrishnan U. [1], Mountain J. L. [1]. (1) Stanford University

STR polymorphisms are often the markers of choice to estimate intraspecific divergence time due to their high mutation rate. However, these markers are characterized by high levels of homoplasy, often leading to inaccurate inference of population divergence. A SNPSTR is a genetic system that consists of an STR polymorphism closely linked (<500bp) to one or more single nucleotide polymorphisms (SNPs). We use coalescent-based simulations in the context of models of demographic history to compare divergence time estimates based on SNPSTR haplotype frequencies and STR frequencies. We demonstrate that estimates of divergence time based on STR variation on the background of a derived SNP allele are more accurate (3-7% bias for SNPSTR vs 11-20% bias for STR) and more precise than STR based estimates, conditional on a recent SNP mutation. These results hold for models involving complex demographic scenarios with gene flow, population expansion, and population bottlenecks. Varying the timing of the SNP revealed that estimates of divergence time are sensitive to SNP age. However, varying both mutational properties of STR loci and SNP age demonstrated that multiple independent SNPSTR systems provide less biased estimates of divergence time. In light of our simulations, we interpret estimates from data for human populations.

RAMOS, Edward

Identification of Functional Cis-Regulatory Modules by Comparative Analysis of *Drosophila melanogaster* and *Drosophila pseudoobscura*.

Ramos E. [1], Corces V. G. [1], Lai Z. C. [2]. (1) Johns Hopkins University, (2) The Pennsylvania State University

Cis-regulatory modules (CRMs) are known to control when, where, why and how a gene is turned on. It is common knowledge that enhancers, repressors, insulators and other regulatory elements work in conjunction to operate the cell's gene activity. In this study, we took a molecular approach to identify cis-regulatory elements. We identified regulatory modules for the *Drosophila melanogaster yan* gene and conducted a thorough analysis of these enhancers. Comparative analysis of these functional domains between *Drosophila melanogaster* and *Drosophila pseudoobscura* show a high level of conservation. This level of conservation is only seen within these functional domains and not within intervening sequences. We then looked to see if this conservation specificity would hold true for other genes. We inspected the enhancers of several genes and found that they also contain a high degree of conservation at functional domains. We believe that this data validates the idea that studies of conserved non-coding sequences can be used as a way to target cis-regulatory modules within an organism genome.

RAND, David

The human mitochondrial genome and proteome show opposing departures from neutral evolution

Rand D. M. [1], Cezairliyan B. E. [1], Aleghan Z. [1]. (1) Brown University

A consistent departure from neutrality in animal mtDNA is an excess of amino acid polymorphism within species relative to divergence between species. This has been explained by selection against mildly deleterious nonsynonymous changes that are more likely to be polymorphic than fixed between species. This departure from neutrality should hold for other classes of functionally distinct nucleotide changes. We test this prediction in the complete mitochondrial genomes of hominoids by comparing ratios of polymorphism to divergence for preferred and unpreferred synonymous changes, for synonymous and nonsynonymous changes, and for conservative and radical amino acid changes. MK tests confirm a significant excess of nonsynonymous polymorphism. Synonymous site tests indicate weak negative selection. In contrast, conservative/radical tests reveal an excess of radical fixations, indicating positive selection unique to the human mitochondrial proteome. Lineage specific non-neutral fixations of distinct functional classes are also observed. These data are not compatible with the same distribution of selection coefficients for the human mitochondrial genome and proteome and cannot be reconciled by nearly neutral models with changing population sizes. We suggest that in human history selection on mitochondrial genome expression has been negative while selection on mitochondrial protein function has been positive.

RAQUEL, Tavares

Comparative analysis of nuclear receptors expression patterns

Tavares R. [1], Robinson-Rechavi M. [2], Laudet V. [1]. (1) ENS Lyon LBMC Equipe Struc.Evol.Rec.Nucl.Hormones, (2) Joint Center for Struc. Genomics UCSD Burnham Institute

Nuclear hormone receptors (NRs) are transcription factors which mediate the action of a wide variety of ligands and regulate many aspects of development and physiology. They have been shown to be good markers of genome duplication, harbouring a strong phylogenetic signal. This makes them ideal tools to understand the evolution of genes after duplication. We have compared the expression patterns of NRs in vertebrates, using EST and *in situ* hybridization data. Our main goals are to address functional and evolutionary questions specific to this gene family and to test models of expression divergence after gene duplication and/or speciation. Furthermore, our analysis will be used for the definition of an automatic method of comparison of expression patterns in gene families.

RATERMAN, Denise

Phylogenetic relationships of the cuscuses and brushtail possums (Marsupialia: Phalangeridae)

Raterman D. M. [1], Hamilton A. T. [2], Springer M. S. [1]. (1) University of California - Riverside, (2) Lawrence Livermore National Laboratory

The family Phalangeridae is comprised of approximately two-dozen extinct and extant species that include the brushtail possums (*Trichosurus*), scaly-tailed possums (*Wyulda*) and cuscuses (*Phalanger*, *Strigocuscus*, *Spilocuscus*, and *Ailurops*). This study uses molecular sequence data from the nuclear gene BRCA1 and the mitochondrial genes 12s ribosomal RNA and tRNA-valine to elucidate the relationships within this group. Of particular interest is the placement of *Ailurops ursinus*, the bear cuscus. This species has been considered the most plesiomorphic sister group to all other phalangerids based on morphological characteristics. However, phylogenetic results from this study suggest that Trichosurins are the basal group while *Ailurops* is sister to *Phalanger* and *Spilocusus*. This study also confirms the results of other molecular studies that place *Phalanger gymnotis* within the genus *Phalanger*. This species had previously been classified as *Strigocuscus* based on morphological data.

RAY, David

Platyrrhine phylogenetics as revealed by mobile element insertions

Ray D. A. [1], Laborde M. [1], Landry K. [1], Xing J. [1], Collier S. [1], Hedges D. J. [1], White B. [1], Batzer M. A. [1]. (1)

Louisiana State University
Using SINEs as phylogenetic markers is promising because the integration of a particular element at a location in the genome is irreversible and essentially random. These attributes make analysis of SINEs as phylogenetic characters an essentially homoplasy-free affair. Alu elements are primate-specific SINEs that make up a large portion of the human genome and are widespread in other primates. Using a combination wet-bench/computational approach we analyzed 200+ Alu insertion loci specific to the genomes of nine New World primates. We used ~150 of these loci to investigate branching order and have produced a cladogram that supports a basal position for the Pitheciidae (Titi and Saki monkeys) and a sister relationship between the Atelidae (Spider monkeys) and the Cebidae (Marmosets and Owl monkeys). These relationships are supported with a homoplasy index of under 0.01. This picture of New World monkey evolution represents the most robust molecular phylogeny to date.

REAL, Leslie

Molecular Ecology of Rabies Virus

Real L. A. [1]. (1) Emory University, Department of Biology
Rabies virus, a single-stranded negative-sense RNA virus, occurs in both epidemic and endemic conditions in different host species across North America and Europe. I will review our labs characterization of the spatial molecular evolution of rabies virus and the spatial population structure of the virus in endemic versus epidemic regions of the United States and Canada. Earlier suggestions that rabies virus may show local ecological variants must be replaced by simpler spatial hypotheses consistent with isolation-by-distance processes rather than local adaptation. Epidemic rabies in red foxes clearly indicates isolation-by-distance while endemic raccoon rabies variants show homogeneous spatial distribution. The difference between spatial organization in these two regions may be the joint effect of local movement of pathogens and host-specific transmission dynamics. Application of this kind of spatial analysis may be essential in identifying origins or testing hypotheses on disease emergence, e.g. the current Ebola outbreak in West Africa among lowland gorillas.

REYNOLDS, Noah

Evolutionary history of antifreeze glycoproteins in Gadidae inferred from molecular phylogenetic analysis.

Reynolds N. M. [1], Cheng CH. C. [1]. (1) University of Illinois
Gadidae (cod fishes) is a family of Gadiformes, and consists of a diverse assemblage of about 50 species classified into 3 subfamilies. Many members of the Gadidae family have evolved antifreeze glycoproteins (AFGP) as an adaptation to polar freezing conditions. To gain insight into the evolutionary mechanism and history of the AFGP gene in the gadid lineage, a well resolved phylogeny is needed. Phylogenetic relationships within the family Gadidae were investigated using the complete coding sequence of the mitochondrial genes NADH dehydrogenase II and cytochrome oxidase I, employing maximum parsimony, maximum likelihood, Bayesian and distance methods. Phylogenetic reconstruction shows the presence of AFGPs in two distinct clades, suggesting that AFGP may have separately evolved twice within the family. A continued effort is underway to produce a large data set from a much larger sampling of gadoid taxa to gain a reliable inference into the evolution of AFGPs in the gadoids.

RICHARDSON, Aaron O.

Horizontal transfer of a *cis*-spliced group II intron from a eudicot to a monocot, *Hedychium coronatum*

Richardson A. O. [1], Qiu Y. L. [2], Palmer J. D. [1]. (1) Indiana University, (2) University of Michigan
Qiu and Palmer (JME, in press) examined *cis*- and *trans*-splicing of the mitochondrial group II intron nad1.i728 in 439 land plants, and inferred 15 shifts from *cis* to *trans*-splicing and two apparent transitions from *trans*- to *cis*-splicing during the course of land plant evolution. The two reversions to *cis*-splicing are in *Pontederia cordata* (Pontederiaceae) and *Hedychium coronarium* (Zingiberaceae), two unrelated monocots nested deep within a clade of *trans*-splicing eumonocots. Through sequencing and subsequent

phylogenetic analysis of nad1.i728 (and its encoded matR gene), we show that the Pontederia nad1.i728 is probably *trans*-spliced, with a mitochondrial rearrangement most likely responsible for the *cis*-spliced Southern Blot pattern observed. *Hedychium*, however, appears to have acquired a *cis*-spliced nad1.i728 from a eudicot donor.

RICKER, Jill

Strong evidence for extensive regulatory subfunctionalization driving the maintenance of duplicated genes in regulatory gene families of *Arabidopsis thaliana*

Ricker J. M. [1], Cui L. [1], Zhang Q. [1], Zhang X. [1], Wall K. [1], Altman N. [1], Ma H. [1], Leebens-Mack J. [1], dePamphilis C. [1]. (1) The Pennsylvania State University

Gene duplication is highly prevalent in plant nuclear genomes, and plays an important role by providing raw material for molecular evolution. Gene duplicates have a variety of fates - nonfunctionalization, subfunctionalization, and neofunctionalization - but the relative importance of each fate is uncertain. We identified 151 clusters of paralogous genes from a data set of 1700 *Arabidopsis thaliana* regulatory genes using single linkage clustering based on a blastp search. ANOVA was used to identify gene, tissue, and gene by tissue effects using whole-genome microarray data from 6 wild-type tissues. Approximately 90% of paralogs have statistically significant gene by tissue effects, which is indicative of subfunctionalization and/or neofunctionalization. We conclude that regulatory subfunctionalization is the predominant pathway for maintained duplicated regulatory genes in angiosperms and there should be a general trend of increasing specificity through each round of duplication.

RIESEBERG, Loren

Hybridization and the evolution of phenotypic diversity in sunflowers

Rieseberg L. H. [1]. (1) Biology Department, Indiana University

Hybridization is frequent in many organismal groups, but its role in adaptation is poorly understood. In sunflowers, species found in the most extreme habitats are ancient hybrids, and new gene combinations generated by hybridization are speculated to have contributed to ecological divergence. This possibility was tested through phenotypic and genomic comparisons of ancient and synthetic hybrids. Most trait differences in ancient hybrids could be re-created by complementary gene action in synthetic hybrids and were favored by selection. The same combinations of parental chromosomal segments required to generate extreme phenotypes in synthetic hybrids also occurred in ancient hybrids. Thus, hybridization facilitated ecological divergence in sunflowers.

RIFKIN, Scott

The neutral rate of genome-wide gene expression evolution in *Drosophila melanogaster*

Rifkin S. A. [1], Houle D. [2], White K. P. [1], Kim J. [3]. (1) Yale University, (2) Florida State University, (3) University of Pennsylvania

Using whole-genome microarrays and a set of mutation accumulation lines, we measured the mutational heritability of genome-wide gene expression at two times during the onset of *Drosophila melanogaster* metamorphosis. Mutational heritability is an estimate of the rate of evolution of a quantitative character solely due to mutation and drift and can be used to calibrate tests for selection. While estimates of mutational heritability for gene expression differ between genes, they are broadly consistent with estimates for other characters in Drosophila. We discuss the relationships between gene function and neutral rates of gene expression evolution, between sensitivities to genetic and environmental perturbation, and between genetic architecture and evolvability of gene expression. We also use this data to investigate the evolutionary forces shaping species differences in gene expression at the start of metamorphosis.

RISPE, Claude

Comparative study of the evolution of codon usage in endosymbiotic bacteria (hosted by insects) and

their free-living relatives.

Rispe C. [1], Shaber J. [2], Moya A. [2], Delmotte F. [1]. (1) Institut National de la Recherche Agronomique, (2) Universitat de Valencia, Spain

We have studied the body of data constituted by 5 recently sequenced genomes of endosymbiotic bacteria inhabiting insect cells (from aphids, tse-tse flies, and carpenter ants). We present results on the evolution of codon usage in these species, which is compared to that in free-living relatives (e.g. *Escherichia coli*). We found that adaptive codon bias almost disappeared in all endosymbionts, due to heavy mutational bias. However, we found strong correlations between codon bias in *E. coli* genes and different characteristics of their homologs in symbiotic bacteria: propensity to be lost (since the last common ancestor, which has been reconstructed), evolutionary rates. Finally, we use this data to weigh the possibility that codon bias in free-living bacteria is influenced by selection for accuracy (and not just rapidity) of the process of translation. This is done by comparing the level of evolutionary constraint on sites using optimal versus sub-optimal codons.

RISSLER, Heather

High-light stress responses in *Euglena gracilis*: evolution of photoprotective and light-harvesting strategies in secondarily derived plastids

Rissler H. M. [1], Durnford D. G. [1]. (1) University of New Brunswick

The acquisition of plastids by primary and secondary endosymbiosis and subsequent transfer of genes from the endosymbiont to host genome introduced unique challenges in that photosynthetic eukaryotes must coordinate the expression of separate genomes in order to effectively balance light harvesting and energy dissipation. As part of the Protist EST Program (PEP), we have characterized EST libraries from low-light acclimated and high-light stressed cultures of *Euglena gracilis*. Comparisons of EST distribution between libraries have revealed putative novel photoprotective proteins, while both bioinformatics and biochemical analyses indicate that high-light stress responses in *E. gracilis* are subject to predominantly post-transcriptional regulatory mechanisms. Cross-species comparisons of light stress responses have identified trends in the evolution of photoprotective and light harvesting strategies in photosynthetic eukaryotes. Using a bioinformatics approach, we are investigating how several biochemical challenges were met following secondary endosymbiosis, including the mechanism of protein import into complex plastids and the evolution of heme and chlorophyll biosynthetic pathways. We are further utilizing EST sequence data to examine the evolutionary history of *E. gracilis* and secondarily derived plastids.

ROCA, Alfred

Mesozoic origin for West Indian insectivores

Roca A. L. [1], Bar-Gal G. K. [2], Eizirik E. [2], Helgen K. M. [3], Maria R. [4], Springer M. S. [5], O'Brien S. J. [2], Murphy W. J. [6]. (1) SAIC-Frederick and National Cancer Institute, (2) National Cancer Institute, (3) University of Adelaide, (4) Parque Zoologico Nacional Dominican Republic, (5) University of California Riverside, (6) SAIC-Frederick

Highly endangered solenodons, endemic to Cuba (*Solenodon cubanus*) and Hispaniola (*S. paradoxus*), comprise the only two surviving species of West Indian insectivores. Combined gene sequences (13.9 kb) from *S. paradoxus* established that solenodons diverged from other eulipotyphlan insectivores 76 million years ago (Mya) in the Cretaceous. Though not incompatible with origins via dispersal, the divergence date is consistent with a mainland-Antillean vicariant origin in the Mesozoic for the solenodon lineage, and with potential proto-Antillean survival following the Chicxulub catastrophic bolide impact in the Caribbean. Sequence of 1.6 kb of mtDNA from *S. cubanus* indicated a deep family-level divergence of 25 million years versus the congeneric *S. paradoxus*, consistent with vicariant origins as tectonic forces separated Cuba and Hispaniola. Efforts to prevent extinction of the two surviving solenodon species would conserve an entire lineage as old or older than many mammalian orders.

ROELANTS, Kim

Multi-gene evidence for Pangaean diversification of crown-group frogs

Roelants K. [1], Bossuyt F. [1]. (1) Vrije Universiteit Brussel

DNA sequence analyses have repeatedly challenged morphological evidence for a paraphyletic assemblage of archaeobatrachian lineages at the base of the anuran tree. We explored this controversy by phylogenetic analysis of nearly 4,000 bp derived from six mitochondrial and three nuclear genes. Our results reject the reciprocal monophyly of Archaeobatrachia and Neobatrachia, and instead support a basal divergence of amphicoelous frogs (Leiopelmatidae and Ascaphidae). Subsequent divergence age estimates document the rise of extant frog lineages in the Triassic, approximately 55 million years prior to their appearance in the fossil record. Their existence on Pangaea gains support from the observation that the timing of three splits between Laurasia- and Gondwana-associated families coincides with the initial rifting of both landmasses. These findings constitute the first evidence that at least seven major crown-group frog lineages were already established by the Early Jurassic.

ROGOZIN, Igor

Evolution of splice signals and protosplice sites

Rogozin I. B. [1], Sverdlov A. V. [1], Babenko V. N. [1], Koonin E. V. [1]. (1) NCBI/NLM/NIH

A comparison of the nucleotide sequences around the splice junctions that flank old (shared by two or more major lineages of eukaryotes) and new (lineage-specific) introns in eukaryotic genes reveal substantial differences in the distribution of information between introns and exons. Old introns have a lower information content in the exon regions adjacent to the splice sites than new introns, but have a corresponding higher information content in the intron itself. This suggests that introns insert into non-random (proto-splice) sites but, during the evolution of an intron after insertion, the splice signal shifts from the flanking exon regions to the ends of the intron itself. Accumulation of information inside the intron during evolution suggests that new introns largely emerge de novo, rather than through propagation and migration of old introns. The information obtained from this analysis was further employed for reconstruction of ancient protosplice sites.

ROKAS, Antonis

More sequence or more taxa? The effect of taxon sampling on genome-scale datasets

Rokas A. [1], Carroll S. B. [1]. (1) University of Wisconsin-Madison

A major question in phylogenetics is the relative contribution of taxon sampling and sequence data set size to confidence in phylogenetic inference. The strategy favored by most phylogeneticists is increasing the number of taxa. This approach has gained support by simulation studies which have demonstrated that this is crucial in phylogenies containing long branches. To examine the trade-offs involved in increasing the number of taxa while keeping the amount of sequence data stable and vice versa, we have taken advantage of published genomic data available from 17 closely related yeast species. We have devised multiple tests to evaluate the relative merits of each approach under a variety of conditions using biological sequence data on a genomic scale. Our data argue for a very significant effect of data set size on the confidence of the inferred phylogeny.

ROONEY, Alejandro P.

Amino Acid-Rich Selection on Protein-Coding Genes

Rooney A. P. [1]. (1) USDA-ARS

The neutral model of molecular evolution predicts that amino acid composition and nucleotide composition are driven by the underlying genomic GC content, as a result of mutation bias. In this study, analyses of several different kinds of proteins show that they undergo an unusual form of selection known as amino acid-rich selection, in which the frequencies of adaptively important amino acids are maintained at relatively high levels. This form of selection is unusual because it represents a form of

purifying selection that does not act on the specific sequence of amino acids but on their overall frequencies instead.

ROSENBERG, Michael

Fidelity of Sequence Alignment Procedure in Establishing Site Homology in Comparative Genomics

Rosenberg M. S. [1]. (1) Arizona State University

Sequence alignment is the process of identifying positions between two homologous DNA sequences that have descended from a common ancestor. Success of comparative genomic investigations depends on the accuracy of the predicted site homologies. I find that the proportion of truly homologous positions aligned correctly declines rapidly towards zero once the identity between sequences falls below 65%. Sequences with greater than 50% site divergence cannot be distinguished from random, as the alignment procedures artificially increase identity beyond this threshold. The addition of intermediate sequences in a multiple alignment can improve the accuracy of homology identification, depending on the distance among the original sequences and the phylogenetic positions of the new sequences. These results indicate that alignments of genomic regions for species much more divergent than human and mouse are likely to contain a large fraction of spuriously homologous positions.

ROSSITER, Wayne

PLA2 venom loci show evidence for positive selection in Sistrurus rattlesnakes with different diets

Rossiter W. D. [1], Gibbs H. L. [1]. (1) The Ohio State University, Dept. of Evolution Ecology

Previous studies suggest that genes coding for snake venom proteins are under strong diversifying selection at the molecular level. However, there have been few attempts to examine evolutionary changes in specific loci in closely related species with different diets. Here, we characterize the evolution of PLA2 venom loci in the eastern massasauaga rattlesnake (*Sistrurus catenatus*) and the pigmy rattlesnake (*S. miliarus*), two congeneric species that specialize on different prey. We identified four putative loci in *S. catenatus* and two loci in *S. miliaris*. Analyses of patterns of substitution in the coding regions of these loci revealed Ka/Ks ratios > 1.3 for three of eight possible interspecific comparsions providing evidence for positive selection at the molecular level at some of these loci. Because venom proteins play a key role in allowing snakes to utilize different prey this system offers the potential for understanding the genetic basis of adaptive variation at the molecular level.

ROTHENBURG, Stefan

Identification of a new Z-DNA binding protein (PKZ) from zebrafish - new insights into the evolution of Z-DNA binding proteins

Rothenburg S. [1], Deigendesch N. [2], Koch-Nolte F. [2], Haag F. [2], Schwartz T. [3], Lowenhaupt K. [4], Rich A. [4]. (1) Institut für Immunologie, Universitätsklinikum Ham, (2) Inst. f. Immunologie, Universitätsklinikum Hamburg, (3) Laboratory of Cell Biology, The Rockefeller University, (4) Department of Biology, Massachusetts Institute of Technology

The Za motif, which binds tightly and specifically to left-handed Z-DNA, is found in a family of proteins, comprising the RNA editing enzyme ADAR1 (found in vertebrates), Z-DNA Binding Protein1 (ZBP1) (so far found only in mammals) and the poxvirus virulence factor, E3L. We cloned a new member of this family termed Protein Kinase containing Z-DNA binding domains (PKZ) from Zebrafish. Two regions with homology to Za are present at the N-terminus. While the overall sequence identity of the Z-DNA binding domains is only about 20%, critical amino acids that contact Z-DNA or make up the hydrophobic core are highly conserved. The C-terminal part of PKZ shows homology to the double-stranded RNA-dependent protein kinase (PKR). Five different splice variants have been identified including variants in which the N-terminal Za domains or the C-terminal kinase domain are missing. The ability of E3L to bind Z-DNA is essential for viral virulence, and the Z-DNA binding protein, E3L, from poxviruses is an antagonist of mammalian PKR. The identification of a gene combining Za domains with a PKR-like

domain further strengthens the hypothesis that Z-DNA binding proteins are involved in the host response to viruses and that poxviruses evolved E3L to prevent this response.

SAITOU, Naruya
Human and hominoid specific changes and conservation during primate evolution
Saitou N. [1], Kitano T. [1], Kim C. [1], OOta S. [1], Sumiyama K. [1], Ezawa K. [1]. (1) National Institute of Genetics
Any genetic changes occurred in one particular evolutionary lineage can be detected if we compare that genome with those of phylogenetically closely related species. Great apes (chimpanzee, bonobo, gorilla, and orangutan) are the best source of comparison for detecting human specific changes. We thus compared human and great ape 103 protein-coding sequences to decipher human specific changes after the human lineage diverged from the common ancestor of human/chimpanzee. We were also able to detect hominoid specific conserved regions as well as human specific changes through comparing the Hox A gene cluster. This "simultaneous multigenome comparison" technique should be the model for comparative genomic studies.

SALCEDO, Tovah
Nucleotide variation on the X chromosome in a natural population of house mice, *Mus domesticus*
Salcedo T. [1], Nachman M. W. [1]. (1) University of Arizona
Patterns of DNA sequence variation have been well-documented in populations of both humans and *Drosophila melanogaster*, but have not been characterized in natural populations of *Mus*. Laboratory strains of mice serve as models for many complex traits in humans, but the relationship of genetic variation among laboratory strains to genetic variation in the wild is poorly understood. Here, we describe patterns of nucleotide variation at two X-linked loci (*Dmd* and *Amelx*) among 60 wild-caught *Mus domesticus* and eight commonly-used inbred strains. The Duchenne muscular dystrophy gene (*Dmd*) spans 2.4 Mb of the X chromosome in humans and mice, and is located approximately in the center of the mouse X chromosome. The amelogenin gene (*Amelx*) spans 5.3 kb, and is located at the telomeric end of the mouse X chromosome. In mice, *Dmd* experiences a low level of recombination, whereas *Amelx* experiences a high level of recombination. In general, levels of nucleotide variability and patterns of linkage disequilibrium observed at these genes are more similar to patterns seen in humans than in flies, consistent with the presumed smaller effective population size for mice and humans compared to *D. melanogaster*.

SALZBURGER, Walter
Adaptive evolution and color genes in cichlid fishes
Salzburger W. [1], Braasch I. [1], Meyer A. [1]. (1) Department of Biology, University Konstanz
The adaptive radiations of cichlid fishes in East Africa are well known for their spectacular diversity and their astonishingly fast rates of speciation. We present a molecular phylogeny and life-history trait reconstruction of East African haplochromine cichlids, the most species-rich cichlid lineage that includes the entire assemblages of lakes Victoria and Malawi. We show that in the ancestor of the haplochromines key-innovations in behavior (maternal mouthbrooding), morphology (egg-spots) and a sexually selected trait (color polymorphism) arose, that might be responsible for the haplochromines' propensity for explosive speciation. Focusing on the cichlid's diversity in coloration, we also determined DNA sequences of putative color-genes and analyzed their molecular evolution.

SALZMAN, Yael
Adaptation by gene disruption in D.melanogaster: gene loss leads to pesticide resistance
Salzman Y. T. [1], Davis J. C. [1], Petrov D. A. [1]. (1) Stanford University
Many transposable elements (TEs) in D. melanogaster are likely to be deleterious due to the induction of ectopic recombination among dispersed TE copies. Such TEs are unlikely to reach high population frequencies unless they generate independent adaptive effects. In a screen of TEs present in the sequenced D. melanogaster genome we have identified and studied one such unusually frequent Doc

element insertion. This Doc insertion appears to severely disrupt or even knock out a gene coding for a putative choline esterase gene. We present population genetic data supporting the idea that the Doc insertion has been advantageous and phenotypic assay data suggesting that this mutation confers pesticide resistance. We will also describe the evolution of this gene and the gene family that contains it. The evolutionary patterns are suggestive of the studied gene undergoing adaptive evolution even prior to the insertion of the Doc element.

SAMPEDRO JIMENEZ, Javier

Microsynteny as a tool to study the comparative evolution of the expansin superfamily in Arabidopsis and rice

Sampedro Jimenez J. [1], Cosgrove D. J. [1], dePamphilis C. W. [1], Lee Y. [2], Kende H. [2]. (1) Penn State University, (2) Michigan State University

Expansins are wall-loosening proteins that function in plant growth and other processes. We used the rice and Arabidopsis genomes to study how the expansin superfamily has evolved since the last common ancestor. Arabidopsis has 36 expansin genes in 4 families while rice has 56 genes in the same 4 families. Phylogenetic analysis shows most Arabidopsis and rice genes on separate branches, but it is unable to resolve the early branching pattern. Interspecies microsynteny was detected between 22 Arabidopsis genes and 37 rice ones. Combining both methods we traced the 92 genes from these species back to a minimum of 16 ancestral genes in the common ancestor. For Arabidopsis we reconstructed the growth of the superfamily under the assumption of 3 whole genome duplications. The result was an estimated 64 gene births and 44 gene deaths since the divergence from rice. Tandem duplications on the other hand have been more important for gene births in rice (18 cases) than in Arabidopsis (8).

SANCHEZ, Maribel

Identification of Cell Cycle Control Proteins in the Genome of Malaria Parasite *Plasmodium falciparum*

Sanchez M. [1], Wang X. [2], Erickson B. K. [1], Wang Y. [1]. (1) University of Texas at San Antonio, (2) Astrazeneca Pharmaceuticals

Plasmodium falciparum, the causative agent of human malaria, has a dynamic life cycle in the mosquito vector and human hosts. During the intraerythrocytic stage, the parasite undergoes complex and atypical cell cycles. Cell cycle control proteins that are essential in parasite biology, but distantly related to host proteins, are believed to be potential therapeutic targets. Based on the hypothesis that the cell cycle control machinery in *P. falciparum* resembles, at some level, its counterpart in a variety of eukaryotes, we identified a series of putative cell cycle control proteins, using the well-characterized yeast cell cycle system as a template. The core components in this derived draft of parasite machinery include cyclin, cyclin-dependent kinases, transcription factors, and DNA repair proteins, which may play roles in a cascade of signal transduction pathways. The "missing links" in the cell cycle regulation map may represent parasite lineage-specific regulators.

SANTINI, Simona

Organization and nucleotide composition of Tilapia *Hox* genes: implications for the evolution of *Hox* clusters in fish.

Santini S. [1], Bernardi G. [1]. (1) Stazione Zoologica

Hox genes encode DNA binding proteins that specify cell fate in the anterior-posterior axis of metazoan animal embryos. While the *Hox* genes content in mammalian genomes is homogeneous, i.e. each cluster contains the same genes among the different mammalian species, this does not happen in ray-finned fish, in which both the number and organization of *Hox* genes and even the number of *Hox* clusters are variable. Ray-finned fish are believed to have undergone a further genome duplication that led to the presence of 8 *Hox* clusters (four twin pairs) in their ancestor. Here we describe the Tilapia (*Oreochromis niloticus*) *Hox* genes set in terms of gene content, clusters organization and nucleotide composition and

compare it with those of pufferfish and zebrafish. We observed that in all these fish, when paralogous genes are conserved in both the twin clusters, the gene which i) belongs to the less gene-rich (less conserved) cluster or ii) is a pseudogene or iii) has a reduced field of embryonic expression, has a lower G-C content. The relationship between the decrease of G-C content and the loss of conservation and function of one of the paralogous genes from twin clusters is discussed.

SATTA, Yoko

Out of Africa with a human specific pseudogene

Hayakawa T. [1], Aki I. [2], Takahata N. [2], **Satta Y.** [2]. (1) Departments of Medicine and Cellular and Molecular Medicine, University of California, San Diego, (2) Dept. of Biosystems Science., Graduate University for Advanced Studies (Sokendai)

Sialic acids may play important roles in cell-cell communication or pathogen infection. There are two types of sialic acids, Neu5Gc and Neu5Ac, in mammals. Exceptionally humans are deficient in Neu5Gc because of loss of gene (*CMAH*) function that encodes the enzyme converting Neu5Ac to Neu5Gc. This loss of function resulted from an exon deletion which took place ~3.2 million years (Myr) ago. We sequenced a ~7.5 kb *CMAH* intron region that encompasses the deleted exon for 130 chromosomes (24 African and 106 non-African) with different ethnic origins. The exon deletion is shared by all chromosomes examined. The genealogical analysis revealed ~3 Myr age of the most recent common ancestor and several African specific lineages. Based on these observations, we discuss the process of spreading the *CMAH* pseudogene from Africa to the rest of the world and the possible effect of natural selection involved.

SAUNDERS, Matthew

G6PD Deficiency, a High Frequency Enzymeopathy in Kurdish Jews: Effects of Demography or Natural Selection by Malaria?

Saunders M. A. [1], Kim J. [1], Hammer M. F. [1], Nachman M. W. [1]. (1) University of Arizona

An allele of glucose-6-phosphate dehydrogenase (G6PD) commonly found in the Middle East, *G6PDmed*, is associated with a severe clinical phenotype, yet the allele is under positive natural selection by malaria in some human populations. *G6PDmed* is found in the Kurdish Jews at a frequency of 0.70, the highest known frequency of G6PD deficiency in humans. Although this population may have experienced a bottleneck, resulting in an extreme frequency of *G6PDmed*, a study of unlinked nucleotide data from *Beta-globin* does not show evidence of a strong founder effect. To understand the relative contributions of demography vs. positive selection in shaping nucleotide variability at *G6PD* in Kurdish Jews, we resequenced *G6PD*, four windows each of >3 kb spanning 1.8 Mb around *G6PD*, as well as 2 unlinked loci in a sample of 40 individuals of this population. Patterns of linkage disequilibrium and nucleotide diversity suggest a substantial contribution of natural selection to the high frequency of *G6PDmed* in Kurdish Jews.

SAWAI, Hiromi

The tempo and mode of *HLA* evolution mediated through the dual function

Sawai H. [1], Takahata N. [1], Satta Y. [1]. (1) Dept. of Biosystems Science, Graduate University for Advanced Studies

In the human genome, there are six functional *HLA* class I genes (*HLA-A*, *-B*, *-C*, *-E*, *-F*, and *-G*) and four pseudogenes (*HLA-H*, *-J*, *-K* and *-L*). In general, *MHC* class I gene family evolves according to the birth-and-death process. Although the gain (duplication) process of *HLA* genes has been well focused, little attention has been paid to the loss process. Here we date the pseudogenization of *HLA-H*, *-J*, *-K* and *-L* based upon the nucleotide sequence data. It is estimated that the *HLA-H*, *-J*, *-K* and *-L* loci were inactivated 6.6, 14.7, 36.6 and 46.6 million years (myr) ago, respectively. Although the standard errors are relatively large, examination of the orthologous sequences in *Macaca mulatta* supports these datings of *HLA* pseudogenizations. It is argued that there have been six to eight functional class I genes in the *HLA*

region at any time over the past 50 myr and that this constancy results from dual functions of MHC molecules.

SCALLY, Mark

Phylogenetic comparison of mammalian BRCA1 sequences and the relevance to human disease.

Scally M. [1], Burk-Herrick A. [1], Amrine-Madsen H. [2], Stanhope M. J. [2], Springer M. S. [1]. (1) University of California Riverside, (2) GlaxoSmithKline

The Breast and Ovarian Cancer Susceptibility gene (BRCA1) has been implicated in approximately 40% of hereditary breast cancer cases. While mutations in the 5' and 3' regions are known to cause cancer, little is known about the effect of mutations elsewhere in the gene. The central exon 11 comprises 60% of the coding region of the gene and features 2 nuclear localization signals and other purported binding regions. The homologous sequences of exon 11 for 132 mammals, including additional primate and marsupial representatives are reported here in an effort to further understand the evolution of the BRCA1 gene in mammals. An evolutionary comparison is made to assess potential oncogenic risk at sites where missense mutations have been reported in humans. Possible selection differences among separate phylogenetic lineages of mammals are also examined in an effort to strengthen the inferences made towards understanding human disease causing factors.

SCHLENKE, Todd

QTL analysis of DDT and malathion resistance in Drosophila melanogaster.

Schlenke T. A. [1], Clark A. G. [1]. (1) Cornell University

Drosophila has been exploited as a model system for understanding the molecular basis of insecticide resistance, but successes have come from treating resistance as a dichotomous trait caused by single mutations of major effect. Using a set of 135 recombinant inbred D. melanogaster lines that were previously genotyped for 152 markers across the genome, we sought to identify quantitative trait loci (QTL) responsible for genetic variation in DDT and malathion resistance. There were significant differences among the lines in both DDT and malathion resistance, and significant QTL were located on chromosome 2 for DDT resistance, and on chromosomes 2 and 3 for malathion resistance. Although there was a weak correlation between DDT and malathion resistance, no significant QTL were shared across the two treatments. Interestingly, a QTL for DDT resistance was identified at cytological position 48, where two gene thought to confer resistance to DDT (Cyp6g1 and Cyp12d1) are located. Although previous work has shown that insertion of an Accord transposable element in the 5' regulatory region of Cyp6g1 is strongly associated with increased resistance to DDT, this insertion was not segregating in the lines tested.

SCHMIDT, Deena

Adaptive evolution drives the diversification of zinc-finger binding domains.

Schmidt D. [1], Durrett R. [1]. (1) Cornell University

The human genome is estimated to contain 700 zinc-finger genes, which perform many key functions including regulating transcription. The dramatic increase in the number of these genes as we move from yeast to *C. elegans* to Drosophila and to humans, as well as the clustered organization of these genes in humans suggests that duplication has played an important role in diversifying this gene family. Using likelihood methods developed by Yang and his coauthors, we have investigated four gene clusters and found signs of positive selection. Due to the limitations of time we will discuss results only for a cluster of genes on the p arm of chromosome 19 near the centromere, which are closely related but have no identified orthologs in mouse.

SCHUENZEL, Erin

Identification of shifts in natural selection and indications of host adaptation in the plant pathogen *Xylella fastidiosa*

Schuenzel E. L. [1], Nunney L. [1]. (1) University of California, Riverside
The bacterium *Xylella fastidiosa* has a number of genetically distinct host-plant strains, of which two have been completely sequenced. Using data from these genomes and two draft genomes, we are investigating the genetic differences between these host strains to identify the genes responsible for host adaptation, by identifying the signature of adaptive evolution. We propose a "top-down" log-linear analysis based on a maximum likelihood estimation of the number of nonsynonymous and synonymous substitutions in pairs of lineages. The test is applied to pairs of lineages joined at progressively deeper nodes. For each pair, the analysis simultaneously tests three null hypotheses, $w = 1$, $w1 = w2$, where $w = KA/KS$, and $K1 = K2$ (the relative rate test), where Ki is the combined (nonsynonymous and synonymous) substitution rate. These test, respectively, for adaptive evolution, a shift in natural selection, and unequal evolution rate. We used this method to detect the signature of natural selection for all open reading frames that occur in the four *X. fastidiosa* genomes.

SCHWARZ, Dietmar

Ecologically mediated hybrid speciation in a diploid, bisexual animal

Schwarz D. [1], McPheron B. A. [1]. (1) The Pennsylvania State University
Hybridization as a process in animal speciation has long been regarded as irrelevant. No plausible mechanism was available by which diploid, bisexually reproducing hybrids would become reproductively and ecologically isolated from their parents. Here we provide evidence for the hybrid origin of the "*Lonicera* Fly" (LF), a new population within the *Rhagoletis pomonella* (Diptera: Tephritidae) species complex found on introduced honeysuckle, *Lonicera* spp. Based on eight nuclear markers and mtDNA sequences we show that the LF formed by hybridization between the sympatric *R. mendax* and *R. zephyria*. The two parents are host-specific fruit parasites and hybrids have only been collected on honeysuckle. The analysis of multilocus genotypes shows that mating on honeysuckle is random, indicating some degree of LF isolation. Host race formation is a plausible mechanism for the reproductive isolation of the LF and theory suggests that hybridization relaxes the conditions for sympatric speciation.

SEO, Tae-Kun

The changing pattern of absolute rates of synonymous and nonsynonymous substitution during mammalian evolution: Analysis of mitochondrial protein-coding genes

Seo T. [1], Kishino H. [2], Thorne J. L. [3]. (1) North Carolina State University, USA, (2) University of Tokyo, Japan, (3) North Carolina State University
There are many possible sources of variation in evolutionary rates among lineages. Biological sources of rate variation may affect synonymous and nonsynonymous rates differently. There are conventional approaches that distinguish between synonymous and nonsynonymous substitution, but these can estimate only amounts of substitution and they thereby confound time and rate. To disentangle rates and times, fossil evidence or other information external to sequence data is needed. In this talk, we discuss our Bayesian method for estimating divergence times and investigating patterns of synonymous and nonsynonymous rate change. We focus on the evolution of mammalian mitochondrial protein-coding genes and the distinct patterns of synonymous and nonsynonymous rate change.

SHEDLOCK, Andrew M.

Reptile genome structure inferred from BAC-end sequence analysis of the American Alligator and Painted Turtle.

Shedlock A. M. [1], Zhao S. [2], Shetty J. [2], Edwards S. V. [1]. (1) Harvard University Dept. OEB, (2) The Institute for Genomic Research
Comparative genomic information on non-avian reptiles is presently lacking in the literature despite practical interest in making large-scale comparisons between mammals and other model vertebrates such as the chicken and pufferfish. To infer general structural features of reptile genomes and thereby reduce this knowledge gap, we analyzed paired sequence reads for ~3800 clones from BAC libraries of the

American alligator (*Alligator mississippiensis*) and painted turtle (*Chrysemys picta*). No significant contamination from *E. coli* or vectors was apparent. We report results from a batch analysis of clone sequences characterizing the density and distribution of repetitive elements, exons and introns, and DNA strings observed and discuss prospects for using large-scale reptilian BAC-end sequence data for phyloinformatics and reconstruction of ancestral states in amniote genome evolution.

SHIU, Shin-Han
Global Patterns of Gene Family Expansion in the Human Lineage After Its Divergence from Mouse
Shiu S. H. [1], Pan R. [1], Byrnes J. K. [1], Tzeng Y. H. [1], Rosner M. R. [1], Li W. H. [1]. (1) University of Chicago
To understand the processes of duplicates retention at the whole-genome level, we analyzed human and mouse families and defined the ancestral family size and gene gains/losses subsequent to the divergence of these mammals. Interestingly, the ancestral family size is positively correlated with the number of gene gained. However, the extent of gene gain in human is not correlated with that in mouse. In addition, we found that synonymous (Ks) and non-synonymous (Ka) substitution rates are elevated in expanded clades with multiple human and one mouse gene as compared to unexpanded clades with 1:1 relationship between these mammals. The mean Ka/Ks value of expanded clades is also significantly higher than that of the unexpanded, indicating that expanded clades experienced not only higher mutation rates but also periods of relaxation of selection or positive selection. Interestingly, expansion is evident in genes involved in defense responses, responses to chemical substances and pathogens, cell-cell adhesion, and olfaction. In contrast, genes involved in intracellular processes such as transport, signaling networks, and transcriptional regulations in general do not reside in expanded clades. These findings suggest functional bias in duplicates retained and we hypothesize that some of these categories with biased representation may be similar among organisms.

SILVA, Joana
Recent Expansion of Transposable Elements Leads to Dramatic Increase in Genome Size
Silva J. C. [1], Bidwell S. J. [1], Carlton J. M. [1]. (1) TIGR
Parabasalids are part of a group of taxa that diverged early in the evolution of eukaryotes. Their main representative, the protist *Trichomonas vaginalis*, is a human extracellular obligate parasite. Sequence data generated by the *T. vaginalis* Genome Sequencing Project suggest a highly repetitive genome of over 100 Mb, an unusually large size for a parasitic protist. We show that the repetitive component is composed of several dozen repeat families, which make up over 1/3 of the genome. These families are characterized by a few hundred copies, with average pairwise differences within families of less than 5%. Preliminary characterization of these repeats reveals the presence of transposable elements, several hundred rDNA genes and many unknown repeats. Our results strongly suggest that the relatively large genome size of *T. vaginalis* is a very recent acquisition, which has resulted from the simultaneous expansion of many repeat families. These results are discussed in the context of the association between the parasite and its human host.

SIMILLION, Cedric
Building genomic profiles for uncovering segmental homology in the twilight zone
Simillion C. A. [1], Vandepoele K. [1], Saeys Y. [1], Van de Peer Y. [1]. (1) Ghent University
The identification of homologous regions within and between genomes is essential when studying genome structure and evolution. Different methods already exist that detect homologous regions automatically. A specific group of these methods is based on identifying chromosomal regions showing conservation of gene order and content. This approach has proven useful for detecting homology between highly divergent chromosomal regions. However, until now, detecting homology using such map-based approaches required significant colinearity between individual segments. We present a novel method that creates profiles combining the gene order and content information of multiple mutually homologous genomic segments. These profiles can be used to detect segments that show significant

colinearity with the entire profile but not necessarily with individual segments. When applying this new method to the genomes of *Arabidopsis* and rice, we find evidence for ancient duplication events in the rice genome.

SIMON, Dawn
Divergent Histories of Group I Introns in Lichen Fungi
Simon D. M. [1], Moline J. [1], Helms G. [2], Friedl T. [2], Bhattacharya D. [1]. (1) University of Iowa, (2) University of Goettingen
The wide, but sporadic distribution of group I introns in protists, plants and fungi, as well as in eubacteria, has been interpreted as evidence for rampant lateral transfer. A realistic estimate of the level of horizontal intron transfers in nature can potentially be obtained by focusing on closely related species or genera. We have chosen one of the most group I intron-rich lineages known, the lichen family Physciaceae. In these analyses, we have used robust phylogenetic methods and a large data set containing 62 novel large subunit (LSU) rRNA group I introns to address the relevance of horizontal transfer within this monophyletic family of lichenized fungi. Interestingly, we find support for five horizontal transfers into homologous sites between species, but do not find strong evidence supporting transposition events into non-homologous sites. This is in contrast to previous work with Physciaceae small subunit (SSU) rRNA group I introns where there is strong support for multiple ectopic transpositions. This may be due to an increased number of positions in the SSU which are suitable for insertion and/or retention of introns. In contrast, the LSU may have comparatively few acceptable positions.

SIMONELIC, Kevin
Fitness consequences of intraspecific hybridization in *Caenorhabditis elegans*
Simonelic K. M. [1], Denver D. R. [1], Lynch M. [1]. (1) Indiana University
Caenorhabditis elegans is a globally distributed, androdioecious nematode species composed of self-fertile hermaphrodites and rare males that mate with hermaphrodites. The nuclear and mitochondrial phylogenies of *C. elegans* natural isolates have nearly identical topologies and reveal the presence of two major, divergent intraspecific clades. This finding suggests a minimal occurrence of intraspecific hybridization in *C. elegans*, plausibly due to reduced hybrid fitness. A comparative assay was performed to directly examine the fitness differences among laboratory-generated intraspecific *C. elegans* hybrids and uncrossed parental controls. Preliminary analyses suggest a fitness reduction in intraspecific hybrids, and may point to the presence of Dobzhansky-Muller interactions or other hybrid breakdown mechanisms acting intraspecifically in *C. elegans*.

SINCLAIR, Colleen
Molecular analysis of the telomere regions in scleractinian coral, *Acropora surculosa*
Sinclair C. S. [1], Richmond R. H. [2], Ostrander G. K. [3]. (1) Towson University, (2) Kewalo Marine Laboratory, (3) Johns Hopkins University
The terminal ends of vertebrate chromosomes are protected by tandem repeats of the (TTAGGG)n sequence. Recent studies have reported identical sequence repeats in a number of aquatic invertebrates including clams, mussels, echinoderms, and snails. Southern blot analysis confirmed the presence of the vertebrate telomere repeats, (TTAGGG)n, in the genome of the scleractinian coral, *Acropora surculosa*. Treatment with *Bal*31 exonuclease revealed progressive shortening of the DNA fragments positive for the (TTAGGG)n sequence, supporting location of the repeats at the chromosome ends. The presence of the vertebrate telomere repeats in corals is further evidence that the (TTAGGG)n sequence is highly conserved among a divergent group of species and suggests a much older ancestory than first proposed.

SINGH, Nadia
Genomic Heterogeneity in Neutral Substitutional Patterns in *Drosophila melanogaster*
Singh N. D. [1], Arndt P. F. [2], Petrov D. A. [1]. (1) Dept. of Biological Sciences, Stanford University, (2) Max

Planck Institute for Molecular Genetics,Berlin
The study of evolutionary forces acting on genes and intergenic sequences requires detailed knowledge of neutral patterns of nucleotide substitution. Here we present the first genome-wide assessment of neutral substitutional patterns across various chromatin and functional domains in the *Drosophila melanogaster* genome. We estimated substitutional patterns using several thousand fragments of DNAREP1, the most common transposable element in the *D. melanogaster* genome. Among other findings, our analyses show that neutral substitutional patterns differ between heterochromatic and euchromatic sequences, possibly due to a recombination-associated mutational bias towards G's and C's in this species. In addition, we confirm that the codon bias differences in regions of low and high recombination are only partially due to variation in neutral patterns of substitution. We will also discuss implications of our results for the studies of selection acting on intergenic and intronic sequences in Drosophila.

SINHA, Neelima

The role of KNOX and PHAN in compound leaf development

Sinha N. R. [1], Champagne C. E. [1], Chung K. H. [1], Goliber T. E. [1], Townsley B. T. [1], Koenig D. P. [1], Gerttula S. M. [1], Schwartz R. D. [1], Uchida N. [1], Bharathan G. [2]. (1) U. C. Davis, (2) SUNY, Stonybrook
The Class I Knotted-like homeobox (KNOX 1) genes are highly expressed in the shoot apical meristem but not expressed in the emerging leaf primordium in tobacco, maize, or Arabidopsis. In tomato, KNOX1 expression (LeT6, TKN1) is seen in the early leaf primordium (Chen et al. 1997; Hareven et al. 1996). We have analyzed compound leaf producing shoot apices in clades with independently derived compound leaves and shown that with one exception (a derived clade in the Fabaceae) compound leaves always show expression of KNOX genes (Bharathan et al., 2002). In the derived pea clade the LFY/FLO gene regulates this function of generating leaf complexity. While KNOX genes appear to be important for generating leaf complexity (except in a derived clade in the Fabaceae) we find that other genes like PHANTASTICA might play a role in determining the form of the compound leaf generated. Transgenic plants overexpressing antisense PHAN suggest that PHAN, by modulating dorsiventrality, has a role in regulating the number of leaflets and their placement in a compound leaf. In Neobeckia aquatica, leaves with different morphologies are produced depending on environmental conditions. The expression differences between these two phenotypic states are also being explored.

SIRVIÖ, Anu

Mapping the genome of ant species *Acromyrmex echinatior*

Sirviö A. M. [1], Pamilo P. [1], Gadau J. [2], Page Jr. R. E. [3]. (1) University of Oulu, (2) University of Wurzburg, (3) University of California Davis
In contrast to humans, genetic model organisms and domesticated plants and animals, few genomic maps exist for insects in spite of the large number of species. Here we present and discuss our work on genomic map sizes of social insects. The honeybee genome has been shown to have an exceptionally large size, caused by a high recombination frequency. We constructed genomic map for leaf-cutter ant *Acromyrmex echinatior* of sub-family Myrmicinae by using 96 males produced by one queen. We used 242 AFLP-markers when constructing map for the species. Here we compare the genomic features of ant species to those of other social insects.

SNOEYENBOS-WEST, Oona

Understanding Genome Complexity and Protein Evolution in Ciliates: insights from Mitochondria and Nuclei.

Snoeyenbos-West O. L. [1], Cardillo J. C. [1], Wilkins E. [1], Lasser E. [1], Katz L. A. [1]. (1) Smith College, Department of Biological Sciences
Ciliates have dual genomes contained within both a somatic macronucleus (MAC) and a germline micronucleus (MIC). Only the MAC is transcriptionally active and it divides by amitosis rather than

mitosis. We are assessing the evolutionary implications of this unusual genome architecture. We are testing the hypothesis that genome duality is responsible for the heterogeneous and unusually elevated rates of amino acid substitutions we find in ciliate nuclear-encoded protein genes. We posit that protein coding genes can explore 'protein space' through mutations in the MIC that are effectively 'hidden' from selection in the processed MAC genome. As a result, rapid divergence among paralogous genes is facilitated by differential selection. To test this hypothesis we are analyzing structural RNA and protein coding genes from both mitochondrial and nuclear genomes in a variety of ciliate species. Data analyzed to date for SSU rRNA genes in ciliate mitochondria indicate that a phylogentic signal is retained as resulting phylogenies are concordant with those obtained from nuclear SSU rRNA genes. This provides a strong framework by which to analyze the unusual mode of protein evolution that occurrs in ciliates, which in turn may help reveal the potential evolutionary costs and benefits of genome duality.

SNOEYENBOS-WEST, Oona

Molecular Phylogeny of Phyllopharyngean Ciliates and their Group I Introns

Snoeyenbos-West O. L. [1], Cole J. [2], Campbell A. [1], Coats W. [3], Katz L. A. [1]. (1) Smith College, Department of Biological Sciences, (2) American Type Culture Collection, (3) Smithsonian Environmental Research Center

We analyze small subunit ribosomal DNA (ssu-rDNA) sequences to evaluate both the monophyly of the ciliate class Phyllopharyngea de Puytorac et al., 1974, and relationships among subclasses. Classifications based on morphology and ultrastructure divide the Phyllopharyngea among four groups, the cyrtophorids, chonotrichs, rhynchodids and Suctoria. Our analyses of ssu-rDNA genealogies derived from sequence data collected from diverse members of the class Phyllopharyngea provide strong support for the monophyly of the Phyllopharyngea, and show that the chonotrichs emerge from within the cyrtophorids. Suctorian budding types are monophyletic, and exogenous budding appears to be basal. Further, we report the discovery of a group I intron at position 891 in the Suctoria *Acineta* sp. and *Tokophrya lemnarum*, and a second group I intron at position 1506 in *T. lemnarum*. These introns represent only the second example of group I introns in a ciliate ribosomal gene, since the first discovery of ribozymes in the LSU rRNA gene of *Tetrahymena thermophila* twenty years ago. Phylogenetic analyses of Group I introns suggest a complex evolutionary history involving either multiple loses or gains of introns within endogenously budding Suctoria.

SONG, Nan

Accurate classification of multidomain protein families

Song N. [1], Davis G. B. [1], Durand D. [1]. (1) Carnegie Mellon University

Identifying a protein's family is important to understanding its function and evolutionary history. Sequence comparison methods are widely used to determine whether two proteins belong to the same family. However, proteins from two different families with complex domain architectures can have significant sequence similarity due to a shared domain, in spite of having distinct evolutionary histories or functional roles. Although additional criteria have been proposed to address this difficulty, the problem of reliable classification of multidomain proteins remains open. We introduce a novel method based on protein neighborhood, the set of proteins with significant matches to a query protein. Our analysis of mouse protein sequences suggests that neighborhood similarity is a stronger indicator of shared family than sequence similarity for multidomain proteins. Our method can successfully identify homologs in protein families with complex domain architectures, such as kinases and trypsins.

SORENSON, Michael

Accelerated mtDNA rate in brood parasitic finches: a case of nearly neutral evolution?

Sorenson M. D. [1], Balakrishnan C. N. [1], Mercer D. M. [1]. (1) Boston University

Two lineages of brood parasitic finches in Africa, the genera Anomalospiza and Vidua, are sister groups that diverged from each other perhaps as long as 20 million years ago and therefore have been obligate

brood parasites for a long time. The social system of the Vidua parasites and the genetic structure of extant species suggests a dynamic history involving repeated cycles of host colonization, speciation, and extinction. If so, this lineage may have an accelerated rate of molecular evolution due to a higher fixation rate for nearly neutral mutations. We tested this hypothesis by sequencing the entire mitochondrial genomes of Vidua chalybeata, Anomalospiza imberbis , two estrildid finches (representing the sister group of parasitic finches), two ploceid finches, and Prunella montanella , a more distant outgroup. Rates of molecular evolution and the ratio of non-synonymous to synonymous substitutions were both higher in the parasitic finch lineage than in estrildid finches.

STAJICH, Jason
A Comparative Study of Fungal Introns
Stajich J. E. [1], Dietrich F. S. [1]. (1) Duke University
The origin and maintenance of introns in Eukaryotes is an important question in the study of genome evolution. We have undertaken a survey of introns within the fungal kingdom to study the history of their positional conservation, gain, or loss within orthologous genes. We have utilized available draft and finished fungal genome sequences to compare how intron size, splice site conservation, intron distribution within coding sequences, total number of introns, and proportion of intron phases have evolved among the Ascomycetes and Basidiomycetes. The yeast *Saccharomyces* cerevisiae has relatively few introns (<300) in its entire genome while the Basidiomycete *Cryptococcus* neoformans averages almost 5 introns per gene and the Euascomycete *Neurospora* crassa averages almost 2 per gene. We have been able to identify a set of positionally conserved introns among major fungal clades and a smaller number conserved throughout the fungi. We will discuss the relationship between the roles introns play in these genomes and their patterns of evolution.

STEEN, Tomoko
Co-evolution of Controversial Theories
Steeen T. Y. [1]. (1) The Library of Congress
Co-evolution means, as you know, the joint evolution of two or more closely interacting species. If one evolves, this change affects the selection pressure operating between the two or more individuals, so the other(s) also evolve(s). Would three controversial theories evolve together just like co-evolution of flowers, bees, and humming birds? In my talk, I will discuss the evolution of the molecular clock hypothesis in this context—its relationship with the neutral theory and nearly neutral theory, two other major controversial theories in 20th century evolutionary biology.

STEFANOVIC, Sasa
Transcompartmental gene duplication in action - the *cox2* case in legumes (Phaseoleae, Fabaceae)
Stefanovic S. [1], Palmer J. D. [2]. (1) Indiana University, Department of Biology, (2) Indiana University
The transfer of mitochondrial genes to nucleus occurred on a massive scale early in evolution, and is therefore of fundamental importance to all eukaryotes, but continues to a significant extent only in plants. Several disparate legume species from tribe Phaseoleae were shown recently to retain intact and expressed *cox2* gene in both compartments. This one known case of transcompartmental gene duplication is being further explored using a two-pronged approach. The "in-width" approach provided additional cases of inactivation, both mitochondrial and nuclear. Inclusion of these and other relevant species in phylogenetic analyses, in conjunction with additional phylogenetic markers, resulted in an improved phylogeny of tribe Phaseoleae. This will increase the statistical power and enable addressing the basic question whether the likelihood of *cox2* inactivation is independent of its compartmental location. In parallel, "in-depth" RT-PCR assays are conducted In those species that retained both copies across a spectrum of different tissues addressing the question about the fixation, redundancy, and persistence of both genes.

STOEBEL, Daniel

Two ways linkage affects the rate of neutral substitution

Stoebel D. M. [1], Last M. S. [2], Brisson D. [1], Dykhuizen D. E. [1]. (1) Stony Brook University, (2) University of California, Davis

The rate of neutral evolution is often used as a baseline against which to examine patterns of evolution caused by positive or negative selection. Substitutions do not occur in a vacuum; they occur in the context of the entire genome. Earlier analyses of the effect of linked selection on neutral mutations showed that the probability of fixation of a neutral mutation by hitchhiking is the same as the probability of fixation by drift, and argued that this meant that the rate of neutral substiution was uneffected by hitchhiking. We do not dispute that the probability of fixation by hitchhiking is the same as that by drift; we argue that the marginal probability of fixation by either drift *or* hitchhiking is not same as the probability of fixation by either force alone. We show that under strong-selection, weak-mutation assumptions, hitchhiking increases the rate of fixation of neutral alleles. Furthermore, if the locus undergoing adaptive evolution can be described by a model of adaptation towards an optimum, we show that the expected rate of neutral substitution will slow over time.

SU, Bing

Accelerated evolution of the PACAP precursor gene during human origin

Wang Y. Q. [1], Qian Y. P. [2], Yang S. [1], Shi H. [1], Liao C. H. [1], Zheng H. K. [3], Wang J. [3], Lin A. [4], Cavalli-Sforza L. [4], Underhill P. [4], Chakraborty R. [2], Jin L. [2], **Su B.** [1]. (1) Kunming Institute of Zoology, CAS, (2) University of Cincinnati, (3) Huada Genomics Institute, CAS, (4) Stanford University

Pituitary adenylate cyclase-activating polypeptide (PACAP) is a neuropeptide that is widely expressed in numerous tissues of vertebrate animals, especially in the central nervous system. PACAP is involved in regulating many important biological processes, such as reproduction, development and neuronal signal transduction. The sequence of PACAP is extremely conserved across vertebrate species, indicating a strong functional constraint during the course of evolution. However, here we show that the PACAP precursor gene underwent an accelerated evolution in the human lineage. The protein substitution rate in humans since they diverged from chimpanzees is at least seven times faster than those of the other mammal species. Eleven human-specific amino acid changes were identified in the 176 amino acid PACAP precursor while they are conserved from murine to African apes. The neutrality test demonstrated that the rapid evolution of human PACAP precursor is due to strong Darwinian positive selection. Protein structure analysis revealed gains and losses of functional motifs in the human PACAP precursor, and a putative novel neuropeptide might have originated during human origin, and contributed to the formation of human mentality.

SUBRAMANIAN, Sankar

Gene Expression Intensity Shapes Evolutionary Rates of the Vertebrate Proteins

Subramanian S. [1], Kumar S. [1]. (1) Arizona State University

Natural selection leaves its footprints on protein coding sequences by modulating their evolutionary rates. In highly expressed genes in invertebrates, these footprints are seen in the higher codon usage bias and lower synonymous divergence. Here we have examined how the rate of protein sequence evolution is modulated by the amount (intensity) of gene expression in the vertebrate genome. To understand how natural selection operates on vertebrate genes created in earlier and later phases of animal evolution, we have contrasted patterns of mouse proteins that have homologs in invertebrate or protist genomes (Precambrian genes) with those that do not (vertebrate specific genes). We find that the intensity of gene expression relates inversely with the rate of protein sequence evolution. The most highly expressed genes have the lowest total number of substitutions per polypeptide, showing the cumulative affects of purifying selection on individual amino acid replacements. Pre-Cambrian genes exhibit a more pronounced difference in protein evolutionary rates (up to 3 times between the genes with high- and low

expression levels) than the vertebrate-specific genes, which appears to be due to a narrower breadth of expression of the vertebrate-specific genes.

SUZUKI, Yoshiyuki

New methods for detecting positive selection at single amino acid sites

Suzuki Y. [1]. (1) National Institute of Genetics

For inferring positive selection at single amino acid sites, Suzuki and Gojobori (1999) developed a method (SG method) for comparing rates of synonymous and nonsynonymous substitutions at each codon site using ancestral codons at interior nodes of phylogenetic tree as inferred by maximum parsimony method. In SG method, however, selective neutrality cannot be tested at sites, where only termination codons are inferred at interior nodes or number of equally parsimonious inferences of ancestral codons exceeds 10,000. Here I present a modified SG method which is free from these problems by using distance-based Bayesian method for inferring the single most likely ancestral codon from 61 sense codons. In computer simulations and real data analyses, modified method showed a higher overall efficiency than original method particularly at highly polymorphic sites, indicating that modified method is useful for inferring positive selection at sites where neutrality cannot be tested by original method. I also discuss that p-distance is preferable to number of synonymous substitutions for inferring phylogenetic tree in SG methods, and present a maximum likelihood method for detecting positive selection at single amino acid sites, which produced reasonable results in real data analyses.

SWIGONOVA, Zuzana

Structure and evolution of the chromosomal region containing the *r/b* gene of rice, maize, and sorghum

Swigonova Z. [1], Messing J. [1]. (1) Waksman Institute, Rutgers University

Sequence comparison of chromosomal regions bearing the *r/b* gene revealed extensive similarity between rice and sorghum that share eight collinear genes and are equally poor in transposable elements. Retrotransposons identified within the duplicated regions of maize, occupying 55% and 76% of the genomic sequences, inserted mostly within the last 2 million years. Five tandem genes shared by rice and sorghum are present within the two homoeologous regions of maize, two are present as homoeologous duplicates. Phylogenetic analysis of *r/b* sequences revealed that the grass ancestral *r* gene amplified in parallel in rice, sorghum, and maize and that the maize *r1* and *b1* genes are allelic descendants of one ancestral gene. Another conserved gene, *cisZOG*, exhibited parallel amplification across lineages and extensive deletion of gene copies in the maize genome. Diploidization of tetraploid maize seems to result in a mosaic retention of at least one copy of the ancestral gene.

TAKAHASHI, Aya

A high frequency null mutant of an odorant-binding protein, *Obp57e*

Takahashi A. [1], Takano-Shimizu T. [1]. (1) Dept. of Population Genetics, NIG, Mishima, Japan

We have found a null mutant of an odorant-binding protein, *Obp57e*, in *Drosophila melanogaster*. This frameshift mutation is at a high frequency in Kyoto population, and is also present in Taiwan and Africa. We have sequenced ~1.5kb region including the tandemly duplicated gene, *Obp57d*, from 16 inbred lines sampled in Kyoto, Japan. The analyses showed a peak nucleotide diversity and a strong linkage disequilibrium around this mutation. This pattern suggests an elevated mutation rate or an influence of balancing selection in this region. The former is not supported from the level of nucleotide divergence between *D. melanogaster* and *D. simulans*. Thus, this presence/absence polymorphism may be due to balancing selection, which took advantage of the relatively weak functional constraint in members of a large gene family. In addition, *Obp57d* gene region showed an excess of high frequency derived mutants that is consistent with a pattern predicted under positive natural selection.

TAKEZAKI, Naoko

The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of 44 nuclear genes

Takezaki N. [1], Figueroa F. [2], Zaleska-Rutczynska Z. [2], Takahata N. [3], Klein J. [4]. (1) National Institute of Genetics, (2) Max-Planck-Institut für Biologie, (3) The Graduate University for Advanced Studies, (4) Department of Biology, The Pennsylvania State University

The origin of tetrapods is a major outstanding issue in vertebrate phylogeny. Each of the three possible principal hypotheses (coelacanth, lungfish, or neither being the sister group of tetrapods) has found support in different sets of data. In an attempt to resolve the controversy, sequences of 44 nuclear genes encoding amino acid residues at 10,404 positions were obtained and analyzed. However, this large set of sequences did not support conclusively one of the three hypotheses. Apparently, the coelacanth, lungfish, and tetrapod lineages diverged within such a short time interval that at this level of analysis their relationships appears to be an irresolvable trichotomy.

TAMURA, Koichiro

Accuracy of extra-large NJ trees with simultaneous estimation of pairwise distances

Tamura K. [1], Nei M. [2], Kumar S. [3]. (1) Tokyo Metropolitan University, (2) Pennsylvania State University, (3) Arizona State University

Current efforts to reconstruct the tree of life and histories of multigene families demand the inference of very large phylogenies consisting of thousands of sequences. For such cases the Neighbor-Joining (NJ) method is useful due to its computational speed and accuracy. As datasets grow, however, the fraction of the tree space examined by the NJ algorithm becomes miniscule. Therefore we examine the accuracy of NJ method for inferring very large phylogenies using computer simulations. We also present a new likelihood method that estimates all pairwise distances simultaneously. This method for distance matrix estimation corrects up to 60% of the errors of NJ method. Our simulation results show that a 128-times difference in the number of sequences (32 to 4096) caused only a 5% average reduction in the accuracy of NJ method even in the presence of extreme transition/transversion biases, highly skewed nucleotide compositions, and extensive variation of evolutionary rate among lineages.

TANAKA, Hiroshi

Analysis of Within-patient Drug Resistance Evolution by Reconstructing Longitudinal Phylogenetic Tree of HIV-1 pol Gene under HAART

Ren F. [1], Sugiura W. [2], Hasegawa N. [1], Matusda M. [2], **Tanaka H.** [1]. (1) Tokyo Medical and Dental University, (2) National Institute of Infectious Diseases

We have previously proposed an algorithm for longitudinal analysis of within-patient viral evolution and successfully reconstructed a longitudinal phylogenetic tree of 24 HIV-1 env variants obtained from a patient who had not received drug treatment. However, viral evolution in patients who are undergoing drug therapy, especially HARRT (Highly Active Anti-Retroviral Therapy), is much more complicated because of the selective pressure from the anti-viral drugs. Here we report that we have improved our algorithm and applied it to a large data set (541 viral samples) of the HIV-1 pol gene obtained from a patient who had been undergoing HAART over 3 years. By using the revised method, we drew a longitudinal phylogenetic tree that not only clarified the evolutionary relationship of the 541 viral variants, and also demonstrated the dynamic change of the viral population that responded to the change of the anti-HIV drugs. The results contribute to our understanding of the mechanism of drug resistant mutation acquisition.

TANAKA, Tsuyoshi

Evolution of vitamin B6 (Pyridoxine) metabolism by gain and loss of genes

Tanaka T. [1], Tateno Y. [1], Gojobori T. [1]. (1) National. Institute of Genetics.

Vitamin B6 (VB6) is a cofactor of the many diverse enzymes. While most unicellular organisms and plants

can biosynthesize VB6 by themselves, animals such as insects and mammals can not, and must take it as nourishment. To understand the evolutionary process of VB6 metabolism from the viewpoint of gain and loss of genes, we compared VB6 biosynthetic pathways among 122 species whose genome sequences were completely determined. As a result, we have found that every gene in VB6 biosynthesis was lost more than once in the whole evolutionary lineages of the 122 species. We have also found the following two points in the evolution of VB6 biosynthesis: (1) The breakdown of VB6 biosynthesis has independently occurred at least three times in animal lineages, and (2) a de novo pathway was created by the existence of pdxB in gamma-proteobacteria. These findings indicate that the evolutionary process of VB6 metabolism has been extraordinarily dynamic with respect to gain and loss of genes.

TAYLOR, James

Are transitions at CpG dinucleotides replication dependent?

Taylor J. P. [1], Makova K. D. [1]. (1) The Pennsylvania State University

Low evolutionary divergence between human and chimpanzee allows more reliable analysis of mutations that occur at high rates. Using genome wide alignments between these two species we compared rates of transitions at CpG dinucleotides for autosomes and sex chromosomes in two sets of presumed neutral sequences: the first constructed by excluding known and predicted coding regions and CpG islands, the second consisting of ancestral repeats predating primate/rodent divergence. We found that in both sets CpG transition rates were higher on Y and lower on X compared to autosomes, and the male-to-female mutation rate ratio was approximately the same when calculated for each sample set and all three possible comparisons between X, Y, and autosomes. This result is consistent with a male mutation bias at these sites. Since the male germ-line undergoes more replications than the female germ-line, this suggests that CpG transitions in primates are replication dependent.

TEELING, Emma

Conservation of the X chromosome across Eutheria; 105 million years of conservation revealed through comparative radiation hybrid maps.

Teeling E. C. [1], Crumpler N. [1], Page J. [2], O'Brien S. J. [2], Murphy W. J. [1]. (1) SAIC- Frederick Inc., National Cancer Institute, (2) National Cancer Institute

Detailed genomic maps are only available for members of two (Euarchontoglires, Laurasiatheria) out of the four superordinal clades of mammals. We are generating radiation hybrid (RH) derived physical-maps for members of the two remaining unrepresented eutherian clades: Afrotheria (African elephant) Xenarthra (nine-banded armadillo). It has been estimated using molecular data that all eutherians last shared a most recent common ancestor 105 million years ago. The X chromosome is considered one of the most conserved mammalian chromosomes in terms of structure and gene content. To investigate if this conservation has been maintained throughout all of Eutheria, homologous ordered markers on the X-chromosome RH maps were compared across cat, pig, dog, human, macque, horse, rat, mouse, cow, armadillo and elephant. Shared syntenic gene associations, and shared chromosome breakage points, conserved across all four superordinal groups of mammals for the past 105 million years were identified

TERRY, Philip

Evolution of 3-isopropylmalate dehydrogenase.

Terry P. M. [1], Moriyama H. [1]. (1) University Nebraska-Lincoln

In excess of 150 protein sequences for a family of decarboxylating dehydrogenases which include those for 3-isopropylmalate, isocitrate, and tartrate are now available for study of evolutionary, sequence, structure, and function relationships among species. Among them, 3-isopropylmalate dehydrogenase or (IPMDH) is well-studied biophysically and biochemically.

To analyze sequence variation in IPMDH among the available sequences, we created multiple sequence alignments (MSA), using as input, a set of BLASTP hits (E value < e-14) resulting from an IPMDH as query. Gaps and substitutions in columns of the MSA were compared with available structures from the

PDB to validate the alignment of sequences in the MSA. We projected biochemical knowledge of IPMDH to the MSA to validate the alignments.

TETZLAFF, Karin

Dynamics of rDNA in 400 generations of *Drosophila melanogaster*

Tetzlaff K. R. [1]. (1) University of Rochester

The rDNA loci of eukaryotes are comprised of tandem units, each containing a set of rRNA genes, and often inserted with retrotransposable elements. These units undergo recombinations that homogenize the rRNA genes and eliminate insertions. To better understand this mechanism of concerted evolution, we characterized the rDNA locus of 19 *D. melanogaster* lines which were separated from a homogenized stock 15 years ago (400 generations). We found that during this period the percent of inserted rDNA units inserted by R1 and R2 elements now ranged from 40-70% of the locus, and the size of the locus itself varied 2-fold. Comparison of the intergenic spacer (IGS) showed little change in the rDNA units on the X chromosomes but substantial differences on the Y. Correlations between IGS length variants and neighboring R2 insertions were also examined. These findings provide new insight into mechanisms of concerted evolution and how retrotransposable elements survive in the rDNA locus.

THEISSEN, Guenter

Evolution of MADS-box genes controlling flowering in plants

Theissen G. [1], Kaufmann K. [1], Nutt P. [1], Simon H. [1]. (1) University of Jena

Trying to understand the evolution of genes and of complex organismic structures are both ambitious scientific goals. Understanding the relationship between the two processes, however, is an even more considerable intellectual challenge. Heterochronic and homeotic meristem and organ identity genes belonging to the MADS-box gene family sculpt the structure of extant angiosperm inflorescences and flowers. Changes in these genes may thus have significantly contributed, directly or indirectly, to the origin of biological complexity and diversity on our planet. The role of MADS-box gene phylogeny in the evolution of flowering plants is thus an attractive system for studying the interrelationships between the phylogeny of developmental control genes and the evolution of morphological structures. The evolution of MADS-box genes in plants is characterized by the frequent generation of paralogous genes followed by non-, neo- and subfunctionalization events. Neo- as well as subfunctionalization events included changes in regulatory as well as in coding regions. Case examples from studies in diverse seed plants suggest that spatiotemporal changes in MADS-box gene expression patterns and in MADS-domain protein interactions played important roles during the origin of floral evolutionary novelties.

THORNBURG, Bart

Transposable Elements as a Source of Mammalian Transcription Factors

Thornburg B. [1], Gotea V. [1], Makalowski W. [1]. (1) The Pennsylvania State University

Repeatitive elements compose nearly 50% of the human genome. Once thought of as "junk" DNA, these sequences have recently been more and more regarded as genomic treasures, as evidence of their coding capacity and regulatory function started to emerge. The purpose of the presented study was to determine if classes of transcription factors have a proclivity to bind to a certain type of repeated element. 2000 base pairs upstream of the transcription start site of each annotated human gene were scanned for mobile elements and possible transcription factor binding sites overlapping these elements. These results were compared against MATCH data for randomly generated repeats of the same length and GC content. Comparative statistical analysis will reveal whether or not the different types of repeats (SINEs, LINEs, LTRs, etc) have a propensity to contain transcription signals for specific classes of binding factors.

THORNE, Jeffrey

Protein evolution with dependence among codons due to tertiary structure

Robinson D. M. [1], Jones D. T. [2], Kishino H. [3], Goldman N. [4], **Thorne J. L.** [1]. (1) North Carolina State

University, (2) University College London, (3) University of Tokyo, (4) European Bioinformatics Institute
The relationship between phenotype and survival of the genotype is central to both genetics and evolution. We have been developing novel statistical tools for studying this relationship at the molecular level. Protein-coding DNA sequences serve as the genotype whereas protein tertiary structures represent a fundamental unit of phenotype. We rely upon the observation that protein sequences change more quickly over time than do tertiary structures. As a surrogate for fitness, we use measures of sequence structure compatibility that have been designed for protein fold recognition. Incorporation into evolutionary models of dependence among codons due to protein tertiary structure leads to diverse applications, including enhanced statistical inference of adaptive evolution.

TILLICH, Michael
Plant Organellar RNA Editing – On the Evolution of Editing Sites and *trans* Factors
Tillich M. [1], Poltnigg P. [1], Funk H. [1], Schmitz-Linneweber C. [2], Maier R. M. [1]. (1) Philipps-Universität Marburg, Zellbiologie, (2) University of Oregon, Institute of Molecular Biology
In plastids and mitochondria of vascular plants C-to-U RNA editing at specific nucleotide positions restores evolutionary conserved codon identities, and in all cases analysed, editing is crucial for the function of the affected proteins. In plastids and probably also in mitochondria, editing sites are recognized by site-specific, nuclear encoded proteins (*trans* factors) which seem to form evolutionary units with their respective target sites. RNA editing displays high evolutionary variability reflected in species-specific patterns of editing sites and corresponding *trans* factors. The rather constant number of total editing sites in organellar genomes of different species indicates a balanced gain and loss of editing sites during evolution. Loss of an editing site seems to be accompanied by loss of the corresponding *trans* factor. Beside its unknown origin, it is still puzzling why organellar RNA editing is evolutionary preserved, since the C-to-U conversions posttranscriptionally derived on the RNA level could also have been achieved by C-to-T point mutations in the organellar genomes.

TING, Chau-Ti
Evolution of the male accessory gland protein gene *Acp26Aa* in *Drosophila mauritiana*
Ting C-T. [1], Poh Y-P. [1], Lin Y. P. [2], Tsaur S-C. [2]. (1) Tsing-Hua University, (2) Academia Sinica
Many genes pertaining to male reproductive functions have been shown to evolve rapidly between species due to the action of positive selection. The accessory gland protein gene *Acp26Aa* and the speciation gene *OdsH* of *Drosophil*a are two such examples. From our previous studies, an unprecendented very high level of amino acid polymorphism in the N-terminal quarter of *Acp26A*a exon 2 in *D. mauritiana* has been observed. In the case of *OdsH*, *D. mauritiana* shows excess fixed differences within *D. simulans* clade. In order to test whether *D. mauritiana* has experienced fast evolving can be attributed to lineage-effect. Testis ESTs were collected for comparing the nonsynonymous substitution rate and synonymous substitution rate with *D. melanogaster* and *D. pseudoobscura*, respectively. These data suggest *D. mauritian*a has been under more intense sexual selection as a result of reinforcing lineage-specific effect.

TORGERSON, Dara
Enhanced Adaptive Evolution of Sperm-Expressed Genes on the Mammalian X Chromosome
Torgerson D. G. [1], Singh R. S. [1]. (1) McMaster University
Genes on the mammalian X chromosome may be under unique evolutionary pressure due to hemizygous expression in males. Selective constraints could be more pronounced in X-linked genes, as any recessive deleterious mutation would be immediately expressed in males and could be removed more efficiently from a population. Alternatively, if a recessive mutation were beneficial, an immediate exposure to selective forces could be advantageous, and may facilitate adaptive evolution. We tested whether 1) genes on the X chromosome have a higher incidence of positive selection than those on the autosomes, and 2) whether positive selection acts more strongly on X-linked genes in terms of the number of codons

affected. Using a comparison of 85 orthologous sperm-expressed and tissue-specific genes from human, mouse, and rat, we suggest that hemizygous expression facilitates adaptive evolution, and that X-linkage enhances the strength of adaptive evolution in positively selected sperm-expressed genes.

TURNER, Elizabeth

Reproductive isolation and reinforcement in *Neurospora*: mapping complex traits in a model filamentous fungus

Turner E. [1], Taylor J. W. [1]. (1) Plant & Microbial Biology, UC Berkeley

Reproductive isolation between species in the genus *Neurospora* is a polygenic trait with high heritability. To study the evolutionary genetics of interspecies isolation and intraspecies cohesion in two taxa pairs, (1) *N. crassa* and *N. intermedia,* and (2) the moderately genetically and reproductively isolated A and C lineages of *N. crassa,* we used AFLP markers to create linkage maps that form the bases of quantitative trait loci analyses. Mating assays with genotyped f1 hybrids identified loci that are significantly associated with the strength of reproductive isolation between f1 hybrids and *N. crassa* or *N. intermedia* tester strains from different geographic regions. In natural strains interspecies reproductive isolation is strongest between individuals originating in sympatry, a pattern consistent with the reinforcement of reproductive isolation barriers by natural selection. We have identified loci that are correlated with a sympatry associated barrier between *Neurospora* species.

URRUTIA, Araxi

Codon usage in human genes under: Beyond GC content and context mutations

Urrutia A. O. [1], Kalyanaraman V. [1], Kumar S. [1]. (1) Arizona State University

Unequal use of alternative codons has been observed in all kinds of organisms. In non-vertebrate organisms codon usage bias has been associated with levels of gene expression. Genes with higher activity have higher bias in their use of codons. In mammals, however the relationship between codon bias and expression has not been proved. Although in mammalian genes codon usage bias has also been observed, in these species, codon bias is strongly influenced by genomic base composition variations and context mutation patterns. Some attempts have been made to delucidate the extent of codon bias once correcting for these effects but these are either based on small samples or have failed to account for both of the effects (composition and context mutations) together. Therefore, it remains unknown what is the true extent of codon preference in the human genome and its possible selective origin. Here we present the analysis of codon bias in over 30000 human genes, correcting for both background base composition and context dependent mutations at the same time. We also analyze whether any residual bias is related to levels of gene activity.

UYENOYAMA, Marcy K.

Maximum-likelihood estimation of rates of recombination within mating type regions

Uyenoyama M. K. [1], Takebayashi N. [2], Newbigin E. [3]. (1) Duke University, (2) University of Alaska, (3) University of Melbourne

Features common to many mating type regions include recombination suppression over large genomic tracts. Model systems for homomorphic self-incompatibility (SI) in flowering plants share these characteristics. We introduce a maximum likelihood method for the exact computation of the joint probability of numbers of neutral mutations segregating at the determinant of mating type and at a linked marker locus. The underlying Markov model incorporates strong balancing selection into a two-locus coalescent. We apply the method to the estimation of the rate of recombination between a marker locus and S-RNase, the determinant of SI specificity in pistils of Nicotiana alata. Even though the sampled haplotypes show complete allelic linkage disequilibrium and recombinants have never been detected, a highly significant deficiency of synonymous substitutions at the marker compared to S-RNase suggests a history of recombination. We estimate a recombination rate of perhaps three orders of magnitude greater than the rate of synonymous mutation.

VAN DE LAGEMAAT, Louie

The human and chimpanzee genomes differ by over 8500 retroelement insertions

van de Lagemaat L. N. [1], Medstrand P. [2], Svenback D. [2], Mager D. L. [1]. (1) BC Cancer Research Centre, Vancouver, Canada, (2) Faculty of Medicine, Lund University, Lund, Sweden

We used the chimpanzee draft sequence to conduct a genome-wide analysis of retroelement insertional differences between human and chimpanzee. We detected over 8500 such events, 70% of them in human. Interesting findings revealed by this analysis: 1) As with older ERVs, new ERV insertions in gene introns, which likely contain strong promoters and polyadenylation signals, are significantly more often oriented antisense to the gene's transcriptional direction. 2) Alu elements, which lack strong pol II promoters and poly-A signals, are evenly distributed with respect to genes. 3) New L1-dependent non-L1 insertions are much more numerous in human, although L1 insertions are comparable in number. 4) As reported before for a smaller dataset, the chimpanzee-specific L1 insertions were significantly shorter that those in human. This work suggests that retroelements have been siginificantly more active in human since divergence from chimpanzee.

VAN TUINEN, Marcel

Using genetics to study effect of environmental change on biotic evolution: from micro to macro

van Tuinen M. [1], Ramakrishnan U. [1], Hadly E. A. [1]. (1) Stanford University

For decades, attempts to put genetic results into evolutionary or climatic frameworks have been rather descriptive, based on scanty data and assumed simple evolutionary models. Evolutionary geneticists now face a major opportunity when integrating across the rapidly increasing amount of genetic data and existing biologic scenarios based on ecology, fossils or climate models. Although genetic data acquisition and analysis, and climatic modeling have improved tremendously, several limitations remain, especially at the interface of deep and recent time. Here, instead of exclusively focusing on the merits of genetics in reconstructing evolutionary history, we highlight the often-overlooked reflective tool of the iterative use between genes and history by merging data from two promising genetic methods with paleontologic (fossil) and ecologic (life history) information.

VARKI, Ajit

Multiple differences in Sialic Acid Biology between Humans and Great Apes

Varki A. [1]. (1) UC San Diego

Sialic acids (Sias) attached to the ends of sugar chains on vertebrate cell surfaces interact with Sia-recognizing receptors of extrinsic and intrinsic origin, such as the Siglecs. A Sia called Neu5Gc is widely expressed in mammals, but not humans. This results from an inactivating mutation the human *CMAH* gene, mediated by an *Alu* replacement event ~2.5-3 mya. Additional human-specific changes in Sia biology include: a build-up of the precursor Sia Neu5Ac; differences in expression and distribution of Siglec-1 (sialoadhesin), a macrophage receptor recognizing Neu5Ac and not Neu5Gc; a mutation in Siglec-XII, changing an arginine required for Neu5Gc recognition in apes; a switch in Sia linkage on some cell types; placental Siglec expression; and, changes in Neu5Gc versus Neu5Ac preferences and expression patterns of some Siglecs. Since <60 genes are involved in synthesis, recognition and turnover of Sias, these changes may be of significance for human evolution, and have implications for the human condition, such as susceptibility or resistance to pathogens; signaling by immune system receptors; placental function; expression of oncofetal antigens; and, consequences of dietary intake of Neu5Gc. Down-regulated expression of *CMAH* in other mammalian brains suggests that this mutation could have also played some role in brain evolution.

VEERAPPAN, Chendhore

Molecular evolutionary analysis of the SET domain protein families in fungal genomes

Veerappan C. S. [1], Avramova Z. [1], Moriyama E. [1]. (1) University of Nebraska , Lincoln

The SET domain is approximately a 130 amino acid motif identified in plants, animals, and yeast, and

considered to be associated with eukaryotic functions. These proteins both activate and repress gene transcription mechanisms. Proteins in different families contain unique sets of other domains that are not shared between different families. In order to elucidate evolutionary relationships and distributions of this protein family across eukaryotes, we are conducting large-scale searches from various fungal genomic databases as well as protozoan and other eukaryotes. Our results indicate that some SET-domain protein groups unique to filamentous fungal species. Phylogenetic analysis shows that these proteins can be classified based on their internal architectures of SET domain sequences.

VEKEMANS, Xavier

Identification of the targets of balancing selection through analyses of trans-specific polymorphisms at the self-incompatibility gene *SRK* in genus Arabidopsis

Castric V. [1], Charlesworth D. [2], Schierup M. H. [3], **Vekemans X.** [1]. (1) Université de Lille 1, (2) University of Edinburgh, (3) University of Aarhus

Balancing selection at the self-incompatibility locus in plants (S-locus) causes the persistence of many allelic lineages over extended time periods. This process explains the many observed cases of trans-specific polymorphisms, and unusually high levels of nucleotide divergence between S-alleles within species. It also implies that diversity analyses cannot pinpoint nucleotides targeted by selection. We tested an approach aiming at identifying nucleotide sites that are not targets of selection by comparing pairs of trans-specific alleles in recently derived species, assuming they have retained the same specificity. We investigated variation in *Arabidopsis halleri* at *SRK*, the S-locus female specificity determining gene. We analysed trans-specific polymorphisms by comparing *A. halleri* S-alleles with published *A. thaliana* and *A. lyrata* sequences. Among 16 putative alleles identified in *A. halleri*, nine were similar to previously identified *A. lyrata* alleles while one was closest to the *A. thaliana* allele. We compared the location of the nucleotide substitutions that occurred between each pair of trans-specific alleles with those occurring among alleles within species, and with sites exhibiting excess non-synonymous substitutions. We also compared the substitution rate along the branches connecting the trans-specific alleles at *SRK* with that for other members of the gene family and unrelated "reference" genes.

VERNOT, Benjamin

Notung 2.0: A Program for Evaluating and Optimizing Gene Duplications and Losses in Gene Trees.

Vernot B. H. [1], Durand M. D. [1]. (1) Carnegie Mellon University

Gene tree analysis is a powerful approach to studying the history and timing of duplications in the evolution of a gene family. However, if a gene tree is unrooted or contains edges with low bootstrap values, more than one interpretation is possible. Furthermore, for genome scale studies, a large number of trees must be considered. Software tools to automate this analysis and to manage, score and display alternate interpretations are required. We present an exploratory analysis tool to evaluate alternate hypotheses in terms of the number of gene duplications and losses that occurred. Our program can also be used to analyze a large number of rooted gene trees automatically. When tested on gene phylogenies presented in recent articles on large scale duplication, our program consistently yielded results that agreed with the interpretations of the original publications.

WADHAWAN, Samir

Overlapping Coding Regions in Mammalian Genomes

Wadhawan S. R. [1]. (1) The Pennsylvania State University

May a mammalian gene encode more than one protein by using alternative (out of phase) reading frames (ARFs)? ARFs are commonly found in compact viral genomes but thought to be without precedence in higher eukaryotes. Recent studies have identified two mammalian genes encoding functional proteins using ARFs. Are there other mammalian genes containing ARFs? Is there an evidence for a "hidden proteome" consisting of previously undetected ARF-encoded polypeptides? What are the patterns of co-

evolution between two proteins encoded by the same region of DNA? To answer these intriguing questions we performed a genome wide survey and found 81 human genes having an ARF in addition to the canonical coding region. In each case this arrangement is conserved in at least two orthologs from non-primate mammals. These important findings may affect our understanding of the mammalian gene organization and evolution.

WALL, Kerr

PlantTribes: A global classification for *Arabidopsis* and Rice Proteins

Wall P. K. [1], Leebens-Mack J. H. [1], Altman N. [2], Albert V. [3], Field D. [4], dePamphilis C. W. [1]. (1) Department of Biology, Penn State University, (2) Department of Statistics, Penn State University, (3) The Natural History Museum, University of Oslo, (4) Centre for Ecology and Hydrology, Oxford

With rapidly growing numbers of whole genome and EST sequences in our public databases, sequence-based protein classification systems are providing foundations for gene annotation, functional genomics, and comparative investigations of gene and genome evolution. We used the clustering program TribeMCL to place all genes in the inferred *Arabidopsis* and Rice proteomes into putative gene families. Classifications have been constructed using three clustering stringencies, and jackknife analyses were used to test the strength of support for each cluster. The results of these analyses provide insights into the *Arabidopsis* and Rice genomes, gene family evolution, and the propagation of functional domains within and among gene families. We introduce our PlantTribes database as a useful scaffold for sorting genes from other plant species into objectively defined clusters of *Arabidopsis* and Rice genes that can be aligned and analyzed in formal phylogenetic analyses.

WANG, Wen

Origin of new genes revealed by young genes

Wang W. [1], Long M. [2]. (1) Kunming Institue of Zoology, Chin. Acad. Sci., (2) University of Chicago

How new genes originate in genomes is a fundamental question in evolution. The conspicuous difference of gene numbers in various organisms shows the importance of this process. In addition, the knowledge of how new genes are created in nature is also useful in designing new gene products. However, empirical study on origin of new genes has been highly challenging, as it necessitates the discovery of young genes because the early evolutionary signatures of old genes has often been obscured by subsequent evolutionary process. We have successfully developed an efficient experimental approach to identify young genes. Application of this approach to search new genes in the eight Drosophila species of the D. melanogaster subgroup has yielded promising results. These young genes, including sphinx and monkey king genes, has shed insight into understanding of this fundamental process.

WANG, Guanfang

Genome-wide analysis of Arabidopsis cyclins and comparative phylogenetic analysis of plant cyclin-like proteins

Guanfang Wang [1], Hongzhi Kong [1], Yujin Sun [1], Xiaohong Zhang [1], Wei Zhang [1], Naomi Altman [2], Hong Ma [1]. (1) Department of Biology, The Pennsylvania State University, (2) Department of Statistics, The Pennsylvania State University

Cyclins and cyclin-dependent kinases are key regulators of eukaryotic cell cycle progression. Various evidences have suggested that plant cyclins have unique features. However, information on plant cyclins is rather limited compared to animal and yeast. To gain a better understanding of plant cyclins and to obtain clues about their potential functions, we searched exhaustively for cyclin genes in Arabidopsis and other vascular plants. Based on phylogenetic analysis, we defined 10 classes of plant cyclins, 4 of which are plant-specific. Microarray and RT-PCR analyses were also performed to obtain global expression profiles of Arabidopsis cyclin genes. Expression patterns and phylogenetic analyses of Arabidopsis cyclin genes suggest potential gene redundancy among closely related members. Our study provides an

opportunity to rapidly assess the position of plant cyclin genes in terms of evolution and classification, serving as a guide for further functional study of plant cyclins.

WANG, Xiaoxia

Testing the chromosomal speciation hypothesis for humans and chimpanzees

Wang X. [1], Podlaha O. [1], Zhang J. [1]. (1) University of Michigan

Fixed differences of chromosomal rearrangements between isolated populations may promote speciation by preventing between-population gene flow upon secondary contact. This chromosomal speciation hypothesis thus predicts more rapid genetic divergence on rearranged than colinear chromosomes. A number of studies of fungi, plants, and animals, including limited genetic data of humans and chimpanzees, support the hypothesis. Here we reexamine the hypothesis for humans and chimpanzees with substantially more genomic data than were used previously. No difference is observed between rearranged and colinear chromosomes in the level of genomic DNA sequence divergence between species. The same is also true for protein sequences. When the gorilla is used as an outgroup, no acceleration in protein sequence evolution associated with chromosomal rearrangements is found. Furthermore, divergence in expression pattern between orthologous genes is not significantly different for rearranged and colinear chromosomes.

WATABE, Teruaki

V3 loop conformations of HIV-1 and structural features associated with viral tropism

Watabe T. [1], Kishino H. [2], Kitazoe Y. [1]. (1) Kochi University, (2) University of Tokyo

The principal neutralizing determinant of HIV-1 has been mapped to a 35-amino acid residue in the V3 region of gp120. The amino acid sequence of V3 loop is highly variable, especially in the regions flanking the highly conserved central GPG part. The hexapeptide GPGRAF including the central tripeptide elicits the antibodies that neutralize all HIV-1 isolates that contain the hexapeptide and do not neutralize those lacking GPGRAF. Holmes et al. (1992) have analyzed the 24 amino acid sequences of V3 loop and observed that the same hexapeptide motif (GPGSAV) has evolved convergently. This provides strong evidence that selective processes determine the evolutionary fate of sequence variants in the V3 region. We have developed a pairwise potential technique to examine conformations for given sequences, based on the Knowledge Based Hamiltonian (M. Sasai, 1995). We compare the conformations preferred by the 24 sequences of V3 loop and discuss structural features associated with viral tropism.

WEBSTER, Matthew

Fixation biases affecting human SNPs

Webster M. T. [1], Smith N. G. [1], Ellegren H. [1]. (1) Uppsala University

Under neutrality all classes of mutation have an equal probability of becoming fixed in a population. Here we analyze the frequency distributions of >5000 human SNPs, providing evidence of biases in the process of fixation of certain classes of point mutation most likely attributable to biased gene conversion (BGC). The results indicate an increased fixation probability of mutations that result in the incorporation of a G:C base pair. BGC has the potential to explain both the existence of isochores and the compositional asymmetry in mammalian transcribed regions. The existence of fixation biases should be incorporated into realistic models of neutral evolution.

WEINREICH, Daniel

Rapid evolutionary escape from a local fitness peak in large populations

Weinreich D. M. [1]. (1) Harvard University

In the early 1930's, Sewall Wright observed that fitness interactions between loci in the genome can result in local peaks on the fitness landscape, and the problem of how multiple mutations that are individually deleterious but jointly beneficial may fix remains a central problem of evolutionary genetics. Theoretical attention has focused on small populations, in which low fitness mutational neighbors of the peak

genotype are more likely to fix, owing to genetic drift. Using simple, haploid population genetic model I show analytically that as effective population size increases, the expected time to escape from a local peak across a valley two mutations wide quickly approaches a quantity proportional to the reciprocal of population size, a sharply declining function. This result is robust to the depth and width of the fitness valley, fitness gain across the valley, mutation rate and modest recombination, and is well corroborated by computer simulation.

WESSLER, Susan

Transposable Elements: Make New Genes and Change the Old

Wessler S. R. [1], Jiang N. [1], Bao Z. [2], Zhang X. [1], Eddy S. [2]. (1) University of Georgia, (2) Washington University

Transposable elements (TEs) were discovered in maize as the agents responsible for the unstable mutant alleles of many genes. These alleles were the starting point for the cloning of the first plant DNA transposons, which turned out to be the founding members of several transposon families including hAT (Ac/Ds), CACTA (Spm/En), and MULEs (Mutator-like elements) in addition to the first miniature inverted repeat transposable element (MITE). However, without genome sequence or a facile transformation protocol, the characterization of maize DNA transposons at a whole genome level has been impossible. Fortunately, rice has these resources in addition to harboring representatives of most of the TEs found in maize and in other large-genome cereal grasses.

Without prior genetic evidence for active DNA TEs, the two draft genomic sequences of rice (from subspecies japonica and indica) were analyzed using a program called RECON in order to identify all repeat families (including TEs). The RECON output has been the basis for many projects; two will be discussed in detail. One involves the identification of the first active DNA transposon family in rice (Ping/Pong) and the first active MITE in any organism (mPing). Since domestication, the mPing element has amplified almost 1000 fold and, in doing so, has become a significant factor in generating genomic diversity in this selfing plant. The second project was initiated to determine the prevalence of TE-mediated gene creation on a whole-genome level. Remarkably, we find that a significant fraction of the annotated genes in the rice genome were created though the activity of the rice Mutator family.

WHITEHEAD, Andrew

Evolutionary analysis of variation in global gene expression among Killifish populations

Whitehead A. [1], Crawford D. L. [1]. (1) University of Miami, RSMAS

It is probable that many heritable differences between taxa that are responsible for physiological adaptation are manifest as differences in gene expression. Killifish populations are distributed from Maine to Florida along one of the steepest thermal gradients in the world. We applied functional genomics using a cDNA microarray to test for differences in gene expression among populations. Extensive technical and biological replication allowed for the detection of significant differences in gene expression as small as 1.2-fold between groups. Phylogenetic autocorrelation uses genetic distance as a covariate to habitat temperature, and was applied to distinguish expression differences that are most likely due to random genetic drift from those most likely due to selection pressure. Phylogenetic analyses coupled with rigorous microarray experimental design and analysis of variance offer a strong approach for identifying adaptively important polymorphisms.

WILDMAN, Derek

Coincident amino acid and gene expression changes in nuclear encoded subunits of primate aerobic energy metabolism genes

Wildman D. E. [1], Uddin M. [1], Opazo J. C. [1], Goodman M. [1], Grossman L. I. [1]. (1) Wayne State University

Aerobic energy metabolism is regulated by cytochrome *c* oxidase (COX). In anthropoid primates, the genes for 70% of COX subunits have elevated Ka/Ks ratios. We sequenced one of these genes, *COX5A*, in

all major primate lineages. Ka is notably elevated in the human lineage since its divergence from the chimpanzee lineage. Humans and chimpanzees diverge five times more from one another than do mice and rats which share 100% identity in the mature polypeptide. Because the relatively enlarged anthropoid neocortex requires an extensive supply of aerobic energy, we hypothesize that changes in *COX5A* sequence promote efficient electron transfer in the brain of primates and especially in humans. To examine this idea we measured *COX5A* expression levels in primate brains. Humans express more of the gene than all other primates, due perhaps to the different *cis*-regulatory elements that are found in the *COX5A* promoters of humans and other primates.

WILLIAMS, Barry

Understanding the meaning of protein sequence evolution

Williams B. L. [1], Selegue J. E. [2], Carroll S. B. [3]. (1) Univeristy of Wisconsin, (2) University of Wisconsin - HHMI, (3) Universit of Wisconsin - HHMI

There is currently a disconnect among biologists as to the meaning of protein divergence. Evolutionary biologists debate whether amino acid replacements are neutral, nearly neutral, or fixed by selection. Molecular biologists study mutations in conserved sites with large fitness effects, but haven't considered the functional contribution of non-conserved sites. Ecologists measure large selection coefficients on natural phenotypic variation, but have yet to understand the nature of underlying mutations. One common element that is missing is knowledge of the fitness effects of amino acid replacements fixed among species. The evolution of yeasts provides a unique opportunity to examine such effects. We created several orthologous gene replacements and developed a sensitive competition assay to measure their relative fitness in vivo. The data show that even large numbers of substitutions are collectively of slight effect and could only be fixed by selection in very large populations.

WILLIAMSON, Scott

Inferring selection from the frequency spectrum of polymorphic sites using non-stationary population genetic models

Williamson S. [1], Bustamante C. D. [1]. (1) Cornell University

The frequency spectrum of polymorphic sites, or site-frequency spectrum, describes the relative frequency of rare, intermediate, and common mutations in a sample. Past studies have shown that statistical methods based on the full site-frequency spectrum have extraordinary power to detect even very weak selection. The chief disadvantage of these methods is that they are extremely sensitive to demographic, as well as selective, forces. Here we address this problem by developing a Poisson Random Field model that incorporates non-stationary population dynamics. If data are available from putatively neutral genomic regions (e.g. variation in non-coding regions or pseudogenes), then this model can be used to correct for the effect of demography on the site-frequency spectrum. Monte Carlo methods are used to address how robust this method is to different types of demographic forces. We demonstrate the application of this method using a large data set of human Single Nucleotide Polymorphisms.

WILSON, Alex

How divergent are they? Investigating the potential for cross-species application of pea aphid microarrays in the green peach aphid.

Wilson A. C. [1], Moran N. A. [2]. (1) Center for Insect Science, University of Arizona, (2) EEB, University of Arizona

Genomic level studies promise a synthetic understanding of biology. The International Aphid Genomics Consortium aims to develop genomic tools for the pea aphid, *Acyrthosiphon pisum*. This initiative includes submission of a genome sequencing white paper and development of first-generation microarrays. Whilst the pea aphid is the primary aphid used in developmental, host specialization and symbiosis studies, many aspects of aphid biology and evolution are better understood in other aphid species. One such species is the green peach aphid, *Myzus persicae*. We are interested in the potential application of *A. pisum*

microarrays to the study of genome degradation and rearrangement in long-term asexual lineages belonging to the *M. persicae* group. For this reason we have estimated DNA sequence divergence between *A. pisum* and *M. persicae* by obtaining DNA sequences from *M. persicae* for a range of *A. pisum* ESTs that showed high protein homology to known proteins in the drosophila genome.

WITTKOPP, Patricia

The genetic basis of divergent gene expression: cis and trans

Wittkopp P. J. [1], Haerum B. K. [1], Clark A. G. [1]. (1) Cornell University

Differences in gene expression play a central role in evolution. Both *cis*- and *trans*-regulatory changes contribute to divergent gene expression, but their respective contributions remain largely unknown. Using closely related *Drosophila* species, we examined *cis*- and *trans*-regulatory changes underlying interspecific expression differences. Functional *cis*-regulatory changes were detected by comparing the relative abundance of *D. melanogaster* and *D. simulans* transcripts in F1 hybrids, whereas differences in *trans*-regulatory activity were inferred by comparing the ratio of allelic expression in hybrids to the ratio of gene expression between species. We found that interspecific expression differences were almost always accompanied by changes in *cis*-regulatory function. In addition, *trans*-regulatory divergence affected approximately half of the genes we examined. These data contrast with recent studies of intraspecific gene expression, which found that expression variation was caused primarily by differences in *trans*-regulation. Distinct patterns of *cis*- and *trans*-regulatory changes appear to underlie expression differences within and between species.

WLASIUK, Gabriela

The *agouti* locus and coat color variation in the rock pocket mouse, *Chaetodipus intermedius*

Wlasiuk G. [1], Nachman M. W. [1]. (1) University of Arizona, EEB Department.

Coat color polymorphism in the rock pocket mice, *Chaetodipus intermedius*, presents a useful system for studying the genetic basis of adaptation and making direct connections between phenotype and genotype. In most populations, these mice are light-colored and live on light-colored rocks, while in others, mice live on dark lava and are correspondingly dark-colored, presumably an adaptation for crypsis. Previously, we showed that mutations in the melanocortin-1-receptor locus (*Mc1r*) are responsible for the light/dark phenotypic differences observed in one population but not in others. This suggests that dark color has arisen independently in different populations through changes at different genes. Here, we explore the role of *agouti* in dark color in three populations from New Mexico. In laboratory mice, an *agouti* mutation known as "black and tan" produces a dark-dorsal and light-ventral phenotype very similar to the dark phenotype of *C. intermedius*. This mutation has been molecularly characterized and results from an insertion of 6 kb that disrupts one of the promoters of *agouti*. Here, we characterize the promoter region of *agouti* in *C. intermedius*. We then conduct association studies with SNPs in this region in animals from several populations to determine if this gene is involved in the dark/light polymorphism.

WOLF, Yuri

Quantitative genomics: connections between phenotypic and evolutionary measures.

Wolf Y. I. [1], Koonin E. V. [1], Carmel L. [1]. (1) NCBI/NLM/NIH

Numerous reports in the literature explore pairwise relationships between various kinds of quantitative genomic measures. Some connections (e.g., the negative correlation between the expression level and evolution rate) seem to be well-established, whereas the status of others (e.g., the connection between the protein-protein interactions and evolution rate) is under investigation. Here we present preliminary results of a multivariate analysis of different measures of evolution rate, phyletic distribution, gene expression, and knockout fitness effect. We introduce the notion of "social status in the genome community" as a means for generating null hypotheses on the relationships between quantitative

genomic variables. The connections between all explored variables are relatively weak, albeit statistically significant.

WOLFE, Ken

Ancient genome duplication and adaptive gene order evolution in yeasts

Byrne K. P. [1], Scannell D. R. [1], Wong S. [1], Gordon J. L. [1], **Wolfe K. H.** [1]. (1) Trinity College Dublin

Ascomycete yeasts are a model system for investigating eukaryotic gene order and chromosomal evolution. I will discuss two findings:

(1) Rates and patterns of gene loss after polyploidy: A whole-genome duplication occurred in the common ancestor of several yeast species. Subsequently, most of the duplicated genes were quickly lost again. However, about 32% of loci were still duplicated at the time of speciation between *S. castellii* and *S. cerevisiae*, as compared to 14% remaining duplicated in each of those species today. Independent gene losses occurred at 200 loci where *S. cerevisiae* and *S. castellii* retain alternative (paralogous) single copies of genes. These differential gene losses result in many essential single-copy genes being located at non-syntenic positions in the two species' genomes, ensuring sterility of hybrids and supporting a chromosomal model of speciation in paleopolyploids.

(2) An example of natural selection acting to relocate genes. Gene order along chromosomes is generally well conserved among ascomycetes. However, the DAL cluster of allantoin metabolism genes was assembled recently, by dramatic relocation of genes from other scattered sites in the genome. We suggest that this dramatic rearrangement was driven by natural selection for increased ability to grow in low-oxygen environments.

WON, Yong-Jin

Cichlids of Lake Malawi

Won Y. [1], Sivasundar A. [1], Hey J. [1]. (1) Department of Genetics, Rutgers University

To study mechanisms of speciation in cichlid fish from Lake Malawi, we have examined multilocus of nuclear DNA sequence haplotypes together with linked microsatellites or short tandem repeats (STRs). These 'HapSTRs' offered a higher resolution of recent events associated with the high mutation rate of STRs, together with the advantages of low homoplasy of unique nuclear DNA sequence ranging approximately 400~600 bp. A Markov chain Monte Carlo method was applied to the genetic data to estimate the posterior probability distribution of model parameters under an 'isolation with migration' (IM) model. We estimated divergence time, migration rate and effective population sizes of *Labeotropheus fuelleborni* and *Labeotropheus trewavasae*.

WONG, Alex

Evidence for positive selection on genes expressed in the reproductive tract of female *Drosophila melanogaster*

Wong A. [1], Swanson W. J. [2], Wolfner M. F. [1], Aquadro C. F. [1]. (1) Cornell University, (2) University of Washington

Male-female evolutionary conflict and sexual selection are often invoked to explain the rapid, adaptive evolution of reproductive proteins in *D. melanogaster* and other species (Swanson and Vacquier, 2002). However, most studies have focused on male contributions to reproduction. A complete understanding of the evolutionary dynamics of reproductive proteins will also require the study of female reproductive molecules. Accordingly, we examined the evolution of several proteins expressed in the female reproductive tract, whose sequence signatures predict secretion or cell surface location, suggesting that they could be involved in male-female interactions at the molecular level. We have sequenced several such genes in up to eight *Drosophila* species each. Using maximum likelihood analyses (PAML; Yang, 1997), we detect a strong signature of positive selection on a subset of codons in at least two of the genes studied. Several genes have also shown patterns of nucleotide polymorphism consistent with recent adaptive fixations within at least one species (*D. melanogaster* and/or *D. simulans*). Our finding of strong

evidence for positive selection at several candidate female reproductive genes sets the stage for critical tests of the roles of sexual antagonism and sexual selection in driving the evolution of reproductive proteins.

WOOLFIT, Megan
Population size and molecular evolution on islands
Woolfit M. [1], Bromham L. [1]. (1) University of Sussex
The nearly neutral theory predicts that the rate and pattern of molecular evolution will be influenced by effective population size (Ne) because in small populations more mutations are expected to drift to fixation. This important prediction has not been widely empirically tested because of the difficulty of identifying comparisons between lineages which differ only in Ne. Island endemic species provide an ideal test of the effect of Ne on the tempo and mode of molecular evolution because species restricted to islands frequently have smaller Ne than their mainland relatives. We have selected 75 phylogenetically independent comparisons of island and mainland taxa, including vertebrates, invertebrates and plants, from 18 different island groups. By analyzing a wide range of DNA sequences including nuclear and mitochondrial genes, we show that island species have significantly higher ratios of nonsynonymous to synonymous substitution rates, as predicted by the nearly neutral theory.

WU, Gang
Synthetic Gene Designer: Bioinformatics tools to improve heterologous gene expression.
Wu G. [1], Freeland S. J. [1]. (1) University of Maryland Baltimore County
One limiting factor hindering heterologous gene expression (expressing genes in a non-native genome) is the different codon usage patterns across genomes. Site-directed mutagenesis has been the commonest way to remove problematic codons. More recently, some software allows users to generate a theoretical optimal gene by optimizing every codon; users then synthesize this gene by assembly PCR. However, this may be too simplistic: highly expressed native genes are not exclusively composed of optimal codons. One theory to explain the use of rare and suboptimal codons is that occasional codons may cause a translational pause that is important for correct folding of nascent polypeptides. We therefore hypothesize that a heterologous gene which matches "natural" codon bias patterns of the new genome will express better than either its wild-type counterpart or simplistically "optimized" equivalents. We have initiated experiments to test this hypothesis using a gene implicated in prostate cancer that has defied traditional heterologous gene expression methods. To inform these experiments, we are developing a web-based software package, the "Synthetic Gene Designer", to redesign any gene for heterologous expression by matching its native profile favored/disfavored codons to their equivalents in the new genomic context.

XING, Jinchuan
Molecular Phylogeny of Cercopithecidae (Old World Monkeys) as inferred by *Alu* insertions
Xing J. C. [1], Wang H. [1], Han K. D. [1], Hedges D. J. [1], Ray D. A. [1], Huang C. [1], Batzer M. A. [1]. (1) Louisiana State University
Alu elements are the most successful Short Interspersed Elements (SINEs) in primate genomes. Over the last 65 million years, *Alu* elements have accumulated in primate genomes via a process termed retroposition. Family- and subfamily-specific *Alu* insertions offer several advantages in the study of primate phylogeny: *Alu* elements are identical by descent and essentially homoplasy free; they have known ancestral state (absent); *Alu* insertions are also easy to genotype using a PCR assay. More than 100 *Alu* insertion loci specific to twelve Cercopithecid species representing nine different genera were identified by both computational and experimental methods. We used these loci to investigate branching order in Cercopithecidae. This resulting cladogram supports the relationships presented in most other molecular and morphology data. In addition, this study of Cercopithecidae phylogeny represents the

most robust molecular phylogeny to date and illustrates the utility of mobile element insertions as cladistic markers in studies of non-Hominid primate evolution.

YAN, Li

Rapid Diversification of Antimicrobial Peptides in Bombina Toads

Lee W. H. [1], **Li Y.** [1], Lai R. [1], Li S. [1], Zhang Y. [1], Wang W. [1]. (1) Kunming Institute of Zoology, CAS

Antimicrobial peptides secreted by the skin of many amphibians play an important role in the innate immunity defense. We identified 39 different antimicrobial peptides cDNA clones from a skin cDNA library of the Chinese red belly toad (Bombina maxima). Each of the cDNAs codes for two kinds of antimicrobial peptides, maximin and maximin H. Among these cDNAs, we found that the rate of non-synonymous substitution (dN) in both the maximin and maximin H domains significantly exceed the rate of synonymous substitution (dS), whereas the same pattern was not observed in other regions, such as the signal and propiece regions. In light of the demand of defending various pathogens, it seems that antimicrobial peptide genes in the Bombina toads have been experiencing rapid diversification driven by Darwinian selection. To understand the evolution process in the antimicrobial peptide gene family, we constructed gene trees for the maximin, maximin H and the remaining structural regions, respectively. The clustering relationships among these genes disclosed by the three trees vary remarkably. Together with the observation that the dS of the maximin region is different from other regions, this result suggest that domain shuffling or gene conversion among these genes has frequently happened.

YOON, Hwan Su

The Origin of Minicircle Genes in the Dinoflagellate Algae

Yoon H. S. [1], Hackett J. D. [1], Bhattacharya D. [1]. (1) University of Iowa

Dinoflagellates present many interesting and perplexing issues in plastid evolution. These algae are the main source of toxic red tides in the ocean and contain different types of plastids that show unique characters such as the type of accessory pigment, the number of plastid membranes, the presence of Form-II rubisco, and minicircle-encoded photosynthetic genes. Although a red algal origin has been postulated for the plastid in the dominant peridinin-containing dinoflagellates, this issue remains unclear because of limited taxon sampling and highly divergent minicircle gene sequences. To enable a better understanding of dinoflagellate plastid evolution, we determined the sequence of five minicircle-encoded plastid genes in peridinin-containing dinoflagellates and their homologs from fucoxanthin-containing taxa as well as from chromist and red algae. Maximum likelihood and Bayesian analysis using the data set of five concatenated proteins show that the plastid sequences in peridinin-containing dinoflagellates are nested within the stramenopiles, whereas these genes in fucoxanthin-containing dinoflagellates branch within the haptophyte algae. These results are consistent with host phylogenies of alveolates and begin to clarify the complex evolutionary history of dinoflagellate plastids.

ZAHN, Laura

Evolution of MADS-box genes in the Angiosperms

Zahn L. M. [1], Arrington J. M. [1], Hu Y. [1], Landherr L. [1], Leebens-Mack J. [1], dePamphilis C. [1], Becker A. [2], Theissen G. [3], Ma H. [1]. (1) The Pennsylvania State University, (2) School of Biological Sciences Monash University, (3) Lehrstuhl für Genetik,Friedrich-Schiller-Univ.

The MADS-box genes are a family of transcription factors of which several sub-families are important in floral organ identity in the flowering plants. This study investigated the phylogenetics and expression of MADS-box genes in a basal eudicot California Poppy (Eschscholzia californica). The eudicots, including the Ranunculales, compromise 75% of all flowering plants. It has been hypothesized that the basic floral Bauplan is flexible through the early branching eudicot lineages including the Ranunculales, but floral morphology became canalized in the later branching "core eudicots" which then gave rise to the vast majority of eudicot species. This suggests that examination of the MADS-box genes within the

Ranunculales provide a key evolutionary milieu from which to determine what derived characters have become fixed in the core eudicots and which pleisomorphic characters have been retained.

ZHANG, Peng
Different Evolutionary Patterns between Young Duplicate Genes in the Human Genome
Zhang P. [1], Gu Z. [1], Li W. H. [1]. (1) Dept. of Ecology and Evolution, University of Chicago
We extracted 250 independent pairs of young duplicate genes from the human genome. Nearly 60% of the pairs have evolved at significantly different rates from each other at the amino acid level. More than 25% of these gene pairs also showed significantly different Ka/Ks ratios. Duplicate pairs with different rates of amino acid substitution tend to differ also in the Ka/Ks ratio and the fast-evolving copy tends to have a slightly higher Ks than the slow-evolving one. Over half of the fast-evolving copies have accumulated amino acid substitutions evenly across the protein sequences, whereas most of the slow-evolving copies exhibit uneven substitution patterns. Our results suggest that after duplication, one copy evolves faster than the other and accumulates amino acid substitutions evenly across the sequence while the other accumulates amino acid substitutions unevenly. Such different evolutionary patterns may be largely due to different functional constraints on the two copies.

ZHANG, Jianzhi
Parallel gene duplication and adaptive evolution of a digestive enzyme in leaf-eating monkeys
Zhang J. [1]. (1) University of Michigan
A subfamily of Old World monkeys, colobines are unique primates that use leaves rather than fruits and insects as their primary food source; these leaves are then fermented by symbiotic bacteria in the foregut. Similar to ruminants, colobines recover nutrients by breaking and digesting the bacteria with various enzymes, including pancreatic ribonuclease (RNASE1), which is secreted from the pancreas and transported into the small intestine to degrade RNAs. We here report that independent duplications of the RNASE1 gene occurred in Asian and African colobines and that the duplicated genes acquired enhanced catalytic activities against bacterial RNA. At the same time, the ability to degrade double-stranded RNA, a non-digestive activity characteristic of primate RNASE1, has been lost in the new genes, indicating functional specialization and relaxation of purifying selection. The finding of independent duplications and parallel functional changes in two colobine lineages provide strong evidence for the role of gene duplication in organismal adaptation.

ZHANG, Xian
Evidence for R2 element activity in *Drosophila simulans* population
Zhang X. [1], Eickbush T. H. [1]. (1) University Of Rochester
R2 is a non-LTR retrotransposable element that site-specifically inserts into the 28S genes of the rDNA locus. The element was found in all lineages of arthropods as well as in other phyla. Many questions remain concerning the frequency and pattern of R2 retrotransposition and how the element co-evolves with the rDNA locus. We used a PCR assay to monitor 5' truncated R2 elements in highly inbred isofemale lines of *Drosophila simulans* derived from a natural population. While in some lines all flies had identical R2 truncation profiles, other lines showed variation in the pattern of R2 insertions. In lines with variation, we observed R2 insertions and deletions in a single generation, suggesting active retrotransposition. R2 activity was maintained through many generations, and was not correlated with the size of the rDNA locus. Experiments are underway to determine why levels of R2 activity vary between lines.

ZHAO, Zhongming
Neighboring-nucleotide effects on single nucleotide polymorphisms (SNPs) in the mouse genome
Zhao Z. [1]. (1) Virginia Commonwealth University
Substitution patterns and their neighboring-nucleotide effects are important for understanding the

mutational mechanisms. In this study, we analyzed the neighboring-sequence context of the largest public collection of SNPs (433,192 biallelic SNPs in the dbSNP database) across the mouse genome. Large neighboring-nucleotide biases relative to the genome average were observed at the two immediate adjacent sites and small biases seemed to extend to 20 bases away from the substitution site. Further examination of the six categories of substitutions indicated that transitions were strongly affected by the hypermutability of CpG and the effects on transversions were complex. The A+T content at the two adjacent sites had an influence on the proportion of transversions, which was also observed in humans and plants. Finally, short sequences containing dinucleotide CpG were overrepresented at the polymorphic sites. In comparison, the neighboring bias patterns in the mouse and human genomes are generally same; however, extent of the biases is significantly less in mice.

ZHAXYBAYEVA, Olga

Cladogenesis, Coalescence and the Evolution of the Three Domains of Life

Zhaxybayeva O. [1], Gogarten J. P. [1]. (1) University of Connecticut

Using a simple model with constant number of species and rates of speciation equaling rates of extinction we explore the large-scale structure of the tree of life, and consequences from horizontal gene transfer for the concept of a most recent common ancestor of all living organisms (cenancestor). A simple null hypothesis based on coalescence theory already explains some features of the tree of life. It does not appear warranted to invoke more complex hypotheses involving bottlenecks and extinction events, to explain the features of the tree of life that are compatible with the null hypothesis. Simulations of genes and organismal lineages suggest that there was no single common ancestor that contained all the genes ancestral to the ones shared between the three domains of life. Rather each contemporary molecule has its own history and traces back to an individual molecular cenancestor. These molecular ancestors were likely to be present in different organisms and at different times.

ZHU, Lan

A Composite Likelihood Ratio Test for Detecting Natural Selection from Site Frequency Data

Zhu L. [1], Bustamante C. D. [1]. (1) Cornell University

We present a novel composite likelihood ratio test (LRT) for detecting natural selection from population genetic sequence data. The method uses the likelihood functions of Hartl, Moriyama, and Sawyer (1994) for inference on the selection parameter and corrects for non-independence among sites by application of coalescent simulations with recombination. Using coalescent and forward simulations with selection and recombination, we investigate the power, robustness, and accuracy of our test under a wide range of demographic, selective, and linkage assumptions. We also explore the effects of selection on the accuracy of various estimators of the local population recombination rate. We find that the test has excellent power to detect weak negative selection, and moderate power to detect positive selection. The test is quite robust to the underlying recombination rate, but not to certain demographic scenarios which will be discussed. Lastly, we apply our method to a database of >400 human genes with SNPs found by direct sequencing and identify several genes that are putative subject to weak negative or positive selection.

ZUFALL, Rebecca

Developmental Genome Processing and Protein Evolution in the Ciliate *Chilodonella uncinata*

Zufall R. A. [1], Katz L. A. [1]. (1) Smith College

Ciliates are microbial eukaryotes with dual genomes, one present in the transcriptionally-inactive germline micronucleus (MIC) and the other in the somatic macronucleus (MAC). In the development of the MAC from the MIC, ciliates process their genomes by chromosomal fragmentation, excision of internal excised sequences (IESs), and amplification of chromosomes. *Chilodonella uncinata* is in a class of ciliates, Phyllopharyngea, that undergo extensive processing to generate MACs containing thousands of gene-sized chromosomes. Previous analyses suggest that sequences involved in this processing are highly variable among ciliate lineages. In this study, we examine cis-acting signals involved in the elimination of

IESs in *C. uncinata* in order to understand the phylogenetic level at which processing signals are conserved. In addition, we are testing the hypothesis that the differential selection on dual genomes in ciliates allows unusually rapid divergence among paralogs of protein coding genes.

ZUFALL, Rebecca
The genetic basis of parallel evolution in flower color in *Ipomoea*
Zufall R. A. [1], Rausher M. D. [2]. (1) Smith College, (2) Duke University
Examining independent origins of phenotypically similar traits allows for unique insights into the processes of adaptation. In order to determine whether similar traits in related species have arisen due to shared selective pressures acting on homologous genes, we examine the adaptive significance, biochemical bases, and genetic underpinnings of parallel evolutionary transitions to red flowers in the genus *Ipomoea* (morning glories). A comparative analysis of floral morphology suggests that red flowers have evolved independently multiple times in this genus in response to selection by bird pollinators. A comparison of the genetics underlying these independent derivations of red flowers reveals that a diversity of mechanisms are responsible for these transitions in flower color. In addition, in one of these species, *I. quamoclit*, we find that there have been redundant mutations in the transition to red flowers indicating that a reversion in this character is very unlikely.

ZUPUNSKI, Vera
Adaptive evolution in snake venom Kunitz/BPTI protein family
Zupunski V. [1], Kordis D. [2], Gubensek F. [2]. (1) University of Ljubljana, Slovenia , (2) Jožef Stefan Institute, Slovenia
Snake venoms are rich sources of numerous different proteins encoded by functionally diversified multigene families. Many serine protease inhibitors were isolated from the venoms of Viperidae and Elapidae snakes, they belong to the bovine pancreatic trypsin inhibitor (BPTI) family. Snake Kunitz/BPTI homologues have diverse functions acting as protease inhibitors or neurotoxins. *Vipera ammodytes* Kunitz/BPTI inhibitors comprise a multigene family with non-coding regions more conserved than coding regions. In snake Kunitz/BPTI homologues the nonsynonymous-to-synonymous rate ratio is significantly greater than 1, indicating adaptive evolution. The diversification of the snake Kunitz/BPTI inhibitors, especially their inhibitor binding loop, resulted in new protein functions (dendrotoxins) or changed specificities toward proteases in prey. Despite the extensive studies on Kunitz/BPTI proteinase inhibitors, this study is the first conclusive evidence of the adaptive evolution in the highly conserved BPTI fold.

Author Index

(Single names and those in parentheses are presenting authors)

Boissinot S
Boore JL (Francino P)
Boore JL (Dehal P)
Boore JL
Boore J (Leebens-Mack J)
Bosdet I (Gonzalez J)
Bossau B (Andersson S)
Bossuyt F (Mannaert A)
Bossuyt F (Roelants K)
Bousquet J (Guillet-Claude C)
Braasch I
Braasch I (Salzburger W)
Brenner SE (Engelhardt B)
Brenner SE (Hill E)
Brenner SE
Briscoe AD
Briscoe DA (Frankham R)
Brisson D (Stoebel D)
Bromham L (Woolfit M)
Brooks DJ
Brown CW (Ong H)
Browning TL
Brunner S (Rafalski A)
Bulazel K
Burge CB (Friedman B)
Burk-Herrick A
Burk-Herrick A (Scally M)
Bustamante CD (Williamson S)
Bustamante CD (Jensen J)
Bustamante CD (Zhu L)
Bustamante CD
Byrne KP (Wolfe K)
Byrnes JK (Shiu S)
Caetano-Anolles D
Caetano-Anolles G
Cai JJ
Callinan PA (Hedges D)
Callinan PA (Batzer M)
Calteau A
Calvete O (Gonzalez J)
Campbell A (Snoeyenbos-West O)
Carboni A
Cardillo JC (Snoeyenbos-West O)
Carlton JM (Silva J)
Carmel L (Wolf Y)
Carney SL
Carrasco H (Gaunt M)
Carroll SB (Rokas A)
Carroll SB (Hersh B)
Carroll SB (Williams B)
Casals F (Gonzalez J)
Casillas S (Negre B)
Castillo-Davis CI (Hartl D)
Castillo-Davis CI (Achaz G)
Castric V (Vekemans X)
Catania F
Cavalcanti AR
Cavalcanti AR (Liang H)
Cavalcanti A (Landweber L)
Cavalieri D (Landry C)
Cavalli-Sforza L (Su B)
Cezairliyan BE (Rand D)
Chaix R
Chakraborty R (Su B)
Chalkia D
Chalkia D (Lin Z)
Champagne CE (Sinha N)
Chang WJ (Landweber L)
Charlesworth B (Haddrill P)
Charlesworth D (Vekemans X)
Charlesworth D (Kamau E)
Cheng CC (Ota T)
Cheng C
Cheng CHC (Reynolds N)
Chiang DY (Gasch A)
Chiaromonte F (Makova K)
Cho S (Ellis R)
Cho S
Cho S
Chung KH (Sinha N)
Chung W
Clark AG (Wittkopp P)
Clark AG (Payseur B)
Clark AG (McGraw L)
Clark AG (Schlenke T)
Clark AG (Montooth K)
Clark AG (Lazzaro B)
Clark AG
Clarke BB (Crouch J)
Coat G (McNeal J)
Coats W (Snoeyenbos-West O)
Coffin JM (Achaz G)
Cole J (Snoeyenbos-West O)
Colgan DJ (Belov K)
Collier S (Ray D)
Collins LJ (Penny D)
Corces VG (Ramos E)
Cordaux R
Cordaux R (Hedges D)
Cosgrove DJ (Sampedro Jimenez J)
Costas B (Doherty M)
Costas B (Pirog K)
Coyle S (Kroll E)
Coyne JA (Greenberg A)
Cracraft J (Barker K)
Cracraft J
Crandall KA (Pascual M)
Crawford AJ
Crawford DL (Whitehead A)
Crawford DL (Phinchongsakuldit J)

Crawford DL
Crease TJ
Croom HB (McNeal J)
Crouch JA
Crumpler N (Teeling E)
Crumpler NJ (Murphy W)
Cui L (Ricker J)
Cui L
Cui L (Beckmann K)
Cui L (Leebens-Mack J)
Cusack BP
Cutter AD
Dagan T
Das J
Daugherty CH (Miller H)
David RM (Ko W)
David VA (Murphy W)
Davis GB (Song N)
Davis JC (Salzman Y)
Davis RW (Dean J)
de Jong P (Negre B)
de Jong P (Gonzalez J)
de Jong WW (Madsen O)
Dean EJ
Dehal PS
Deigendesch N (Rothenburg S)
Del Bianco F (Carboni A)
Delaney TK
Delmotte F (Rispe C)
Delprat A (Gonzalez J)
Denawa MD (Okamura H)
Denawa M
Denver DR
Denver DR (Simonelic K)
dePamphilis CW (Cui L)
dePamphilis C (Zahn L)
dePamphilis CW (Kong H)
dePamphilis CW (Sampedro Jimenez J)
dePamphilis CW (McNeal J)
dePamphilis C (Ricker J)
dePamphilis CW (Wall K)
dePamphilis CW
dePamphilis C (Beckmann K)
dePamphilis C (Beckmann K)
dePamphilis CW (Leebens-Mack J)
Dermitzakis ET
Dewar K (Kitano T)
Dietrich FS (Stajich J)
Doherty M
Doniger SW
Doolittle WF
Druzhinina IS
Dunbar HE (Moran N)
Dunn B (Kroll E)
Durand D (Song N)
Durand D (Raghupathy N)
Durand MD (Vernot B)
Duret L (Meunier J)
Durnford DG (Rissler H)
Durrett R (Schmidt D)
Dushoff J (Fraser H)
Dyer KA
Dykhuizen DE (Stoebel D)
Eanes WF (Merritt T)
Eddy S (Wessler S)
Edwards SV (Shedlock A)
Eickbush TH (Zhang X)
Eisen MB (Gasch A)
Eisen MB (Moses A)
Eisen MB (Pollard D)
Eisen MB (Fay J)
Eizirik E (Roca A)
Eldridge MD (Browning T)
Eldridge MD (Bulazel K)
Eldridge MD (Belov K)
Eldridge MD (Ferreri G)
Ellegren H (Webster M)
Ellis RE
Ellis RE (Cho S)
Emerson JJ
Engelhardt BE
Erickson BK
Erickson BK (Sanchez M)
Esteves PJ
Excoffier L (Dermitzakis E)
Ezawa K
Ezawa K (Saitou N)
Fain MG
Fay JC (Doniger S)
Fay JC
Feijao A (Battistuzzi F)
Ferdig MT
Ferrand N (Melo-Ferreira J)
Ferrand N (Esteves P)
Ferrand N (Geraldes A)
Ferreri GC. C (Marzelli M)
Ferreri GC (Bulazel K)
Ferreri GC
Feschotte C (Pritham E)
Feschotte C
Feschotte C (Cedric F)
Field D (Wall K)
Figueroa F (Takezaki N)
Filipski A (Kumar S)
Firestein S
Fisher CR (Carney S)
Fleischer RC (Ketcham K)
Flockerzi A (Mayer J)
Foissner W (Griffin A)
Fong A (Archibald J)

Lim S (McNeal J)
Lin A (Su B)
Lin CF
Lin C-F (Ko W)
Lin YP (Ting C)
Lin Z
Lindberg D (Boore J)
Lipps HJ (Landweber L)
Long M (Emerson J)
Long M (Wang W)
Lowenhaupt K (Rothenburg S)
Lu G
Lu Y
Lu Y (Hoberman R)
Lucas OJ
Luo J
Lynch M (Simonelic K)
Lynch M (Denver D)
Ma H (Kong H)
Ma H (Zahn L)
Ma H (Nam J)
Ma H (Lin Z)
Ma H (Ricker J)
Ma H (Cui L)
Ma H (Beckmann K)
Macey JR (Boore J)
Mack JA
MacKinnon M (Archibald J)
Madsen O
Mager DL (van de Lagemaat L)
Mager DL (Baillie G)
Mahfooz NS (Popadic A)
Maier RM (Tillich M)
Makalowski W (Gotea V)
Makalowski W (Thornburg B)
Makalowski W (Lin Z)
Makalowski W (Chalkia D)
Makalowski W (Lin C)
Makarova KS
Makova KD (Taylor J)
Makova K (Goetting-Minesky P)
Makova KD
Maldarelli F (Achaz G)
Maltsev N (Prachumwat A)
Man O (Gilad Y)
Mannaert A
Mao X
Margulies EH (Hayakawa T)
Maria R (Roca A)
Mariño-Ramírez L (Jordan K)
Marland E (Prachumwat A)
Marquard L (Madsen O)
Martins LO
Marzelli ME
Marzelli M (Ferreri G)
Masta S (Boore J)
Mateos M (Lawton B)
Matheson JW (Penny D)
Mathews S
Matusda M (Tanaka H)
Matzkin LM (Merritt T)
Mayer J
McCullough HL (Fay J)
McCusker JH (Dean J)
McDonald JF (Ganko E)
McGrath CL
McGraw LA
McHugh DE (Bauchet M)
McManus GB (Pirog K)
McManus G (Doherty M)
McNeal JR
McNeal J (Leebens-Mack J)
McPheron BA (Barr N)
McPheron BA (Schwarz D)
Meade K (Bauchet M)
Medina M (Boore J)
Medstrand P (van de Lagemaat L)
Medstrand P (Mayer J)
Meese EU (Mayer J)
Mei R (Halder I)
Meisel RP
Mellors JW (Achaz G)
Melo F (Opazo J)
Melo-Ferreira JF
Mena-Ali JI
Mercer DM (Sorenson M)
Merritt TJ
Merriwether DA (Norton H)
Messing J (Swigonova Z)
Mestres F (Pascual M)
Metcalfe CJ (Bulazel K)
Meunier J
Meyer A (Salzburger W)
Meyer A (Braasch I)
Meyer A (Barluenga M)
Meyer A
Meyer A (Luo J)
Meyer CI
Mgone CS (Norton H)
Mikkelsen T (Hayakawa T)
Mile MA (Gaunt M)
Miles GA (Gaunt M)
Miller HC
Miller ST (Das J)
Mineta K (Gojobori T)
Mirkin BG (Makarova K)
Mishra PK
Miyake T (Ota T)
Moline J (Simon D)
Mollenbeck M (Landweber L)

Wang CG (Lu G)
Wang H (Xing J)
Wang J (Su B)
Wang W (Yan L)
Wang W
Wang X (Sanchez M)
Wang X
Wang X (Erickson B)
Wang YQ (Su B)
Wang Y (Sanchez M)
Wang Y (Erickson B)
Ward S (Cutter A)
Watabe T
Watabe T (Kitazoe Y)
Webster MT
Wei Zhang (Wang G)
Weinreich DM
Weise V (Panopoulou G)
Weiss KM (Kawasaki K)
Weitz JS (Cavalcanti A)
Wells MA (Goodisman M)
Wendel JF (Adams K)
Wessler SR (Pritham E)
Wessler SR (Feschotte C)
Wessler SR
Wheeler DE (Goodisman M)
Whelan S (Bofkin L)
Whelan S (Goldman N)
White B (Ray D)
White KP (Rifkin S)
Whitehead A
Wilcox JL (Moran N)
Wildman DE
Wilkins E (Snoeyenbos-West O)
Williams BL
Williamson S
Wilson AC
Wittkopp PJ
Wlasiuk G
Wolf YI (Jordan K)
Wolf YI (Makarova K)
Wolf YI
Wolf YI (Koonin EV)
Wolfe KH
Wolfe KH (Cusack B)
Wolfner MF (McGraw L)
Wolfner MF (Wong A)
Won Y
Wong A
Wong LC (Landweber L)
Wong S (Wolfe K)
Woo PCY (Cai J)
Woodruff RC (Gong Y)
Woodruff RC (Gu S)
Woodward R (Belov K)
Woolfit M
Wu CI (Greenberg A)
Wu G
Wyman SK (Leebens-Mack J)
Xiaohong Zhang (Wang G)
Xing J (Ray D)
Xing J (Batzer M)
Xing J (Cordaux R)
Xing JC
Xing J (Hedges D)
Yamazaki T (Ishiyama H)
Yamazaki T (Goto H)
Yang S (Su B)
Yang S (Makova K)
Yeo M (Gaunt M)
Yoon HS (Hackett J)
Yoon HS
Young G (Mower J)
Young ND (McNeal J)
Yuen KY (Cai J)
Yujin Sun (Wang G)
Zahn LM
Zaleska-Rutczynska Z (Takezaki N)
Zenger KR (Browning T)
Zhai SK (Esteves P)
Zhang HM (Gu X)
Zhang J (Cho S)
Zhang J
Zhang J (Wang X)
Zhang J (Grus W)
Zhang J (Podlaha O)
Zhang P
Zhang Q (Ricker J)
Zhang X
Zhang X (Ricker J)
Zhang X (Firestein S)
Zhang X (Firestein S)
Zhang X (Wessler S)
Zhang Y (Yan L)
Zhang Y (Leebens-Mack J)
Zhang ZQ (Gu X)
Zhao S (Shedlock A)
Zhao Z
Zhaxybayeva O
Zheng HK (Su B)
Zhu CT (Merritt T)
Zhu L
Zufall RA
Zufall RA
Zufall RA (McGrath C)
Zupunski V

Subject Index

(Names of presenting authors in parentheses)