MOLECULAR EVOLUTION

Organizers: Michael Clegg and Stephen O'Brien

February 27-March 6, 1989

Plenary Sessions Page

Poster Sessions

Molecular Evolution

Genome Evolution-I

CF 001 DECIPHERING OF CODING SEQUENCES ACCORDING TO THE PRINCIPLE OF PERIODIC-TO-CHAOTIC TRANSITION, Susumu Ohno, Department of Theoretical Biology, Beckman Research Institute of The City of Hope, Duarte, CA 91010. Each coding sequence is a finite resource as to the number and composition of four bases. Accordingly, the excessive recurrence of one base dimer entails the noticeable under-representation by the other, so that if the former is the same in most, if not all, of the coding sequences, the latter too must necessarily be the same in all. Indeed, the dimer analysis of divergent coding sequences of varied base compositions derived from plants to man revealed that two dimers, TG and CT, always recur at the rates far higher than expected and that these excesses are invariably compensated by gross underrepresentation by two other dimers; TA and CG. In those coding sequences encoding proteins rich in secondary structures, two excessive dimers invariably combined to make C T G, the most frequently recurred base trimer, thus, accounting not only for the abundance of Leu in these proteins but also for the universal preponderance of C T G among six Leu codons. The paucity of T A accounts for the fact that the content of Tyr in most proteins remains below 5%. Arg contents of proteins, on the other hand, can be higher in spite of the similar paucity of C G, this is because of the presence of A G A and A G G as additional Arg codons. In most coding sequences, C T G is incorporated in the primordial repeating units which are most often tetramer (e.g., G C T G) or pentamer (G C C T G). The original periodicity has been degenerating in accordance with the rule of periodic-to-chaotic transition by the golden mean; the tetrameric periodicity degenerating into heptameric and then monodecameric periodicities.

Evolutionary Analysis of Molecular Data

CF 002 VARIATION IN SEQUENCE DISTANCES. B.S. Weir, Department of Statistics, North Carolina State University, Raleigh, NC 27695.

Advances in molecular biology have made it easier to sequence more than one representative of a given region from a species or population. Statistics constructed from the observed base sequences can be used to make inferences about evolutionary events. This paper is concerned with finding the sampling properties of such statistics, so that the quality of evolutionary inferences may be assessed. Emphasis is given to the variance of a measure of distance between two sequences, and all sources of variation are identified. These sources include both "statistical" sampling, which is under the control of the experimenter, and "genetical" sampling, which results from the choice of genes transmitted from parent to offspring. The latter component comes into play when more than one sequence is available from a population. A general variance formula follows from that for the average heterozygosity over a number of loci, and includes terms for populations, individuals within populations, loci and loci by individuals within populations. The magnitudes of the components follow from the theory of descent measures, and allow the effects of different sampling strategies to be evaluated.

Reconstructing Population History From Molecular Data

CF 003 GENE TREES AND ORGANISMAL HISTORIES-- A PHYLOGENETIC APPROACH TO POPULATION BIOLOGY, John C. Avise, Department of Genetics, University of Georgia, Athens, GA 30602. A "gene tree" is the phylogeny of alleles or haplotypes for any specified stretch of DNA. Surveys of haplotype diversity in the rapidly-evolving mitochondrial DNA (mtDNA) molecule have provided the first extensive empirical data suitable for estimation of gene trees on a microevolutionary (intraspecific) time scale. The relationship between phylogeny and the geography of populations constitutes the phylogeographic pattern for any species. Concordances in the phylogeographic patterns across species almost certainly evidence the effects of shared historical biogeographic factors in moulding intraspecific pedigrees.

These conclusions will be supported by examples from ongoing studies of mtDNA phylogeography of a number of species in the southeastern United States. Marine or estuarine species as diverse as the Seaside Sparrow, Horseshoe Crab, American Oyster, and various fishes, exhibit geographically-concordant phylogenetic breaks distinguishing most Atlantic Coast from Gulf Coast populations. The phylogenetic breaks often localize to the area of Cape Canaveral, Florida. These shared patterns are probably due to historical barriers to gene flow related to changes in sea level, ocean currents, and habitat availability associated with Pleistocene glaciations. Populations that were formerly in continuous contact may have been isolated by emergence of the Florida Peninsula. The effects of Pleistocene events are also registered in the mtDNA genomes of a number of freshwater fish species, whose populations consistently show major east-west phylogeographic breaks in the area of the Apalachicola or Alabama/Tombigbee drainages. Interdrainage dispersal of freshwater fishes from eastern and western Pleistocene refugia probably accounts for these results.

Most previous concerns with the geographic distributions of genetic characters have emphasized explanations involving either natural selection or genetic drift. The phylogenetic content of mtDNA provides a new perspective on these deliberations. Study of the phylogeographic relationships of conspecific populations should be of general significance to evolutionary biology by: (1) providing a phylogenetic backdrop against which to interpret the distributions of morphological, behavioral, or other traits; (2) facilitating reconstruction of the biogeographic histories of regional biotas; and (3) providing a bridge between the nominally separate disciplines of systematics and population genetics.

CF 004 AN EVALUATION OF PHYLOGENIES OF DROSOPHILA SPECIES GROUPS BASED ON GENETIC DISTANCES, Ross J. MacIntyre, Glen E. Collier and David Featherston,Section of Genetics and Development, Cornell University, Ithaca, NY 14853, and Department of Biology, Illinois State University, Normal, IL 61761

The accuracy of genetic distance based on phylogenetic trees for species groups within the genus Drosophila has been assessed by comparing the topologies of such trees to topologies of trees based upon other kinds of data for these same species. The alternative data types include cytology, hybridization tests, comparative morphology, and zoogeography. The tree topologies were compared by calculating a "coefficient of distortion" (Camin and Sokal, 1965 and Farris, 1973) for each comparison and by determining the probability that the difference (d) seen for each comparison is due to chance (Penney et al., 1982). A table of random coefficients of distortion has been derived by computer simulations. In only four of the seventeen comparisons of Drosophila species groups does the coefficient of distortion exceed 0.4 and the probability of d greater than 0.057. For two of these four cases, the placement of a single species is responsible for virtually all of the coefficient of distortion. The remaining two cases involve relatively large numbers of species (11 and 16) and rely upon genetic distance estimates based upon relatively few loci (7 and 10). The role that these and other factors may play in determining the accuracy of tree topology is discussed. These comparisons support the contention that genetic distance based trees are useful tools for the evolutionary biologist and at this point in time certainly as useful as trees based on DNA sequence information.

Molecular Evolution

Molecular Diversity in Populations

CF 005 CONTRASTING LEVELS AND PATTERNS OF DNA SEQUENCE VARIATION IN DROSOPHILA SPECIES: INFERENCES ON ROLES OF SELECTION AND DRIFT, Charles F. Aquadro, Section of Genetics and Development, Cornell University, Ithaca, NY 14853. Recent analyses of naturally occurring DNA sequence and restriction map variation in several species of *Drosophila* have shown that DNA variation is 4-6 times higher in *D. simulans* and *D. pseudoobscura*, compared to *D. melanogaster*. This difference is particularly striking in light of the contrast between the closely related species *melanogaster* and *simulans*: the latter species has been found to be less polymorphic and geographically differentiated for allozymes and morphology. Transposable elements have also been found to contribute significantly less to naturally occurring sequence variation in both *D. simulans* and *D. pseudoobscura* compared to *D. melanogaster*. These contrasts serve as a warning that we must be cautious in making generalizations from results in D. melanogaster. The striking differences in levels of DNA restriction site, protein and transposable element polymorphism between particularly *D. melanogaster* and *D. simulans* has been hypothesized to be due to different effective population sizes in the species and slight deleterious selection acting on the majority of protein variants and transposable element insertions (Aquadro, Lado and Noon, 1988, Genetics 119:875-888). Tests of this hypothesis using DNA sequence data from the *rosy* and *period* genes will be discussed. These tests are based on the comparision of frequencies of synonymous to nonsynonymous substitutions within and between the two species, and on the frequency spectra of sequence variants within each species. In addition, comparison of intra- to interspecific variation indicates differences in constraint, mutation rate and natural selection all contribute to strongly nonrandom distributions of restriction site and insertion/deletion polymorphisms across genes and chromosomes in all species.

CF 006 MOLECULAR GENETIC ANALYSIS OF LION SOCIAL STRUCTURE, Craig Packer, Department of Ecology and Behavioral Biology, University of Minnesota, Minneapolis, MN 55455; Dennis A. Gilbert, Stephen J. O'Brien, National Cancer Institute, Frederick, MD 21701-1240.

Lions live in stable social groups ("prides") that consist of 1-18 adult females, their dependent offspring and a coalition of 1-9 males. Based on 22 yrs of behavioral and demographic data collected in the Serengeti National Park and Ngorongoro Crater, Tanzania, lion biologists have made the following assertions concerning genetic relationships of pridemates: 1) Males of the resident coalition sire all cubs born during their tenure. 2) Females of the same pride are always related. 3) Only a proportion of male coalition partners are related. Blood samples were collected from approximately 200 individuals in order to test these hypotheses, and also to determine the extent of genetic differentiation between prides in the same population and between adjacent populations. Isozyme polymorphisms were insufficient to deduce paternity, but were useful in measuring the genetic differentiation between the Serengeti and Ngorongoro populations. Using hypervariable probes developed in the domestic cat, DNA fingerprinting was employed to measure the reproductive success of resident males. The DNA techniques could specify paternity, even when coalition partners were closely related. The analysis showed that while the majority of cubs were fathered by the resident coalition, a small proportion of cubs were fathered by extra-pride males. Further, several cubs were found to be unrelated to all of the adults in the pride and thus were apparently adopted by the females of the pride. Finally, littermates included half-siblings, full siblings and identical twins. Therefore DNA analysis has revealed patterns of relatedness far more complex than previously supposed. Its utility in inferring lower degrees of relatedness between pridemates and members of adjacent prides will also be discussed.

Interaction of Viral and Host Genomes in Evolution

CF 007 GENETIC EPIDEMIOLOGY OF A NEUROTROPIC RETROVIRUS IN LC WILD MICE, Murray B. Gardner, M.D., Department of Medical Pathology, University of California, Davis School of Medicine, Davis, CA 95616

A fatal hind leg paralysis occurs spontaneously in a small minority (~10%) of aging, lymphoma-prone wild mice (Mus musculus) from an isolated squab farm near Lake Casitas (LC) in southern California. The disease, called "Spongiform Polioencephalopathy" is caused by an infectious, ecotropic murine leukemia virus (MuLV) acquired by maternal congenital infection (1). The disease is characterized by a noninflammatory, nonimmunogenic, direct retrovirus injury to anterior horn neurons in the lumbosacral spinal cord. Susceptibility to the neurologic disease in wild mice depends primarily on inheritance of an ecotropic MuLV resistance gene, initially called Akvr-1^R, segregating in this outbred population (2). The allele frequency of Akvr-1^R is 0.47 in LC mice and the genotype frequencies observed do not vary significantly from expectations of the Hardy-Weinberg equilibrium. Thus, the probable frequency of LC mice that contain at least one Akvr-1^R allele is 0.72; these mice would be expected to be resistant to infection with the ecotropic, paralytogenic MuLV. This restriction gene has been molecularly cloned and sequenced; it represents a defective endogenous retrovirus with a complete envelope gene which expresses a glycoprotein (gp70) related to the ecotropic MuLV of inbred AKR mice (3). This glycoprotein blocks the ecotropic MuLV receptors on the cell surface thus interfering with entrance of ecotropic MuLVs which all use the same receptor (4). The Akvr-1^R allele does not block another class (amphotropic) of weakly lymphomogenic, nonparalytogenic MuLV, highly prevalent in LC wild mice. The Akvr-1^R gene has not been found in North American laboratory mice, presumably because of the narrow genetic base from which they were derived. The Akvr-1 restriction gene is allelic with and phenotypically and sequence identical to the FV-4 MuLV restriction gene previously mapped to chromosome 12 in Japanese wild mice (Mus molossinus) (5-7). This chromosomal locus was probably acquired in recent times by interbreeding of these feral species.

1. Gardner, M.B. Rev. Infect. Dis. 7:99-110, 1985.
2. Gardner, M.B., S. Rasheed, B.K. Pal, J.D. Estes, S.J. O'Brien. PNAS 77:531-535, 1980.
3. Dandekar, S., P. Rossito, S. Pickett, G. Mockli, H. Bradshaw, R. Cardiff, M. Gardner. J. Virol. 61:308-314, 1987.
4. Rasheed, S., M.B. Gardner. Int. J. Cancer 31:491-496, 1983.
5. O'Brien, S.J., E.J. Berman, J.D. Estes, M.B. Gardner. J. Virol. 47:649-651, 1983.
6. Odaka, T.H., H. Ikeda, H. Yoshibura, K. Moriwaka, S. Suzuki. JNCI. 57:1123-1127, 1981.
7. Ikeda, H.F., F. Laigret, M.A. Martin, R. Repaske. J. Virol. 768-777, 1985.

CF 008 **THE EVOLUTION AND MAINTENANCE OF TRANSPOSABLE ELEMENTS IN BACTERIAL POPULATIONS**, Bruce R. Levin, Department of Zoology University of Massachusetts, Amherst MA 01003

Transposable genetic elements, "transposons" are DNA molecules that are capable of moving to different places on the chromosomes of bacteria or their plasmids and prophage, or between chromosomes and these accessory elements. Is this capacity for transposition sufficient to explain the evolution and maintenance of bacterial transposons? Theoretical studies suggest that there are conditions where transposons that move by a replicative process can be maintained by hitchhiking on conjugative plasmids, even when these transposons confer a disadvantage on their host. However, these conditions for maintaining transposons as "parasites" in single populations are restrictive and, even when they obtain, the rate of ascent of these parasitic genetic molecules appears to be too low to account for their evolution and persistence. In this report, I consider alternative models for establishment and maintenance of transposons in the absence of positive selection for their carriage. I discuss the fit of these theoretical results to the dynamics of transposons in experimental populations of E. coli and the distribution of these moveable genetic elements in E. coli from natural sources.

Evolution of Organelle Genomes

CF 009 UNSTABLE CHLOROPLAST GENOMES AND TRANSFER OF GENE FUNCTIONS FROM THE CHLOROPLAST TO THE NUCLEUS, Jeffrey D. Palmer, Sandra L. Baldauf, Patrick J. Calie and Claude W. dePamphilis, Department of Biology, University of Michigan, Ann Arbor, MI 48109

The chloroplast DNAs of land plants are, with rare exception, highly conserved in size, gene content, and gene order. We have identified two independent lineages of flowering plants - one photosynthetic (geraniums, Pelargonium) and the other nonphotosynthetic (beechdrops, Epifagus) - whose chloroplast DNAs deviate from this conservative pattern and have sustained major rearrangements affecting the structure and function of important plastid genes. The most spectacular rearrangement in geranium involves the fragmentation of the rpoA gene, which encodes the alpha subunit of chloroplast RNA polymerase. About 40% of the normal rpoA coding sequences is deleted from geranium chloroplast DNA, and the remaining 60% of this 1 kb gene is fragmented into five short segments scattered over a 6 kb region of the chromosome. We cannot find an intact rpoA gene in either the mitochondrial or nuclear genomes, raising the question of whether any functional alpha subunit is produced in geranium.

Physiological and biochemical measurements indicate a complete absence of photosynthetic activity in the root parasite Epifagus. Yet this plant has rudimentary plastids containing high levels of plastid DNA. However, the plastid genome has sustained massive deletions, including the loss of such photosynthetic genes as rbcL. More surprisingly, the genome also appears to lack a number of genes encoding subunits of the transcriptional and translational apparatus. These preliminary results suggest the possibility that the plastid DNA in Epifagus is in essence a vestigial "pseudogenome". If so, then one can infer that plastid genes are not essential for any metabolic processes in the plastid other than photosynthesis and gene expression.

We have found the first cases of chloroplast genes that have been transferred to the nucleus as functional entities recently, i.e. during land plant evolution. The gene rpl22, encoding chloroplast ribosomal protein L22, is encoded in the nucleus in all legumes, but in the chloroplast in all other plants. Hence, this gene was transferred from the chloroplast to the nucleus about 50 million years ago, in the common ancestor of legumes. The gene tufA, which encodes the chloroplast protein synthesis factor EF-Tu, was transferred to the nucleus about 400 million years ago, in the common ancestor of land plants. As part of their functional adaptation to the genetic environment of the nucleus/cytoplasm, the nuclear forms of these genes have acquired new upstream sequences, including promoters and transit peptides.

Evolution of Immune Response

CF 010 HLA EVOLUTION, Walter F. Bodmer and Julia G. Bodmer, Imperial Cancer Research Fund, London

The HLA region is now known to be about 3,500 kb long and, in addition, to the Class I, Class II and some complement genes, has now been shown to contain many other genes with apparently unrelated functions. Only some of the Class I, Class II, and to a lesser extent complement genes show high levels of polymorphism. This is most probably generated by immune response differences associated with resistance to infectious diseases. Some, but by no means all, alleles of the HLA system are old, based on analysis of nucleotide pair differences. Detailed studies of sequence differences between alleles, in relation to the structure of the Class I and Class II molecules, leads to an insight into possible influences on immune response. The fact that some of the extracellular matrix receptors are members of the immunoglobulin super gene family, which includes, HLA Class I and Class II products, supports the suggestion that the common origin of this family is through functions required for specific cellular interactions during differentiation and development, before the evolution of an immune system.

Prokaryotic Origins

CF 011 ORIGIN AND EVOLUTION OF SEX IN PROKARYOTES, Richard E. Michod
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, 85721. Sex involves two basic components (i) recombination, in the sense of breakage and rejoining of DNA (or equivalent) molecules and (ii) outcrossing, which refers to the property that the two molecules involved in recombination come from different individuals. Sex exists in a variety of different forms in prokaryotes, from recombination after multiple infection in viruses to transformation in bacteria. Arguments from first principles, as well as examples from the simplest living organisms, will be presented in support of the hypothesis that sex arose very early in the history of life and is currently maintained for the function of genetic repair (1, 2). Experimental tests of this hypothesis in the case of transformation in *Bacillus subtilis* show that competent (sexual) wild-type cells survive UV damage better than noncompetent cells, if they are allowed to undergo transformation with homologous DNA after the UV treatment (3). The advantage of competent cells goes away if they are transformed before UV treatment. These results persist in various strains which are deficient in excision repair or SOS-like repair but go away if the competent cells are transformed with non-homologous plasmid DNA (4). These results support the hypothesis that transformation evolved for the function of bringing DNA into the cell for use as template in recombinational repair.

1. Bernstein, H., H. C. Byerly, F. A. Hopf, R. E. Michod. 1984. Origin of sex. *J. Theor. Biol.* 110:323-351.
2. Bernstein, H., F. A. Hopf, R. E. Michod. 1987. The molecular basis of the evolution of sex. *Advances in Genetics*, 24:323-370.
3. Michod, R. E. , M. F. Wojciechowski and M. A. Hoelzer. 1988. DNA repair and the evolution of transformation in the bacterium *Bacillus subtilis*. *Genetics* 118:31-39.
4. Wojciechowski, M. F., M. A. Hoelzer, R. E. Michod. DNA repair and the evolution of transformation in *Bacillus subtilis*. II. role of inducible repair. *Genetics*, In press.

Transposon Evolution

CF 012 TRANSPOSABLE ELEMENTS IN NATURAL POPULATIONS, Brian Charlesworth,
Department of Ecology and Evolution, The University of Chicago, Chicago, IL 60637

Transposable elements are present in the genomes of organisms from bacteria to man. They are characterized by their presence in multiple, dispersed copies within their host genomes. Mathematical models of their population dynamics indicate that transpositional increase in the mean number of copies per individual of members of a given family can be opposed by a variety of forces, including self-regulation of the rate of transposition, selection against insertional mutations, and selection against chromosome rearrangements induced by recombination between homologous elements located at different chromosomal sites. If these forces are sufficiently powerful, then the mean copy number is much less than the number of occupable chromosomal sites, so that the mean element frequency per site in the population is low. Data from bacteria, yeast and *Drosophila* show this pattern. Quantitative analyses of population data on *Drosophila* from *in situ* hybrization studies indicate that transposition must in fact be actively opposed, and that selection against insertional mutations is unlikely to be the primary agent. There is a large excess of elements at the base of the X chromosome, and a smaller excess at the tip, both regions of reduced recombination. This suggests that the last of the above mechanisms of copy number control may be operating. The data also show that the sites of occupation by elements appear to be randomly distributed along the chromosome, except where recombination is infrequent, and that (as postulated in the models) there are no correlations in element frequencies between closely linked sites. There are, however, correlations between different element families in the sites that are occupied.

CF 013 EVOLUTION OF *P* ELEMENTS IN THE GENUS *DROSOPHILA* , Margaret G. Kidwell, Kenneth R. Peterson and Stephen B. Daniels, Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721. The *P* family of transposable elements has a broad but patchy distribution in the genus *Drosophila* (1, 2). Complete autonomous elements are restricted to some populations of *D. melanogaster* and probably one or two species of the distantly-related *willistoni* subgroup. Smaller sequences with homology to *P* have a much broader distribution within the genus and are presumably relics of a deletion process. We have cloned and sequenced a complete *P* element (P_w) from a recently-derived natural population of *D.willistoni* . The DNA sequence was identical to that published for a complete *D. melanogaster P* element (3), except for an A to G transition at position 32 which is located upstream of the first exon. Following microinjection, P_w has been shown to be functional in *D. melanogaster* using several criteria for mobility. These results are in sharp contrast to those obtained with *P* elements from another *willistoni* subgroup species, *D. nebulosa,* which had diverged more than 5% from the *D.melanogaster P* element and could not produce transposase (4). An explanation of the near identity of the *D. melanogaster* and *D.willistoni P* elements in terms of extreme conservation of sequence homology during the estimated 50 million years since the divergence of the two species is not well supported by data on the phylogenetic distribution of the element family. An alternative hypothesis of recent interspecific transfer of *P* between *D.willistoni* and *D. melanogaster* is consistent with the phylogenetic distribution and is strongly supported by earlier published data on the temporal and geographical distribution of this element in *D. melanogaster* (5).

References: (1) Lansman, R. A., Stacey, S. N., Grigliatti, T. A. and Brock, H. W., 1985 Nature 318: 561-563; (2) Daniels, S. B., Strausbaugh and L. D., 1986 J. Mol. Evolution 23: 138-148; (3) O'Hare, K., Rubin, G. M., 1983 Cell 34: 25-35; (4) Lansman, R. A., Shade, R. O., Grigliatti, T. A. and Brock, H. W., 1987 Proc. Nat. Acad. Sci., U. S. 84: 6491-6495; (5) Kidwell, M. G., 1983 Proc. Natl. Acad. Sci. USA 80: 1655-1659.

CF 014 THE DISTRIBUTION AND CONTAINMENT OF DROSOPHILA DNA PARASITES, Charles H. Langley, Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709. Transposable elements of many different families are found throughout the euchromatic regions of the Drosophila genomes in natural populations of several species. The average number along the chromosome ranges from 0.0001 to 0.02 per kilobase. Transposable element insertions are typically found in the intergenic regions of the DNA. The frequencies in the population of the individual insertions are generally quite low. In a few instances, apparently identical insertions have been recovered from natural populations. These cases are compared to the situation in mammalian populations where virtually all insertions are fixed in the population and even the genus. Several lines of evidence support the hypothesis that the main mechanism responsible for the containment of transposable element copy number is the aneuploidy associated with unequal recombination arising from exchange between transposable elements at different sites in the genome.

Genome Evolution-II

CF 015 PHYLOGENY RECONSTRUCTION FROM DNA COMPARISONS, Charles G. Sibley, Tiburon Center for Environmental Studies, San Francisco State University, Tiburon, CA 94920

DNA-DNA hybridization measures the degree of base sequence similarity between the single-copy DNAs of different species. Congruence between DNA comparisons and evidence from independent sources is probably the best test that the correct phylogeny has been inferred. In studies of the living birds of the world the DNAxDNA results are highly congruent with evidence from morphology for recently diverged taxa. The degree of congruence between DNA hybridization evidence and traditional ideas of phylogenetic relationships declines as the genetic distance between taxa increases. Presumably this is due to the combined effects of convergence and divergence which make the interpretation of morphological characters more difficult. It may also be due, in part, to experimental errors correlated with increasing genetic distance. Examples will be presented.

Retrovirus Evolution

CF 016 RETROVIRUS GENE TRANSFER BETWEEN SPECIES AND THE EVOLUTION OF VIRAL GENES IN MAMMALS, Raoul E. Benveniste and G.L. Wilson, Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, MD 21701

Endogenous retroviruses are present in the somatic and germ cell DNA of all animals of a species, and have remained stable enough to make it possible to detect related nucleic acid sequences among mammalian species that diverged millions of years ago. There have also been several examples of retrovirus transfer between species, with subsequent incorporation into the germ line.

A transfer of retrovirus from the ancestors of squirrel monkeys to the ancestors of skunks was shown initially by the hybridization of a squirrel monkey retrovirus DNA probe to skunk cellular DNA. The lack of hybridization to the skunk's supposedly close relatives, the mustelids, led to an examination of carnivore phylogeny by DNA hybridization. Skunks (Mephitis spp and Spilogale spp) are shown to have diverged from other mustelids 40 m.y.b.p. suggesting that genetically distinct skunk ancestors existed throughout the Oligocene. The other carnivore families (canids, ursids, pinnipeds, procyonids, viverrids, hyaenids, felids) are shown to have a nearly contemporaneous appearance approximately 40 m.y.b.p. The viverrids and mustelids contain taxa with ancient divergence times, whereas the other families contain species that have more recent origins.

The transfers of retroviruses between mammalian species and a geographic component present in primate retrovirus evolution places an evolutionary perspective on the AIDS infection in man and on the presence of simian immunodeficiency viruses in several African primate species.

CF 017 THE REGULATORY GENES OF THE HUMAN T CELL LEUKEMIA VIRUS HTLV-1 AND HUMAN IMMUNODEFICIENCY VIRUS HIV-1, William Haseltine, Dana-Farber Cancer Institute, Division of Human Retrovirology, 44 Binney Street, Boston, MA

The human retroviruses differ from other retroviruses previously characterized in mammals by the existence of multiple regulatory genes. These genes govern the expression of viral proteins. These genes may also be important for the establishment of latency and the release from latency during the course of natural infection. Additionally, the regulatory genes they have pleotropic affects on cellular as well as viral genes. Alteration of cellular gene function via viral trans-acting genes may account for some of the pathogenic properties of the human retroviruses.

Genome Evolution

CF 100 THE COMPLETE cDNA SEQUENCE FOR CHICKEN PROTHROMBIN: IMPLICATIONS FOR THE EVOLUTION OF PROTHROMBIN. DAVID K. BANFIELD, DAVID M. IRWIN, DAN A. WALZ *, AND ROSS T. A. MacGILLIVRAY DEPARTMENT OF BIOCHEMISTRY, UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER, CANADA V6T 1W5 AND * DEPARTMENT OF PHYSIOLOGY, WAYNE STATE UNIVERSITY, DETROIT, MICHIGAN.
Prothrombin is synthesized in the liver as a prepro-protein, and it is a member of a family of trypsin-like serine proteases. The active protease thrombin, has a critical role the final step of the coagulation cascade. Like many of the proteins involved in hemostasis, prothrombin contains a number of distinct structural and functional domains. These domains include: a propeptide required for recognition by a vitamin K-dependant carboxylase, a region containing gamma carboxylated glutamic acid residues (Gla domain), two Kringle domains, and the serine protease domain. Comparisons of the predicted amino acid sequence of chicken prothrombin with those of human and bovine prothrombin reveal several interesting regions of conservation. Among the regions of high amino acid identity are the propeptide (11 of 17 residues identical), the Gla domain (26 of 34), and the A chain region of thrombin (16 of 18). Extensive amino acid identity is found in the active site region (a block of 17 identical residues) as well as the primary substrate binding pocket (a block of 27 identical residues). In addition, the amino acid motif L-ENYCRNPD found in both Kringles of chicken prothrombin is identical to the consensus sequence of all other Kringles sequenced to date, however, the human and bovine Kringle sequences (L-ENFCRNPD) vary slightly. The high degree of amino acid sequence identity of prothrombin in these three species strongly suggests the molecule is under rigorous functional constraints. To better understand the biological and evolutionary significance of these regions of high amino acid identity, we are presently characterizing prothrombin from the Pacific Hagfish, Eptatretus stouti.

CF 101 RESTRICTION ENZYME MAPPING OF RIBOSOMAL DNA TO DISTINGUISH BETWEEN FASCIOLID (LIVER FLUKE) SPECIES, David Blair and Donald P. McManus, Department of Pure and Applied Biology, Imperial College, Prince Consort Road, London SW7 2BB, Great Britain.

Recognition sites for 9 different restriction endonucleases were mapped on rDNA genes of fasciolids. Fragments were separated in agarose gels, transferred to nylon membranes by Southern blotting, and probed sequentially with 3 different cloned probes derived from, and known to span between them, the entire rDNA repeat unit in Schistosoma mansoni. Eighteen recognition sites were mapped for Fasciola hepatica and 17 for F. gigantica and Fascioloides magna. Each species had no more than 2 unique recognition sites; the remainder being common to one or both of the other species. No intraspecific variation was noted in F. hepatica (11 samples mapped: 2 from Northern Ireland, 2 from England, 1 from Spain, 1 from Hungary 3 from Australia, 1 from Mexico, 1 from New Zealand). Likewise, no intraspecific variation was noted in F. gigantica (2 samples: Indonesia and Malaysia). Only one sample of Fascioloides magna was available. One specimen of Fasciola sp. from Japan (specific identity regarded as uncertain in the literature) yielded a restriction map identical to that of F. gigantica.

Almost all cut sites occurred in or near the putative rRNA coding regions. The non-transcribed spacer region had few or no cut sites despite the fact that this region is at least 30% of the entire repeat. The spacer probably consists of multiple short subrepeat units. Because of the lack of informative variable cut sites in the non-transcribed spacer and the conserved nature of sites in the coding regions, restriction maps are likely to be of limited value in systematic studies on digeneans (flukes).

CF 102 MOLECULAR ANALYSIS OF THE HISTONE MULTIGENE FAMILY IN Drosophila: THE willistoni SPECIES GROUP, Laurine Bow and Linda Strausbaugh, Department of Molecular and Cell Biology, The University of Connecticut, Storrs, CT 06269-3125

The *willistoni* species group (genus *Drosophila*, subgenus *Sophophora*) contains 23 named species in two subgroups, *willistoni* and *bocainensis*. The *willistoni* subgroup contains six sibling species (*D. equinoxialis, D. insularis, D. paulistorum, D. pavlovskiana, D. tropicalis* and *D. willistoni*) that have nearly identical external morphologies. We have examined the molecular characteristics of the histone multigene family in this subgroup, and have also included *D. sucinea* of the *bocainensis* subgroup as an outgroup member. The results of these studies reveal a number of interesting facts. The histone repeats occur in tandem arrangement and display remarkably uniform size of approximately 5kb, with little length heterogeneity within or between species. The organization of the repeating unit is the same as that previously described for other members of the genus. Despite the fact that the histone proteins are among the most highly conserved proteins known, we find that the repeating unit demonstrates extensive restriction fragment length polymorphisms, even between sibling species. Consistent with the phenomenon of concerted evolution, the vast majority of repeats show identity for any particular RFLP. Based on the restriction endonucleases that reveal the most polymorphisms and the positions of the recognition sites, we conclude that most changes occur in coding portions of the repeats. We postulate that the basis for the RFLPs are neutral mutations. The suitability of histone repeats as DNA markers for insect biosystematics will be discussed.

CF 103 PHYLOGENETIC RELATIONS AMONG TETRAHAYMENA SPECIES DETERMINED BY DAN SEQUENCE ANALYSIS, Clifford F. Brunk and Lori A. Sadler, Biology Department, UCLA, Los Angeles, CA 90024. The Tetrahymena pyriformis complex is composed of our 30 species that are virtually indistinguishable at the morphological level and highly divergent at the molecular level. We have determined the phylogenetic relationships among most of these species by comparing the DNA sequence of a homologous region from each species. This region includes portions of histones H311 and H411 as well as the intergenic region between the genes. The region of interest was amplified from each species using the polymerase chain reaction, cloned and sequenced. An analysis of the aligned sequences provides phylogenetic relationships among the species. In general these relationships are consistent with the relationships previously determined for 13 species by comparing rRNA sequences. Although the region of the genome we have analyzed is much shorter (566 bp) than the rRNA sequences analyzed (1753 bp) there are six times as many nucleotide substitutions in this region. In spite of divergence among different species, specific portions of the intergenic region have a high degree of sequence conservation, which implies functional constraint. The histone sequences are more divergent than expected. A comparison of the orthologous histone genes and the paralogous histone genes from Tetrahymena thermophila suggests some mechanism of concerted evolution may be operating.

CF 104 EVOLUTION OF β-GLUCURONIDASE REGULATION IN THE GENUS *MUS*
Robin M. Bush, Kenneth Paigen, Genetics Dept., Univ. of California, Berkeley, CA, 94720.

Differences in the developmental timing and tissue-specific expression of gene products are likely to be at least as significant in the course of evolution and speciation as changes in the physical properties of proteins themselves. In the mouse kidney, the β-glucuronidase gene codes for an enzyme thought to be involved in the production of agression phermones which are released into the urine. The hormonal regulation of β-glucuronidase involves induction by androgen in the presence of growth hormone, and repression of induction by estrogen. Variation in the hormonal regulation of the β-glucuronidase gene has been found among tissues and among species within the genus *Mus*. Sources of variation in regulation, and the relationship of the pattern found among species is discussed in light of current thought on murine phlyogenetics.

CF 105 STRONG FUNCTIONAL GC-PRESSURE IN A LIGHT REGULATED NUCLEAR GENE FROM MAIZE AS REVEALED BY PSEUDOGENES, R.Cerff[1], F.Quigley[1], H.Brinkmann[1] and W.Martin[2] 1) Lab.Biol.Moléc.Vég., CNRS UA 1178, Université Joseph Fourier, F-38042 Grenoble, France; 2) MPI für Züchtungsforschung, D-5000 Köln 30, FRG.

It has been suggested that the high G+C content of inducible housekeeping genes from monocotyledonous plants may be due to overall genomic constraints, G+C rich "isochores" surrounding these genes, rather than to functional constraints exerted at the expression level. The light regulated nuclear gene encoding subunit A of chloroplast glyceraldehyde-3-phosphate dehydrogenase (GapA) from maize is extremely G+C rich (67 % in codons) [1,2]. We analysed the genomic surrouning of this gene (4.3 kb) and the structure of two recent pseudogenes. Our results show that the high G+C content of the maize GapA gene is maintained independently of the surrounding noncoding sequences, which are G+C poor (42 % G+C), and only as long as the gene encodes a functional protein. After nonfunctionalisation GapA pseudogenes rapidly loose G+C mainly due to enhanced turnover of CpG and CpXpG methylation sites. These results suggest that the maize GapA gene is under strong functional GC-pressure reflecting the combined effect of separate constraints at both the transcriptional (CpG rich island) and translational (G+C rich codons) level. They further indicate that GapA pseudogenes are methylated and that methylation was either the cause or the immediate consequence of their nonfunctionalisation.

1. Brinkmann,H., Martinez,P., Quigley,F., Martin,W. and Cerff,R.(1987) J. Mol. Evol.**26**, 320-328. 2. Quigley,F., Martin,W. and Cerff,R. (1988) Proc. Nat. Acad. Sci. USA **85**, 2672-2676.

CF 106 THE INVOLUCRIN GENE OF PAN IS MORE CLOSELY RELATED TO THAT OF GORILLA THAN TO THAT OF HOMO. P. Djian, J. Teumer, H. Tseng, and H. Green, Department of Cellular and Molecular Physiology, Harvard Medical School, Boston, MA 02115. Most of the coding region of the gene coding for involucrin, an epidermal protein, is composed of repeats of a 10 codon sequence (the modern segment). Each repeat can be assigned to one of 2 classes A or B, according to the first 3 codons. The pattern of these repeats can be used to establish correspondence of the repeats of different higher primates. The modern segment can be divided into 3 regions designated as early, middle and late. The early region (10 repeats at the 3' end of the modern segment) is very similar in all anthropoids so far examined. The middle region (18-19 repeats) is similar in human, gorilla, chimp and orang, and must have been created in a common ancestor of the hominoids. In the late region, located at the 5' end of the modern segment, the number and pattern of repeats is quite different between human and the other hominoids. However the late regions of Pan and Gorilla are closely related. Part of the late region of these 2 species was created in a common ancestor, and after their separation, the Gorilla gene continued to add new repeats. Similarities in the middle regions of Pan and Gorilla also support a closer relatedness of these 2 species to each other than to Homo. Since the modern segment has evolved by the addition of repeats at precise locations, the expansion of the gene appears to be a directed, not a random process.

CF 107 NEW ALU REPEATS IN THE HUMAN GENOME: AN EVOLUTIONARY NOVELTY IN HUMAN PHYLOGENY, Achilles Dugaiczyk and Susan C. Ryan, Department of Biochemistry, University of California, Riverside, California 92521.
We discovered the presence of a new Alu repetitive DNA element within intron 4 of the human α-fetoprotein (AFP) gene; this element is absent from the same position in the gorilla AFP gene. The human Alu element is flanked by 12 base pairs direct repeats which presumably arose by way of duplication of the intronic target site AGGATGTTGTGG at the time of the Alu insertion. In the gorilla, only a single target site is recognized, which is identical to the terminal repeat flanking the human Alu element. We conclude that this Alu element was inserted into the human genome at a time postdating the divergence of the human and gorilla lineages, and hence represents an evolutionary novelty in human phylogeny. Our conclusion about a recent origin of this Alu element draws further support from the recognition that members of the Alu family can be subdivided into four subclasses (Willard et al., J. Mol. Evol. 26, 180; 1987; Britten et al. PNAS 85, 4770; 1988), each being inserted into the host genome at different times of its evolutionary history. The Alu element in the human AFP gene belongs to the youngest of the four subfamilies. This is already the second finding of Alu repeats being present in one lineage of the higher primates and absent from others. The first one was the finding by Trabuchet et al. (J. Mol. Evol. 25, 288; 1987) of an Alu element within the globin genes in the gorilla; this repeat is absent from the same position in the human, chimpanzee, and macaque genomes. We postulate that, unlike nucleotide substitutions, spreading of Alu repeats through the genomes of species is an irreversible process which could provide molecular markers for tracing phylogenies.

CF 108 ESTABLISHMENT OF GENOMIC AND DEVELOPMENTAL ARCHITECTURES AS A SOURCE OF NOVELTY AND CONSTRAINT: THE CAMBRIAN METAZOAN RADIATION Douglas H. Erwin, Department of Geological Sciences, Michigan State University, East Lansing, MI, 48824
Most metazoan phyla and classes first appeared during the relatively brief Cambrian Metazoan Radiation c. 600 million years ago. The development of such complex metazoa required the construction of cell:cell interactions, regulatory patterns and developmental mechanisms beyond those required by the unicellular and multicellular precursors of metazoa. Construction of phylogenies based on shared regulatory and developmental elements not only contain more information than traditional molecular phylogenies, but also allow mapping of the apparent acquisition of such elements. Analysis based on theoretical models of regulatory patterns indicates that repatterning of control networks (and thus greater potential for large-scale morphologic change) would be easier early in the history of such networks before the networks rigidified through the development of secondary and tertiary connections. This proposal has several consequences: 1) Genomic and developmental patterns which evolved prior to the radiation will be common to most metazoa, while the many subsequent innovations associated with specific body plans will be unique to the phyla or classes in which they first appeared. 2) The genomic events leading to the production of novel body plans may have only been possible in the simpler genetic architectures of the latest Precambrian, suggesting that mechanisms of molecular evolution may be time-specific.

CF 109 MOLECULAR SYSTEMATICS AND GOELDI'S MARMOSET

T. Fanning and L. Forman, National Cancer Institute, Bethesda, MD 20892 and National Zoo, Washington, DC 20008.

The position of Goeldi's marmoset (Callimico geoldii) in the taxonomy of neotropical primates is controversial. Many investigators place C. goeldii in the family Callitrichidae (tamarins and marmosets), while others suggest that it belongs in the family Cebidae (monkeys). Because many of its features are intermediate between the Callitrichidae and Cebidae, it has also been suggested that C. goeldii be placed in its own family, Callimiconidae.

We have cloned and characterized a number of repetitive DNAs from neotropical primates, including the LINE-1 transposable element and tandemly repeated satellites. LINE-1 restriction maps suggest that C. goeldii is more closely related to the Callitrichidae than to the Cebidae. Based upon sequence data from a satellite unique to C. goeldii, we estimate that the C. goeldii lineage diverged from other neotropical primates at least 30 million years ago.

CF 110 ALTERATIONS IN DNA REPLICATION AFTER INTRODUCTION OF 7SL RNA GENES INTO HUMAN CELLS, M. Fordis, B. Helmly, K. Sakamoto, W. Holter, T. Howard and B. Howard, Laboratory of Molecular Biology, National Cancer Institute, Bethesda, MD 20892. Repetitive elements may play a role in the evolutionary restructuring of the eucaryotic genome. For example, members of the Alu family of repeats are involved in recombination and transposition and implicated, by several lines of evidence, in DNA replication. We wished to investigate any role such sequences may play in DNA replication and cell growth but were daunted by the difficulty of selecting a functionally active element from the 300,000 or more Alu repeats scattered throughout the human genome. However, since the Alu family is thought to descend from the functional 7SL RNA family of genes, Alu members may retain some of the properties demonstrable in studies of the ancestral genes. Therefore 7SL genes were first selected for study in a newly developed transient analysis system where cells expressing exogenous DNA can be identified and DNA replication measured. Experiments were performed in HeLa cells by cotransfecting candidate sequences along with a marker gene encoding the p55 (TAC) subunit of the human interleukin2 receptor. Two to three days after transfection, cells were briefly exposed to BUDR and then harvested and stained with anti-TAC and anti-BUDR antibodies. Flow cytometry was used to identify the transfected cells and to measure the incorporation of BUDR in cells expressing the TAC antigen. In three separate experiments the introduction of either a wild-type, genomic 7SL gene or a 7SL cDNA produced statistically significant ($p<.05$) increases in mean BUDR incorporation to levels that were 140% and 126%, respectively, of that seen in cells transfected with a pUC19 control plasmid. By comparison, in two experiments the introduction of an anti-proliferative gene, β-interferon, decreased mean BUDR incorporation to 83% of control. The increased BUDR incorporation associated with introduction of the 7SL genes is not explained by increased cell division (one experiment) nor is it accompanied by a statistically significant change in surface receptor expression and therefore, may represent interactions of these genes with the replication machinery.

CF 111 STUDIES OF DNA TOPOISOMERASES INDICATE CRITICAL STEPS OF CELLULAR EVOLUTION,

Patrick Forterre, Institut de Microbiologie d'Orsay, Universite Paris-sud, 91405 Orsay, France

DNA topoisomerases determine the three-dimentional structure of the DNA message and solve the topological problems encountered in DNA replication, transcription and recombination. Topoisomerases are classified as type I and II according to their mechanisms of action which involve either a transient single-stranded or a double-stranded DNA break. DNA topoisomerases II are homologous in eubacteria, eukaryotes and archaebacteria indicating that their common ancestor appeared after the RNA/DNA worlds transition. The absence of DNA gyrase in eukaryotes (a eubacterial type II enzyme which introduces negative superturns) and the non-homology between eukaryotic and eubacterial type I DNA topoisomerases indicate a drastic shift between eukaryotic and prokaryotic control of DNA topology. Recently, a type I DNA topoisomerase which introduces positive superturns, reverse gyrase, has been discovered in some thermophilic archaebacteria. New data suggest that reverse gyrase is present in only one of the two archaebacterial branches and could be related to the eubacterial topoisomerase I. These data indicate that another drastic change in the mechanisms which control DNA topology occur at the time of the divergence between the two branches of archaebacteria.

CF 112 ANALYSIS OF Bm1 MIDDLE REPETITIVE ELEMENT TRANSCRIPTS FROM *Bombyx mori*, Guang P. Gao, Jose L. Soto, Jean G. Cantave, Leticia R. Vega, Rene J. Herrera, Department of Biological Sciences, Florida International University, University Park, Miami FL 33199

Several families of middle repetitive elements are known to be transcribed. They tend to remain in the nucleus and they are rapidly degraded. Some of these middle repetitive families are transcribed as part of large transcripts containing mRNA. The Bm1 family of middle repetitive sequences found in the silkmoth, *Bombyx mori*, exhibits homology with U1 snRNA, tRNA and *Alu* sequences. The Bm1 family of repetitive elements was found to be transcribed and recently we have determined that the transcripts remain mainly in the nucleus. The transcripts differ in size and they are approximately the size of the Bm1 consensus sequence (approximately 250 nucleotides). Therefore, it seems that Bm1 is not transcribed as part of larger transcripts containing mRNA sequences. Another possibility is that larger transcripts containing mRNA sequences are rapidly processed. There are approximately 1.3 x 10^9 repetitive element transcripts per ug of total nuclear RNA. Since the Bm1 consensus sequence exhibits a 61% homology with *Drosophila* U1 snRNA, we believe that the Bm1 family may have evolved from U1 sequences, probably U1 pseudogenes.

CF 113 **ANALYSIS OF THE EVOLUTION OF PROTON PUMPING ATPases ALLOWS CLOCK-RATE INDEPENDENT ROOTING OF THE "THREE-KINGDOM" TREE: EUKARYOTES SHARE A MORE RECENT COMMON ANCESTOR WITH Sulfolobus THAN WITH Eubacteria**

Johann Peter Gogarten, Henrik Kibak and Lincoln Taiz, Biology Department, Thimann Laboratories, University of California, Santa Cruz, California 95064

Catalytic subunits (≈70kDa) of proton pumping ATPases of the vacuolar type (vacuoles, *trans*-golgi, lysosomes, endosomes, chromaffin granules) show a low but significant homology to the β-subunits (catalytic) of F_oF_1-ATPases (*Eubacteria*, chloroplasts, mitochondria), and a much higher similarity (>50% identity) to the α-subunit of the H^+-ATPase from the plasmalemma of *Sulfolobus acidocaldarius*.

The same order of similarity holds for noncatalytic subunits (F_oF_1: α-subunits; V-type: ≈60kDa subunits, *Sulfolobus* β). The noncatalytic subunits themselves are homologous to the above mentioned catalytic subunits (paralogous), i.e. they evolved from the same gene after gene duplication. As this gene duplication occurred in the common ancestor of the three groups (eukaryotes, eubacteria and *Sulfolobus*), the noncatalytic subunits can serve as outgroup for phylogenetic trees based on the catalytic subunit and *vice versa*; i.e.: the place where the paralogous subunits join the tree of the orthologous subunits gives the position of the last common ancestor (root).

CF 114 THE rRNA MULTIGENE FAMILY IN HUMAN AND APE EVOLUTION, Iris L. Gonzalez and James E. Sylvester, Hahnemann University, Pathology Department, Philadelphia, PA 19102.

rRNA genes been the prototype for the study of the evolution of tandemly-arranged multigene families, and are also useful for phylogenetic studies over long and short evolutionary periods. Our laboratory has studied both of these aspects.

Although the rRNA genes are subject to concerted evolution, we have characterized human intra-species sequence and length variability in both the coding and spacer regions. This variability arises from base changes, replication-slippage and from cross-over events; the latter also promote the process of concerted evolution. Here we demonstrate the independent insertion and the concerted evolution of four Alu elements in the non-transcribed spacer. Frequent recombination keeps the rRNA repeat closest to the telomere similar to the internal repeats and confirms the observations of Worton, et al., concerning the homogeneity of the distal ends of the acrocentric chromosomes.

On the phylogenetic front, we have sequenced 3500 bases from the 28S gene and internal transcribed spacer in human and great apes. Our results show that human and chimpanzee are most closely related.

CF 115 CODON PAIR UTILIZATION IN *E. coli* IS NON-RANDOM, George A. Gutman, G. Wesley Hatfield, Department of Microbiology and Molecular Genetics, University of California, Irvine, CA 92717. We have examined the pattern of codon pair utilization among 237 *E.coli* protein coding genes. We find that their use is highly non-random, many pairs being highly over- or under-expressed compared with their random expectations. This effect is independent of non-randomness in amino acid pair utilization (which itself is highly evident); it is much weaker when non-adjacent codon pairs are examined, and virtually disappears when pairs separated by two or three intervening codons are evaluated. A high degree of directionality is evident in this bias; any particular codon which participates in many non-random pairs will tend to make *both* over- and under-represented pairs, but preferentially as a right- *or* left-hand member. There appears to be a relationship between codon pair utilization and level of gene expression: genes encoding proteins expressed at low levels tend to favor less abundant but more highly over-represented codon pairs, relative to more highly expressed genes.

The non-random utilization of codon pairs may be a consequence of their effects on translational efficiency, which in turn may be related to the compatibility of adjacent aminoacyl-tRNA isoacceptors at the A and P sites of a translating ribosome. We suggest that non-random codon pair utilization is the consequence of protein-coding regions of an organism co-evolving with its protein synthetic machinery, and that this co-evolution is responsible for the species-specific patterns of the number, abundance, and chemical and structural modifications of tRNA isoacceptors, as well as codon and codon pair utilization. Such co-evolution could provide each organism the ability to optimize and regulate translational efficiency while imposing minimal constraints on the structure and function of the encoded proteins.

CF 116 8,000 YEAR OLD HUMAN DNA, William W. Hauswirth*, Cynthia D. Dickel*, Glen H. Doran+, and David N. Dickel+, *Department of Immunology and Medical Microbiology, University of Florida, College of Medicine, Gainesville, Florida 32610 and +Department of Anthropology, Florida State University, Tallahassee, Florida 32306.

The Windover Archeological Site in East-Central Florida has yielded the skeletal remains of 166 individuals. Within 91 crania, material identifiable by gross anatomical features and remanent cellular structure as brain has been recovered and dated at 8,000 ±200 BP. DNA can be isolated and the existence of human mitochondrial DNA demonstrated by Southern blotting. About 10% of six-base restriction sites are cleavable, short DNA fragments are clonable at low efficiency and the human DNA is amplifiable by PCR.

This find may represent a significant genetic resource for the evolutionary community. We therefore wish to encourage discussion of several important issues including; 1) relative value of DNA of this antiquity for evolutionary and population studies; 2) additional criteria for determining its potential as a genetic resource; 3) distribution to the scientific community, including questions concerning under what conditions material should be distributed, how much and in what state samples should be provided, and how best to preserve a portion of each available brain for the future. We solicit suggestions on these and any related issues so that optimal use can be made of this apparently unique but limited resource.

CF 117 CORRELATION OF GENE CONVERSION AND CODON BIAS IN THE EVOLUTION OF DUPLICATED DROSOPHILA GENES, Donal A. Hickey, Veronique Payant, Yves Genest and Bernhard Benkel, Department of Biology, Univ. of Ottawa, Ottawa, Canada K1N 6N5. Recently, there have been a number of demonstrations, using laboratory experiments, that intrachromosomal recombination between duplicated DNA sequences can lead to gene conversion by heteroduplex DNA repair. Other experiments have shown that there is a bias in heteroduplex repair which favors the formation of GC base pairs. The results of these experiments suggest that, during the course of evolution, duplicated sequences would undergo repeated rounds of gene conversion and that, in so doing, they would become increasingly GC-rich. This latter effect would inevitably bias the codon usage within such genes. We report on a "natural experiment" in Drosophila evolution which bears out these predictions. We show that a pair of closely-linked amylase genes undergoes gene conversion within the melanogaster species group. Moreover, the third codon position within these genes shows an extreme GC bias, in contrast to third codon position of the homolgous sequence in other insect species. These results indicate that some instances of codon bias may result primarily from the mechanics of DNA repair processes.

CF 118 THEORY OF REVERSE TRANSCRIPTASE-MEDIATED EXON MOBILIZATION.

Clague P. Hodgson, Labs of Molecular and Developmental Biology/Dairy Science, The Ohio Agricultural Research and Development Center, The Ohio State University, Wooster, OH. 44691.

Since much of the DNA of animal and plant genomes has apparently arisen by reverse transcription of cellular and virally derived RNAs, logic suggests that genomes would protect themselves from the effects of retroposition and/or utilize retroviral mechanisms to mobilize genes during evolution. Rampant reverse transcription can be lethal to cells and organisms, while special types of reverse transcriptase-mediated events might be selected for because of the evolutionary advantage which they confer by rapidly mobilizing DNA sequences. Junk DNA (the 'C' value paradox), polyadenylation, and split genes provide; 1) a neutral buffer against lethal insertions, 2) a barrier to random priming of reverse transcription of cellular mRNAs, and 3) the opportunity for enhanced mobilization of exons, respectively.

The selective mechanism involves 3'-poly(A) priming of cDNA synthesis in vivo from oligo(U) tracts sequestered in noncoding regions of RNA. Oligo(U)-rich templates tend to be localized to intronic regions of hnRNAs of animal species. The degree to which species segregate oligo(U) may reflect their implementation of selective reverse transcription.

CF 119 ROLES FOR GENE AMPLIFICATION AS WELL AS REGULATORY AND STRUCTURAL GENE CHANGES IN THE ADAPTIVE EVOLUTION OF RUMINANT STOMACH LYSOZYMES, David M. Irwin and Allan C. Wilson, Department of Biochemistry, University of California, Berkeley, CA 94720.

Although ruminants are lysozyme-deficient in most tissues, while nonruminants express lysozyme in many tissues, sequence and blotting studies show that the number of lysozyme genes has risen from about one to about ten per haploid genome on the lineage leading from nonruminants via primitive ruminants (Tragulina) to advanced ruminants (Pecora). The chief place in which ruminants express lysozyme is in the true stomach. Duplications of this gene have allowed nature to experiment with the sequence and site of expression of at least temporarily non-essential copies of the gene. Duplication was used later as a mechanism to increase the level of gene expression by increasing the number of copies of the now essential gene. In the early stages of adaptation to the ruminant lifestyle, rapid adaptive amino acid sequence evolution optimized lysozyme for function in a new environment, the stomach. More recently this optimized sequence appeared to have evolved at a reduced rate due to an increase in the importance of purifying selection. During the later stages of ruminant evolution the rate of amino acid substitution in stomach lysozyme sequences has been less than half that of typical mammalian lysozymes. Comparison of the cDNA sequences of the multiple lysozymes expressed within cow and sheep stomachs indicates that concerted evolution has homogenized the family of stomach lysozyme genes within these species since their divergence, although one of the cow lysozyme genes (cow 1) appears to have escaped the homogenization process.

CF 120 STRUCTURE AND CHARACTERIZATION OF THE RIBOSOMAL RNA INTERGENIC SPACER FROM CUCURBITA MAXIMA, R. J. Kelly and A. Siegel, Department of Biology, Wayne State University, Detroit, MI 48202

The genes for ribosomal RNA, rDNA, in eucaryotes are clustered into arrays of tandem repeating units, each consisting of a coding sequence for 18S, 5.8S and 25S pre-RNA and an intergenic spacer(IGS). Twenty-nine cloned repeat units of rDNA from C. maxima fall into 6 classes which differ from each other in length and/or nucleotide sequence. Differences in the IGS account for most of the heterogeneity between rDNA repeat units. The nucleotide sequence of the IGS between the 25S and the 18S coding regions of a C. maxima clone which represented the majority class was determined. The IGS has a complex primary structure that resembles IGSs of other species in that it contains regions of highly repetitive DNA. It is composed of 5 repetitive families (A-E) which follow each other in an ordered sequence. The dominant repeat unit (D) consists of 9 tandem 250 nucleotide copies, a characteristic feature of all IGSs examined to date. The cause of IGS length variability can be deduced by comparing the variant restriction map with the standard whose sequence and structure is known.

CF 121 RIBOSOMAL RNA SEQUENCES AND SPONGE PHYLOGENY, Patricia R. Bergquist, Michelle Kelly-Borges and Peter L. Bergquist, Departments of Zoology and Cellular & Molecular Biology, University of Auckland, Private Bag, Auckland, New Zealand. The methodologies of classical and chemical taxonomy have thus far not permitted resolution of the higher order classification of the Demospongiae (Porifera) to a level where there is a universally acceptable arrangement. To move toward such an arrangement we have focussed on one order, the Hadromerida and have used molecular taxonomic techniques to evaluate existing generic and familial assignments. This work, involving comparisons of 18S rRNA sequences, provides the first molecular sequence data for the Demospongiae in which a comparative taxonomic question is addressed. Sequences from hadromerid genera Tethya Polymastia and Aaptos, which fall within two existing families within the Hadromerida have been derived and compared. This information permits current arrangements to be tested.

CF 122 THE IFG FAMILY OF REPETITIVE SEQUENCES: AMPLIFICATION EVENTS AND RESTRICTION FRAGMENT LENGTH POLYMORPHISMS ARE MARKERS FOR PINUS (PINACEAE) EVOLUTIONARY PHYLOGENY. David S. Kossack and Claire Kinlaw, U.S. Forest Service, PSW Experimental Sta., Berkeley, Ca. 94701. A family of repetitive sequences has been found that has had multiple amplification events and shows restriction fragment length polymorphisms (RFLP) within the family Pinaceae. The IFG repetitive sequences, IFG PI - 1 & 2, were originally cloned from a band visible in an agarose gel of Pinus lambertiana (Sugar Pine) DNA digested with the restriction enzyme Eco RI. Additional IFG$^+$ repeats (IFG Pr) have been isolated from a P. radiata (Monterey Pine) EMBL-4 genomic library using IFG PI-1 as a probe. When IFG PI-1 is used as a probe for DNA hybridizations with species in the order Coniferales only trees in the family Pinaceae hybridize with the probe. Within the family Pinaceae copy number varies from <10 copies/genome in Cedrus and Psuedotsuga to ~10^4 copies/genome in Pinus. A restriction fragment length polymorphism (RFLP) is present in the Picea. Within the genus Pinus copy number of IFG PI-1 repetitive sequence is consistently ~10^4 copies/genome. A Hind III RFLP is consistent with separation of the genera into two subgenera, Strobus and Pinus. When IFG Pr is used only hybridizations with the subgenera Pinus show high copy number (>1000). Trees in the subgenera Strobus show a copy number of about 10. P. monophylla (Pinyon Pine) is an exception. This species is grouped in the subgenera Strobus but has IFG PI-1 Hind III RFLP consistent with the subgenera Pinus. This species also lacks any detectable signal when probed with IFG Pr. The evolutionary relationship of many of the species of the Pinus such as those in the subsections Cembroides (Pinyon Pines), Balfourianae (Foxtail and Bristlecone Pines) and Oocarpae (Monterey and Bishop Pines) remains unresolved. An evolutionary phylogeny based on the IFG repetitive sequence indicates that progenitors of the subgenus Strobus separated from the subgenus Pinus and P. monophylla before the amplification of IFG PI-1 containing the Strobus Hind III RFLP. The IFG PI-1 repeat containing the Pinus Hind III RFLP was also amplified after the separation of Strobus but before the separation of P. monophylla from Pinus. The amplification of IFG Pr then occurred in Pinus after the separation of P. monophylla.

CF 123 ALU ELEMENTS ORIGINATE IN A SMALL NUMBER OF MASTER SEQUENCES, Damian Labuda & George Striker* (Génétique Médicale, Hôpital Ste-Justine, Université de Montréal, Montréal, P.Q. Canada H3T 1C5; *Max-Planck-Institut für biophysikalische Chemie, 3400-Göttingen, FRG).
A statistical analysis of a set of human Alu elements is based on a published alignment and a recent classification of these sequences, known to be amplified through an RNA intermediate. If the origin of the more than half a million copies of these repeats in the human genome lies in a cascade of retropositions, where amplified sequences serve as templates for further amplifications, then we would expect to find a continuous series of mutations, as a substitution, once introduced, would perpetuate in subsequent amplification rounds. The strategy to test this is to separate the sequences into families, and study the distribution of mutational diversities within the families. For this, consensus sequences of each family are needed, and these must be properly weighted. For, the tenfold greater and unidirectional mutation rate at CpG's requires separate consideration of an independent clock at every stage of analysis. The distributions of the substitutions with respect to these family consensus sequences, taking the CG and the non-CG-nucleotide positions separately, can then be considered. These turn out to lie far closer to the distributions that would be expected if all the elements of a family originated from that single family consensus sequence. Computer analysis of the folding of RNAs derived from these consensus sequences indicates that RNA secondary structure is conserved among the Alu families, suggesting its importance for Alu proliferation. This folding pattern, further substantiated by a number of compensatory mutations, includes secondary structure domains which are homologous to those observed in 7SL RNA and a defined region of interaction between the two Alu subunits. These results are consistent with a model in which a small number of conserved Alu master genes give rise via retroposition to the numerous copies of Alu pseudogenes, that then diversify by random substitution. The master genes appeared at different periods during evolution giving rise to different families of Alu sequences.

CF 124 EVOLUTION OF A TISSUE-SPECIFIC RETROPOSON, John R. McCarrey, Division of Reproductive Biology, The Johns Hopkins University School of Public Health, Baltimore, MD 21205.
Eutherian and metatherian mammals possess two functional loci for the production of the glycolytic enzyme phosphoglycerate kinase (PGK). Pgk-1 is an X-linked gene that is ubiquitously expressed in somatic cells, oogenic cells, and premeiotic spermatogenic cells. Pgk-2 is an autosomal gene expressed specifically in meiotic and postmeiotic spermatogenic cells. The human Pgk-2 gene is a functional retroposon that apparently evolved by reverse-transcriptase-mediated processing of a transcript from the Pgk-1 gene [McCarrey, JR & Thomas, K (1987) Nature 326:501-505]. Initial expression of this retroposon was apparently directed by a Pgk-1-like promoter that was duplicated along with the coding sequence to form the Pgk-2 gene [McCarrey, JR (1987) GENE 61:291-298]. Thus the initial expression of the Pgk-2 gene must have been essentially ubiquitous, and has subsequently evolved to be tissue-specific. A comparison of expression of the Pgk-2 gene in metatherian and eutherian species supports the contention that expression of this locus has become increasingly restricted to late spermatogenic cells, the one cell type in which the Pgk-1 gene is not expressed. This is consistent with the hypothesis that spermatogenesis-specific expression of the Pgk-2 gene is required in eutherian species to compensate for an inactivated Pgk-1 gene in these cells. A detailed analysis of the promoter sequence of the Pgk-2 gene is focused on the critical sequences believed to direct tissue-specific expression. Finally, a comparison of the functional human Pgk-2 retroposon sequence with that of an X-linked processed pseudogene of the Pgk-1 gene illustrates the consequences of divergence of a retroposon from its progenitor locus under the influence of selection at the protein level.

CF 125 EVOLUTIONARY RELATIONSHIPS AMONG SPECIES OF THE TRITICEAE, C. L. McIntyre, P. Reddy, B. C. Clarke and R. Appels, Department of Agronomy, University of Missouri, Columbia, MO 65211 and CSIRO, Division of Plant Industry, GPO Box 1600, Canberra ACT 2601, Australia
Nucleotide sequence information has been obtained from two repeated sequence loci, namely the 5S DNA and Nor, from 15 species in the tribe Triticeae. The data has been used to provide insights into the evolution of these repeated sequence families in addition to phylogenetic relationships between the species examined. An isozyme analysis of 16 putative enzyme loci encompassing these species has also been performed. Both cladistic and phenetic methods have been employed for the three data sets. The evolutionary relationships suggested by the sequence data have been compared and are generally consistent with relationships suggested by the isozyme data and other characters, such as classical systematics and chromosome pairing studies. Apparent inconsistencies can be explained in terms of introgression of chromosome segments from one species to another, and/or amplification/deletion events which are known to occur at repetitive sequence loci.

CF 126 Molecular Evolution of the Small Subunit of Ribulose Bisphosphate Carboxylase. S Berry-Lowe[a], K Rice[b] and R B Meagher[c], Departments of Botany[a] and Genetics[c], University of Georgia, Athens GA. Museum of Comparative Zoology[b], Harvard University, Cambridge, MA. The nucleotide sequences encoding the mature portion of 31 ribulose-1,5-bisphosphate carboxylase small subunit (SSU) genes from 17 genera of plants, green algae and cyanobacteria were examined. The replacement nucleotide substitutions (RNS) among the 465 pair-wise sequence comparisons revealed that the SSU multigene family members within the same species were more similar to each other in RNS than they were to SSU sequences in any other organism. The concerted evolution of independent SSU gene lineages within closely related plant species suggests that homogenization of RNS positions occurs at least once in the life of each genus. A SSU gene tree based on corrected RNS is presented and agrees well with a phylogenetic tree based on morphological and cytogenetic traits for the 17 genera examined. Synonymous nucleotide substitutions, SNS, within SSU sequences occur approximately five times faster than RNS during the initial 70 MY of divergence between any two species. However, the SNS do not occur at five times the level of RNS within a gene family. SNS comparisons within a SSU gene family may show from 1 to 90 times the number of RNS. A mechanism involving gene conversion within the exons followed by drift and selection for biased gene conversion products is presented to explain the concerted evolution of SSU sequences within a gene family. During the initial divergence of SSU sequences in different species, SNS occur at an expected rate of ~6.6 X 10^{-9}/site/yr, similar to the rate of SNS observed in productive animal genes.

CF 127 ANALYSES OF REPETITIVE DNA FAMILIES IN CARNIVORES AND RODENTS, William S. Modi, Thomas G. Fanning and Stephen J. O'Brien, Biological Carcinogenesis and Development Program, Program Resources, Inc., NCI-Frederick Cancer Research Facility, Frederick, MD 21701. The predominant tandem satellite arrays each from the domestic cat and domestic dog were isolated and characterized. Southern blot and in situ chromosome hybridization analyses using the cat probe to other species of felids indicates interspecific variation in copy number and restriction patterns, and that the repeats are localized primarily at telomeric regions of certain autosomes. Similar studies among canids using the dog probe indicate relative interspecific uniformity regarding copy number and restriction patterns. Canid arrays are found principally at autosomal centromeric regions. A tandem satellite array consisting of a 161 bp consensus sequence isolated from the rock vole shows interspecific conservation in restriction pattern but differences in copy number. A highly amplified, sex-chromosome specific repetitive element, also derived from the rock vole, appears to be organized as a long interspersed repeat and is absent in the genomes of closely related species.

CF 128 NUCLEAR VS. MITOCHONDRIAL RIBOSOMAL RNA RELATIONSHIPS GIVE CONTRASTING VIEWS OF THE NUMBER OF ORIGINS OF VERTEBRATE PARASITISM IN TRYPANOSOMATID PARASITES, Kimberlyn Nelson[1], Jay DeSalvo[2], Elizabeth Zimmer[2], Stephen M. Beverley[1], [1]Dept. of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston 02115 and [2] Dept. of Biochemistry, Lousiana State University, Baton Rouge, 70803. Trypanosomatid protozoan parasites exhibit two basic lifestyles: monogenetic genera such as Crithidia and Leptomonas which parasitize invertebrates, and digenetic genera such as Leishmania and Trypanosoma which additionally parasitize vertebrates, often causing disease. The number of origins of the digenetic lifestyle is currently a matter of debate, with estimates ranging from once to as many as five events. We are pursuing a molecular phylogenetic approach to this problem, by determining and analyzing small and large subunit nuclear ribosomal RNA sequences of 11 taxa. We are employing rRNA sequences from the related family Bodonidae to root the molecular tree. Our data, currently including approximately 400 bases per taxon, suggest that the molecular tree for nuclear rRNA differs substantially from that calculated for small subunit mitochondrial rRNA sequences by Lake et al (PNAS 85:4779, 1988). The nuclear tree suggests that multiple shifts have occurred, whereas the mitochondrial tree suggests a single origin of vertebrate parasitism. We are now obtaining additional sequence data and exploring the effects of asymmetrical base composition and methods of tree construction, to confirm the substantial differences in the two molecular trees. This finding would have several interesting implications concerning the evolution of nuclear and organellar genes.

CF 129 SHAPING OF THE GENOMES OF SALMONID SPECIES DURING EVOLUTION Norihiro Okada, Institute of Biological Sciences, University of Tsukuba Tsukuba city 305, Japan Previously, we demonstrated that the O.keta Pol III/SINE is derived from a lysine tRNA pseudogene. To examine the mechanism of amplification and dispersion of Pol III/SINEs in salmonid species, we determined several sequences. In the case of S. leucomaenis, which belongs to the different genus from that of O.keta, the Pol III/SINE has high over-all sequence homology with that of O.keta. Especially, the tRNA-related block of Pol III/SINE has 94% homology with that of O.keta, and has over-all homology with a lysine tRNA (72% homology). The fact that it contains a CCA sequence provides direct evidence that the tRNA-related block retrotransposed via tRNA. On the other hand, the Pol III/SINE in O. masou, which belongs to the same genus as O. keta, has less homology with that of O. keta. However, the tRNA-related region of O.masou is homologous to a lysine tRNA, although the extent of homology is not so high. Therefore, the extent of homology between the Pol III/SINEs in salmonid species does not correlate to the genetic distance of these species. Taken together, it is proposed that the tRNA-related block of each species is derived from a lysine tRNA and that a combined unit of tRNA-related and tRNA-unrelated blocks was independently generated and amplified at the time of generation of each salmonid species. The units of O.keta and S.leucomaenis were originally the same, and the present Pol III/SINEs of these species were generated by the second amplification event. A model of formation of the salmonid Pol III/SINEs is presented.

CF 130 EVOLUTION IN REPLETA GROUP DROSOPHILA: INFERENCES FROM A PHYLOGENETIC ANALYSIS OF MITOCHONDRIAL AND RIBOSOMAL RESTRICTION SITE DATA, Linda K. Park, Department of Biology, Washington University, St. Louis, MO 63130. The repleta group is the largest continental group in the genus Drosophila, comprising over 76 species in five major subgroups. Almost the entire radiation of this group has been confined to the New World. I have constructed restriction maps of the mitochondrial genomes of 30 repleta group species, using 15 enzymes. Combining this data with sequence data, and restriction data of the nuclear rDNA, I have constructed a phylogeny for these taxa. Comparisons will be made to results from similar studies conducted with Hawaiian Drosophila, and the role of population level processes on the differences in molecular level evolution between the continental-evolved species and the island-evolved species will be examined.

CF 131 STRUCTURAL CHARACTERIZATION INCLUDING REACTIVE SITE REGION OF THE MAJOR PLASMA SERPINS OF HORSE AND WALLABY, Scott D. Patterson, Kevin Bell and Denis Shaw¶, Dept. Physiology and Pharmacology, Univ. of Queensland, St. Lucia, 4067 and ¶J.C.S.M.R., Aust. National University, Canberra, 2601, AUSTRALIA. The major plasma serpins (SERine Protease INhibitors) of the horse and tammar wallaby are highly polymorphic and show similarities to the major human plasma serpin, α_1Pi. Structural characterization has been performed on these inhibitors using electrophoretic methods including 'in situ' oxidation/complex formation, limited 1D peptide and 'epitope' maps and blotting onto either glass fibre or PVDF for amino acid microsequence analysis. The patterns of all four equine loci and the single wallaby locus proteins displayed similar peptide and 'epitope' maps to human α_1Pi and the structural studies grouped the equine loci into two pairs (Spi 1&2, 3&4). N-terminal sequence analyses confirmed these results and revealed that the wallaby and equine inhibitors were able to be aligned with human α_1Pi. The reactive sites of equine Spi 1 and the wallaby proteins showed close sequence identity with that of human α_1Pi with a higher degree of positional identity than was found at the N-terminus. These proteins could be termed METserpins based on the identification of Met as their P_1 residue. Two other equine proteins were identified as an ILEserpin (Spi 2) and an ARGserpin (Spi 4). A V8 protease-overlap peptide was not obtained for Spi3. The most interesting feature of this study is the high degree of structural and reactive site similarity between proteins from diverse evolutionary origins.

CF 132 RESTRICTION ENZYME ANALYSIS OF MITOCHONDRIAL DNA FROM TWO SPECIES OF MICROTINE RODENT WITH COMMENTS ON THE ORIGIN OF MICROTUS BREWERI. Carleton J. Phillips, Dorothy E. Pumo and Carol Ann Briskey, Department of Biology, Hofstra University, Hempstead, NY 11550. Mitochondrial DNA (mtDNA) was isolated from two microtine rodents, Microtus pennsylvanicus and M. breweri. Specimens of the former were from Long Island, Connecticut, Rhode Island, and Maine. The latter species is only known from Muskeget Island, off the coast of Massachusetts. Purified mtDNA was digested with restriction enzymes, end-labelled, and separated by electrophoresis. Microtine mtDNA was cloned and compared to laboratory mouse mtDNA by means of Southern blots. Five mtDNA genotypes (L1-5), exhibiting 0.37-3.3% nucleotide sequence divergence, were isolated from M. pennsylvanicus. A single genotype, B1, was isolated from M. breweri. The B1 mtDNA was most similar to the L3 mtDNA from M. pennsylvanicus; indeed, the two differed by only two surveyed EcoR I sites hypothesized to be in the 16S rRNA gene. Phylogeographic analysis of mtDNA thus genetically links the endemic island species, M. breweri, to a maternal lineage presently represented in M. pennsylvanicus living in northwestern Maine. The morphologically distinctive M. breweri probably evolved from animals stranded on Muskeget Island by the last glacial retreat at the end of the Pleistocene. (NSF grant BBS-8609231.)

CF 133 THE ISOLATION OF UNIQUE SEQUENCE DNA CLONES FROM THE AVIAN SEX CHROMOSOMES AND THEIR APPLICATION TO PROBLEMS OF POPULATION BIOLOGY AND MOLECULAR EVOLUTION, Thomas W. Quinn, Fred Cooke*, Allan C. Wilson, and Bradley N. White*, Department of Biochemistry, University of California, Berkeley, CA 94720; *Department of Biology, Queen's University, Kingston, Ontario, Canada, K7L 3N6. We have isolated a segment of unique sequence DNA which was shown to originate from the Z chromosome of the Snow Goose by the presence of two copies in males versus one copy in females, as well as by the presence of single alleles at two polymorphic Taq I sites in females. This clone also has female-specific homology to a 2.1 Kb Taq I fragment of the W chromosome, and will allow W-specific clones to be recovered from a Snow Goose library. Such clones will allow several concurrent studies to be done including: 1) The development of a molecular technique for sex identification which is based on dot blot hybridization, and which has many advantages over those currently in use; and 2) In Aves, the W chromosome undergoes fewer rounds of replication per generation than either the Z (approx. 9 times fewer) or the autosomal (approx. 7 times fewer) chromosomes in germ line cells. By comparing DNA sequence differences between species at varying taxonomic distances, it will be possible to test what effect, if any, the number of rounds of replication per generation has on the rate of DNA sequence evolution.

CF 134 LOCALIZATION OF FOUR GENES ON CHINESE HAMSTER CHROMOSOMES 4, 6, AND 7 BY IN SITU HYBRIDIZATION: RELATIONSHIP OF GENE MAPPING TO G-BAND HOMOLOGY IN RODENTS, Mazin B. Qumsiyeh and D. Parker Suttle, Department of Biochemical and Clinical Pharmacology, St. Jude Children's Research Hospital, Memphis, TN 38101

In situ hybridization was used to map four genes on the Chinese Hamster (Cricetulus griseus, CGR) chromosomes. The results show that transferrin (TF) and uridine monophosphate synthase (UMPS) are localized to the distal portions of 4q and 4p, respectively. The gene for topoisomerase II is localized to CGR 7q21-23. We also confirm the location of the adenosine deaminase (ADA) gene (previously mapped by somatic cell hybrid panels) to chromosome 6 and further sublocalize it to bands 6q1.4-1.7. We have used high resolution G-banded karyotypes of the hamster (Cricetidae), the mouse (Muridae), and numerous Sigmodontidae, Gerbillidae and Arvicolidae to document homology and rearrangement in the linkage groups corresponding to CGR 4, 6 and 7 in these families of muroid rodents. Data on G-banding and gene mapping in rodents were then compared to those for primates especially the recently mapped genes for UMPS, TF (on HSA 3), ADA (HSA 21), and topoisomerase II (on HSA 17). We concluded that the segment representing HSA 17, MMU II, and CGR 7d is an ancient linkage group that has undergone few rearrangements during mammalian evolution. Other conclusions from combining comparative G-band analyses and gene map data for rodents and primates are: 1) careful analyses of G-band patterns in different rodent families allows for identification of homology and thus accurate prediction of gene map positions, and 2) comparative gene assignments based on synteny without a careful study of chromosome evolution and G-band homology can be misleading.

CF 135 SATELLITE DNA FAMILIES IN BOVIDAE

Sohail Qureshi and R.D. Blake
Department of Biochemistry, University of Maine, Orono, ME 04469

Satellite DNAs constitute the largest and least understood block of sequences in the eukaryotic genome. We have been involved in a study of several satellite DNAs in the genomes of calf (Bos taurus), its close relative the bison (Bison bison) and a cross between the two ("beefallo"). While representatives of the three principal satellite families are present in all three animals, their amounts, consensus sequences, sequence divergences within and between populations and chromosomal locations show significant differences. Satellite I (d=1.715g/cc) and satellite III (d=1.711g/cc) are present in Bos by more than twice that in bison, while the differential for satellite II (d=1.720g/cc) is reversed. Sequence divergence within a satellite population and between populations in the three animals have been determined from their consensus sequences, from ΔTm and from melting profile broadening in high resolution curves. The extent of differences and divergence levels in these satellites can be attributed to the homogenizing effects of unusual post-replication events (eg gene conversion) is estimated in the context of a model. (This work supported by MAES, Project No. 08405).

CF 136 EXON-STRUCTURE RELATIONSHIPS AND EVOLUTION OF REGULATORY CALCIUM BINDING PROTEINS,Malathi,R.,Mahendra,G.& Yathindra,N. Dept. of Genetics & Dept. of Biophysics, Univ. of Madras, India. Regulatory calcium binding proteins such as calmodulin, troponin-c and parvalbumin are believed to be genetically related and could have evolved as a consequence of duplication or triplication of a gene coding for a single calcium binding site. Gene sequences and three dimensional structures of these proteins have been analysed using distance diagonal plot and other structural features to examine if exon-intron arrangement and their reshuffling could have occurred within themselves during their evolution. Gene sequences of all these proteins are interrupted in identical and most conserved, regions namely the calcium binding loops, indicating that introns are not inserted randomly in nature. No strong correlation between exonic sequences and structural domains are found. Similar analysis are carried out on dehydrogenases and carbonic anhydrases.

CF 137 STRUCTURAL CHARACTERIZATION OF THE MOUSE MULTIDRUG RESISTANCE MDR1 GENE. Martine Raymond and Philippe Gros, Department of Biochemistry, McGill University, Montreal, Canada H3G 1Y6. Multidrug resistance in cultured cells is often associated with the amplification of a small gene family termed mdr. Sequence analysis of a full-length mdr1 cDNA clone suggests that the mammalian mdr1 gene originates from the duplication of an ancestral gene homologous to the energy-coupling subunit encoding genes of bacterial periplasmic transport systems. Each mdr1 duplicated half includes 6 transmembrane domains and a consensus ATP-binding site. We have isolated the entire mouse mdr1 gene in cosmid and bacteriophage clones and determined its structural organization. The transcription initiation site has been mapped by S1 nuclease protection. Nucleotide sequencing of the promoter region has allowed the identification of a putative TATA and a canonical CAAT box. The gene spans 68 kb of genomic DNA and is split into 27 exons encoding the 4342 nucleotides mdr1 mRNA. Projection of the intron positions on the sequence of the mdr1 protein shows that most introns are located between or near the ends of predicted structural protein domains (transmembrane loops, ATP-binding folds). Genomic organization of the duplicated halves of the gene has been compared: preservation of the exon/intron arrangement parallels the degree of homology in the coding sequence. Particularly, the highly homologous ATP-binding sites are each specified by three conserved exons. Identical or similar positioning of some introns implies that the mammalian mdr1 gene arose by duplication of an ancestor gene already containing introns. (Supported by MRC, NCI and NSERC)

CF 138 REGULATORY ALTERATIONS IN GENE EXPRESSION HAVE GENERATED MORPHOLOGICAL DIVERSITY OF THE MOTH CHORION, Jerome C. Regier, Center for Agricultural Biotechnology of the Maryland Biotechnology Institute & Department of Entomology, University of Maryland, College Park, MD 20742

Within the Lepidopteran superfamily called Bombycoidea, choriogenesis entails formation of a lamellar framework, followed sequentially by lamellar expansion and densification. Near the end of choriogenesis, surface structures called aeropyle crowns may be formed, depending on the particular species. We have isolated cloned copies of the RNAs and genes that encode known protein components of aeropyle crowns in the polyphemus silkmoth. In the cecropia silkmoth, aeropyle crowns are absent. Homotypic cloned sequences from cecropia were selected by cross-hybridization under stringent conditions with aeropyle crown-encoding clones from polyphemus. These clones were then used as species-specific probes. Inter-specific comparisons reveal that changes in the abundance and in the timing of chorion gene expression are, at least in part, responsible for the presence/absence of aeropyle crowns. This type of analysis has been extended to other superfamilies where differences in earlier modes of chorion morphogenesis (relative to that in the Bombycoidea) are observed. The generality of the importance of regulatory alterations in generating morphological diversity of the Lepidopteran chorion will be tested.

CF 139 MOLECULAR AND CHROMOSOMAL EVOLUTION IN THE MAMMALIAN ORDER *PERISSODACTYLA*

O.A. Ryder, R.E. Benveniste[§], M. George, Jr.[†], L.G. Chemnick, M.L. Houck, and A.T. Kumamoto. Center for Reproduction of Endangered Species, San Diego Zoo, San Diego, CA 92112; [§]Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, MD 21701; [†]Dept. of Biochemistry, Howard University, Washington, D.C. 20059. Nine species of horses, four species of tapirs and five species of rhinoceroses comprise the extant perissodactyls. Their common ancestor is believed to have been present in the early Eocene (ca. 50-55 MYBP). The fossil record has suggested the existence of separate lineages leading to extant horses and to the tapirs and rhinos. Divergence times from ancestral forms and phylogenetic relationships within each family have not been clear from fossil data. Additional insights concerning the evolution and speciation within these groups are desirable in the context of making conservation management decisions as over 70% of the extant perissodactyl species are currently threatened or endangered. The significance of the many named geographical and/or morphological subspecies ascribed to equid and rhinocerotid species can most usefully be discussed from an evolutionary genetical perspective. Our collaborative study brings together DNA-DNA hybridization data, the results of comparative cytogenetic studies, and the analysis of mitochondrial DNA cleavage patterns in the first attempt to provide a molecular phylogeny of the order. The chromosomal and molecular data are broadly congruent with the phylogeny of perissodactyls based on the fossil record. Rates of chromosomal evolution vary within and among the horse, tapir and rhinoceros families. These data will be presented as well as suggested guidelines for conservation management of gene pools. Supported by the Zoological Society of San Diego, NIH grant GM23073, and the Institute of Museum Services.

CF 140 DIFFERENTIAL EXPRESSION OF MYELIN PROTEINS IN THE BRAIN OF SHARKS AND MAMMALS,

Raul A. Saavedra, Lance Fors, Ruedi Aebersold and Leroy Hood, Division of Biology, California Institute of Technology, Pasadena, CA 91125

Myelin of the vertebrate nervous system is a dynamic structure that provides electric insulation and thus facilitates the rapid conduction of nerve impulses along axons. The major myelin proteins in the central nervous system (CNS) of mammals are proteolipid protein (PLP) and myelin basic protein (MBP), whereas in the peripheral nervous system (PNS) the major proteins are P_o and MBP. Sharks are the oldest known living vertebrates that have myelin organized as a concentric multilamellar structure around the nerve fibers; and therefore, shark is the system of choice to study the evolution of the myelin organelle. We isolated the major CNS myelin proteins from the shark Heterodontus francisci and obtained N-terminal and internal fragment sequences. The sequence data unequivocally shows that the major proteins of shark CNS myelin are similar to mammalian P_o and MBP. We screened a shark brain cDNA library using oligonucleotide probes reversed translated from the protein sequences, and isolated cDNA clones for both P_o and MBP. The shark P_o cDNA clones code for a protein that is 46% similar to its mammalian counterpart. Shark and mammalian P_o have an extracellular, a transmembrane, and an intracellular domain. The extracellular domains of P_o in both species appear to be organized as immunoglobulin-like domains that mediate homotypic interactions. The shark MBP cDNA clones code for a protein that is 44% similar to mammalian MBP. The MBPs of both species show conserved interspersed regions and are present in multiple forms that arise by alternative splicing of a single transcript. These structural analyses indicate that the complexities seen in mammalian myelin arose early during vertebrate evolution.

Abstract Withdrawn

CF 142 RECENTLY AMPLIFIED ALU FAMILY MEMBERS SHARE A COMMON PARENTAL ALU SEQUENCE, Valerie K. Slagel, Mark A. Batzer and Prescott L. Deininger, Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, New Orleans, LA 70112. Three polymorphic Alu subfamily members, representing recent amplification events, share a common parental Alu sequence. This data indicates that one, or a few Alu family members are dominating the amplification process and the vast majority are not actively involved in retroposition. We estimate that there are 500-1000 members of this most recently amplified subfamily in the human genome. The consensus sequence for the recent members is 1.8% divergent from the subfamily consensus sequence. To identify recent members, a specific oligonucleotide probe was used to screen a human genomic library under stringent hybridization conditions. These members are being further characterized with regard to their distribution throughout primates.

CF 143 AN ANALYSIS OF THE HETEROGENEITY OF CODON USAGE BETWEEN DIFFERENT HUMAN GENES. John Sved, School of Biological Sciences A12, Sydney University, NSW 2006, Australia and Richard Cowan, CSIRO Division of Biotechnology, North Ryde, NSW, Australia. Although there are known biases towards particular codons in different organisms, there are also conspicuous differences between genes within organisms. The high number of genes sequenced makes humans ideal material for an analysis of heterogenity of codon usage. We have shown that for each of the 18 amino acids with multiple codons there is a high level of heterogeneity of codon usage between genes. We have repeated the analysis for frameshifts of both one or two bases, and find that although there is still considerable heterogeneity, it is neither as high overall nor as uniform over the different amino acids. This suggests that at least a substantial component of the heterogeneity is due to properties of the codons themselves rather than to general properties of the DNA, e.g. variation in CpG dinucleotide frequency. We examine the question of whether the heterogeneity could be attributable to systematic use of different codons to modulate the speed of translation, as suggested by Grosjean and Fiers for prokaryote genes.

CF 144 GATA/GACA REPEAT SEQUENCES ARE INVOLVED IN ALLIGATOR, FISH AND SHRIMP SEX DETERMINATION, Martin L. Tracey, Peter A. Pechan and Case K. Okubo, Department of Biological Sciences, Florida International University, University Park, Miami Fl 33199. Various investigators have hypothesized that GATA/GACA repeat sequences are involved in sex determination. Others have suggested that this is not the case. We have probed genomic DNA and/or RNA isolated at differing developmental stages in order to assess the taxonomic breadth of GATA/GACA repeat correlations with sexual differentiation. The protogynous reef fish *Anthias squamipinnis* exhibits distinct bands in male DNA digested with a variety of restriction enzymes; these bands are not seen in females. The serially hermaphroditic shrimp *Lysmata wudermanni* gives similar results. Northern blots of *Alligator mississippiensis* provide further evidence of sex specific transcription. Thus we conclude that GATA/GACA repeats are generally involved in sex determination, because they exhibit sex specific differences in a variety of organisms.

SUPPORTED BY NIH RR08205

CF 145 THE MODERN SEGMENT OF INVOLUCRIN WAS CREATED DURING HIGHER PRIMATE EVOLUTION BY STEPWISE, SUCCESSIVE DUPLICATIONS, H. Tseng, J. Teumer and H. Green, Department of Cellular and Molecular Physiology Harvard Medical School; 25 Shattuck St. Boston, MA 02115 The protein involucrin is a product of terminal differentiation in the epidermal cell and related cell types. By comparing the nucleotide sequence of the involucrin genes of different animals, it is clear that the gene has undergone unusual evolution in the primates. The coding region of the gene contains an ancestral segment, most of which is common to the lemur, the pig and all three higher primates examined, and a species-specific segment of repeats derived from the ancestral segment. In the human, the segment of repeats (the modern segment) consists of 39 repeats of a 10 codon sequence. The involucrin genes of the lemur and the pig possess repeats derived from another sequence at a different location in the ancestral segment. Therefore, the modern segment of repeats must have been created in the higher primates after their divergence from the prosimians. Involucrin genes from two other higher primates - gorilla and owl monkey - have now been sequenced. All possess a modern segment composed of two types of repeats, designated A and B, distinguished by the first three amino acids of the repeat. The modern segment can be divided into three regions - early, middle and late. The early region is shared by all three higher primate species and thus must have originated in a common ancestor of the anthropoids. The middle region is present in human and gorilla genes, but not in the owl monkey and is probably specific to the hominoids. The late region is unique to each species examined. The modern segment has been generated by an apparently non-random process of successive duplication that has continued over a large part of the evolutionary history of the higher primates.

CF 146 HOMINOID IMMUNOGLOBULIN ALPHA GENES: MULTIPLE RECOMBINATIONAL EVENTS IN THEIR HINGE REGIONS, Shintaroh Ueda and Shohji Kawamura, Department of Anthropology, Faculty of Science, The University of Tokyo, Tokyo 113, Japan
The heavy-chain constant region genes of the human immunoglobulin are located on chromosome 14 in the order 5'-Cμ-Cδ-Cγ3-Cγ1-Cε2-Cα1-ψCγ-Cγ2-Cγ4-Cε1-Cα2-3'. Immunoglobulin epsilon and alpha genes were isolated from chimpanzee, gorilla, orangutan, gibbon and crab-eating monkey DNAs and their structures were compared with the human counterparts.
Multiple deletions and duplications seem to have taken place in both genes during hominoid evolution, especially in the hinge regions of Cα genes; humans, chimpanzees, gorillas and gibbons had two Cα genes in their genomes, while orangutans and Old World monkeys had one Cα gene. Moreover, there were great varieties in length of their hinge regions.
Assuming that the branching order is Old World monkeys-gibbons-orangutans-gorillas-chimpanzees-humans, the simple evolutionary steps are the following; duplication of Cα gene took place independently in the gibbon lineage, and independent deletions or duplications of a part of the hinge region of Cα gene took place in each lineage of Old World monkeys, gibbons and orangutans. In the common ancestor of humans, chimpanzees and gorillas, duplication of DNA containing Cε and Cα genes happened. Then, deletion of a part of the hinge region took place to yield the Cα2 gene. After divergence of gorillas, there was duplication of the hinge region in the Cα1 gene of the common ancestor of humans and chimpanzees.
These results suggest that the hinge region of Cα gene is a hot spot for DNA rearrangement.

CF 147 HOMOGENIZATION BY CONVERSIONS IN WHEAT rDNA SPACERS, Itamar Vinizky, Rudy Appels, and Jan Dvorak, Department of Agronomy and Range Science, University of California, Davis, CA 95616. Wheat (AA BB DD genomes) Nor loci coding for 18S-5.8S-26S rRNA (rDNA) is a favorable system for studying the evolution of tandemly arranged repeated sequences. We have cloned and sequenced the spacer of one randomly selected spacer of a rDNA unit for the Nor-D3a allele in T. tauschii, the ancestor of wheat D genome and compared it with the spacer sequence of the unit from the same allele present in T. aestivum. The two alleles have been isolated by at least 8,000 years since the origin of wheat. They have identical structure of which the spacer consists of nine repeats tandemly arranged and each is 120 bp long. The two clones differed by only three mutations in its repeated array which is 1,082 bp long. In one region, a single mutation relative to the consensus sequence appeared to be transferred into another repeat, and in another region a cluster of three mutations appeared to be transferred. Since the two rDNA clones have identical numbers of repeats in the spacer, it is unlikely that the two cases of homogenitation were due to unequal crossing over within the array. The data indicate that they should have arisen by sequence conversions between repeats. It is concluded that in wheat the principal mechanism of homogenization in the repeated array of the spacer is nucleotide sequence conversions.

CF 148 INTRAGENERIC VARIATION IN REPETITIVE SEQUENCES ISOLATED BY PHYLOGENETIC SCREENING OF MAMMALIAN GENOMES, Holly A. Wichman, Thomas Payne and Tod W. Reeder, Dept. of Biological Sciences, Univ. of Idaho, Moscow, ID 83843. Phylogenetic (Φ) screening is a procedure developed for the isolation of rapidly evolving repetitive sequences. The concept behind the Φ-screen is straightforward: sequences which are repetitive in the genome of one species but are absent or in low copy number in the the genome of a related species are likely to have undergone rapid evolution. Both retro-transcripts and tandemly repeated sequences have been isolated using this procedure. We have screened two mammalian genomes with the Φ-screen procedure. The deer mouse (Peromyscus leucopus) genome was screened using house mouse as an outgroup, and the zebra (Equus zebra) genome was screened using rhinoceros as an outgroup. We have isolated many more clones containing rapidly evolving repetitive sequences in the zebra Φ-screen than in the deer mouse Φ-screen. In addition, some repetitive families isolated from the zebra genome show both extensive variation in copy number and restriction fragment length polymorphisms (RFLPs) within the genus Equus. In contrast, tandemly repeated sequences isolated from the deer mouse genome show some variation in copy number within the genus Peromyscus, but only minor RFLPs. Mys, a family of retrotransposons isolated from Peromyscus, shows major RFLPs between species of this genus.

CF 149 MOLECULAR EVOLUTION OF HIGHLY REPEATED DNA ELEMENTS IN HUMAN FILARIAL PARASITES, Steven A. Williams and Daniel J. Freedman, Department of Biological Sciences, Smith College, Northampton, MA 01063 and Program in Molecular and Cellular Biology, University of Massachusetts, Amherst, MA 01003.
We have cloned and characterized members of a repeated DNA sequence family found in filarial parasites of the genus Brugia. Several species in this genus are the causative agents of lymphatic filariasis and elephantiasis in more than 100 million persons living in the tropics. We have designated this repeat the HhaI repeat family because it is cleaved by that restriction endonuclease in all species of Brugia examined to date. These repeats are 322 base pairs in length, are organized in long head-to-tail tandem arrays, and make up 12% of the total Brugia genome. Due to the high copy number of this repeat (30,000 copies per haploid genome), cloned copies of this sequence can serve as extremely sensitive probes for the detection of Brugia in blood samples or in mosquitoes.

More than sixty copies of this repeat have been cloned and sequenced from various species and strains of Brugia. These data are being used to estimate genetic variation within and between populations and to construct phylogenetic trees. Restriction endonuclease mapping, DNA sequencing, Southern blot analysis and pulsed-field gel electrophoresis techniques are being used to study the organization of the HhaI repeat clusters on Brugia chromosomes. The organization of these clusters is being compared between Brugia species in an effort to deduce the evolutionary history of the HhaI repeats.

CF 150 GENERATION OF A NEW-LENGTH ALLELE AT A VNTR LOCUS WITHOUT EXCHANGE OF FLANKING DNA SEQUENCE, Roger Wolff[1], Yusuke Nakamura[2] and Ray White[1,2], [1]Department of Cellular, Viral and Molecular Biology, University of Utah, Salt Lake City, UT 84132, [2]Howard Hughes Medical Institute, University of Utah, Salt Lake City, UT 84132.
Variable-number-tandem-repeat (VNTR) DNA markers are contributing new power to human genetic studies because their hypervariable nature allows individualization at the DNA level. The practical value of VNTR markers has been well established for genetic linkage mapping, forensic biology, paternity testing, and monitoring of bone marrow transplants. A popular hypothesis attributes generation of variability at VNTR loci to unequal exchange between homologous chromosomes at meiosis; predicting that flanking markers will be recombinant. We report here the observation of a newly generated length-allele at VNTR locus D17S5, and characterize it in terms of unequal exchange between homologous chromosomes. Sequencing of the two maternal alleles from which the new allele was derived revealed base substitutions between the alleles flanking the VNTR. Upon sequencing the daughter's new allele, it became apparent that contrary to the prediction of the hypothesis, the new allele was generated by loss of one perfect 70-bp repeat unit, without exchange of flanking DNA sequences. The results are consistent with models that employ sister chromatid exchange, polymerase slippage, and loopout deletion. However, we cannot rule out the possibilities of a double crossing-over event, or a gene conversion that allows a single crossing-over event to be resolved into parental flanking markers. Study of more events that generated new alleles will be required before we can distinguish among the various models.

Evolution of Prokaryotes and Viruses

CF 200 RESOLUTION OF RECENT EVOLUTIONARY DIVERGENCES AMONG *E. Coli* FROM RELATED LINEAGES. Robert D. Arbeit, Michel Arthur, Roberta Dunn, Robert K. Selander, and Richard Goldstein. Departments of Medicine and Molecular Epidemiology, Boston University Schools of Medicine and of Public Health, Boston, MA 02118 and Department of Biology, Pennsylvania State University, University Park, PA, 16802.

E. coli isolates, both commensal and pathogenic, are distributed among independently evolving cell lines. These diverging lineages have been resolved and analyzed using multilocus enzyme electrophoresis. *E. coli* infecting the urinary tract and other extraintestinal sites belong to a limited number of lineages. As a result, isolates obtained at widely different times and places are often indistinguishable and, therefore, epidemiologic relationships among specific isolates cannot be resolved. We analyzed uropathogenic *E. coli* by digesting total chromosomal DNA with eight-base pair restriction enzymes and resolving the resulting fragments using pulse field gel electrophoresis (PFGE). All isolates obtained from different patients, including isolates belonging to the same or closely related lineages, had distinct and unique restriction fragment profiles. In contrast, isolates obtained from different sites of infection within the same patient had identical restriction profiles, indicating that these polymorphisms were stable *in vivo* over short time periods. This analysis (i) defines a previously unresolved level of genetic variability indicating recent, divergent evolution and is therefore useful for epidemiologic analysis, and (ii) provides genetic evidence that *E. coli* infection is due to expansion by a single strain.

CF 201 YELLOW FEVER VIRUS EVOLUTION: COMPARATIVE ANALYSIS OF THE NUCLEOTIDE AND DEDUCED AMINO ACID SEQUENCES OF THE STRUCTURAL GENES OF TWO GEOGRAPHICALLY ISOLATED STRAINS. Mary E. Ballinger and Barry R. Miller, Division of Vector Borne Diseases, Center for Infectious Diseases, Centers for Disease Control, PHS, HHS, Fort Collins, Colorado 80522

Geographically distinct isolates of yellow fever virus (YFV) have been shown to vary in certain biological and physical characteristics. Specifically, studies have shown differences between South American and African YFV strains in their virulence for mice and in the physical properties of their envelope glycoprotein. Additionally, genetic heterogeneity has been demonstrated between such strains by oligonucleotide fingerprinting. In order to investigate the molecular basis for these variations, we have determined the nucleotide sequence of the structural genes of a Peruvian YFV isolate (1899/81) and compared it with the published (Hahn, et al.) sequence of the corresponding region of the African YFV Asibi strain. Additionally, a comparison was made of the deduced amino acid sequences of the two isolates. These YFV strains have evolved in separate ecological systems for presumably hundreds of years and have thus responded to different selection pressures imposed by environment, invertebrate vectors, and vertebrate hosts. As may be expected between RNA virus strains which have evolved in such varied ecosystems, we observed a great deal of hetergeneity at the nucleotide level. However, the deduced amino acid sequences of the two strains were remarkably homologous.

CF 202 TAXONOMIC RELATIONSHIPS BETWEEN *THERMUS* SPECIES AS DEDUCED BY NUMERICAL TAXONOMY AND RIBOSOMAL RNA ANALYSIS, PL Bergquist[1], JA Hudson[2], H Morgan[2], C Wallace[1] and D Penny[3], [1]Auckland University, Auckland, [2]Waikato University, Hamilton and [3]Massey University, Palmerston North, New Zealand.

The bacterial genus *Thermus* currently comprises two subgroups, one with organisms growing optimally at 70°C and another at 55-60°C, based on phenotypic, chemotaxonomic and DNA/DNA hybridization data. The higher temperature group of *T. aquaticus*-like strains has been subdivided by numerical phenetic techniques into groups which appear to reflect their geographical source and may demonstrate a requirement for the creation of new species. Correlations between these groups and taxonomic status as deduced from 16S ribosomal RNA sequences will be presented.

CF 203 UNUSUAL GENE ORGANIZATION OF THE RIBOSOMAL RNA OPERON IN THE INTRACELLULAR BACTERIUM RICKETTSIA BELLII, Paul A. Fuerst and Jonathan B. Clark, Department of Molecular Genetics, The Ohio State University, Columbus, OH 43210.
Rickettsia are obligate intracellular parasites of eukaryotic cells, most commonly associated with arthropod hosts, and are members of the α-subgroup of the purple bacteria. Studies in our laboratory indicate that the genome of Rickettsia possesses as single rRNA cistron. Sequences of the 16S rRNA gene have been collected for phylogenetic analysis. In the course of these studies, nucleotide sequences flanking the 16S rRNA gene have also been determined. In one species, R. bellii, these flanking sequences suggest that the structure of the rRNA operon is unusual compared to other eubacteria. The 3'-flanking sequences of the 16S gene have been determined for about 1 kb following the gene. This sequence contains no tRNA, nor was the beginning of the 23S gene located. Southern blots of the 16S, 23S and 5S rRNA genes from R. bellii, R. prowazekii and R. montana indicate that the genes are linked closely in the latter two (as seen in other eubacteria), but that the 16S rRNA gene is not linked closely to the 23S-5S cluster in R. bellii. The 3'-flanking sequence from R. bellii contain a 14bp perfect inverted repeat sequence 405 bases downstream from the end of the 16S rRNA gene. **This element occurs at least 20 times in the genome of R. bellii but does not occur** in other species of Rickettsia. Other studies have indicated that repeated genetic sequences are rare in Rickettsia. **The relationship of this element with the 23S-5S rRNA gene cluster** is being investigated to determine whether it could be implicated in a genomic rearrangement which has separated the rRNA genes. The structure of the flanking sequences raises the possibility that the different rRNA genes in R. bellii are not cotranscribed.

CF 204 MOLECULAR POPULATION STUDIES OF INDIGENOUS SOUTHERN AFRICAN GEMINI VIRUSES. Ralph Kirby, Fiona Hughes and Edward Rybicki, Department of Microbiology, Upper Campus, University of Cape Town, Rondebosch, Cape Town, R.S.A.
Restriction enzyme polymorphism analysis of the dsDNA replicative form of a variety of indigenous southern African gemini viruses was carried out. A novel gemini virus, sugar cane streak virus (SCSV) was identified. A wider variety of maize streak viruses (MSV) than expected were found. The data was analysed to give a possible evolutionary tree for the known gemini viruses.

CF 205 THE PHYLOGENETIC RELATIONS OF DNA-DEPENDENT RNA POLYMERASES TESTIFY TO THE CHIMERIC ORIGIN OF THE EUKARYOTIC NUCLEAR GENOME AND THE COHERENCE OF THE ARCHAEBACTERIA. Hans-Peter Klenk [1], Peter Palm [1], Gabriela Pühler [1], Felix Gropp [1], Henrik Leffers [2] and Wolfram Zillig [1], [1] Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG, [2] Biostructural Chemistry, Kemisk Institut, Aarhus University, DK-8000 Aarhus C, Denmark.
The amino acid sequences of the A and C subunits of the DNA-dependent RNA polymerases of the archaebacteria *Sulfolobus acidocaldarius*, *Halobacterium halobium* and *Methanobacterium thermoautotrophicum*, were aligned with 11 corresponding sequences including representatives of the three eukaryotic nuclear DNA-dependent RNA polymerases and eubacterial β' subunits. Unrooted phylogenetic dendrograms were derived by distance matrix and parsimony methods including the evolutionary parsimony algorithm developed by J. A. Lake. The distance matrix dendrogram was subjected to various branch length corrections. The results obtained by the different methods testify for the coherence of the archaebacteria. The eukaryotic polymerase I lineage shares a common bifurcation with the eubacterial β' lineage and appears to arise separately from the lineage of the eukaryotic polymerases II und III. This suggests a chimeric origin of the eukaryotic nuclear genome.

CF 206 IS NATURAL SELECTION THE COMPOSER AS WELL AS THE EDITOR OF GENETIC VARIATION?
B. Levin[1], D. Gordon[1] and F.M. Stewart[2], [1]Department of Zoology, University of Massachusetts, Amherst, MA 01003 and [2]Department of Mathematics, Brown University, Providence, RI 02912
In a recent article; Cairns, Overbaugh and Miller (NATURE, 335:142-145) present evidence they interpret as support for the hypothesis that E. coli "have mechanisms for choosing which mutations occur". In this report we examine two lines of their evidence for this phenomenon of 'dirercted mutation'; i) the distribution of mutants from independent (fluctuation test) cultures; and ii) the continuous occurrence of new mutants in stationary phase cultures maintained in the presence of the agents selecting for their ascent. We present the results of the analysis of a model of random mutation that allows for multiple mutant states, fitness and mutation rate differences among mutant and parental populations and changes in mutation rate with the physiological state of the bacteria (growth conditions). We consider the fit of this theory to the distribution of revertants in experiments with three negative mutations in E. coli K-12, a nonsense (amber) mutation in lacZ, an amber mutation in the trp gene, and a frameshift mutation at lacZ locus. We present evidence that almost all the "mutants" accumulating in stationary phase, selecting medium were originally present in the fluctuation test cultures. While it is not possible to formally rule out some directed mutations, the results of our study indicate the vast majority (possibly all) of mutations in these experiments can be accounted for by random mutation with the environment acting solely as a selection agent.
*EcLF is an acronym for the E. coli Liberation Front. Members of the EcLF contributing to this endeavor include R. Evans, D. Gordon, D. Lebel, B. Levin, L. Simonsen (Dept. of Zoology, University of Massachusetts, Amherst MA 01003) and F. Stewart (Dept. of Mathematics, Brown University, Providence, RI 02912).

CF 207 **Evolution of AIDS viruses**, **Wen-Hsiung Li and Masako Tanimura, Center for Demographic and Population Genetics, University of Texas, Houston, TX 77030, and Paul M. Sharp, Department of Genetics, Trinity College, Dublin 2, Ireland.**
The acquired immune deficiency syndrome (AIDS), caused by a retrovirus called Human Immunodeficiency Virus (HIV), has become a pandemic. A knowledge of the rate of nucleotide substitution in HIV and the history and pattern of spread of the virus is important for understanding the epidemiology and pathogenesis of AIDS and for developing therapies and vaccine strategies. A new model has been developed and used to estimate the substitution rates in various regions in the HIV genome. The rate of nonsynonymous (amino acid changing) substitution is lowest in the regions coding for the capsid proteins and the reverse transcriptase, being ~ 1.7×10^{-3} nucleotide substitutions per site per year. The nonsynonymous rate is extremely high (14×10^{-3}) in the hypervariable regions of the envelope gene, suggesting extremely rapid change in viral antigenicity. The nonsynonymous rates in the other coding regions are between 3×10^{-3} and 7×10^{-3}. The average synonymous rate for the HIV genome is 10×10^{-3}. Evidence is provided for a case of recombination between different HIV strains. Our analysis suggests that the AIDS virus had existed in Central Africa before 1960 and spread to North America before the mid 1970's. The evolutionary relationships among HIV isolates are inferred from nucleotide sequence data and the result is consistent with the view that AIDS spread from Haiti to the United States. Our analysis suggets that HIV-1 and HIV-2 diverged about 150 years ago and that the African green monkey virus (SIV_{AGM}) is slightly closer to HIV-1 than to HIV-2.

CF 208 RAPID EVOLUTION IN GENOME-LIMITED CELLS (THE MYCOPLASMAS), Jack Maniloff, Department of Microbiology and Immunology, University of Rochester, Medical Center Box 672, Rochester, NY 14642.
The mycoplasmas are a group of microorganisms with the smallest reported cell and genome sizes. These cells arose by degenerate evolution from Gram-positive bacteria. The earliest phylogenetic event was a genome reduction and loss of cell wall (characteristic of all mycoplasmas), producing saprophytic mycoplasmas with genome sizes about 1500 kb. Subsequent branching led to a sterol-requirement (characteristic of many mycoplasmas), and cells with genome sizes about 1500 kb and plant and insect habitats. At least three nodes on this plant-insect mycoplasma branch led to mycoplasmas with genome sizes about 700 kb. Since there were multiple origins of the small genome mycoplasmas and these sublines all have genomes about 700 kb, this must represent the minimal genetic capacity for free-living cells. Cells with such small genomes may have higher mutation rates than larger genome size bacteria. To examine this possibility, the transition:transverison ratio has been analyzed in 5S rRNA and tRNA sequences. This ratio is several-fold higher in mycoplasmas than in phylogenetically related bacteria. Hence, mycoplasmas must be relatively deficient in proofreading and/or DNA repair, and have an increased mutation rate. However, the sequence analyses also show the mycoplasma mutation rate is only slightly faster than in bacteria. This indicates that, although mycoplasmas have a fast mutation rate, their fixation probability must be smaller than in bacteria. The attrition in genome size during mycoplasma evolution may have led to a reduction in nonfunctional DNA, thereby constraining the fixation probability.

CF 209 EVOLUTION OF THE GENE ENCODING THE GLYCOPROTEIN OF VESICULAR STOMATITIS VIRUS NEW JERSEY, Stuart T. Nichol[1], Joan E. Rowe[1], and Walter M. Fitch[2], [1]Cell and Molecular Biology Program, School of Veterinary Medicine and Department of Microbiology, University of Nevada, Reno, NV 89557 and [2]Molecular Biology, Ahmanson Center for Biological Research, University of Southern California, Los Angeles, CA 90089-1340.
The surface glycoprotein (G) of vesicular stomatitis virus (VSV) is the major protein to which neutralizing antibodies are elicited. We have previously shown by T1 RNase fingerprinting that a considerable extent of genetic diversity exists amongst natural isolates belonging to the VSV New Jersey (NJ) serotype. To determine the extent of variability and evolution of the G gene and its encoded protein, a nucleotide sequence analysis of virus G genes was performed. Based on earlier T1 fingerprinting results, 34 VSV NJ natural variants were selected for primer extension dideoxynucleotide sequencing analysis of virus G genes. Approximately 54 Kb of nucleotide sequence information has been obtained, representing the complete G gene sequence for each virus isolate. A considerable number of base differences were detected on comparison of these viruses. An analysis of the pattern of nucleotide and predicted amino acid variability, and the evolutionary significance of these changes relative to other virus proteins will be presented.

CF 210 MOLECULAR EVOLUTION OF DNA-DEPENDENT RNA POLYMERASE OF THE ARCHAEBACTERIUM SULFOLOBUS ACIDOCALDARIUS, Gabriela Pühler and Wolfram Zillig, Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG.
DNA-dependent RNA polymerases are multi-component enzymes existing in all organisms. The ubiquity and the fundamental role of these enzymes make them well suited, beside the ribosomes, as an independent molecular clock for determining phylogenetic relations. For this purpose, both the organization of the genes of their components and the derived amino acid sequences have been compared. The genes coding the three largest subunits (B, A, C) of the RNA polymerase of the archaebacterium Sulfolobus acidocaldarius DSM 639 were cloned and sequenced. Amino acid sequence comparisons of the S.acidocaldarius enzyme with yeast, Drosophila and mouse RNA polymerases I, II and III showed an extensive homology between the combined Sulfolobus subunits A and C and the largest eukaryotic subunit (A). The Sulfolobus B subunit is highly homologous to the second largest eukaryotic subunit (B) in all these cases. Considerably less homology could be found between the Sulfolobus subunits and the E. coli subunits β' and β. On the other hand, the RNA polymerase genes of S. acidocaldarius (B, A and C) are organized in a similar though not identical way as the homologous rpoB (β) and C (β') genes of E. coli. Transcription initiation sites were identified by S_1-mapping and these structures were then compared with known eukaryotic and eubacterial promoter sequences, the recognition sites for DNA-dependent RNA polymerase.

CF 211 MISMATCH-STIMULATED ANTIRECOMBINATION BY LONG PATCH MISMATCH REPAIR(LPMR) AS THE MOLECULAR MECHANISM OF SPECIATION. C. Rayssiguier and M. Radman.Laboratory of mutagenesis,Institut Jacques Monod, C.N.R.S. University Paris 7, Paris (France).
The prediction of M. Radman (Mismatch repair and fidelity of genetic recombination,1988, Genome,in press) that LPMR is the primary cause of sterility in interspecific and intergeneric crosses was tested in bacterial crosses. The time of divergence between the two genera E.Coli and Salmonella typhimurium has been estimated by Ochman and Wilson(Evolutionary history of enteric bacteria,1987, In:Escherichia coli and Salmonella typhimurium,vol 2, edited by Neidhardt,American Soc. for Microbiol. Washington,D.C.,pp1649-1654) to 120 to 160 millions years;conjugational crosses between the two give a frequency of integration of donor DNA of 10^{-8} per donor cell as compared to 10^{-1} in intraspecific crosses(Baron et al., 1968 Intergenericbacterial matings,Bacteriol. Rev. 32, 362-369). The prediction that mismatch repair deficiency should activateintergeneric recombination was fulfilled: when restriction deficient F⁻ Salmonella typhimurium recipient of E.coli Hfr,DNAismismatchrepairdeficient the frequency ofintergeneric recombinants increases by 10^3 to 10^7 fold.MutL and S mutants give the strongest effect then mutH and mutU. This effect is not observed if the recA mutation is introduced into the recipients. The implication of this observation on the molecular mechanism of speciation through the antirecombinogenic effect of the E.coli LPMR (MutHLSU system) will be discussed.

Population Inferences and Data Analysis

CF 300 CONSERVATION OF AMINO ACID SEQUENCES IN ALBUMIN: IS ALBUMIN AN ESSENTIAL PROTEIN? Michael E. Baker, Dept. of Medicine, M-023, University of California, San Diego, La Jolla, CA 92093 It is puzzling that albumin, the most abundant protein in vertebrate serum, is thought to be a non-essential protein. The principle basis for this notion is that "analbuminemic" humans and rats appear to be otherwise healthy, apparently because other serum proteins take over albumin's actions to bind and transport hydrophobic ligands and to maintain osmotic pressure in blood. However, analbuminemic humans have 10 μg/ml albumin in their serum, so they are not truly analbuminemic. Moreover, 10 μg/ml albumin is high compared to many essential hormones and growth factors. This suggested to us that albumin may indeed have essential functions, in which case, one would expect to find parts of the protein under constraints as far as changes in the amino acid sequence are concerned. To test this, we compared the amino acid sequences in each exon of human albumin and alpha-fetoprotein (AFP), two paralogous proteins, that separated from a common ancestor about 400 Myr ago. We find that the last 3 exons, containing 133 residues, are 50% identical, without the need to insert any gaps in the alignment. For comparison, the alpha and beta chains of human hemoglobin, 2 paralogous proteins that have been separate for about 400 Myr, are 45% identical. Thus, the last 3 exons of albumin and AFP are changing more slowly than hemoglobin, suggesting that this part of albumin and AFP has some essential, albeit as yet unidentified, function.

CF 301 THE USE OF NUCLEAR AND MITOCHONDRIAL DNA POLYMORPHISMS TO INVESTIGATE THE POPULATION GENETICS, EPIDEMIOLOGY AND EVOLUTION OF OPHIOSTOMA ULMI - THE CAUSATIVE AGENT OF DUTCH ELM DISEASE, Malcolm R. Bates, Kenneth W. Buck, Department of Pure and Applied Biology, Imperial College of Science, Technology and Medicine, London SW7 2BB, UK and Clive M. Brasier, Forest Research Station, Alice Holt Lodge, Farnham, Surrey, GU10 4LH, UK. The current devastating epidemics of Dutch elm disease in North America, Europe and South East Asia are caused by the emergence and spread of a highly pathogenic, aggressive sub-group of the fungus Ophiostoma (=Ceratocystis) ulmi, which is rapidly replacing the 'old' non-aggressive sub-group, believed to be responsible for milder disease epidemics earlier in the century. The aggressive sub-group has previously been divided into the Eurasian (EAN) and North American (NAN) races on the basis of a range of biological characteristics. RFLP analysis of both nuclear and mitochondrial DNA efficiently distinguishes between EAN, NAN and non-aggressive isolates and allows an assessment of their genetic variability. Their was a significantly greater variation in the mitochondrial DNA within the NAN race than within the EAN race or non-aggressive sub-group. In addition both the nuclear and mitochondrial data suggest a single origin for the aggressive races.

CF 302 MOLECULAR EVOLUTION OF THE WIDELY DISTRIBUTED MARINE FISH HAKE (*Merluccius*). Inga I. Becker, Ralph Kirby and Stuart Grant, Department of Microbiology, Upper Campus, University of Cape Town, Rondebosch, Cape Town, R.S.A.
Mitochondrial DNA from four species of *Merluccius* were mapped using a variety of restriction enzymes. This study showed that the sympatric species, *Merluccius capensis* and *Merluccius paradoxus* found off the coast of southern African coast were distinct species which separated about 5.8 million years ago. A proposed evolutionary tree and migration route for *Merluccius capensis*, *Merluccius paradoxus*, *Merluccius productus* and *Merluccius australis* will be presented. DNA sequence data on the organisation of the *Merluccius* mitochondrial genome will also be presented

CF 303 MOLECULAR EVOLUTION OF BIOLOGICAL "CLOCK" GENES IN DROSOPHILA, Eldredge Bermingham and Charles Aquadro, Section of Genetics and Development, Cornell University, Ithaca, NY 14853. The *period* locus affects several circadian phenotypes and male courtship song in *Drosophila*. Published work on this locus has focused on nucleotide variants distinguishing mutant clock phenotypes from a single "wild-type" phenotype. We will report results characterizing DNA sequence variation among several alleles of per isolated from natural populations of both *Drosophila melanogaster* and *D. simulans*. These data permit direct analysis of the types of sequence variants underlying extensive restriction site polymorphism our lab has found at this locus in natural populations of these species. In addition, comparison of synonymous to nonsynonymous substitution rates within and between these two species allow a direct test of Ohta's model of slightly deleterious alleles and molecular evolution as proposed for contrasting patterns of DNA and protein polymorphism revealed by restriction site analyses of *per* and other loci in these two species by Aquadro, Lado and Noon (1988, Genetics 119:875-888). The observation that motivated this hypothesis is the contrasting levels of DNA versus protein polymorphism in these sibling species. Populations of *simulans* show less protein polymorphism and differentiation than *melanogaster*, yet *simulans* is four to six times more variable than *melanogaster* at the DNA restriction map level at several unlinked genes. Our study of DNA sequence variation at the per locus within and between species of *Drosophila* will also allow us to carry out direct tests of the relative contributions of balancing selection, varying selective constraint and varying mutation to the high levels of nucleotide polymorphism observed at the per locus compared to flanking sequences, and will contribute to our understanding of the evolution of male courtship song and circadian rhythms.

CF 304 EFFECTS OF LINKAGE ON RATES OF MOLECULAR EVOLUTION

C. William Birky, Jr. and J. Bruce Walsh, Department of Molecular Genetics, The Ohio State University, Columbus, OH 43210 and Department of Ecology and Evolution, The University of Arizona, Tucson, AZ 85721. The fixation of mutant alleles at a site may be affected by the occurrence of selected mutations at linked sites, as seen in the phenomenon of hitchhiking. The consequences of hitchhiking for genetic diversity are well known. In contrast, little is known about the consequences of linkage for rates of base pair substitution. We used computer simulations to show that complete linkage to either advantageous or detrimental mutations does not affect the substitution rate of neutral mutations at a site. However, linkage to either advantageous or detrimental mutations in the background decreases the substitution rate of advantageous mutations, and increases the substitution rate of detrimental mutations, at a site (Birky & Walsh 1988 Proc. Nat Acad Sci. USA 85:6414). The results can be understood in terms of the Hill-Robertson effect, in which selection reduces the effective population size. We are doing further simulations to determine if these effects are significant for very small genomes such as those in organelles, where linkage may be complete or nearly so but the total number of genes, and hence the total background mutation rate, is very low. We are also testing the strength of the effect on very large genomes (as in eukaryotic nuclei) with realistic frequencies of recombination.

CF 305 AVERAGE VALUES OF A DISSIMILARITY MEASURE NOT REQUIRING SEQUENCE ALIGNMENT ARE TWICE THE AVERAGES OF CONVENTIONAL MISMATCH COUNTS REQUIRING SEQUENCE ALIGNMENT. B. Edwin Blaisdell, Linus Pauling Institute of Science and Medicine,Palo Alto, CA 94306. Three measures of sequence dissimilarity have been compared on a computer generated model system in which substitutions in random sequences were made at randomly selected sites with randomly selected characters. The three measures were the conventional mismatch count between aligned sequences and two measures not requiring prior sequence alignment. These two measures were the squared Euclidean distance between vectors of counts of t-tuples (t=1 to 6) of characters (MDD) and counts of characters not covered by word structures of statistically significant length common to the two sequences. Average MDD distances were found to be two times average mismatch counts for all values of t from 1 to 6 and all degrees of substitution from one per sequence to so many as to produce effectively random sequences. This simple relation held independently of sequence length and of sequence composition. The relation was confirmed by exact results on small model systems and by formal asymptotic results in the limit of very sparse substitutions and in the limit of two random sequences. The coefficient of variation for MDD distances was greater for mismatch counts for singlets but both measures approached the same low value for sextets. Needleman-Wunsch alignment produced incorrect mismatch counts at higher degrees of substitution. Application of the Jukes-Cantor asymptotic adjustment produced increasingly bad results with increasing degrees of substitution. Average CLW distances for a variety of common word structures were more or less parallel to MDD distances for appropriately long t-tuples. These results on model systems supported the validity of the two dissimilarity measures not requiring sequence alignment that was found in earlier work on natural sequences.

CF 306 INFERENCE OF DNA HAPLOTYPES FROM PCR-AMPLIFIED POPULATION SAMPLES, Andrew Clark, Department of Biology, Penn State University, University Park, PA 16802.

Suppose one has a pair of primers for a single copy gene and wishes to obtain information on sequence variation in a population using PCR amplification of genomic DNA. If it is possible to isolate a single haploid genome, by assaying haploid tissue (as in many plants) or by genetic manipulations (as with Drosophila), then haplotypes can be obtained unambiguously. If diploid and possibly heterozygous material must be used, then all of the polymorphic sites will lead to ambiguities in inference of haplotypes from sequencing gels. In particular, if a sample shows n polymorphic sites, then there are 2^n possible haplotypes. With a population sample, some haplotypes will be obtained unambiguously from individuals that are homozygous at all polymorphic sites. In many cases, haplotypes of heterozygotes can be obtained by "subtracting off" other known haplotypes. By a series of such subtracting operations, it may be possible to assign all haplotypes in a sample, but this solution is often not unique. The probability of success of this approach will be analyzed theoretically and an algorithm for extracting haplotypes will be presented. The success of the method depends on the level of sequence variability, the level of heterozygosity, the degree of linkage disequilibrium, the sample size, and the length of the DNA segment.

CF 307 TAXONOMIC RELATIONSHIPS AMONG ISOLATES OF THE FUNGAL GENUS BASIDIOBOLUS REVEALED BY RESTRICTION ANALYSIS OF RIBOSOMAL DNA, Rex T.Nelson, Bienvenido Yangco, Diane TeStrake, and Bruce J. Cochrane, Department of Biology and College of Medicine, University of South Florida, Tampa, FL 33620 The genus Basidiobolus is a widespread taxon, that includes strains associated with human phycomycoses, as well as nonpathogenic ones that are saprophytic in reptiles and amphibians. Based upon morphological criteria, the existence of four species (B. haptosporus, B. meristosporus, B. microsporus, and B. ranarum) has been proposed; however, the validity of this classification has been questioned. We have cloned 9 kb of ribosomal DNA from B. ranarum and employed it as a probe to analyze restriction patterns from pathogenic and saprophytic isolates. Patterns observed were highly heterogeneous, and indicate the existence of extensive rDNA polymorphism within isolates. Most pathogenic, isolates previously grouped into three species, are indistinguishable. B. microsporus is clearly distinct from this group, as are two saprophytic isolates. Thus, we can group the genus into three monophyletic groups. Digestion of DNA with some enzymes reveals pathogen-specific fragments that may prove useful in diagnostic applications. These results, combined with others involving immunological and allozyme analyses, suggest that existing taxonomies based on morphological criteria may overestimate the taxonomic diversity among pathogenic isolates and underestimate the extent of genetic diversity in the genus as a whole.

CF 308 AMINO ACID POLYMORPHISM AT THE ESTERASE-6 LOCUS OF *DROSOPHILA MELANOGASTER*, Peter H. Cooke and John G. Oakeshott, Division of Entomology, CSIRO, Canberra, Australia. High resolution electrophoresis has revealed ten allozymes of esterase-6 in *Drosophila melanogaster*. The sequences of 13 isolates of the *Est6* gene covering all ten allozymes were obtained and fifty-two nucleotide differences were found. Sixteen of these cause amino acid replacements, of which three result in charge differences whose size and direction are consistent with the electrophoretic mobilities of the allozymes in which they occur. The smeared electrophoretic phenotype of one allozyme can be explained by the loss of a cysteine residue involved in a disulphide bridge. Several minor mobility variants within the major F and S electrophoretic phenotypes differ by amino acid differences which are generally conservative for charge but not for some other properties like size, polarity or hydrophobicity. Four amino acid differences are found among different isolates of the same allozymes and, overall, 12 amino acid haplotypes occur among the 13 isolates sequenced. Nevertheless, the most common variants within F and S are only distinguished by two amino acids (Asn/Asp at 237 and Thr/Ala at 247) and these are the most likely targets for the selection underlying complementary latitudinal clines in F and S frequencies.

CF 309 THE ANATOMY OF MOLECULAR ADAPTATION: RECENT DUPLICATION AND RAPID DIVERGENCE OF P LYSOZYME. Gino A. Cortopassi and Allan C. Wilson, Department of Biochemistry, University of California, Berkeley, CA 94720 A cDNA sequence was obtained for the Paneth-cell form of mouse prelysozyme, designated P, by enzymatic amplification and direct sequencing. The P coding sequence differs from that of M, the macrophage prelysozyme, by 21 base substitutions, only 5 of which are synonymous. This observation and the occurrence of base differences at 5% of the positions in the 5' and 3' non-coding regions implies that M and P genes are related by a duplication that took place 5 to 10 million years ago, i.e. after the rat and mouse lineages diverged. Phylogenetic analysis, conducted with rat lysozyme as an outgroup, suggests that while 11.5 amino acid replacements happened on the P lineage, there were only 3.5 on the mouse M lineage. The possibility that positive selection contributed to accelerated amino acid replacement on the P lineage is suggested by the high ratio of replacements to synonymous substitutions (15:4). The non-random distributions of nucleotide mutations with respect to individual codons, amino acid dispensability, and effect on protein net charge are consistent with the view that positive selection has driven the evolution of P lysozyme.

CF 310 MOLECULAR BASIS OF EVOLUTIONARY ADAPTATION IN TWO LATITUDINALLY EXTREME POPULATIONS OF *FUNDULUS HETEROCLITUS*. Douglas L. Crawford and Dennis A. Powers, Department of Biology, Johns Hopkins University, Baltimore, MD, 21218. Adaptation to the physical environment occurs at all levels of biological organization: behaviorally, morphologically, physiologically and biochemically. The heart form of lactate dehydrogenase (LDH-B) in *Fundulus heteroclitus* has two allelic isozymes which show clinal variation in gene frequency: *Ldh-B*^{b} is virtually fixed in northern populations while *Ldh-B*^{a} predominates in southern populations. These two allelic isozymes are kinetically different at there respective TM. Additionally, as shown in this study, the concentration of LDH enzyme and mRNA is approximately twice as great in the northern population. The increase in the northern LDH-B enzymes appears to be a function of the amount of mRNA and the differences in both these measures of LDH-B is apparently genetically determined. Because of the differences in kinetic parameters and the enzyme concentration, the predicted *in vivo* reaction velocity is about the same for either population at their respective TM. Thus, there appears to be evolutionary tendency to maintain a constant reaction velocity to adapt to different thermal environments.

CF 311 EVIDENCE FOR RECOMBINATION IN SEQUENCES OF *ROSY* ALLELES FROM WILD-TYPE LINES OF *DROSOPHILA MELANOGASTER*. Daniel Curtis, Steven H. Clark*, Arthur Chovnick* and Welcome Bender. Dept. of Biochemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, and *Mol. Cell. Biol., University of Connecticut, Storrs, CT, 06268.
We have isolated and sequenced a 7.3 kb *Hind* III fragment from 7 unrelated lines of *Drosophila melanogaster*, with 3 more lines in progress. This DNA includes the entire *rosy* gene, which encodes the enzyme xanthine dehydrogenase (XDH), and part of the 5' flanking gene *l(3)S12*. The alleles share polymorphisms in patterns that are best explained as resulting from gene conversion events. We have also studied experimentally derived recombination events at *rosy*. Using a selective system, we recovered crossovers and conversions at *rosy*, and we cloned and sequenced these genes through the regions of recombination. Gene conversion tracts averaged 1.2 kb in length (range: 200 bp to 3.1 kb), and were continuous in nature. Although the wild-type sequences are highly polymorphic, few of the polymorphisms are unique to a single allele. Rather, different alleles share blocks of continuous sequence identity with other alleles in the sample. The length and continuity of these blocks suggest that they represent conversion events. These features of the sequences, as well as a comparison of recombination rates at *rosy* with estimated mutation rates, suggest that recombination plays a major role in generating haplotype diversity.

CF 312 THE CHARACTERIZATION OF "DNA FINGERPRINT" LOCI IN CULTIVATED RICE, John F. Dallas, Department of Agronomy, Curtis Hall, University of Missouri, Columbia, MO 65211.
The areas of plant genetics in which "DNA fingerprinting" techniques may be useful are (a) the identification of cultivars and seed-purity tests in hybrid crops, (b) the estimation of genetic diversity in closely-related cultivars and germplasm accessions, and (c) the characterization of mating patterns in natural populations. The human minisatellite probes 33.6 and 33.15 (Jeffreys et al., Nature 314, 67-73, 1985) cross-hybridize to genomic DNA of the Asian and African cultivated rices, Oryza sativa and O. glaberrima. Both probes detect many restriction fragment length polymorphisms (RFLPs) in cultivars of both species (Dallas, Proc. Natl. Acad. Sci.(USA) 85, 6831-6835, 1988). The analysis of RFLPs detected by 33.6 within and between cultivars of O. sativa suggests that most of this polymorphism occurs between cultivars. The inheritance of fragments detected by 33.6 in F_2 progeny of two O. sativa cultivars suggests that the fragments are homozygous in each parent and occur at several unlinked loci. Analysis of 33.6-detected fragments in the progeny of plants regenerated from tissue culture suggests that these fragments are somatically stable. Calculation of the probability of a chance match between the 33.6 hybridization patterns of two cultivars suggests that these patterns are "DNA fingerprints". Comparison of the hybridization patterns detected by 33.6 and those detected by several single-copy rice RFLP probes suggests that the former probe detects "hypervariable" regions in the rice genome, implying that the 33.6-like rice sequences are involved in the generation of at least some of the polymorphisms.

CF 313 **IS THERE A FINITE UNIVERSE OF EXONS? THE EXTENT OF EXON SHUFFLING IN MODERN PROTEINS.** Robert L. Dorit, Lloyd H. Schoenbach and Walter Gilbert. Dept. of Cellular and Developmental Biology, Harvard University, Cambridge, MA. 02138.
We propose that the extent of exon shuffling in the evolution of modern proteins can be estimated by examining the amino acid similarities that exist between individual exons. We consider pairs of exons revealing sequence similarity beyond what would be expected by chance to be putatively homologous.

In order to gauge the frequency of exon shuffling, we have generated a complete exon database using all sequences available in the Genbank and EMBL data sets. This data set was purged of identical exons; approximately 2000 distinct exons were subsequently compared. We have also carried out a variety of Monte Carlo simulations using sets of randomized exons. These simulations allow us to establish a p-value for any given pairwise exon comparison.

The exon homologies we have obtained were grouped into two distinct categories:
1) similar exons within a single protein, suggesting duplication or intra-protein shuffling;
2) similar exons found in non-homologous proteins, suggesting early exon shuffling.

We argue that exon shuffling in unrelated proteins suggests that exons preexist these proteins. Furthermore, the frequency of "shuffled" exons may reflect the size of the ancestral exon pool.

CF 314 POWER OF MITOCHONDRIAL DNA SEQUENCES FOR RESOLVING BIRD PHYLOGENY, Scott V. Edwards*, Peter Arctander, and Allan C. Wilson, Departments of Biochemistry and *Zoology, University of California, Berkeley, CA 94720.
We used the polymerase chain reaction (PCR) to amplify and sequence a 225-base-pair region of the cytochrome b gene of mitochondrial DNA (mtDNA) in several oscine and suboscine passerine birds. The sequences were readily alignable at both the nucleotide and amino acid levels. Of the 75 amino acids encoded by this region, 10 were variable among the sequences. In each of two comparisons involving oscine and suboscine species, 8 nucleotide substitutions were responsible for the observed amino acid replacements, and in one comparison between two oscine species 5 replacement substitutions had occurred. In these two sets of comparisons we estimate that 4.7 and 2.9% of the replacement sites, respectively, have been substituted. These values suggest that the sequences are not in the multiple-hit zone with respect to replacement sites. Therefore, despite the high rate of evolution at silent sites displayed by animal mtDNA, the use of replacement substitutions of mtDNA coding sequences may have considerable power to resolve the phylogeny of the orders of living birds.

CF 315 THE EVOLUTIONARY HISTORY OF THE MOUSE *t* COMPLEX AS REVEALED BY NATURALLY OCCURRING MOSAIC *t* HAPLOTYPES, Mark A. Erhart, Sandra J. Phillips and Joseph H. Nadeau, The Jackson Laboratory, Bar Harbor, ME 04609. Two of the least understood aspects of the mouse *t* complex are its origin and evolutionary history. The fact that *t* haplotypes are found in *Mus* species other than *Mus domesticus* may be due to either a more ancient origin of the *t* complex or a recent introgression of *t* haplotypes from *M. domesticus* to other *Mus* species. We are using DNA probes for several unrelated loci within the *t* complex to analyze resriction fragment variation in wild-derived mice. We have previously identified a number of wild-derived haplotypes that are mosaic for wild-type and *t*-specific restriction fragment patterns. Some of the same mosaic *t* haplotypes were found in different geographic regions and in both *M. domesticus* and *M. musculus*. This suggested that these mosaic *t* haplotypes were ancestral in origin, pre-dating the divergence of these two species. To test this notion, we have performed a high resolution analysis of two unrelated loci, *Pim-1* and *Crya-1*, located within the distal region of the *t* complex. Several of the mosaic *t* haplotypes were analyzed in order to establish the evolutionary relationships among haplotypes found in different *Mus* species and populations.

CF 316 EVOLUTIONARY ANALYSIS OF ALLELIC DIVERSITY AT THE HLA CLASS II DQ and DR LOCI IN PRIMATES USING DNA AMPLIFICATION, Ulf B. Gyllensten,[1,2] Deval A. Lashkari,[1] Ines Ezcurra[1] and Henry A. Erlich.[1] 1. Department of Human Genetics, Cetus Corporation, Emeryville, CA 94608. 2. Department of Medical Genetics, Biomedical Center, University of Uppsala, Uppsala, Sweden.
The extensive sequence polymorphism observed at human class II loci may either have been generated during the lifetime of the species, or alternatively, could have arisen prior to speciation and been maintained in the contemporary human population. To examine this issue, the polymorphic second exons of the of the DQα, DQβ and DRβ I-IV loci were enzymatically amplified by the polymerase chain reaction (PCR) method, using PCR primers to conserved sequences from a variety of primate species. Comparison of the human alleles with those of Old World species (chimpanzee, pygmy chimpanzee, lowland gorilla, baboon, rhesus monkey and langur) as well as those of New World species (capuchin monkey and marmoset) reveals:

i). The major allelic types at the DQ and DR loci were present at least 5 million years ago in the ancestral species that gave rise the chimpanzee, gorilla and human lineages. Certain allelic types can be identified as far back as 20 million years ago.

ii). Based on the ratio of replacement to silent changes it appears that some alleles have been subjected to selection conserving the amino acid sequence.

iii). The haplotypic combinations of DQα and DQβ alleles in non-human primates are not as restricted as found in humans. However, the association of DQβ and DRβ alleles appears to have been conserved in all hominoids.

CF 317 MULTILOCUS THEORY AND THE ANALYSIS OF INTRAGENIC VARIATION, Ian R. Franklin, CSIRO, Division of Animal Production, PO Box 239, Blacktown, NSW, Australia, 2148.
While historical circumstance and other random effects account for much of patterns in gene sequence variation, large population size, small recombination coefficients and the opportunity for interactions among signal and coding sequences allow for selective influences on haplotype frequencies. Here I discuss some models of intragenic interactions and the role of multi-locus theory in interpreting gene sequence data, particularly from Drosophila.

CF 318 PARENTAGE ANALYSIS IN BIRDS AND MAMMALS USING MURINE MHC PROBES, H. Lisle Gibbs, Peter T. Boag, Yves Plante and Bradley N. White, Department of Biology, Queen's University, Kingston, Ontario K7L 3N6 CANADA.

Probes derived from murine Major Histocompatibility (MHC) genes are shown to be excellent sources of highly polymorphic genetic markers, useful for parentage analysis in a wide range of wild vertebrate species, ranging from mice to several passerine bird species. Results are presented for pedigrees of meadow voles and polygynous family groups of red-winged blackbirds. In both cases virtually all unrelated adults appear to be heterozygous, and no alleles are present at frequencies greater than 5%. Unlike banding patterns typical of "genetic fingerprints" derived from repetative sequence probes, restriction fragment profiles in our MHC systems are relatively easy to quantify. This permits their storage in a database for subsequent retrieval, essential to searches for a biological parent in population studies involving uncertain paternity or maternity.

CF 319 THE DEVELOPMENT OF FELINE-SPECIFIC HYPERVARIABLE PROBES AND THEIR APPLICATION TO ANALYZING THE GENETIC STRUCTURE OF WILD LION POPULATIONS, Dennis A. Gilbert, Craig Packer and Stephen J. O'Brien, BCD Program, PRI, National Cancer Institute, Frederick, MD 21701. Using five different 14-18 b.p. oligonucleotide sequences (derived from consensus human Variable Number Tandem Repeat (VNTR) sequences) as hybridization probes, we have screened genomic libraries prepared from DNA isolated from three unrelated domestic cats. For each oligonucleotide probe, an average of 150 positive clones per genome equivalent were obtained. One hundred and ten of these positive clones were screened for their ability to detect hypervariable DNA sequences when hybridized to genomic DNA of several domestic cats. Fifteen of these clones were chosen based on their detection of highly polymorphic DNA sequences in the domestic cat. On average these clones detect 15-30 bands per individual; and, in an analysis of 40 randomly-collected domestic cats, they recognize 80-120 unique bands. The frequency of these unique bands is on average 0.08 (range 0.02-0.5), and gives a banding pattern that is completely individual-specific. Two of the probes which displayed a high degree of polymorphism in an analysis of captive lions (*Panthera leo*) were chosen to analyze the genetic structure of lion prides in the Serengeti National Park. Using these probes, we were able to estimate quantitatively, the extent of kinship between individual members of various prides. Because these probes were capable of detecting individual-specific banding patterns in all lions, we are now assessing the reproductive success of resident males in several prides in an attempt to precisely define pedigree and social structure of free ranging lion prides.

CF 320 INTRASPECIFIC MITOCHONDRIAL DNA VARIATION IN MIGRATORY AND NON-MIGRATORY POPULATIONS OF SONG SPARROW (Melospiza melodia), Matthew P. Hare and Gerald F. Shields, Institute of Arctic Biology, University of Alaska, Fairbanks, AK 99775. Song Sparrows are extremely polytypic across North America. Four migratory and four resident subspecies are presently recognized in Alaska based on size, beak morphology, and plumage differences. We performed restriction fragment analysis of mitochondrial DNA (mtDNA) comparing variation within and between resident Aleutian Island and migratory South Central Alaskan populations. Mitochondrial DNA of 39 birds is homogeneous overall, although differences in the distribution of mtDNA types and the amount of variation within a locality exist between the resident and migratory populations.

CF 321 ESTIMATE OF GENETIC DISTANCE OF ORANG UTAN SUBSPECIES BASED ON ISOZYME AND TWO-DIMENTIONAL ELECTROPHORESIS, Dianne N. Janczewski, David Goldman and Stephen J. O'Brien, Program Resources Inc., National Cancer Institute, Frederick MD 21701

The orang utan (*Pongo pygmaeus*) is currently divided into two subspecies. *Pongo pygmaeus pygmaeus* resides on Borneo and *P. p. abelii* resides on Sumatra. These populations have been geographically separated by for at least 10,000 years. At present, there is no known route of gene flow between the two populations except through rehabilitated individuals which have been released back into the wild over the last several decades. Differentiation of the two subspecies has been shown using morphological and behavioral characters and they can be distinguished by karyotype. Few studies exist which estimate the actual amount of genetic variation that exists between the populations. In the present study, we compared Nei genetic distances between the great apes and humans using 44 isozyme loci. In addition, we developed a new data set of genetic distances based on the relative mobility of 458 soluble fibroblast-proteins which are resolved by two-dimensional gel electrophoresis. The two distance matrices were used to address the phylogenetic affinity of orang utan subspecies relative to other apes particularly the two chimpanzee species *Pan paniscus* and *Pan troglodytes*. Both data sets were in relative agreement in deriving the following conclusions: 1 The orang utan subspecies distances were approximately 10x closer to each other than they are to the African apes. 2 The orang utan subspecies distance is nearly as great as the distance between species of chimpanzees. Comparison of the genetic distances from this study with other subspecies distance estimates, done in our lab under identical conditions, shows that the distance between Bornean vs. Sumatran orang utan is 5-10 times the distance measured from the other subspecies including lions, cheetahs, tigers and tamarins. Evolutionary topologies were constructed from the distance data using both cladistic and phenetic methods. The resulting trees indicate that man and chimpanzees derived from a common ancestor subsequent to their divergence from the common ancestor with gorilla.

CF 322 ASSUMPTION AND EFFICIENCY OF THE EVOLUTIONARY PARSIMONY METHOD OF PHYLOGENETIC ANALYSIS, Li Jin and Masatoshi Nei, Center for Demographic and Population Genetics, University of Texas at Houston, Houston, TX 77225 Lake (1987) proposed a new method of constructing a phylogentic tree called the evolutionary parsimony method. This method is primarily applied to four DNA sequences and utilizes information on the transition/transversion bias in nucleotide subtitution. The actual procedure is to compute three quantities, X, Y, and Z, and to determine which of them is significantly different from 0. If one of them is significant, the tree corresponding to the quantity is regarded as the correct one. We examined the theoretical basis of this method and obtained the following conclusions. (1) For a full description of Lake's model of nucleotide substitution, six parameters are required. (2) When nucleotide substitution follows Kimura's 2-parameter model, Lake's assertion that Y and Z remain 0 if the "X" tree is correct holds. (3) In other cases, the expectations of Y and Z can be nonzero even if the "X" tree is the correct one. Particularly when the rates of substitution to two transversion states are unequal, Y or Z can be larger than X. (4) Even when nucleotide substitution follows the 2-parameter model, the evolutionary parsimony method is not always superior to other tree-making methods (e.g., maximum parsimony method, neighbor-joining method) in obtaining the correct tree. Part of the reason is that the evolutionary parsimony method does not utilize all information on nucleotide substitution.

CF 323 A STUDY OF OVER 100 MARKERS IN SIX POPULATIONS FAVOURS A PRIMARY SPLIT BETWEEN AFRICA AND THE REST OF THE WORLD, K.K. Kidd,[1] J.R. Kidd,[1] J. Rogers,[1] A.M. Bowcock,[2] J.M. Hebert,[2] A. Lin,[2] J.M. Mountain[2] and L.L. Cavalli-Sforza,[2] [1]Dept. of Human Genetics, Yale University, New Haven, CT 06510 and [2]Dept. of Genetics, Stanford University Medical Center, Stanford, CA 04305. One hundred seven DNA markers from 69 genes or chromosomal regions have been investigated in six populations. Of the 69 loci, 42 were known genes and 27 were anonymous DNA segments. Lymphoblastoid cell lines have been established for all the populations under study: Mbuti pygmies from the Ituri forest in N.E. Zaire, Biaka pygmies from the Central African Republic, Nasioi Melanesians from Bougainville in the Solomon Islands, Chinese born in mainland China (primarily the south), Japanese and Caucasoids of northern or central European origin. The majority of these RFLPs were polymorphic in all populations, suggesting that they predate the diversification of modern Homo sapiens. The variation between populations (measured by F_{st}) is highly significant. When known genes and random segments were compared, there was no significant difference in the F_{st} values, which averaged .15 on a fraction of the markers fully analyzed so far, or in the average heterozygosity (28%). The F_{st} distance between all populations considered in pairs, and averaged over all loci, favours a primary split between Eurasia and Africa, in agreement with the most recent data from non-DNA and mitochondrial sources. Multi-site haplotypes give higher F_{st} values than the separate markers, although similar conclusions can be drawn. Work supported by NSF grant BNS-8619703 and NIH grant GM020467. We are indebted to the many individuals who helped us collect these samples.

CF 324 GENETIC VARIATION WITHIN ASEXUAl LINEAGES OF TARAXACUM OFFICINALE, Lynn M. King and Barbara A. Schaal, Department of Biology, Washington University, St. Louis, MO 63130. Genetic variation was studied within asexual lineages and among clonal genotypes of Taraxacum officinale (Wigg.) using ribosomal DNA (rDNA), chloroplast DNA and alcohol dehydrogenase (Adh). North American T. officinale is triploid and reproduces asexually through seed. A total of 723 asexually produced offspring from 32 parents (26 genotypes) was surveyed. Variation in rDNA was found between parents and offspring in two families. The variant offspring are characterised by loss of an EcoRI restriction site and changes in two restriction fragment lengths in rDNA. In one family, 35 offspring from a single capitulum of the parent had non-parental rDNA. In the other family, 1/26 offspring from one capitulum of the parent had non-parental rDNA and non-parental Adh restriction fragment lengths. Somatic mutation before the development of the capitulum could produce an entire family of variant offspring, while non-sexual recombination during the development of the asexual seed could produce a single variant offspring within the progeny array. In both cases, the F_2 generation from the variant offspring inherited the variant rDNA. These results suggest that somatic recombination and gene amplification generates genotypic diversity among asexual T. officinale.

CF 325 AMPLIFICATION AND DIRECT SEQUENCING OF DIVERGENT mtDNAs VIA THE POLYMERASE CHAIN REACTION, **Thomas D. Kocher, W. Kelley Thomas, Axel Meyer, Scott V. Edwards, Svante Pääbo, Francis X. Villablanca and Allan C. Wilson, Department of Biochemistry, University of California, Berkeley, CA 94720.** **Primers that amplify three regions of human mitochondrial DNA via the polymerase chain reaction amplified the corresponding regions from 99 other species, including mammals, birds, amphibians, fishes, and some invertebrates. Amplification and direct sequencing were possible using unpurified mtDNA from nanogram samples of fresh specimens and microgram amounts of tissues preserved for months in alcohol or decades in the dry state. These sequences provide a consistent metric for phylogenetic comparisons ranging from the population to the interclass level in vertebrates. This molecular method appears to have major advantages over those used before in comparative genetics and taxonomy.**

CF 326 A PHYLOGEOGRAPHIC SURVEY OF MITOCHONDRIAL DNA RELATEDNESS IN THE COYOTE, Niles E. Lehman, Robert K. Wayne, and Charles E. Taylor, Department of Biology, UCLA, Los Angeles, CA 90024 We have analyzed over 200 coyotes (Canis latrans) from all parts of their geographic range in North America for restriction fragment length polymorphisms in mitochondrial DNA. The analysis has been done at three geographic levels in an attempt to ascertain the nature and degree of population genetic structuring. The levels are i) at the broad-scale level, across the entire range of the species; ii) at the state level, across the counties of California; and iii) at the local level, across the Los Angeles area basin, particularily in the Santa Monica Mountains. Our study has revealed some interesting patterns of mtDNA haplotype distribution, such as the apparent high level of variability within California as compared to the rather low levels outside of the state. We are attempting to reconcile our findings with similar mtDNA studies of related canids, as well as with fine-scale DNA fingerprinting analyses of our coyote samples. Finally, we are engaged in a computer simulation study with the goal of modelling the transmission and evolution of mtDNA haplotypes within a defined population.

CF 327 PATTERNS OF DISPERSAL AND THE GEOGRAPHIC STRUCTURE OF GENE GENEALOGIES, Joseph E. Neigel, Department of Biology, The University of Southwestern Louisiana, Lafayette, LA 70504. A phylogenetic approach was used to examine geographic distributions of gene genealogies that arise from organismal dispersal. The models began with assumptions similar to those used in classical isolation-by-distance models, but were constructed in terms of gene phylogenies rather than gene frequencies. Both analytical probability models and computer simulations were developed. Results from the models indicate that a relationship between genetic distance and geographic distance could be used to estimate average per generation dispersal distances. In addition, the models are used to examine the effects of limited barriers to dispersal on the formation of geographic breaks in genetic variation and on concordance among the genealogies of unlinked loci.

CF 328 ANCIENT DNA AS A MEANS TO STUDY MOLECULAR EVOLUTION, Svante Pääbo, W. Kelley Thomas, Nancy L. Shearin*, Dennis H. O'Rourke* and Allan C. Wilson, Department of Biochemistry, University of California, Berkeley, CA 94720; *Department of Anthropology, University of Utah, Salt Lake City, UT 84112. DNA can be extracted from soft tissue remains that vary in age from four to 13,000 years. The DNA obtained is invariably of a low average molecular size and damaged by oxidative processes, that render conventional cloning difficult and error prone. However, the polymerase chain reaction can be used to study short sequences from extracts of old DNA. This opens up the possibility of using museum collections and archaeological finds to study molecular evolution over time. Results from extinct species as well as Egyptian and American human mummies will be presented.

CF 329 STATISTICAL TESTS OF PHYLOGENIES BASED ON GENETIC DISTANCES, Pekka Pamilo, Department of Genetics, University of Helsinki, Arkadiankatu 7, SF-00100 Helsinki, Finland
Methods used for estimating the confidence of UPGMA phenograms based on molecular genetic data are examined with special emphasis in defining the specific hypotheses tested. The methods based on internodal variances or on bootstrapping over characters are compared by simulating evolution of a DNA sequence by random nucleotide substitutions in a model with three species trifurcating at the same point. The bootstrap method seems slightly better in this comparison. Weighting of OTUs when constructing the phenogram is also examined. A method which weights each OTU according to the estimated independent evolutionary information, a modified WPGMA method, appears slightly better than UPGMA in estimating the branching points and branch lengths. The methods are applied to the data on restriction sites in mtDNA of eight Hawaiian Drosophila species (DeSalle, 1984). The significancies of the clusters among this group of species differ from those reported earlier.

CF 330 LARGE ESTIMATED DIFFERENCES BETWEEN MTDNA SEQUENCES WITHIN A MAMMALIAN SPECIES MAY BE THE RESULT OF GENE REARRANGEMENT. Dorothy E. Pumo, Carleton J. Phillips, Hugh H. Genoways, Michael Amella, and Colleen O'Connor, Department of Biology, Hofstra University, Hempstead, NY 11550, and University of Nebraska State Museum, University of Nebraska-Lincoln, Lincoln, NE 68588-0338.
MtDNA analysis is useful for answering phylogeographic questions. Many investigators assume that mitochondrial gene order is conserved among mammals. Several complete mtDNA sequences support this. We previously reported large (>8%) intraspecific sequence divergence in the fruit bat, Artibeus jamaicensis, based on mapped data for 9 restriction enzymes. Typically, values for such intraspecific comparisons are less than 5%. Although we tested several explanations for the large divergence value, only one hypothesis appears consistent with our data. The data suggest that a rearrangement occurred. Two Sst II sites (separated by ~ 1600 bp) occur in the rRNA genes in vertebrate mtDNA. When restriction maps of mtDNA from A. jamaicensis SV and J groups are aligned to maximize site similarity, the paired SstII sites are separated by ~ 3000 bp suggesting that the rRNA genes are in different relative positions. Examination of neighboring restriction sites suggests that the rRNA genes were inverted during the translocation. Comparison of a "reconstructed" SV genome with a J genome yields an estimated sequence divergence <4%. (NSF grant BBS-8609231.)

CF 331 THE GENOMIC STRUCTURE AND POPULATION GENETICS OF THE EST 6 LOCUS IN *DROSOPHILA*, Rollin C. Richmond, James Brady, Chris Collet, Peter Cooke, Karen Nielsen, John G. Oakeshott, and Robyn J. Russell, Department of Biology, Indiana University, Bloomington, IN 47405; CSIRO Division of Entomology, Canberra, ACT 2601, Australia. The Esterase 6 (*Est-6*) locus of *Drosophila* codes for an esterase which exhibits high levels of variability in its amino acid sequence. Moreover the pattern of *Est-6* expression has changed dramatically even in closely related species. Genomic clones containing sequences homologous to an *Est-6* cDNA clone were isolated from a *D. melanogaster* genomic library. Comparison of the genomic and cDNA sequences revealed that the *Est-6* gene comprises two exons separated by a short intron. Further sequencing revealed the presence of a tandem duplication (denoted *Est-P*) 197 bp 3'of the *Est-6* gene. The two genes have a similar genomic structure and show similarities of 64% and 60% at the DNA and protein levels, respectively. Presumptive 5' regulatory sequences of *Est-P* overlap at least the 3' non-coding region of *Est-6*. Differences in the timing of the expression of these two loci suggest different physiological functions for the products of the two genes. Population genetic analysis of 13 fully sequenced *Est-6* alleles has identified 52 nucleotide polymorphisms of which 16 result in amino acid sequence polymorphisms. Two of these polymorphisms are likely targets for the action of selection. Cloning and sequencing of the homologous region from *D. pseudoobscura* reveals that it consists of three related esterase genes. Transformation of *D. melanogaster* with each of the three *pseudoobscura* esterase loci shows that only one of the three corresponds to a previously known esterase locus.

CF 332 NUCLEOTIDE POLYMORPHISM AT THE XDH LOCUS IN DROSOPHILA PSEUDOOBSCURA, Margaret A. Riley, Mary Ellen Powers and Richard Lewontin, Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138. 58 isochromosomal lines sampled from two natural populations of D. pseudoobscura were examined using four-cutter restriction mapping. A 4.6 Kb region of the xanthine dehydrogenase locus was probed and 65 of 135 sites scored were polymorphic. This predicts that on average every eighth base pair would be polymorphic in this region for the genes surveyed if polymorphisms occurred randomly along the coding region. 49 distinct haplotypes were recognized in the 58 lines examined. The most common haplotype obtained a frequency of only 5%. Very high levels of recombination can be inferred from the presence of all four gametic types in the data set. 296 pairwise comparisons reveal the presence of recombinational events. The level and pattern of nucleotide polymorphism at the Xdh locus suggests that this gene may be evolving in a predominately neutral fashion.

CF 333 INFORMATION PROPERTIES OF MAXIMUM LIKELIHOOD DNA PHYLOGENIES.
Kermit Ritland[1] and Michael Clegg[2], [1]Department of Botany, University of Toronto, Toronto M5S1A1 Canada, [2]Department of Botany and Plant Sciences, University of California, Riverside CA 92521.

The information properties of the maximum likelihood phylogeny of a small number of DNA sequences, wherein the number of substitutions along each branch are estimated, was studied via inversion of Fisher information matrices. The number of substitutions represent branch length and "evolutionary distance". It was assumed that base substitution followed Hasagawa's (1985) model, but with no substitution bias and with equality of base frequencies. These simple cases illustrate properties expected of larger phylogenies (but the information matrix does not include the error due to incorrect assignment of sequences to tips).

The following properties were among those found. (1) When two sequences are compared, inclusion of a third sequence of unknown distance provides almost no additional information about evolutionary distance between the first two sequences (if the third sequence is of known distance, information is provided about this distance). (2) In a three-sequence phylogeny, an "optimal" sequence divergence (which minimizes the error of relative branch lengths, or phylogeny "shape") occurs at a quite large 30% sequence divergence from tip to node. This optimum occurs because as sequences diverge, the increased information about distance provided by an increasing numbers of "single-hit" changes is offset by decreased information caused by more "multiple hits". (3) In a three-sequence phylogeny, increasing the distance of the third sequence (an "outgroup") always decreases the information about the ratio of first two branch lengths. (4) In three- and higher-sequence phylogenies, the bases (A, T, C or G) at each site along the ancestral node sequence(s) can be inferred *in probability*. This in turn allows inference of numbers of substitutional changes at each site for (a) each branch of the phylogeny and (b) the entire phylogeny, but not without some problems, which are discussed. An example of such inference is given for the chloroplast rbcL gene.

CF 334 MOLECULAR GENETICS OF PRE-COLUMBIAN SOUTH AMERICAN MUMMIES,
Peter K. Rogan*, National Cancer Institute-Frederick Cancer Research Facility, Frederick, MD 21701, and Joseph J. Salvo, GE Corporate Research & Development, Schenectady, NY 12301.

We have isolated human deoxyribonucleic acid (DNA) from the mummified remains of pre-Columbian South American Indians. Ten different individuals from two coastal Chilean cultures separated geographically and temporally (Maitas: 700 years b.p., Camarones: 500 years b.p.) were selected for our preliminary screen. High molecular weight DNA (>12000 base pairs) was recovered from nine of the specimens. Southern blot hybridization indicated that each of these samples contained human DNA of both genomic and mitochondrial origin. Further analysis of ancient nucleic acid was hampered by low DNA yields and lesions that appear to inhibit its efficient replication *in vitro*. These difficulties were overcome in part by first repairing the DNA, and then amplifying specific target sequences with a polymerase chain reaction protocol. Our approach produced a high level of amplified genomic and mitochondrial sequences in most of the samples as confirmed by Southern blot analysis.

*This research was sponsored in part by the National Cancer Institute, DHHS, under contract No. N01-CO-74101 with Bionetics Research, Inc.

CF 335 RESTRICTION FRAGMENT LENGTH POLYMORPHISMS IN A WILD POPULATION OF YELLOW BABOONS FROM TANZANIA, Jeffrey Rogers and Kenneth K. Kidd, Dept. of Human Genetics, Yale University School of Medicine, New Haven, CT 06510. Traditional electrophoretic studies of protein variation in animal populations found that large mammals exhibit substantially less variation per locus than other vertebrates and invertebrates (Nevo 1978, Theor Pop Biol 13:121). Studies of DNA polymorphism in human populations using RFLPs have revealed significantly higher levels of variation at the DNA level (Jeffreys 1979, Cell 18:1, Bowcock et al. 1988, Gene Geog. XX:xx), but thorough interpretation of the observed levels of variation requires quantification of the amount of DNA polymorphism across a number of species. We have captured, bled and released 135 wild yellow baboons (Papio hamadryas cynocephalus) from six social groups within a single Mendelian population, extracted genomic DNA and screened for RFLPs using human cDNA clones as probes. A preliminary screen for polymorphisms used a panel of 9 to 18 unrelated baboons digested with from 6 to 8 restriction enzymes. The first five probes tested (loci: APOB, REN, HEXB, ARG1, AT3) detected a total of 15 distinct RFLP systems (probe-enzyme combinations), including 2 with more than two alleles. Using Ewens et al. 1981 (PNAS 78:3748) equations, preliminary minimum estimates of the proportion of polymorphic nucleotides at these loci vary from 1.5% to 2.5%. We are currently typing the entire sample of 135 individuals for these RFLP systems and will report estimates of within-population variation at each locus as well as overall levels of heterozygosity and between-social group differentiation.

CF 336 RATE OF SYNONYMOUS SUBSTITUTION DIFFERS BETWEEN DUPLICATE GENES OF THE ALCOHOL DEHYDROGENASE LOCUS OF Drosophila, Stephen W. Schaeffer, Department of Biology, The Pennsylvania State University, University Park, PA 16802. Intraspecific and interspecific comparisons of the nucleotide sequences of the duplicate genes in the alcohol dehydrogenase region within the obscura group of Drosophila were used to test predictions of the neutral theory of molecular evolution. The nucloetide sequence comparisons showed that: (1) the frequency of synonymous substitutions is greater than the frequency of amino acid replacements; (2) the frequency of substitutions differed among synonymous, intron, and noncoding sites; (3) the frequency of substitutions in synonymous sites differed between the duplicate genes; and (4) regions that diverged rapidly between species are highly polymorphic within species. These results are consistent with this region evolving according to the predictions of the neutral theory of molecular evolution. The heterogeneity in the rate of synonymous substitution between the duplicate genes must indicate differences in mutation rate or selective constraint.

CF 337 MOLECULAR GENEOLOGY MAY BE AN ILLUSION CREATED BY THE CONSTRUCTION PRINCIPLE OF THE PREMORIDAL GENOME, Christian Schwabe, Department of Biochemistry and Molecular Biology, Med. Univ. of S.C., Charleston, SC 29425
Manifest endorsement of chance as the prime force in abiogenesis makes a single point origin a conditio sine qua non for the Neo-Darwinian paradigm of molecular evolution. I have explored a system wherein chance has been replaced by chemical boundary determinism and the single origin by the proposition that life started from many surviving origins. While the deterministic properties of atoms are responsible for the well-known results of the Miller-Urey experiment the unique properties of nucleic acids (DNA or RNA) to form quasi crystals of complimentary (as opposed to identical) particles fall into the same category only at a higher level of organization. The discovery by Ohno of heptameric and lower order repeats in modern genomes suggests that the primordial gene was assembled by oligomer rather than monomer polymerization. Using "Ohno pieces" and a computer I have reproduced repetitive polymers with long open reading frames and highly variable amino acid sequences. Slight changes in initial conditions led to greater homology and to domain similarity in the various artificial proteins whereas other starting peptides produced stop codons with such frequency that coding regions constructed from these oligomers would have eliminated themselves, or remained as pseudogenes. In conjunction with a stochastic cell formation mode the system mimics evolutionary trees, accounts for the discrepancies between paleontological and molecular data and eliminates the need to postulate that increased complexity has come about by a random process such as mutation.

CF 338 MOLECULAR GENETIC VARIATION IN DROSOPHILA ANANASSAE, Wolfgang Stephan and Charles H. Langley, Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709
We have surveyed 60 X-chromosomes from three natural populations for restriction map variation at the Bar, forked, and vermilion loci. The level of variation in the vermilion region is reduced (relative to forked and Bar). This is in agreement with similar observations from D. melanogaster, and can be attributed to "hitch-hiking". Furthermore, we found strong evidence for population subdivision.

CF 339 THE HUMAN GENE MAPPING LIBRARY: A RESOURCE FOR STUDIES OF MOLECULAR EVOLUTION, POPULATION GENETICS, AND COMPARATIVE MAPPING

J. Claiborne Stephens, Rowena K. Track, Iva H. Cohen and Kenneth K. Kidd, Howard Hughes Medical Institute Human Gene Mapping Library, 25 Science Park, New Haven CT 06511.

The Human Gene Mapping Library maintains five interconnected databases, viz, MAP -- a catalog of almost 4000 known human genes and anonymous DNA segments that have been mapped to chromosomes or chromosomal regions; RFLP -- descriptions of over 1400 loci that exhibit DNA variation detected by restriction endonuclease analysis; PROBE -- details on the size, vector, availability, and other characterizations of probes that have been useful in linkage or mapping studies; CONTACT -- an on-line address book for researchers and others interested in human gene mapping and variation; and LIT -- over 10,000 literature references for entries in the other databases. Information of particular interest to researchers in molecular evolution or population genetics includes 1) RFLP allele frequencies for diverse human populations, when known; 2) full characterization of each RFLP system (e.g., probe-enzyme combination, special conditions used); 3) cross-references to GenBank®; and 4) indications on cloning and availability for each probe. All entries in MAP, RFLP, PROBE, and LIT are mutually cross-referenced to help the researcher obtain relevant information from more than one database. Interactive searches are through Telenet. The Human Gene Mapping Library is funded by the Howard Hughes Medical Institute. Access is currently free of charge.

CF 340 POPULATION DISTRIBUTION OF A 9-BP DELETION IN HUMAN MTDNA: IMPLICATIONS FOR THE COLONIZATION OF THE PACIFIC, M. Stoneking, A. Salam M. Sofro*, and A.C. Wilson, Dept. of Biochemistry, Univ. of California, Berkeley, CA 94720 and *Dept. of Biochemistry, Faculty of Medicine, Gadjah Mada University, Yogyakarta, Indonesia. The loss of one of two copies of a 9-bp repeated sequence in a small noncoding region of human mitochondrial DNA (mtDNA) is a marker for populations of Asian origin. We screened 146 individuals from six islands in eastern Indonesia for the presence of this 9-bp deletion by enzymatic amplification of the mtDNA segment. Amplification products were sized and visualized by ethidium bromide staining of agarose gels; direct sequencing of several individuals confirmed the loss of the 9-bp repeat. The frequency of the deletion ranged from 8% to 32% in Indonesian populations. We and others have shown that the deletion occurs also in coastal Papua New Guinea, Melanesia, and is nearly fixed in Polynesia, but it is absent from Australia and the highlands of New Guinea. This distribution is consistent with previous suggestions of at least two migrations from Asia through the South Pacific: an early migration of individuals without the deletion that populated Australia and New Guinea, followed by a later migration of individuals with the deletion through Indonesia and coastal and island Melanesia. Polynesians are derived almost exclusively from this latter migration.

CF 341 CLASSICAL PLANT TAXONOMIC AMBIGUITIES EXTEND TO THE MOLECULAR LEVEL, Michael Syvanen and Hyman Hartmen, Department of Medical Microbiology and Immunology, School of Medicine, University of California at Davis, Davis, CA 95616. Botanists have long had difficulty in classifying flowering plants into unambiguous higher taxonomic groups (1,2,3). The problem in plant taxonomy is the widespread existence of convergences, or more generally homoplasies, of gross morphological characters (3,4,5). We have addressed whether plants are qualitatively different from animals in this regard by determining the molecular phylogenies of the strictly comparable cytochromes from plants and vertebrates. On the basis of a cladistic analysis from 26 plant species, compared to that from 27 vertebrate species, we find that while the vertebrate sequences yield reasonably well defined minimal trees that are congruent with the biological tree, the plant sequences yield multiple minimal trees that are not only highly incongruent with each other, but none of which is congruent with any reasonable biological tree. That is, the plant sequence set is much more homoplastic than that of the animal. However as judged by the relative rate test, the extent of divergence and degree of functional constraint, cytochrome c evolution in plants does not appear differently from that of vertebrates. The fact that homoplasy in plants is seen with both morphological characters and neutral molecular characters strongly suggests that the homoplasy is not the result of natural selection but rather is the result of basic genetic mechanisms. The most likely explanation is horizontal gene spread among the flowering plants.

CF 342 A DIACHRONIC STUDY OF MITOCHONDRIAL DNA LINEAGE MAINTENANCE IN NATURAL POPULATIONS, W. Kelley Thomas, Svante Paabo, Francis X. Villablanca and Allan C. Wilson, Department of Biochemistry, University of California, Berkeley, CA 94720

The advent of the polymerase chain reaction has opened up the possibility of sequencing DNA from old tissue specimens. Combined with the presence of museum specimens collected in population samples over the last 100 years, this new technique has opened up the possibility of following genetic changes in populations over long periods of time. In this study, specimens of the panamint kangaroo rat Dipodomys panamintinis from localities representing each of three subspecies were examined. A total of 63 museum specimens originally collected and prepared as dried skins in 1911, 1917 or 1937 were included. For each specimen a 250 base pair segment of the mitochondrial D-loop was amplified by the PCR and directly sequenced. The same populations were then re-sampled today and compared to the old ones. The possibility of directly studying the history of molecular variation allows us to gain insights into the relative effects of factors such as population size and migration on the maintenance of variation in natural populations.

CF 343 A NOVEL QUADRUPLE ζ-GLOBIN GENE ARRANGEMENT IN HUMANS: MECHANISTIC & EVOLUTIONARY IMPLICATIONS, Elizabeth A. Titus and John A. Hunt,Department of Genetics, John Burns School of Medicine, University of Hawaii, Honolulu, HI 96822.

In screening Laotian families for α-thalassemia by Southern blot technique, we found several anomalous ζ-globin haplotypes. The ζ-globin genes encode α-like embryonic hemoglobin subunits in the α-globin multigene complex on chromosome 16. There are normally two ζ-globin genes in this cluster: 5' ζ_2 and 3' $\Psi\zeta_1$. In our study, a novel quadruple ζ-globin arrangement was found segregating in one family; the triple ζ-globin gene arrangement was found in three unrelated Laotian families. Eco R1, Bam H1, Bgl II and Hind III digests hybridized to a ζ-globin gene probe revealed anomalous bands in the ~10 kb range consistent with reduplication of the ζ-globin region. This haplotype interpretation was confirmed by Southern blot analyses using double digestions hybridized to a cDNA ζ-gene probe, and Pvu II digests probed with a 5'-$\Psi\zeta_1$ intergenic fragment. In addition to the numerous length variants in the quadruple ζ-globin haplotype, a polymorphic BamH1 site was identified in the intergenic region between two of the zeta genes. Restriction fragments generated from these anomalous ζ-globin chromosomes are compatible with a homologous but unequal crossover mechanism. The repetitious nature of the ζ-globin introns could facilitate misalignments responsible for length mutations and multiple and diminished gene arrangements in the α-globin cluster. Frequencies of triple ζ-globin genes as high as 17% in Southeast Asians has been reported (Fucharoen and Winichagoon. 1987.*Hemoglobin* 11(1): 65), indicating a homologous recombination event involving a triple ζ-globin and "normal" α-globin chromosome is highly probable. The high frequency of rearranged ζ-globin genes in the South-East Asians may reflect a selective advantage, as in the proposed correlation with malaria resistance (Flint et al., 1987. *Nature* 321: 744) ; alternatively, it may reflect a high rate of DNA rearrangements in this region of the α-globin complex.

CF 344 HUMAN MITOCHONDRIAL DNA SEQUENCE POLYMORPHISM IN AFRICA: THE ORIGIN OF PYGMIES Linda Vigilant and Allan C. Wilson, Biochemistry Department, University of California, Berkeley, CA 94720

Enzymatic amplification and direct sequencing of D loop region mtDNA reveals the history of human populations in Africa. Populations surveyed are primarily hunter-gatherers and include the Pygmies of Zaire and of the Central African Republic, the Zhun/twa and Herero-Banderu of Southern Africa, and the Tanzanian Hadza. Samples of single plucked hairs contained sufficient DNA for analysis. A preliminary tree constructed by parsimony analysis of more than fifty phylogenetically informative sites reveals the presence of two main mtDNA lineages in Pygmies, and within each there is a distinct clustering of Zaire Pygmy mtDNA types apart from those of the CAR. The 27 pygmy individuals are represented by only 11 mtDNA types, and the nearest neighbor of each of the lineages is a Black American. The parsimony tree will also allow estimation of the age of the Pygmy lineages and of the number and ages of the lineages leading to other African populations.

CF 345 VARIABLE RATES OF SEQUENCE EVOLUTION AMONG HIGHLY DIVERGENT mtDNA HAPLOTYPES OF THE EAST AFRICAN BLACK-BACKED JACKAL, CANIS MESOMELAS ELONGAE, Robert K. Wayne, Pieter W. Kat, Todd K. Fuller, Derek Girman, Blaire Van Valkenburgh, Niles Lehman, and Stephen J. O'Brien, Department of Biology, U.C. Los Angeles, Los Angeles, CA 90024. Debates about the relative constancy of molecular evolution usually focus on DNA or protein sequence differences among distinct taxa with divergence times ranging from 5 to 500 million years. Intraspecific variability in rate is rarely considered. We test the assumption of intraspecific rate constancy by comparing mtDNA sequence divergence, as inferred from restriction fragment length polymorphisms, between mtDNA haplotypes of the African black-backed jackal, Canis mesomelas, and those of two other jackal species, C. aureus, the golden jackal and C. adustus, the side-striped jackal. Our results are unusual for several reasons. First, a large sequence divergence of 6.6% is found between black-backed jackal mtDNA haplotypes. Previous, intraspecific studies of terrestrial mammals have not reported values greater than 5%. Second, these highly divergent haplotypes are from the same locality whereas in past studies large divergence values generally occur among geographically distant mtDNA haplotypes. Third, the extent of sequence divergence between mtDNA haplotypes of two other jackal species and black-backed jackal mtDNA haplotypes differs by as much as 2-3 fold, thus indicating a substantial heterogeneity in the rate of sequence evolution among black-backed jackal mtDNA haplotypes. The results are difficult to reconcile with the predictions of the neutral model of evolution.

CF 346 STANDARDBRED STALLION GENE TRANSMISSION: EVIDENCE FOR SELECTION IN TROTTING HORSES, Lowell R. Weitkamp, Jean W. MacCluer, Sally A. Guttormsen and Rose H. King, Division of Genetics, University of Rochester Medical Center, Rochester, NY 14642 and Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX 78284
For genes selectively neutral in the population as a whole, it is nevertheless plausible that one allele may mark a haplotype with a higher selective value than the alternative haplotype in a specific individual. That is, there may be differential fitness for marker genes in the context of the genetic combinations available to the gametes of that individual. To test the hypothesis that individual differences in transmission ratios may occur for genes that segregate according to Mendelian expectations, the transmission ratios of alleles at 12 protein marker loci were computed individually for Standardbred stallions in a genealogy of 5392 phenotyped horses. Among 42 trotting sires (diagonally opposite legs move forward together) there were 4598 offspring genotypes informative for transmission of one or the other sire's genes; among 69 pacing sires (legs on same side of body move forward together) there were 6676 informative offspring-genotypes. Over all loci there was significant gene transmission distortion for trotting sires ($p=0.0019$) but not for pacing sires ($p=0.99$). The transmission distortion in trotters was due to sire-specific effects ($p=0.0024$) and not to increased transmission of one or the other allele of a given heterozygous genotype. To our knowledge, this is the first report in a mammalian population (trotting horses) of sire-specific gene transmission distortion *taken over all typed loci.* Individual-specific, non-random transmission of homologous chromosomes may provide a mechanism for selection to operate without requiring differential fitness of specific alleles or genotypes.

Organelle and Transposon Evolution; Evolution of the Immune System

CF 400 A MULTI-COPY REGULATORY ELEMENT IN WHEAT MITOCHONDRIAL DNA
Linda Bonen and Sharon Bird, Department of Biology, University of Ottawa, Ottawa, Canada, K1N 6N5
Plant mitochondrial genomes evolve rapidly through DNA rearrangements, yet individual genes are very conservative in nucleotide sequence. In a search for signals involved in gene expression, we have examined sequences flanking a number of wheat mitochondrial protein-coding genes and compared them to those of other plants. In many instances, non-coding sequences are completely unrelated. Moreover, the breakpoint in homology sometimes occurs within the coding region so that predicted protein sequences differ between closely-related plants. The wheat mitochondrial genes for COXII, ATP6 and the plant-specific ORF25 are preceded by a highly-conserved block of approximately 200 bp that is present at several other genomic locations but does not directly precede the genes for COXI, CYB, NAD5 or ATP9. This conserved block contains the promoter motif, TATAGTA, as determined by primer extension and guanylyltransferase capping studies. Southern hybridization analysis indicates that sequences related to the 200 bp block are multiply-represented in several monocot, but not dicot, mitochondrial DNAs. However, none of the genes cited above is directly preceded by this conserved block in maize. Thus, mitochondrial DNA rearrangements alter the expression signals of very closely-related plant genes.
Supported by the Natural Sciences and Engineering Research Council of Canada.

CF 401 MITOCHONDRIAL DNA CONTROL REGION SEQUENCES: IMPLICATIONS FOR HOMINOID PHYLOGENY
Jeffrey L. Boore and Wesley M. Brown, Department of Biology, University of Michigan, Ann Arbor, MI 48109-1048, U.S.A. Molecular studies of hominoid phylogeny have established the African genera *Homo* (man), *Pan* (chimpanzees), and *Gorilla* as the most closely related, with the genera *Pongo* (orangutan) and *Hylobates* (gibbons) being successively less related. Attempts to resolve relationships within the African genera have been less successful, due in part to an insufficient number of informative characters (restriction sites, nucleotides, and amino acids) on which to base a statistically robust conclusion. Even comparisons of mitochondrial DNA (mtDNA), which evolves approximately ten times faster than single-copy nuclear DNA, have failed to adequately resolve this branching order. However, more rapidly evolving DNA sequences may yet provide the degree of resolution that is necessary to resolve this problem.

The Control Region (CR), the only large non-coding region in animal mtDNA, evolves several times faster than the coding regions. We recently determined CR sequences for *Pan* (*P. troglodytes* and *P. paniscus*) and *Gorilla* and compared them with the *Homo* CR sequence, but only limited phylogenetic inference was possible due to the lack of an outgroup. For the present study, we determined the *Pongo* CR sequence to provide an outgroup. Comparison of the aligned CR sequences supports a *Homo-Pan* clade in three ways: (1) their sequences are most similar; (2) their transition:transversion ratios are highest; and (3) portions of the CR sequence present in both *Pan* and *Homo* are absent in both *Gorilla* and *Pongo*, suggesting that these portions were acquired by the immediate, common ancestor to *Pan* and *Homo*. While these sequences could have been deleted independently in *Gorilla* and *Pongo*, this appears to be less likely.

CF 402 MITOCHONDRIAL DNA IN THE PINE WEEVILS: CHARACTERIZATION OF AN UNUSUALLY LARGE GENOME, Thomas M. Boyce, Michael E. Zwick, and Charles F. Aquadro, Genetics and Development, Cornell University, Ithaca, NY 14853. Mitochondrial DNA of higher animals has been described as an example of extreme efficiency in gene arrangement and use. Where exceptionally large size molecules have been found (>20 Kb), all have occurred as rare variants within a species, suggesting that these variants arise infrequently and do not persist for long periods in evolutionary time. In contrast, all individuals of at least three species of pine weevil (Curculionidae: *Pissodes*) possess a mitochondrial genome of unusually large size (30 to 36 Kb). The molecule owes its large size to a dramatically enlarged A+T-rich region (15 to 20 Kb). Gene content and order outside of this region appear to be similar to that found in *Drosophila* and other insects. A series of 0.8 to 1.9 Kb repeated sequences occur adjacent to the large A+T region and have perhaps played a role in the generation of the large size as well as a high frequency of size variant heteroplasmy (ca. 99% of weevils sampled in all three species). As many as five distinct size classes may occur in an individual weevil. The persistence of such a large size through two speciation events and the abundance of size variants within individuals indicates that these molecules may not be subject to strong selection for efficiency of replication, i.e. the so called "race to replicate."

CF 403 HUMAN MITOCHONDRIAL DNA MUTATIONS IN POLYNESIAN POPULATIONS DETECTED RAPIDLY BY PCR AMPLIFICATION, Rebecca Cann and Olga Rickards, Department of Genetics, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI 96822 We have instituted a screen of genetic variability in isolated human populations in the Pacific at selected mitochondrial DNA loci, concentrating on partial cytochrome B, 12s rRNA, and D-Loop sequences. In addition, we have investigated a highly Asian-specific length polymorphism associated with a repetitive sequence near the end of the CO11 gene (8250-8300). Sources of DNA for this study were total genomic DNAs and hair samples, and amplification utilized primers designed by Kocher, Thomas, Meyer, Edwards, Paabo, Villablanca, and Wilson of UC Berkeley as "universal" for their high degree of sequence conservation throughout known metazoan mtDNAs. Direct sequencing of amplified products gives estimates of nucleotide diversity consistent with fine-scale restriction endonuclease cleavage maps for humans, but also allows internal calibration of the time scale for human mitochondrial DNA sequence evolution based on known colonization times for parts of the Pacific. Based on mtDNA sequences, at least 2 separate maternal lineages reached precontact Hawaii. Cook Islanders, Tahitians, and Native Hawaiians in addition reveal the process of genetic drift operating in mtDNA evolution, as maternal lineages are lost to periodic extinction events. We are attempting now to recover mtDNA from ancient Hawaiian materials in order to estimate the degree of lineage replacement in the last 800 years.

CF 404 EVOLUTION AND DISTRIBUTION OF RETROTRANSPOSABLE ELEMENTS SPECIFIC TO THE 28S RIBOSOMAL GENES. Thomas H. Eickbush, John L. Jakubczak, Yue Xiong and William D. Burke. Department of Biology, University of Rochester, Rochester, NY. 14627 USA

A fraction of the 28S ribosomal genes in many insects are interrupted by the sequence specific retrotransposable elements R1 and R2. The remarkable specificity of these elements is at least partly a result of their encoding sequence specific integrase functions. Because R1 and R2 insert at unique target sites that are only 74 bp apart in the 28S gene, it is possible to assay their presence in widely different organisms even though the level of sequence homology does not permit cross-hybridization using nucleic acid probes. We have detected R1 and R2 elements in a large number of Dipteran and Lepidoteran species, and in at least one species from a number of other insect orders. It would appear that these elements are well adapted for their existence in the rDNA genes of species throughout the class Insecta. Sequence comparison of these elements in distant species reveals the same genomic organization as well as several conserved coding regions which may represent protein domains essential to the propagation of the elements. Our data suggests that R1 and R2 have existed as distinct, non-recombining elements for an extensive evolutionary period. R1 and R2 represent the only instances where the same transposable elements have been traced across a broad range of species. Finally, we have also conducted a comparison of the amino acid sequences of R1 and R2 with other reverse transcriptase containing sequences. Our results reveal that all reverse transcriptase containing elements can be divided into two major groups. The first group is composed of retroviruses, certain DNA viruses, and transposable elements containing long terminal repeats (LTRs). R1 and R2 are grouped with a second class of transposable elements, including I, G, F and Jockey of Drosophila and mammalian L1 sequences. All of these elements lack LTRs. Surprisingly, these non-LTR elements contain a higher degree of similarity to fungal mitochondrial class II introns than to LTR containing retrotransposons.

CF 405 ROCK WALLABIES INTRODUCED TO HAWAII IN 1916 REVEAL LIMITED MITOCHONDRIAL DNA SEQUENCE VARIABILITY, Leslyn A. Hanakahi and G. Kaui Wong, Department of Genetics, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI 96822

Genetic variability is often lost when animal populations undergo bottlenecks associated with range expansion or contraction. The unintentional introduction of a single pair of rock wallabies (Petrogale penicillata) to the Hawaiian island of Oahu has allowed a natural experiment to progress- to the point that this stable population has been estimated by conventional census techniques to now number about 250 animals. MtDNA sequences were amplified from total genomic DNAs prepared from tail blood samples of 8 animals via the polymerase chain reaction. All 12S rRNA and Cytochrome B sequences examined so far are identical, and work is in progress on amplified D-loop sequences. Molecular evidence is consistent so far with our knowledge of a single founder female who was mitochondrially homogeneous in DNA sequence. Although at least 20 generations have followed from the first introduction, human urban expension has limited the number of available wallaby habitats and opportunities for successful dispersal to adjacent areas from the original Ewa-Kalihi cliffside appears low. We are attempting to identify the Australian population which served as the source for this successful introduction, even though Australian animals are now largely endangered in their original ranges. Supported by NIH-MARC Grant No. GM07684.

CF 406 A QUANTITATIVE SELFISH RETROTRANSPOSON MODEL FOR MAMMALIAN LINES ONE, Stephen C. Hardies and Brad A. Rikke, Department of Biochemistry, Univ. of Texas Health Science Center at San Antonio, TX 78284

LINES ONE is a family of mammalian retrotransposons including both defective and potentially active elements. We have found an excess of point mutations in certain LINES ONE lineages which we propose are due to reverse transcription errors. Using the fidelity of typical RNA-directed DNA polymerases in combination with neutral theory, we calculate the mean time between retrotranspositions in these lineages to be 50,000 years. The following support this interpretation: 1) All observed duplicative events (indicated by the sampling of both progeny) fall on these accelerated lineages. 2) The accelerated lineages show selective pressure on the LINES ONE reading frames, which are thought to encode a reverse transciptase. And most importantly, 3) the mean time for loss of activity of an individual element, calculated from the size of its reading frames and the neutral rate for point mutations, is 80,000 years. The close correspondence between the calculated frequencies of retrotransposition and of inactivation support a SELFISH RETROTRANSPOSON model wherein active elements duplicate just often enough to offset the loss due to point mutations. Further development of this model provides 1) observational evidence that lineages were lost in selfish competition with the currently active lineages, 2) an estimate for the extant number of fully active elements, 3) a prediction that some defective elements still express defective protein, and 4) a proposal for a clearance mechanism and a P-element-like regulatory mechanism that would keep the process in a steady state. We have also used synthetic oligonucleotides as hybridization probes to monitor the spread of sequence variants within individual rodent species, including *Mus domesticus* and *Mus spretus*.

CF 407 MITOCHONDRIAL DNA USED TO TRACK THE DISPERSAL OF A PREFERRED MENU ITEM IN WAIKIKI RESTAURANTS: MAHI-MAHI AS AN EXAMPLE OF GENETIC TAGGING WITH MTDNA, Christopher E. Herzig, Department of Genetics, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI 96822

International predation pressure on pelagic marine species in Hawaii has created concern about the possibility of overfishing in a variety of large-bodied and long-lived fish. Mahi-mahi (Coryphaena hippurus) has a world-wide distribution in tropical and subtropical waters, but precise identification of separate stocks has proven problematic using conventional starch gel electrophoretic methods. As an attempt to develop genetic methods for tagging these animals, mtDNA sequences were amplified via the Polymerase Chain Reaction (PCR) from total genomic DNAs prepared from larvae, fingerlings, and fin clip samples of a tank-reared population used for aquaculture development. In a comparison of about 330 bp of 12S rRNA sequence, all inbred fish examined (n=3) were found to be identical. Presently we are sequencing different samples of wild caught Mahi-mahi in hopes of finding polymorphism. In an attempt to find more intraspecific polymorphism, we are also working on amplifying via PCR the D-loop region and the Cytochrome B gene. If sequence differences can be found then specific probes could be designed to identify different stocks. The ability to screen large samples of wild caught fish with stock specific probes will greatly improve the understanding of population structures.

CF 408 **EVOLUTION OF THE PROTONMOTIVE CYTOCHROME b.** **Neil Howell, Department of Radiation Therapy, The University of Texas Medical Branch, Galveston, TX 77550.**

The amino acid sequences of the protonmotive cytochrome *b* from seven representative and phylogenetically diverse species have been compared to identify protein regions or segments which are conserved during evolution. The sequences analyzed included both prokaryotic and eukaryotic examples as well as mitochondrial cytochrome *b* and chloroplast b_6 proteins. The principal conclusion from these analyses is that there are five protein regions - each comprising about 20 amino acid residues - which are consistently conserved during evolution. These regions are evident despite the low density of invariant amino acid residues. The two most highly conserved regions, spanning approximately consensus residues 130-150 and 270-290, are located in extra membrane loops and are hypothesized to constitute part of the Q_o reaction center. The intramembrane, hydrophobic protein regions containing the heme-ligating histidines are also conserved during evolution. It was found, however, that the conservation of the two protein segments extramembrane to the histidine residues ligating the low potential *b*566 heme group showed a higher degree of sequence conservation. The location of these conserved regions suggests that these two extramembrane segments are also involved in forming the Q_o reaction center. A protein segment putatively constituting a portion of the Q_i reaction center, located approximately in the region spanned by consensus residues 20-40, is conserved in species as divergent as mouse and Rhodobacter. However, it shows substantially less sequence conservation in the chloroplast cytochrome b_6. The catalytic role of these conserved regions is strongly supported by the locations of residues which are altered in mutants resistant to inhibitors of cytochrome *b* electron transport.

CF 409 NUCLEOTIDE VARIATION AMONG POPULATIONS, SUBSPECIES AND SPECIES OF ECTOTHERMS, David L. Jameson, Osher Laboratory of Molecular Systematics, California Academy of Sciences, Golden Gate Park, San Francisco, California 94118. Closely related populations of ectotherms were examined to determine the relation between stage of speciation and the amount and kind of variation in the mitochondrial genome. Purified mitochondrial DNA from 362 individuals was digested with from from 7 to 10 restriction endonucleases. The sites were mapped for 12 species, 20 subspecies and 83 populations. Seven species of treetoads were polymorphic for 87.9% (58/66) of the sites, two species of killifish, 96.2% (52/54), two species of rattlesnake 86.8% (33/38). One tetraploid grey treetoad differs from the diploid in only 2 of 39 sites while another differs by 14 of 41 sites. The proportion of sites polymorphic across 5 subspecies of treetoads was 42.0 (21/50), 4 subspecies of killifish, 42.4 (22/52), and three species of toads, 47.4 (18/38). In summary 48 populations had only 1 haplotype, 15 had two, 6 had 3, 4 had 4, and 2 had 8 haplotypes. While in rattlesnakes 18 localities had only one haplotype, and five had two the populations of the toad had an average of 3.6 haplotypes/locality. Within population variation can be explained by population size, distribution and behavioral patterns. The isolation of subspecies and species can explain the consistency in the results at these levels.

CF 410 GENOMIC DISTRIBUTION OF COPIA-LIKE TRANSPOSABLE ELEMENTS IN SOMATIC TISSUES AND DURING DEVELOPMENT OF DROSOPHILA MELANOGASTER

Junakovic N.(1), Di Franco C.(1), Pisano C.(2), Dimitri P.(3), Gigliotti S.(4)/ (1) Centro Acidi Nucleici C.N.R., (2) Centro di Genetica Evoluzionistica C.N.R., (3) Dipartimento di Genetica e Biologia molecolare, Università "La Sapienza, 00185 Roma, Italy, (4) I.I.G.B. Via Marconi 10, Napoli, Italy.

That transposable elements are stable in somatic tissues has been reported at the early stages of studies on transposition. Subsequently, somatic rearrangements have been described in Drosophila melanogaster (Cell 44,7), Drosophila mauritiana (Science 235,1636) and Caenorhabditis elegans (Cell 39,599). We have compared by the Southern technique the genomic distribution of the elements of copia, 412,B 104,mdg 1, mdg 4, 1731 and 297 transposon families in spermatozoa, embryos, 1st.,2d.,3d. instar larvae, salivary glands, brains and adults of two related Drosophila lines. No detectable heterogeneity was observed among the samples compared suggesting that somatic transposition should be a rare event in Drosophila melanogaster.

CF 411 **THE MITOCHONDRIAL SMALL rRNA GENE OF EUGLENA GRACILIS: EVOLUTIONARY IMPLICATIONS, Mary Rose Lamb, Michael Gilmore and Jerry Stultz, Department of Biology, University of Puget Sound, Tacoma, WA 98416. On the basis of cristae structure and biochemical comparisons of cytochrome c from a number of protistans, a polyphyletic origin has been proposed for mitochondria (Stewart & Mattox, J. Mol. Evol. 21: 54-57, 1984). Two types of cristae have been observed in the mitochondria of protistans, tubular and lamellar cristae. Any one organism has only one of these structural types. These mitochondrial structures are thought to be diagnostic of important evolutionary divergences in the protistans because they are conserved in monophyletic groups and correlate with other characteristics that are used to group protistans. Do mitochondrial genomes also suggest a polyphyletic origin for the organelle? To answer this question, we are determining the sequence of the small ribosomal RNA from a colorless mutant of Euglena gracilis, a protistan with lamellar cristae. This sequence will be compared with the published sequences for mitochondrial rRNAs from organisms with lamellar cristae (e.g. Chlamydomonas) and those with tubular cristae (e.g. Tetrahymena, Paramecium) for features which correlate with the mitochondrial structure.**

CF 412 DISPERSAL OF LINE1 INTERSPERSED REPEAT FAMILY MEMBERS IN MICE,

Sandra L. Martin, Department of Cellular and Structural Biology, University of Colorado Health Sciences Center, Denver, CO 80262. L1, or LINE1, is a repetitive DNA family found in all mammals. Although individual elements of L1 may be more than 6 kb long in mice, most copies interspersed throughout the genome are truncated at their 5'ends. DNA sequence data collected from a sub-region of L1 from three species of mice indicates that the L1-family members within each species are evolving in concert. In addition, L1 sequences evolve as if they are under selection for protein-coding function. Thus, it appears that L1 is a large family of truncated pseudogenes derived from a smaller number of functional genes. Other structural features suggest that the apparent concerted evolution of L1 is due to a constant dispersal of new elements. Perhaps dispersal is mediated by protein products encoded by the two open reading frames found in long L1 elements, and these proteins interact with full-length L1 transcripts which are also necessary intermediates in the generation of new L1 insertions. If the concerted evolution of L1 is due to constant insertion of new elements, the specific, functional expression of L1 must occur in germ cells or early in embryogenesis. Thus, the production of L1 mRNA may be tightly controlled as a function of mouse development. We have examined a number of mouse tissues and cell lines for L1 expression. The distribution of specific L1 transcripts will be discussed as it relates to L1 dispersal.

CF 413 MITOCHONDRIAL GENES AMPLIFIED VIA THE POLYMERASE CHAIN REACTION: DYNAMICS OF NUCLEOTIDE SUBSTITUTIONS AND EVOLUTION OF CICHLID FISHES A. Meyer, T.D. Kocher and A.C. Wilson Dept. of Biochemistry, Univ.of California, Berkeley, CA 94720.
Parts of the mitochondrial 12S rRNA and the cytochrome b genes of cichlid fishes were amplified with human primers and sequenced directly. 20 taxa of cichlid fishes from Africa and the Neotropics were examined and their phylogenetic relationships investigated. Sequence divergence is about twice as fast in the cytochrome b as in the 12sRNA gene. The evolutionary patterns of sequence divergence was found to be the same in these morphologically fast evolving fishes, as it was for published cases of mammals. In the cytochrome b gene we found that there is a strong transition bias for closely related taxa and C-T changes outnumber G-A by far. Replacement changes were rare. Aminoacid changes matched predictions based on the function of cytochrome b. The 12s sequences diverge at faster rates in loops than in stem regions. Also here transitions are much more common than transversions. Surprisingly, preliminary results suggest that some African cichlids are more closely related to Neotropical cichlids than to other Afircan cichlids. Sequencing via the polymerase chain reaction in combination with direct sequencing is superior to traditional restriction fragment analysis in molecular studies of evolution.

CF 414 THE EVOLUTIONARY ORIGIN OF THE GLYCOSOME IN TRYPANOSOMA BRUCEI, Paul Michels[1], Martine Marchand[1] and Rik Wierenga[2], 1) International Institute of Cellular and Molecular Pathology, Brussels, Belgium and 2) University of Groningen, The Netherlands. In Trypanosomatidae the glycolytic pathway is organized in a unique manner: most of its enzymes are localized in a peroxisome-like organelle, called the glycosome. The evolutionary origin of peroxisomes is not yet established: the organelle, which does not contain any DNA, could either be derived from other intracellular membranous structures in the ancestral eukaryotic cell or from a domesticated endosymbiont. We have addressed this question by an analysis of the genes for five glycolytic enzymes that are found in the glycosome of Trypanosoma brucei: glucosephosphate isomerase (PGI), aldolase, triosephosphate isomerase (TIM), glyceraldehydephosphate dehydrogenase (GAPDH) and phosphoglycerate kinase (PGK). Moreover, the genes for the cytosolic pyruvate kinase and the cytosolic isoenzymes of GAPDH and PGK were studied. Comparison of the predicted amino-acid sequences of the glycosomal proteins with the sequences of the cytosolic isoenzymes and those of corresponding proteins of other organisms has revealed some features unique to the organellar proteins. These are clusters of positive residues and short sequence elements. They are presumably involved in the routing of these proteins to the glycosome and/or their functioning within the organelle. However, these features do not allow a conclusion about the evolutionary origin of the glycosome. A phylogeny reconstruction using the available sequences of the glycolytic enzymes, both glycosomal and cytosolic, supports the conclusion obtained from the analysis of ribosomal RNAs, that the Trypanosomatidae diverged very early from the eukaryotic branch of the phylogenetic tree. None of the glycosomal proteins displayed specific prokaryotic features. This could either mean that the glycosome did not originate from a prokaryotic endosymbiont or that the organellar proteins have lost their prokaryotic nature. A striking difference could however be observed between T.brucei's two GAPDH isoenzymes. The mere 50 % sequence identity between these two proteins indicates a separate origin. Whether this is the result of an endosymbiontic origin of the glycosome, or has to be attributed to other ways of horizontal gene transfer, remains to be established.

CF 415 PHYLOGENETIC RELATIONSHIP OF PROCHLOROPHYTES TO CYANOBACTERIA AND CHLOROPLASTS BASED ON SEQUENCE ANALYSIS OF *psbA* GENES IN *PROCHLOROTHRIX HOLLANDICA*, Clifford W. Morden and Susan Golden, Department of Biology, Texas A&M University, College Station, Texas 77843.
Prochlorothrix hollandica is an oxygen-evolving photosynthetic prokaryote, termed a prochlorophyte, characterized by the presence of a chlorophyll *b*- containing light-harvesting antenna rather than phycobilisomes as in cyanobacteria. Overall pigment composition and thylakoid membrane structure suggest an intermediate nature of *P. hollandica* between cyanobacteria and the chloroplasts of higher plants. The *P. hollandica psbA* genes which encode the photosystem II 32 kd thylakoid protein, D1, were cloned and sequenced. There are two *psbA* genes present in *P. hollandica* which encode an identical amino acid sequence. The *P. hollandica psbA* sequences were compared to those reported from cyanobacteria, a green alga, a liverwort, and higher plants. Like all chloroplast *psbA* genes, there is a seven amino acid gap near the carboxy terminus of the derived protein relative to the protein predicted by cyanobacterial genes. This indicates that *P. hollandica* is a part of the lineage that led to chloroplasts in green plants after a divergence from cyanobacteria. This hypothesis is also supported by cladistic analysis of derived D1 amino acid sequences from *psbA* genes of eleven taxa. Research efforts are continuing with the analysis of other genes encoding proteins of the photosystem II reaction center and the large subunit of ribulose-bisphosphate carboxylase.

CF 416 INSERTION ELEMENTS AND PLASMID DYNAMICS IN GAS VACUOLE DEFICIENT MUTANTS OF HALOBACTERIUM HALOBIUM

Felicitas Pfeifer, Ulrike Blaseio and Mary Horne, Max-Planck-Institut für Biochemie, D-8033 Martinsried, Fed. Rep. of Germany

The major plasmid, pHH1 (150 kb), of the archaebacterium Halobacterium halobium contains copies of various insertion elements (ISH) that through transposition or spontaneous recombination events lead to mutations at frequencies as high as 10-2 affecting gas vacuole production. Two types of mutational events were observed to affect the expression of the plasmid-borne copy of a gas vacuole protein encoding gene. These mutations involved either the integration of an insertion element near the gene or a deletion event encompassing the entire vac gene region. The deletion formation within pHH1 involved a recombination between one copy of ISH2 or ISH27 and a short DNA-sequence homology. Further analysis of the plasmid dynamics detected a transposition burst of ISH27. At least three unique though highly homologous copies of ISH27 transposed during the development of the culture. This ISH27 element family of H. halobium is related to the family of ISH51 elements characterized in Haloferax volcanii.

Horne, MC, Englert, C, and Pfeifer, F (1988) MGG 213:459-464
Pfeifer, F, Blaseio, U, and Ghahraman, P (1988) J Bacteriol 170:3718-3724

CF 417 RESTRICTION ENZYME ANALYSIS OF MITOCHONDRIAL DNA FROM TWO SPECIES OF MICROTINE RODENT WITH COMMENTS ON THE ORIGIN OF MICROTUS BREWERI. Carleton J. Phillips, Dorothy E. Pumo and Carol Ann Briskey, Department of Biology, Hofstra University, Hempstead, NY 11550. Mitochondrial DNA (mtDNA) was isolated from two microtine rodents, Microtus pennsylvanicus and M. breweri. Specimens of the former were from Long Island, Connecticut, Rhode Island, and Maine. The latter species is only known from Muskeget Island, off the coast of Massachusetts. Purified mtDNA was digested with restriction enzymes, end-labelled, and separated by electrophoresis. Microtine mtDNA was cloned and compared to laboratory mouse mtDNA by means of Southern blots. Five mtDNA genotypes (L1-5), exhibiting 0.37-3.3% nucleotide sequence divergence, were isolated from M. pennsylvanicus. A single genotype, B1, was isolated from M. breweri. The B1 mtDNA was most similar to the L3 mtDNA from M. pennsylvanicus; indeed, the two differed by only two surveyed EcoR I sites hypothesized to be in the 16S rRNA gene. Phylogeographic analysis of mtDNA thus genetically links the endemic island species, M. breweri, to a maternal lineage presently represented in M. pennsylvanicus living in northwestern Maine. The morphologically distinctive M. breweri probably evolved from animals stranded on Muskeget Island by the last glacial retreat at the end of the Pleistocene. (NSF grant BBS-8609231.)

CF 418 MITOCHONDRIAL DNA DIVERSITY AMONG POPULATIONS AND CLOSELY RELATED SPECIES OF ECONOMICALLY IMPORTANT INSECTS. Richard L. Roehrdanz and Carol F. Schmidt. Insect Biochemistry and Molecular Genetics, Biosciences Res. Lab., USDA, ARS, Fargo, ND 58105. Mitochondrial DNA analysis can be used to provide information on the population diversity, evolution and possible dispersal patterns among insect species that are a threat to agriculture. Some of these insects may be targets of genetic or biological control where the existence of great genetic diversity could adversely impact control programs. Screwworm flies from Texas to Costa Rica exhibit wide dispersal of 2 mtDNA morphs while 14 other morphs are limited to a single locale. The number of different lineages equals about half the number of lines examined but the estimated sequence divergence is fairly low (<1%) for most pairs. The screwworm has also been compared with two other species of Calliphorid flies and the house fly. All three Calliphorids have sequence divergences of >10% vs. the house fly. Screwworm vs. secondary screwworm difference is 5.6%. Both are from the western hemisphere and considered to be in the same genus. The hairy maggot blowfly, from Australia and Asia, is a member of an old world genus, and has a 6.8% difference with the secondary screwworm but >10% difference with the screwworm. Another pest insect under study is the boll weevil. Only four population lines have been examined, but they represent a wide geographic area (Mississippi, Arizona, Mexico, El Salvador) and three different host plants. Although each line is a distinct morph, the per cent sequence divergence is relatively small. A potentially interesting question to ask with this species is whether the shift from native plants to cultivated cotton has been accompanied by a significant reduction in the number of mtDNA lineages.

CF 419 EVOLUTION OF MITOCHONDDRIAL DNA IN CANADA GEESE, Gerald F. Shields, Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, Alaska 99775-0180.
All subspecies of Canada geese (Branta canadensis) possess unique types of mitochondrial DNA as assessed by restriction fragment analysis. Two mitochondrial DNA clones are obvious: the first includes all of the large-bodied subspecies (canadensis, interior, maxima, moffitti, fulva, occidentalis and parvipes), while the second includes all of the small-bodied subspecies (leucopareia, minima, taverneri and hutchinsii). If a 2.0% rate of divergence/million years is accurate for mtDNA evolution within this group, then large-bodied and small-bodied birds diverged initially about 900,000 years ago. Intrasubspecific divergence of mtDNA is essentially zero, and this has lead us to predict that subspecies of Canada geese are founded by small numbers of females and that populations since subspecific differentiation have experienced numerous bottlenecks. A model for radiation of the species across the continent is proposed.

CF 420 MOLECULAR EVOLUTION OF HOBO TRANSPOSABLE ELEMENTS IN DROSOPHILA, Gail M. Simmons, Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709. Population genetics theory now includes a body of work concerning the evolution of transposable element families in eukaryotic hosts. Models of copy number regulation have been amenable to experimental analysis in Drosophila, using *in situ* hybridization to identify element insertion sites. However, very little data is available to test models of sequence divergence in transposable elements. I have isolated members of the hobo transposable element family from EMBL4 libraries of wild-caught *D. melanogaster* and its close relatives, *D. simulans* and *D. mauritiana*. Using synthetic oligonucleotide primers (based on the hobo element sequenced by Streck, MacGaffey and Beckendorf) I have sequenced four full-length hobo elements and have compared their sequences to the published sequence. Sequence variation is apparent even among elements from the same species. The data are applied to the model of Hudson and Kaplan to calculate the time back to the most recent common ancestor for these elements. The results are discussed in relation to the known evolutionary history of the species.

CF 421 EVOLUTION OF THE MITOCHONDRIAL SMALL RIBOSOMAL SUBUNIT IN INSECTS AND ITS IMPLICATION FOR PHYLOGENETIC RECONSTRUCTIONS, Chris Simon[1] and Svante Pääbo[2], 1) Department of Zoology, University of Hawaii, Honolulu, HI 96822 and 2) Department of Biochemistry, University of California, Berkeley, CA 94720. In the last several years many population biologists have used nucleotide sequence information inferred from restriction site mapping of mitochondrial DNA to produce phylogenetic hypotheses. The advent of the polymerase chain reaction (PCR) has made direct sequencing, in which 300-600 sequential base pairs are read at once from specific amplified regions, feasible on the scale necessary for phylogenetic studies. PCR allows the rapid selection and isolation of DNA regions of interest and simplifies the preparation of DNA for sequencing. Data from direct sequencing are less ambiguous and supply more information for a given level of effort than data from restriction mapping. PCR and chain termination sequencing were used to examine a section of the small ribosomal subunit in several insect genera. Comparative studies of this region reveal varying levels of conservation. This conservation is related to secondary structure and imposes constraints which influence 1) compensatory mutations in helical stems, 2) transition/transversion ratios, and 3) percent of A-T nucleotides. The relationship between these three characteristics and level of conservation is discussed. Finally the implications of these data for multiple hits corrections necessary for the calculation of evolutionary rates are explored. Characterization of the variability in this region has another function. Regions identified as variable can be used to construct evolutionary trees for closely related species, and regions identified as conserved can be used to construct evolutionary trees for distantly related species.

CF 422 MITOCHONDRIAL DNA AND ALLOZYMES PROVIDE INSIGHT ABOUT THE PALEOZOOGEOGRAPHY OF *FUNDULUS HETEROCLITUS*. Michael W. Smith, Mark C. Glimcher, Robert W. Chapman and Dennis A. Powers. Department of Biology, Johns Hopkins University, Baltimore, MD. About 10,000 years ago the Atlantic coastline, including the Chesapeake and Delaware bays, was flooded as glaciers melted and sea levels rose. The receding glaciers opened new habitats along the Atlantic coast and its estuaries. Our previous work identified an apparent zone of secondary intergradation between 'northern' and 'southern' races of *Fundulus heteroclitus* along the northern New Jersey coast. To test whether that locality was the initial zone of secondary intergradation prior to the last glaciation (i.e., 20,000 years ago), mitochondrial DNA haplotypes and allozyme genotypes were determined for fish from the Chesapeake bay, Delaware bay and three smaller estuaries in the Carolina's. Populations in the upper Chesapeake and Delaware bays, and their tributaries contained haplotypes similar to fish from northern coastal locations, while populations throughout the rest of these bays were primarily southern coastal haplotypes. Fish from estuaries south of the Chesapeake bay had mtDNA haplotypes similar to southern coastal populations.

Since the Chesapeake and Delaware bays did not exist during the last glacial period, the presence of northern mtDNA haplotypes in the upper Chesapeake bay over 100 km south of the coastal intergrade zone, suggests that extensive migration and colonization has taken place since the last glaciation. Two hypotheses could account for these data. First, if the glaciers depressed the land mass sufficiently to allow the newly formed Hudson and Susquehanna rivers to be connected via the glacial watershed, then the more freshwater and cold tolerant northern race could have invaded the estuaries being formed. In other words, the northern race essentially inoculated the tributaries of the Chesapeake and Delaware bays, while the southern race invaded via coastal waters. Alternatively, the initial coastal zone of secondary intergradation could have been at or below the ancestral mouth of the Chesapeake bay prior to the last glaciation (i.e., 20,000 years ago). As the glaciers retreated, the northern race may have lead an invasion of both races up the coast and into the newly forming Chesapeake and Delaware bays.

CF 423 THE SAME DELETION EVENT IS RESPONSIBLE FOR THE LOSS OF THE INVERTED REPEAT IN PEA AND BROAD BEAN CHLOROPLAST GENOMES, André Steinmetz and Françoise Herdenberger, Institut de Biologie Moléculaire des Plantes, Université Louis Pasteur, 67000 Strasbourg (France)

The chloroplast genome of nearly all higher plants is characterized by the presence of an inverted repeat structure which varies in length from 10 to 28 kbp and contains a variable number of genes including the ribosomal RNA operon. Only a few species lack the inverted repeat structure: they are mostly members of the legume family and include pea and broad bean. The site of deletion of one inverted repeat unit was recently identified in pea. It is flanked by the *trnH* and *ndh5* genes which are separated by a 200 bp DNA sequence that includes two short DNA fragments derived from the *rbcL* and *psbA* genes. We have recently sequenced the corresponding region of the broad bean chloroplast genome and we found the same organization as in pea. This observation as well as the high sequence homology found in the duplicated fragments of the *rbcL* and *psbA* genes as well as in the noncoding regions suggests that pea and broad bean diverged approximately 10 million years ago from a common ancestor that already lacked the inverted repeat structure.

CF 424 ISOLATION OF A NOVEL SPECIFIC FRACTION OF CHROMOSOMAL DNA FROM HUMAN, DROSOPHILA AND PLANT CELLS, Nickolai A. Tchurikov, Natalia A. Ponomarenko and Nickolai I. Barbakar, Department of Nucleic Acids Biosynthesis, Engelhardt Institute of Molecular Biology, USSR Academy of Sciences, Vavilov str., 32, 117984, Moscow B334, USSR

Total cellular unbroken DNA samples from human, Drosophila, Arabidopsis, tomato and maize were isolated inside agarose gels and used in pulsed field gel (PFG) electrophoresis (LKB Pulsaphor system). In conditions appropriate for separation of λ-ladder, S. cerevisiae chromosomes (250 sec. pulses) and S. pombe chromosomes (75 min pulses), a heterogeneous fraction of DNA from human, Drosophila, Arabidopsis, tomato and maize DNA-agarose plugs migrating mainly in a region of 100-200 kb was observed. This fraction (forum DNA, fDNA) makes up ~5% of total DNA. Random cloned fragments of fDNA were used for probing of genomic DNA blots. It was found that fDNA is enriched with some genomic sequenses (especially some transposable elements in experiments with Drosophila tissue culture cells), depleted with transcribed DNA sequences and demonstrates some tissue specificity in one and the same individual. Both the length and the sequence distribution in fDNA suggest that it is excised in a non-random way. We propose that fDNA comes from some higher chromosomal structures which mainly possess 100-200 kb stretches of DNA.

CF 425 MUTATIONAL DYNAMICS IN NATURAL POPULATIONS: GENETIC AND MOLECULAR ANALYSES OF P DNA ELEMENT DISTRIBUTION IN KENYA POPULATIONS OF DROSOPHILA MELANOGASTER AND CLUSTERS OF MUTATION IN EVOLUTIONARY THEORY, R. C. Woodruff[1], J. N. Thompson, Jr.[2], H. N. B. Gopalan[3], W. A. Ngure[3], E. S. Norris[1], and A. A. Szekely[1]. [1]Dept. Biol. Sci., Bowling Green St. Univ., Bowling Green, OH, [2]Dept. of Zoology, Univ. of Oklahoma, Norman, OK, [3]Dept. of Botany, Univ. of Nairobi, Nairobi, Kenya. Although it is clear that mutation is the ultimate source of genetic variation, in the short run how important is mutation to the process of adaptive evolution? For example, mutation rate is often thought to be a predictable and relatively stable factor in evolution. However, we would like to discuss two factors that do affect mutations rates and to emphasize that individual mutation events not only vary, but can vary in ways that are difficult to predict. First, spontaneous premeiotic mutations in D. melanogaster occur much more often than generally assumed; approximately 20% of spontaneous mutations occur in clusters. Such a cluster occurring in a small population, in an inbred population, or in a population in which offspring are from a few parents might have an enhanced likelihood of rapidly becoming polymorphic and contributing to the evolution of the population. Transposable DNA elements, which are known to be major cause of spontaneous genetic change in many organisms and which can cause mutation clusters, are present in high frequencies in some natural populations. We have observed that the majority of 76 natural population lines collected in 15 different sites in Kenya contain complete P elements that synthesize transposase. However, many of these lines have low P activity and this may be due to a recent invasion of some East African populations by P transposons. (Supported in part by Fulbright Research Award 87-47122 to RCW).

CF 500 THE ROLE OF INTRAGENIC SEGMENTAL EXCHANGE IN GENERATING *Aβ* POLYMORPHISM *Stefen A. Boehme and Edward K. Wakeland*, Department of Pathology, University of Florida, Gainesville, Florida.

The class II molecules encoded by the murine major histocompatibility complex are extremely polymorphic cell surface glycoproteins that play a crucial role in controlling the immune response by presenting antigen to regulatory T-lymphocytes. We are defining the role of intragenic segmental exchange in the generation of MHC class II gene polymorphisms using *Aβ* as a model. To this end, we have examined the structural polymorphism of *Aβ* in 13 standard laboratory inbred and 19 wild-derived *H-2* homozygous mouse strains with five Aβ-specific monoclonal antibodies. Thirteen *Aβ* alleles were distinguished with this panel of monoclonal antibodies. Interestingly, three of the *Aβ*-specific monoclonal antibodies exhibited an antithetical relationship for 25 of the 32 strains examined. These two forms correlate with specific structural forms of the *Aβ*1 exon. Alleles that encode molecules which are 10.2.16 -, 25-9-17 +, 34-5-3 + have a two codon deletion in the $A\beta_1$ exon while alleles which encode molecules that are 10.2.16 +, 25-9-17 -,34-5-3 - have the undeleted form. We have previously organized these alleles into 3 evolutionary lineages on the basis of their intron structures. By comparing the intron lineages of these alleles with the structural form of the $A\beta_1$ exon, 28% (7/25) of the alleles are possible candidates to have undergone inter-lineage intragenic segmental exchange in the $A\beta_1$ exon. This prediction has been shown to be valid by nucleotide sequence in the case of $A\beta^b$ and $A\beta^{nod}$. Nucleotide sequencing analysis of the Aβ1 and Aβ2 exons is underway to confirm the results of this serological analysis. (Supported by AI-17966)

CF 501 EVOLUTION OF THE CLASS I ALLELES OF THE MAJOR HISTOCOMPATIBILITY COMPLEX IN HOMINOIDS, David A. Lawlor, Frances E. Ward* and Peter Parham, Department of Cell Biology, Stanford University School of Medicine, Stanford, CA 94303 and *Department of Microbiology and Immunology, Duke University Medical Center, Durham, NC 27710 The class I loci of the major histocompatibility complex (MHC) of man are characterized by a extraordinary degree of allelic diversity. The products of 85 alleles of the A, B, and C loci have been serologically identified and it is estimated that the total number of alleles is in the hundreds. Until recently, it was not known whether the alleles arose following the splitting of the human lineage or were passaged intact from the ancestral species. Our analysis of the MHC of the chimpanzee, *Pan troglodytes*, supports the latter hypothesis since there is a high degree of similarity with man; some human alleles are more similar to chimp than they are to other human alleles. We have expanded our initial study to include two other hominoids, *Gorilla gorilla* and *Pongo pygmaeus*. Complementary DNA clones from the chimp, gorilla and orangutan libraries hybridizing to class I heavy chain and β-2-microglobulin probes have been purified and inserts will be subcloned into M13 for sequence analysis. The species will be assessed for degree of allelic diversity and presence of alleles shared by all or some of the hominoid members. Potential utilization of the sequences for derivation of a hominoid phylogenetic tree will be explored.

CF 502 EVIDENCE FOR OVERDOMINANCE AND REPRODUCTIVE SELECTION IN THE MAINTENANCE OF MHC POLYMORPHISMS IN SEMI-NATURAL POPULATIONS OF *MUS*

Wayne K. Potts, C. Jo Manning, Edward K. Wakeland, Department of Pathology, University of Florida, Gainesville, FL 32610

Genes of the major histocompatibility complex (MHC) are the most polymorphic (coding) loci known for vertebrates. The essential role of MHC gene products in antigen recognition during the initiation of immune responses has led to the assumption that parasite driven selection must be involved in the maintenance of MHC polymorphisms. However, other non-disease based mechanisms such as disassortative mating preferences and selective abortion have also been suggested. To distinguish between these hypotheses we are measuring components of fitness in semi-natural, captive, populations of *Mus domesticus*. These populations carry four MHC haplotypes (derived from inbred lines) on wild genetic backgrounds. More females settled in the territories of MHC heterozygous males than expected from random association of founders ($p<.001$; males, n=37; females, n=96). This was due both to a higher probability of heterozygous males acquiring a territory ($p<.005$) as well as mean harem sizes that were two times larger than territorial homozygous males (n.s.). In terms of MHC haplotype sharing between prospective mates, female settlement in male territories did not differ from random. However, the production of MHC heterozygous offspring exceeded random mating expectations ($p<.001$, n=194). These results suggest that both overdominance (among males) and reproductive mechanisms operate in the maintenance of MHC polymorphisms in *Mus*.