**2nd INTERNATIONAL CONFERENCE/WORKSHOP**

# Genomic Impact Of Eukaryotic Transposable Elements

**FEBRUARY 6 – 10, 2009**

**ASILOMAR, PACIFIC GROVE, CALIFORNIA, USA**

2ND INTERNATIONAL CONFERENCE/WORKSHOP

# Genomic Impact Of Eukaryotic Transposable Elements

**Organizer:**

**Jerzy Jurka**

*Genetic Information Research Institute*
*Mountain View, California, USA*

**ASILOMAR 2009**

*These abstracts should not be cited in bibliographies. Material contained herein should be treated as personal communication and should be cited as such only with consent of the author.*

*Please note that recording of oral sessions by audio, video or still photography is strictly prohibited except with the advance permission of the author(s) and organizers.*

### 2nd International Conference and Workshop

# Genomic Impact of Eukaryotic Transposable Elements

### Organizer: Jerzy Jurka

## Friday, February 6, 2009

| | | |
|---|---|---|
| 15:00-18:00 | REGISTRATION (Phoebe A. Hearst Social Hall) | |
| 18:00-19:00 | *Dinner* (Crocker Dining Hall) | |
| 19:30-23:00 | Warm-up party/poster previews/preparation of audio-visual (Fred Farr Forum/Kiln) | |

The following equipment will be provided in all sessions: an LCD projector, a laser pointer and a microphone. Speakers should load their talks at Fred Farr Forum in the evening preceding the presentations. There will be a limited time for last-minute testing (30 min. before the morning session and during breaks). Equipment for 35 mm slides WILL NOT be provided at this meeting.

## Saturday, February 7, 2009

| | | |
|---|---|---|
| 7:30-8:30 | *Breakfast* (Crocker Dining Hall) | |
| 8:00-9:00 | REGISTRATION (Phoebe A. Hearst Social Hall) | |
| 9:00-9:10 | **Jerzy Jurka** – Opening remarks | |
| 9:10-9:50 | **Roy Britten** – History and relationships of TEs | p.1 |
| 9:50-10:20 | **David Haussler** - Evidence that transposons shaped vertebrate gene regulatory networks | p.2 |
| 10:20-10:50 | *Coffee-break* (Fred Farr Forum)/**Group photo** | |
| 10:50-11:20 | **Gill Bejerano** - A study of co-option in the human genome | p.3 |
| 11:20-11:50 | **Michael Savageau** - Quantitative evolutionary design of gene circuits: lessons from Bacteria | p.4 |
| 12:00-13:00 | *Lunch* (Crocker Dining Hall) | |
| 13:30-14:00 | **Juergen Brosius** - The retro-look of genomes | p.5 |
| 14:00-14:30 | **Norihiro Okada -** Multiple SINE insertions made our brain mammalian? | p.6 |
| 14:30-15:00 | **Chantal Vaury** - Insulators brought by TE contribute to gene regulation in relationship with nuclear architecture | p.7 |
| 15:00-15:30 | **Dixie Mager -** Complex epigenetics of mammalian endogenous retroviruses | p.8 |
| 15:30-16:00 | *Coffee-break* (Fred Farr Forum) | |
| 16:00-16:30 | **Cedric Feschotte** - Mammalian transposable elements and the emergence of lineage-specific functions | p.9 |
| 16:30-17:00 | **King Jordan -** Transposable elements, chromatin and gene regulation | p.10 |
| 17:00-17:30 | **Alan Weiner** - Human PGBD3: a piggyBac transposon that is both good and bad for us? | p.11 |

| | | |
|---|---|---|
| 18:00-19:00 | *Dinner* (Woodlands) | |
| 19:15-19:45 | **John Moran -** Studies of a human transposable element | p.12 |
| 19:45-20:15 | **Sandy Martin** - Mutational Analyses of ORF1p Function in LINE-1 Retrotransposition | p.13 |
| 20:15-20:45 | **Thomas Eickbush** - Regulating the expression of R2 elements within the nucleolus | p.14 |
| 20:45-23:00 | *Happy Hours* / **Poster session** – odd numbers (Kiln) | |

## Sunday, February 8, 2009

| | | |
|---|---|---|
| 7:30-8:30 | *Breakfast* (Crocker Dining Hall) | |
| 9:00-9:30 | **Holly Wichman** - Comparative biology of L1 elements | p.15 |
| 9:30-10:00 | **Prescott Deininger** - Cellular responses to damage by non-LTR retroelements | p.16 |
| 10:00-10:15 | **Nicolas Gilbert** - LINE-1 mediated mobilization of snRNA | p.17 |
| 10:15-10:30 | **Eric Devor** - A microRNA incubator on the marsupial (Monodelphis domestica) X-chromosome was created via L1 transposon-mediated serial duplication | p.18 |
| 10:30-10:45 | **Corrado Spadafora** - Functional roles for LINE1-encoded reverse transcriptase in embryonic development and tumor progression | p.19 |
| 10:45-11:00 | *Coffee-break* (Fred Farr Forum) | |
| 11:00-11:30 | **Josefa Gonzalez** - High rate of recent transposable element-induced adaptation in Drosophila melanogaster | p.20 |
| 11:30-12:00 | **Mark Batzer** - Mobile elements and primate genomic variation | p.21 |
| 12:00-13:00 | *Lunch* (Crocker Dining Hall) | |
| 13:30-14:00 | **Zoltan Ivics** - Transposon-host cell interactions in the regulation of Sleeping Beauty transposition | p.22 |
| 14:00-14:30 | **Regina Baucom** - The evolution, activities and specificities of transposable elements in grass genomes | p.23 |
| 14:30-14:45 | **Marie-Angele Grandbastien** – Transposable elements and TE-gene associations in the transcriptome of Solanaceae species | p.24 |
| 14:45-15:00 | **R. Keith Slotkin** - Germ cell-specific activation and silencing of transposable elements in plants | p.25 |
| 15:00-15:15 | **Vincent Colot** - A role for RNAi in the correction of transposable element methylation and silencing defects in Arabidopsis | p.26 |
| 15:15-15:30 | **Ken Naito** - Live watching TE burst: behavior and impact of a rice transposon mPing | p.27 |
| 15:30-15:50 | *Coffee-break* (Fred Farr Forum) | |
| 15:50-16:20 | **Arian Smit** - Evolution of transposable elements in mammals and birds | p.28 |
| 16:20-16:50 | **Andrew Shedlock** - The Anolis genome assembly: balancing mammalian and avian perspectives on amniote repeat evolution | p.29 |
| 16:50-17:20 | **Vladimir Kapitonov** - Stories of outrageous horizontal transfer of DNA transposons | p.30 |
| 17:20-17:50 | **Irina Arkhipova** - Massive Horizontal gene transfer in bdelloid rotifers | p.31 |
| 18:30-22:30 | *Dinner at Monterey Bay Aquarium* (buses depart at 18:30) | |

## Monday, February 9, 2009

| | | |
|---|---|---|
| 7:30-8:30 | *Breakfast* (Crocker Dining Hall) | |
| 9:00-9:30 | **Hadi Quesneville** - REPET: pipelines for the identification and annotation of transposable elements in genomic sequences | p.32 |

| | | |
|---|---|---|
| 9:30-10:00 | **David Pollock** - Identifying repeat-derived regions in the twilight zone | p.33 |
| 10:00-10:30 | **Marcelo Bento Soares** - High-throughput sequence-based epigenomic analysis of human Alu repeats | p.34 |
| 10:30-10:45 | *Coffee-break* (Fred Farr Forum) | |
| 10:45-11:00 | **Guillaume Bourque** - Evolution of the mammalian transcription factor binding repertoire via transposable elements | p.35 |
| 11:00-11:15 | **Victor Zhurkin** - The tumor suppressor protein p53 binding sites in mammalian genomes are related to transposons | p.36 |
| 11:15-11:30 | **Tomasz Zemojtel** - Methylation and deamination of CpGs generates p53 binding sites on a genomic scale | p.37 |
| 11:30-11:45 | **Geoff Faulkner** - The regulated retrotransposon transcriptome of mammalian cells | p.38 |
| 11:45-12:00 | **Sebastian Szpakowski** - Methylation profiling: repetitive elements and disease progression | p.39 |
| 12:00-13:00 | *Lunch* (Crocker Dining Hall) | |
| 13:30-14:00 | **Marlene Belfort** - Stressed retrotransposons on the move | p.40 |
| 14:00-14:15 | **Nancy Craig** - Characterization of an active mammalian DNA transposon | p.41 |
| 14:15-14:30 | **Gerald Schumann** - 5'-transduced SVA retrotransposon groups spread efficiently throughout the human genome | p.42 |
| 14:30-14:40 | **Dustin Hancks** - Exon-trapping mediated by the SVA retrotransposon | p.43 |
| 14:40-14:55 | **Hidenori Nishihara** - Retroposon analysis reveals simultaneous divergence of the placental mammalian ancestor possibly triggered by continental divisions | p.44 |
| 14:55-15:05 | **Kamal Rawal** - ELAN: A server based tool for genome wide analysis of mobile genetic elements | p.45 |
| 15:05-15:30 | *Coffee-break* (Fred Farr Forum) | |

## Mini-workshop on Classification and Nomenclature of TEs

Moderators: **Pierre Capy, Cedric Feschotte, Jerzy Jurka, Vladimir Kapitonov, Arian Smit**

| | | |
|---|---|---|
| 15:30-16:00 | **Pierre Capy** - Classification of transposable elements within the Tc1/mariner super-family | p.46 |
| 16:00-16:30 | **Ruth Seal/Jens Mayer** (joint presentation) - The need for an approved nomenclature for (human) endogenous retroviruses | p.47 |
| 16:30-16:35 | **Jonas Blomberg** - Optimal taxonomical markers for ERVs | p.48 |
| 16:35-16:40 | **Francois Sabot** - Non-autonomous and complete derivate elements: how to classify them? | p.48 |
| 16:40-16:45 | **Irina Arkhipova** - Superfamily-specific reference datasets as possible aids in resolving classification problems | p.48 |
| ~16:15-17:30 | Open floor: nominations and election of the International Committee for Classification of TEs (ICCTE). | |
| 18:00-19:00 | *Dinner* (Woodlands) | |
| 19:30-23:00 | Happy Hours / **Poster session** – even numbers (Kiln) | |

## Tuesday, February 10, 2009

| | |
|---|---|
| 7:30-8:30 | *Breakfast* (Crocker Dining Hall) |
| 12:00-13:00 | *Lunch* (Crocker Dining Hall) |

**Due to high interest in the "Genomic Impact of Eukaryotic Transposable Elements", a special issue of Gene devoted to this topic will be published after the conference. Deadline for manuscript submissions is April 30, 2009. Details will be posted on the conference website http://www.girinst.org/conference/Asilomar-2009/index.html after the conference.**

## POSTER PRESENTATIONS

(1) **Population structure in Medicago truncatula inferred from MITE insertion polymorphisms** p.49
*Dariusz Grzebelus*, Mirosława Gładysz, Anna Gambin*

(2) **REPET: pipelines for the identification and annotation of transposable elements in genomic sequences** p.50
*Timothee Flutre*, Elodie Duprat, Hadi Quesneville*

(3) **Changes in transposable element activity and the epigenomic consequences in immortalized plant cells** p.51
*Milos Tanurdzic*, Matthew Vaughn and Robert Martienssen*

(4) **Tol1 and Tol2 of medaka: two similar hAT-family elements at different evolutionary stages in a single fish genome** p.52
*Akihiko Koga*

(5) **Regulatory Impact of MusD Retroelements** p.53
*Tuğçe Aktaş*, Marc Friedli, François Spitz*

(6) **A recent adaptive transposable element insertion near highly conserved developmental loci in Drosophila** p.54
*Josefa González*, J. Michael Macpherson, and Dmitri A. Petrov*

(7) **Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition** p.55
*Masaki Kajikawa*, Jun Suzuki, Katsumi Yamaguchi, Noritaka Adachi, Hideki Koyama, Shunichi Takeda and Norihiro Okada*

(8) **Identifying cis-acting elements capable of epigenetic regulation** p.56
*Monica Sentmanat*, Sarah Elgin*

(9) **Transposable elements and young sex chromosomes in dioecious plant Silene latifolia** p.57
*Eduard Kejnovsky*, Roman Hobza, Zdenek Kubat, Tomas Cermak, Jiri Macas, Alex Widmer and Boris Vyskot*

(10) **Serum albumin intron-1 of western Palearctic water frogs contains a non-LTR CR1-like retrotransposon** p.58
*Jörg Plötner*, Frank Köhler, Thomas Uzzell, Peter Beerli, Hansjürg Hotz, Robert Schreiber, and Gaston-Denis Guex*

(11) **Short retroelement could induce RNAi leading to silencing of a relative LINE** p.59
*Nickolai A. Tchurikov* and Olga V. Kretova*

(12) **The Origination and Evolution of Primate Retrogenes** p.60
*Qu Zhang*, Amir Karger, and Maryellen Ruvolo*

(13) **Assessment of the extent of substitution rate variation of Long Terminal Repeat sequences in Oryza sativa and Oryza glaberrima** p.61
*Andrea Zuccolo*, Aswathy Sebastian, Yeisoo Yu and Rod A. Wing*

(14) **Evolutionary dynamics of the LTR retrotransposon roo inferred from twelve complete Drosophila genomes** p.62
*Nicole de la Chaux* and Andreas Wagner*

(15) **Evolution of Can-SINEs in Felidae and other Carnivora** p.63
*Kathryn Walters*, Marc Allard, Diana Johnson, Jill Pecon-Slattery*

(16) **BNR – a LINE family with an ORF1 RNA-binding motif is present in a variety of higher plants and represents a novel L1 subclade** p.64
*Tony Heitkam* and Thomas Schmidt*

(17) **LTR retrotransposons in filamentous fungi.** p.65
*A. Muszewska*, M. Grynberg and A. Paszewski*

(18) **Reconstructing functional, virtual transposable elements from dense contigs of BAC-end sequences in Genbank** p.66
*Howard Laten*, Lauren Mogil, LaBianca Wright, Darya Isayeva, Donald Jonker, Eftychia Gouvas, Edward Badal, and Christi Lindorfer*

(19) **Retrotransposon nik: an infective retrovirus?** p.67
*Adriana Ludwig*

(20) **Mammalian imprinted genes derived from retrotransposition are regulated by maternal methylation and exert monoallelic gene expression on their hosts** p.68
*R.J. Oakey*, R.B. McCole, R. Schulz and A.J. Wood.*

(21) **A bioinformatic analysis of the association between small RNAs and transposable elements in Arabidopsis** p.69
*Alexis Sarazin*

(22) **Giant genomes of Fritillaria lilies are dominated by LTR retrotransposons** p.70
*Katerina Ambrozova, Jiri Macas, Ilia J. Leitch, Martin A. Lysak*

(23) **Lack of Evidence that Repeats in Short Indels in Tetrahymena deletion elements Form Parts of Longer Repeats.** p.71
*Piroska Huvos*, Michael Wainer, undergraduates: Jason Maas, David Kluge, Amanda Harvey, Kelly Ming, Josh Newby, Richa Ailinani, Jason Horton, Dan Scott*

(24) **Ancestral polymorphism or frequent hybridization? Curious presence of L1 retroelements in Mus spretus, M.macedonicus and M.musculus** p.72
*Václav Janoušek*, Pavel Munclinger, Barbara Dod, Pierre Boursot*

(25) **Characterization of centromeric retrotransposons in diverse plant species** p.73
*Pavel Neumann*, Alice Navratilova, Veronika Steinbauerova, Eduard Kejnovsky, Boris Vyskot, Eva Hribova, Jiri Macas*

(26) **Genetic and epigenetic characterization of Alu repeats in normal and cancer cells** p.74
*Mireia Jordà*, Berta Martín, Jairo Rodríguez, Víctor Barrera, Miguel A. Peinado*

(27) **Discreet charm of neglected markers: SINE insertion polymorphisms in Western Palaearctic house mice** p.75
*Pavel Munclinger*, Václav Janoušek, Lenka Kubešová, Barbara Dod, Pierre Boursot*

(28) **RetroTector online, a rational tool for analysis of retroviral elements in small and medium size vertebrate genomic sequences.** p.76
*Göran Sperber, Anders Lövgren, Nils-Einar Eriksson, Farid Benachenhou and Jonas Blomberg**

(29) **Genome wide analysis of Sleeping Beauty (SB), piggyBac and Tol2 transposon mediated integration in human primary T cells** p.77
*Xianzheng Zhou*

(30) **Inter- and intra-family variation in transposable element dynamics among natural populations of a cyclical parthenogen, Daphnia pulex** p.78
*Sarah Schaack*, Abby Wolf, Ellen Pritham, and Michael Lynch*

(31) **Discovery of the first primate endogenous lentivirus** p.79
*Clément Gilbert*, David Maxfield, Cédric Feschotte*

(32) **Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods** p.80
*John K. Pace II, Clément Gilbert*, Marlena S. Clark, and Cédric Feschotte*

(33) **The Contribution of L1 Transcription to Variation in the Human Transcriptome** p.81
*Sanjida H Rangwala*, Lili Zhang and Haig H. Kazazian Jr.*

(34) **Discovery and Annotation of TEs on VectorBase** p.82
*Ryan C. Kennedy*, Maria F. Unger, Scott Christley, Jenica L. Abrudan, Neil F. Lobo, Greg Madey, Frank H. Collins*

(35) **Integration Mechanism of the Site Specific Non-LTR retrotransposon R2Bm** p.83
*Christi Hull* and Shawn Christensen, Ph.D.*

(36) **Target Site Recognition of Site Specific Non-LTR retrotransposons** p.84
*Blaine Thompson* and Shawn Christensen*

(37) **Transposable elements characterization using the REPET framework in four insect genomes** p.85
*Emmanuelle Permal*, Timothée Flutre, Hadi Quesneville*

(38) **Retrotransposon nik: an infective retrovirus?** p.86
*Adriana Ludwig* 1; Vera Lúcia da Silva Valente, Elgion Lúcio da Silva Loreto*

(39) **Differential insertion of transposable elements in Anopheles gambiae M & S genomes** p.87
*Jenica L. Abrudan*, Ryan C. Kennedy, Maria F. Unger, Michael R. Olson, Scott J. Emrich, Frank H. Collins, Nora J. Besansky*

(40) **The landscape of transposable elements in the soybean genome** p.88
*Jianxin Ma*

(41) **Bioinformatic detection and annotation of non-LTR retrotransposons in the Culex quinquefasciatus mosquito genome.** p.89
*Maria F. Unger, Ryan C. Kennedy, Jenica L. Abrudan, Frank H. Collins*

(42) **The relation of short motif overrepresentation in the human genome and the fragmentation of segmental duplications and transpositions** p.90
*Abanish Singh*, Umeshkumar Keswani, David Levine, Nikola Stojanovic*

(43) **Transposons Shape the Architecture of Plant Genomes** p.91
*Heidrun Gundlach*, Georg Haberer, Rémy Bruggmann and Klaus F.X. Mayer*

(44) **Automatic identification of non-LTR retrotransposons in genomic sequences** p.92
*Mina Rho* and Haixu Tang*

(45) **Transposable elements and evolutionary plasticity of the malaria mosquito genome** p.93
*I.V. Sharakhov*, A. Xia, M.V. Sharakhova, S.C. Leman, Z. Tu*

(46) **Searching for horizontal transfer of transposable elements within grass genomes** p.94
*Aaron O. Richardson* and Susan R. Wessler*

(47) **Different contribution of retroelements to intergenic transcription units** p.95
*Saneyuki Higashino*, Hidemasa Bono, Kunio Kikuchi and Yasunori Aizawa*

(48) **Evidence for Co-Evolution between Human MicroRNAs and Alu-repeats** p.96
*Stefan Lehnert*, Peter Van Loo, Pushpike J. Thilakarathne, Peter Marynen, Geert Verbeke and Frans C. Schuit*

(49) **Inherent promoter bidirectionality facilitates escape from epigenetic silencing during parasitic DNA integration in mammalian genomes** p.97
*Paul Kalitsis and Richard Saffery*

(50) **Conserved Structure of Vertebrate Retroviral LTRs; Detection of Single LTRs in Genomic Data** p.98
*Farid Benachenhou*, Patric Jern, Vidar Blikstad, Merja Oja, Göran Sperber, Panu Somervuo, Samuel Kaski and Jonas Blomberg*

(51) **Two distinct CACTA transposon subfamilies create phenotypic and genomic diversity in soybean** p.99
*Gracia Zabala*, Lila O. Vodkin*

(52) **New superfamilies of eukaryotic DNA transposons and their internal divisions** p.100
*Weidong Bao*, Matthew G. Jurka, Vladimir V. Kapitonov, Jerzy Jurka*

(53) **Mutator as a probe for variations in the epigenetic landscape of plant genomes** p.101
*Damon Lisch*

(54) **Ecg-1 elements create genetic diversity in asexual root-knot nematodes** p.102
*Stephen M. Gross*, Rushi A. Shah, and Valerie M. Williamson*

(55) **Unraveling the DNA repeat content of four Apicomplexan species** p.103
*Assiatu B Barrie*, Cedric Feschotte, Jessica Kissinger and Ellen J. Pritham*

(56) **Locus-specific hyper- and hypo-methylation of mouse B1 elements in male germ cells** p.104
*Kenji Ichiyanagi*, Junko Kitayama, and Hiroyuki Sasaki*

(57) **Analysis of the repetitive DNA in the 1RS chromosome of Rye (Secale Cereale)** p.105
*S. Fluch, D. Kopecky, R. Boonruangrod, M. Berenyi*, B. Klauninger, J. Dolezal, T. Lelley, K. Burg*

(58) **Amlification of an endogenous retrovirus on the chicken W-chromosome** p.106
*T. Revay* and A. Hidas*

(59) **Junk lends a hand: Transposable elements contributing to genetic variation among isolates of Trichomonas vaginalis** p.107
*Claudia P. Marquez* And Ellen J. Pritham*

(60) **Rice transposable elements: Facts and Stats** p.108
*Francois Sabot*

(61) **Exon-trapping mediated by the SVA retrotransposon** p.109
*Dustin C. Hancks*, Adam D. Ewing and Haig H. Kazazian Jr*

(62) **VisualRepbase: an interface for the study of occurrences of transposable element families** p.110
*Sebastien Tempel*

(63) **A microRNA incubator on the marsupial (Monodelphis domestica)X-chromosome was created via L1 transposon-mediated serial duplication** p.111
*Eric J. Devor*, Paul B. Samollow and Andrew S. Peek*

(64) **Intronic Transposable Elements and Their Host Genes** p.112
*Ying Zhang* and Dixie Mager*

(65) **Genomic Drive: A Powerful Engine of Evolution** p.113
*Keith R. Oliver * and Wayne K. Greene*

(66) **Clustering of some SINEs in gene-rich genome regions may be caused by positive selection of specific genome rearrangements facilitating transcription** p.114
*Usmanova N. and *Tomilin N.*

# Author Index

*Abrudan*, Jenica L., p. 82, 87, 89*
*Adachi, Noritaka, p. 55*
*Ailinani, Richa, p. 71*
*Aizawa, Yasunori, p. 95*
*Aktaş*, Tuğçe, p. 53*
*Allard, Marc, p. 63*
*Ambrozova*, Katerina, p. 70*
*Arkhipova*, Irina R., p. 31,48*
*Arndt, Peter F., p. 37*
*Badal, Edward, p. 66*
*Bailey, Arnold D., p. 11*
*Bao*, Weidong, p. 100*
*Barrera, Víctor, p. 74*
*Barrie*, Assiatu B., p. 103*
*Batzer*, Mark A., p. 21, 42*
*Baucom*, Regina, p. 22*
*Beauregard, Art, p. 40*
*Beerli, Peter, p. 58*
*Bejerano*, Gill, p. 3*
*Belancio, V.P., p. 16*
*Belfort, Marlene, p. 40*
*Benachenhou*, Farid, p. 76, 98*
*Bennetzen, Jeff, p. 22*
*Berdasco, Maria, p. 26*
*Berenyi*, M., p.105*
*Besansky, Nora J., p. 87*
*Bhattacharya, Alok, p. 45*
*Bhattacharya, Sudha, p. 45*
*Blikstad, Vidar, p. 98*
*Blomberg*, Jonas, p.48, 76, 98*
*Boccara, Martine, p. 26*
*Bonaldo, Maria, de F., p.34*
*Bono, Hidemasa, p. 95*
*Boonruangrod, R., p. 105*
*Bourque*, Guillaume, p. 35*
*Boursot, Pierre, p. 72, 75*
*Brandsma, Janet, p. 39*
*Britten*, Roy, p. 1*
*Brosius*, Juergen, p. 5*
*Bruggmann, Rémy, p. 91*
*Brunsfeld, John, p. 15*
*BUI, Quynh Trang, p. 24*
*Burg, K., p. 105*
*Burke, William D., p. 14*
*Bush, Kiana A., p. 15*
*Cantrell, Michael, p.15*
*Capy*, Pierre, p. 46*
*Carninci, Piero, p. 38*
*Carter, C. C., p. 12*
*Cermak, Tomas, p. 57*
*Chalamcharla, Venkata, p. 40*
*Chen, Min, p. 39*
*Christensen, Shawn, p. 83, 84*
*Schröder*, Christiane p.67*
*Christley, Scott, p. 82*
*Bleidorn, Christoph p. 67*
*Chung, Ho-Ryun, p. 37*
*Ciaudo, Constance, p. 26*
*Clark, Marlena S., p. 80*
*Collins, Frank H., p. 82, 87, 89*
*Collins, K. L., p. 12*
*Colot*, Vincent, p. 26*
*Coros, Colin, p. 40*
*Costa, Jose, p. 39*
*Craig*, NL, p. 41*
*Cruaud, Corinne, p. 26*
*Cui, Feng, p. 36*
*Damert, Annette, p. 42*
*de la Chaux*, Nicole, p. 62*
*Deininger*, P., p. 16*
*Devor*, Eric J., p. 18, 111*
*Dod, Barbara, p. 72, 75*
*Dolezal, J., p. 105*
*Dolezel, Jaroslav, p. 73*
*Dorlhac De Borne, François, p. 24*
*Duprat, Elodie, p. 50*
*Dyer, Andrew, p. 39*
*Eickbush*, Thomas H., p. 14*
*Eickbush, Danna G., p. 14*
*Elgin, Sarah, p. 56*
*Emrich, Scott J., p. 87*
*Erickson, Issac K., p. 15*
*Eriksson, Nils-Einar, p. 76*
*Esteller, Manel, p. 26*
*Estep, Matt, p. 22*
*Estil, Jamie, p. 22*
*Evans, James, p. 13*
*Ewing, Adam D., p. 43, 109*
*Fan, Hua-Ying, p.11*
*Faulkner*, Geoffrey J., p. 38*
*Feschotte*, Cedric, p. 9, 79, 80, 103*
*Fluch, S., p. 105*
*Flutre*, Timothée, p. 32, 50, 85*
*Fraga, Mario, F., p. 26*
*Friedli, Marc, p. 53*
*Gambin, Anna, p. 49*
*Garcia-Perez, J. L., p. 12*
*Gilbert*, Clément, p. 79, 80*
*Gilbert*, Nicolas, p. 17*
*Gladyshev, Eugene A., p. 31*
*Gładysz, Mirosława, p. 49*
*Goldman, Stewart, p. 34*
*González*, Josefa, p. 20, 54*
*Gouvas, Eftychia, p. 66*
*Grandbastien*, Marie-Angèle, p. 24*
*Gray, Lucas T, p. 11*
*Greene, Wayne K., p. 113*
*Grimmond, Sean M, p. 38*
*Gross*, Stephen M., p. 102*
*Grynberg, M., p. 65*
*Grzebelus*, Dariusz, p. 49*
*Guex, Gaston-Denis, p. 58*
*Gundlach*, Heidrun, p. 91*
*Haberer, Georg, p. 91*
*Hancks, Dustin C., p. 43, 109*
*Harvey, Amanda, p. 71*
*Haussler, David, p. 2*
*Heitkam*, Tony, p. 64*
*Heredia, Fabiana, p. 26*
*Hidas, A., p. 106*
*Higashino*, Saneyuki, p. 95*
*Hobza, Roman, p. 57*
*Horton, Jason, p. 71*
*Hotz, Hansjürg, p. 58*
*Howell, Eric, p. 15*
*Hribova, Eva, p. 73*
*Hua-Van, Aurélie, p. 46*

*Hull*, Christi, p. 83*
*Huvos*, Piroska, p. 71*
*Ichiyanagi*, Kenji, p. 104*
*Isayeva, Darya, p. 66*
*Ivics*, Zoltán, p. 23*
*Izsvák, Zsuzsanna, p. 23*
*Janoušek*, Václav, p. 72, 75*
*Jern, Patric, p. 98*
*Johnson, Diana, p. 63*
*Jonker, Donald, p. 66*
*Jordà*, Mireia, p. 74*
*Jordan*, King, p. 10*
*Jurka, Jerzy, p. 100*
*Jurka, Matthew G., p. 100*
*Jursch, Tobias, p. 23*
*Kajikawa*, Masaki, p. 55*
*Kalitsis*, Paul, p. 97*
*Kapitonov*, Vladimir V., p. 30, 100*
*Karger, Amir, p. 60*
*Kaski, Samuel, p. 98*
*Kazazian Jr., Haig H., p. 43, 81, 109*
*Kejnovsky*, Eduard, p. 57, 73*
*Kennedy*, Ryan, C., p. 82, 87, 89*
*Keswani, Umeshkumar, p. 90*
*Keyes, Amanda, p. 15*
*Kiełbasa, Szymon M., p. 37*
*Kikuchi, Kunio, p. 95*
*Kissinger, Jessica, p. 103*
*Kitayama, Junko, p. 104*
*Klauninger, B., p. 105*
*Kluge, David, p. 71*
*Koga*, Akihiko, p. 52*
*Köhler, Frank, p. 58*
*Kopecky, D., p. 105*
*Kowalski, Diane, p. 39*
*Koyama, Hideki, p. 55*
*Kretova, Olga, V., p. 59*
*Kubat, Zdenek, p. 57*
*Kubešová, Lenka, p. 75*
*Kulpa, D. A., p. 12*
*Lage, José M., p. 39*
*Laten*, Howard, p. 66*
*Lehnert*, Stefan, p. 96*
*Leitch, Ilia J., p. 70*
*Lelley, T., p. 105*
*Leman, S.C., p. 93*
*Lenkov, Kapa, p. 20*
*Levine, David, p. 90*
*Lindorfer, Christi, p. 66*
*Lipatov, Mikhail, p. 20*
*Lisch, Damon, p. 101*
*Lizardi, Paul M., p. 39*
*Lobo, Neil F., p. 82*
*Loo, Peter Van, p. 96*
*Loreto, Elgion Lúcio, da Silva, p. 86*
*Lövgren, Anders, p. 76*
*Lowe, Craig, p. 2*
*Löwer, Johannes, p. 42*
*Löwer, Roswitha, p. 42*
*Ludwig*, Adriana, p. 86*
*Lynch, Michael, p. 78*
*Lysak, Martin A., p. 70*
*Ma, Jianxin, p. 22, 88*
*Maas, Jason, p. 71*
*Macas, Jiri, p. 57, 70, 73*
*Macpherson, J. Michael, p. 20, 54*
*Madey, Greg, p. 82*
*Mager*, Dixie L., p. 8, 112*
*Maksakova, Irina A., p. 8*
*Marquez, Claudia P., p. 107*
*Martienssen, Robert, p. 25, 51*
*Martin*, Sandra L., p. 13*
*Martín, Berta, p. 74*
*Martinez, Armando R., p. 15*
*Marynen, Peter, p. 96*
*Maunoury, Nicolas, p. 24*
*Maxfield, David, p. 79*
*Mayer*, Jens, p. 47*
*Mayer, Klaus F.X., p. 91*
*McCole, R.B., p. 68*
*Meselson, Matthew, p. 31*
*Ming, Kelly, p. 71*
*Mitra, Rupak, p. 41*
*Mogil, Lauren, p. 66*
*Moran*, J. V., p. 12*
*Morell, M., p. 12*
*Munclinger*, Pavel, p. 72, 75*
*Muszewska*, A., p. 65*
*Naito, Ken, p. 27*
*Navratilova, Alice, p. 73*
*Neumann*, Pavel, p. 73*
*Newby, Josh, p. 71*
*Newman, John C, p. 11*
*Nishihara*, Hidenori, p. 44*
*O'Shea, K. S., p. 12*
*Oakey*, R.J., p. 68*
*Oja, Merja, p. 98*
*Okada*, Norihiro, p. 6, 55*
*Oliver*, Keith R., p. 113*
*Olson, Michael R., p. 87*
*Orlando, Valerio, p. 38*
*Pace II, John K., p. 80*
*Paszewski, A., p. 65*
*Pavelitz, Tom, p. 11*
*Pecon-Slattery, Jill, p. 63*
*Peddigari, Suresh, p. 13*
*Peek, Andrew S., p. 18, 111*
*Peinado, Miguel A., p. 74*
*Permal*, Emmanuelle, p. 85*
*Petrov*, Dmitri A., p. 20, 54*
*Piazza, Carol Lyn, p. 40*
*Plötner, Jörg, p. 58*
*Pollock, David, p. 33*
*Poulain, Julie, p. 26*
*Pritham, Ellen J., p. 78, 103, 107*
*Quesneville*, Hadi, p. 24, 32, 50, 85*
*Raiz, Julija, p. 42*
*Rajaram, Veena, p. 34*
*Tiedemann, Ralph p. 67*
*Ramaswamy, Ram, p. 45*
*Rangwala*, Sanjida H., p. 81*
*Rawal*, Kamal, p. 45*
*Reiss, Daphne, p. 8*
*Reuter, Max, p. 8*
*Revay, T., p. 106*
*Rho*, Mina, p. 92*
*Richardson*, Aaron O., p. 94*
*Rodríguez, Jairo, p. 74*
*Rouault, Jacques-Deric, p. 46*
*Roudier, François, p. 26*
*Rubinstein, Jill, p. 39*
*Ruvolo, Maryellen, p. 60*
*Sabot*, François, p. 48, 108*
*Saffery, Richard, p. 97*
*Samollow, Paul B., p. 18, 111*
*Sarazin, Alexis, p. 26, 69*
*Sasaki, Clarence, p. 39*
*Sasaki, Hiroyuki, p. 104*
*Savageau*, Michael, p. 4*
*Schaack*, Sarah, p. 78*

*Schmidt, Thomas, p. 64*
*Schorn, , Andrea, p. 23*
*Schreiber, Robert, p. 58*
*Schuit, Frans C., p. 96*
*Schulz, R., p. 68*
*Schumann*, Gerald G., p. 42*
*Sciamanna, Ilaria, p. 19*
*Scott, Dan, p. 71*
*Scott, LuAnn, p. 15*
*Seal*, Ruth, p. 47*
*Sebastian, Aswathy, p. 61*
*Sentmanat*, Monica, p. 56*
*Shah, Rushi A., p. 102*
*Sharakhov*, I.V., p. 93*
*Sharakhova, M.V., p. 93*
*Shedlock*, Andrew M., p. 29*
*Singh*, Abanish, p. 90*
*Sirotin, Michael V., p. 36*
*Slotkin*, R. Keith, p. 25*
*Smit*, Arian, p. 28*
*Smith, Christina, p. 34*
*Smith, Dorie, p. 40*
*Soares*, Marcelo B., p. 34*
*Somervuo, Panu, p. 98*
*Spadafora*, Corrado, p. 19*
*Sperber, Göran, p. 76, 98*
*Spitz, François, p. 53*
*Hartmann, Stefanie, p. 67*
*Steinbauerova, Veronika, p. 73*
*Stojanovic, Nikola, p. 90*
*Sun, Xueguang, p. 39*
*Suzuki, Jun, p. 55*
*Szpakowski*, Sebastian, p. 39*
*Takeda, Shunichi, p. 55*
*Tang, Haixu, p. 92*
*Tanurdzic*, Milos, p. 25, 51*
*Tchurikov*, Nickolai A., p. 59*
*Teixeira, Felipe Karam, p. 26*
*Tempel, Sebastien, p. 110*
*Thilakarathne, Pushpike, J., p. 96*
*Thompson*, Blaine, p. 84*
*Tomilin, N., p. 114*
*Tomita, Tadanori, p. 34*
*Tran, Joseph, p. 24*
*Tu, Z., p. 93*
*Unger*, Maria F., p. 82, 87, 89*
*Usmanova, N., p. 114*
*Uzzell, Thomas, p. 58*
*Valente, Vera Lúcia, da Silva, p. 86*
*Vaughn, Matthew, p. 25, 51*
*Vaury*, Chantal, p. 7*
*Verbeke, Geert, p. 96*
*Vingron, Martin, p. 37*
*Vitte, Clementine, p. 22*
*Vitullo, Patrizia, p. 19*
*Vodkin, Lila O., p. 99*
*Voinnet, Olivier, p. 26*
*Vyskot, Boris, p. 57, 73*
*Wagner, Andreas, p. 62*
*Wainer, Michael, p. 71*
*Walisko, Oliver, p. 23*
*Wallace, N., p. 16*
*Walters*, Kathryn, p. 63*
*Wang, Hui, p. 42*
*Wang, Min, p. 34*
*Wang, Ting, p. 2*
*Wang, Yongming, p. 23*
*Weiner*, Alan M, p. 11*
*Wessler, Susan R., p. 27, 94*
*Wichman*, Holly A., p. 15*
*Widmer, Alex, p. 57*
*Williamson, Valerie M., p. 102*
*Wincker, Patrick, p. 26*
*Wing, Rod A., p. 61*
*Wolf, Abby, p. 78*
*Wood, A.J., p. 68*
*Wright, LaBianca, p. 66*
*Xia, A., p. 93*
*Xie, Hehuang, p. 34*
*Xing, Jinchuan, p. 42*
*Yamaguchi, Katsumi, p. 55*
*Yang, Lixing, p. 22*
*Yu, Yeisoo, p. 61*
*Zabala*, Gracia, p. 99*
*Zemojtel*, Tomasz, p. 37*
*Zhang*, Qu, p. 60*
*Zhang*, Ying, p. 8, 112*
*Zhang, Feng, p. 27*
*Zhang, Lili, p. 81*
*Zhao, Hongyu, p. 39*
*Zhou*, Xianzheng, p. 77*
*Zhou, Jun, p. 14*
*Zhurkin, Victor B., p. 36*
*Zuccolo*, Andrea, p. 61*

# SPEAKER ABSTRACTS

# History and relationships of TEs

*Roy Britten*

California Institute of Technology, Pasadena, CA 91125

Start with the first recognition of repeated sequences. Proving that agar was not an accelerator of DNA reassociation and showing the rate of reassociation of the mouse satellite.

Recent sequence comparisons using Wublast2. A probe was made of the best matching regions of the human genome to the complete set of vertebrate TEs. This probe (WVH) consists of 1800 human sequences. Comparing WVH with the human genome shows sequence similarity to about 69% of human sequence. This was confirmed by comparison with repeat masked human DNA from UCSC. The additional human TE regions beyond the customary 45 to 50% (e.g. Landers *et al*) are presumably the “fossils” of ancient TE insertions that are still recognizable.

The human genome is made up of “fossil” sequences of genes as well as TEs. A set of human coding sequences identified as refMrnas consists of the mRNAs from 27,000 genes. Using Wublast the refMrnas were compared with the human genome. Wublast recognizes imperfectly matching sequences with greater than 50% match. The result was that 87% of the human genome matches these mRNA sequences with a range of precision ranging from 50 to 100%. There is an approximately flat distribution of percent match. The best model is that the genome is mostly made up sequences that are similar to those of present day mRNAs but with a wide range of divergences. 96% of the TE locations were within the regions of similarity to the mRNAs

Finally some comments on the concept that the origin of eukaryotes was the result of the evolution of the TEs that were capable of help in the formation and control of the large number of novel genes required for the origin and evolution of the eukaryotes.

## Evidence that transposons shaped vertebrate gene regulatory networks

*David Haussler*

Center for Biomolecular Science & Engineering, UC Santa Cruz, Santa Cruz, CA 95064

Roy Britten and Eric Davidson proposed nearly 40 years ago that the spread of repetitive DNA may play a key role in the evolution of gene regulatory networks, expanding on an earlier line of investigation initiated by Barbara McClintock. We will discuss recent evidence that supports this hypothesis.

Joint work with Ting Wang and Craig Lowe.

# A study of co-option in the human genome

*Gill Bejerano*

Department of Developmental Biology and Department of Computer Science, Stanford University, CA 94305-5329, USA

To probe the role of transposable elements in gene regulation, we examine related co-option events to see whether and how their intrinsic properties contribute to their novel function(s).

## Molecular Phenotypes in the Design Space of Prophage Lambda

*Michael Savageau*

Department of Biomedical Engineering, University of California at Davis, 451 Health Sciences Drive, Davis, Ca 95616, USA

One of the major unsolved problems of modern biology is deep understanding of the relationship between the information encoded in the genome of an organism and the phenotypic properties manifested by that organism. Although this problem is often portrayed as if the task were to find a more or less direct link between these two levels, on closer examination the relationship is far more layered and complex. Detailed studies at each of the many intervening levels have revealed an enormous diversity of molecular elements and circuits. We are just beginning to understand the functional implications of these variations and to grasp the factors that have influenced their evolution. At each of the intervening levels there are major challenges to characterizing the phenotypes and elucidating the design principles. Although there are some intuitive notions of what is meant by phenotype at the level of the organism, it is far from clear what this term means at the biochemical level, and our understanding of molecular design principles is in its infancy. My colleagues and I have previously described design principles that are readily reveled by representation of molecular systems in an appropriate design space. I will describe a generic approach to the construction of such a design space in which qualitatively distinct phenotypes can be identified and counted, their fitness analyzed and compared, and their tolerance to change measured. I will illustrate the approach by treating regulation of prophage lambda induction.

# The retro-look of genomes

*Brosius J*

Institute of Experimental Pathology, ZMBE, University of Münster, Von Esmarch-Str. 56, 48149 Münster, Germany.

RNA.world@uni-muenster.de

Many multicellular organisms generate large amounts of superfluous DNA, chiefly by the unabating conversion of RNA to DNA by reverse transcription. When we extrapolate, not 45% but (almost) the entire genome must be derived from transposed elements most of which are RNA-derived. The majority of these sequences is and will remain devoid of function, evolve neutrally and, after 150-250 million years of mutational onslaught, will not be discernible any more. Exaptation (co-optation), for example, as (parts of) protein coding genes, genes encoding non-protein coding RNAs (npcRNAs), as well as regulatory elements can occur at any stage of decay. Often, retroposed elements are not "ready to use" cassettes with functional elements but require further mutational changes that can happen almost immediately after insertion, or tens of millions years later. An example for exonization of an Alu element is given that experienced a number of changes over time and whose inclusion into mature alternative mRNAs even is contingent upon Adenosine → Inosine editing at the RNA level (1).

Concerning unusual SINE elements, a case of an efficiently retroposed small nucleolar RNA (snoRNA) fused to a segment of an RTE element in platypus is presented (2).

Finally, by using L1MB5 retroposed elements as insertional markers, a phylogenetic tree is proposed that suggests a soft trifurcation of Boreotheria, Afrotheria and Xenartra at the very base of placental mammal evolution (3).

1) Möller-Krull, M., Zemann, A., Roos, C., Brosius, J., Schmitz, J. (2008) Beyond DNA: RNA editing and steps toward Alu-exonization in primates. J. Mol. Biol. 382, 601-609.
2) Schmitz, J., Zemann, A., Churakov, G., Kuhl, H., Grützner, F., Reinhardt, R., Brosius, J. (2008) Retroposed SNOfall – a mammalian-wide comparison of platypus snoRNAs. Genome Res. 18, 1005-1010.
3) Kriegs, J.O. (2007) Retroposed elements- witnesses of the evolutionary history of placental mammals. Thesis. Westfälische Wilhelms University of Münster, Germany.

## Multiple SINE insertions made our brain mammalian?

*Norihiro Okada*

Tokyo Institute of Technology, Tokyo, Japan

Retroposons such as short interspersed elements (SINEs) and long interspersed elements (LINEs) are the major constituents of higher vertebrate genomes. Although there are many examples of retroposons acquiring function, none have been reported to be involved in the creation of morphological innovations specific to a certain taxonomic group. We previously characterized ~100 copies of a SINE family, AmnSINE1, present as a part of conserved non-coding elements (CNEs) in mammalian genomes, proposing that they have acquired genomic functionality, or were exapted after their retroposition, in a common ancestor of mammals to gain characteristics specific to mammals (). Here we refined 124 total loci, several of which were analysed further. Using a mouse enhancer assay, we clearly demonstrate that one SINE locus, AS071, 230 kbp from the gene Fgf8 (fibroblast growth factor 8), is an enhancer that recapitulates Fgf8 expression in two forebrain regions, namely diencephalon and hypothalamus of the developing forebrain. Our gain of function analysis revealed that expression of FGF8 in the diencephalon controls patterning of thalamic nuclei, which are a relay center of the neocortex, suggesting its role in mammalian specific forebrain patterning. Furthermore, we demonstrated that the locus, AS021, 392 kbp from the gene Satb2, controls gene expression in lateral telencephalon, which is suggested as one of the signaling centers during development (2). Recently, we characterized several more loci which function as enhancers for genes that are expressed in forebrain, suggesting important roles of SINEs in developing the neuronal network highly organized specific to mammals and introduced by exaptation of AmnSINE1 in a common ancestor of mammals.

(1) Functional non-coding sequences derived from SINEs in the mammalian genome. (2006) Nishihara, Smit and Okada. Genome Res. 16, 864-874

(2) Possible involvement of SINEs in mammalian-specific brain formation. Sasaki et al. (2008) Proc. Nall. Acad. Sci. USA 105, 4220-4225

# Insulators brought by TE contribute to gene regulation in relationship with nuclear architecture

*Chantal Vaury*

Unité INSERM U384, 63000 Clermont-Ferrand, France

Eukaryotic regulation of gene transcription is controlled by cis-regulatory elements that reside upstream, downstream or within introns of genes. However, the connection between the organization of chromatin inside the nucleus and regulation of gene expression has also emerged as a key component to establish a proper differentiated and developmental state in eukaryotes. Such a compartmentalization implies the existence of regulatory elements that enforce the functional independence of distinct chromatin domains. Chromatin insulators or boundary elements have been implicated in the establishment of this compartmentalization, as they may be involved in segregating independent chromosomal domains.

We have identified an insulator in the Long Terminal Repeat (LTR) of Idefix, a retroelement from Drosophila melanogaster. When positioned between an enhancer and a promoter, Idefix insulator is capable of disrupting enhancer-promoter communication, a property referred to as enhancer-blocking. This enhancer blocker effect is accomplished without inactivation of the intrinsic properties of any of the regulatory elements, implying that the insulator disrupts signaling between the enhancer and the promoter. Combining transgenic experiments and three-dimensional fluorescent in situ hybridization (3D IFISH)-immunoassays, we found that the presence of two copies of Idefix insulator increases the insulator ability of just one. However, when these two copies are fused to two copies of a fragment taken from the 5'UTR of Idefix, the enhancer-blocking activity of Idefix LTR is completely lost. The gene is then correctly expressed as if the communication between the enhancer and its target gene was re-established. Interestingly, we found that the transcriptional activity of the gene is correlated with its displacement from within the nucleus toward the nuclear periphery. The molecular mechanisms that might allow reactivation of the target gene by the enhancer when they are directed to the nuclear membrane will be discussed in light of our last data.

Our results put in light mechanisms by which transposable elements and their cis-regulatory sequences contribute to genome control and phenotypic variations by acting on nearby regulatory sequences, by establishing long-range interaction and three dimensional organization of chromatin within the nucleus

# Complex epigenetics of mammalian endogenous retroviruses

*Dixie L. Mager*1, Irina A. Maksakova1, Max Reuter2, Ying Zhang1 and Daphne Reiss1*

1Terry Fox Laboratory, BC Cancer Agency, Vancouver, BC, Canada and 2Univ. College London, London, UK

Endogenous retroviruses (ERVs) affect host genes by providing transcriptional regulatory signals and disrupting splicing. Since ERVs are targets of epigenetic silencing, their epigenetic state may also influence host gene transcription. For instance, cases have been reported of mouse ERVs exerting variable effects on genes in isogenic animals, with the magnitude of the effect depending on the variable state of ERV methylation. Such cases have prompted the suggestion that ERVs or other TEs may be epigenetic mediators of phenotypic variability. However, the prevalence of variably silenced ERVs and the underlying mechanisms of this phenomenon have not been explored. Our work on one human ERV family did not reveal evidence of variable epigenetic silencing as we found that the methylation pattern for a given insertion is similar between cells and individuals, suggesting that ERV methylation and the extent of its variability can themselves vary in a systematic way. We have now investigated DNA methylation patterns of the active MusD/ETn family of mouse ERVs and have observed that patterns of methylation depend on characteristics of the subfamily to which copies belong. One characteristic is the internal (non-LTR) sequence, with autonomous MusD elements attracting more silencing marks than non-autonomous ETnII elements, despite their nearly identical LTRs. Another factor seemingly involved in establishment of LTR methylation is the strength of the LTR promoter. We found that variable methylation is a frequent phenomenon but restricted to the ETnII subfamily that differs from the ETnI subfamily primarily by their strong LTR promoters. We propose that the excess of variance in methylation observed for ETnII is linked to their higher affinity for transcription factors (TFs). In this scenario, TFs protect LTRs from methylation while small RNAs induce it, and actions of these opposing forces lead to a variegated pattern of methylation, as they prevail in a stochastic way.

## Mammalian transposable elements and the emergence of lineage-specific functions

*Cedric Feschotte*

University of Texas at Arlington, Department of Biology, Arlington, TX 76019, USA

Change and innovation in regulatory systems are thought to be of major importance for the emergence of biological novelties and the evolution of species. How these changes have come about remains poorly understood, but it is clear that they have necessitated the rewiring of existing genetic networks and the appearance of new regulatory proteins and cis-acting DNA elements. In this talk, I will revisit earlier theoretical models (e.g. 1,2) invoking that genomic repeats, and in particular transposable elements, have provided a rich supply of material for the assembly and tinkering of regulatory systems (for review, ref. 3). I will present the results of evolutionary, genetic and biochemical analyses supporting the idea that TEs have been a continuous source of functional sequences, both coding and non-coding, prior to and throughout mammalian evolution. Evidence is accumulating that the bulk of functional sequences derived from TEs have fueled regulatory evolution. I will argue that mammalian DNA transposons occupy a disproportionately important place in lineage-specific exaptations by virtue of their life cycle, their propensity for horizontal transmission and some functional predispositions.

1) Britten, R. J. & Davidson, E. H. (1971) Q. Rev. Biol. 46: 111–138
2) Georgiev G.P. (1984) Eur. J. Biochem. 145:203-220
3) Feschotte C. (2008) Nat. Rev. Genet. 9:397-407

## Transposable elements, chromatin and gene regulation

*King Jordan*

Georgia Institute of Technology, Atlanta, Georgia 30332

Transposable elements (TEs) constitute a vast percentage of eukaryotic genomes, and they influence the function of their host genomes in a number of different ways. We have found a connection between human TEs and the epigenetic regulation of chromatin represented by two major mechanisms, and we show that these TE-related epigenetic processes have a demonstrable effect on gene expression in both normal and cancerous cells. First, TEs bind nucleosomes with differential affinities thereby modulating access to genomic DNA. Clusters of genes with similar TE-promoter profiles are co-regulated and show distinct tissue-specific patterns of gene expression. Secondly, TEs are enriched in various histone tail modifications, and combinations thereof, that specify different chromatin states. Histone tail modification combinations that repress tissue-specific gene expression are depleted for certain TE families, particularly ancient families, whereas combinations of modifications that active expression are enriched for the same TE families. Thus, TEs are targets, and perhaps initiators, of epigenetic modifications that have functionally relevant effects on human gene expression.

## Human PGBD3: a piggyBac transposon that is both good and bad for us?

*Arnold D Bailey, John C Newman, Lucas T Gray, Tom Pavelitz, Hua-Ying Fan1, and Alan M Weiner**

Department of Biochemistry, School of Medicine, University of Washington, Seattle, WA; 1Fox Chase Cancer Center, Philadelphia, PA

Cockayne syndrome (CS) is a devastating progeroid syndrome usually caused by mutations in the Cockayne syndrome Group B gene (CSB, aka ERCC6) which encodes a SWI/SNF-like DNA-dependent ATPase and chromatin-remodeling protein required for transcription-coupled repair (TCR) of UV-induced and oxidative DNA damage (1). We recently found that a piggyBac transposable element called PGBD3 integrated into intron 5 of the 22 exon CSB gene >43 Mya before marmosets diverged from humans. As a result, the CSB gene generates three proteins: CSB, a CSB-PGBD3 fusion protein in which the first 5 exons of CSB are alternatively spliced to the PGBD3 transposase, and solitary PGBD3 transposase transcribed from an internal promoter in exon 5 (2). Genetic and evolutionary arguments suggest that this CSB-PGBD3 fusion protein is advantageous in the presence of functional CSB protein, but harmful in its absence. Using a host cell reactivation assay, we now show that introduction of the CSB-PGBD3 fusion protein into the CSB null cell line UVSS1KO facilitates repair of UV-damaged DNA, but interferes with repair of oxidatively damaged DNA. As repair of accumulated cytotoxic DNA damage is thought to play a central role in CS, these data argue that the conserved CSB-PGBD3 fusion protein plays a role in both health and CS disease. Importantly, piggyBac3 also gave rise to 1,000 nonautonomous internally-deleted piggyBac3 elements called MER85s, last mobilized by piggyBac3 about 35 Mya. We find that the CSB-PGBD3 fusion protein and solitary PGBD3 transposase bind to many consensus MER85s, in vitro and in vivo, and that many MER85 elements are located within 1 Mb of CSB-regulated genes (1,2). We propose that the CSB-PGBD3 fusion protein mediates a MER85-based regulatory network, much as the SETMAR protein binds to dispersed mariner repeats (Cordaux et al., 2006).

(1) Newman, Bailey, and Weiner (2006) Cockayne syndrome group B protein (CSB) plays a general role in chromatin maintenance and remodeling. Proc Natl Acad Sci 103, 9613-9618.

(2) Newman, Bailey, Fan, Pavelitz, and Weiner (2008) An abundant evolutionarily conserved CSB-PiggyBac fusion protein expressed in Cockayne syndrome. PLoS Genet 4, e1000031.

# Epigenetic silencing of LINE-1 retrotransposition events in human embryonic carcinoma cell lines

*J. L. Garcia-Perez1, M. Morell2, D. A. Kulpa3, C. C. Carter4, K. L. Collins3,4,5, K. S. O'Shea2, and J. V. Moran*1,3,4,6*

Departments of Human Genetics (1), Cell and Developmental Biology (2), Internal Medicine (3), Cellular and Molecular Biology (4), and Microbiology and Immunology (5), University of Michigan Medical School, Ann Arbor MI. Howard Hughes Medical Institute (6)

Human LINE-1 retrotransposons are abundant mobile genetic elements that comprise approximately 17% of genomic DNA. Ongoing LINE-1-mediated retrotransposition events continue to impact the genome and are estimated to be responsible for ~1/1000 disease producing mutations in man. Despite their mutagenic potential, relatively little is known about host mechanisms that act to regulate and/or restrict LINE-1 retrotransposition. In this presentation, we will describe a series of experiments demonstrating that engineered LINE-1 retrotransposons, as well as a zebrafish LINE-2 retrotransposon, can be efficiently silenced in various human cultured embryonic carcinoma (EC) cell lines either during or shortly after integration. Remarkably, treating cells containing silenced LINE-1 retrotransposition events with histone deacetylase inhibitors leads to the reactivation of a retrotransposed indicator cassette, and chromatin immunoprecipitation experiments suggest that silencing of the retrotransposed cassette is associated with post-translational histone modifications. We also present evidence that the silencing of human LINE-1 elements in the PA-1 embryonic carcinoma cell line is attenuated upon differentiation and that this process appears to differ from previously reported retroviral silencing mechanisms in mouse embryonic stem cells and mouse embryonic carcinoma cells. We speculate that the epigenetic silencing of de novo LINE-1 insertions could represent a novel host mechanism to restrict retrotransposition in mammalian genomes.

# Mutational Analyses of ORF1p Function in LINE-1 Retrotransposition

*Sandra L. Martin*1,2, Suresh Peddigari1 and James Evans1*

(1) Department of Cell and Developmental Biology and (2) Molecular Biology Program
University of Colorado School of Medicine

The LINE-1 (L1) element of mammals encodes two proteins that are essential for retrotransposition, ORF1p and ORF2p. Early sequence comparisons revealed the homology of ORF2p to known endonucleases and reverse transcriptases; subsequent mutational analyses confirmed these roles for ORF2p in L1 retrotransposition. The lack of homology of the ORF1p to proteins with known function, however, has made its role in L1 retrotransposition enigmatic. Mouse ORF1p forms a stable homotrimer via an N-terminal, coiled-coil domain and binds nucleic acids via a C-terminal domain that is conserved among ORF1 proteins from other mammalian L1s. Protease sensitivity studies revealed two structured regions in the C-terminal domain of mouse ORF1p, M and RID, interrupted by an unstructured domain (linker). In vitro, the protein has two activities: ORF1p binds RNA with low nanomolar affinity and is a robust nucleic acid chaperone. Few ORF1p mutations have been studied in detail, but mutations in both the coiled-coil domain and the RID can have dramatic effects on nucleic acid chaperone activity and retrotransposition without affecting the affinity of the protein for RNA. Here, we report the effect of mutations of 12 additional conserved residues in the M domain, flexible linker and RID domains on retrotransposition. The in vitro biochemical properties of those with significant effects on retrotransposition are under investigation and will be described in detail. Current evidence is consistent with there being at least two independent nucleic acid interaction domains in mouse L1 ORF1p.

# Regulating the expression of R2 elements within the nucleolus

*Danna G. Eickbush, Jun Zhou, William D. Burke and Thomas H. Eickbush**

Department of Biology, University of Rochester, Rochester, New York, 14627

R2 non-LTR retrotransposable elements exclusively insert into the tandemly organized 28S rRNA genes. R2 elements have persisted for hundreds of millions of years, suggesting that organisms find it extremely difficult to rid their rDNA loci of these elements. Our studies of R2 regulation have indicated that in Drosophila simulans active retrotransposition is correlated with the level of R2 transcripts. Nuclear run-on transcription experiments revealed the R2 transcript levels correlate with differential transcription of R2-inserted genes rather than degradation of the R2 transcripts. Finally, crosses between active and inactive lines indicated that the major component of this transcriptional control is linked to the rDNA locus. To determine how R2 elements survive in a natural population, rDNA loci were sampled from two populations of animals. In both populations, about half of the rDNA loci supported no R2 transcripts and no retrotransposition and half supported variable levels of R2 transcripts and retrotransposition events. Structural analysis of the rDNA loci revealed that R2 activity did not correlate with R2 number or rDNA locus size. Instead loci with no R2 activity had at least one large contiguous block of rDNA units free of R2 insertions. We propose a model in which only a small domain of the rDNA locus is activated by cells for transcription. If R2 inserted units are excluded from this domain, then R2 activity is prevented. However, the recombinations that result in the concerted evolution of the rDNA locus continually redistribute the individual units. Thus R2-inserted units are periodically reactivated for transcription enabling the long-term survival of active R2.

# Comparative Biology of L1 Elements

*Holly A. Wichman*, Michael Cantrell, LuAnn Scott, Amanda Keyes, Eric Howell, Issac K. Erickson, Kiana A. Bush, Armando R. Martinez and John Brunsfeld*

Department of Biological Sciences and Initiative for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844-3051

Most of what we know about L1s in mammals comes from a few well-studied systems and from whole genome sequencing projects. From this we are left with the following impressions: L1s have persisted in the mammalian genome since before the divergence of placental mammals from marsupials, but are not found in monotremes. There have been differences of opinion concerning the current activity level of L1 in some species, but in general they are thought to have remained active in most lineages. There many potentially active L1 sequences within a mammalian genome, yet within each speices L1s do not diverge; rather they exist as a single lineage over millions of years of evolution. Finally, they have an unusual base composition. The two L1 genes are unusually adenine-rich – about 40% A on the coding strand. To examine the generality of these impressions, we have undertaken a broad survey of L1 activity in both placental mammals and marsupials. We have examined nearly 150 species representing all orders of placental mammals and nearly all orders of marsupials. We find that persistent L1 activity is the rule, but that L1 extinctions have occurred multiple times in the history of mammals. Multiple active lineages are not as rare as previously thought, but single active lineages and A-rich genes are a common feature of L1s in both placental mammals and marsupials.

## Cellular responses to damage by non-LTR retroelements

*N. Wallace, V.P. Belancio, and P. Deininger**

Tulane Cancer Center, New Orleans, LA

LINE-1 elements comprise 17% of the human genome, which translates into about 5x105 L1 copies, the majority of which are defective. The expression and activity of these elements contribute to human genomic instability both through insertional mutagenesis, as well as causation of double-strand breaks in DNA. Many people make the assumption that these elements are expressed solely in the germline, as this is the location that is most critical to their life cycle and evolution. We have shown that some somatic tissues demonstrate high levels of L1 expression, bringing up the strong possibility that they can adversely affect human health during the lifetime of the individual.

Because of their negative impact on the genome and human health, these elements have co-evolved with cells to have a plethora of regulatory mechanisms to ensure that they are not too deleterious to their host. It is well known that their transcription is heavily regulated by methylation. In addition, their transcripts are subject to premature polyadenylation and aberrant splicing events that severely regulate the amounts of full-length transcripts produced. Similarly, there is evidence that RNAi and surveillance by Apobec 3 proteins may also regulate L1 activity. We have now shown that the ERCC1/XPF heterodimer that is normally associated with the nucleotide excision DNA repair pathway also is highly effective at removing the partially inserted L1 cDNA from the genome. Thus, surveillance against L1 insertion occurs at almost all levels of its life cycle. In addition, there appear to be a wide range of cellular responses to the DNA damage caused by L1 elements that lead to cell cycle checkpoints, cell death, cellular senescence and in many ways resulting in an altered response to L1 expression.

## LINE-1 mediated mobilization of snRNA

*Nicolas Gilbert*

Institut de Génétique Humaine, CNRS, UPR 1142, 34396 Montpellier cedex 5, France

Long Interspersed Element-1 (LINE-1 or L1) sequences comprise ~17% of human DNA and ongoing L1 retrotransposition continues to impact genome evolution. The L1-encoded proteins also can mobilize other cellular RNAs (e.g., Alu retrotransposons, SVA retrotransposons, and U6 snRNAs), which comprise ~13% of human DNA. Recently, we have demonstrated that the trans-mediated mobilization of non-L1 RNAs can occur by either template choice or template switching mechanisms.
Here, by in silico analysis of primate genomes, we have observed variability in the insertion signature of L1-mediated snRNA pseudogenes. This variability suggests recruitment of cellular RNA at various steps of the retrotransposition cycle, in the cytoplasm or in the nucleus. Moreover, by comparing different mammalian genomes (from placental mammals to monotreme) we observed a wide variability of the L1 retrotransposition dynamics. This could reflect divergence in efficiency, processivity and/or affinity of the L1 RNP complex to perform retrotransposition. Host factors could also influence the variability between genomes.

# A microRNA incubator on the marsupial (Monodelphis domestica)X-chromosome was created via L1 transposon-mediated serial duplication

*Eric J. Devor*(1), Paul B. Samollow(2) and Andrew S. Peek(1)*

(1)Molecular Genetics and Bioinfomatics, Integrated DNA Technologies, Coralville, Iowa 52241
(2)Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas 77843, USA

The origin of microRNAs (miRNAs), small (21nt to 23nt) non-coding regulatory RNAs, is a topic of interest in evolutionary biology as well as in functional genomics. There is mounting evidence to support a view that these and, perhaps, other non-coding RNAs arise from transposons though the mechanisms that are, as yet, unclear (Borchert et al., 2006). We have discovered a closely related family of 39 microRNAs spanning just over 100Kb of the X-chromosome of the grey, short-tailed South American opossum Monodelphis domestica (Mikkelsen et al., 2007). Detailed analysis of this region indicates that this family was created via a series of duplication events that were mediated by the presence of L1 transposons flanking the pre-miRNA. Further, there is some evidence that the ancestral miRNA itself evolved out of an L1 sequence.

Here we present the complete anatomy and evolution of this miRNA family including evidence that these miRNAs are in the process diverging and that at least two have already become pseudogenized.

Borchert GM, Lanier W, and Davidson BL 2006 Nat Struct Mol Biol. 13:1097-101.
Mikkelsen TS, Wakefield MJ, Aken B, et al. 2007 Nature 447: 167-177.

# Functional roles for LINE1-encoded reverse transcriptase in embryonic development and tumor progression

*Ilaria Sciamanna, Patrizia Vitullo and Corrado Spadafora**

Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy. Phone: +39 06-49903117; Fax: +39 06-49903672; e-mail: cspadaf@tin.it

In mammalian cells, endogenous reverse transcriptase (RT) is encoded by two families of retrotransposons, i.e. LINE-1 (Long Interspersed Nuclear Elements) and endogenous retroviruses (ERVs).

Growing evidence from our laboratory suggest that LINE-1-encoded RT plays regulatory roles in early embryonic development and in cell transformation.

We have inhibited the endogenous RT activity using two pharmacological RT inhibitors currently employed in AIDS treatment, nevirapine and efavirenz; we have found that this causes early developmental arrest at 2- or 4-cell stages. A similar arrest was also observed after down-regulation of expression of a highly expressed LINE-1 family by microinjecting LINE-1 antisense oligonucleotides into mouse zygotes. Embryo arrest by either drug- or antisensense-mediated RT inhibition is associated with subverted gene expression profiles.

In parallel studies, we have inhibited RT activity in a variety of human cancer cell lines, again using either RT inhibitory drugs or RNA interference-mediated silencing of a highly active human LINE-1 family; both approaches yielded a significant decrease in cell proliferation and promoted differentiation. Concomitant with this, we noticed global modifications in chromatin architecture, associated with profound alterations in both gene and miRNAs expression profiles. On discontinuation of RT inhibition, however, transformed cells resumed their original proliferation rate and dedifferentiated features, suggesting that RT acts at an epigenetic level. RT inhibitory treatment also proved effective in antagonizing the progression of human tumors inoculated in nude mice, yet tumor progression was resumed upon discontinuation of the treatment.

Together, these data indicate that a RT-dependent epigenetic machinery acts as a global regulator of gene and miRNA expression in cell growth and suggest that endogenous RT may be regarded as a novel target in anti-cancer therapy.

## High Rate of Recent Transposable Element-Induced Adaptation in Drosophila melanogaster

*Josefa González*, Kapa Lenkov, Mikhail Lipatov, J. Michael Macpherson, Dmitri A. Petrov**

Department of Biology, Stanford University, Palo Alto, California, United States of America

Although transposable elements (TEs) are known to be potent sources of mutation, their contribution to the generation of recent adaptive changes has never been systematically assessed. In this work, we conduct a genome-wide screen for adaptive TE insertions in Drosophila melanogaster that have taken place during or after the spread of this species out of Africa. We determine population frequencies of 902 of the 1,572 TEs in Release 3 of the D. melanogaster genome and identify a set of 13 putatively adaptive TEs. These 13 TEs increased in population frequency sharply after the spread out of Africa. We argue that many of these TEs are in fact adaptive by demonstrating that the regions flanking five of these TEs display signatures of partial selective sweeps. Furthermore, we show that eight out of the 13 putatively adaptive elements show population frequency heterogeneity consistent with these elements playing a role in adaptation to temperate climates. We conclude that TEs have contributed considerably to recent adaptive evolution (one TE-induced adaptation every 200–1,250 y). The majority of these adaptive insertions are likely to be involved in regulatory changes. Our results also suggest that TE-induced adaptations arise more often from standing variants than from new mutations. Such a high rate of TE-induced adaptation is inconsistent with the number of fixed TEs in the D. melanogaster genome, and we discuss possible explanations for this discrepancy.

Citation: Gonzalez J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA (2008) High rate of recent transposable element–induced adaptation in Drosophila melanogaster. PLoS Biol 6(10): e251.

# Mobile elements and primate genomic variation

*Mark A. Batzer**

Department of Biological Sciences, Louisiana State University, 202 Life Sciences Building, Baton Rouge, Louisiana 70803, USA

Primate mobile elements (SINEs, LINEs and SVA) belong to discrete subfamilies that can be differentiated from one another by diagnostic nucleotide substitutions. An analysis of several recently integrated mobile element lineages was undertaken to assess mobile element associated primate genomic diversity. Our screening of the mobile elements resulted in the recovery of a number of "young" Alu, L1 and SVA elements with different distributions throughout the primate lineage. Many of the mobile elements recovered from the human genome were restricted to the human lineage, with some elements that were polymorphic for insertion presence/absence in diverse human populations. Some of the mobile elements recovered from the human lineage also resided at orthologous positions in non-human primate genomes. Sequence analysis demonstrated that these mobile elements were the products of gene conversion events of older pre-existing elements, independent parallel forward insertions of older elements in the same short genomic region, or authentic shared phylogenetic characters. The distribution of Alu, L1 and SVA elements throughout various primate genomes makes them useful tools for resolving human population genetic relationships and non-human primate systematic relationships. We have also identified genomic deletions associated with the retrotransposition and insertion of recently integrated mobile elements in primate genomes and with post insertion recombination events between mobile elements. These genomic deletions are another source of mobile element associated genetic variation within the primate lineage.

# Transposon-Host Cell Interactions in the Regulation of Sleeping Beauty Transposition

*Zoltán Ivics*, Oliver Walisko, Tobias Jursch, Andrea Schorn, Yongming Wang, Zsuzsanna Izsvák*

Max Delbrück Center for Molecular Medicine, 13092 Berlin, Germany

The mobility of transposable elements in natural populations is usually strictly regulated in order to preserve genomic stability. It is believed that the fine control of transposon movement is brought about by both transposon- and host cell-encoded factors and mechanisms. The Sleeping Beauty (SB) transposon is a reconstructed element, the first ever shown to be active in any vertebrate-derived cell. SB not only represents a powerful gene vector system for genomic manipulations in vertebrate species, but has also been serving as a useful experimental system to address transposon-host cell interactions at the molecular level. We have established that, in addition to the element-encoded transposase, cellular factors are involved in SB transposition and its regulation. The transposase has to be expressed, it needs to interact with the transposon DNA, a complex in which excision takes place has to form and the excised transposon needs to interact with the target DNA for insertion. In the context of a living cell, all of these processes are under regulation of the transcriptional apparatus, chromatin structure, accessory proteins that influence DNA topology, factors that interact with the transposase and modulate cellular mechanisms including the cell-cycle and DNA repair as well as DNA-binding proteins that tether the transpositional complex to certain chromosomal sites. The Sleeping Beauty transposon has been, and will continue to be, a major experimental tool to elucidate many of these mechanisms.

# The evolution, activities and specificities of transposable elements in grass genomes

*Jeff Bennetzen1, Regina Baucom*1, Matt Estep2, Jamie Estil2, Jianxin Ma1, Clementine Vitte1 and Lixing Yang1*

Departments of Genetics1 and Plant Biology2, University of Georgia, Athens, GA 30602, USA

Grass genomes vary tremendously in genome composition, and most of this variation is caused by differences in the abundance of transposable elements (TEs). We have characterized the processes and rates of TE amplification and removal in grasses, with special focus on LTR retrotransposons and Helitrons in Zea and Oryza. The results presented will describe the great diversity of these elements in the complex maize genome, will indicate lineage-specific and region-specific differences in activity, will examine the nature of selection on individual elements, and will exemplify general TE effects on their host genomes.

## Transposable elements and TE-gene associations in the transcriptome of solanaceae species

*BUI Quynh Trang1, MAUNOURY Nicolas1, TRAN Joseph2, DORLHAC DE BORNE François3, QUESNEVILLE Hadi4 and GRANDBASTIEN Marie-Angèle1**

1: Biologie Cellulaire, IJPB, INRA-Versailles 78026 Versailles, France
2: Biologie des Semences, IJPB, INRA-Versailles 78026 Versailles, France
3: Tobacco Institute, Imperial Tobacco Group, 24100 Bergerac, France
4: Unité de Recherches en Génomique-Info (URGI), 91034 Evry cedex, France

Transposable elements (TEs) are a source of structural and functional diversification. Their impact on genome function include gene disruption, chromosomal landscape remodelling and modulation of gene expression via generation of alternative promoters or production of chimeric co-transcripts with adjacent genes. In order to evaluate the importance of TEs and TE-gene associations in plant transcriptomes, we have surveyed ESTs collections available for two model Solanaceae species, tomato (Solanum lycopersicum) and tobacco (Nicotiana tabacum). Although the TE content of both species is only partially known, TE-related ESTs represent at least 2% of their transcriptomes, with LTR-retrotransposons being predominant. In both species, TE-related transcripts are over-represented in stress conditions such as in vitro cell cultures. In tobacco, chimeric co-transcripts are surprisingly abundant, especially for the well characterized Tnt1 LTR-retrotransposon, known to be specifically activated in stress conditions. We have identified many co-transcripts originating from the Tnt1 LTR transcription start and extending into dowstream adjacent sequences, including genic sequences. These co-transcripts, potentially promoted from the Tnt1 LTR, are predominantly detected in stress-associated ESTs libraries such as those produced from senescent tobacco leaves. We are presently investigating the potential impact of these chimeric co-transcripts on the expression of adjacent sequences. These data indicate that Tnt1 is frequently co-transcribed with adjacent cellular sequences, and may act as an intermediate or "sensor" of specific stimuli, able to translate and redirect messages towards adjacent cellular functions in a large range of stress conditions.

## Germ cell-specific activation and silencing of transposable elements in plants

*R. Keith Slotkin*, Milos Tanurdzic, Matthew Vaughn and Robert Martienssen*

Cold Spring Harbor Laboratory, NY 11724, USA

Primary targets of gene silencing are transposable elements (TEs), mobile fragments that can generate mutations and genomic stability when active. Mammalian TEs are targeted and silenced specifically in the germline by piRNAs. In the model plant Arabidopsis, we have found that TEs are also specifically silenced in germ cells, however by a mechanistically distinct pathway. Plant germ cells are not set-aside early in development, but rather differentiate late from somatic cells. The haploid products of meiosis also further divide by mitotic divisions in plants, generating multicellular (haploid) germ cell structures called gametophytes. In the plant male gametophyte (pollen), heterochromatic silencing modifications such as DNA methylation and small RNAs are erased from the pollen vegetative nucleus, a companion nucleus that controls the development of the pollen grain, but does not contribute DNA to the next generation. The activation of TEs in pollen results in the production of a distinct size class of novel small RNAs, which are enriched in sperm cells. These results suggest that TE reactivation in the vegetative nucleus of pollen targets TE silencing in sperm cells.

# A role for RNAi in the correction of transposable element methylation and silencing defects in Arabidopsis

*Felipe Karam Teixeira (1,2), Fabiana Heredia (1), Alexis Sarazin (2), François Roudier (1,2), Martine Boccara (1,2), Constance Ciaudo (3,4), Corinne Cruaud (5), Julie Poulain (5), Maria Berdasco (6), Mario F. Fraga (6), Olivier Voinnet (3), Patrick Wincker (5), Manel Esteller (6), Vincent Colot* (1,2)*

1) Unité de Recherche en Génomique Végétale, CNRS UMR8114, INRA UMR1165, Université d'Evry Val d'Essonne, 91057 Evry cedex, France.
2) CNRS UMR8186, Département de Biologie, Ecole Normale Supérieure, 75230 Paris cedex 05, France.
3) Institut de Biologie Moléculaire des Plantes, CNRS UPR 2357, 67084 Strasbourg cedex, France.
4) CNRS UMR218, Institut Curie, 75248 Paris cedex 05, France.
5) Génoscope, CEA - Institut de génomique, 91057 Evry cedex, France.
6) Cancer Epigenetics Laboratory, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain.

E-mail: colot@biologie.ens.fr

DNA methylation is essential for silencing transposable elements and some genes in higher eukaryotes, implying that this modification must be tightly controlled. However, accidental changes in DNA methylation can be transmitted through mitosis, as in cancer, or meiosis, leading to epiallelic variation. We have uncovered the existence of an efficient mechanism that protects against transgenerational loss of DNA methylation and silencing in Arabidopsis. This mechanism is specific to the subset of heavily methylated repeats that are targeted by the RNAi machinery. It does not spread into flanking regions, is usually progressive over several generations, and faithfully restores wild-type methylation as well as silencing over target sequences, in an RNAi-dependent manner. These findings suggest an important role for RNAi in protecting genomes against long-term epigenetic defects.

## Live watching TE burst: Behavior and impact of a rice transposon mPing

*Ken Naito*1, Feng Zhang1, and Susan R. Wessler*

Department of Plant Biology ,University of Georgia, Athens, Georgia 30602

It has long been proposed that transposable elements (TEs) have TEs have been a rich source of material for the assembly and tinkering of eukaryotic gene regulatory systems, and have played key roles in evolution of organisms. Recent comparative genomics researches have accumulated many examples of TE-derived sequences under purifying selections.
However, to experimentally demonstrate this theory, obtaining a material where a TE burst is currently taking place.
Recently we revealed a rice strain, EG4, contains ~1,000 copies of a transposon “mPing” while most other strains have less than 50. We also found mPing is still increasing its copy number by ~40 per plant per generation.
In this study, we have analyzed mPing insertion sites in 24 EG4 plants derived from a single progenitor. We identified 934 insertions shared by all plants (old insertions) and 734 individual-specific (de novo) insertions. Regardless whether insertions are new or old, they are very rare in exons and are clearly overrepresented within 1kb upstream of transcription start sites. In addition, microarray analysis indicated that mPing insertions were significantly more associated with upregulation of proximal genes than downregulations (95 vs. 32). We also found there are many motifs of binding sites for transcription factors involved in stress responses. Subsequent qPCR analysis revealed ~10% of mPing associated genes acquired stress inducibility, indicating some of those cis-elements might be functional.
Taken together, we conclude mPing can rapidly amplify in the host genome because; (1) it hardly inserts into exons, (2) they are mostly (~80%) neutral to host gene expressions, and (3) if it affects transcriptions, upregulation, which is less harmful to the host, is much more dominant than downregulation. However, mPing does introduce numbers of cis-elements across the genome and put multiple genes under the same regulation, such as stress responses.

# Evolution of transposable elements in mammals and birds

*Arian Smit*

Institute for Systems Biology, Seattle, WA 98103-8904

# The Anolis genome assembly: Balancing mammalian and avian perspectives on amniote repeat evolution

*Andrew M. Shedlock*

Dept. of Organismic and Evolutionary Biology Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138

The imminent completion of the Anolis lizard genome assembly provides the first comprehensive, detailed view of the non-avian genomic landscape in Reptilia, the sister group of Mammalia. Squamate reptiles are arguably more physiologically, developmentally, and taxonomicaly diverse than mammals, and, unlike birds, their genome structure exhibits far more complexity in terms of both the diversity and activity of the repetitive landscape. The abundance and distribution of interspersed and tandem repeats in Anolis are more mammal-like than bird-like in their evolutionary profiles but include substantial reptile-specific diversification of transposable elements (TEs) not visible in either mammals or birds. It is likely that Anolis offers only a glimpse of the extensive clade-specific retropositional dynamics that have facilitated squamate diversification. The positional information in the Anolis assembly also offers a rich opportunity to examine the relationship between TEs and microchromosome formation and the impact of non-coding mobile DNA on the evolution of amniote transcriptome structure and function. A summary of the abundance and distribution of TEs in Anolis will be presented within a phylogenomic context that informs human/mammalian as well as avian views on the major role mobile elements have played in driving vertebrate genotypic evolution.

# Stories of outrageous horizontal transfer of DNA transposons

*Vladimir V. Kapitonov*

Genetic Information Research Institute, Mountain View, California 94043

DNA transposons are viewed now not just as simple parasites but rather as an important evolutionary force that has shaped genetic complexity of their hosts. It was recognized in the past that some DNA transposons were involved in events of horizontal transfer. However, the extent of this involvement is still not clear. Here, I report several examples of recent and ancient horizontal transfer of DNA transposons.

Based on cross-genomic comparison of transposons identified in different species, it is evident that the hAT5-5_NM and hAT5-1_BF DNA transposons have been transferred horizontally just a few million years ago (mya) between the sea anemone (Nematostella vectensis) and lancelet (Branchiostoma floridae), which diverged from their last common ancestor over 600 mya.

As it follows from computational studies of a recently discovered Chapaev superfamily, a family of these transposons (Chapaev3-1_ET) was introduced horizontally 40-60 mya into the lesser hedgehog tenrec (Echinops telfairi) genome from reptiles. The Chapaev3-1_ET consenus sequence is 97% identical to Chapaev3-3N1_AC from in the lizard Anolis carolinensis genome. Among >20 mammals, whose genomes have been sequenced, tenrec is the only species containing Chapaevs. It appears that most Chapaev3 transposons, including those from hydras, flat worms, and leeches, have evolved through horizontal transfer in a “big bang”-like scenario. For instance, Chapaev3s form a very compact cluster, when the average identity between the transposase protein sequences is ~50%, which is higher than in other metazoan transposons.

Massive horizontal transfer was also behind evolution of the Charlie3 transposon (originally described as a primate-specific family of hAT transposons). Apparently, Charlie3 was introduced horizontally ~80 mya in the lamprey Petromyzon marinus and five separate mammalian lineages ancestral to the modern (1) primates, (2) tree shrew, (3) guinea pig, (4) flying fox, and (5) ruminants/dolphins/whales.

# Massive Horizontal Gene Transfer in Bdelloid Rotifers

*Eugene A. Gladyshev(1), Matthew Meselson(1), and Irina R. Arkhipova(1,2)**

(1)Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA; (2)Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA 02543, USA

Horizontal gene transfer in metazoans is generally regarded as a rare phenomenon, unless the donor species are closely associated with the recipient by virtue of endosymbiosis or parasitism. Contrary to this expectation, we find that bdelloid rotifers, microscopic invertebrates for which males and meiosis are unknown, harbor numerous genes of foreign origin, concentrated mainly in telomeric regions and accompanied by transposable elements (TEs) of diverse types, both intact and decayed, that may have also arrived horizontally. Within these regions, approximately one-third of the genes that are known in other taxa appear to be of non-metazoan origin and to have come from bacteria, fungi and plants. While some are defective, others, including genes of bacterial origin, are intact and transcribed, contain functional spliceosomal introns, and express active enzymes when introduced into E. coli. Most of these genes are “operational” rather than “informational”, coding for relatively simple enzymatic functions. In contrast to telomeric regions, bdelloid proximal gene-rich regions are similar to those of model invertebrates in lacking foreign genes, but unlike them are depleted in mobile elements. We hypothesize that the highly unusual bdelloid lifestyle, characterized by repeated episodes of desiccation and recovery, which are likely accompanied by DNA breakage and repair and disruption of membrane integrity, facilitates uptake and incorporation of foreign genes, especially at deprotected chromosome ends, and that the capture and functional assimilation of exogenous genes may represent an important evolutionary force shaping bdelloid genomes.

References:

Gladyshev E.A., Meselson M., Arkhipova I.R. (2008). Science 320:1210-1213.

# REPET: pipelines for the identification and annotation of transposable elements in genomic sequences

*Timothée Flutre and Hadi Quesneville**

URGI, INRA, Centre de recherches de Versailles, route de St Cyr, 78026 Versailles, France

We present REPET, a framework displaying two pipelines, TEdenovo and TEannot. At some steps, the pipelines launch in parallel several different prediction programs and combine their results afterwards to improve the accuracy and exhaustiveness of the detection.
The first phase corresponds to the de novo approach, i.e. the definition of consensus corresponding to TE families. The TEdenovo pipeline begins by searching for repeats via a self-alignment of the input genomic sequences. The resulting high-scoring segment pairs are then clustered. Finally a multiple sequence alignment is performed for each cluster, from which a consensus is built. When applying TEdenovo on D. melanogaster with the most efficient combination of programs, 119 TE families are found from the 123 that are present in the sequence. In this genome sequence, 77 and 55 families can be found with respectively at least one or two complete copies. With our pipeline we found 50 consensus correctly reconstructed as a complete copy: 91% (50/55) of the families that could theoretically be reconstructed by the method. These consensus reveal the complex and intricate evolution of different TE families inside the same genome such as chimeric subfamilies.
In the second phase, these consensus are used as a library to mine the genome and detect TE copies, from full ones to fragmented and nested copies. The TEannot pipeline aligns a TE library with genomic sequences using several programs. The matches are statistically filtered and then combined all together. Simple short repeats are annotated along the way. Join procedures are applied to connect nested fragments belonging to the same copy.

# Identifying repeat-derived regions in the twilight zone

*David Pollock*

Department of Biochemistry and Molecular Genetics, University of Colorado, Denver

Traditional alignment-based methods for identifying transposable elements in genomes are not particularly effective at identifying old elements, fragments of elements, or elements without well-defined families of consensus sequences. A large amount of the genome is almost certainly repeat-derived, but not positively identified. The statistical nature of this large multiple comparison problem and analyses of the ineffectiveness of alignment algorithms indicate that previous repeat-derived composition estimates are grossly wrong. Analysis of Alu and MIR element distributions using a new "element-specific" derivative of the evolutionary-based P-cloud repeat detection method indicates that this could be a rapid and effective way to obtain accurate estimates of the structure of the transposable element "twilight zone". Half of all MIR sequence in the human genome, for example, may be currently un-annotated.

# High-throughput Sequence-based Epigenomic Analysis of Human Alu Repeats

*Hehuang Xie1,3, Min Wang1,3, Maria de F. Bonaldo1,3, Christina Smith1, Veena Rajaram4, Stewart Goldman2,3, Tadanori Tomita5 and Marcelo B. Soares*1,3*

1Cancer Biology and Epigenomics Program, Children's Memorial Research Center; 2Division of Hematology/Oncology/Transplantation, 3Department of Pediatrics; 4Division of Anatomic Pathology, Department of Pathology; 5Division of Pediatric Neurosurgery, Department of Surgery; Feinberg School of Medicine, Northwestern University, Chicago IL 60614-3394, USA

DNA methylation is the only known covalent modification of mammalian genomic DNA, predominately occurring in CpG dinucleotides. Over 50% of CpG dinucleotides in the human genome are found within repeat elements. However, no method is currently available for large-scale ascertainment of CpG methylation of repetitive sequences. Toward this goal, we have developed a bisulfite sequencing-based strategy for large-scale parallel determination of the status of CpG methylation of thousands of repeats and their flanking sequences. To achieve that, we developed a computation algorithm to design primers allowing the simultaneous amplification of large sets of repeat elements from bisulfite converted genomic DNA. Using one single primer set, we derived CpG methylation data from 31,178 Alu elements and their 5' flanking sequences, altogether representing more than 4 Mb of the human brain epigenome. Importantly, we have successfully utilized this strategy to generate epigenomic representations of pediatric ependymomas in an effort to elucidate the pattern of global loss of methylation in this cancer while identifying epigenomic profiles that might be predictive of clinical behavior.

# Evolution of the mammalian transcription factor binding repertoire via transposable elements

*Guillaume Bourque*

Genome Institute of Singapore
bourque@gis.a-star.edu.sg

Identification of lineage-specific innovations in genomic control elements is critical for understanding transcriptional regulatory networks and phenotypic heterogeneity. We analyzed, from an evolutionary perspective, the binding regions of seven mammalian transcription factors (ESR1, TP53, MYC, RELA, POU5F1-SOX2 and CTCF) identified on a genome-wide scale by different chromatin immunoprecipitation approaches and found that only a minority of sites appear to be conserved at the sequence level. Instead, we uncovered a pervasive association with genomic repeats by showing that a large fraction of the bona fide binding sites for five of the seven transcription factors (ESR1, TP53, POU5F1-SOX2 and CTCF) are embedded in distinctive families of transposable elements. Using the age of the repeats, we established that these Repeat-Associated Binding Sites (RABS) have been associated with significant regulatory expansions throughout the mammalian phylogeny. We validated the functional significance of these RABS by showing that they are over-represented in proximity of regulated genes and that the binding motifs within these repeats have undergone evolutionary selection. Our results demonstrate that transcriptional regulatory networks are highly dynamic in eukaryotic genomes and that transposable elements play an important role in expanding the repertoire of binding sites.

## The tumor suppressor protein p53 binding sites in mammalian genomes are related to transposons

*Feng Cui, Michael V. Sirotin and Victor B. Zhurkin**

National Cancer Institute, NIH, Bethesda, MD 20892, USA

The p53 protein is involved in amazingly complicated regulatory network, mediating expression of ~1000 human genes. p53 binds DNA with extremely degenerate sequence specificity. A typical p53 binding site (p53BS) comprises two decamers RRRCWWGYYY separated by a variable spacer, S. Human genome contains ~1,000,000 potential p53BS, most of them with unknown functional significance. The p53-induced transactivation is in the focus of experimental studies; however, the p53-mediated repression of transcription and replication is also critical for the tumor suppression activity of p53.

Earlier, we found that a significant fraction of p53BS in human genome belongs to transposons from the families LTR10, MER61 and MLT1H. Here, we compare occurrences of p53BS in the human and mouse genomes, using the length of the spacer between two p53 half-sites as a 'critical' parameter. We show that the genome-wide distribution of the spacer, S, is drastically different for the two species. In human genome, the two strongest peaks in this distribution (S=0 and S=3 bp) originate due to Alu transposons. Alu subfamilies are characterized by various length of the spacer: AluSg and AluSp predominantly have spacer S=0; whereas AluJb, AluSq and AluSx have S=3 bp. Importantly, the p53BS with S=0 occur mostly in the vicinity of genes up-regulated by p53, while the p53 sites with S=3 bp are found close to the down-regulated genes.

The number of putative p53BS embedded in Alu elements is enormous, ~200,000. Comparing these sites with the consensus Alu sequences, we found numerous substitutions of CATG (in p53BS) for CGCG or CACG (in consensus). Therefore, we suggest that one of the mechanisms responsible for the abundance of p53 sites in Alu elements is based on the CG to CA:TG mutation (involving cytosine deamination). Our observations may be important for understanding the origins of evolution of the p53 regulatory network.

# Methylation and deamination of CpGs generates p53 binding sites on a genomic scale

*Tomasz Zemojtel*, Szymon M. Kiełbasa, Peter F. Arndt, Ho-Ryun Chung, Martin Vingron*

Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestrasse 73, D-14195 Berlin, Germany

Formation of transcription factor binding sites is a key evolutionary process. Here, we show that methylation and deamination of CpGs constitutes a vehicle that generates in vivo p53 binding sites predominantly in the Alu family of SINE transposable elements but also in non repetitive DNA. Based on this observation, we propose that the highly mobile Alu elements spread p53 binding sites in a species-specific manner.

# The regulated retrotransposon transcriptome of mammalian cells

*Geoffrey J Faulkner*1, Valerio Orlando2, Sean M Grimmond1 and Piero Carninci3*

(1)Expression Genomics Laboratory, Institute for Molecular Bioscience, University of Queensland, Queensland, Australia; (2)Dulbecco Telethon Institute at Institute of Genetics and Biophysics, National Research Council (IGB CNR), Via Pietro Castellino Street 111, 80131 Naples, Italy; (3)Omics Science Center, RIKEN Yokohama Institute, Yokohama, Kanagawa 230-0045, Japan

Retrotransposons have long been regarded as a source of transcriptional control in eukaryotes, though the precise mechanism of this regulation has largely remained elusive. In the past year, several prominent works have provided a snapshot of the regulatory potential of retrotransposons, including control of chromatin structure1 and the production of endogenous siRNAs2,3. Here, we present a genome-wide, high-throughput sequence tag survey of transcription initiation from within mammalian retrotransposons. Cap Analysis Gene Expression (CAGE) revealed nearly 250,000 novel transcription start sites in mouse and human retrotransposons, with approximately half having the potential to act as alternative promoters of adjacent genes. From a regulatory standpoint, the expression of nearly 20,000 retrotransposon transcription start sites in mouse and human was strongly positively correlated with the expression of known genes near which they map. These promoters included a cohort of 300 putative boundary elements. A further 2,000 retrotransposons produced bidirectional transcription suitable to generate siRNA precursors. Finally, we discovered a novel mechanism by which transcribed retrotransposons may target protein-coding transcripts. Overall, the extent of this transcription far exceeded prior estimates, with a common theme of strong tissue specificity and association with nearby protein-coding genes, suggesting that the global transcription of retrotransposons is a novel and substantial factor in the output and regulation of the mammalian transcriptome.

1. Lunyak, V.V. et al. Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science 317, 248-51 (2007).
2. Ghildiyal, M. et al. Endogenous siRNAs Derived from Transposons and mRNAs in Drosophila Somatic Cells. Science (2008).
3. Watanabe, T. et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature (2008).

# Methylation profiling: repetitive elements and disease progression

*Sebastian Szpakowski*1,2, Xueguang Sun2, José M. Lage2, Jill Rubinstein1,2,3, Andrew Dyer2, Min Chen4, Diane Kowalski2, Janet Brandsma2, Hongyu Zhao4, Clarence Sasaki3, Jose Costa2, Paul M. Lizardi1,2*

Yale University School of Medicine: (1)Interdepartmental Program in Computational Biology and Bioinformatics, (2)Department of Pathology, (3)M.D.-Ph.D. Program, 4Department of Surgery

Globally, dysregulation of methylation of DNA domains has been liked to disease phenotypes. To facilitate a high-throughput, genome-wide and local study of the methylation patterns of the human genome, we created a custom microarray platform. The microarray is capable of profiling the methylation status at 339,314 uniquely identifiable genomic loci nearby, or overlapping with, genes, repetitive elements (annotated using RepeatMasker and Tandem Repeat Finder), microRNAs or un-annotated CpG rich regions. A collection of 35 head and neck squamous cell carcinoma (HNSCC) samples and 18 morphologically "normal," tumor-adjacent tissue samples were used to generate methylation profiles using the custom microarray. In addition, 10 buccal scraping samples from unrelated healthy individuals were used to provide a set of reference-normal methylation profiles. The collection was validated using bisulfite sequencing for 207 loci selected at random from the collection, yielding a concordance with the microarray measurements close to 88%. The microarray results are also highly reproducible based on technical replicate analyses performed for several samples.
Initially, we found 15,587 informative probes that significantly distinguish tumor from tumor-adjacent profiles. Unexpectedly, besides being enriched in certain categories of genes, the list of informative probes was significantly more enriched in certain sub-families of repetitive elements, such as primate Line-1s or Alus. We observe that the methylation levels differ significantly among various subfamilies of repetitive elements from the same family, correlating with evolutionary age, length and potential for biological activity of the repetitive element. Our findings indicate that there may exist various modes of subfamily-specific epigenetic control for repetitive elements. Furthermore, a dysregulation of epigenetic control of specific subsets of repetitive elements may play an important role in a tumor's evolution towards the most optimal phenotype.

## Stressed out retrotransposons on the move

*Marlene Belfort*, Art Beauregard, Venkata Chalamcharla, Colin Coros, Carol Lyn Piazza, Dorie Smith*

Wadsworth Center, NYSDOH, and Dept. Biomedical Sciences, SUNYA, 150 New Scotland Avenue, Albany, NY 12208, USA

Retrotransposons move via RNA intermediates and usually encode the agent of their mobility, reverse transcriptase. They also rely on host factors to proliferate. We use group II introns in bacteria as the paradigm to study the dynamics between target-primed retrotransposons and their hosts1. The Ll.LtrB intron recruits Escherichia coli polymerases, nucleases and DNA ligase to complete the retromobility process2. We used a genetic screen to identify other E. coli functions involved in retromobility of the Ll.LtrB intron. Disruptions that increased or decreased retrohoming levels into the E. coli chromosome were isolated. These functions included factors involved in RNA processing, DNA replication, energy metabolism and global regulation3. We characterized an rne mutant that regulates RNase E expression and elevates retrohoming and retrotransposition. The stimulatory effect of the mutation on retromobility results from intron RNA accumulation. These results suggest that RNase E, which is the central component of the RNA degradosome, regulates retrohoming in response to cellular physiology. Next we focused on genes encoding enzymes that catalyze synthesis of global regulators cAMP and ppGpp, which elevate group II intron mobility. These small molecules program genetic transitions between nutrient excess and starvation. Accordingly, we demonstrated that glucose depletion of wild-type cells and cAMP supplementation of mutants stimulated retromobility. Likewise, amino acid starvation, which induces the alarmone ppGpp, activated retromobility. In both cases, retrotransposition to ectopic sites was favored over retrohoming. Interestingly, these stimulatory effects are mediated at the level of the DNA target, rather than of expression of the retroelement. Thereby, during metabolic stress, cAMP and ppGpp control group II intron movement in concert with the cell's global circuitry, stimulating genetic diversity.

1) Beauregard, A., et al. Annu. Rev. Genetics. (2008) 42, 587-617.

2) Smith, D., et al. Genes Dev. (2005) 19, 2477-2487.

3) Coros, C.J.. et al. RNA (2008) 14, 2634-2644.

# Characterization of an Active Mammalian DNA Transposon

*NL Craig* & Rupak Mitra*

Howard Hughes Medical Institute, Dept of Molecular Biology & Genetics, Johns Hopkins University School of Medicine. Baltimore MD 21205, USA

Analysis of the human, mouse and other mammalian genomes has provided no evidence for the movement of DNA cut & paste transposons within the last 35-40 million years although such elements do transpose when introduced into mammalian cells. However, bioinformatic analysis (1,2) of the little brown bat genome (Myotis lucifugous) has suggested that several types of DNA transposons have been recently active, i.e. within 1 million years ago.

We have now demonstrated that a bat DNA cut & paste piggyBac element is active in human cells, bat cells and in the yeast S. cerevisiae at levels only slightly less than that of the well-studied insect piggyBac element. Thus at least one mammal contains a currently active DNA transposon. It will be interesting to probe what features of the bat have allowed this active DNA transposon to persist and whether other mammalian genomes also contain active DNA elements.

1. Pritham EJ, Feschotte C. (2007) Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus. Proc Natl Acad Sci U S A. 104:1895.

2. Ray DA, Feschotte C, Pagan HJ, Smith JD, Pritham EJ, Arensburger P, Atkinson PW, Craig NL. (2008) Multiple waves of recent DNA transposon activity in the bat, Myotis lucifugus. Genome Res. 18:717.

# 5’-transduced SVA retrotransposon groups spread efficiently throughout the human genome

*Julija Raiz1, Annette Damert1, Johannes Löwer1, Hui Wang3, Jinchuan Xing4, Mark A. Batzer2, Roswitha Löwer1 and Gerald G. Schumann1**

1Section PR2/Retroelements, Paul-Ehrlich-Institut, Paul-Ehrlich-Strasse 51-59, D-63225 Langen, Germany
2Department of Biological Sciences, Biological Computation and Visualization Center, Center for Bio-Modular Multiscale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA
3Human Genome Sequencing Center, Baylor College of Medicine, N1519 One Baylor Plaza, Houston, TX 77030, USA
4Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA

SVA elements represent the youngest family of hominid non-LTR retrotransposons which alter the human genome continuously, and they stand out due to their organization as composite elements.
In order to draw conclusions on the assembly process that led to the current organization of SVAs, their transcriptional regulation, and the mechanism of SVA retrotransposition, we assessed differences in the structures of the 116 SVA elements located on chromosome 19. We identified seven structural SVA variants including novel variants like retrotransposed solitary VNTR regions, termed SVA2 elements, 3’-truncated and 5’-transduced elements.
We found that chromosome 19 is harbouring members of a remarkably successful human-specific 5’ transduction group of SVA_F elements encompassing ~84 elements genome-wide. We assigned those elements to a separate SVA subfamily termed F1. An ancient retrotransposition event that led to transcriptional control of a 5’-truncated SVA copy by the testis-specific promoter of the MAST2 gene served as primal source element of this 5’-transduction group. The variety of secondary 5’ transduction events found in SVA_F1 members indicates transcriptional control of their source elements by many different external cellular promoters. The presence of those source elements in the human genome, their ongoing transcription confirmed by EST data, as well as the finding that several SVA_F1 subfamily members showed insertion polymorphisms indicates that members of this subfamily are currently mobilized in the human genome. We also demonstrate experimentally in tissue culture that mobilization of SVA reporter elements is mediated in trans by the human LINE-1 protein machinery in tissue culture.
Taken together, our results not only allow to draw conclusions on the SVA assembly process, but also demonstrate that transcriptional control of SVA source elements by external promoters is prevalent. SVA subfamily F1 is of very recent evolutionary origin and some of its members are serving themselves as source elements for ongoing retrotransposition.

# Exon-trapping mediated by the SVA retrotransposon

*Dustin C. Hancks*, Adam D. Ewing, and Haig H. Kazazian Jr*

Department of Genetics, University of Pennsylvania

The great majority of human retrotransposable elements are inactive, however, both inactive and active retrotransposons drive genome evolution and may influence transcription through various mechanisms. Little is known about the SVA element, a non-coding RNA, which is one of three retrotransposon families still active in the human genome. We report the identification of a new subfamily of SVA, which appears to have been formed by an alternative splicing event, where the first exon of the Mast2 gene spliced into an intronic SVA at a site which resembles a splice acceptor sequence in the SVA and subsequently retrotransposed. Upon performing molecular and computational experiments, we have identified many functional splice acceptor sites within several different transcribed SVAs across the genome. We propose that SVA is mimicking a gene-trap in order to mediate its transcription by any means necessary. Furthermore, this result implies that SVA elements residing within introns of genes in the same orientation may disrupt normal gene transcription and that SVAs may alter the transcriptome, thereby altering genome evolution.

# Retroposon analysis reveals simultaneous divergence of the placental mammalian ancestor possibly triggered by continental divisions

*Hidenori Nishihara*

Tokyo Institute of Technology

Due to recent developments in molecular phylogenomics, all the extant orders of placental mammals are grouped into three lineages: Afrotheria, Xenarthra and Boreotheria, which originated in Africa, South America and Laurasia, respectively. In spite of this advancement, however, the divergence order of these three lineages remains unsolved. Here, we performed extensive retroposon analysis using mammalian genomic data. Surprisingly, we identified a similar number of informative retroposon loci that support each of three possible phylogenetic hypotheses: the basal position for Afrotheria (22 loci), Xenarthra (25 loci), and Boreotheria (21 loci). This result indicates that the divergence of the placental common ancestor into the three lineages occurred almost simultaneously. Thus, we re-evaluated recent geological data to establish that complete separation between Africa and South America occurred at 120 ± 10 Ma. Furthermore, the exact timing of continental rifting of Europe from Africa (estimates range from 148-110 Ma) remains unclear. These data do not exclude the possibility that Pangea was divided into Laurasia, Africa and South America simultaneously at 120 Ma. Our retroposon data and revised geological estimates suggest that the simultaneous divisions of continents leading to isolated Africa, South America and Laurasia caused concomitant divergence of the ancient placental ancestor into three lineages, Afrotheria, Xenarthra, and Boreotheria, around 120 Ma.

# ELAN: A server based tool for genome wide analysis of mobile genetic elements

*Kamal Rawal*1,5, Sudha Bhattacharya 2, Alok Bhattacharya1,3, Ram Ramaswamy 1,4*

1 School of Information Technology
2 School of Environmental Sciences
3 School of Life Sciences.
4 School of Physical Sciences,
Jawaharlal Nehru University, New Delhi 110 067, India, 5. JIIT University, India.

Mobile genetic elements occupy significant proportion of eukaryotic genomes. They are involved in number of important cellular functions. ELAN is a suite of tools for genome wide analysis of mobile genetic elements. It finds distribution and nature of mobile genetic elements. DNA SCANNER is a part of ELAN which analyses insertion sites of mobile genetic elements for the presence of various physicochemical signals. Insertion Site Finder (ISF) is a machine learning tool which incorporates information derived from DNASCANNER and uses support vector machines to classify DNA sequences into insertion sites and non insertion site classes. ELAN has been applied to wide variety of organisms. It has identified distributions of several mobile elements such as Alu in various organisms such as Human, Mouse, Drosophila, E. histolytica etc. DNA SCANNER has identified common set of statistically important signals flanking insertion sites in various genomes suggesting common insertion mechanism operating in wide variety of organisms. ISF has emerged to be an important tool for insertion site prediction as it has shown high accuracy levels (65-90%). The dataset and information derived during analysis will serve as bench marking resource in future for various analyses. Large data has been organized into web portal as well as relational database named as InSiDe which is available online at http://nldsps.jnu.ac.in/bioit/ccbb/elan.html. Experiments were conducted in E. histolytica to validate computational findings.

References:

1. Identification of Insertion hot spots for non-LTR retrotransposons: Computational and Biochemical application to Entamoeba histolytica. Nucleic Acid Research, 34, 2006, 5752-5763.
2. he LINEs and SINEs of Entamoeba histolytica: comparative analysis and genomic distribution. Journal of Experimental Parasitology, 110, 2005, 207-213.

## Automatic classification of mariner-like elements

*Jacques-Deric Rouault, Aurélie Hua-Van, Pierre Capy**

Laboratoire Evolution, Génomes et Spéciation, CNRS, 91190 Gif-sur-Yvette, and Université Paris-Sud 11, France

A classification is required as soon as the number of items that we have to deal with grows until it becomes too large. This can be illustrated by the classification of transposable elements (TEs). While the upper part of the classification of TEs (from classes to superfamilies or families) is regularly updated, the lower levels are investigated much less.
The method proposed uses the nucleotidic sequences (functional or not, complete or partial), and was tested on the mariner family which is probably one of the most well known. The classification process is based on a distance matrix built from pairwise alignments from which the two closest groups are pooled at each step using an aggregation algorithm. This process is iterated until all elements belong to the same final group.

At any level of the aggregation process, the central element of a group is defined to minimize the sum of the distances to the other members of the group. The Highest Internal Distance (HID) of the group is the distance between the central element and the farthest element of the group, and the Lowest External Distance (LED) is the smallest distance between the central element of the group and the closest element not belonging to the group. A group is considered valid if its HID is lower than its LED. The valid groups of the higher levels are named Clans.

Application to the mariner family leads us to define 12 subfamilies consistent with the phylogenies obtained from protein sequences. In a few cases, a subfamily corresponds to a Clan (Capitata, Irritans, Rosa, Elegans), but in general a subfamily is a pool of several Clans (Cecropia, Mellifera, Mauritiana, Briggsae). The rate of error from a posteriori re-assignment is lower than 1% and only occurs in invalid groups.

# The need for an approved nomenclature for (human) endogenous retroviruses

*Jens Mayer[*1] and Ruth Seal[*2] on behalf of the HGNC[2]*

1 Department of Human Genetics, University of Saarland, 66421 Homburg/Saar, Germany
2 HUGO Gene Nomenclature Committee, European Bioinformatics Institute, Cambridge, United Kingdom

At the moment, there is no standardized nomenclature used to represent human endogenous retroviruses (HERVs), and ERVs in general. This lack of standardization means that it is difficult to perform effective searches of the literature for this field, that there is a risk of misinterpretation of data, and that there is little guidance for the naming of newly-discovered ERVs. Additionally, there are a number of problems with the various nomenclature systems that are currently in use. For example, many HERV symbols incorporate family designations, but the designations used in genome databases often differ from those that appear in the literature, and some families have been reclassified over time, resulting in changes in name usage.
A successful nomenclature needs to be stable and acceptable to as many researchers in the field as possible to ensure that the approved symbols are widely used in publications and talks. Also, it is important that the same nomenclature system is supported by the International Committee for the Taxonomy of Viruses (ICTV), RepBase, and the International Committee on Classification of Transposable Elements. Additionally, the HUGO Gene Nomenclature Committee (HGNC) is particularly interested in standardizing names for proviruses that have been shown to be transcriptionally active within the human genome. Approved HERV symbols need to include, at minimum, a root symbol that represents the (H)ERV, perhaps including agreed-upon family designations, and a unique identifier to represent each particular provirus within the genome. The community will need to decide on how much information can feasibly and accurately be included in each symbol. We will present some ideas for discussion on how to achieve an approved nomenclature for HERVs.

## Optimal taxonomical markers for ERVs

*Jonas Blomberg*

## Non-autonomous and complete derivate elements: how to classify them?

*Francois Sabot*

## Superfamily-specific reference datasets as possible aids in resolving classification problems

*Irina Arkhipova*

# POSTER ABSTRACTS

## (1)

## Population structure in Medicago truncatula inferred from MITE insertion polymorphisms

*Dariusz Grzebelus1, Mirosława Gładysz1, Anna Gambin2*

1 Department of Genetics, Plant Breeding and Seed Science, Agricultural University of Krakow, Krakow, Poland
2 Institute of Informatics, Warsaw University, Warsaw, Poland

Site-specific transposon insertion polymorphisms can be easily identified using simple PCR assays, provided that sequences flanking these insertions are known. In the genome of Medicago truncatula we identified a group of miniature inverted repeat elements (MITEs) related to the previously described transposons of the PIF/Harbinger superfamily. Initially, we mined ca. MITE 40 insertion sites in the reference genome of cv. A17 'Jemalong'. In order to avoid bias, we than searched for insertion sites in other M. truncatula ecotypes. Using iPCR, we identified and sequence-characterized seven insertion sites that were unoccupied in the reference genome. Amplification of sequence-characterized MITE insertion sites across a set of 25 M. truncatula core colection ecotypes showed that roughly half of all assayed insertion sites were fixed monomorphic insertions, the remaining half being polymorphic. In the latter group, usually the two expected variants, i.e. amplicons originating from the occupied and the unoccupied sites could be found. However, amplicons of different length were also present with low frequency and lack of amplification (null-allele) was observed in several cases. A subset of insertions producing clear and consistent polymorphic profiles were chosen for inferring population structure of the species. The ecotypes representing M. truncatula genetic diversity were divided into three groups (K = 3) using Structure v. 2.2. One of these groups comprised ecotypes from the Western Mediterranean, while the remaining two groups had broader and partially overlapping geographical range. The research project was funded by the Polish Ministry of Science and Higher Education grant no. N301 036 31/1203, for the years 2006-2008.

## (2)

# REPEAT: pipelines for the identification and annotation of transposable elements in genomic sequences

*Timothee Flutre (*), Elodie Duprat, Hadi Quesneville*

E-mail: timothee.flutre@versailles.inra.fr

Transposable elements (TEs) account for a large part of most eukaryotic genome sequences and have a huge impact on genome structure, function and evolution. We present REPET a framework displaying two parallelized pipelines, TEdenovo and TEannot, currently used in 15 genome projects (plants, animals and fungi). The first phase corresponds to the de novo approach, i.e. the definition of consensus corresponding to TE families. Then, in the second phase, these consensus are used as a library to mine the genome and detect TE copies.
The TEdenovo pipeline searches for repeats via a self-alignment of the input genomic sequences, then clusters the resulting high-scoring segment pairs, and finally build consensus from multiple sequence alignments. At each step, several programs are combined to improve the efficiency. When applied on D. melanogaster, 85% of the consensus are matching with already known TEs, less than 10% matching with several of them, hence giving a hint about their specificity. Moreover, consensus from TEdenovo correspond to subfamilies rather than families. Hence such consensus, notably the chimeric ones, reveal the complex, intricate evolution of different TE families inside the same genome.
The TEannot pipeline (Quesneville et al., 2005) aligns a TE library with genomic sequences using several programs. The matches are statistically filtered and then combined all together. Simple short repeats are annotated along the way. A long join procedure is applied to connect nested fragments belonging to the same copy. Finally, the annotation is delivered in common formats allowing their subsequent analysis in genome browser and annotation editor.

Quesneville, H., C. M. Bergman, O. Andrieu, D. Autard, D. Nouaud, M. Ashburner, and D. Anxolabéhère (2005, July). Combined evidence annotation of transposable elements in genome sequences. PLoS Computational Biology 1 (2).

## (3)

# Changes in transposable element activity and the epigenomic consequences in immortalized plant cells

*Milos Tanurdzic*, Matthew Vaughn and Robert Martienssen*

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY11724

Plant cells grown in culture exhibit genetic and epigenetic instability similar to animal cell lines and cancer cell lines. Using a combination of microarray and ChIP-seq chromatin profiling and RNA-seq transcriptome profiling approaches we have mapped the location and abundance of histone and DNA modifications as well as strand-specific transcripts in a continuously proliferating, de-differentiated cell suspension culture of Arabidopsis. We have found genome-wide epigenetic changes within the euchromatin, which becomes hypermethylated in cell culture, with numerous epigenetically modified genes targeted by the heterochromatic silencing machinery and TE-produced siRNAs acting in trans. In contrast, the heterochromatin itself, composed mainly of various transposable element (TE) repeats undergoes dramatic and very precise mosaic DNA hypomethylation with transcriptional activation of specific transposable elements in culture. High throughput sequencing of small interfering RNA (siRNA) revealed that those TEs activated in culture have increased levels of 21nt siRNA, often at the expense of the 24nt siRNA class. In contrast, TEs that remain transcriptionally silent, which match the predominant 24nt siRNA class, do not change significantly in their siRNA profiles. These results implicate RNAi and chromatin modifications in epigenetic restructuring of the genome following activation of TEs in immortalized cells and a role that TEs can play in creating mitotically stable epialleles.

## (4)

# Tol1 and Tol2 of medaka: two similar hAT-family elements at different evolutionary stages in a single fish genome

*Akihiko Koga*

Kyoto University, Japan

The Tol1 and Tol2 elements are both members of the hAT family, residing in the genome of the medaka fish Oryzias latipes. They share many structural similarities, including full-length sequences of 4.4 kb and 4.7 kb, transposase enzymes consisting of 685 aa's and 851 aa's, and terminal inverted repeats of 14/14 bp and 17/19 bp, respectively. Both elements also have target site duplications of 8 bp. However, they exhibit contrasting characteristics in genomic organization. More than 90% of extant Tol1 copies are shorter than 2.0 kb in length, due mainly to internal deletions, while a similarly defective copy of Tol2 has not been found among more than 400 copies examined. An obvious difference is also seen in their distribution among species, Tol1 occurring in all ten medaka species examined but Tol2 present in only two closely related species. These features of the two elements are reminiscent of a structural transition observed with the P element of Drosophila. This element is considered to have been introduced into the fly genome by horizontal transmission. During or after their proliferation, mutational changes by internal deletion accumulated and deprived them of the ability to produce a functional transposase. The status of Tol2 can be thought to correspond to that of P at an early stage of its history, and Tol1 to present-day P copies. Experiments have shown that Tol2 can be transposed with only the terminal regions as long as the transposase is present. It is thus a reasonable prediction that defective Tol2 copies will eventually occur within the medaka genome and subsequently increase tin frequency over time.

## (5)

# Regulatory Impact of MusD Retroelements

*Tuğçe Aktaş*1, Marc Friedli2, François Spitz1*

1 European Molecular Biology Laboratory, Developmental Biology Unit, Heidelberg, Germany
2 Department of Genetic Medicine and Development, University of Geneva Medical School and University Hospitals of Geneva, Switzerland

Transposable elements are traditionally perceived as "junk DNA" but they are also a major evolutionary force in shaping genes and genomes. We have taken advantage of the mouse Dactylaplasia mutants to study the functional impact of ERVs on neighboring genes. Two distinct insertions of MusD subfamily-retrotransposons in the same region on chromosome 19 have been associated with dominantly inherited limb malformations. One of these insertion occurred in the intron of Fbxw4, leading to abortive transcription of this gene. In addition, this insertion disrupts an evolutionary conserved enhancer that drives gene expression notably in brain and neural tube and which may control the adjacent Fgf8 gene. These findings illustrate the pleiotropic loss-of - functions associated with ERV insertions. Yet, these “local” and first-line effects cannot account for the limb phenotype, which is caused by a gain-of-function mechanism and is also observed with another MusD insertion occurring at ~ 55kb from the previous one, outside of Fbxw4. Interestingly, a very similar human limb malformation (SHFM3) is caused by large tandem duplications comprising about 500kb of the orthologous locus, with breakpoints in Fbxw4. These data suggest that MusD elements may contain sequences that can promote a complex reshuffling of the cis-regulatory network of the Dac locus. We will present data addressing these novel gene regulatory and possibly insulator activities of MusD elements. Intriguingly, certain mouse strains contain an unlinked modifier locus that protects them from the manifestation of the Dac phenotype. We have observed that this is associated with a strong methylation of MusD 5’LTR, which is absent in phenotypic mutants. Our findings show that MusD elements impinge neighboring genes in a variety of manners and may behave as epigenetically regulated “controlling” elements.

## (6)

# A recent adaptive transposable element insertion near highly conserved developmental loci in Drosophila

*Josefa González*, J. Michael Macpherson, and Dmitri A. Petrov*

Department of Biology, Stanford University, Stanford, USA

A systematic identification of adaptive insertions revealed that Transposable Elements (TEs) are a considerable source of recently adaptive mutations in the Drosophila genome(González et al 2008). We analyze in depth, one of these putatively adaptive insertions: Bari-Jheh. This transposon is inserted in the 0.7kb intergenic region between Jheh2 and Jheh3 genes both of them involved in the degradation of Juvenile Hormone. We sequenced the 7kb region flanking the insertion, including the whole coding region of both genes, and show that Bari-Jheh is likely to be the mutation causing the adaptive sweep identified in this region. We performed allele-specific expression analysis and found that the insertion is associated with a down-regulation in the expression of both Jheh2 and Jheh3 genes. These changes in gene expression seem to have subtle phenotypic consequences. We found that under specific environmental conditions, flies with the insertion have lower viability and higher developmental time compared to flies without the insertion. Altogether, these results suggest that Bari-Jheh insertion is highly likely to be an adaptive mutation. We finally looked for evidence of recurrent adaptive events in this region of the genome by analyzing population genetics data of the genes flanking the insertion both in D.melanogaster and D. simulans and conclude that the adaptive insertion of Bari-Jheh is an extremely unusual event in the history of Jheh loci. We discuss the implications of these findings for the study of adaptation since the inference of selection is often based on the assumption that adaptation is recurring at the same loci or even at the same sites.

González et al 2008. PLoS Biology 10(6).

## (7)

# Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition

*Masaki Kajikawa1*, Jun Suzuki1, Katsumi Yamaguchi1, Noritaka Adachi2, Hideki Koyama2, Shunichi Takeda3 and Norihiro Okada1*

1 Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Japan
2 International Graduate School of Arts and Sciences, Yokohama City University, Japan
3 Graduate School of Medicine, Kyoto University, Japan

Long interspersed elements (LINEs) are transposable elements that proliferate within eukaryotic genomes, making a large impact on eukaryotic genome evolution. LINEs mobilize via a process called retrotransposition. Although the role of the LINE-encoded protein(s) in retrotransposition has been extensively investigated, the participation of host-encoded factors in retrotransposition has not been well studied. To investigate the involvement of host factors in LINE retrotransposition, we examined retrotransposition frequencies (RFs) of two structurally different LINEs—zebrafish ZfL2-2 and human L1—in knockout chicken DT40 cell lines deficient in genes involved in the non-homologous end-joining (NHEJ) repair of DNA and in human HeLa cells treated with a drug inhibiting NHEJ. Deficiencies of NHEJ proteins decreased RFs of both LINEs in these cells, suggesting that NHEJ is generally involved in LINE retrotransposition. Characterization of ZfL2-2 insertions in DT40 cells further supported the involvement of NHEJ in LINE retrotransposition. Taken together, our data suggests that the NHEJ proteins are components of the retrotransposition intermediate at the target site, integrating LINE insertions efficiently into the host chromosomal DNA.

## (8)

# Identifying cis-acting elements capable of epigenetic regulation

*Monica Sentmanat*, Sarah Elgin*

Department of Biology, Washington University, St Louis, MO

Heterochromatin is defined as the densely staining, late-replicating, gene deserts of the genome. Correct assembly of heterochromatin is critical for fundamental biological processes such as regulated gene expression, mitosis, and chromosome stability. Assembly begins with histone deacetylation concomitant with di- and tri-methylation of histone H3 at K9; this modified site is bound by heterochromatin protein 1 ( HP1). Heterochromatin predominates at the centromere and telomeres of chromosomes—regions abundant in transposable elements and other repeats. Transcription of these sequences has been found to be a platform for assembly of heterochromatin through RNAi in S. pombe and may play a critical role in A. thaliana, and D. melanogaster. However, the mechanism and target sequences remain largely ill defined.

Recently, 1360, a DNA transposable element, has been implicated as a target for heterochromatin formation in D. melanogaster1. The presence of a single 1360 element was found sufficient to promote position-effect variegation (PEV) at an hsp70-white reporter close to centromere 2L. Position effect variegation has been shown to correspond to compact chromatin packaging that depends on heterochromatin component HP1, making PEV a suitable marker for heterochromatin assembly2. To determine what minimal feature(s) of 1360 (among other elements) are sufficient to promote heterochromatin formation, candidate features will be site-specifically inserted upstream of the hsp70-white reporter at 1360 variegating sites using phiC31-integrase. Elements capable of inducing position effect variegation (PEV) of hsp70-white will be selected for further analysis. Currently we are testing reporter insertion sites for sensitivity to 1360.

1. Haynes KA, Caudy AA, Collins L, Elgin SC. Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. Curr Biol 2006; 16:2222-7.
2. Wallrath LL, Elgin SC. Position effect variegation in Drosophila is associated with an altered chromatin structure. Genes Dev 1995; 9:1263-77.

## (9)

# Transposable elements and young sex chromosomes in dioecious plant Silene latifolia

*Eduard Kejnovsky*1, Roman Hobza1, Zdenek Kubat1, Tomas Cermak1, Jiri Macas2, Alex Widmer3 and Boris Vyskot1*

1 Laboratory of Plant Developmental Genetics, Institute of Biophysics, Brno, Czech Republic
2 Laboratory of Molecular Cytogenetics, Institute of Plant Molecular Biology, Ceske Budejovice, Czech Republic
3 ETH Zurich, Institute of Integrative Biology, Plant Ecological Genetics, Zurich, Switzerland

Sex chromosomes are unique parts of genomes, especially the Y chromosome, often largely non-recombining. Different evolutionary forces are in action in non-recombining regions – dominant are degeneration of genes and accumulation of repetitive DNA sequences. Dioecious plant Silene latifolia (white campion) has evolutionary young sex chromosomes (10mya) in contrast to old mammalian sex chromosomes (100-200mya).
We systematically studied main repetitive DNA sequences in S. latifolia and tested their accumulation on the Y chromosome, the largest chromosome in this genome, by fluorescence in situ hybridization (FISH). Only Ty1/copia elements were accumulated on the Y chromosome while other elements were dispersed on all chromosomes. Similarly, recently we found accumulation of tandem repeats, microsatellites and chloroplast DNA sequences on the Y chromosome in S. latifolia. Surprisingly, Ty3/gypsy elements of Ogre type were present on all chromosomes but absent in non-recombining parts of Y chromosome. On the Y chromosome Ogre-like elements were present only in pseudoautosomal region (PAR) indicating that Ogre-like element colonizes only recombining regions of genome. We can speculate that Ogre-like elements are in S. latifolia spread in conjunction with recombination machinery or, alternatively, are active only in females.
We microdissected Y and X chromosomes and autosomes and amplified various TEs. Comparison of TEs from different chromosome revealed their higher intrachromosomal similarity, most evident in Y chromosome. We explain this phenomenon by intrachromosomal gene conversion working more strongly on TEs located on the Y chromosome.

## (10)

# Serum albumin intron-1 of western Palearctic water frogs contains a non-LTR CR1-like retrotransposon

*Jörg Plötner1*, Frank Köhler1, Thomas Uzzell2, Peter Beerli3, Hansjürg Hotz1,4, Robert Schreiber1, and Gaston-Denis Guex4*

E-mail: joerg.ploetner@museum.hu-berlin.de

A 5' truncated non-LTR CR1-like retrotransposon, named RanaCR1, was identified in the serum albumin intron-1 (SAI-1) of at least seven species of western Palearctic water frogs. Like other CR1 elements, the 3' end of RanaCR1 is defined by a perfect octameric repeat. As a result of the transposition event the target site was duplicated flanking both the 5' and 3' ends of RanaCR1. Based on the carboxy-terminal region (CTR) of ORF2 and the highly conserved 3' untranslated region (3' UTR), RanaCR1 is closely related to CR1-elements found in the genome of Xenopus tropicalis. Length variation of water frog SAI-1 sequences is caused by deletions that extend in some cases beyond the 5' or 3' ends of RanaCR1, probably a result of negative selection against the insertion event. Unlike other CR1 elements, RanaCR1 contains a CA microsatellite upstream of the octameric repeats in its 3' UTR. The low nucleotide diversity of the 3' UTR compared to the CTR suggests that this region still has a functional role, probably within a species-specific transcriptional network. Both SAI-1 and RanaCR1 sequences support earlier hypotheses for water frog systematics based on mtDNA and protein electrophoretic data.

## (11)

# Short retroelement could induce RNAi leading to silencing of a relative LINE

*Nickolai A. Tchurikov* and Olga V. Kretova*

Engelhardt Institute of Molecular Biology Russian Academy of Sciences, Moscow, Vavilov str.32, 119991, Russia

Suffix and F element in Drosophila genome provide the first described example of short and long non-LTR retroelements sharing common sequences (Tchurikov et al, 1986; Di Nichera and Casari, 1987). Suffix corresponds to a 265-bp stretch at the 3' polyadenylated region of F element. Suffix does not have a sequence with similarity to the pol III promoter and thus do not belong to SINEs. This element was originally cloned from the cut locus (Tchurikov et al., 1982). In one Drosophila line, several 265-bp conserved suffix copies, flanked by different sequences, have been isolated (Tchurikov et al., 1986).

The separate copies of suffix are far more actively transcribed than their counterparts on the F element. Transcripts from both strands of suffix are present in RNA preparations during all stages of Drosophila development, providing the potential for the formation of double-stranded RNA and the initiation of RNA interference (RNAi).
Using in situ RNA hybridization analysis, we have detected the expression of both sense and antisense suffix transcripts in germinal cells. These sense and antisense transcripts are colocalized in the primary spermatocytes and in the cytoplasm of the nurse cells, suggesting that they really form double-stranded RNA. We performed further analyses of suffix-specific small RNAs using northern blotting and the nuclease protection assays and detected both siRNAs and piRNAs.

We further found by 3' RACE that in pupae and ovaries, F element transcripts lacking the suffix sequence are also present. Our data provide direct evidence that suffix-specific RNAi leads to the silencing of the relative LINE, F element, and suggests that SINE-specific RNA interference could potentially downregulate a set of genes possessing SINE stretches in their 5' or 3' non-coding regions in different genomes (concerted silencing).

References:

1. Tchurikov NA, Ebralidze AK, Georgiev GP (1986) The suffix sequence is involved in processing the 3' ends of different mRNAs in Drosophila melanogaster. EMBO J 5:2341-2347.
2. Di Nocera PP, Casari G (1987) Related polypeptides are encoded by Drosophila F elements, I factors, and mammalian L1 sequences. Proc Natl Acad Sci USA 84:5843-5847.
3. Tchurikov NA, Naumova AK, Zelentsova ES, Georgiev GP (1982) A cloned unique gene of Drosophila melanogaster contains a repetitive 3' exon whose sequence is present at the 3' ends of many different mRNAs. Cell 28:365-373.

## (12)

# The Origination and Evolution of Primate Retrogenes

*Qu Zhang*1, Amir Karger2, and Maryellen Ruvolo1*

Dept. of Biological Anthropology1, FAS Center for Systems Biology2, Harvard University

Novel genes created by retroposition have recently been shown to play essential roles in the evolution of lineage-specific phenotypic traits. Here, we present the first study that systematically surveys the impact of retroposition events in four primate genomes (human, chimpanzee, orangutan, and macaque). Using more extensive and more conservative strategies than previous studies, we found greater than 3,000 to 4,000 retrocopies in each species. Almost 3% of annotated genes are retrogenes in primates. Among these retrogenes, one third are generated by ancient retroposition events that occurred hundreds of millions of years ago, and these are presumably maintained by purifying selection. Zinc finger genes are particularly susceptible to retroposition, and human chromosome 19 (on which a large cluster of zinc finger genes is found) contains the greatest number of parental genes giving rise to functional retrogenes. Chromosome X also has an excess of retrogene parents, and this is possibly related to escape from X chromosome inactivation. Dating the retrocopies in a phylogenetic context reveals a peak of retroposition events around the divergence time of hominoids and Old World monkeys, more recent than previously reported. The rate of retrocopy recruitment is estimated to be 20-30 per million years, of which 2-3 copies are functional retrogenes. Molecular evolutionary analyses reveal that a relatively large fraction of retrogenes is quickly evolving under positive selection. Additionally, retrogenes are overrepresented in the functional category transcription factors, suggesting that they may play an important role in the fine-scale tuning of the transcriptome during evolution. Finally, the functional annotation of the macaque genome likely suffers from errors, as many pseudogenes have probably been incorrectly assigned as virtual genes.

## (13)

# Assessment of the extent of substitution rate variation of Long Terminal Repeat sequences in Oryza sativa and Oryza glaberrima

*Andrea Zuccolo*, Aswathy Sebastian, Yeisoo Yu and Rod A. Wing*

Arizona Genomics Institute, Department of Plant Sciences, BIO5 Institute, University of Arizona, Tucson, AZ 85721

Long Terminal Repeat retrotransposons (LTR-RTs) are widespread and ubiquitous in the plant kingdom where they constitute significant portions of several genomes. Useful information about the "history" of these elements in a genome is provided by the comparative study of their insertion times that can be inferred through the comparisons of the two retrotransposon LTRs, if the appropriate mutation rate is known.

Over the past several years, different mutation rates have been proposed for LTRs in crop plants. However very little is known about the extent of the mutation rate variation and the factors contributing to this variation, so the rates currently used are generally considered rough estimators of the actual rates.

To evaluate the extent of substitution rate variation in LTRs, we mined all orthologous LTRs on the chromosome 3 short arms of both Oryza sativa and Oryza glaberrima species. Seventy orthologous LTRs were isolated: since they were present in the common ancestor before the two species separated, the mutations in these regions have accumulated during the time elapsed from speciation event. This applies to all the orthologous sequences mined. This gives the opportunity to study the variation of LTR substitution rate in different elements across the short arm of the chromosome 3. For comparison purposes we investigated a similar amount of not repeat related sequences collected near the orthologous LTRs.

We demonstrated that the extent of the substitution rate variation in LTRs is greater than 5 fold, is positively correlated with GC content, and is negatively correlated with LTR-RT position along the chromosome. We confirmed that in the vast majority of cases, that LTRs mutate faster than their corresponding not repeat related neighboring sequences. Finally we discussed the effects of methylation on LTR mutation rate variation.

## (14)

# Evolutionary dynamics of the LTR retrotransposon roo inferred from twelve complete Drosophila genomes

*Nicole de la Chaux* 1, 2 and Andreas Wagner 1, 2, 3*

1: University of Zurich
2: The Swiss Institute of Bioinformatics
3: The Santa Fe Institute

Transposable elements (TEs) are mobile DNA sequences that can multiply in the genome, and that are major contributors to eukaryotic genome organization. They can be transferred both vertically from parent to offspring and horizontally, between organisms of the same or different species. While trans-species horizontal gene transfer is well-known in prokaryotes, there are fewer candidate examples in eukaryotes. We here study the phylogenetic distribution and evolution of the roo retrotransposon in 12 completely sequenced genomes of the fruit fly genus Drosophila. Roo is the most abundant retrotransposon in D. melanogaster. Its evolutionary origins and dynamics are thus of special interest for understanding the evolutionary history of Drosophila genome organization. We identified a total of 157 roo copies, 56 of which were previously unidentified copies that occur in 7 of the 12 genomes. Genomes rich in roo elements experienced recent transposition bursts.
The phylogenetic tree of roo is not congruent with the phylogenetic tree of Drosophila, which may indicate horizontal gene transfer among Drosophila species. This hypothesis, while not unequivocally supported by the data, is strengthened by amino acid divergence, and the solo-LTR distribution between roo elements from different Drosophila species.

## (15)

# Evolution of Can-SINEs in Felidae and other Carnivora

*Kathryn Walters*, Marc Allard*, Diana Johnson*, Jill Pecon-Slattery***

* The George Washington University, Department of Biological Sciences, Washington DC
** National Cancer Institute, Laboratory of Genomic Diversity, Frederick, MD

One major application of the rich resources now available from genome sequence projects is the resolution of complex mammalian taxonomy and systematics. SINEs comprise a class of genomic markers that have been lauded as 'perfect phylogenetic characters'. However, the practicality of using specific insertion loci as species markers is only now being fully explored. Through mining the canine and feline whole genome projects for SINEs unique to carnivores (or CAN-SINES), fifty insertion loci specific to the Felis catus genome were identified. The phylogenetic utility, accuracy and precision of these loci are assessed. Presence/absence status of thirty-seven loci has been determined for all extant members of Felidae. All of these loci are conserved among the Felidae lineages, and several are conserved amongst members of the Feliformia suborder. However, many novel Can-SINE insertion loci specific to certain exotic species were also discovered, the majority of which are consistent with existing phylogenetic information although some challenge current taxonomy. In addition, instances of presence/absence polymorphism within the Lynx genus were also found. To evaluate phylogenetic signal contained within DNA sequences from SINE insertions, twenty of the conserved loci were sequenced across Felidae. Distinct patterns of sequence variation were observed and phylogenies constructed with this data are largely congruent with current hypotheses of Felidae evolution. Thus, SINEs are appropriate tool for phylogenetic assessment, however analysis of these elements must be completed with an understanding of the caveats of transposable element behavior.

## (16)

# BNR – a LINE family with an ORF1 RNA-binding motif is present in a variety of higher plants and represents a novel L1 subclade

*Tony Heitkam* and Thomas Schmidt*

Dresden University of Technology, Institute of Botany, Chair of Plant Cell and Molecular Biology, Dresden, Germany

We report the identification and isolation of BNR, a novel LINE family from sugar beet (Beta vulgaris), which aquired an ORF1 structure different from all previously characterized plant or vertebrate LINEs. A conserved secondary structure motif, the RNA recognition motif, substitutes the zinc finger motif typical for plant LINEs and might have taken over its RNA-binding function.
The reference element BNR1 has a length of 6700 bp and an 8 bp poly(A)-tail, codes for two ORFs and created a 16 bp target site duplication upon integration. The specific transposition event of BNR1 could only be detected in two cultivars, indicating a relatively recent transposition of a probably still active element. BNR-like LINEs are predominantly localized in the subterminal heterochromatic regions of all sugar beet chromosomes and are widely spread and highly diverse with 64 % sequence identity in the genus Beta.
Homology searches revealed the presence of similar LINE-families in the genome of several higher plants such as poplar, lotus and soybean. These LINEs possess an RNA recognition motif in the ORF1. Additionally, when performing a phylogenetic analysis, their ORF2 sequences form a distinct branch in the L1 clade separating them from all other previously characterized plant LINEs. These similarities indicate a common origin of the BNR-like elements. Therefore, we suggest that these elements form a L1 subclade which we have designated “RRM subclade”.

## (17)

# LTR retrotransposons in filamentous fungi.

*A. Muszewska*, M. Grynberg and A. Paszewski*

Institute of Biochemistry and Biophysics, PAS, Pawinskiego 5A, 02-106 Warsaw, Poland

LTR retrotransposons are mobile elements that are characterized by the ability to replicate through reverse transcription and by the presence of flanking LTRs (long terminal repeats). The coding regions consist of gag genes related to structural proteins and pol genes crucial for transposition. LTR retrotransposons are grouped into families according to the reverse transcriptase amino-acid sequence similarity and by the order of encoded pol genes. Both Pseudoviridae (Ty1/copia) and Metaviridae (Ty3/gypsy) families are present in fungi. The number of encoded LTR retrotransposons is variable among fungal species.
Using LTR_harvest, LTR_finder, LTR_struct and blast programs we identified about 2000 complete LTR retrotransposons in 53 sequenced fungal genomes. Most transposons predicted and analyzed in our project were not annotated before. Predicted coding regions in all identified transposons have been analyzed. In our study we show that human and animal pathogenic fungi from the Onygenales family encode an elevated number of LTR retrotransposons. Four pathogenic species (Coccidioides immits, C. posadasii, Ajellomyces capsulatus and Paracoccidiodes brasiliensis) have genomes abundant in intact LTR retrotransposons whereas their close non-pathogenic relative, Uncinocarpus reesei, has only 3 complete elements in its genome. Pyrenophora tritici-repentis encodes the higest number of LTR retrotransposons from all analyzed genomes. This phenomenon can be related to non-effective genome defense mechanisms. Not only closely related species but even strains of the same species show different LTR retrotransposons content.

## (18)

# Reconstructing functional, virtual transposable elements from dense contigs of BAC-end sequences in Genbank

*Howard Laten*, Lauren Mogil, LaBianca Wright, Darya Isayeva, Donald Jonker, Eftychia Gouvas, Edward Badal, and Christi Lindorfer*

Department of Biology, Loyola University Chicago, Chicago, IL USA 60626

Retrotransposons constitute the majority of the protein coding regions of most eukaryotic genomes. Most genomes carry tens to hundreds of copies of multiple families but many, if not most, of these carry disabling mutations, including large indels. Regions rich in these elements are virtually ignored in all but the most complete genome sequencing projects. When individual retrotransposon genes are encountered, they are usually incorrectly translated by standard gene prediction programs because of numerous frameshift mutations and large indels. We have shown that many repetitive DNA families can be scavenged and pieced together from hundreds of short overlapping fragments that exist on separate clones that have been deposited in Genbank databases containing single-pass, BAC-end sequences. We have devised an in silico strategy to recover and reconstruct elements from long consensus sequences by building dense contigs ranging up to 14,000 bp. The results are hypothetical ancestral sequences that encode fully functional elements with intact open reading frames and identifiable cis-regulatory elements that can be phylogenetically characterized with respect to previously characterized elements.

## (19)

# Pre-selection of SINE-Loci for Pinnipeds phylogeny using CENSOR

*Christiane Schröder*[1], *Stefanie Hartmann*[2], *Christoph Bleidorn*[1] *& Ralph Tiedemann*[1],

[1]Evolutionary Biology/Zoology, Institute of Biochemistry and Biology, University of Potsdam, Germany, [2]Bioinformatics, Institute of Biochemistry and Biology, University of Potsdam, Germany

While the monophyly of the pinnipeds and the differentiation between phocids and otariids is well established, the placement of the walrus within the pinnipeds is continuously under debate. Among the discussed hypotheses, the walrusses have been considered to (1) represent the sister group to the otariids, (2) group within the family Otariidae, (3) constitute the sister group to the phocids, or (4) represent a taxon within the phocids. So far, sequence based phylogenies could not unambiguously resolve this issue. Short interspersed elements (SINEs) have been proposed as an essentially homoplasy-free phylogenetic character. However, their subsequent use has been limited due to the difficulty to establish SINE loci in non-model organisms. Making use of available genomic resources, we screened complete genomes of two carnivores (cat and dog) for SINE-containing introns. Using perl scripts, we first identified and selected for all introns ranging from 200-1000 bp. Using the program CENSOR (ftp://ftp.ncbi.nlm.nih.gov/repository/repbase/SOFTWARE/) we then identified loci containing SINEs, and based on comparisons between cat and dog sequence information we selected SINEs that are likely to have arisen in the Caniformia lineage. Thus we identified 130000 Introns in the dog genome, and of these are 45000 containing SINEs. Using this strategy, we were able to establish dozens of new SINE loci for non-model organisms.
By PCR and sequencing, we assessed absence/presence of these SINEs in a representative carnivoran taxon set. We discuss implications of our new data for pinniped phylogeny.

## (20)

# Mammalian imprinted genes derived from retrotransposition are regulated by maternal methylation and exert monoallelic gene expression on their hosts

*R.J. Oakey*, R.B. McCole, R. Schulz and A.J. Wood.*

Department of Medical and Molecular Genetics, King's College London

Imprinted genes undergo epigenetic modifications during gametogenesis, which lead to transcriptional silencing of either the maternally- or the paternally-derived allele in the subsequent generation. Previous work has suggested an association between imprinting and the products of retrotransposition, but the nature of this link is not well defined. In the mouse, four imprinted genes have been described that originated by retrotransposition and overlap CpG islands that undergo methylation during oogenesis. Nap1l5, U2af1-rs1, Inpp5f_v2 and Mcts2 are likely to encode proteins and share two additional genetic properties: they are located within introns of host transcripts and are derived from parental genes on the X chromosome. The orthologous human retrogenes NAP1L5, INPP5F_V2 and MCTS2 are also paternally expressed. The striking correlation between imprinting and X chromosome provenance suggests that retrotransposed elements with homology to the X chromosome can be selectively targeted for methylation during mammalian oogenesis. Furthermore, the host gene generally falls victim to imprinted expression as a consequence of intronic retrogene gene imprinting, but this victimisation does not extend beyond the two genes. This is in contrast to the organisation of the more complex imprinted domains, within which multiple mechanisms can act. A consequence of the arrangement of genes into intronic-host pairs is that of marked transcript diversity via the process of alternative polyadenylation.

## (21)

# A bioinformatic analysis of the association between small RNAs and transposable elements in Arabidopsis

*Alexis Sarazin*

E-mail: sarazin@biologie.ens.fr

Two major classes of small RNAs exist in Arabidopsis, one composed of a small number of highly abundant, mainly 21-nt microRNAs (miRNAs), and the other, of a large number of low abundance, mainly 24-nt small interfering RNAs (siRNAs). These 24-nt siRNAs correspond almost exclusively to repeat elements and are thought to act as guides for the methylation of these sequences. Thus, 24-nt siRNAs are an essential component of the silencing system that keeps transposable elements in check. Our recent findings indicate however that the 24-nt siRNA producing pathway acts primarily to enable efficient and faithful remethylation of its target sequences after accidental DNA methylation loss, rather than to maintain high methylation levels once they have been established. Moreover, we have found that only a subset of normally methylated repeats are associated with 24-nt siRNAs and hence are capable of remethylation. We now wish to identify the determinants that result in the accumulation of siRNAs over some methylated TEs only. To this end, we are conducting a systsematic bioinformatics analysis, combining published small RNA deep sequencing data, genome-wide DNA methylation profiles and a detailed annotation of TEs fragments. Results of this analysis will be presented.

## (22)

# Giant genomes of Fritillaria lilies are dominated by LTR retrotransposons

*Katerina Ambrozova1, Jiri Macas2, Ilia J. Leitch3, Martin A. Lysak1*

1 Department of Functional Genomics & Proteomics, Institute of Experimental Biology, Masaryk University, Czech Republic
2 Biology Centre ASCR, Institute of Plant Molecular Biology, Ceske Budejovice, Czech Republic
3 Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey, UK

The genus Fritillaria comprises species with the largest genome so far reported for any plant species, and at the same time, it provides extensive genome size variation (30,000-127,000 Mb/C). The genus diverged from Lilium c. 12 million years ago and subsequently diversified into two clades within Eurasia followed by a later dispersion of some species to North America. It is known, that genome size variation in plants is caused mainly by large differences in proportion of various repetitive sequences. Hence, the genus Fritillaria represents an excelent model for comparative studies on the evolution of repetitive sequences in giant genomes. To identify most abundant dispersed repetitive sequences we have constructed genomic DNA libraries from one species of each group selected so that both have approximately equal genome size (~ 45,000 Mb): Fritillaria imperialis (Eurasia) and F. affinis (North America). Based on the screening of both libraries with total genomic DNA as a probe, we selected four clones from each library containing highly repeated regions and sequenced them. Our preliminary data suggest that the percentage of different classes of repetitive elements does not differ significantly between the two species, and Ty3/gypsy-type LTR retrotransposons were identified as a dominating component in Fritillaria genomes. In both species, a Ty3/gypsy-type element related to the del1-46 retrotransposon isolated from Lilium henryi has been found exceptionally abundant. We also identified several families of Ty1/copia type LTR retrotransposons and one type of non-LTR retrotransposon. Further studies aims to characterize selected repetitive elements in a wide range of Fritillaria species and analyze their contribution to genome size variation across the genus. This work was in part supported by a research grant no. MSM0021622415.

**(23)**

## Lack of Evidence that Repeats in Short Indels in Tetrahymena deletion elements Form Parts of Longer Repeats.

Piroska Huvos*1, Michael Wainer2, undergraduates Jason Maas1, David Kluge,1 Amanda Harvey1, Kelly Ming1, Josh Newby2, Richa Ailinani2, Jason Horton2, Dan Scott2

1 Department of Biochemistry and Molecular Biology,
2 Dept of Computer Science
Southern Illinois University

There are over 6000 internally eliminated DNA sequences (deletion elements, or IESs) in the Tetrahymena genome that are deleted in a programmed fashion during the development of a transcriptionally active macronucleus from a transcriptionally inactive germline micronucleus. IESs are usually AT - rich and have a high repeat content. The sequences of only ten individual IESs are known: no common sequence features can explain their reproducible developmental deletion. Recently, based on several results, a homology and small RNA-based mechanism has been proposed for the efficient elimination of IES elements. To investigate the mechanism of generation of IESs in Tetrahymena, in a previous study I compared the sequences of selected IESs amongst different T. thermophila strains and Tetrahymena species. Three indels of 200-1800 bp were detected among strains of T. thermophila in four IES elements. A 500 bp and a 600 bp indel were found to be repeated a few hundred times in the genome. Although transposon-like sequences have been found in Tetrahymena by others, the short repeats participating in insertion-deletions and found inside IESs did not seem to be identifiable as transposons. We therefore attempted to determine whether the short repeats represent parts of longer ones that may have been truncated during their dispersal throughout the genome. We examined the 500 bp and the 600 bp indels as well as another 630 bp repetitive section. Preliminary data indicate that the sequences flanking the repeats cannot be regarded as members of a single sequence family, i.e. there is no evidence for the short repeats being part of longer ones. We also initiated a limited analysis of the genomic sequence data to examine the distribution of the repeats between macronuclear and IES sequences. Tools for further characterization of these sequences are sought.

## (24)

# Ancestral polymorphism or frequent hybridization? Curious presence of L1 retroelements in Mus spretus, M.macedonicus and M.musculus

*Václav Janoušek1, Pavel Munclinger1, Barbara Dod2 a Pierre Boursot2*

1 Deparment of zoology, Faculty of Science, Charles University in Prague, Viničná 7, 128 44 Praha 2, CZ
2 Institut des Sciences de l'Evolution, UMR5554, CNRS, Université Montpellier 2, Place E. Bataillon, F-34095 Montpellier, France

Retroelements constitute a powerful new tool for systematic biology and population genetics. Markers based on the presence/absence of a SINE or LINE element at a particular location are considered as essentially homoplasy free. In other words, the presence of an element represents identity by descent, since the probability of two different similar elements integrating independently in the same chromosomal location is negligible. While retrotransposon have been frequently used to address various questions concerning human origins and demography the mouse retroelements are far less studied than their human counterparts. This holds true especially for insertion polymorphisms in wild mice populations, an untouched field of research. The most abundant mouse LINEs are L1s. In order to test suitability of L1s as markers for mouse population studies we selected 196 recently active L1 loci using comparison of two house mouse inbred strains (C57BL/6J and MSM/Ms) and tested their presence/absence in closely related mouse species Mus spretus, M.macedonicus, and three subspecies of M. musculus (musculus, domesticus and castaneus). Unexpectedly 22 loci showed presence/absence polymorphism in M. musculus supspecies and simultaneously were also present in M.spretus or M.macedonicus. The observed pattern can be explained by parallel insertions, deletions by non-homologous recombination, interspecies hybridization or retained ancestral polymorphism. Since parallel insertions and retroelement deletions are not frequent phenomena we favour the hybridization and/or retained ancestral polymorphism as the most probable explanations.

## (25)

# Characterization of centromeric retrotransposons in diverse plant species

*Pavel Neumann (1), Alice Navratilova (1), Veronika Steinbauerova (1,2), Eduard Kejnovsky (3), Boris Vyskot (3), Eva Hribova (4), Jaroslav Dolezel (4), Jiri Macas (1)*

(1) Biology Centre ASCR, Institute of Plant Molecular Biology, Ceske Budejovice, Czech Repuplic
(2) University of South Bohemia, Faculty of Science, Ceske Budejovice, Czech Republic
(3) Institute of Biophysics ASCR, Brno, Czech Republic
(4) Institute of Experimental Botany ASCR, Olomouc, Czech Republic

Centromeric retrotransposons represent a distinct clade among chromoviruses. Their typical feature is a strong preference for integration into (peri)centromeric chromatin. Although chromoviruses are widespread among eukaryotes, this particular clade is unique for plants. So far, members of centromeric retrotransposons were described in detail in only a few species, most of which belonged to Poaceae family. Although evolutionary analysis of chromoviruses based on reverse transcriptase domain sequences implied presence of centromeric retrotransposons in a broader spectrum of plant species, the respective elements remained uncharacterized. Importantly, a clear proof of (peri)centromeric localization has also been missing for most of the uncharacterized elements. In this work we showed that centromeric retrotransposons are most likely widespread at least among angiosperms although their copy numbers vary greatly among different species. The structure of centromeric retrotransposons seems to be highly conserved. They possess a primer binding site complementary to tRNAmet and a single ORF encoding full Gag-Pol polyprotein. The ORF is extended into 3' LTR which seems to be a unique feature of centromeric retrotransposons. The C-terminus of the polyprotein contains a chromodomain which is supposed to allow targeted integration into (peri)centromeric regions. The localization within the (peri)centromeric chromatin was confirmed using in situ hybridization in case of all tested elements although some of them invaded also other parts of chromosomes e.g. subtelomeric regions.

## (26)

# Genetic and epigenetic characterization of Alu repeats in normal and cancer cells

*Mireia Jordà*, Berta Martín, Jairo Rodríguez, Víctor Barrera, Miguel A. Peinado*

Institut de Medicina Predictiva i Personalitzada del Càncer (IMPPC) , Barcelona (Spain)

Cumulated evidences indicate that the complexity of living organisms is not just a direct outcome of the number of coding sequences but repetitive elements may play a key role. On the other hand, a large body of data reveals that information other than the encoded within the DNA sequence, termed epigenetic, is required. Methylation of the cytosine within the CpG dinucleotide is the most common epigenetic modification in mammals and is considered a mark of long-term inactivation. Most of methylated cytosines are found in repetitive elements, being Alu sequences the most abundant. Alu's contain up to 33% of the total number of CpG sites and are highly methylated in most somatic tissues. Nevertheless, a fraction of Alu's remains unmethylated in normal cells and this proportion is increased during aging and cancer.
We have developed a novel methodology to quantify and identify unmethylated Alu sequences (Rodriguez et al., Nucleic Acids Res 36:770-776, 2008). These studies have revealed that normal colon epithelial cells contain about 25,000 unmethylated Alu's per haploid genome, while in tumor cells this figure is almost twice (p=0.004). Moreover, we have identified about 100 individual Alu elements exhibiting full or partial unmethylation in normal colonic mucosa and/or in colorectal carcinoma cells. Analysis of unmethylated elements at genetic and epigenetic level and comparison with randomly selected Alu's indicate differential traits of unmethylated Alu's regarding sequence, chromatin structure and genomic context. Results strongly suggest that unmethylated Alu elements constitute a defined compartment of "Aluome" with specific sequence properties located in active chromatin domains and that may play a regulatory role. Further studies on chromatin structure and expression of genes flanking unmethylated Alu's are being done to elucidate their functional implications and to understand the epigenetic changes taking place during the malignant transformation.

This work was supported by grants from the Spanish Ministry of Science and Innovation (SAF2008/1409 and Consolider Ingenio 2010 CSD 2009/49)

## (27)

# Discreet charm of neglected markers: SINE insertion polymorphisms in Western Palaearctic house mice

*Pavel Munclinger1*, Václav Janoušek1, Lenka Kubešová1, Barbara Dod2 a Pierre Boursot2*

1 Deparment of zoology, Faculty of Science, Charles University in Prague, Viničná 7, 128 44 Praha 2, CZ
2 Institut des Sciences de l'Evolution, UMR5554, CNRS, Université Montpellier 2, Place E. Bataillon, F-34095 Montpellier, France

The interest in the wild mouse genome has recently experienced resounding renaissance. Nevertheless, despite the availability of a complete mouse genome sequence our understanding of the genetic processes at a population level is still restricted to studies based on a very limited number of markers representing a small fraction of the genome. Besides several allozyme, RFLP and simple PCR loci mouse populations at a larger geographical scale are studied mainly using mitochondrial DNA. However, as the gene tree and population tree can differ substantially the history of a single locus does not necessarily reflect the true history of species or populations. Retroelement insertion polymorphisms (polymorphisms consisting of the presence/absence of an element at a particular chromosomal location) represent homoplasy free markers widely dispersed in genome that can serve as a desirable alternative to the mitochondrial DNA. The recently active LINE and SINE retroposons in mouse comprise mainly L1, and B1 and B2. We studied presence/absence polymorphism of 50 B2 and B1 loci in more than 100 Western Palaearctic house mouse individuals. 23 loci were found to be subspecies (Mus musculus domesticus /M. m. musculus) diagnostic markers. Genetic clustering algorithms detected clusters of individuals corresponding to subspecies and continental regions. However, several individuals were miss-assigned which probably suggest human mediated long distance migrations. The complex situation in Iran supports the view that in this region several mouse lineages meet and probably interbreed.

## (28)

# RetroTector online, a rational tool for analysis of retroviral elements in small and medium size vertebrate genomic sequences.

*Göran Sperber1, Anders Lövgren2 , Nils-Einar Eriksson2, Farid Benachenhou3 and Jonas Blomberg3**

1 Physiology unit, Dpt of Neuroscience
2 Linnaeus Centre for Bioinformatics, Biomedical Centre
3 Section of Virology, Department of Medical Sciences
Uppsala University, Uppsala Sweden.

The rapid accumulation of genomic information in databases necessitates rapid and specific algorithms for extracting biologically meaningful information. More or less complete retroviral sequences, also called proviral or endogenous retroviral sequences; ERVs, constitute 5-50% of vertebrate genomes. After infecting the host, these retroviruses have integrated in germ line cells, and have then been carried in progeny genomes for up to several 100 million years. A better understanding of structure and function of these sequences can have profound biological consequences.

RetroTector© is a platform-independent Java program for identification and characterization of proviral sequences in vertebrate genomes. The full version requires a local installation with a MySQL database. Although not overly complicated, the installation may take some time. We have now created a "light" version of RetroTector©, (RetroTector online; ROL) which does not require specific installation procedures, and which can be accessed via the world wide web.

ROL (http://www.neuro.uu.se/fysiologi/jbgs) was implemented under the Batchelor web interface (A Lövgren et al, unpublished). It allows both GenBank accession number, file and FASTA cut-and-paste admission of sequences (5 to 10 000 kilobases). Up to ten submissions can be done simultaneously, allowing batch analysis of <= 100 Megabases. Jobs are shown in an IP-number specific list. Results are downloadable as text files, and can be viewed with a stand-alone program, RetroTectorViewer.jar (downloadable from the same site), which has the full graphical capabilities of the basic RetroTector© program. Thus, a detailed analysis of any retroviral sequences found in the submitted sequence is graphically presented, and can be exported in standard formats. A complete analysis of a 1 Megabase sequence is complete in under 10 minutes. Nonretroviral repetitive sequences in the submitted sequence can be masked before analysis, using host genome specific "brooms".

RetroTector online is a rational tool for retrovirological and genomic work.

## (29)

# Genome wide analysis of Sleeping Beauty (SB), piggyBac and Tol2 transposon mediated integration in human primary T cells

*Xianzheng Zhou*

E-mail: zhoux058@umn.edu

We have reported that the Sleeping Beauty (SB) transposon system can mediate genomic integration and long-term reporter gene expression in human peripheral blood (PB) T cells. Both PB and umbilical cord blood (UCB) derived T cells after SB mediated therapeutic gene expression can kill leukemia and lymphoma cells in vitro and in mice. In this study, we directly compared the genomic integration efficiencies and transposition site preferences of SB, piggyBac, and Tol2 in both PB and UCB T cells. It was found that piggyBac demonstrated the highest efficiency of stable gene transfer in PB T cells, whereas SB and Tol2 mediated intermediate and lowest efficiencies, respectively. Southern blot analysis demonstrated that integrants could be detected in all three transposon transfected T cell clones. Using recoverable cassette constructs derived from each transposon, we sequenced approximately 3,000 integration sites in PB and UCB T cells from two different donors and found as follows. (1) There was no significant difference between PB and UCB T cells transfected by each transposon. In addition, variation between each donor was minimal. Therefore, all the DNA sequences from the same transposon transfected cells were pooled together for comparison. (2) Integration by SB, piggyBac, and Tol2 occurred in all the chromosomes with no preference. (3) Integration sites by piggyBac and Tol2 were mainly localized near the transcriptional start site, whereas SB integration sites were randomly localized within 5Kb upstream and downstream regions of the transcriptional start site. (4) In agreement of the finding (3), SB integrations were random, whereas piggyBac and Tol2 integrations were close to CpG islands and DNase hypersensitive sites. These results suggest that SB might be safer than piggyBac and Tol2 in T cell engineering and imply that SB mediated T cell engineering for leukemia therapy could be moved to clinical trials.

## (30)

# Inter- and intra-family variation in transposable element dynamics among natural populations of a cyclical parthenogen, Daphnia pulex

*Sarah Schaack*(1), Abby Wolf(1), Ellen Pritham(2), and Michael Lynch(1)*

1 Indiana University
2 University of Texas Arlington

The evolutionary dynamics of transposable elements (TEs) are thought to vary among types of TEs, among hosts, and as a function of specific characteristics of the host environment. In order to investigate the inter- and intra-family variation in TE dynamics and to examine the role of recombination in TE evolution, we surveyed 6 DNA transposon families, representing 4 super-families, in 57 populations of Daphnia pulex. These populations include sexuals (where meiotic recombination happens approximately every year) and asexuals (where sexual reproduction has been lost).
We find a greater mean number of insertions in sexual populations for some, but not all, TE families. The differences observed (for Tc1-2, Pokey, and Helitron1 and Helitron2; but not for Tc1-1 or hAT) suggest that the pattern is not consistent among families within a super-family, and instead suggests that there is some other explanatory variable. We observed that the proportion of occupied sites varies little among sexual populations for each family. However in asexuals the proportion of occupied sites decreases when mean copy number of the family is low. Low copy number families (LCNFs) may represent recent invaders, or TEs against which selection is particularly strong. The higher relative difference in site occupation between sexuals and asexuals for LCNFs presents another line of evidence suggesting these families–in particular–experience stronger negative selection. In addition, the relative difference in the proportion of singletons between sexuals and asexuals also increases for LCNFs. Lastly, we observe higher levels of polymorphism (based on mean pairwise distances) in sexual versus asexual populations among all families. Given the strong relationship with copy number, but not phylogeny, these data suggest evolutionary relatedness does not govern trends in TE dynamics for these families. Instead, contrasting patterns among sexuals and asexuals reveals negative selection against different families may explain major trends.

## (31)

# Discovery of the first primate endogenous lentivirus

*Clément Gilbert*, David Maxfield, Cédric Feschotte*

Department of Biology, University of Texas at
Arlington, P.O. Box 19498, Arlington, TX 76019

Lentiviruses are the focus of intense study since the discovery of their causal association with AIDS. Here we report on the characterization of the first primate endogenous lentivirus, called LELV, for Lemur Endogenous Lentivirus, in two genera of Malagasy lemurs, Microcebus and Cheirogaleus. Phylogenetic analyses indicate that LELV is most closely related to the HIV/SIV clade of simian lentiviruses. In addition to the predicted gag, pol, env and rev genes present in all lentiviruses, LELV contains a dUTPase domain, which is found in non-primates lentiviruses but not in modern simian lentiviruses and two putative accessory genes located between pol and env with no similarity to those of other lentiviruses. Thus LELV is best viewed as a molecular fossil representing a structural intermediate between primate and non-primate lentiviruses. We used three different approaches to estimate the timing of LELV germline integration in the Microcebus lineage, which converge to a period of 5 to 11 million years ago, i.e., after this genus diverged from Cheirogaleus around 23 mya. These data, together with the high level of sequence identity of LELV between the two lemur genera, suggest independent endogenization events. These results provide evidence that lentiviruses have infected prosimian species and successfully infiltrated their germline, and they reveal that primates have been exposed to lentiviruses for a much longer time than previously inferred based on sequence comparison of circulating lentiviruses.

## (32)

# Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods

*John K. Pace II, Clément Gilbert*, Marlena S. Clark, and Cédric Feschotte*

Department of Biology, University of Texas at Arlington, P.O. Box 19498, Arlington, TX 76019

Horizontal transfer (HT) is central to the evolution of prokaryotic species. Selfish and mobile genetic elements, such as phages, plasmids, and transposons, are the primary vehicles for HT among prokaryotes. In multicellular eukaryotes, the prevalence and evolutionary significance of HT remain unclear. Here, we identified a set of DNA transposon families dubbed SPACE INVADERS (or SPIN) whose consensus sequences are _96% identical over their entire length (2.9 kb) in the genomes of murine rodents (rat/mouse), bushbaby (prosimian primate), little brown bat (laurasiatherian), tenrec (afrotherian), opossum (marsupial), and two non-mammalian tetrapods (anole lizard and African clawed frog). In contrast, SPIN elements were undetectable in other species represented in the sequence databases, including 19 other mammals with draft whole-genome assemblies. This patchy distribution, coupled with the extreme level of SPIN identity in widely divergent tetrapods and the overall lack of selective constraint acting on these elements, is incompatible with vertical inheritance, but strongly indicative of multiple horizontal introductions.We show that these germline infiltrations likely occurred around the same evolutionary time (15–46 mya) and spawned some of the largest bursts of DNA transposon activity ever recorded in any species lineage (nearly 100,000 SPIN copies per haploid genome in tenrec). The process also led to the emergence of a new gene in the murine lineage derived from a SPIN transposase. In summary, HT of DNA transposons has contributed significantly to shaping and diversifying the genomes of multiple mammalian and tetrapod species.

## (33)

# The Contribution of L1 Transcription to Variation in the Human Transcriptome

*Sanjida H Rangwala*, Lili Zhang and Haig H. Kazazian Jr.*

Department of Genetics, University of Pennsylvania School of Medicine,
Philadelphia, PA 19104

LINEs (Long Interspersed Nuclear Elements) are a class of non-LTR retroelement present throughout eukaryotic evolution. Mobilization of LINEs can have a dramatic impact on the organization and content of genomes, often resulting in insertional mutagenesis, misregulation, or sequence shuffling of essential genomic regions. L1 is the major family of human LINE elements, making up close to 20% of the human genome. Because only elements that produce RNA can move, we have taken an approach that focuses on expression polymorphisms in order to highlight insertions that are likely to have a direct impact on shaping the genome. We have developed variations of 3' and 5' RACE to identify unique sequence information from the insertion site of expressed L1 elements. 3' RACE tagging has been used to identify over 175 unique transcript sites, most of which were not previously represented in the EST databases. Selected elements have been examined in a panel of normal individuals in order to determine the extent of expression polymorphism present in human populations. L1 expression in the absence of mobilization might also be important in de-repressing the chromatin of the genomic region, resulting in a more permissive transcriptional environment. Therefore, we are also characterizing the transcriptional start sites of expressed L1 elements as well as the expression of adjacent protein-coding genes in L1 expressing and non-expressing individuals. This research should reveal a deeper understanding of the ecology of L1 elements in the genome.

## (34)

# Discovery and Annotation of TEs on VectorBase

*Ryan C. Kennedy*, Maria F. Unger, Scott Christley, Jenica L. Abrudan, Neil F. Lobo, Greg Madey, Frank H. Collins*

Eck Institute for Global Health, University of Notre Dame

Although transposable elements (TEs) were discovered over 50 years ago, the robust discovery of them in newly sequenced genomes remains a difficult problem. Numerous types with different structural characteristics, sequence degradation, multiple insertions within existing elements, and co-option by the organism's regulatory system are some of the issues confounding the discovery process. We have developed an automated pipeline employing a homology-based approach, complemented with de novo- and structure-based approaches, to discover and annotate TEs in invertebrate genomes. Once fully automated, our pipeline will be integrated with VectorBase, an NIAID Bioinformatics Resource Center for invertebrate vectors of human pathogens, to produce a first-pass discovery and annotation of TEs for newly sequenced genomes. Currently hosting five organisms with more on the way, VectorBase provides the Ensembl genome browser, computational tools and other data specific to the study of invertebrate vectors. The annotation component of our pipeline includes enhancements to the Ensembl genome browser, elevating the importance of TEs by displaying genomic location, structural details, alignments with consensus TEs, and homology with other organisms. VectorBase has developed a community annotation system whereby the research community can upload annotation corrections to genes for curation and broad dissemination; we plan to extend this to TEs. We hope this will provide an invaluable resource for researchers studying the biology of TEs and their genomic impact.

## (35)

# Integration Mechanism of the Site Specific Non-LTR retrotransposon R2Bm

*Christi Hull and Shawn Christensen, Ph.D.*

Biology Department, University of Texas Arlington

Non-LTR retrotransposons integrate into new chromosomal sites by reverse transcribing the element RNA directly at the site of insertion using an exposed chromosomal free 3'-OH to prime cDNA synthesis, a process known as target primed reverse transcription (TPRT). While TPRT is well characterized, how elements complete integration remains largely a mystery. For example, how the second DNA strand is generated is unresolved. The R2 element from Bombyx mori (R2Bm) has proved to be a useful model for biochemically dissecting non LTR retrotransposition. Two subunits of R2, in the form of an RNP, are known to site specifically bind target DNA in opposite orientation to each other, nick the DNA strand used to prime cDNA synthesis, perform cDNA synthesis, and nick the opposing DNA strand. The R2 reverse transcriptase is able to catalyze DNA templated DNA polymerization and displace RNA from DNA-RNA hybrids; the activities required for second strand synthesis. We are generating integration competent RNPs by complexing R2 protein with mini R2 RNAs; these RNPs are then used in in vitro transposition assays. Preliminary experiments indicate that integration events are occurring but are inefficient. An integration competent RNP is expected to be two R2 proteins bound to a single element RNA. It is hoped that using purified RNPs, rather than an undefined mixture of RNPs, will increase integration events. Further, a phased nucleosome can exists at the site of insertion (in vivo and in vitro). It is possible that integration can be improved by using histone bound target DNA.

## (36)

# Target Site Recognition of Site Specific Non-LTR retrotransposons

**Blaine Thompson and Shawn Christensen†, PhD*

Biology Department, University of Texas Arlington
†corresponding author, *poster presenter

R2 and NeSL-1 are members of the site specific non-LTR retrotransposons which can copy (retrotranspose) through an RNA intermediate into either the 28S rDNA or the Spliced Leader-1 (SL-1) locus, respectively. It has been shown that the R2 element from Bombyx mori (R2Bm) requires two subunits of protein, one bound upstream of the insertion site and one bound downstream of the insertion site, to integrate into DNA. R2 and NeSL have similar structures and will be used to study how site specific elements recognize target DNA. R2 elements have a variable number of nucleic acid binding motifs in their N-terminal region, yet they all bind 28S rDNA. We are investigating the mechanistic implications of this variability. The NeSL element has two nucleic acid binding motifs in its N-terminal domain, we show that these are involved in target recognition. In R2Bm, the animo terminal domains are responsible for binding the downstream subunit to the target DNA. The domain responsible for binding the upstream R2 subunit is unknown, but we have identified a putative carboxyl terminal myb domain which may fulfill this function.

NeSL-1 contains a Ubiquitin like protease domain. We have expressed, purified, and tested the protease. The protease efficiently removes Sumo, a common post-translational modifier of protein function, from Sumoylated proteins invitro. While the exact function of the protease during transposition remains unknown, we believe that the protease likely desumylates host proteins to either gain access to the DNA or to combat host response (i.e., host parasite conflict).

## (37)

# Transposable elements characterization using the REPET framework in four insect genomes

*Emmanuelle Permal*, Timothée Flutre, Hadi Quesneville*

Unité de Recherche en Génomique et Informatique – UR1164, INRA,
Route de St Cyr – RD 10 78000 Versailles, France.
Contact: Emmanuelle.Permal@versailles.inra.fr

Transposable elements (TEs) are key elements of genome plasticity and account for a large part of many eukaryotic genomes. Their annotations in several genomes is then of major interest. We applied REPET (Quesneville et al. 2005), an efficient transposable element annotation framework divided in two complementary pipelines, on four insect genomes : Acyrthosiphon pisum, Spodoptera frugiperda, Helicoverpa armigera and Bombyx mori.
We were interested in any characteristics that can be revealed by the comparison of the TE distribution of these genomes.
Using methods for de novo TE identification and characterization, the REPET framework (Quesneville et al. 2005) analysed the genomes by:

(i) Searching repeats with BLASTER for an all-by-all genome self comparison,
(ii) Grouping results using three clustering methods: GROUPER, RECON and PILER,
(iii) Building one consensus per group with a multiple sequence alignment program,
(iv) Classifying each consensus according to structural and coding TE features.

Then the annotation pipeline of the REPET framework annotated the four genomes with the TE library produced de novo. This was done by:

(i) Detecting the TEs with BLASTER, RepeatMasker and Censor softwares
(ii) Finding the satellites with RepeatMasker, TRF and Mreps softwares.

In parallel, consensus sequences representing ancestral copies of TEs subfamilies were clustered into groups to identify TE families by using the GROUPER clustering method.

The REPET framework provided de novo TE libraries, delivered good quality annotations of TEs along the genome, allowed us to evaluate the distribution of TE families, and gave a hint on the dynamics of these genomes.

Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D et al. (2005) Combined evidence annotation of transposable elements in genome sequences. PLoS computational biology 1(2): 166-175.

## (38)

# Retrotransposon nik: an infective retrovirus?

*Adriana Ludwig* 1; Vera Lúcia da Silva Valente. 1; Elgion Lúcio da Silva Loreto,1,2*

1 PPG Genética e Biologia Molecular, UFRGS
2 Departamento de Biologia, UFSM

The insect endogenous retrovirus nik, also named gypsy5, was discovered by the analysis of Drosophila melanogaster genome sequence. Aiming to understand the evolutionary history of this element, we conducted in silico searches in the twelve Drosophila genomes. Sequences homologous to nik were found only in the melanogaster subgroup species. Interestingly, there are few copies of this element in these genomes. The presence of intact copies in D. simulans and D. melanogaster suggests that this element may be still active. To further define the distribution of nik and infer its phylogenetic relationship a survey by PCR amplification was performed in 74 Drosophilidae species. Amplification products of expected length (550 bp) were obtained in several species of the melanogaster subgroup (D. melanogaster, D. simulans, D. teissieri, D. yakuba, D. erecta, D. sechelia, D. santomea and D. mauritiana) and in more distant Neotropical species from subgenus Drosophila, such as D. crocina, D. tripunctata, D. gasici and D. polymorpha, as well as in Scaptodrosophila latifasciaeformis. The phylogenetic analysis revealed that the members of the nik family can be grouped into at least five genetically differentiated clusters. In general, the similarity degree of the studied sequences and their discontinuous distribution among species make it highly unlikely that these sequences diverged at the same time as their host taxa, suggesting the occurrence of multiple nik horizontal transfers among Drosophilidae species. Preliminary assay of infection showed that an "empty" strain can acquire nik element when maintained on medium containing homogenized pupae from a strain detaining the element nik. Our results have raised the question if nik element could be an infectious retrovirus, like gypsy, which can easily leave a species and infect another one. Further analysis will help us answer this intriguing issue.

## (39)

# Differential insertion of transposable elements in Anopheles gambiae M & S genomes

*Jenica L. Abrudan*, Ryan C. Kennedy, Maria F. Unger, Michael R. Olson, Scott J. Emrich, Frank H. Collins, Nora J. Besansky*

Eck Institute of Global Health, University of Notre Dame

Mosquitoes in the Anopheles gambiae species complex are the major vectors of malaria in Africa. The original A. gambiae genome sequenced was the PEST strain, which was later discovered to be a composite of the A. gambiae M and S forms. These 2 sympatric forms demonstrate reproductive isolation and are believed to be incipient or different species. They have been individually sequenced recently, so we are performing computational analysis of the three genomes to identify sequence differences. We hypothesize that transposable elements may be influencing the speciation of A. gambiae.

Insertions of transposable elements have been associated with alterations in chromosome structure, recombination, replication, and gene regulation. Recent studies have indicated the existence of “speciation islands” and numerous genes differentially expressed across multiple developmental stages between the M and S forms though many of those genes lie outside of the “speciation islands” implying there are more causal factors to be discovered.. We have identified sequences differently inserted between the M & S genomes relative to PEST. We then identified the subset of those sequences that contain transposable elements using a discovery pipeline we have developed. We are currently using this subset of data to identify those sequences that are in close proximity (~1kb) to gene elements, and will perform experiments designed to measure the expression levels of those genes. We hope to find a correlation between the differentially inserted transposons and the observed gene expression differences .

## (40)

# The landscape of transposable elements in the soybean genome

*Jianxin Ma*

E-mail: maj@purdue.edu

The draft sequencing of the soybean genome has provided a unique opportunity to study structural and evolutionary dynamics of transposable elements in this economically important legume crop species. Using a combination of structure-based analysis and homology-based comparison, we have identified 12600 intact LTR-retrotransposons and 18300 solo-LTRs, which are classified into >500 distinct families with the copy numbers ranging from 1 to 1253. Of these families, 60% are gypsy-like elements, and the rest are copia-like elements. These elements, together with numerous truncated fragments or remnants, make up ~35% of the soybean genome. Based on the chromosomal distribution of these elements and their association with soybean centromere satellite repeats, three centromere-enriched families have been identified. Our data suggest that the majority (78%) of the intact LTR-retrotransposons were amplified in the past 3 million years, and 43% of them have undergone bursts in the past 1 million years. On the other hand, rapid removal of retrotransposon DNA by unequal recombination and illegitimate recombination has played a major role in counteracting the expansion of soybean genome caused by rapid proliferation of LTR-retrotransposons. Currently, we are identifying DNA transposons, such as Mutator and CACTA, and Helitron elements, and have obtained a preliminary collection. These transposable element datasets will facilitate the annotation of soybean genes and lay a foundation for further study of the organization, structure and evolution of the soybean genome.

## (41)

# Bioinformatic detection and annotation of non-LTR retrotransposons in the Culex quinquefasciatus mosquito genome.

*Maria F. Unger*, Ryan C. Kennedy, Jenica L. Abrudan, Frank H. Collins*

Eck Institute of Global Health, University of Notre Dame

We have conducted an extensive computational analysis of the Culex genome to find and annotate a specific subfamily of the TEs: Class-I non-long terminal repeat retrotransposons (non-LTRs), by building a semi-automated pipeline.

Initially we conducted BLAST searches to find the similarity to the known non-LTRs using amino acid sequences of Reverse-Transcriptase (RT) of known non-LTRs as the starting queries. Consequently Blast-hits (DNA sequences) were combined and extracted utilizing PERL scripts, to obtain non-LTR candidates of Culex. These sequences were than assembled using SEQMAN module of DNA-STAR, manually truncated, adjusted, and annotated.

Annotation was done by two steps: 1.- we annotated all the sequences using BLAST to nr database (NCBI), and identified some of Culex non-LTR consensuses as belonging to known non-LTR families; 2.- we conducted phylogenetic analysis on all Culex non-LTRs, allowing us to further successfully annotate our consensus sequences. Some of the elements were deteriorated and not possible to classify as a specific clade.

Upon completing preliminary annotation a copy number of each element in the genome within the threshold was found. Comparison between Aedes aegypti, Anopheles gambiae and Culex, has shown different non-LTR family composition, suggesting a different evolutionary development of these species.

## (42)

## The relation of short motif overrepresentation in the human genome and the fragmentation of segmental duplications and transpositions

*Abanish Singh*, Umeshkumar Keswani, David Levine, Nikola Stojanovic*

Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019, USA

Regulatory signals in DNA have been studied for decades, and over the last few years (especially after the sequencing of several major eukaryotic genomes has been completed) the search techniques became primarily computational. They focused on sequence features, such as periodicity or motif overrepresentation, or relied on phylogenetic conservation, homology, and data mining. Finding putative functional elements through motif overrepresentation has been moderately successful, but usually suffered from many false positives. We have thus undertaken a study [1] of the distribution of short motifs in repeat-masked human genomic sequences, and revealed a remarkable microrepetitive pattern of many motifs occurring for an order of magnitude more frequently than expected by chance, in random sequences as well as in regions immediately upstream of known genes.

Since only about half of the human genome is considered repetitive, and a very small fraction appears to be under functional constraint, we became interested in the origins of the presumably unique non-functional sequence. It was intuitive that this sequence also derives from duplication activity, with degenerated copies, unrecognizable by current tools, leaving short remnant motifs dramatically overrepresented throughout the genome. In an attempt to prove this, we have developed software [2] to associate the co-occurring short statistically overrepresented motifs in the human genome into larger elements.

Our program was able to identify about 4.58% of presumably unique human sequence as repetitive, and build the consensuses of these elements, many repeated hundreds of times. Only a small fraction (around 20%) of them has been previously identified as segmental duplications, and we are currently working on the characterization of these elements. So far only a few have shown the hallmark features of transposons.

References:

[1] Singh, A., C. Feschotte and N. Stojanovic (2007) A study on the repetitive structure and distribution of short motifs in human genomic sequences. International Journal of Bioinformatics Research and Applications 3(4), 523-535.

[2] Singh, A., C. Feschotte and N. Stojanovic (2007) Micro-repetitive structure of genomic sequences and the identification of ancient repeat elements. In Proceedings of BIBM 2007, The IEEE International Conference on Bioinformatics and Biomedicine, 165-171.

## (43)

# Transposons Shape the Architecture of Plant Genomes

*Heidrun Gundlach 1, Georg Haberer 1, Rémy Bruggmann 2 and Klaus F.X. Mayer 1*

1 MIPS/IBIS Inst. for Bioinformatics and System Biology, Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Neuherberg, Germany (http://mips.gsf.de)
2 Plant Genome Initiative at Rutgers (PGIR), Waksman Institute, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA (http://pgir.rutgers.edu)

Transposable elements constitute a major part of plant genomic DNA and are often responsible for drastic increases in genome size. Their overall amount ranges from 15 to over 90 %. The large genome size variations even between related taxonomic groups pose many intriguing questions regarding the concerted evolution of transposons and host plant.
Within the family of grasses diploid genome sizes vary between 0.3 Gb up to ~8 Gb. All grasses have evolved within the last 60 mio years from a common ancestor and have retained large stretches of syntenic regions, with still conserved gene order. However the intergenic space has evolved much more rapidly under the influence of different transposon families.

We have established workflows and analyses protocols for an exhaustive structural annotation not only of the genic, but also of the repetitive space of plant genomes. An in-depth repeat annotation has to deal with element identification, defragmentation and the reconstruction of insertion events. Our repeat annotation concept is based on mips-REcat, a generic repeat element classification catalog, and mips-REdat, an exhaustive database of plant repeat elements. The detection layer of our ANGELA pipeline (Automated Nested Genetic ELement Analysis) combines intrinsic repeat detection approaches with homology based methods. The processing layer integrates genes or other additional data. It handles element overlaps, followed by the identification of nested structures and the timing of LTR retrotransposon insertion events. Implemented standard evaluations cover composition tables, copy numbers, target-insert pair preferences or insertion age distributions. Heat maps are used for the visualisation of chromosome organization. The Apollo synteny viewer allows a detailed exploration of syntenic regions and their breakpoints. The presentation will compare the influence of transposons on the chromosomal architecture in different plant species and give insights into the fine structure of selected syntenic regions.

## (44)

# Automatic identification of non-LTR retrotransposons in genomic sequences

*Mina Rho* and Haixu Tang*

School of Informatics, Indiana University, Bloomington, IN 47408, USA

Non-long terminal repeat (non-LTR) retrotransposons are a class of mobile genetic elements (MGEs), which play important roles in the evolution of eukaryotic genomes. As more genomes have been fully sequenced, the computational methods for genome-wide identification of MGEs are necessary for genome annotation and comparative studies. In this presentation, we describe a computational framework to identify non-LTR retrotransposons in the whole genomic sequences by using a generalized hidden Markov model (HMM), which allows the emission of sequences with flexible lengths defined by a duration density distribution. In order to classify the identified non-LTR retrotransposons into fourteen clades, in the model we defined separate hidden states for each clade. In general, the hidden states were used to represent the sequences encoding protein domains and the linker regions (between domains). These domain encoding regions were modeled by profile HMMs, where as the linker regions were modeled by Gaussian Bayes classifiers. Our model was tested on two genomic sequences of Drosophila melanogaster and Daphnia pulex. The non-LTR retrotransposons in the D.melanogaster genome have been collectively identified a total of 118 intact elements in the previous research. Our approach described here alone found 104 of these known elements and classified them into three clades, I, Jockey, and R1, automatically. Notably, our model identified significantly larger numbers (93) of intact non-LTR retrotransposons in the D. pulex genome, compared with the results from RepeatMasker
against the current version of RepBase Update. The most abundant clade in D. Pulex is L2, which comprise approximately half of the identified non-LTR retrotransposons.
Finally, we will also discuss the phylogenetic analysis carried out with the
identified retroelements in the D. pulex genome.

## (45)

# Transposable elements and evolutionary plasticity of the malaria mosquito genome

*I.V. Sharakhov*, A. Xia, M.V. Sharakhova, S.C. Leman, Z. Tu*

Virginia Tech, Blacksburg, VA, USA

An. gambiae, An. funestus, and An. stephensi are primary malaria vectors and are members of different series of the subgenus Cellia. These species are highly polymorphic for inversions which are associated with ecological adaptations and distributed non-randomly on five chromosome elements. The goal of this study was to determine the rates of inversion fixation and to identify molecular features associated with the rearrangement breakpoints. We used 231 uniformly distributed DNA markers for comparative mapping of An. gambiae, An. funestus, and An. stephensi. We used the Nadeau and Taylor model to find the expected length of conserved synteny regions. The Bayesian analysis showed that our data fit the random breakage model. The analysis using the Genome Rearrangements In Man and Mouse (GRIMM) program revealed that the X chromosome has the highest rate of inversion fixation whereas autosomes vary in the inversion density: 2R>2L>3R>3L. Another remarkable observation was a significant positive correlation between polymorphic inversions and fixed inversion on autosomes. The analysis of the An. gambaie genome identified a significant negative correlation between the number of fixed inversions and the density of Matrix/Scaffold Attachment Regions (M/SARs) suggesting a role of nuclear architecture in determining the chromosome specificity of rearrangement rates. M/SARs can potentially mediate an interaction of specific chromosome sites with a nuclear envelope and affect inter-chromosomal interactions. In addition, we found a positive correlation between the rates of inversion fixations and the simple repeat content on five chromosomal arms. The comparative analysis of breakpoint regions and synteny blocks revealed significant at least two-fold enrichment of AT-repeats, inverted repeats, and transposable elements in breakpoints. Interestingly, there is 7-fold difference in density of retrotransposons between breakpoint regions and synteny blocks for arms 2L, 3R, and 3L. These data suggest involvement of simple DN!
A repeats and transposable elements in facilitating rearrangements.

## (46)

# Searching for horizontal transfer of transposable elements within grass genomes

*Aaron O. Richardson* 1 and Susan R. Wessler 1*

Department of Plant Biology, University of Georgia, USA

Successful horizontal transfer requires that DNA physically travel from one organism to another and then is incorporated into the recipient genome. Transposable elements (TEs) are good candidates because they have active mechanisms to insert themselves into new sites in a genome. Numerous compelling examples of TE horizontal transfer have been studied in animal genomes, but the extent of the phenomenon is unclear in plants. I use computational approaches to identify and analyze transposable elements from sequence data available for members of the grass family (Poaceae). As an example, my maximum likelihood analyses reveal evidence for at least five active CACTA elements at the base of the AA, BB, and CC clade within the genus Oryza. The descendent elements of each active ancestor largely recapitulate the organismal phylogeny, but one O. sativa CACTA element does not fit this pattern and is therefore a candidate horizontal transfer. Ongoing work seeks to refine the provenance of this element. I am also using molecular evolution and population genetic analyses to better understand how selection shapes TE evolution. One O. sativa-specific clade of CACTA elements appears to have undergone a recent expansion to several hundred elements with over 99% similarity at the nucleotide level. These elements were subject to purifying selection during the expansion, with a Ka/Ks ratio of 0.5. However a comparison with sister conspecific elements that split prior to the expansion shows an excess of fixed coding differences separating the burst and non-burst clades, which may be due to a history positive selection. I hope to use such techniques to identify sites associated with increases and decreases in element proliferation, and use the evolutionary inference to guide wet lab experiments into the genetics of element activity.

## (47)

# Different contribution of retroelements to intergenic transcription units

*Saneyuki Higashino*1, Hidemasa Bono2, Kunio Kikuchi1 and Yasunori Aizawa1, 3*

1 Graduate school of Bioscience and Biotechnology, Tokyo Institute of Technology, Kanagawa, Japan
2 Database Center for Life Science, Tokyo, Japan
3 Center for Biological Resources and Informatics, Tokyo Institute of Technology, Kanagawa, Japan

Retrotransposons have a great impact on their host genome. Over 40% of the human genome is composed of retrotransposon-derived elements (retroelements, REs). Human REs are mostly incapable of additional retrotransposition but in some cases known to function as gene regulatory elements: promoters, enhancers, splice sites, polyadenylation (polyA) signals and so on. However, the vast majority of REs are spread in the intergenic region and regarded as non-functional sequences.

On the other hand, recent mammalian transcriptome studies demonstrated more intergenic regions were transcribed than had been expected. Although these intergenic transcripts are less conserved than known genes, some of them were identified as functional transcripts. This implicates that these intergenic transcripts may serve as seeds of species-specific genes.

Considering these facts, it could be speculated that REs contribute to ITU formation and evolution and, if this is the case, that the contribution mode may be distinct from that to known genes. To test these hypotheses, we extracted REs that provide polyA signals to intergenic transcription units (ITUs). Theses REs have more than one canonical polyA signal (AATAAA or ATTAAA) that add 3' polyA tails to at least two mRNAs and/or ESTs in the UCSC Genome Browser. The resultant 150 intergenic REs were compared with REs providing polyA signals for known genes regarding RE subclass distribution. As expected by previous reports, both polyA RE groups were dominated by L1s. However strikingly, L2, MIR and ERV1 showed biased distribution: L2 and MIR more in known genes and ERV1 more in ITUs. This may reflect their different contributions to gene evolution in the human genome history. In this poster presentation, we will demonstrate up-to-date results about the interesting subclass distribution and also discuss general significance of intergenic REs

## (48)

# Evidence for Co Evolution between Human MicroRNAs and Alu-repeats

*Stefan Lehnert*1, Peter Van Loo2,3,4, Pushpike J. Thilakarathne5, Peter Marynen2,3, Geert Verbeke5 and Frans C. Schuit1*

1 Gene Expression Unit, Department of Molecular Cell Biology, Katholieke Universiteit Leuven
2 Department of Molecular and Developmental Genetics, VIB; 3Department of Human Genetics, Katholieke Universiteit Leuven
4 Bioinformatics group, Department of Electrical Engineering, Katholieke Universiteit Leuven
5 Biostatistical Centre, U.Z. Sint-Rafaël, Katholieke Universiteit Leuven, Leuven, Belgium

Address correspondence and reprint requests to Frans C. Schuit, Herestraat 49, P.O. Box 901, B-3000 Leuven, Belgium. E-mail: frans.schuit@med.kuleuven.be

We propose that Alu repeats, the most abundant repetitive elements in the human genome are evolutionary interacting with microRNAs, small RNAs that alter gene expression at the post-transcriptional level. Base-pair complementarity could be demonstrated between the seed sequence of a subset of human microRNAs and Alu repeats that are integrated parallel (sense) in mRNAs. The most common target site coincides with the evolutionary most conserved part of Alu. A primate-specific gene cluster on chromosome 19 encodes the majority of miRNAs that target the most conserved sense Alu site. The individual miRNA genes within this cluster are flanked by an Alu-LINE signature, which has been duplicated with the clustered miRNA genes. Gene duplication events in this locus are supported by comparing repeat length variations of the LINE elements within the cluster with those in the rest of the chromosome. Thus, a dual relationship exits between an evolutionary young miRNA cluster and their Alu targets that may have evolved in the same time window. One hypothesis for this dual relationship is that these miRNAs could protect against too high rates of duplicative transposition, which would destroy the genome.

## (49)

# Inherent promoter bidirectionality facilitates escape from epigenetic silencing during parasitic DNA integration in mammalian genomes

*Paul Kalitsis1 and Richard Saffery2*

1 Chromosome and Chromatin Research
2 Epigenetics Research, Murdoch Childrens Research Institute, Department of Paediatrics, University of Melbourne, Parkville, Melbourne 3052, Victoria, Australia

It is becoming clear that many genes in mammalian genomes are arranged in a bidirectional manner sharing a common promoter and regulatory elements. This is especially common amongst promoters containing a CpG island. The reason for such an arrangement remains unclear but in many cases it appears to be related to a common gene function. The majority of CpG island-rich promoters exist in an unmethylated state associated with constitutive transcription and a predicted “open” chromatin structure. Here we provide data that single gene CpG island promoters can be “hijacked” by transposable elements thus creating novel genes and bidirectional pairs in the genome. Often this is associated with an increase in CpG island length and transcriptional activity in the antisense direction. From a list of over 60 protein-coding genes derived from transposable elements in the human genome and 40 in the mouse, we have found that a significant proportion are orientated in a bidirectional manner with a neighbouring gene. This suggests that the selective force that shields the endogenous CpG containing promoter from epigenetic gene silencing appears to extend to the incoming foreign DNA with the unexpected stabilisation and transcription of such DNA elements in the genome. Over time host genomes “domesticate” such elements to produce novel functions often essential for proper mammalian development.

# (50)

# Conserved Structure of Vertebrate Retroviral LTRs; Detection of Single LTRs in Genomic Data

*Farid Benachenhou1*, Patric Jern1†, Vidar Blikstad1, Merja Oja3, Göran Sperber2, Panu Somervuo3, Samuel Kaski3, and Jonas Blomberg1*

1 Section of Virology, Dept. of Medical Sciences, Uppsala University, Uppsala, Sweden
2 Unit of Physiology, Dept. of Neuroscience, Uppsala University, Uppsala, Sweden
3 Helsinki Institute for Information Technology, Dept. of Computer Science, University of Helsinki and Laboratory of Computer and Information Science, Helsinki University of Technology, Helsinki, Finland
†Present address: Dept. of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, MA, USA.

Retroviral LTRs, paired or single, are plastic structures which contribute profoundly to vertebrate genomic and transcriptional diversity. However, detection and alignment of single LTRs is a bioinformatic challenge. In this study, LTR sequences of Human, Mouse and Chicken genomes, mainly obtained as consensus sequences from RepBase or via the RetroTector©1 program, were used for training of Hidden Markov models (HMMs). Five more or less specialized HMMs (General Vertebrate, Human MMTV-like;hml, general Betaretroviral, Lentiviral and Gammaretroviral) yielded Viterbi alignments, allowing detection of common consensus structures (match states). Interestingly, the match states of the five HMMs could be arranged into six modules with small internal and longer intermodule insert states. The modules were, in U3: "TG", "Intermediate", "Primary A-rich" (containing TATAA); in R: "Secondary A-rich" (containing AATAAA); and in U5: "T-rich" and "CA". ORFs of around 100 nt and longer were detected in betaretroviruslike LTRs, with predicted amino acid sequences of previously unknown proteins. They either occured between the first and second module (like MMTV sag) or between the primary and secondary A-rich modules, at the approximate U3-R border, maybe allowing function in spite of the disruption of LTR structure by an ORF. Conservation conforming to stem-loop structures were seen in the R and U5 regions. A few transcription factor binding sites were conserved. The HMMs could be used for detection of single LTRs, with up to 87% sensitivity (hml model), when compared with RepeatMasker data from human chromosome 19. When all five HMMs were combined, 71%(specificity 20%), or 53%(specificity 74%) were found, with low and high thresholds, respectively. The HMMs thus allow ab initio detection of LTRs. Better knowledge of LTR structure and function from our models improves the understanding of genomic evolution and basics of retroviral biology like LTR plasticity, control of the sometimes bidirectional promoters and novel! retroviral proteins.

1. Sperber GO, Airola T, Jern P, Blomberg J. Automated recognition of retroviral sequences in genomic data – RetroTector. Nucleic Acids Res. (2007) 35:4964-4976 (2007).

## (51)

# Two distinct CACTA transposon subfamilies create phenotypic and genomic diversity in soybean

*Gracia Zabala*, Lila O. Vodkin*

Department of Crop Sciences, University of Illinois, Urbana, Illinois 61801. USA

Despite the abundance of retrotransposon like sequences in the soybean genome, only CACTA type elements have been shown to be the cause of phenotypic variation in a handful of soybean mutant alleles characterized at the molecular level. Results of two complex insertions will be presented. Tgm-Express1, a 5.7-kb transposon in intron 2 of the F3H wp allele, contains five unrelated host gene fragments. RT-PCR derived cDNAs of this wp allele, represented a multiplicity of processed RNAs varying in length and sequence that included some identical to the correctly processed mature F3H transcript with three properly spliced exons. The five gene fragments carried by the Tgm-Express1 were processed through complex alternative splicing as additional exons of the wp transcript. Thus, the gene fragments carried by the Tgm inverted repeat ends appear to be retained as functional exons within the element. The spliceosomes then select indiscriminately the canonical intron splice sites from a pre-mRNA to assemble diverse chimeric transcripts from the exons contained in wp (1, 2). A second, 20.5 kb insertion, Tgmt*, in the F3'H gene t* allele had the molecular structure of an autonomous transposon of the CACTA family. RT-PCR derived cDNAs uncovered a large precursor mRNA encoding a transposase with a tnp2 and TNP1 domains as well as alternatively spliced smaller transcripts resembling the TNPA-mRNA generated by the En-1 element of maize. Based on divergence of two required determinants for CACTA transposon excision, Tgmt* defined the existence of two subtypes of CACTA transposon families in soybean the other being Tgm1-7. Portions of both Tgm subfamilies transposases were found in moderately high copy number in the soybean genome (3). In all, the CACTA transposons of the soybean appear to be an active force in genome rearrangement with the potential to create novel chimeric genes that influence genomic and proteomic diversity.

1. Zabala, G., Vodkin, L.O. (2005). The wp mutation of Glycine max carries a gene-fragment-rich-transposon of the CACTA superfamily. The Plant Cell 17: 2619-2632.
2. Zabala, G., Vodkin, L.O. (2007). Novel exon combinations generated by alternative splicing of gene fragments mobilized by a CACTA transposon in Glycine max. BMC Plant Biology 7: 38. PMCID: PMC1947982.
3. Zabala, G., Vodkin, L.O. (2008). A putative autonomous 20.5 kb-CACTA transposon insertion in an F3'H allele identifies a new CACTA transposon subfamily in Glycine max. BMC Plant Biology (in revision).

## (52)

# New superfamilies of eukaryotic DNA transposons and their internal divisions

*Weidong Bao, Matthew G. Jurka, Vladimir V. Kapitonov, Jerzy Jurka*

Genetic Information Research Institute, 1925 Landings Drive, Mountain View, CA 94043, USA

Despite their enomous diversity and abundance, all known eukaryotic DNA transposons belong to only 15 superfamilies. Here, we report two new DNA transposon superfamilies, named Sola and Zator. Sola transposons encode DDD-transposases and are flanked by 4-bp target site duplications (TSDs). Elements from Sola superfamily are widely distributed in species from bacteria through protists to plants and metazoans, and they can be divided into three distinct groups named Sola1, Sola2 and Sola3. Individual Sola elements from each group show little sequence similarity to each other, and also have different termini and target site preferences. However, Sola elements from the three groups converged into a single superfamily when cross-compared by PSI-BLAST within GenBank protein sequences. The DDD transposase sequences encoded by Sola transposons are not similar to known transposases. We also report Zator, a second superfamily, with 3-bp TSDs. The Zator superfamily is relatively rare in eukaryotic species, and it has evolved from a group of bacterial transposases, transposase_36 superfamily, also known as rhodopirellula transposase (Pfam07592).

## (53)

# Mutator as a probe for variations in the epigenetic landscape of plant genomes

*Damon Lisch*

E-mail: dlisch@berkeley.edu

Evidence from a wide variety of species suggests that there is dramatic variation in the epigenetic landscape of genomes. The most obvious examples are regions of heterochromatin surrounding the centromeres in many species, but there is certainly more subtle epigentic variation throughout genomes that influence local patterns of gene expression. In plants, there is very little known about such variation. We are using the Mutator system of transposons in maize to uncover a particular form of epigenetic variation that influences the propensity of a locus to remain epigenetically silenced over time. Using a unique, two component transposon system, we can trigger epigenetic silencing of the Mutator regulatory transposon MuDR, segregate away the trigger locus, and observe the heritability of MuDR silencing in subsequent generations. Using this system, we have uncovered dramatic differences in the heritability of the silenced state of MuDR element depending on chromosomal position. We suggest that these differences reflect underlying variation in the propensity of different regions of the maize genome to maintain heritably silenced states of gene expression. Our aim is to use the Mutator system as a probe for such variation by examining heritability of silencing of MuDR elements at a variety of positions. Finally, we discuss the possibility that transposons may have captured sequences important for "forgetting" in order to increase their activity levels in the absence of silencing triggers. Although a great deal is now known about the ability of genomes to recognize and silence transposons, very little is known about the ways in which transposons can avoid or reverse silencing. We suggest that such mechanisms could dramatically influence the impact of transposons on host genome evolution by modification of local epigenetic states.

## (54)

# Ecg-1 elements create genetic diversity in asexual root-knot nematodes

*Stephen M. Gross*, Rushi A. Shah, and Valerie M. Williamson*

Department of Nematology, University of California, Davis

Meloidogyne javanica is a parthenogenic root-knot nematode, an obligate endoparasite of plant roots. Despite their lack of sexual reproduction to generate genetic diversity, M. javanica populations are polyphagous—capable of parasitizing many plant species—and thus have poorly understood mechanisms to create genetic variation [1]. The tomato gene Mi-1 confers resistance to parasitism by M. javanica. Previous work [2] found a spontaneous derivative of the avirulent VW4 isolate, named VW5, capable of surmounting Mi-1 mediated resistance. AFLP analysis demonstrates VW5 is isogenic to VW4, except for the deletion of a DNA sequence named Cg-1. Though it encodes no obvious protein, RNAi silencing of Cg-1 in VW4 individuals confers the ability to parasitize Mi-1 tomato plants. Here, we present additional analysis of the Cg-1 sequence leading us to hypothesize it is a novel non-autonomous Class II transposon we have named Ecg-1. Using primers designed from the terminal inverted repeats of Ecg-1, we isolated a putative autonomous Ecg-1 element through PCR. We are using the recently published genome of a related nematode to identify additional members of the Ecg-1 transposon family. In addition, we are exploring whether an Ecg-1 product directly interacts with Mi-1 protein in tomato to trigger resistance, or whether Ecg-1 regulates a cis-linked avirulence factor that is directly recognized by Mi-1. Ecg-1 transposons provide a model in which to study genome change and adaptation in asexual species.

References:

1. Castagnone-Sereno, P., Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. Heredity, 2006. 96(4): p. 282–9.
2. Gleason, C.A., Q.L. Liu, and V.M. Williamson, Silencing a candidate nematode effector gene corresponding to the tomato resistance gene Mi-1 leads to acquisition of virulence. Mol Plant Microbe Interact, 2008. 21(5): p. 576–85.

## (55)

# Unraveling the DNA repeat content of four Apicomplexan species

*Assiatu B Barrie*1, Cedric Feschotte1, Jessica Kissinger2 and Ellen J. Pritham1*

1 Department of Biology, University of Texas, Arlington, TX 76019, USA
2 Department of Genetics, University of Georgia, Athens, GA 30602

All known Apicomplexans are intracellular eukaryotic pathogens and cause devastating diseases, such as malaria and toxoplasmosis. The complete genomes of fourteen Apicomplexan species have been sequenced, assembled and made publicly available. The genomes range in size from ~8.74 Mb to 63.50 Mb revealing that these genomes are in flux, which is not surprising since pathogens and their hosts are in a constant arms race for survival. Genome size variation in many eukaryotes is largely shaped as the result of transposable element proliferation. Indeed, transposable elements make up the largest and most dynamic component of multi-cellular and many unicellular eukaryotic genomes. To date, there is little evidence supporting the presence of transposable elements in the genomes of any Apicomplexan genome despite the near ubiquity of transposable elements in the eukaryotic tree of life. It remains unclear whether the reported dearth of transposable elements is a result of a lack of rigorous scrutiny or the result of, yet so far unknown mechanisms that act to prevent transposable element proliferation in these genomes. To investigate the presence of transposable elements in the phylum, Apicomplexa, a comprehensive analysis of four Apicomplexan genomes (Toxoplasma gondii, Plasmodium falciparum, Cryptosporidium parvum, and Theileria parva) was undertaken. The repetitive content of these species were determined computationally, using a de novo repeat annotation program called Repeatscout. The manual classifications of these results are underway and will be discussed in light of genome evolution in this medically important phylum.

## (56)

# Locus-specific hyper- and hypo-methylation of mouse B1 elements in male germ cells

*Kenji Ichiyanagi*, Junko Kitayama, and Hiroyuki Sasaki*

Division of Human Genetics, National Institute of Genetics, JAPAN

In mammals, epigenetic information is reprogrammed in the germ line. Thus, DNA methylation marks at CpG sites are largely erased in the early stage of the germ-line development. In male, LINEs and LTR elements soon become re-methylated by the stage of spermatogonia, and keep hyper-methylated through meiosis to sperm formation. On the other hand, it has been reported that B1 SINE elements remain relatively unmethylated status. These studies employed bulk PCR amplification of bisulfite-treated genomic DNA, and the results were interpreted based on that unmethylated CpG sites are converted into TpG whereas methylated CpG sites are not. However, this method underestimates the level of methylation, because substantial fraction of genomic B1 copies carry TpG substitution. Moreover, the inter-loci variance of DNA methylation is unknown. In this study, we investigated the DNA methylation status of >60 loci of B1 elements individually, rather than in bulk, in sperm and liver (somatic reference). In contrast to the previous reports, most (~85%) of B1 copies were hyper-methylated in sperm. We also revealed that several copies are reproducibly unmethylated in sperm. Such hypo-methylation seems germ line-specific because these copies, as well as sperm-methylated copies, were all hyper-methylated in liver. Statistical analysis suggests that B1 copies embraced in promoter regions tends to escape DNA methylation in the male germ cells.

## (57)

## Analysis of the repetitive DNA in the 1RS chromosome of Rye (Secale Cereale)

*S. Fluch, D. Kopecky, R. Boonruangrod, M. Berenyi*, B. Klauninger, J. Dolezal, T. Lelley, K. Burg*

Dept. Biogenetics/natural resources, PICME, Austrian Research Centers GmbH – ARC, 2444 Seibersdorf, Austria

Presenting author Maria.berenyi@arcs.ac.at

Rye (Secale cereale) is grown extensively as a grain and forage crop. It is a member of the wheat tribe (Triticeae) and is closely related to barley and wheat. Numerous known and unknown wheat lines worldwide are carrying the 1BL/1RS wheat-rye translocation making the short arm of the rye chromosome 1 (1RS) an integrated part of the wheat germplasm. This 1RS is especially interesting for breeders as it has been shown to confer characteristic traits to the wheat grain. 1RS is available as separate telocentric chromosome in Chinese Spring/Imperial ditelo1RS wheat line allowing the isolation of this chromosome by flow sorting.
The 1RS chromosomal DNA has been subject to 454 FLX and Titanium sequencing runs revealing 317Mbp of sequence information resulting in 0,8x coverage of the chromosome. Using different software packages we were able dissect the data set and to categorize and classify the repeat classes present in 1RS. Using a publicly available DB for repeat identification we found that over 69% of the sequences fall into various known repeat classes. Further analysis of the remaining 30% of the sequences by clustering and contig analysis revealed the presence of additional repeat classes not described previously in literature.
For the first time we will present a detailed repeat analysis of a single chromosome of rye, thus providing valuable knowledge on the structure of repeat classes in a single chromosome of a member of the Triticeae.

## (58)

# Amlification of an endogenous retrovirus on the chicken W-chromosome

*T. Revay* and A. Hidas*

Research Institute for Animal Breeding and Nutrition, Godollo (Hungary)

The impact of the sex specific DNA sequences made us to study a sex-specific RAPD (random amplified polymorphic DNA) fragment found during the investigations of Hungarian indigenous chicken breeds. A 900 bp long fragment showed very strong amplification in females, suggesting an origin from the repetitive satellite DNA rich W-chromosome. After gel electrophoresis, the fragment was cut, purified and cloned into plasmid vector and transformed into host E. coli. Recombinant plasmid was used as a probe for Southern hybridization on the RAPD pattern to verify the origin of the clone. It recognizes not only the fragment cut, but gave a same-sized, weak signal in males too. This latter fragment was cloned and sequenced as the female ones. The male and female sequences showed almost complete homology. The sequence was used for searching the chicken genome sequence and Repbase, and high homology was found with the GGERVL-A endogenous retrovirus not only on the sex chromosomes, but on several autosomes too. It seems that the strong RAPD amplification was due to the difference in copy number between the sexes, which was verified by real-time PCR experiments. The specific amplification of the retrotransposon on the chicken W-chromosome was investigated by FISH.

## (59)

# Junk lends a hand: Transposable elements contributing to genetic variation among isolates of Trichomonas vaginalis

*CLAUDIA P. MARQUEZ* AND ELLEN J. PRITHAM*

Department of Biology, University of Texas, Arlington, Arlington, TX 76019, USA

Protozoan pathogens contribute to the death of millions each year, thus exacting a tremendous socioeconomic impact. To gain an understanding of the factors that contribute to the emergence of infectious diseases, it is critical to understand the factors that allow these organisms to adapt quickly to new environments and hosts. The draft genome sequence of one such protozoan, Trichomonas vaginalis was recently completed and the assembly was hindered by what appears to be a recent explosion of a large variety of DNA transposons (representing ten different superfamilies). A comprehensive analysis of the transposable element (TE) landscape of T. vaginalis completed in our lab, led to the identification of 105 different families of TEs. Masking the genome sequence with our repeat library revealed that 49% of the genome is made up of related TE sequences. To understand the contribution of TEs in altering the overall architecture of the T. vaginalis genome, we have selected a number of recently active TE families to measure the level of polymorphism between different isolates of T. vaginalis and its closely related sister species, T. tenax. We are using transposon display to understand the level of TE-generated polymorphisms between isolates to determine if TEs are active contributors to genetic variation between strains and to identify active transposable elements in the T. vaginalis genome.

## (60)

# Rice transposable elements: Facts and Stats

*Francois Sabot*

E-mail: francois.sabot@ird.fr

Cultivated rice genome (Oryza sativa L.) is 380Mb long, and is composed of more than 40% of identifiable transposable elements (TEs). Those are from each classical types identified in plant, i.e. LTR retrotransposons, LINEs, SINEs, TIRs (and MITEs), and even Helitrons. LTR retrotransposons are the main component of this repeated genomic fractions. Even being largely represented, only a few elements are known to be unambigously active, transpositionally speaking. Here we will present which are identified such, and how they have been identified. For LTR retrotransposons, are identified: Tos17 (Copia), Lullaby (Copia) and Houba (Copia). For LINEs: Karma.For TIRs: mPing/Pong (Tourist MITE/PIF), dTok (hAT) and Dart (hAT). The methods used for their identification and the assessment of their transposition are various: cDNA identification/Northern blot, Mutation analysis, AFLP, Transposon Display, microarray.Here we will compare those approach, their advantages and their limits.

## (61)

# Exon-trapping mediated by the SVA retrotransposon

*Dustin C. Hancks*, Adam D. Ewing and Haig H. Kazazian Jr*

Department of Genetics, University of Pennsylvania,

The great majority of human retrotransposable elements are inactive, however, both inactive and active retrotransposons drive genome evolution and may influence transcription through various mechanisms. Little is known about the SVA element, a non-coding RNA, which is one of three retrotransposon families still active in the human genome. We report the identification of a new subfamily of SVA, which appears to have been formed by an alternative splicing event, where the first exon of the Mast2 gene spliced into an intronic SVA at a site which resembles a splice acceptor sequence in the SVA and subsequently retrotransposed. Upon performing molecular and computational experiments, we have identified many functional splice acceptor sites within several different transcribed SVAs across the genome. We propose that SVA is mimicking a gene-trap in order to mediate its transcription by any means necessary. Furthermore, this result implies that SVA elements residing within introns of genes in the same orientation may disrupt normal gene transcription and that SVAs may alter the transcriptome, thereby altering genome evolution.

## (62)

# VisualRepbase: an interface for the study of occurrences of transposable element families

*Sebastien Tempel*

Genetic Information Research Institute, Mountain View, CA 94043, USA

Repbase is a reference database of eukaryotic repetitive DNA, which includes prototypic sequences of repeats and basic information described in annotations. Repbase already has software for entering new sequence families and for comparing the user's sequence with the database of consensus sequences. We describe the software named VisualRepbase and the associated database, which allow for displaying and analyzing all occurrences of transposable element families present in an annotated genome. VisualRepbase is a Java-based interface which can download selected occurrences of transposable elements, show the distribution of given families on the chromosome, and present the localization of these occurrences with regard to gene annotations and other families of transposable elements in Repbase. In addition, it has several features for saving the graphical representation of occurrences, saving all sequences in FASTA format, and searching and saving all annotated genes that are surrounded by these occurrences. VisualRepbase is available as a downloadable version. It can be found at http://www.girinst.org/VisualRepbase/index.html.

## (63)

# A microRNA incubator on the marsupial (Monodelphis domestica)X-chromosome was created via L1 transposon-mediated serial duplication

*Eric J. Devor*(1), Paul B. Samollow(2) and Andrew S. Peek(1)*

(1)Molecular Genetics and Bioinfomatics, Integrated DNA Technologies, Coralville, Iowa 52241
(2)Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas 77843, USA

The origin of microRNAs (miRNAs), small (21nt to 23nt) non-coding regulatory RNAs, is a topic of interest in evolutionary biology as well as in functional genomics. There is mounting evidence to support a view that these and, perhaps, other non-coding RNAs arise from transposons though the mechanisms that are, as yet, unclear (Borchert et al., 2006). We have discovered a closely related family of 39 microRNAs spanning just over 100Kb of the X-chromosome of the grey, short-tailed South American opossum Monodelphis domestica (Mikkelsen et al., 2007). Detailed analysis of this region indicates that this family was created via a series of duplication events that were mediated by the presence of L1 transposons flanking the pre-miRNA. Further, there is some evidence that the ancestral miRNA itself evolved out of an L1 sequence.

Here we present the complete anatomy and evolution of this miRNA family including evidence that these miRNAs are in the process diverging and that at least two have already become pseudogenized.

Borchert GM, Lanier W, and Davidson BL 2006 Nat Struct Mol Biol. 13:1097-101.
Mikkelsen TS, Wakefield MJ, Aken B, et al. 2007 Nature 447: 167-177.

## (64)

# Intronic Transposable Elements and Their Host Genes

*Ying Zhang* and Dixie Mager*

Terry Fox Laboratory, BC Cancer Research Centre, Vancouver, BC, Canada

In humans and mice, transposable elements (TEs) comprise nearly half of the genomic DNA and much work has been done to analyze TE distribution patterns on a whole genome scale. However, factors governing the interactions between intronic TEs, which might be more influential to the host, and the genes in which they reside are still largely unknown. This research is focused on TEs located close to/within host genes, and several candidate factors that might influence TE-host interactions have been examined. The first factor analyzed was the gene/intron size. While it seems natural to expect more intronic TE insertions when the gene is large, details about this property have not been carefully examined. We found a positive linear correlation between TE-coverage and gene/intron size and found that some apparent over-representations of TEs in genes of certain functional classes can be attributed solely to gene size. We also examined some exceptional cases. The second factor investigated was the distance from intronic TEs to intron-exon boundaries. Since TEs can carry cryptic splicing signals, the closeness of TEs to genic splice sites may disrupt normal splicing of host genes. Indeed our study showed a significant drop of TE density near intron-exon boundaries, as well as differences between different TE types. For example, we found evidence suggesting that, within 2 kb of an exon, intronic human LTR elements are much less well tolerated compared with Alu elements. The third factor examined was the conservation level of host genes and their corresponding TE densities. We hypothesized that genes of different conservation levels may show different degrees of tolerance of TE insertions, and this was also confirmed along with more details. Other factors such as gene expression breath and levels have also been examined. A comprehensive analysis of combining all factors will be presented.

## (65)

# Genomic Drive: A Powerful Engine of Evolution

*Keith R. Oliver 1* and Wayne K. Greene 2*

1. School of Biological Sciences and Biotechnology, Murdoch University, Western Australia.
2. School of Veterinary and Biomedical Sciences, Murdoch University, Western Australia.

e-mail: w.greene@murdoch.edu.au

There is much evidence that Transposable Elements (TEs) create what we are terming 'Genomic Drive.' This term recognizes that TEs render germ line genomes flexible and dynamic, making them efficient drivers of evolution. Much of evolution is driven by natural selection acting on the progeny of organisms with TE generated genomic changes (1). TEs often create genetic changes consisting of more orderly re-arrangements of the genome, than the genomic corruptions of some other mutagens. The cellular mechanisms to control TEs are important in both Genomic Drive and epigenetics.

Genomic Drive complements all known mechanisms of evolution. It acts in several ways: active (transposition of TEs) and passive (ectopic recombinations of TEs). TE sequences can also be co-opted for other functions, and can alter gene regulation (2).

If in a lineage the proliferation of TEs in the germ line is not adequately controlled, the lineage could become extinct, due to genomic chaos resulting in unviable progeny phenotypes. However, if effective TEs are very scarce then lineages are likely to become static, rare, or extinct, due to inability to adapt to change. Where there is a balance between TE activity and TE control, the level of variation is optimal for evolution and lineages are adaptable, fecund, and taxonate readily. In the short term TEs may lower the fitness of some individuals, but in the long term Genomic Drive raises the fitness of lineages.

Often, successive waves of modified, or novel, endogenous or exogenous TEs infiltrate germ line genomes of some lineages, increasing Genomic Drive and taxonation. Cellular controls and degradation of the TEs impede the function of Genomic Drive over time. The result is periods of rapid taxonation alternating with periods of slow taxonation or stasis, i.e. punctuated equilibrium (3). Prolonged stasis can result in relict taxa or "living fossils."

References

1. Kazazian, H.H., Jr. (2004) Mobile elements: drivers of genome evolution. Science 303: 1626-1632.
2. Feschotte, C. (2008) Transposable elements and the evolution of regulatory networks. Nature Reviews
   Genetics 9: 397-405
3. Gould, S.J. and N. Eldridge (1977) Punctuated equilibria: the tempo and mode of evolution reconsidered.
   Paleobiology 3: 115-151.

## (66)

# Clustering of some SINEs in gene-rich genome regions may be caused by positive selection of specific genome rearrangements facilitating transcription

*Usmanova, N. and Tomilin, N.**

Institute of Cytology, Russian Academy of Sciences, 194064 St.Petersburg, Russia

Mammalian SINEs accumulate in GC-rich gene rich genome regions and retroelements of the LINE family L1 accumulate in AT-rich gene poor regions but mechanism of this clustering remains unknown. Recently integrated Alu elements and de novo L1 insertions in HeLa cells are distributed randomly, regardless of the GC content of the surrounding DNA, indicating that post-insertion rearrangements play a role in genome shaping during evolution. Whether these rearrangements are neutral or they are driven by natural selection? Negative selection against the presence of L1 in gene rich regions can be driven by extensive methylation of L1 DNA and repression of adjacent genes, but accumulation of SINEs in gene-rich regions is suggested to be selectively neutral. It may be due to an excess of SINE-promoted intrachromosomal segmental duplications (ISDs) over deletions in gene-rich regions which should lead to even co-amplification of all retroelements located in GC-rich DNA. To examine this hypothesis we studied distribution of SINEs along some mouse and human chromosomes which differ strongly in ISDs. In the human chromosomes 20 and X in which non-redundant ISD duplications differs ~6 times we found similar clustering of Alu repeats in gene-rich regions. The mouse chromosomes 18 and X in which non-redundant ISD duplications differ >40 times show similar clustering of B1 and B2 repeats. Alu repeats also show weak co-clustering with the MIR family in human chromosomes although this family is associated with GC rich DNA These results argue against the role of ISDs in the clustering of some SINEs in gene-rich regions. We believe that positive selection of specific genome rearrangements leading to clustering of some SINEs near genes is involved, which may be caused by SINE-associated transcription regulatory elements. Many SINEs contain binding sites for transcription complex TFIIIC which can function as an insulator against spread of repressive histone modifications.

## (67)

# ELAN: A server based tool for genome wide analysis of mobile genetic elements

*Kamal Rawal*1,5, Sudha Bhattacharya 2, Alok Bhattacharya1,3, Ram Ramaswamy 1,4*

1 School of Information Technology
2 School of Environmental Sciences
3 School of Life Sciences.
4 School of Physical Sciences,
Jawaharlal Nehru University, New Delhi 110 067, India, 5. JIIT University, India.

Mobile genetic elements occupy significant proportion of eukaryotic genomes. They are involved in number of important cellular functions. ELAN is a suite of tools for genome wide analysis of mobile genetic elements. It finds distribution and nature of mobile genetic elements. DNA SCANNER is a part of ELAN which analyses insertion sites of mobile genetic elements for the presence of various physicochemical signals. Insertion Site Finder (ISF) is a machine learning tool which incorporates information derived from DNASCANNER and uses support vector machines to classify DNA sequences into insertion sites and non insertion site classes. ELAN has been applied to wide variety of organisms. It has identified distributions of several mobile elements such as Alu in various organisms such as Human, Mouse, Drosophila, E. histolytica etc. DNA SCANNER has identified common set of statistically important signals flanking insertion sites in various genomes suggesting common insertion mechanism operating in wide variety of organisms. ISF has emerged to be an important tool for insertion site prediction as it has shown high accuracy levels (65-90%). The dataset and information derived during analysis will serve as bench marking resource in future for various analyses. Large data has been organized into web portal as well as relational database named as InSiDe which is available online at http://nldsps.jnu.ac.in/bioit/ccbb/elan.html. Experiments were conducted in E. histolytica to validate computational findings.

References:

1. Identification of Insertion hot spots for non-LTR retrotransposons: Computational and Biochemical application to Entamoeba histolytica. Nucleic Acid Research, 34, 2006, 5752-5763.
2. he LINEs and SINEs of Entamoeba histolytica: comparative analysis and genomic distribution. Journal of Experimental Parasitology, 110, 2005, 207-213.

## (68)

# The evolution of two partner LINE/SINE families in the first reptilian genome of Anolis carolinensis

*Oliver Piskurek*, Hidenori Nishihara, Norihiro Okada*

Department of Biological Sciences, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259-B21 Nagatsuta-cho, Midori-ku, Yokohama 226-8501, Japan

Transposable elements (TEs) have been characterized in a number of vertebrates, including whole genomes of mammals, birds, and fishes. The Anolis draft assembly provides the first opportunity to study retroposons in a reptilian genome. Here, we identified and reconstructed a number of retroposons based on database searches: Sauria SINEs (Piskurek et al. 2006), 5S-Sauria SINE chimeras, Anolis SINE 2, Anolis LINE 2, Anolis LINE 1, Anolis CR 1, and a chromodomain-containing Ty3/Gypsy LTR element. We focused on two SINE families and their partner LINE families (Anolis Sauria SINE/Bov-B LINE and Anolis SINE/LINE 2). We demonstrate that each SINE/LINE pair is distributed similarly and that the evolutionarily youngest Sauria SINE sequences evolved as part of novel rolling-circle transposons. The evolutionary time frame when Sauria SINEs/Bov-B LINEs were less active in their retrotransposition is characterized by a retrotransposition burst of Anolis SINE/LINE 2 elements. We also characterized the first full-length chromoviral LTR element in amniotes. This newly identified chromovirus has been very well preserved in the Anolis genome. TEs in the Anolis genome account for approximately 20% of the total DNA sequence, whereas the proportion is more than double that in many mammalian genomes in which such elements have important biological functions. Nevertheless, 20% TE coverage is sufficient to predict that Anolis retroposons and other mobile elements also may have biologically and evolutionarily relevant functions. The SINEs and LINEs and other ubiquitous genomic elements characterized in the Anolis genome will prove very useful for studies in comparative genomics, phylogenetics, and functional genetics.

Piskurek O, Austin CC, Okada N, 2006. Sauria SINEs: novel short interspersed transposable elements that are widespread in reptile genomes. J. Mol. Evol. 62, 630-644.