## OCCASION

This publication has been made available to the public on the occasion of the 50th anniversary of the United Nations Industrial Development Organisation.

## DISCLAIMER



## FAIR USE POLICY



## CONTACT

Please contact publications@unido.org for further information concerning UNIDO publications.

For more information about UNIDO, please visit us at www.unido.org

UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION
Vienna International Centre, P.O. Box 300, 1400 Vienna, Austria
Tel: (+43-1) 26026-0 · www.unido.org · unido@unido.org

20756

INTERNATIONAL SOCIETY
OF
MOLECULAR EVOLUTION

# WORKSHOP
# ON
# OPEN QUESTIONS IN MOLECULAR EVOLUTION

**Guanacaste, Costa Rica, April 18-23, 1994**

**Organizers : Giorgio Bernardi, Morris Goodman, Barry Hall, Gabriel Macaya and Mary-Jane West-Eberhard**

**A Workshop sponsored by the Sloan Foundation, the International Center for Genetic Engineering and Biotechnology, the International Union of Biochemistry and Molecular Biology, the Committee on Genetic Experimentation of the International Council of Scientific Unions, Springer Verlag and Consejo Nacional para Investigaciones Cientificas y Tecnologicas, Costa Rica**

## TABLE OF CONTENTS

## INTRODUCTION

The Workshop on "Open questions in molecular evolution" was organized with the purpose to discuss a number of basic questions in molecular evolution which are still open and debated. These questions concerned : Genome dynamics and mechanisms in molecular evolution, selection-induced mutations, isochores and the selectionist-neutralist controversy, mammalian phylogeny, protein evolution and the antiquity of introns and the evolution of viral sequences.

The Workshop took place between April 18-23 at the Hotel Las Espuelas, near Liberia, in the province of Guanacaste, Costa Rica. The hotel is located in the proximity of two conservation areas, Tempisque and Guanacaste (Santa Rosa Park). This location proved to be ideal, because it allowed personal contacts and short visits to the conservation areas in the little time left free by a rather intense schedule.

Three features worth mentioning of the Workshop were the following (i) the participation of most scientists who are active in the areas dealt with in the Workshop; these included two Nobel Laureates : Werner Arber and Walter Gilbert; (ii) the participation of a number of young researchers from Costa Rica and Latin America; (iii) a series of evening lectures on general subjects ranging from evolution history to natural history of Costa Rica. The total number of participants was about 90.

## Sessions

### Monday afternoon (April 18)

*Genome dynamics and mechanisms in molecular evolution*

**Chairman : W. Gilbert**

**W. Arber** (Basel)
"Gene functions involved in the biological evolution of prokaryotes"

**P. Miramontes** (Montreal) sc
"DNA sequence analysis shows distinctive evolutive pathways between eukaryotes and prokaryotes"

**G. Dover** (Leicester)
"There's a lot more to molecular evolution than selection and drift"

**E. Zuckerkandl** (Palo Alto)
"Molecular pathways to parallel evolution"

**S. Saccone** (Bari)
"The use of glutamine synthetase enzymes for tracing the tree of life"

**C. Lanave** (Bari) sc
"Glutamine synthetase gene evolution in bacteria

**P. Leon** (San José)
"Genome evolution in the RNA world and the emergence of DNA"

**Chairman : Giorgio Bernardi**

**Evening lecture : R. Keynes** "Erasmus Darwin's *Temple of Nature*"

### Tuesday (April 19)

*Genome dynamics and mechanisms in molecular evolution* (cont.)

**Chairman : G. Dover**

**R.J. Britten** (Corona del Mar)
"Evolutionary selection against change in many alu repeat sequences interspersed through primate genomes"
"The importance of the Gypsy/Ty3 retrotransposons" (poster)

**N. Okada** (Yokohama)
"On the possibility of horizontal transmission of SINEs"

**N. Junakovic** (Roma) sc
"Polymorphisms in the genomic distribution of transposable elements in inbred Drosophila lines"

**P. Sniegowski** (East Lansing) sc
"Containment of retrotransposon copy number in Drosophila : a test of a current model"

sc : short communication

*Selection-induced mutations*

**Chairman : W. Arber**

**J. Cairns** **(Oxford)**

"The origins of the controversy about adaptive mutation"

**P. Sniegowski** **(East Lansing)**

"Adjustment of mutation rates : proximate or ultimate causation?"

**B. Hall** **(Rochester)**

"Selection-induced mutations"

*Isochores and the selectionist-neutralist controversy*

**Chairman : E. Trifonov**

**Giorgio Bernardi** **(Paris)**

"Isochores and the evolution of the vertebrate genome"

**S. Karlin** **(Stanford)**

"Patchiness of DNA and implications for sequence modeling"

**J. Fickett** **(Los Alamos)**

**G. Macaya** **(San José)**

**J.L. Oliver** **(Granada)**

"A model to generate nucleotide sequences with long-range correlations"

**Chairman : J. Gillespie**

**T. Ikemura** **(Mishima)**

"Complete form of PAB (XY pseudoautosomal boundary-like sequence exists near boundary of long-range G+C% mosaic domains in the human MHC locus"

**H. Musto** **(Montevideo)** **sc**

"Compositional patterns of nuclear encoded genes of *Trypanosoma brucei* and *T. cruzi*"

**W.-H. Li** **(Houston)**

"Molecular evolution of Y- and X-linked genes"

**K. Wolfe** **(Dublin)**

"Molecular evolution of imprinted genes"

**A.L. Hughes** **(Penn State)**

"The evolution of functionally novel proteins after gene duplication"

**Giacomo Bernardi** **(Pacific Grove)**

"Molecular adaptations to a changing environment in the killifish *Fundulus heteroclitus*"

**Chairman : B. Hall**

**Evening lecture : W. Provine**

"Neutral theories of molecular evolution in historical perspective"

## Wednesday (April 20)

*Isochores and the selectionist-neutralist controversy* (cont.)

**Chairman : T. Ikemura**

**N. Sueoka** **(Boulder)**

"Directional mutation pressure and dynamics of molecular evolution : parity rules of DNA base composition at equilibrium and strand biases"

"Further support of difference in directional mutation pressure as a major underlying cause of isochore formation in multicellular eukaryotes"

**G. Holmquist** **(Duarte)**

"Patterns of damage, repair, and mutation along the mammalian genome"

**W.-H. Li** **(Houston)** **sc**

"A mutation model for the origin of GC-rich isochores in vertebrate genomes"

**A. Marin** **(Sevilla)** **(poster)**

"Differences in DNA repair, G+C content and expression level among *E. coli* genes"

**A. Eyre-Walker** **(Piscataway)**

"The evolution of synonymous codon bias in enteric bacteria"

**Chairman : T. Ohta**

**W. Fitch** **(UCLA)**

"The molecular clock may be better than you think"

**J. Gillespie** **(Davis)**

"Substitution processes in molecular evolution"

**T. Ohta** **(Mishima)**

"Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory "

**C. Gautier** **(Lyon)**

"Genomic regional structures and substitution rate in vertebrates"

**Giorgio Bernardi** **(Paris)**

"Natural selection and random drift in synonymous positions of mammalian genes"

**T. Steen** **(Ithaca)** **(poster)**

"Where do we place the nearly neutral theory ? : the development of the nearly neutral theory and its position in the controversy"

**Chairman : M. Goodman**

**Evening lecture : R. Lewin**

## Thursday (April 21)

*Mammalian Phylogeny*

**Chairman : W.-H. Li**

**M. Goodman** **(Detroit)**

"DNA sequence evidence on primate phylogeny"

**U. Arnason** **(Lund)**

"A molecular view of pinniped relationships with particular emphasis on the true seals"

**J. Beintema** **(Groningen)** **(poster)**

"Evolution of arthropod hemocyanin and insect storage proteins (Hexamerins)"

**H.J. Breukelman** **(Groningen)** **(poster)**

"Comparing nucleic acid sequences of different ruminant secretory ribonuclease genes"

**W.W. de Jong** **(Nijmegen)**

"Molecular evolution of the vertebrate eye lens"

**D. Graur** **(Tel Aviv)**

"A protocol for higher-level phylogenetic reconstructions with an example concerning the ordinal phylogeny of eutherians"

**Chairman : W. Fitch**

**D. Hewett-Emmett** **(Houston)**

"Conservation and convergence in the evolution of the carbonic anhydrase gene families"

**R. Honeycutt** **(Washington)**

"Mammalian mitochondrial DNA evolution : a comparison of rodents and artiodactyls"

**J.C. Mounolou** **(Gif-sur-Yvette)**

"Changes in genetic diversity of european rabbit populations during historical times"

**G. Pesole** **(Bari)**

"The use of a metazoan mt database (MmtDB) for studying the evolution of the D-loop region in humans"

**J. Pettigrew** **(Brisbane)**

"Flying DNA"

**K.D. Tartof** **(Philadelphia)**

"The evolution of mammals as inferred from syntenic genetic maps"

**Chairman : M.J. West-Eberhard**

**Evening lecture : R. Gamez**

**Friday (April 22)**

*Protein evolution - Antiquity of introns*

**Chairman : E. Zuckerkandl**

**E. Trifonov (Rehovot)**

" Protein sequences as fossil record of early gene fusion"

**S. Ohno (Duarte)**

"Active sites of ligands and their receptors are made of common peptides that are also found elsewhere"

**T. Gojobori (Mishima)**

"Domain evolution with special reference to kringle structures"

**R. Cerff (Braunschweig)**

"The chimeric nature of nuclear genomes and the antiquity of introns as demonstrated by the GADPH gene system"

**J.M. Logsdon Jr. (Bloomington)**

"The recent origin and punctuated evolution of nuclear spliceosomal introns"

**M. Go (Nagoya)**

"Origin of introns : were the original blocks of proteins encoded by exons ?"

**W. Gilbert (Cambridge, Ma.)**

"The antiquity of introns"

*Evolution of viral sequences*

**Chairman : T. Gojobori**

**W. Fitch (UCLA)**

"Evolutionary rates in Influenza and VSV are ecologically controlled"

**S. Karlin (Stanford)**

"New methods in molecular evolutionary comparisons"

**M. Roossinck (Ardmore)**

"Helper virus-specific mutation and selection in the evolution of cucumber mosaic cucumovirus satellite RNAs"

**P. Sharp (Nottingham)**

"Mosaic evolutionary histories of 'AIDS' viruses"

**Chairman : G. Macaya**

**Evening lecture : H. Hoenigsberg**

**Saturday (April 23)**

Excursion

INTERNATIONAL SOCIETY
OF
MOLECULAR EVOLUTION

# WORKSHOP
# ON
# OPEN QUESTIONS IN MOLECULAR EVOLUTION

## Guanacaste, Costa Rica, April 18-23, 1994

**Organizers : Giorgio Bernardi, Morris Goodman, Barry Hall, Gabriel Macaya and Mary-Jane West-Eberhard**

## LIST OF SPEAKERS & PARTICIPANTS


Dr. W. Arber
Abteilung Mikrobiologie
Biozentrum der Universität
Klingelbergstrasse 70
4056 Basel (Switzerland)
Tel. 41- 61 267 21 30 - Fax 41- 61 267 2118

Dr. U. Arnason
Department of Genetics
Division of Evolutionary Molecular Systematics
University of Lund
Sölvegatan 29
223 62 Lund (Sweden)
Tel. 46- 46 109259 - Fax 46- 46 147874

Dr. Giacomo Bernardi
Hopkins Marine Station - Stanford University
Department of Biological Sciences
Pacific Grove, CA. 93950-3094 (U.S.A.)
Tel. 408 655 6210 - Fax 408 375 0793 - E-mail "Bernardi@leland.Stanford.EDU"

Dr. Giorgio Bernardi
Laboratoire de Génétique Moléculaire
Institut Jacques Monod
2, Place Jussieu
75005 Paris (France)
Tel. 43 29 58 24 - Fax 44 27 79 77 - E-mail"Bernardi@Citi2.Fr"

Dr. J. Beintema
University of Groningen
Department of Chemistry
Nijenborgh 4
9747 AG Groningen (The Netherlands)
Tel. 31-50 634209 - Fax 31-50 634165

Dr. H.J. Breukelman
University of Groningen
Department of Chemistry
Nijenborgh 4
9747 AG Groningen (The Netherlands)
Tel. 31-50 634209 - Fax 31-50 634165

Dr. R.J. Britten
California Institute of Technology
Kerckhoff Marine Laboratory
101 Dahlia Street
Corona del Mar, CA. 92625 (U.S.A.)
Tel. 714 675 2159 - Fax 714 675 1837 - E-mail"rbritten@darby.bio.uci.edu"

Dr. S. Caccio
Laboratoire de Génétique Moléculaire
Institut Jacques Monod
2, Place Jussieu
75005 Paris (France)
Tel. 44 27 51 12 - Fax 44 27 79 77 - E-mail"Bernardi@Citi2.Fr"

Dr. J. Cairns
Clinical Trial Service Unit &
ICRF Cancer Studies Unit
Radcliffe Infirmary
Oxford OX2 6HE (England)
Tel. 44-865 57241 - Fax 44- 865 58817

Dr. R. Cerff
Institut für Genetik - Biozentrum
Technische Universität Braunschweig
Spielmannstrasse 7
38106 Braunschweig (Germany)
Tel. 531 391 5771 - Fax 531 391 5765 - E-mail"rcerff@dbstu1.rz.tu.bs.de"

Dr. W. De Jong
Department of Biochemistry
University of Nijmegen
P.O. Box 9101
6500 HB Nijmegen (The Netherlands)
Tel. 31- 80 616 848 - Fax 31- 80 540 525 - E-mail"u62900.at.hnykun11"

Dr. S.J. de Souza
Ludwig Institute for Cancer Research
Rue Antonio Prudente 109-4 Andar
01509-010 Sao Paulo, SP (Brasil)
Fax 55-11 270 7001

Dr. G. D'Onofrio
Stazione Zoologica
Villa Comunale 1
80121 Naples (Italy)
Fax 39-81 764 1355

Dr. G. Dover
Department of Genetics
University of Leicester
Leicester LE1 7RH (U.K.)
Fax 44-71 261 6162

Dr. L. Duret
Laboratoire de Biométrie
Génétique et Biologie des Populations
U.R.A. 243 - Université Claude Bernard
69600 Villeurbanne (France)
Tel. 72 44 81 42 - Fax 78 89 27 19

Dr. A. Eyre-Walker
Department of Biological Sciences
Rutgers University
Piscataway, NJ 08855-1059 (U.S.A.)
Tel. 908 932 0623 - Fax 908 932 5870

Dr. J.W. Fickett
Theoretical Biology & Biophysics
Group T-10, Mail Stop K710
Los Alamos National Laboratory
Los Alamos, New Mexico 87545 (U.S.A.)

Dr. W. Fitch
Department of Ecology & Evolutionary Biology
716 Engineering Building
University of California - Irvine
Irvine, CA. 92717 (U.S.A.)
Fax 714 725 2181

Dr. C. Gautier
Laboratoire de Biométrie
Génétique et Biologie des Populations
U.R.A. 243 - Université Claude Bernard
69600 Villeurbanne (France)
Tel. 72 44 81 42 - Fax 78 89 27 19 - E-mail"cgautier@biomserv.univ.lyon1.fr"

Dr. W. Gilbert
The Biological Laboratories
Department of Cellular & Developmental Biology
Harvard University - 16 Divinity Avenue
Cambridge, MA. 02138 (U.S.A.)
Tel. 617 495 0760 - Fax 617 496 4313 - E-mail"gilbert@chromo.harvard.edu"

Dr. J. Gillespie
Section of Evolution and Ecology
University of California
Davis, CA. 95616 (U.S.A.)
Tel. 916 752 6018 - Fax 916 752 1449 - E-mail"jhgillespie@ucdavis.edu"

Dr. M. Go
Nagoya University
Chikusa-ku, Furo-cho
Nagoya 464-01 (Japan)
Tel. 81 52 789 2976 - Fax 81 52 789 2977- E-mail "d42807b@nucc.cc.nagoya-u.ac.jp"

Dr. T. Gojobori
National Institute of Genetics
Shizuoka-Ken
Mishima 411 (Japan)
Fax 81- 559 - 75 6040

Dr. M. Goodman
Department of Anatomy & Cell Biology
Wayne State University School of Mecicine
430 Lande
Detroit, MI. 48201 (U.S.A.)
Tel. 313 577 1138 - Fax 313 577 1080 - E-mail"mg@tree.roc.wayne.edu"

Dr. D. Graur
National Center for Biotechnology Information
National Library of Medicine - Building 38A
8600 Rockville Pike
Bethesda, MD. 20894 (U.S.A.)

Dr. B. Hall
Department of Biology
University of Rochester
Hutchinson Hall
Rocheser, N.Y. 14727 (U.S.A.)
Tel. 716 275 0721 - Fax 716 275 2070 - E-mail"bhall@bgh.rochester.edu"

Dr. D. Hewett-Emmett
The University of Texas
Health Science Center at Houston
P.O. Box 20334
Houston, Texas 77225 (U.S.A.)
Tel. 713 792 4585 - Fax 713 792 4595

Dr. G. Holmquist
Beckman Research Institute of the City of Hope
Department of Biology
1450 E. Duarte Road
Duarte, CA. 91010 (U.S.A.)

Dr. R. Honeycutt
Department of Fish and Wild Life Science
Texas A & M University
College Station, Texas 77843-2258 (U.S.A.)
Tel. 409 845 5777 - Fax 409 845 3786 - E-mail "E070rh@tamvl.tamu.edu"

Dr. T. Ikemura
National Institute of Genetics
Yata 1, Shizuoka-ken
111 Mishima 411 (Japan)
Tel. 81- 559 75 0771 - Fax 81-559 71 36510

Dr. N. Junakovic
Centro Acidi Nucleici
Dipartimento di Genetica e Biologia Molecolare
Università "La Sapienza"
00185 Roma (Italy)
fax 39-6 444 0812

Dr. S. Karlin
Department of Mathematics
Stanford University
Stanford, CA. 94305-2125 (U.S.A.)
Tel. 415 723 2204 - Fax 415 725 2040 - E-mail"fd.zgg@forsythe.stanford.edu"

Dr. R. Keynes
Physiological Laboratory
Downing Street
Cambridge CB2 3EG (U.K.)
Fax 44-223 333 840

Dr. C. Lanave
Istituto di Chimica Biologia
Università di Bari
Via Amendola 165/A
70126 Bari (Italy)
Tel. 80 544 3305 - Fax 80 544 3317 - E-mail"lanave@mvx36.csata.it"

Dr. R. Lewin
390 Broadway
Cambridge, Ma. 02139 (U.S.A.) Tel. & Fax 617 492 3082

Dr. P. Leon
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José (Costa Rica) - Fax 506-2 24 61 58

Dr. W.-H. Li
The University of Texas
Health Science Center at Houston
Graduate School of Biomedical
Sciences - P.O. Box 20334
Houston, Texas 77225 (U.S.A.)
Tel. 713 792 4619 - Fax 713 792 4615 - E-mail"li@gsbs18.gs.uth.tmc.edu"

Dr. J. Logsdon
Indiana University - Department of Biology
Jordan Hall 142
Bloomington, Indiana 47405 (U.S.A.)
Tel. 812 855 2549 - Fax 812 855 6705 - E-mail"jlogsdon@bio.indiana.edu"

Dr. G. Macaya
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José (Costa Rica)
Fax 506-2 24 61 58 - E-mail "G.macaya@ucrvm2.ucr.ac.cr"

Dr. L.P. Moreno
Laboratorio de Bioquimica de Virus
Instituto de Biotecnologia Universidad Nacional de Colombia
Santa Fe, Bogota 14490 (Colombia)
Fax 57-1 269 8164

Dr. A. Marin
Departemento de Genetica
Facultad de Biologia
Apartado 1095
41080 Sevilla (Spain)
tel. 34-5 455 7113 - fax 34-5 455 7104 - E-mail"anmarin@cica.es"

Dr. P. Miramontes
Département de Biochimie
Université de Montréal
C.P. 6128, succ. centre-ville
Montréal, Québec (Canada)
Fax 514 343 2210 - E-mail "miramonp@ere.umontreal.ca"

Dr. D. Mouchiroud
Laboratoire de Biométrie - Génétique et Biologie des Populations
U.R.A. 243 - Université Claude Bernard
69600 Villeurbanne (France)
Tel. 72 44 81 42 - Fax 78 89 27 19 - E-mail"cgautier@biomserv.univ.lyon1.fr"

Dr. J.-C. Mounolou
Centre de Génétique Moléculaire - C.N.R.S.
91198 Gif-sur-Yvette (France)
Fax 69 07 53 22

Dr. H. Musto
Departamento de Genética - Facultad de Medicina
Gral. Flores 2144
Montevideo 11800 (Uruguay)
Tel. 598-2 949 562 - Fax 598-2 949 563 - E-mail "hmusto@ibiol.edu.uy"

Dr. S. Ohno
Beckman Research Institute of the City of Hope
1450 East Duarte Road
Duarte, CA. 91010-0269 (U.S.A.)
Tel. & Fax 818 357 1929

Dr. T. Ohta
National Institute of Genetics
Yata 1, Shizuoka-ken
111 Mishima 411 (Japan)
Tel. 81- 559 75 0771 - Fax 81- 559 71 3651 - E-mail"tohta@ddbj.nig.ac.jp"

Dr. N. Okada
Faculty of Bioscience and Biotechnology
Tokyo Institute of Technology
4259 Nagatsuta-cho, Midori-ku
Yokohama 227 (Japan) Tel. & Fax 81 45 923 1136

Dr. J.L. Oliver
Dpto. de Genetica - Facultad de Ciencias
Universidad de Granada
18071 Granada (Spain)
Tel. 34 58 24 32 61 - Fax 34 58 24 40 73 - E-mail "oliver@ugr.es"

Dr. G. Pesole
Istituto di Chimica Biologia
Università di Bari
Via Amendola 165/A
70126 Bari (Italy)
Tel. 80 544 3305 - Fax 80 544 3317 - E-mail"lanave@mvx36.csata.it"

Dr. J. Pettigrew
The University of Queensland
Vision, Touch & Hearing Research Centre
Brisbane Qld 4072 (Australia)
Tel. 61-7 365 3842 - Fax 61-7 365 4522 - E-mail"j.pettigrew@uthrc.uq.edu.au"

Dr. W. B. Provine
Section of Ecology & Systematics
Cornell University
Carson Hall
Ithaca, N.Y. 14853 (U.S.A.)
Fax 812 855 6705

Dr. L. Rodriguez
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José (Costa Rica) Fax 506-2 24 61 58

Dr. M. Roossinck
The Noble Foundation
Division of Plant Biology - P.O. Box 2180
Ardmore, OK 73402 (U.S.A.)
Tel. 405 223 5810 - Fax 405 221 7380 - Email"roossinck@aardvark.nes.uoknor."

Dr. C. Saccone
Istituto di Chimica Biologia - Università di Bari
Via Amendola 165/A
70126 Bari (Italy)
Tel. 80 544 3305 - Fax 80 544 3317 - E-mail"lanave@mvx36.csata.it"

Dr. P. Sharp
Department of Genetics - Queen's Medical Building
University of Nottingham
Nottingham NG7 2UH (U.K.)
Tel. 44- 602 709 263 - Fax 44- 602 709 906

Dr. P. Sniegowski
Center for Microbial Ecology
Michigan State University
East Lansing, MI. 48824-1325 (U.S.A.)
Tel. (19)1- 517 355 3728 - Fax (19)1- 517 353 2917 - E-mail"psnieg@ant.css.msu.edu"

Mrs. T. Y. Steen
Ecology & Evolutionary Biology
237E Corson Hall
Cornell University
Ithaca, N.Y. 14850 (U.S.A.)
Fax 607 255 6044 - E-mail "ts25@cornell.edu"

Dr. N. Sueoka
Department of Molecular, Cellular & Developmental Biology
University of Colorado
Boulder, CO. 80309-0347 (U.S.A.)
Fax 303 492 0388 - E-mail "sueoka@beagle.colorado.edu"

Dr. E. Trifonov
Weizmann Institute of Science
Department of Structural Biology
76100 Rehovot (Israel)
Tel. 972-8 324 653 - Fax 972 8 344 105 - E-mail "bptrifo1@weizmann.ac.il"

Dr. M.-J. West-Eberhard
Smithsonian Institution - c/o Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José (Costa Rica) - Fax 506-2 24 61 58

Dr. K. Wolfe
University of Dublin - Department of Genetics
Lincoln Place Gate - Trinity College
Dublin 2 (Ireland)
Tel. (353 1 702 1253 - Fax 353 1 679 8558 -E-mail "khwolfe@unix2.tcd.ie"

Dr. E. Zuckerkandl
Institute of Molecular Medical Sciences
460 Page Mill Road
Palo Alto, California 94306 (U.S.A.)
Fax 415 322 3009

Participating under-graduate students from Costa Rica:

Yolanda Brenes
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica
Phone: (506) 224 67 49
Fax: (506) 224 93 67

Ingrid Rodríguez
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica
Phone: (506) 224 67 49
Fax: (506) 224 93 67

Edward Ruiz
Escuela de Biología
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica

Latin America graduate students in Costa Rica:

Sergio Melgar
Escuela de Biología
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica
Student from Guatemala

Mabel Paz
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica
Phone: (506) 224 67 49
Fax: (506) 224 93 67
Student from Colombia

Ana Xet
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica

Dr. Oscar Rocha
Escuela de Biologia
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica

Dr. Luis Rodriguez
PIET
Escuela de Medicina Veterinaria
Universidad Nacional
Heredia
Costa Rica

Participating graduate students from Costa Rica:

Carolina Bonilla
PIET
Escuela de Medicina Veterinaria
Universidad Nacional
Heredia
Costa Rica

Gustavo Gutiérrez
Instituto de Investigaciones en Salud
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica

Miryam Hernández
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica
Phone: (506) 224 67 49
Fax (506) 224 93 67

Richard Kogel
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica
Phone: (506) 224 67 49
Fax (506) 224 93 67
e-mail: rkogel@cariari.ucr.ac.cr

Dr. William Eberhard
Escuela de Biologia
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica

Dr. Ana Mercedes Espinoza
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica
Phone: (506) 224 67 49
Fax: (506) 224 93 67
e-mail: amespino@cariari.ucr.ac.cr

Dr. Pedro León
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica
Phone: (506) 224 67 49
Fax: (506) 224 93 67

Dr. Jorge A. Lobo
Escuela de Biologia
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica

Dr. Gabriel Macaya
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica
Phone: (506) 224 67 49
Fax: (506) 224 93 67
e-mail: gmacaya@ns.cr // gmacaya@cariari.ucr.ac.cr

Dr. Edgardo Moreno
PIET
Escuela de Medicina Veterinaria
Universidad Nacional
Heredia
Costa Rica

Alejandro Leal
Instituto de Investigaciones en Salud
Universidad de Costa Rica

Ciudad Universitaria Rodrigo Facio
San José
Costa Rica

Marielos Mora
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica
Phone: (506) 224 67 49
Fax: (506) 224 93 67

Bernal Morera
Instituto de Investigaciones en Salud
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica

María Santos
Instituto de Investigaciones en Salud
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica

David Tobin
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa RicaPhone: (506) 224 67 49
Fax: (506) 224 93 67
e-mail: dtobin@cariari.ucr.ac.cr

Heidy Villalobos
CIBCM
Universidad de Costa Rica
Ciudad Universitaria Rodrigo Facio
San José
Costa Rica
Phone: (506) 224 67 49

Fax: (506) 224 93 67
e-mail: hvillalo@cariari.ucr.ac.cr

Sandro José de Souza
Ludwig Institute for Cancer Research
Rue Antonio Prudente 109-4 Andar
01509--010 Sao Paulo, SP
Brasil
Fax 55-11 270 7001
e-mail: ludwig@brfapesp.bitnet // ludwig@fpsp.fapesp.br

Gabriel Moreno-Hagelsieb
Instituto de Biotecnología
Apartado postal 510-3
Cuernavaca
Morelos 62271
México
Phone: 52-73-11 49 00
Fax 52-73-114901 / 52-73-114902 / 52-73-17 23 88
e-mail: moreno@pbr322.ceingebi.unam.mx

Liz Patricia Moreno Fonseca
Laboratorio de Bioquímica de Virus
Instituto de Biotecnología
Universidad Nacional de Colombia
Santa Fé de Bogotá

Colombia
Fax 2698164

Clara Beatriz Ocampo
Fundación CIDEIM
Av. 1a. # 3-03
Cali
Colombia
e-mail: cideimgujccol.bitnet

Eileen Riego
Centro de Ing. Genética y Biotecnología
La Habana
Cuba
Fax 53 7 21 80 70 / 53 7 33 60 08
e-mail: carmas@ingen.cigb.edu.cu

Dr. François Anthony
CATIE
Ap. 59
7170 Turrialba
Costa Rica
Fax: 506 556 15 33

**Title:** Gene functions involved in the biological evolution of prokaryotes

**Authors:** W. Arber

**Institution(s):** Dept. of Microbiology, Biozentrum, University of Basel, Klingelbergstr. 70, CH-4056 Basel

**Abstract:**

In the context of a general overview on molecular mechanisms of microbial evolution we will discuss several genetic systems known to either promote or restrain the generation of genetic variations. Particular attention will be given to functions involved in DNA rearrangements and DNA acquisition.

Sporadic actions of a variety of such systems affecting genetic stability result in an equilibrium of genetic plasticity tolerable to the wealth of microbial populations and allowing for an evolutionary progress needed for a steady adaptation under changing selective forces. Although these evolutionarily relevant biological functions are encoded on the genome of each individual, their actions are exerted to some degree randomly in rare individuals only and therefore become relevant at the population level.

**Title:** A molecular view of pinniped relationships with particular emphasis on the true seals

**Authors:** Úlfur Árnason, Anette Gullberg, Tina Ledje, Suzette Mouchaty

**Institution(s):** Division of Evolutionary Molecular Systematics, University of Lund, Sölvegatan 29, S-223 62 Lund, Sweden

**Abstract:**

Pinniped evolution poses several questions that have not been solved conclusively by classical approaches. The problems include whether the pinnipeds are mono- or diphyletic, the identification of their sister group(s) among terrestrial carnivores, and the relationship between the three extant families. In the present study we examine molecular affinities among the pinnipeds in relation to some chromosomal and zoogeographical data. Analysis of 18 complete sequences of the mitochondrial cytochrome b gene of Phocidae (true seals), Odobenidae (walruses), and Otariidae (sea lions and fur seals), identified the pinnipeds as monophyletic with Odobenidae and Otariidae on a common evolutionary branch. The analysis separated the evolutionary lineages of northern and southern phocids. Both lineages are distinct from the most ancestral phocid genus, Monachus (monk seals), represented by the Hawaiian monk seal. The inclusion of the Hawaiian monk seal in the subfamily Monachinae makes the subfamily paraphyletic. Among the northern phocids, the hooded seal (genus Cystophora, chromosome number 2n=34) is outgroup to the Phoca complex (2n=32). The comparison does not support a generic distinction of Halichoerus within the Phoca complex. The present data suggest that Cystophora and Phoca separated about 6.5 million years ago. Among the southern phocids the close molecular relationship of the Weddell and leopard seals relative to their morphological distinction exemplifies rapid adaptation to different ecological niches. This result stands in contrast to the limited morphological differentation relative to the pronounced molecular distinctions that may occur within the Phoca complex.

**Title:** MOLECULAR EVOLUTION OF RUMINANT, RODENT AND OTHER MAMMALIAN RIBONUCLEASES

**Authors:** Jaap J. Beintema *, Adriana Furia+, Heleen J. Breukelman*, Elena Confalone+, Maria Paola Sasso+ and Nonie van der Munnik*

**Institution(s):** *Department of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands, +Dipartimento di Chimica Organica e Biologica, Università di Napoli, Via Mezzocannone 16, 80134 Napoli, Italy

**Abstract:**

Pancreatic ribonuclease is one of the most intensively investigated digestive enzymes in mammals. However, it differs from other digestive enzymes in having highly different levels of expression in different species and in being found in other tissues and body fluids as well, indicating that the enzyme has other physiological functions besides its role in digestion. High levels of pancreatic ribonuclease occur in ruminants, in species with ruminant-like digestion and in several mammals with cecal digestion. These elevated levels are a response to the necessity of digesting large amounts of ribonucleic acid derived from the symbiotic microflora of the stomach or cecum of these herbivorous mammals.

A most parsimonious tree of primary structures of more than 40 ribonucleases, representing seven mammalian orders, is in reasonable agreement with current phylogenetic opinion about these taxa. More taxa from other orders will be added to the tree to contribute to the solution of open questions in mammalian phylogeny. Recently it has been proposed that the rodents are not monophyletic, but consist of two or more separate taxa, which branched off independently at different times from the mammalian stem (Graur and Li). Ribonuclease studies show strong evidence for the monophyly of six myomorph and for that of seven hystricognath rodent species, but not for that of the two taxa together. Sequence data from more rodent taxa are needed to resolve this question.

Three homologous bovine secretory ribonucleases have been isolated and sequenced. These are the enzymes from the pancreas, brain and seminal vesicles. Analysis of the sequences to derive the most parsimonious tree indicates that these ribonucleases are the products of two gene duplications that occurred in the ancestor of ruminants after divergence from the other artiodactyls. This conclusion has been confirmed by southern blot analysis of DNA of ruminants and other artiodactyls. Sequence studies of ribonuclease genes have already shown the presence of brain-type ribonucleases in the genome of giraffe, deer and *tragulus* (mouse deer) and of seminal vesicle-type ribonucleases in that of giraffe, deer and sheep. However, deletions in the coding regions and substitution of active-site res.dues indicate that the in these ruminants the seminal vesicle-type sequence represents a pseudo gene (see also accompanying contribution by Breukelman et al.).

**Title:** EVOLUTION OF ARTHROPOD HEMOCYANIN AND INSECT STORAGE PROTEINS (HEXAMERINS).

**Authors:** Jaap J. Beintema *, Wytze T. Stam †,Bart Hazes+ and Marten P. Smidt

**Institution(s):** Departments of *Biochemistry, †Marine Biology, and +Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

**Abstract:**

Crustacean and cheliceratan hemocyanins (oxygen transport proteins) and insect hexamerins (storage proteins) are homologous gene products, although the latter do not bind oxygen and do not possess the copper-binding histidines present in the hemocyanins. An alignment of 19 amino acid sequences of hemocyanin subunits and insect hexamerins was made, based on the conservation of elements of secondary structure observed in X-ray structures of two hemocyanin subunits. The alignment was analyzed using parsimony and neighbor-joining methods. Results provide strong indications for grouping together the sequences of the two crustacean hemocyanin subunits, the five cheliceratan hemocyanin subunits and the twelve insect hexamerins. Within the insect clade, four methionine-rich proteins, four arylphorins and two juvenile-hormone-suppressible proteins from *lepidoptera*, and two dipteran proteins form four separate groups. In the absence of an outgroup sequence, it is not possible to present information about the ancestral state from which these proteins are derived. Although this family of proteins clearly consists of homologous gene products, there remain striking differences in gene organisation and site of biosynthesis of the proteins within the cell. Because studies on 18S and 12S rRNA sequences indicate a rather close relationship between insects and crustaceans, we propose that hemocyanin is the ancestral arthropod protein, and that insect hexamerins lost their copper-binding capability after divergence of the insects from the crustaceans.

**Title:** MOLECULAR ADAPTATIONS TO A CHANGING ENVIRONMENT

**Authors:** G. BERNARDI and D. A. POWERS

**Institution(s):** HOPKINS MARINE STATION, STANFORD UNIVERSITY

**Abstract:**

Populations of *Fundulus heteroclitus* are distributed nearly continuously from Canada to Florida. The contact zone between the northern and the southern populations is typified by morphological, behavioral, physiological, mitochondrial DNA (mtDNA), and nuclear gene clines. The LDH-B locus, which encodes for the heart-type lactate dehydrogenase, varies clinally along a latitudinal gradient. Populations at the latitudinal extremes are virtually fixed for two different codominant allozymes ($LDH\text{-}B^a$ in southern populations, $LDH\text{-}B^b$ in northern populations). These LDH-B allozymes are kinetically different for several characteristics such as substrate affinities, reaction rates, heat stabilities, and inhibition constants. Fish with the northern allozyme swim faster than their southern counterparts at 10°C, and the metabolic rates of developing embryos are also different for the two forms. Taken together, these data strongly suggest that this locus is affected by environmental temperature in an adaptationally important manner.

**Title:** COMPARING NUCLEIC ACID SEQUENCES OF DIFFERENT RUMINANT SECRETORY RIBONUCLEASE GENES

**Authors:** Heleen J. Breukelman*, Nonie A.D.H. van der Munnik*, Adriana Furia+, Elena Confalone+, Maria Paola Sasso+ and Jaap J. Beintema*

**Institution(s):** *Department of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands, +Dipartimento di Chimica Organica e Biologica, Università di Napoli, Via Mezzocannone 16, 80134 Napoli, Italy

**Abstract:**

Parsimony analyses of mammalian ribonuclease sequences indicate that the presence of homologous enzymes in bovine pancreas, seminal vesicles and brain may be due to gene duplication events, which occurred after branching of the ancestors of ruminant species from other artiodactyls. This finding has been confirmed by Southern blot analyses of genomic DNA of several mammalian species. By PCR amplification of the coding region for 84 amino acid residues, three different RNase-genes were found in the genome of giraffe. One of the sequences differs at only 4 positions (5 %) from the bovine brain sequence. Two of these differences introduce additional Asn-X-Ser/Thr sequences, which indicates that the giraffe brain RNase may have much more carbohydrate than the bovine enzyme. A second one differs at 7 positions (8 %) from the bovine pancreas sequence and at one position from the previously determined protein sequence. The third sequence differs at 19 positions (23 %) from the bovine seminal sequence. The bovine intersubunit Cys-Cys sequence is replaced by a Phe-Cys sequence. Additional sequence information indicates that this latter sequence is a pseudogene. Recent studies of RNase-genes in hogdeer resulted in two different sequences, which show the highest identity to the brain and seminal ribonuclease-genes of bovine and giraffe. There is strong evidence that this seminal ribonuclease-gene is also a pseudogene. The chevrotain (*Tragulus*) belongs to the earliest diverged taxon of the ruminants, and study of its ribonuclease gene(s) may yield information about the origin of duplicated ribonuclease genes in ruminants. The first RNase-gene sequenced in chevrotain is most similar to those of bovine, giraffe and hogdeer brain ribonucleases. It differs at 20-30 % of the positions from the brain sequences of the three other ruminants.

**Title:** Evolutionary Selection Against Change in Many Alu Repeat Sequences Interspersed Through Primate Genomes

**Authors:** Roy J. Britten

**Institution(s):** California Institute of Technology

**Abstract:**

Mutations have been examined in the 1500 interspersed Alu repeats of human DNA that have been sequenced and are nearly full length. There is a set of particular changes at certain positions that rarely occur (termed suppressed changes) compared to the average of identical changes of identical nucleotides in the rest of the sequence. The suppressed changes occur in positions that are clustered together in what appear to be sites for protein binding. There is a good correlation of the suppression in different positions and therefore the joint probability of absence of mutation at many pairs of such positions is significantly higher than that expected at random. The suppression of mutation appears to result from selection that is not due to requirements for Alu sequence replication. The implication is that hundreds of thousands of Alu sequences have sequence-dependent functions in the genome that are selectively important for primates. In other words, many of them are not purely parasitic DNA sequence elements although their functions cannot yet be identified. In a few known cases Alu inserts have been adapted to function in the regulation of gene transcription.

**Title:** The Importance of the Gypsy/Ty3 Retrotransposons

**Authors:** Roy J. Britten

**Institution(s):** California Institute of Technology

**Abstract:**

Retrotransposons of the Gypsy/Ty3 class have been identified in herring and tunicate DNA as well as echinoderms. The matrix shows the percent AA divergence of the translations of a 525 nt segment of the pol gene coding region.

Spr2 **Strongylocentrotus purpuratus**, sea urchin
39 Por2 **Pisaster ochraceous**, starfish
46 45 Cpr1 **Clupea pallasi**, herring
51 52 41 Cir1 **Ciona intestinalis**, tunicate
52 49 49 55 Por1 **P ochraceous**, starfish
55 60 60 62 62 Spr3 **S purpuratus**, sea urchin
55 51 60 63 61 24 Tgr1* **Tripneustes gratilla**, sea urchin
56 61 60 63 58 26 27 Lvr1* **Lytechinus variegatus**, sea urchin
57 62 60 64 62 27 10 29 Spr1* **S purpuratus**, sea urchin
58 61 62 65 61 15 27 30 31 Spr4 **S purpuratus**, sea urchin
72 72 74 75 72 73 72 72 73 72 Por3 **P ochraceous**, starfish

The underlines separate groups of Gypsy/Ty3 retrotransposons, recognized on the basis of the divergence of this amino acid sequence. The upper group shows surprising similarity between elements from sea urchin, starfish, herring and a tunicate. Asterisks mark previously known elements. Support for the importance of the Gypsy/Ty3 elements includes: 1/Gypsy itself is a retrovirus in Drosophila melanogaster (Kim et al. 1994, PNAS 91:1285-1289, and work by Victor Corces group). 2/The genes in Gyspy/Ty3 class elements are in the same order as simpler retroviruses. 3/Of all known mobile elements this class has the closest sequence similarity to retroviruses. 4/Fish retroviruses are known and the occurrence of a Gypsy element in herring may shed light on the origin of retroviruses. 5/The gene control sites in the LTRs of inserted mobile elements may be of great importance to variation in evolution.

INTERNATIONAL SOCIETY
OF
MOLECULAR EVOLUTION

# WORKSHOP
# ON
# OPEN QUESTIONS IN MOLECULAR EVOLUTION

Guanacaste, Costa Rica, April 18-23, 1994

Organizers : Giorgio Bernardi, Morris Goodman, Barry Hall, Gabriel Macaya and Mary-Jane West-Eberhard

**Title:** The origins of the controversy about adaptive mutations.

**Authors:** John Cairns

**Institution(s):** Clinical Trial Service Unit, The Radcliffe Infirmary, Oxford OX2 6HE, Oxford, England

**Abstract:**

The causes of the mutations that occur in natural populations are of interest not merely in evolutionary biology but also in cancer research. It is now clear that most forms of cancer are due, in large part, to sequence changes in the genes that regulate cell behavior. What is not clear is whether these changes are due to exposure to mutagens. Certainly, the commonest lethal cancers are not more common in people who are deficient in the Uvr pathway for repair and would be expected to behave like Ames tester strains.

This conundrum triggered an investigation into the causes of spontaneous mutations in *Escherichia coli*. It turned out that when selection for novel traits is by selection procedures that are not lethal for non-mutants (i.e., are like the selection that must occur during carcinogenesis), advantageous mutations can arise that would not have arisen in the absence of selection; (this had been suspected by Schrödinger and Delbrück in the 1940s and 50s, but the problem had been allowed to lapse).

As the result of the work of many people during the past 5 years, we now know something about the mechanisms underlying mutation during selection. It is likely that there are several pathways. Certainly it is not difficult to imagine how a cell might produce sequence changes that match changes in its environment, using molecular biological mechanisms that are known to exist, and such a process or processes need not conflict with the Central Dogma of molecular biology.

**Title:** THE CHIMERIC NATURE OF NUCLEAR GENOMES AND THE ANTIQUITY OF INTRONS AS DEMONSTRATED BY THE GAPDH GENE SYSTEM

**Author:** Rüdiger Cerff

**Institution:** Institut für Genetik, Technische Universität Braunschweig, Spielmannstr. 7, D-38106 Braunschweig, Germany

**Abstract:**
Genes for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases of eukaryotes are descendants of a larger ancient gene family that existed in the progenote, the common ancestor of pro- and eukaryotes. During eukaryotic evolution both genes were transferred to the nucleus from the progenitors of present-day chloroplasts and mitochondria, respectively, thereby replacing preexisting GAPDH genes of the host cell. Of the ten intron positions found in chloroplast GAPDH genes five are precisely conserved in glycolytic genes of plants and animals. Three other intron positions differ by only three to eight nucleotides leaving only two introns for which no homologous counterparts were found (so far) in glycolytic GAPDH genes. These findings strongly suggest that introns were present in the oldest genes in agreement with the "exon theory of genes". They also indicate that, in addition to differential loss, "sliding" and "slippage" of introns occurred in GAPDH genes, in regions which do and which do not tolerate insertions/deletions, respectively.

Liaud M-F, Zhang DX, Cerff R (1990). Differential intron loss and endosymbiotic transfer of chloroplast glyceraldehyde-3-phosphate dehydrogenase genes to the nucleus. *Proc Natl Acad Sci USA* 87: 8918-8922.

Martin W, Brinkmann H, Savona C, Cerff R (1993). Evidence for a chimaeric nature of nuclear genomes: Eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenases. *Proc Natl Acad Sci USA* 90: 8692-8696 .

Kersanach R, Brinkmann H, Liaud MF, Zhang DX, Martin W, Cerff R (1994). Five identical intron positions in ancient duplicated genes of eubacterial origin. *Nature* 367:387-389.

Liaud MF, Valentin C, Martin W, Bouget FY, Kloareg B, Cerff R (1994). The evolutionary origin of red algae as deduced from the nuclear genes encoding cytosolic and chloroplast glyceraldehyde-3-phosphate dehydrognases from *Chondrus crispus*. *J Mol Evol* : in press .

**Title:** There's a lot more to molecular evolution than selection and drift.

**Authors:** Professor G. A. Dover

**Institution(s):** Department of Genetics, University of Leicester
University Road, Leicester LE1 7RH, U.K.

**Abstract:**

It is generally assumed that the distribution of mutations through space and time is due to natural selection or neutral drift, acting singly or in combination.

The simplicity of this either/or duality ignores extensive and frequent mechanisms of genomic turnover (transposition, unequal crossingover, slippage, gene conversion, and retrotranposition) which can affect the spread of mutations in populations through the generations (molecular drive). There is widespread evidence that such mechanisms affect the evolution of genes and genomes in major ways. Evidence will be presented reviewing these extensive data in multigene families, internally-repetitive genes and 'single-copy' genes.

Molecular evolution is an outcome of the three most frequent processes for changing the average genotypic composition of a population, (selection, drift and molecular drive). Evidence of subtle interactions between selection and molecular drive, as observed in the phenomenon of 'molecular coevolution' will be presented and discussed.

Recent references

Dover, G.A. (1992) Observing development through evolutionary eyes: a practical approach to molecular coevolution. Bioessays (Special Issue: Evolution and Development) 14:281-287.

Dover, G.A., Ruiz Linares, A. Bowen, T., and Hancock, J.M. (1993) The detection and quantification of concerted evolution and molecular drive. In Methods in Enzymology "Molecular Evolution: Producing the Biochemical Data" (eds. Zimmer, E.A., White, T.J., Cann, R.L. and Wilson A.C.) 224 525-541.

Dover, G.A. (1993) The evolution of genetic redundancy for advanced players. Current Opinions in Genetics and Development 3:902-910.

INTERNATIONAL SOCIETY
OF
MOLECULAR EVOLUTION

WORKSHOP
ON
OPEN QUESTIONS IN MOLECULAR EVOLUTION

Guanacaste, Costa Rica, April 18-23, 1994

Organizers: Giorgio Bernardi, Morris Goodman, Barry Hall, Gabriel Macaya and Mary-Jane West-Eberhard

**Title:** **HOVERGEN: a Database of Homologous Vertebrate Genes**

**Authors:** Laurent Duret and Manolo Gouy
**Institution:** Laboratoire de Biométrie, Génétique et Biologie des Populations
URA CNRS 243 - Université Claude Bernard Lyon 1
43, Bd du 11 Novembre 1918 - 69622 Villeurbanne cedex
FRANCE

**Abstract:** (Presented for the software demonstration session)

Comparison of homologous genes is a major step for many studies related to genome structure, function or evolution. Similarity search programs easily find genes homologous to a given sequence. However, only very tedious manual procedures allow the retrieval of all sets of homologous genes sequenced for a given set of species. Moreover, this search often generates errors due to the complexity of data to be managed simultaneously: phylogenetic trees, alignments, taxonomy, sequences and related information. HOVERGEN helps to solve these problems by integrating all this information. HOVERGEN corresponds to the nuclear vertebrate sequences subset of GenBank, with some data corrected, clarified, or completed, notably to address the problem of redundancy. Coding sequences have been classified in gene families. Protein multiple alignments and phylogenetic trees have been calculated for each family. Sequences and related information have been structured in an ACNUC database which allows complex selections. A graphical interface has been developed to visualize and edit trees. Genes are displayed in color, according to their taxonomy. Users have directly access to all information attached to sequences and to multiple alignments simply by clicking on genes. This graphical tool gives thus a rapid and simple access to all data necessary to interpret homology relationships between genes. HOVERGEN allows to easily select sets of homologous vertebrate genes, and thus is particularly useful for comparative sequence analysis, or molecular evolution studies.

INTERNATIONAL SOCIETY
OF
MOLECULAR EVOLUTION

WORKSHOP
ON
OPEN QUESTIONS IN MOLECULAR EVOLUTION

Guanacaste, Costa Rica, April 18-23, 1994

Organizers: Giorgio Bernardi, Morris Goodman, Barry Hall, Gabriel Macaya and Mary-Jane West-Eberhard

**Title:** **Long Genes are Avoided in G+C-rich Isochores**

**Authors:** Laurent Duret and Dominique Mouchiroud
**Institution:** Laboratoire de Biométrie, Génétique et Biologie des Populations
URA CNRS 243 - Université Claude Bernard Lyon 1
43, Bd du 11 Novembre 1918 - 69622 Villeurbanne cedex
FRANCE

**Abstract:** (Presented for the poster session)

We compared the exon/intron organization of vertebrate genes belonging to different isochores, as predicted by their C+G-content at third codon position. Genes coding for long proteins appear to be preferentially located in G+C-poor isochores. This bias is observed among mammalian, avian and even amphibian genes and is therefore likely to be common to all vertebrates. Our second finding is that introns from warm-blooded vertebrates tend to be longer as the G+C-content of isochores decreases. The ratio introns/coding sequences of human and chicken genes is in average 2.1-2.6 times higher in the G+C-poorest isochores (L1-L2) than in the G+C-richest (H3). We investigated the distribution of interspersed repeats in introns of human genes from different isochores. This analysis confirmed previous studies showing that L1 repeats are absent from H3, however, microsatellites and Alu sequences are found at almost equal frequency in introns from different isochores. Globally, the presence of repeated sequences does not account for the increased introns length in L1-L2 isochores. We also observed that genomic sequences from G+C-poor isochores are clearly under-represented in GenBank, probably because of experimental bias. Therefore, the observed differences in structure between genes from G+C-poor and G+C-rich isochores are probably under-estimated. The relationships between gene structure and global genome organization and evolution will be discussed.

# THE EVOLUTION OF SYNONYMOUS CODON BIAS IN ENTERIC BACTERIA

**Adam Eyre-Walker, Rutgers University, Piscataway, N.J. (USA)**

Many organisms preferentially use a subset of the synonymous codons available to them. In *Escherichia coli* the synonymous codons used are those which bind the commonest tRNA's with normal base pairing. This bias is thought to be caused by selection to maximise the elongation rate, or to minimise the costs of misincorporating the wrong amino acid. However the "optimal" codons are not used at every site in every gene. This is thought to be due to selection being counteracted by either mutation, genetic drift, or conflicting selection pressures. The first two possibilities are tested by considering the relationship between the divergence between *E. coli* and *Salmonella typhimurium* genes and the degree of codon bias. It is shown that genes with high codon bias evolve too slowly for either the mutation or genetic drift hypotheses to be correct. Some examples are given in which conflicting selection pressures have been identified at the start and the end of genes.

**Title:**

# Fortuitous Open Reading Frames, Base Compositional Persistence, and Evolution

**Authors:**

J.W. Fickett[1] and Roderic Guigó[2]

**Institution(s):**

[1]Los Alamos National Laboratory, Los Alamos, NM, USA and [2]University of Barcelona, Barcelona, Spain

**Abstract:**

After describing two results on the genomic distribution of Open Reading Frames (ORFs), one in yeast and the other in human, we will discuss possible implications for the evolution of genes.

In yeast *(Saccharomyces cerevisiae)* splicing is rare and a long (ORF) is often taken as evidence for the existence of a gene. We make this inference more quantitative by considering evidence from the databases and simulation. It is shown that many surprisingly long ORFs may be occurring by chance alone, and that statistical support for the biological function of many ORFs is weaker than has been thought.

In human, we have analyzed nearly 3000 sequences from clones randomly selected (RSCS, for Randomly Selected Clone Sequences) in the course of genome mapping, thus likely representing a fairly unbiased sample of the human genome. We have found that RSCS differ substantially from known DNA sequences in the region of genes. First, Open Reading Frame (ORF) occurrence and coding potential are significantly higher in most database sequences than in the RSCS. Surprisingly, even that fraction of the database sequences near genes but annotated as noncoding exhibits significantly higher coding potential and ORF occurrence than the RSCS. Second, RSCS are AT-rich, whereas the four nucleotides appear with almost equal probability in noncoding sequences from GenBank, and coding sequences are GC-rich.

**Title: Genomic regional structures and substitution rate in vertebrates**

**Authors: C. Gautier**

**Institution(s): Lab. BGBP, Univ Claude Bernard Lyon I (France)**

**Abstract:**

The concept of genomic neighbourhood takes an increasing importance in molecular evolution studies on vertebrates. The belonging to a band in a cytogenetic map or to an isochore in a compositional map implies some characteristics of the genome evolutionary process. Moreover new regional structures seem to be near emergence. Simultaneous management of time, particularly through homologous relationships, and space by the access to comparative genomic mapping is an important methodological challenge addressed by molecular evolution. We present here some computer science approaches that are under development in our team and their use to isochore evolution studies.

We shall discuss particularly the respective roles of natural selection, mutation and statistical artefact in G+C content variations that occur during vertebrate evolution (major and minor shifts) as well as in the substitution rate variability among genes. Among results that we shall discuss we can cite the following ones:

- correlation between silent and non silent divergence,
- estimation of silent divergence acceleration in the rodent (murids) lineage,
- relationship between minor shift and divergence rate.

**DNA Sequence Evidence on Primate Phylogeny**

Morris Goodman, Wayne State University School of Medicine

Cladistic analysis of extensive DNA sequence data provides a picture of primate phylogeny that is largely congruent with the traditional morphological picture but resolves some of the ambiguous features and suggests how those taxonomic groupings in primate classifications that are paraphyletic assemblages or grades could be replaced with monophyletic groups or clades. Molecular and morphological evidences are congruent in grouping Cercopithecoidea (Old World monkeys) and Hominoidea (apes and humans) into Catarrhini, Catarrhini and Platyrrhini (ceboids or New World monkeys) into Anthropoidea, Lemuriformes (Malagasy lemurs) and Lorisiformes (Asian and African lorises and galagos) into Strepsirhini, and Anthropoidea, Tarsioidea (Philippine and Indonesian tarsiers) and Strepsirhini into Primates. With regard to the problematic relationships of Tarsioidea, DNA sequences group it with Anthropoidea into Haplorhini. Moreover, the DNA evidence favors grouping Cheirogaleidae (mouse and dwarf lemurs) close to Lemuridae and Indridae within Lemuriformes in contrast to some morphological studies that favor placing cheirogaleids within Lorisiformes. The DNA studies are also producing a more clearly resolved picture of the phylogenetic systematics of the 16 extant New World monkey genera, and now provide quite significant evidence for sister group relationships within Hominoidea. These hominoid relationships may be defined by a cladistic classification that replaces the paraphyletic Pongidae (which excludes humans) with a monophyletic Hominidae (that includes all extant hominoids, divides this ape-human family into subfamilies Hylobatinae (gibbons) and Homininae, divides Homininae into tribes Pongini (orangutans) and Hominini, and divides Hominini into subtribes Gorillina (gorillae) and Hominina (humans and chimpanzees).

In addition to depicting cladistic relationships, the molecular picture of primate phylogeny reveals striking lineage differences in rates of noncoding DNA evolution. For example, lemuriform and hominoid rates tend to be slow whereas lorisiform and tarsioid rates tend to be fast. Among the 16 genera of ceboids there is some indication that small sized (and shorter generation time) marmosets and squirrel monkeys have faster noncoding DNA rates than the larger sized (and longer generation time) monkeys.

Supported by NSF grant DEB 9116098 and NIH grant HL 33940.

**Title:**
A protocol for higher-level phylogenetic reconstructions with an example concerning the ordinal phylogeny of eutherians

**Authors:**
Dan Graur

**Institution(s):** Department of Zoology, Tel Aviv University, Ramat Aviv 69978, Israel

**Abstract:**

In reconstructing higher-level phylogenetic trees based on molecular data, one has to deal with four possible trade-offs: (1) size of taxonomic sampling versus number of sequences used, (2) number of taxonomic units versus accuracy of phylogenetic reconstruction, (3) size of homologous sequences versus number of taxa, and (4) size of sequences versus appropriateness of outgroup. As a consequence, one usually ends up using only a small fraction of the available data. I present a protocol that attempts to maximize the amount of data used in a phylogenetic study. According to this protocol, for each group of orthologous sequences, we define subsets, each of each is assumed to be monophyletic. For each of the subsets, we reconstruct a phylogenetic tree, and by using an outgroup we choose the slowest evolving taxon from among the taxa within the subset to represent the entire subset. We are, therefore, left with n taxonomic units, each of which represents a large group of taxa. We, then, reconstruct 3 possible maximum parsimony trees for each of the (n!)/(n-3)!3! combinations of three ingroup taxa and the outgroup, and tabulate the results according to the number of informative sites that support a certain clustering of two ingroup taxa to the exclusion of the third. I present preliminary results in which this protocol is used to elucidate the ordinal phylogeny of eutherian mammals. Several taxonomic patterns emerge: (1) Ferungulata, Epitheria, Unguiculata, Anagalidia, Glires, and Rodentia are paraphyletic, (2) Archonta, Tethytheria and Paenungulata are monophyletic, (3) Artiodactyla is paraphyletic but becomes monophyletic if Cetacea is included within it, (4) Artiodactyla is more closely related to Carnivora than it is to Perissodactyla, and (5) it is unlikely that either Insectivora, Pholidota or Edentata represent the most ancient divergence within the Eutheria. I assume that by the time I shall complete this extremely tedious procedure (that unfortunately does not easily lend itself to automation), the phylogenetic tree of the eutherian mammals will either be solved (in a binarily satisfying manner), or only a few alternative trees will be left for further consideration.

Title: ORIGIN OF INTRONS: WERE THE ORIGINAL BLOCKS OF PROTEINS ENCODED BY EXONS?

Authors: Mitiko Gō

Institution(s): Department of Biology, Faculty of Science, Nagoya University, Nagoya 464-01, Japan

Abstract:

Were genes split from the outset (introns early) or did introns spread into eukaryotic genes as selfish elements (introns late)? If introns had a role in the primordial assembly of mini-genes and could speed up evolution, the remnant could be observed in the contemporary protein architecture and function. On the contrary, if introns are late and inserted as selfish elements, the introns had almost nothing to do with protein structure, function and evolution. Our study on a relationship between intron positions and protein modules supports "introns early" scheme. Globular proteins are generally decomposable into small compact pieces consisting of contiguous 10-40 amino acid residues which I call modules [1]. Module organization in various proteins is closely correlated with exon-intron structure of the genes encoding the proteins. This fact suggests that exon-shuffling which might occur during the biological evolution corresponds to module-shuffling in proteins and eukaryotes have retained their many of the original introns and prokaryotes have lost of them. Functional and structural roles of modules in the evolution of primitive proteins will also be discussed [2-4].

[1] Go, M., Nosaka, M.: Cold Spring Harbor Symp. Quant. Biol. **52**, 915-924 (1987).
[2] Noguti, T., Sakakibara, H., Go, M. : Proteins **16**, 357-363 (1993).
[3] Ikura, T. *et al.*: Proteins **16**, 341-356 (1993).
[4] Yanagawa, H. *et al.* : J. Biol. Chem. **268**, 5861-5865 (1993).

Title: DOMAIN EVOLUTION WITH SPECIAL REFERENCE TO KRINGLE STRUCTURES

Authors:

Takashi Gojobori and Kazuho Ikeo

Institution(s):

DNA Research Center, National Institute of Genetics, Mishima 411, Japan.

Abstract:

A group of serine proteases, which are involved in the cascade of blood coagulation and fibrinolysis, consist of several functional domains. A 'kringle' is one of such functional domains. It is composed of approximately 80 amino acids and three disulfide bonds. Recently, the kringle structures have been found in some other proteins which do not have the function as proteases. They include are apolipoprotein(a), hepatocyte growth factor (HGF), and a cell surface protein called ROR. ROR is a kind of tyrosine kinase. To study the evolutionary process and the origin of kringle structures, we constructed the phylogenetic trees by use of the nucleotide sequences of kringle structures, serine protease domains, and tyrosine kinase domains. The phylogenetic tree obtained shows that the ancestral gene of kringle structures appeared about 500 million years. Then, the kringle structure was duplicated and were inserted into other genes. Thus, our molecular evolutionary analysis shows that the domains such as kringle structures evolve dynamically and independently as evolutionary and functional units.

**Title:** Selection-induced mutations

**Authors:** Barry G. Hall

**Institution(s):** Biology Dept., University of Rochester, Rochester, NY 14627 USA

**Abstract:**

Selection-induced mutations are those mutations that occur as specific responses to environmental challenges, and that occur more often when they are selectively advantageous than when they are selectively neutral. Selection-induced mutations have been shown to occur in many loci in both bacteria and yeast, and are thus a general phenomenon. Their occurrence is usually detected by monitoring the continuous appearance of revertant papillae from a thin lawn of non-dividing cells during prolonged, intense selection when the reversion mutation allows the resumption of growth. The specificity of these mutations is demonstrated by showing, for instance, that a strain which requires both tryptophan and proline generates $Trp^+$ reversions during tryptophan starvation, but not during proline starvation.

I will discuss recent work showing that lesions in *uvrA*, *uvrB*, *uvrC*, or *uvrD* dramatically increase the selection-induced mutation rate, but do not increase the random mutation rate that is observed in growing cells; i.e. these genes can be considered as selection-induced specific mutator loci. Evidence will be presented to show that these lesions do not affect the specificity of selection-induced mutations. I will further show that cellular stress plays a direct role in production of these mutations.

Selection-induced mutations have previously been thought to occur only in non-dividing cells. I will present evidence that they also occur in very slowly dividing cells, and that multiple selection-induced mutations that arise under such conditions are a very good model for the mutational origins of tumors and cancers.

**Title:** CONSERVATION AND CONVERGENCE IN THE EVOLUTION OF THE CARBONIC ANHYDRASE GENE FAMILIES.

**Authors:** David Hewett-Emmett[+] and Richard E. Tashian[*]

**Institution(s):** [+]Genetics Center, Univ.Texas Hlth.Sci.Ctr., Houston, TX 77225; [*]Dept. of Human Genetics, Univ. Michigan Med. Sch., Ann Arbor, MI 48109, USA.

**Abstract:**

Two unrelated gene families are now known to encode gene products which catalyze the same reversible $CO_2$ reaction:

$$CO_2 + 2H_2O <-> HCO_3^- + H_3O^+$$

These functionally convergent non-homologous carbonic anhydrase (CA) gene families have been designated by us as the α-CAs and β-CAs.

Mammals have at least seven nuclear-encoded α-CA isozyme genes dispersed to five chromosomal regions. The gene products CA I, II, III and VII are cytosolic; CA IV is membrane-bound; CA V is mitochondrial; and CA VI is extracellular. An eighth related α-CA gene encodes a very conservative protein (CARP) which has two radical active-site replacements, while the green alga *Chlamydomonas reinhardtii* expresses two periplasmic α-CAs. In addition, an inactive CA structure is used as an extracellular (ligand-binding?) module by two receptor-bound protein tyrosine phosphatases (RPTPβ and RPTPγ) and as a cell-surface protein (D6) in the eukaryotic pox virus *Vaccinia*. By contrast, the β-CAs have been found only in the chloroplasts of higher plants and in certain bacteria (*E. coli* and *Synechococcus*). So far no organism expressing both an α- and β-CA has been identified.

Human and rodent (or sheep) sequence pairs are available for each α-CA and the average % protein divergence /mammalian lineage can be estimated:

| Human gene: | *CARP* | *RPTPγ* | *CA7* | *RPTPβ* | *CA3* | *CA2* | *CA1* | *CA5* | *CA6* | *CA4* |
|---|---|---|---|---|---|---|---|---|---|---|
| % Divergence: | 1.1 | 1.7 | 2.5 | 3.5 | 5.0 | 9.5 | 11.0 | 11.9 | 13.4 | 20.0 |

Although CARP, RPTPγ and RPTPβ are not active CAs, they and CA VII (whose function is still unclear) are more highly conserved in mammals than the genes encoding isozymes of known tissue expression and function (CA I-VI), strongly indicating that during evolution these conservative α-CA homologues have been recruited to fulfill important functions. In earlier studies we showed that while the active site of CA II was highly conserved relative to CA I and CA III, the exterior regions of the isozyme were not.

Comparable analyses of the β-CA family are not yet possible without an X-ray structure and knowledge of the active site.

Patterns of induced mutations drive genome evolution?

Gerard P. Holmquist

Beckman Reasearch Institute of the City of Hope Medical Center
Duarte, CA 91010 USA fax 818 358 7703

The grist of genome evolution is spontaneous germline mutations ($10^{-10}$ to $10^{-12}$/base pair per generation). That these are germline replication (polymerization-proofreading-mismatch repair) errors during the reading of undamaged DNA or induced mutations is unclear. The three steps of induced mutagenesis are 1. mutagens (body heat, reactive oxygen species, alkylating agents, ... ) induce DNA lesions (apurinic sites, oxidized bases, alkylated bases,...) 2. most lesions are repaired, 3. unrepaired lesions are often misread, differently by different polymerases, thus producing mutations. The rate of lesion generation/day in a mammalian cell is much higher, $10^{-5}$ / base, than replications errors. These lesions are repaired in 30 min-5 days and their equilibrium concentration, how many per genome are presented to DNA polymerase during S-phase of germline stem cells, is unknown. When replicated, the lesions do not disappear but can persist to cause more mutations during the next round of replication. Because genomes come to lesion equilibrium before a cell cycle is complete, the lesion-induced mutation rate is proportional to cell division rate; Y-linked genes having a disproportionately rapid molecular clock due to many male germ cell divisions does not signify that lesion-induced mutations are insignificant compared to replication error induced mutations. The three steps responsible for induced mutations show patterns which would, after fixation of neutral mutations, produce sequence patterns of mutation bias along the genome. Since the mutational load and substitutional cost for creating any genomic pattern is too great to be explained by natural selection. Finding relations between the patterns of mutation induction and evolutionary patterns of sequence fixation would be a reasonable proof that induced germline mutations contribute to sequence evolution.

Title: Mammalian Mitochondrial DNA Evolution: A Comparison of Rodents and Artiodactyls

Authors: Rodney L. Honeycutt, Michael A. Nedbal, Ron M. Adkins

Institution(s): Texas A&M University, Department of Wildlife & Fisheries Sciences, 210 Nagle Hall, College Station, Texas 77843

Abstract:

The mode and tempo of evolution in two mitochondrial genes, cytochrome *b* and cytochrome *c* oxidase subunit II, were examined in the mammal orders Artiodactyla and Rodentia. When analyzed using maximum parsimony and total substitutions, with no consideration of differential weighting, neither gene performed with a high degree of accuracy in terms of providing consistent phylogenetic results. The phylogenetic inconsistencies observed for both these genes may be the result of several factors including differences in the rate of nucleotide substitution among particular lineages (especially between orders), base composition bias, and different constraints and levels of homoplasy associated with first, second, and third codon positions. We discuss the implications of these findings to the molecular systematics of mammals, especially as they relate to recent hypotheses concerning the polyphyly of the order Rodentia, relationships among the Artiodactyla, and various interordinal relationships.

**Title:** The evolution of functionally novel proteins after gene duplication

**Authors:** Austin L. Hughes

**Institution(s):** The Pennsylvania State University

**Abstract:**

A widely cited model of the evolution of functionally novel proteins (here called the model of mutation during nonfunctionality or MDN model) holds that, after gene duplication, one gene copy is redundant and free to accumulate subsitutions at random. By chance some of these substitutions may suit the protein encoded by such a nonfunctional gene to a new function, which it can subsequently assume. Several lines of evidence contradict this hypothesis: (1) Comparison of expressed duplicate genes from the tetraploid frog *Xenopus laevis* indicates that such genes are subject to purifying selection and thus not free to accumulate substitutions at random. (2) In a number of multi-gene families, there is now evidence that functionally distinct proteins have arisen not as a result of chance fixation of neutral variants but rather as a result of positive Darwinian selection. (3) The phenomenon of gene sharing, in which a single gene encodes a protein having two distinct functions, shows that gene duplication is not a necessary pre-requisite to the evolution of a new protein function. A model for the evolution of new protein function is proposed under which a period of gene sharing ordinarily precedes the evolution of functionally distinct proteins. Gene duplication then permits each daughter gene to specialize for one of the functions of the ancestral gene. However, if the ancestral gene is not bifunctional, either of the following two outcomes is expected to follow gene duplication: (1) one copy will be silenced by a mutation preventing expression; or (2) if both copies continue to be expressed, both will be subject to purifying selection, since a high proportion of nonsynonymous mutations will have a completely or partially dominant deleterious effect.

Title: Complete form of PAB (X Y pseudoautosomal boundary) - like sequence exists near boundary of long-range G+C% mosaic domains in the human MHC locus

Authors:
T. Fukagawa, K. Sugaya, K. Matsumoto, A. Ando+, H. Inoko+, and T. Ikemura

Institution(s):
National Institute of Genetics, Mishima and + Tokai Univ., Kanagawa, Japan

Abstract:

We found human MHC (HLA) is a typical example of long-range G+C% mosaic structures (1-3). HLA is composed of classes I, III and II from telomere to centromere and contiguous classes I and III (a total of 3 Mb) correspond to evidently GC-rich domain and class II (1 Mb), to the domain with reduced G+C%. To precisely locate and characterize boundary of the Mb-level G+C% mosaic domains, cosmid walking from the most centromeric class III gene TN-X to class II and that from the most telomeric DRA to class III were conducted and contiguous clones totally covering the domain boundary were obtained bridging classes II and III. A sharp G+C% transition of the long-range G+C% mosaic domains corresponding to L <-> H isochores (and presumably G <-> R bands) could be precisely defined, and the following organization was revealed.

**(L isochore) PAB-like - LINE cluster - Dense Alu cluster (H isochore)**

Human sex chromosomes are divided into sex-specific and pseudoautosomal regions (PAR). Existence of PAR was deduced from observations of male meiosis, when sex chromosomes pair and form chiasmata. Interface between sex-specific and PAR is the pseudoautosomal boundary (PAB) that is the centromeric limit to recombination in PAR. Interestingly, complete form of PAB-like sequence (about 80% nucleotide identity) was found near the sharp G+C% transition point in HLA. A large number of PAB-like sequences in the human genome could be detected by hybridization with this HLA probe and the complete form of PAB-like sequence (about 600 nt) could be defined by sequencing several of them. (1) J. Mol. Biol. 203, 1 (1988). (2) Genomics 8, 207 (1990). (3) HLA1991 (Oxford Univ. Press) vol. 2, 125.

# Polymorphism in the genomic distribution of various transposable elements in an inbred line of *Drosophila melanogaster*

Nikolaj Junakovic (1), Claude Arnault and Christian Biémont (2)

(1) Centro Acidi Nucleici
Dipartimento di Genetica e Biologia Molecolare
Università "La Sapienza"
00185 Roma, Italy

(2) Laboratoire de Biométrie, Génétique et Biologie des Populations
URA 243, Université Lyon I
69622 Villeurbanne Cedex, France

The genomic distribution of transposable elements (TEs) has been analyzed in a Drosophila line that has been subjected to sib matings for 100 generations and subsequently maintained as a small mass for an additional 100 generations. 48 adult females, individually analyzed, have been sequentially tested by the Southern technique with probes homologous to 9 transposon families. The electrophoretic band patterns observed upon hybridization with mdg1, G, F, and 412 appear homogeneous. The patterns observed upon hybridization with copia, I, jockey, B104 (roo) appear polymorphic to various degrees.

*In situ* hybridization of salivary gland polytene chromosomes from larvae probed with copia and 412 reveals that the insertion copy number and polymorphism are consistent with those detected by the Southern technique.

The high degree of inbreeding of the line, and the comparison of the band patterns for different TEs on a same individual, make any explanation in terms of residual polymorphism and contamination highly unlikely. These results indicate that small mass breeding of a highly-inbred line allows genetic variability associated to TEs to be regenerated. Not all the TEs seem concerned in the same way suggesting high (copia) and very low (mdg1) transposition rates. In that respect, the I element appears more movable than previously estimated.

**Title:** New methods in molecular evolutionary comparisons

**Authors:** Samuel Karlin

**Institution(s):** Stanford University

**Abstract:**

Conventional methods of phylogeny reconstructions from sequence information employ only similarity among alignments of homologous genes or regions. Phylogenetic reconstructions based on parsimony, distance matrices, and maximum likelihood models require prior alignments and independence of neighboring positions. The latter requirement is certainly not valid due to strong local dependencies in DNA composition and/or due to secondary structure constraints when comparing rRNA genes. Another caveat concerns the genetic mosaicism of genes and genomes (especially for bacterial, fungal, and protist species and for viruses) resulting from substantial lateral transfer, transposition, rearrangement, and recombination events. Moreover, different evolutionary relationships often result for the same set of organisms based on analysis of different protein, gene or noncoding sequences.

For large genomic sequences, alignments of the sequences are generally not feasible. We present concepts and methods of both genomic and protein sequence comparisons suitable for establishing evolutionary relationships which do not depend on sequence alignments. Our methods for studies of molecular evolution at the DNA level are based on analysis of residuals of *relative abundance values* of di-. tri-, and tetranucleotides. Genomic sequences are compared with respect to: (i) similarities and differences of oligonucleotide compositional extremes; (ii) relative abundance distances within and between sequences; (iii) partial orderings among genomes of the vector of 16 dinucleotides relative to a set of vector standards; (iv) consistency of an individual sequence with respect to various consensus sequences. Examples will be given to: (a) molecular evolution of herpes viruses; (b) comparisons of genomic sequences within and between protists, fungi, and metazoa; (c) analysis of evolutionary relationships among certain prokaryotes.

Our methods appear to reflect contrasts in DNA sequence structures. The results for large aggregate DNA sequences for collections of species may and often do differ from "plylogenetic reconstructions" based on amino acid sequences of common proteins. It seems meaningful to accept that phylogenies based on DNA sequence structures, on conservation of amino acid sequences, on paleontological or on anatomical characters may reflect real differences in the evolution of these respective organism features.

**Title:** Patchiness of DNA and implications for sequence modeling

**Authors:** Samuel Karlin

**Institution(s):** Stanford University

**Abstract:**

Genomic DNA sequences display compositional heterogeneity on all scales, ranging from difference at the isochore level to local signals. Genomic distances based on dinucleotide relative abundance values of subsamples are significantly smaller within genomes for bacterial, protist, fungal, and protostome species and small but less for vertebrate species and always larger on average between species. This variation ("patchiness") and constancy limits attempts of modeling DNA sequences with simple stochastic processes. We show that in many cases the particular choice of model can greatly influence possible interpretations of sequence data. Four examples are discussed: (i) interpretation of purine/pyrimidine content fluctuations as long-range correlations or as genomic patchiness; (ii) expected oligonucleotide counts calculated under different model assumptions; (iii) expected gene size in random sequences; (iv) differences in distance measures based upon dinucleotide frequencies or residual dinucleotide frequencies. The examples illustrate that mathematical and statistical models can, at best, provide pointers for interacting with sequence data.

INTERNATIONAL SOCIETY
OF
MOLECULAR EVOLUTION

# WORKSHOP
# ON
# OPEN QUESTIONS IN MOLECULAR EVOLUTION

Guanacaste, Costa Rica, April 18-23, 1994

Organizers : Giorgio Bernardi, Morris Goodman, Barry Hall, Gabriel Macaya and Mary-Jane West-Eberhard

**Title:** Erasmus Darwin's 'Temple of Nature'

**Authors:** Richard D. Keynes

**Institution(s):** Physiological Laboratory, Cambridge

**Abstract:**

Enunciation of the theory of common descent is generally attributed to Charles Darwin, when he wrote in the final sentence of *On the Origin of Species*: 'There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that from so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved'. Even if there are still difficulties in fully accepting the theory of natural selection, the existence of the same genetic code for all forms of life has proved the truth of common descent. So the question arises as to where this vital concept originated. The answer appears to be that it had been put forward posthumously in 1803 by Charles's grandfather, Erasmus, in rhyming couplets:

> "Organic Life beneath the shoreless waves
> Was born and nurs'd in Ocean's pearly caves;
> First forms minute, unseen by spheric glass,
> Move on the mud, or pierce the watery mass;
> These, as successive generations bloom,
> New powers acquire, and larger limbs assume;
> Whence countless groups of vegetation spring,
> And breathing realms of fin, and feet, and wing.

Rather few scientific papers are written in verse these days, and Erasmus Darwin's last poem, *The Temple of Nature; or, the Origin of Society*, repays a fresh reading.

Title: Glutamine synthetase gene evolution in bacteria

**Authors:** Cecilia Lanave, Graziano Pesole, Carmela Gissi and Cecilia Saccone

**Institution(s):** CSMME-CNR, Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, via Orabona, 4, 70125 Bari, Italy.

**Abstract:**

In order to extend our previous studies on the evolution of Glutamine Synthetase (GS) enzymes, that previously revealed to behave as a good molecular clock (Pesole et al. 1991), we have analysed all the GS sequences of bacteria available in the database.
The phylogenetic analyses of GS genes, carried out by applying our Stationay Markov Model (Lanave et al 1984) to the second codon positions of 33 bacterial sequences and 3 eukaryotic sequences, show:

- The Gram-negative bacteria cluster together and appear more closely related to high GC gram positives; low GC gram positive bacteria are more closely related to Euryarchaeota-Archaebacteria. Interestingly, both gram negative and high GC gram positive bacterial GS are regulated by the adenylytation/deadenylilation system.

- The distance between GS1 and GS2 isoforms, the former found in Eubacteria and Archaebacteria and the latter in Eukaryotes and in some bacteria, dates 2000±400 m.y. This divergence time might correspond to a possible duplication of the GS gene or alternatively to the split between Prokaryotes and Eukaryotes

- Archaebacteria appear more closely related to Eubacteria than to Eukaryotes. The two phyla of Archaebacteria, namely *Crenarchaeota* and *Euryarchaeota* appear very distantly related.

**Acknowledgments.**
This work was partially financed by MURST, Italy, Progetto Finalizzato Ingegneria Genetica, CNR, Italy and Progetto Finalizzato Biotecnologie e Biostrumentazione, CNR, Italy.

. Pesole, G., M. P. Bozzetti, C. Lanave, G. Preparata and C. Saccone (1991). "Glutamine synthetase gene evolution: a good molecular clock." Proc. Natl. Acad. Sci. USA 88: 522-526.

. Lanave, C., G. Preparata, C. Saccone and G. Serio (1984). " A new method for calculating evolutionary substitution rates." J Mol Evol 20: 86-93.

**Title:** Glutamine synthetase gene evolution in bacteria

**Authors:** Cecilia Lanave, Graziano Pesole, Carmela Gissi and Cecilia Saccone

**Institution(s):** CSMME-CNR, Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, via Orabona, 4, 70125 Bari, Italy.

**Abstract:**

In order to extend our previous studies on the evolution of Glutamine Synthetase (GS) enzymes, that previously revealed to behave as a good molecular clock (Pesole et al. 1991), we have analysed all the GS sequences of bacteria available in the database.
The phylogenetic analyses of GS genes, carried out by applying our Stationay Markov Model (Lanave et al 1984) to the second codon positions of 33 bacterial sequences and 3 eukaryotic sequences, show:

- The Gram-negative bacteria cluster together and appear more closely related to high GC gram positives; low GC gram positive bacteria are more closely related to Euryarchaeota-Archaebacteria. Interestingly, both gram negative and high GC gram positive bacterial GS are regulated by the adenylytation/deadenylilation system.

The distance between GS1 and GS2 isoforms, the former found in Eubacteria and Archaebacteria and the latter in Eukaryotes and in some bacteria, dates 2500±400 My. This divergence time might correspond to a possible duplication of the GS gene or alternatively to the split between Prokaryotes and Eukaryotes

- Archaebacteria appear more closely related to Eubacteria than to Eukaryotes. The two phyla of Archaebacteria, namely *Crenarchaeota* and *Euryarchaeota* appear very distantly related.

**Acknowledgments.**
This work was partially financed by MURST, Italy, Progetto Finalizzato Ingegneria Genetica, CNR, Italy and Progetto Finalizzato Biotecnologie e Biostrumentazione, CNR, Italy.

. Pesole, G., M. P. Bozzetti, C. Lanave, G. Preparata and C. Saccone (1991). "Glutamine synthetase gene evolution: a good molecular clock." Proc. Natl. Acad. Sci. USA 88: 522-526.
. Lanave, C., G. Preparata, C. Saccone and G. Serio (1984). " A new method for calculating evolutionary substitution rates." J Mol Evol 20. 86-93.

# Genome evolution in the RNA world and the emergence of DNA

Pedro Leon, CIBCM, Universidad de Costa Rica, San José

*There was an early period of cellular evolution in which RNA/proteins performed genetic and metabolic functions, before DNA was invented. Catalytic RNA metabolism gave way to protein-based metabolism, as ribosomes and messenger RNAs emerged, and the synthesis of polypeptides, as "statistical proteins" at first became established. The complicated issue of the origin of RNA itself, is not addressed at all in this model.

*dNTPs are derived from NTPs in evolutionary time, as reduction reactions of the 2'-OH, a reaction universally catalyzed by ribonucleotide reductases in all living cells. Consequently, the invention of DNA required at least a ribonucleotide reductase and a reverse transcriptase (RT).

*It is assumed that natural selection operated to favor cell lineages with adaptive features such as the acquisition of a new metabolic pathway (by endosymbiotic capture), or of a new enzyme function in the right environment (by gene transfection, viral transformation, etc.). Those genomes, or more appropriately, those ribotypes that did not carry mutated (specially lethal) functions, left more descendants. Faster and more precise information transfer to the next generation resulted in increasing number of offspring with fewer lethal or deleterious mutations. Improvements in procesivity and fidelity of polymerases; emergence of editing and maturing functions; improvements of ribosome fidelity and growth of the genome were all favored.

*Compartments were required to solve the dual role played by RNA at the onset, as genomic and catalytic macromolecule. Genomic functions included the replication of dsRNA molecules, metabolic functions involved synthesis of (+)RNA sequences, that were "matured" into a single-stranded "active" conformation. The simplest compartment is the membrane (as in bacterial mesosomes), which might have served as compartment for the genetic functions.

*During the transition from the RNA to the DNA world, it seems necessary, that for a while both systems coexisted until the appropriate adaptations had taken place. It is not clear why cells would produce dNTPs, before RT appeared? Explaining the appearance of dNTPs, seems harder than the origin of the RT. Quite likely, this type of enzyme "had been around" as "mutant" of the ancestral polymerases, capable to using dNTPs, if these became available. RNA directed metabolism would have to continue until cells were able to efficiently transcribe (+)RNA sequences from the DNA genome. The question remains as to how dNTPs and NTPs coexisted without altering RNA metabolism, with the incorporation of dNTPs instead of NTPs into catalytic and coding RNAs. This is one of the features considered by the model.

In the RNA world there were probably two kinds of roles played by RNA: genetic and catalytic roles. Because of their greater stability the molecules involved in genetic functions were double-stranded while single-stranded molecules, with a huge number of potential conformations, were the catalytic (and coding) ones. Auto-splicing was one of the earliest activities able to prepare RNA for its metabolic roles, and may have been favored in molecules that ligated into the genome.

It is surmised that a simple RNA duplication mechanism, where both strands were equally replicated, gave way to lineages where the catalytic strand was selectively expressed. A (-)RNA polymerase -presumably recognizing a feature of the 3'-end of the (+)RNA- would synthesize the (-)RNA molecules, while a (+) RNA polymerase copied the other strand. To replicate the (-) strand extensive repair of the (+) strand would be required, with inhibition of splicing and other maturation events that prepared the (+) RNA for metabolic functions. It is not clear if this event was spatially segregated (in a compartment) or temporarily segregated (as in a cell cycle stage), but both are possible choices at present.

Another presumed feature of the RNA world which set the stage for the emergence of DNA, was the ubiquitous presence of RNAses, in addition to the intrinsic reactivity of RNA and to the potential danger of maturing reactions. We presume that the basis of interspecific (or interlineage) competition derives from being able to protect genetic material and destroy that of others for substrates ; ultimately what many viruses do. RNAses probably were the stuff of territorial and predatory war by cell and pathogens of that era. Consequently, for intrinsic and extrinsic reasons, selection would favor cells that had a resilient variant of RNA that would survive RNase attack and still preserve the information in the RNA string. We suggest that by mutating the genomic (-) polymerases, a RT emerged, that was distinct from the (-)RNA transcriptases, in that it could use dNTPs available. This resulted in the synthesis of a (-) DNA strand and the establishment of a (-) stranded genome that could increase in size by ligation of negative molecules. The long range arrangement of the DNA genome, with introns, exons and other sequences may have been established then.

Perhaps a heteroduplex genome arouse during the transition (Leon, Gutierrez, Warnes and Macaya, unpublished). The (+)RNA transcriptases would eventually adapt to transcribing, the heteroduplex genome, making the RNA genome completely dispensable. Double-stranded replication was the last invention in the series and even in modern cells requires a short RNA primer, synthesized by a primase. It is conceivable that this reflects the ancestral transition from the heteroduplex to the dsDNA genome.

Title: A Model for the Origin of $GC$-Rich Isochores and the Correlation of Mutation Rate with $GC$ Content.

Authors: Wen-Hsiung Li and Xun Gu

Institution: Center for Demographic and Population Genetics
University of Texas, P.O.Box 20334 Houston, TX 77225, USA

**Abstract:**

To study the origin of $GC$-rich isochores and the relationships among the mutation rate, the $G+C$ content of the sequence, and the $G+C$ proportion in the nucleotide precursor pool, a model has been developed based on the biochemical kinetics of DNA replication and mutagenesis, including misincorporation and correction. Under the normal physiological conditions of mammalian germ cells, we have obtained the following conclusions: (1) the equilibrium $G+C$ content in a sequence is approximately equal to the $G+C$ proportion in the nucleotide precursor pool and is independent of the next-nucleotide effect, which affects the correction of misincorporation during DNA replication; (2) an inverted-V-shaped distribution of mutation rates with respect to $G+C$ contents is predicted, if the next-nucleotide effect is week; (3) the distribution becomes flatter but the peak at 50% $GC$ is still observed (*i.e.*, inverted-U-shaped ), if the next-nucleotide effect is intermediate; and (4) the peak disappears if the next-nucleotide effect is strong. Our results support the hypothesis that changes in the relative concentrations of nucleotide precursors can cause variations among genes both in mutation rate and in $G+C$ content and that compositional isochores (DNA segments with a homogeneous $G+C$ content) can arise in a genome due to differences in replication times of DNA segments.

**The Recent Origin and Punctuated Evolution of Nuclear Spliceosomal Introns**

John M. Logsdon Jr. and Jeffrey D. Palmer, Dept. of Biology, Indiana University

The Introns-early theory (also known as the "exon theory of genes") posits that nuclear spliceosomal introns were present in large numbers in the common ancestor of all life and were used in the assembly of genes from small exons. The Introns-late theory maintains that nuclear introns are recent, arising late in eukaryotic evolution only. Although the Introns-early scenario is widely accepted, current data do *not* support this position when viewed from a phylogenetic perspective. We have undertaken a survey of the literature on gene structure in eukaryotes, focusing most comprehensively on protists. These data, when projected on a eukaryotic phylogenetic tree, support the contention that nuclear introns have been inserted in eukaryotic genes and that their origins are well within the eukaryotic lineage. Preliminary analyses of intron distribution among the late arising, multicellular eukaryotic lineages comprising animals, fungi and plants point to a highly episodic mode of intron evolution, including periods of both rapid intron gain and loss as well as virtual stasis. In particular, two periods of intron stasis can be discerned, one within vertebrates and the other within seed plants. We have also been studying the TPI gene, which encodes the glycolytic enzyme triose phosphate isomerase and has been a cornerstone of the Introns-early hypothesis. A summation of all of the known intron positions in the gene is purported to reflect the ancestral structure of the gene and the intron/exon structure is thought to strongly coincide with the TPI protein domain structure. We have sequenced TPI genes from three species, *Coprinus cinereus* (mushroom), *Caenorhabdites elegans* (nematode) and *Chlamydomonas reinhardtii* (green alga), in order to test predictions of the Introns-early and Introns-late theories. Each new intron position found not only refutes the belief that all TPI intron positions have been found, but also weakens the assertion that exons encode protein domains in TPI. Five novel intron positions have been found in these three genes; all fall within recognized protein domains. We interpret these results as evidence in favor of the recent acquisition of these TPI introns, providing support for Introns-late.

**Title:** Differences in DNA repair, G+C-content and expresion level among *E. coli* genes

**Authors:**

G Gutiérrez, J Casadesús, JL Oliver[1] and A Marín

**Institution(s):**

Departamento de Genética, Universidad de Sevilla (Spain)
[1]Departamento de Genética, Universidad de Granada (Spain)

**Abstract:**

*E. coli* DNA segments that contain a high frequency of the tetranucleotide CTAG are also rich in the tetramers CTTG, CCTA, CCAA, TTGG, TAGG and CAAG (Group-I tetramers). Conversely, DNA segments lacking CTAG are rich in the tetranucleotides CCTG, CCAG, CTGG and CAGG (Group-II tetramers). These two samples (CTAG-rich and -poor) differ also in G+C-content, frequencies of di- to pentanucleotides, codon usage of the genes harbored therein and amino acid composition of the encoded proteins. Group-I tetramers have in common that they are depleted by very short path mismatch repair (VSP), while Group-II tetramers are favored by VSP activity. The VSP system repairs G:T mismatches to G:C, thereby increasing the overall G+C-content; for this reason the CTAG-rich sample has lower G+C content than the CTAG-poor sample. This compositional heterogeneity can be tentatively explained by a low level of VSP activity on the CTAG-rich sample. A negative correlation is found between the frequency ofGroup-I tetramers and the level of gene expression, as measured by the Codon Adaptation Index. Thus a possible link between the rate of VSP activity and the level of gene expression can be considered.

**Title:** DNA SEQUENCE ANALYSIS SHOWS DISTINCTIVE EVOLUTIVE PATHWAYS BETWEEN EUKARYOTA AND PROKARYOTA

**Authors:** Pedro Miramontes and Germinal Cocho

**Institution(s)** Departement de biochimie; U. de Montréal, Québec and Instituto de Física, UNAM, México.

**Abstract:**

We propose an index of DNA homogeneity (IDH) based on a binary distribution model which quantifies structural and thermodynamic aggregates present in DNA primary structures. Extensive analysis of sequence databases with the IDH uncovers significant constraints on DNA sequence other than those derived from codon usage or protein function. In addition, this index clearly distinguishes between organisms of the eukaryota and prokaryota supereigns and places them in disjoint domains of DNA sequence space. The results support the phenotypic character of the genome.

**Title:** **A gene component of synonymous substitutions**

**Authors:** Dominique Mouchiroud
**Institution:** Laboratoire de Biométrie, Génétique et Biologie des Populations
URA CNRS 243 - Université Claude Bernard Lyon 1
43, Bd du 11 Novembre 1918 - 69622 Villeurbanne cedex
FRANCE

**Abstract:** (Presented for the poster session)

Among mammals, the frequencies of synonymous substitutions between homologous genes show a large variability as estimated from a number of pairwise comparisons (Bernardi et al 1993). The analysis of homologous genes from four species (man, calf, rat and mouse) has indicated that 28% of this variability is gene-specific. Moreover, as expected from previous work (Bernardi et al 1993), this synonymous substitution component is independent of the GC level of isochore class where the gene is located. A significant correlation was also observed between synonymous and nonsynonymous substitution frequencies for two independent pairwise comparisons. Thus, this correlation is not due to properties that depend upon the lineages considered but depends of the gene specific component. Elimination of neighbouring substitutions inside coding sequences, does not greatly modify the correlation. There is no significant difference between the synonymous substitution frequencies of housekeeping and tissue specific genes.Thus , the independence of the synonymous substitution variability from the expression of the coded protein or from the isochore class argues against different selective pressure or variation of the mutation rate as the cause of this gene property. We propose that a novel component of mutation rate variation takes place along the genome . To verify this hypothesis, the study of the synonymous subsitution variability in non coding sequences was undertaken. Preliminary results obtain with a low gene sample are consistent with the regional aspect of the mutation rate.

# Compositional Patterns of Nuclear Encoded Genes of *Trypanosoma brucei* and *T. cruzi*.

We have investigated the compositional distributions of exons and their different codon positions of the nuclear encoded genes of the unicellular parasites *Trypanosoma brucei* and *T. cruzi*. Very large differences between the two species were found in all the properties analyzed. The most striking differences concern the compositional distributions of third codon positions and the large divergence of GC levels in third codon position for homologous genes encoding proteins that are highly conserved in their amino acids sequences. Further, the compositional distribution of third codon positions in the two species suggest that these genomes are compositionally compartmentalized, a feature previously detected by the fractionation of DNA from *T. brucei* and *T. equiperdum*. To investigate whether this compartmentalization is accompained by often different properties of coding sequences, we have analyzed and compared the compositional compartments in the dinucleotide frequencies and amino acid usage of genes divided and pooled according to GC levels in third codon positions. In all cases, compartments displayed remarkable differences, as is the case in highly compartmentalized genomes, like those of warm-blooded vertebrates.

Héctor Musto. Sección Bioquímica, Facultad de Ciencias, Tristán Narvaja 1674, Montevideo 11200, Uruguay; and Depto. de Genética, Facultad de Medicina, Montevideo, Uruguay.

**Title:** ACTIVE SITES OF LIGANDS AND THEIR RECEPTORS ARE MADE OF COMMON PEPTIDES THAT ARE ALSO FOUND ELSEWHERE

**Authors:** Susumu Ohno

**Institution(s):** Beckman Research Institute of the City of Hope
1450 E. Duarte Road, Duarte, California 91010-0269 USA

**Abstract:**

The simultaneous emergence in evolution of a ligand and its receptor might have entailed their active sites to be drawn from the pool of common oligopeptides. This was tested on the principal components of cell-matrix interaction: The RGD (Arg-Gly-Asp) site of matrix proteins and the EKKD (Glu-Lys-Lys-Asp) site of the integrin cell surface receptor. In the 35 diverse proteins scrutinized that totalled 14,806 residues, there were 104 Arg-Gly dipeptides. Most common of the tripeptides beginning with Arg-Gly were Arg-Gly-Leu, Arg-Gly-Gly and Arg-Gly-Asp; each being found in 10 copies. RGD tripeptide being one of the commonest, the fortuitous presence of RGD site was noted in a protease, a pituitary hormone and a viral structural protein. The 35 proteins also contained 121 Lys-Lys dipeptides. Of the tetrapeptides centered by Lys-Lys, the commonest was Lys-Lys-Lys-Lys in 4 copies. The second in rank were Val-Lys-Lys-Leu and Glu-Lys-Lys-Asp; each in 3 copies. The fortuitous presence of EKKD site was noted in three proteins; an intracellular motor protein, a pituitary hormone and a protein of the cerebrospinal fluid. In many instances, protein-protein interaction between the fortuitously present active sites appears to bring about deleterious consequences. Occasionally, however, the fortuitous active site appears to confer a new function to a protein bearing it.

**Title:** Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory

**Authors:**

Tomoko Ohta

**Institution(s):**

National Institute of Genetics, Mishima 411, Japan

**Abstract:**

In order to test the predictions of the nearly neutral theory, the average and variance of synonymous and nonsynonymous substitutions were examined for the mammalian star phylogenies composed of rodentia, artiodaclyla and primates. The sequences of 49 single copy genes were used. The prediction of the nearly neutral theory that the generation-time effect is more pronounced for synonymous substitutions than for nonsynonymous substitutions holds for the average pattern of 49 genes, although there are some notable exceptions. Another important measure for testing the nearly neutral theory is the variation of evolutionary rate. The dispersion index, that is the ratio of the variance to the mean, turns out to be positively correlated with the average substitution number. Through weighting by the lineage effect, the correlation almost disappears for the nonsynonymous substitutions, but not quite so for the synonymous substitutions. After weighting, the dispersion indices of both synonymous and nonsynonymous substitutions are still larger than the value expected under the simple Poisson process. The results indicate that both the systematic bias in evolutionary rate among the lineages and the episodic type of rate variation are contributing to the large variance. The former is more significant to synonymous substitutions than to nonsynonymous substitutions. The isochore evolution may be similar to the synonymous substitutions. The rate and pattern found here are in accord with the nearly neutral theory, such that the relative contributions of drift and selection differ between the two types of substitutions.

**Title: On the possibility of horizontal transmission of SINEs**

**Authors: Norihiro Okada**

**Institiution(s): Faculty of Bioscience and Biotechnology,**
**Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku,**
**Yokohama 227, Japan Tel & Fax 81-45-923-1136**

**Abstract:**

Our group proposed that SINEs are derived from a "strong stop DNA" with a primer tRNA that is an intermediate in the reverse transcription of certain retroviruses (1). This model suggests that a certain group of SINEs may have been generated by horizontal transmission (1).

Recently, our group obtained two experimental data which may suggest horizontal transmission of SINEs. The first finding is as follows. In the subfamily Coregoninae (whitefishes), a new family of SINEs was discovered. Members of this family (designated SRW family) are almost identical to those of the Sma I family, that is restricted to chum salmon and pink salmon in the subfamily Salmoninae. The similarity between these two families is 97%. In contrast, the genetic distance deduced from synonymous changes in a growth hormone gene of chum salmon (Salmoninae) and C. lavaretus (Coregoninae) is about 10%. The result suggests that the Sma I family may have been generated by horizontal transmission of the SRW family, if SINEs have no functions and behave like pseudogenes.

Second, fourteen members of Hpa I subfamilies that had been specifically amplified in particular species were isolated from genomic libraries established for chum salmon, kokanee, coho salmon, masu salmon and steelhead. Alignment of the sequences of these 14 members, three of which were previously demonstrated to have been amplified specifically in certain lineages (2), revealed the presence of five subfamilies with particular diagnostic nucleotides. The amplification of members of the same subfamily in different salmonid lineages and the amplification of members of the different subfamilies in the same salmonid lineage suggest that multiple dispersed loci were responsible for amplification. However, these results are also interpreted on the asumption that SINEs can be transmitted horizontally between species (3).

(1) Ohshima et al., (1993) Proc. Natl. Acad. Sci. USA 90, 6260-6264.
(2) Murata et al., (1993) Proc. Natl. Acad. Sci. USA 90, 6995-6999.
(3) Takasaki et al., submitted to PNAS

**Title:** A model to generate nucleotide sequences with long-range correlations

**Authors:** José L. Oliver[1] and Antonio Marín[2]

**Institution(s):** [1]Dpto. de Genética. Facultad de Ciencias. Universidad de Granada E-18071-Granada (Spain)
[2]Dpto. de Genética. Facultad de Biología. Universidad de Sevilla E-41080-Sevilla (Spain)

**Abstract:**

Two mechanisms have been proposed to explain the long-range correlations recently found in nucleotide sequences: 1) repetitive patterns and 2) compositional patchiness. Here we develop a model simulating the expansion of a family composed of short interspersed repeats within a random sequence of a given nucleotide composition. A probability for nucleotide substitution ('lateral' point mutation) is assigned at each step. The length of both spacers and repeat-units also have a chance to vary during the expansion process. In this way, the initial repeat unit is laterally propagated (5' --> 3'), incorporating random mutational changes at each step; such changes are inherited by the repeat units from subsequent steps. The model proves capable of generating repetitive sequences with long-range correlations, provided the appropriate values for the different parameters are set. In addition, it also allows the generation of repetitive sequences without compositional patchiness but with long-range correlations, thus suggesting that patchiness and long-distance correlations may be uncoupled.

Title: The use of a metazoan mt database (MmtDB) for studying the evolution of the D-loop region in humans

**Authors:** Graziano Pesole, Marcella Attimonelli, Daniela Calò and Cecilia Saccone

**Institution(s):** Dipartimento. di Biochimica e Biologia Molecolare, Università di Bari, via Orabona. 4, 70125 Bari, Italy.

**Abstract:**

The study of genetic variability can be a very powerful tool both for assessing the structure-function model of gene sequences and for reconstructing detailed histories of genes and of organism populations. To this aim, mitochondrial DNA has revealed to be a powerful tool (Pesole et al., 1992). To unravel the genealogical history of our species and study the molecular basis of mitochondrial pathologies many sequences of human mt DNA have been produced, particularly in the hypervariable region of the mt D-loop.

Against this scenario, the need to generate a mtDNA specialized database has proved to be compelling. In our laboratory we have constructed a specialized database of metazoan mt sequences (MmtDB) whose structure is described in this poster. We also present the results obtained analyzing the human D-loop sequence data from MmtDB on the pattern and extent of nucleotide variation in various individuals belonging to several human populations.

**Acknowledgments.**
This work was partially financed by MURST, Italy, Progetto Finalizzato Ingegneria Genetica, CNR, Italy and Progetto Finalizzato Biotecnologie e Biostrumentazione, CNR, Italy.

## Bibliography

Pesole G., Sbisà E, Preparata G and Saccone C. (1992) The evolution of the mitochondrial D-loop region and the origin of modern man. Mol. Biol. Evol. 9(4): 587-598.

# Flying DNA

Title:

John D. Pettigrew

Authors:

Institution(s): Vision Touch and Hearing Research Centre
The University of Queensland 4072 Australia

Abstract:

A vigorous controversy involves the phylogeny of bats. The traditional, majority view holds that megabats (Old World Tropical fruit bats) and microbats (cosmopolitan, insectivorous, echolocating bats) are monophyletic. An opposing view, the "flying primate hypothesis", maintains that megabats are closer to primates than they are to microbats and, therefore, that mammalian flight evolved twice. The bodies of evidence in favour of the two viewpoints were somewhat balanced until 1992, with a host of shared, specialised adaptations in the limbs of megabats and microbats supporting "bat monophyly", and an equally impressive set of shared specialisations in the brains of megabats and primates supporting "flying primates". The availability of DNA sequences from both kinds of bats tipped the balance of opinion in 1992. All studies found more shared substitutions in common between megabat and microbat DNA than between either kind of bat DNA and other mammals. Parsimony analyses showed that monophyly generally gave a tree with shorter steps.

Despite the triumphant tone adopted by these DNA studies, there are a number of problems that cast doubt on the conclusions reached. Four of the studies used non-coding sequences whose alignment was anything but straightforward, particularly when the second problem of base composition effects is taken into account. The substitutions claimed in support of bat monophyly, in all six studies, had a 4:1 bias to AT, the same bias shown by bat DNA. In other words the DNA sequencing studies show what has been known for over 20 years; i.e. bat DNA has a high AT content. They take us no further in trying to decide whether the AT-rich DNA was present in a common flying ancestor or whether it was acquired independently by megabat DNA and microbat DNA. The later possibility is strengthened by the different isochore profiles in the two kinds of bats, by the known association between AT-rich DNA and high levels of aerobic metabolism, and by the fact that bat monophyly finds no support from GC substitutions in DNA nor from the limited protein sequence data available from bats.

**Title:** Helper Virus-Specific Mutation and Selection in the Evolution of Cucumber Mosaic Cucumovirus Satellite RNAs

**Authors:**

M.J. Roossinck[1] and P. Palukaitis[2]

**Institution(s):**

[1]Plant Biology Div., The S.R. Noble Foundation, Ardmore OK 73401;
[2]Dept. of Plant Pathology, Cornell University, Ithaca, NY 14853.

**Abstract:**

Cucumber mosaic cucumovirus (CMV) is a tripartite ss (+) sense RNA virus. CMV often harbors a small molecular parasite known as a satellite RNA (sat-RNA). Sat-RNAs are completely dependent on their helper virus for replication and encapsidation, but do not contribute any essential function to the virus. CMV sat-RNAs are small ss RNA molecules that do not encode any proteins. They are extremely stable, probably due to extensive secondary and tertiary structure. Naturally occurring CMV sat-RNAs share a high degree of sequence identity, although numerous sequence variants have been found. An alignment of over 40 sat-RNA sequences reveals specific areas of hyper-variability.

Previously, upon serial passage of transcript derived from a cDNA clone of the D sat-RNA, pDsat4, with Fny-CMV as helper virus, a specific hypervariable region of the sat-RNA accumulated changes rapidly (Kurath and Palukaitis. 1990. Virology *176*:8-15). In this study we used LS-CMV, and tomato aspermy cucumovirus as helper viruses for passage of pDsat4 transcript. We demonstrate that changes in sat-RNA sequences occur at different rates, and in different regions, dependent upon the helper virus. Furthermore, we have found that different helper viruses exhibit a strong selection pressure for a specific sat-RNA in mixed infections. The region of helper-virus selection has been mapped on the sat-RNA by using recombinants. Future studies include a genetic analysis of the viral sequences responsible for sat-RNA selection.

**Title:** The use of Glutamine Synthetase enzymes for tracing the tree of life

**Authors:** Cecilia Saccone

**Institution(s):** Dipartimento. di Biochimica e Biologia Molecolare, Università di Bari, via Orabona, 4, 70125 Bari, Italy.

**Abstract:**

Glutamine Synthetase (GS), a pivotal enzyme for nitrogen metabolism, is found in at least three distinct forms, namely the GS1, GS2 and GS3. GS1 and GS3 have been found so far only in Prokaryotes, wheras GS2 is primarily associated with Eukaryotes though it is present in some Prokaryotes.

In agreement with our previous description we show that both GS1 and GS2 enzymes behave as good molecular clocks evolving at the same rates at the level of the second codon positions.

The phylogenetic tree of the GS1 isoforms shows Archaebacteria to be more closely related to Eubacteria than to Eukaryotes. This finding is confirmed by the phylogenetic analysis carried out on both large and small subunits of rRNA. However, differently from the rRNA analyses, Crenarchaeota and Euryarchaeota Archaebateria, as well as the high and low GC gram positive bacteria, appear to be polyphyletic. We provide evidence that the observed polyphyly of Archaebacteria might be only apparent, resulting from a gene duplication event preceding the split between Archaebacteria and Eubacteria and followed by the retention of only one isoform in the extant lineages.

Our data which show Eubacteria more closely related to Archaebacteria place the root of the universal tree of life in the eukaryotic line of descent and are in agreement with the evolutionary analyses carried out on DNA topoisomerase II, citrate synthase, HSP70, glutamate dehydrogenase and large and small rRNA subunits. Two other topologies with Eubacteria or Archaebacteria as outgroups, which are supported by few other genes, will be discussed.

**Acknowledgments.**

This work was partially financed by MURST, Italy, Progetto Finalizzato Ingegneria Genetica, CNR, Italy and Progetto Finalizzato Biotecnologie e Biostrumentazione, CNR, Italy.

# MOSAIC EVOLUTIONARY HISTORIES OF "AIDS' VIRUSES

**Paul M. Sharp[1], David L. Robertson[1], Feng Gao[2] and Beatrice H. Hahn[2]**

**Department of Genetics, University of Nottingham, and Department of Medicine and Microbiology, University of Alabama at Birmingham**

HIV-1 and HIV-2 each cause AIDS. They are members of a subfamily of lentiviruses : this subfamily includes several other viruses isolated from non-human primates (SIVs). Phylogenetic analyses of the known primate lentiviruses produce 5 major, approximately equidistant, lineages :

1. HIV-1 plus SIV from chimpanzees.
2. HIV-2 plus SIV from mangabeys and macaques.
3. SIV from African green monkeys.
4. SIV from Syke's monkey.
5. SIV from mandrill.

Within lineages 1, 2 and 3, multiple subtypes can be identified.

In lineage 3, the viruses cluster according to the subspecies of african green monkey from which they were isolated, but within lineages 1 and 2, and indeed over the primate lentivirus phylogeny as a whole, host and virus relationships differ. In addition, a number of viruses have genomes within which different regions have different phylogenies. The evolutionary history of these viruses is thus mosaic, in terms of host-dependent evolution *vs.* cross-species transmission events, and recombination among divergent lineages.

**Title:** Adjustment of mutation rates: proximate or ultimate causation?

**Authors:** Paul D. Sniegowski

**Institution(s):** Michigan State University

**Abstract:**

Several recent reports have claimed that mutations in bacteria and yeast can occur at greatly elevated rates specifically when the mutant phenotypes are advantageous. If true, these results suggest that favorable mutations can spread through populations in proximate response to environmental factors rather than solely through natural selection, as is widely held. The debate over these results has resulted in confusion over terminology and issues, but may have the salutary effect of focusing attention on heretofore unstudied mechanisms of mutagenesis and on evolutionary aspects of mutation. I argue that the evidence for "selection-induced mutations," however, is inconclusive. A number of population-level processes and physiological effects can contribute to an elevated rate of recovery of favorable mutants in the absence of preferential elevation of their rate of occurrence. In the cases that have been re-examined in detail, valid alternatives to selection-induced mutation have been identified. The kind of evidence necessary to document the occurrence of "selection-induced mutation" is discussed. Finally, the adjustment of rates of *random* mutation at specific loci, as an evolutionary (ultimate) response to environmental fluctuations, is discussed. There is considerable circumstantial evidence for such adjustment of mutation rates in pathogenic microbes.

Containment of retrotransposon copy number in Drosophila: a test of a current model. Sniegowski, P. D. and B. Charlesworth. Dept. of Ecology and Evolution, University of Chicago.

**Abstract:** Population studies of the distribution of transposable elements (TEs) on the chromosomes of *Drosophila melanogaster* have suggested that their copy number increase due to transposition is balanced by some form of natural selection. Theory suggests that, as a consequence of deleterious ectopic meiotic exchange between TEs, selection can favor genomes with lower TE copy numbers. This predicts that TEs should be less deleterious, and hence more abundant, in chromosomal regions in which recombination is reduced. To test this, we surveyed the abundance and locations of 10 families of TEs in recombination-suppressing chromosomal inversions from a natural population. The sample of 49 chromosomes included multiple independent isolates of seven different inversions and a corresponding set of standard chromosomes. For all 10 TE families pooled, copy numbers were significantly higher overall within low-frequency inversions than within corresponding regions of standard chromosomes. TEs occupied chromosomal sites at significantly higher frequencies within the *In(3R)Mo* and *In(3R)K* inversions than within the corresponding regions of standard 3R chromosomes. These results are consistent with the predictions of the ectopic exchange model.

**Title:** Where Do We Place the Nearly Neutral Theory?: The development of the nearly neutral theory and its position in the controversy

**Authors:**

Steen, Tomoko Y.

**Institution(s):**

Department of Science & Technology Studies, 726 University Ave.; 237E Corson Hall, Cornell University, Ithaca, NY 14850

**Abstract:**

Understanding the current development of molecular evolutionary biology, the selectionist-neutralist evolutionary controversy is the most crucial issue. Since the advocacy of the neutral theory, much criticism has been raised. The major criticism has been the gap between prediction and actual calculation. The neutral theory predicted the constancy of amino acid substitutions by generation, but calculated constancy by the year. In order to counter-argue this criticism, Tomoko Ohta developed her early version of the nearly neutral theory of molecular evolution in 1973.[1]

There has been, however, much confusion about the difference between the neutral theory and the nearly neutral theory over the last twenty years. Also, several versions of the so-called "nearly neutral theory" exist. Conducting a year of research with Drs. Ohta and Kimura at the National Institute of Genetics and focusing on the development of the nearly neutral theory, I have developed a further understanding of the differences between Ohta's nearly neutral theory and the original neutral theory, as well as of the other versions of the nearly neutral theory.

My poster will be a sketch on the historical development of Ohta's nearly neutral theory: how the theory branched off from the original neutral theory, and on its interaction with the other theories: where the nearly neutral theory should be placed in this controversy. The primary focus of my poster is the historical development of Ohta's nearly neutral theory and the clarification of its position in the neutralist-selectionist controversy.

1. "Slightly deleterious mutant substitutions in evolution" in Nature 246: 96-98.

# Mutation Pressure and Dynamics of Molecular Evolution: Parity Rules of DNA Base Composition at Equilibrium and Strand Biases

Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309-0347

The effect of mutation and selection on the base composition of DNA is dictated by a parity rule that originates from the complementary nature of DNA base pairing (Fig. 1).

Fig. 1

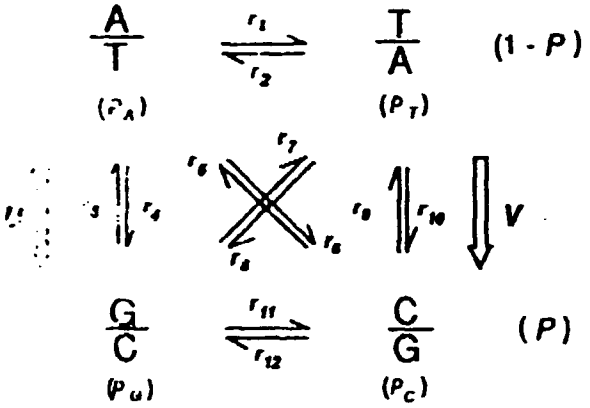
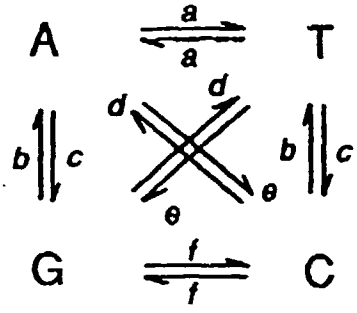



Fig. 2




Here $P_A$, $P_T$, $P_G$, and $P_C$ represent relative frequency of the four nucleotides, where $P_A + P_T = P$, $P_G + P_C = 1$, and $P = P_G + P_C$. $r_1$ to $r_{12}$ are individual conversion rates, and $U$ and $V$ are the respective rates of γ-pair (G/C or C/G) to α-pair (A/T or T/A) and α-pair to γ-pair per nucleotide per time, respectively. Conversion rate is defined as the products of mutation rate ($\mu_i$) and selective coefficient ($s$) as $r_i = \mu_i \cdot s_i$. For neutral nucleotides where s=1, $r_i$ in Fig. 1 can be replaced with $\mu_i$.

- When no strand biases exist, the following rules hold under any conversion biases:
  1. $P_A = P_T$ and $P_G = P_C$ (parity rule). This rule comes from the fact that, without strand biases, always $r_1 = r_2 (= a)$, $r_3 = r_9 (= b)$, $r_4 = r_{10} (= c)$, $r_5 = r_7 (= d)$, $r_6 = r_8 (= \theta)$, and $r_{11} = r_{12} (= f)$ (Fig. 2).
  2. Mutations between isopairs (α-pairs or γ-pairs) do not affect the GC-content.
  3. At equilibrium, $U = r_3 + r_5 + r_7 + r_9$, $V = r_4 + r_6 + r_8 + r_{10}$, and conversion bias is $r_D = V/(U + V)$. This means that isopair mutations do not affect mutational bias and dynamics of $m$ (allopair mutation frequency) and $M$ (cumulative allopair mutations) previously defined (Sueoka, 1993).

- When strand biases exist, the two strands of DNA receive different conversion processes, and as a result the following points become reasonable:
  1. A consistent mutational bias between two strands of DNA in a long stretch of nucleotides is unlikely. On the other hand, selective constraints on the sense strand by the difference in tRNA abundance seems real.
  2. Conversion bias between the two strands of DNA is reflected only in the frequencies of the two bases on the sense strand that are involved in the conversion, and the parity rule ($P_A = P_T$ and $P_G = P_C$) is no longer applicable.
  3. Deviations from $P_A = P_T$ and from $P_G = P_C$ generally provide a measure of selective constraints rather than mutational bias between the two strands of DNA.

**Further Support of Difference in Directional Mutation Pressure as a Major Underlying Cause of Isochore Formation in Multicellular Eukaryotes.**

Hiroshi Sueoka, Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, CO 80309-0347

Previously, I have pointed out that directional mutation pressure may be responsible for the different GC contents of the various domains of chromatids termed as isochores (Bernardi et al. 1985). The proposal was based on the fact that the GC content of the third codon position nucleotides shows remarkably similar features to interspecific variation in bacteria. However, a number of opposite views have been expressed that consider functional selection, instead of directional mutation pressure, could plays the major role in multicellular eukaryotes. Among proposed functional selections, differential usage of synonymous codons due to the varying abundance of isoaccepting tRNA molecules has been the strongest contender because an apparent correlation between the codon usage bias and tRNA abundance does exist. This argument does not, however, explain the following facts: (i) Directionally biased usage of synonymous codons in GC content is consistently found in 57 out of 59 multiple sense codons in different genes of humans (Sueoka, 1992). (ii) Third codon positions consistently reflects directional mutation pressure higher than first and second codon positions. (iii) In spite of the wobbling nature of tRNA, usage of synonymous codons are consistently biased with regard to the overall GC content of each gene in humans as originally shown by the analysis of genes of bacteria with extremely high and extremely low GC content by S. Osawa and his collaborators.

In this talk, I will present another supporting evidence of the directional mutation pressure as playing the major role for intra-specific heterogeneity through the analyses of the correlation between $P_3$'s of two- and four-codon amino acids in various multicellular eukaryotes.

a. In unicellular organisms, the breakdown of the parity rule ($P_A = P_T$ and/or $P_G = P_C$) is generally consistent for each four-codon amino acid in each organism, whereas in multicellular eukaryotes, the breakdown of the parity rule is generally unique in each gene or most likely in the genes in the same isochore of an organism.

b. In spite of the numerous violation of the parity rule among four synonymous codons of various amino acids, individual genes manifest definite influence of the directional mutation pressure when $P_3$ values (GC contents of the third codon letters) of the four-codon amino acids are plotted against those of the two-codon amino acids. As a matter of fact, the regression coefficients are so consistently high ($r = 0.81$ to $0.94$) in different multicellular organisms, the results seem to support the idea that the principles of directional mutation pressure analyses are also applicable to the intraspecific or inter-isochore heterogeneity of the base composition of multicellular eukaryotes.

**Title:** Protein sequences as fossil record of early gene fusion

**Authors:** E. N. Trifonov

**Institution(s):** Department of Structural Biology, The Weizmann Institute of Science, Rehovot 76100, Israel

**Abstract:**

Analysis of length distributions of protein sequences, taken from Swis-Prot database after thorough cleaning of the data, reveals that eukaryotic sequences show clear preferential sizes of 120-126 aa residues as well as twice, trice and four times this value. Similarly, prokaryotic sequences have preferential lengths 148-156 residues, as well as about 300 and 450 aa residues. The size of DNA that would code the elementary units of about 120 (150) aa, i. e. about 400 bp, corresponds to an optimal DNA ring closure size. This suggests that the observed segmented structure of the protein sequences reflects an early stage of protein evolution when early small genes - DNA rings of close-to-optimal size - were joined together in various combinations making new genes of larger sizes. A strong prediction follows from this hypothesis: at the junctions between former genes-segments their initiation triplets (methionines in the protein segments) should be found, those which survived subsequent mutations. Such preference of methionines to positions at multiples of about 120 aa is found indeed. The revealed segmented structure of proteins, apparently, reflects an early stage of genome evolution - transition from disperse independent small DNA rings, including the protein-coding ones, through variety of molecular recombinations, to still independent larger composite molecules, and eventually to only one (few) chromosome(s) of the early genome. The initial inventory of the unit size segments and their reshuffling could have been a powerfool combinatorial source for diversification of the early genomes.

Title: Molecular Evolution of Imprinted Genes

Authors: Kenneth H. Wolfe

Institution(s): Department of Genetics, University of Dublin
Trinity College, Dublin 2, Ireland

FAX +353-1-679-8558 e-mail khwolfe@unix2.tcd.ie

Abstract:

Imprinting is a phenomenon in mammals whereby only one of the two alleles of a gene is expressed; the process has some similarities to X-inactivation. The expressed copy is the maternally-derived allele in some cases and the paternal allele in others. A few genes are imprinted in some tissues whereas both copies are expressed in other tissues. The mechanism of imprinting is thought to involve DNA methylation, which means that the molecular evolution of imprinted genes may differ from that of "normal" genes due to the presence of signals for imprinting.

In a survey of the molecular evolution of a large number of rodent genes [1] it was apparent that some genes had a extraordinarily low rate of silent nucleotide substitution. One of these genes is IGF-II (insulin-like growth factor II), a gene that has been reported to be imprinted in both mice and humans. This has prompted a more detailed examination of evolutionary rates in imprinted genes.

[1] Wolfe KH, Sharp PM (1993) J Mol Evol 37:441-456

**Title:** MOLECULAR PATHWAYS TO PARALLEL EVOLUTION

**Authors:** Emile Zuckerkandl

**Institution(s):** Institute of Molecular Medical Sciences, Palo Alto, California

**Abstract:**

A framework is developed for understanding parallel evolution at the level of gene interaction networks and their relation to morphology. The role of parallel changes in quantitative control of gene expression is emphasized. Three kinds of molecular pathways for parallel evolution are discussed: those dependent upon trends inherent in genome structure; those in part defined by the intervention of environmental agents; and those based on limitations of the spectrum of realizable phenotypes.

## Terminal section

Molecular Evolution is now undergoing a tremendous development. It could, therefore, be foreseen that the Costa Rica Workshop was going to be a success. And, indeed, it was because of the willingness of people, who are supporting different and even opposite viewpoints, to confront their ideas. The setting of the Meeting was ideal to allow such exchanges of ideas and the local organizer, Dr. Gabriel Macaya, deserves to be highly congratulated for an impeccable organization.

The participation of a number of young researchers from Latin America and their lively participation in the debates was a most positive point of the Workshop.

There is no doubt that another such Workshop could most usefully be held again in two year's time.

**An important sequel of the Workshop is a special issue of the Journal of Molecular Evolution which will gather a number of the papers presented at the Workshop and make available to a larger audience many of the ideas that were debated there.**

## FINANCIAL REPORT

This is presented in the following two pages. It deserves the following comments.

1. Living expenses and local transportation. The Costa Rican Company "Horizontes" took care, most efficiently, of the hotel accommodations in San José (2 nights) and in Guanacaste (5 nights) as well as of the transportation from Airport to San José and back and from San José to Guanacaste and back. We paid a total of 50,000 US Dollars to cover all these expenses. The Horizontes invoice is higher, 55,111US Dollars because it includes single-room supplements which were (for the most part) paid by the participants themselves. In three cases (François, Eyre-Waiker and West-Eberhard), expenditure was less than 673 US Dollars per person, because those people did noy stay for the whole duration of the Meeting.

2. Travel expenses were paid by check by G. Bernardi. All check copies are available.

3. Organizational expenses were kept at a minimum level.

4. Ten people only paid registration fees. Invited speakers and Latin American participants did not.

5. We are looking for some additional money (2500 US Dollars) in order to cover the deficit.

**INCOME**

| | |
|---|---|
| Sloan Foundation | 25,000 |
| ICGB | 20,000 |
| COGENE | 11,500 |
| IUBMB | 6,000 |
| Springer-Verlag | 2,500 |
| Registration fees (see attached list) | 8,000 |
| **Total income** | **73,000** |

**EXPENDITURES**

| | |
|---|---|
| Living expenses and local transportation | 50,000 |
| Travel expenses (see attached list) | 19,500 |
| Organizational expenses (mail, fax, telephone, abstract book, secretarial expenses) | 6,000 |
| **Total expenditures** | **75,500** |
| **BALANCE** | **- 2,500** |

**Travel expenses**

| | |
|---|---|
| Arber | 800 |
| Bernardi Giacomo | 700 |
| Bernardi Giorgio | 1300 |
| Britten | 200 |
| Cairns | 800 |
| De Jong | 400 |
| Dover | 800 |
| Eyre-Walker | 200 |
| Fitch | 400 |
| Gilbert | 1000 |
| Gojobori | 800 |
| Goodman | 800 |
| Graur | 200 |
| Hall | 400 |
| Holmquist | 800 |
| Honeycutt | 400 |
| Ikemura | 600 |
| Karlin | 400 |
| Keynes | 800 |
| Li | 200 |
| Logdson | 400 |
| Musto | 800 |
| Okada | 800 |
| Roossinck | 800 |
| Saccone | 400 |
| Sharp | 800 |
| Sniegowski | 400 |
| Sueoka | 800 |
| Trifonov | 1200 |
| Wolfe | 800 |
| Zuckerkandl | 400 |
| **Total** | **19,500** |

**Registration fees**

| | |
|---|---|
| Breukelman | 800 |
| Cerff | 800 |
| Duret | 800 |
| Hewett-Emmett | 800 |
| Junakovic | 800 |
| Lanave | 800 |
| Marin | 800 |
| Oliver | 800 |
| Pesole | 800 |
| Tashian | 800 |
| **Total** | **8,000** |

HORIZONTES
COSTA RICA

Tel. (506) 22-20-22 – Fax (506) 55-45-13
P.O. Box 1780 - 1002 Paseo de los Estudiantes
San José, Costa Rica.

CREDITO

Lic. ICT No. 61
Cédula Jurídica No. 3-101-062240-31

**FACTURA No. 0003279**

| DIA | MES | AÑO |
|---|---|---|
| 29 | Abril | 1994 |

EMPRESA: OPEN QUESTIONS IN MOLECULAR EVOLUTION
CONCEPTO: CONGRESO DE CIENTIFICOS

CODIGO: EVO-424

FECHA SERVICIO: 17/04/94 al 24/04/94

No. FAX: 75

CREDITO
2724 - 3279

| CANTIDAD | DETALLE | | MONTO |
|---|---|---|---|
| 68 | PARTICIPANTES INTERNACIONALES A 673 USD C/U | $ 45,764.00 | 7,056,555.49 |
| 3 | PARTICIPANTES NACIONALES A 518 USD C/U | $ 1,554.00 | 240,636.90 |
| 32 | SINGLE SUPPLEMENTS A 200 USD C/U | $ 6,400.00 | 991,040.00 |
| 1 | ANTHONY FRANCOIS | $ 377.00 | 58,378.45 |
| 1 | ADAM EYRE-WALKER | $ 564.00 | 87,335.40 |
| 1 | MARY JANE WEST-EBERHARD | $ 352.00 | 54,507.20 |
| 1 | GASTOS EXTRA | $ 100.00 | 15,485.00 |
| | Total | $ 55,111.00 | |
| | ULTIMA LINEA | | |
| | RATE OF EXCHANGE : 154.85 | | |

FORMULARIOS STANDARD TEL. 24 46 72

AGENTE

RECIBIDO CONFORME

TOTAL ⟹ ₡ 8,533,938.35

CONDICIONES AL DORSO