1ST INTERNATIONAL CONFERENCE/WORKSHOP

Genomic Impact Of Eukaryotic Transposable Elements

MARCH 31 – APRIL 4, 2006

ASILOMAR, PACIFIC GROVE, CALIFORNIA, USA

1ST INTERNATIONAL CONFERENCE/WORKSHOP

Genomic Impact Of Eukaryotic Transposable Elements

Organizers:

Mark A. Batzer
Prescott Deininger
Jerzy Jurka (Chair)
John Moran

ASILOMAR 2006

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.

1st International Conference and Workshop

Genomic Impact of Eukaryotic Transposable Elements

Organizers:
Mark A. Batzer, Prescott Deininger, Jerzy Jurka (Chair), John Moran

Friday, March 31, 2006

15:00-18:00	REGISTRATION (Phoebe A. Hearst Social Hall)	
18:00-19:00	*Dinner* (Crocker Dining Hall)	
19:00-20:30	Preparation of audio visual (Fred Farr Forum) The following equipment will be provided in all sessions: an LCD projector, overhead projector, 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.	
20:00-23:00	*Reception* (Fred Farr Forum)	

Saturday, April 1, 2006

7:30-8:30	*Breakfast* (Crocker Dining Hall)	
8:00-9:00	REGISTRATION (Phoebe A. Hearst Social Hall)	
9:00-9:30	**Jerzy Jurka** - Introduction to the conference and workshop	
9:30-10:00	**John Moran** - Template-specific reverse transcriptase activity in LINE-1 RNPs	p.1
10:00-10:30	**Prescott Deininger** - Does L1 survive despite, or because of, its incompetence?	p.2
10:30-11:00	*Coffee-break* (Fred Farr Forum)	
11:00-11:30	**Anthony Furano** - The interaction between L1 retrotransposons and their mammalian hosts	p.3
11:30-12:00	**Holly Wichman** - LINE-1 activity and extinction in mammals	p.4
12:00-13:00	*Lunch* (Crocker Dining Hall)	
13:30-14:00	**Haig Kazazian Jr.** - Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity	p.5
14:00-14:30	**Mark Batzer** - Mobile elements and primate genomic variation	p.6
14:30-15:00	**Juergen Brosius** - Mistaken identity - how repetitive elements only indirectly related to retroposons move around the genome	p.7
15:00-15:30	**Norihiro Okada** - Functional non-coding sequences derived from SINEs in the human genomes	p.8
15:30-16:00	*Coffee-break* (Fred Farr Forum)	
16:00-16:30	**Carol Rubin** - Alu repeats: From junk to function	p.9
16:30-17:00	**Nicolai Tomilin** - The role of retrotransposons in the maintenance of heterochromatin	p.10

17:00-17:30	**Carl Schmid** - Functional implications of SINE expression	p.11
18:00-19:00	*Dinner* (Woodlands)	
19:00-19:30	**Andrew Gentles** - Computational reconstruction of transposable elements	p.12
19:30-20:00	**Gabor Toth** - Methods for *de novo* identification of repetitive sequences in newly sequenced genomes	p.13
20:00-20:30	**Arian Smit** - RepeatMasker, FEAST and other tools for analyzing and exploiting repetitive DNA	p.14
20:30-23:00	*Happy Hours* (Woodlands)	

Sunday, April 2, 2006

7:30-8:30	*Breakfast* (Crocker Dining Hall)	
9:00-9:30	**Irina Arkhipova** - Transposons, telomeres and rotifers	p.15
9:30-10:00	**Gill Bejerano** - Origins of ultraconservation and distal cis-regulation in vertebrates	p.16
10:00-10:30	**Sandy Martin** - Single Amino Acid Substitutions in L1 ORF1p with Dramatic Effects on Nucleic Acid Chaperone Activity and L1 Retrotransposition	p.17
10:30-11:00	*Coffee-break* (Fred Farr Forum)	
11:00-11:30	**Daniel Voytas** - Retrotransposon target specificity and genome organization	p.18
11:30-12:00	**Jiri Hejnar** - Human syncytins - an extreme example of transposable element domestication	p.19
12:00-13:00	*Lunch* (Crocker Dining Hall)	
13:30-14:00	**Dusan Kordis** - Enormous impact of retroelements on the genome structure and evolution in land vertebrates	p.20
14:00-14:30	**Dixie Mager** - Effects of LTR elements on mammalian genes	p.21
14:30-15:00	**Alan Schulman** - Parasites and parasites of parasites: Plant retrotransposons and their genomic impact	p.22
15:00-15:30	**John McDonald** - The contribution of LTR retrotransposons to gene evolution: a tale of three genomes.	p.23
15:30-16:00	*Coffee-break* (Fred Farr Forum)	
16:00-16:30	**Horacio Naveira** - Contrasting patterns of sequence turnover of LTR retrotransposons in different eukaryotes	p.24
16:30-17:00	**Marie-Anne Van Sluys** - A genomic approach to depict transcriptionally active transposable elements in sugarcane	p.25
17:00-17:30	**Adam Pavlicek** - Retroposition of processed pseudogenes: the impact of RNA stability and translational control	p.26
18:00-19:00	*Dinner* (Woodlands)	
19:00-19:30	**Wojtek Makalowski** - Validation of diverged repetitive elements using phylogenetic analysis and comparative genomics approach	p.27
19:30-20:00	**Peter Warburton** - Analysis of the relative chronological age of human transposable elements by defragmentation and insertional analysis	p.28
20:00-20:15	**Vini Pereira** - Automated palaeontology of repetitive DNA with REannotate	p.29
20:15-20:30	**Degui Zhi** - Comparative genomics analysis of Alu gene conversions	p.30
20:30-22:30	*Happy Hours* (Fred Farr Forum) **Poster session** - odd numbers (Kiln) Workshop appointments (Oak Shelter)	

Monday, April 3, 2006

7:30-8:30	*Breakfast* (Crocker Dining Hall)	
9:00-9:30	**Pierre Capy** - Dynamics of transposable elements: first steps of invasion and long-term evolution	p.31
9:30-10:00	**Andrew Flavell** - Conflict, compromise or cooperation - Different ways for transposons and genomes to coexist	p.32
10:00-10:30	**Dmitri Petrov** - Population dynamics of a comprehensive set of transposable elements in the D. melanogaster genome	p.33
10:30-11:00	*Coffee-break* (Fred Farr Forum)	
11:00-11:30	**Giorgio Bernardi** - The organization of the human genome: from chromosomal bands to isochores	p.34
11:30-12:00	**Peter Arndt** - Substitution pattern of mammalian transposable elements - element specific, regional, and evolutionary aspects.	p.35
12:00-13:00	*Lunch* (Crocker Dining Hall)	
13:30-14:00	**Cedric Feschotte** Life after death: reincarnation of DNA transposons into genetic networks: A case study in the human genome	p.36
14:00-14:30	**Ning Jiang** - The impact of Mutator-like elements on genome evolution	p.37
14:30-15:00	**Antoni Rafalski** - Helitrons and the evolution of DNA sequence diversity in maize	p.38
15:00-15:30	**Marie-Jose Daboussi** - Transposable elements in filamentous fungi	p.39
15:30-16:00	*Coffee-break* (Fred Farr Forum)	
16:00-16:30	**Andrew Shedlock** - BAC-enabled comparative investigation of CR1 evolutionary dynamics in the reptilian genomic landscape	p.40
16:30-17:00	**Vladimir Kapitonov** - Towards a unified nomenclature and classification of eukaryotic transposable elements	p.41
17:00-17:15	**Emile Zuckerkandl** - Do transposable elements participate in combinatorial epigenetics?	p.42
18:00-19:00	*Dinner* (Woodlands)	
19:00-19:15	**Cristina Vieira** - New regulatory regions of Drosophila 412 retrotransposable element generated by recombination	p.43
19:15-19:30	**Chris Smith** - The Drosophila heterochromatin genome project (DHGP): identifying repeats & using comparative sequence analysis to follow heterochromatin evolution	p.44
19:30-19:45	**Sebastien Tempel** - The combinatorics of helitron termini in *A. thaliana* genome revealed strongly structured superfamilies	p.45
19:45-20:00	**Victor Zhurkin** - The tumor suppressor protein p53 binding sites in the human genome: How are they related to transposons?	p.46
20:00-20:30	**Clark Jeffries** - Hairpin database: why and how?	p.47
20:30-22:30	*Happy Hours* (Fred Farr Forum) **Poster session** - even numbers (Kiln) Workshop appointments (Oak Shelter)	

Tuesday, April 4, 2004

7:30-8:30	*Breakfast* (Crocker Dining Hall)	
8:30 - 9:30	**Mini-workshop on Classification and Nomenclature of TEs** (Fred Farr Forum) Moderators: **Pierre Capy, Cedric Feschotte, Jerzy Jurka, Vladimir Kapitonov**	

	Speakers: **Irina Arkhipova** - Classes, subclasses, superfamilies and clades **Jean-Marc Deragon** - Classification and nomenclature of SINEs in plants **Heidrun Gundlach** - Repeat ontology for automated annotation **Patrik Medstrand** - LTR-based ERV systematics **Ellen J. Pritham** - Maverick classification **Adam Pavlicek** - Standardization of subfamily definition **Daniel Voytas** - Copia/Gypsy/BEL overview
	Each speaker will address two specific issues: (1) what are the classification/nomenclature problems in the particular are of the speaker's expertise, and (2) a proposal how to address them.
9:30-11:00	Short summaries by the moderators, additional talking points and discussion how to constitute an international organization devoted to coordination of the classification and nomenclature of TEs.
11:00-12:00	Nominations and secret ballot (if agreed upon).
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, 2006. Details will be posted on conference website http://www.girinst.org/conference/index.htm after the conference.

POSTER PRESENTATIONS

Author Index

SPEAKER ABSTRACTS

Template Specific Reverse Transcriptase Activity in LINE-1 RNPs

*Deanna A. Kulpa and John V. Moran**

Department of Human Genetics, University of Michigan

LINE-1s (L1s) are non-LTR retrotransposons that mobilize (i.e., retrotranspose) using an RNA intermediate. There are ~517,000 L1s in the average human genome, ~60-100 of which are active. Full length active L1s contain a 5' UTR, two open reading frames (ORF1 and ORF2), and a 3' UTR. ORF1 encodes an RNA binding protein (ORF1p), whereas ORF2 encodes a protein (ORF2p) with endonuclease (EN) and reverse transcriptase (RT) activities. It is hypothesized that only one molecule of ORF2p is synthesized per L1 transcript, whereas ORF1p is produced at greater quantities to coat the L1 RNA. The L1-encoded proteins demonstrate a cis-preference and preferentially associate with their encoding transcript to form a ribonucleoprotein particle (RNP). The L1 RNP forms in the cytoplasm and only after it gains access to the nucleus is the L1 cDNA synthesized by target primed reverse transcription (TPRT).

We previously developed a system to follow the fate of wild type and mutant L1s against a background of endogenously expressed elements in cultured human cells. We demonstrated that wild type ORF1p and L1 RNA co-localize in a cytoplasmic RNP, and discovered two classes of mutants in the ORF1p nucleic acid binding domain that reveal RNP formation to be necessary but not sufficient for L1 retrotransposition. We next wanted to examine ORF2p in the RNP; however, this proved to be difficult presumably because of its low level of expression. To circumvent this problem, we developed a biochemical assay to specifically detect the L1 ORF2p from RNPs via its RT activity, which we termed LEAP, for Line Element Amplification Protocol. Using LEAP, we showed that consistent with cis-preference, the L1 RT prefers its own template and does not promiscuously utilize other cellular RNAs. The L1 RT does not require terminal complementarity between the primer and template to initiate reverse transcription, which is supported by analysis of genomic integration sites in vivo. We also showed that ORF1p is dispensable for the L1 RT activity in vitro. In sum, we have developed an assay to detect the L1 ORF2p in cytoplasmic RNPs, strongly supporting their role as intermediates in L1 retrotransposition. We anticipate the LEAP assay will have future applications in the characterization of other poorly understood steps in TPRT.

Do human L1 elements survive despite, or because of, their incompetence?

Prescott Deininger, Victoria Perepelitsa-Belancio, Megan Whelton, and Stephen Gasior

Tulane Cancer Center, New Orleans, LA 70112

Human L1 elements are the only known autonomous retroelements in humans. They have managed to amplify to 500,000 copies in the human genome. The vast majority of the L1 copies are inactive due to truncations, rearrangements and point mutations, leaving only about 100 having potential activity. L1 elements insert into a new genomic site using an RNA intermediate and a process termed Target-Primed Reverse Transcription (TPRT). During TPRT both strands of the genome must be nicked. We find that expression of L1 creates a very large excess of these DSBs relative to the number of actual L1 inserts formed and causing cellular toxicity and DNA repair responses. Despite multiple types of damage from L1 elements, they persist in their amplification, partly due to the use of multiple mechanisms to limit their expression. Expression of L1 elements commonly increases greatly in tumors. It is well known that methylation of the CpG island in their internal promoter greatly decreases expression. However, significant levels of expression do continue to occur in both the germ line and numerous somatic tissues. We have shown that the vast majority of these transcripts are subject to premature polyadenylation reactions that cause most of the transcripts to be truncated. We have recently found that the majority of transcripts are also subject to RNA splicing events that also serve to limit the number of full-length RNA molecules that can express L1 proteins necessary for retrotransposition. Both polyA sites and splice sites introduced by L1 elements within genes due to new insertions would then be expected to contribute to altered gene expression from the genes in which they insert.

The interaction between L1 retrotransposons and their mammalian hosts

Anthony Furano

NIDDK, National Institutes of Health, Bethesda, MD 20892

Nothing in biology makes sense except in the light of evolution (Theodosius Dobzhansky). Therefore, we used evolutionary analysis to make sense of the interaction between mammalian LINE-1 (L1) retrotransposons and their hosts. If L1 elements are deleterious, selection will favor host mechanisms that control their copy number and ameliorate their deleterious effects. Concomitantly, selection will favor L1 families that escape host inhibition and eliminate L1 elements that are seriously deleterious. The equilibrium between retrotransposon activity and host fitness in non-mammalian organisms resulted in all retrotransposons contributing ~5% of the genome. However, in mammals L1 element activity alone generated 30—40% of the genome. Two other changes accompanied the success of L1 in mammals: Most non-L1 retrotransposons went extinct, and L1 evolution has been largely confined to a single lineage. We suggest that the tolerance of mammals for interspersed repeats provided an environment conducive to the emergence of L1 families active enough to compete with each other, and with other retrotransposons, for common host resources. This competition could explain why L1 became the dominant retrotransposon, and why only one L1 family is active at a time. The latter scenario would account for a single L1 lineage. The occurrence of adaptive evolution in some ancestral primate L1 families that were subject to negative selection implies interaction between these families and its host. We are now analyzing the nature of this interaction using the regions of L1 that have undergone adaptive evolution as “bait” to identify host factors that interact with L1 elements. We also found that the Ta1 family, which is active now in humans, has also exerted a fitness cost. Thus, the evolution of current humans may well be shaped by their response to this L1 family.

LINE-1 activity and extinction in mammals

Holly Wichman

University of Idaho, Moscow, ID 83844

LINE-1 (L1) retrotransposons have played a major role in shaping mammalian genomes. Almost 20% of the mammalian genome consists of L1 sequences, most of which are molecular fossils of ancient retrotranspositions. However, determination of ongoing L1 activity is limited to a few species and is confounded by the large number of L1 pseudogenes in the genome. Because species arising after an L1 extinction event also lack active L1s, deep extinctions could have left many mammals devoid of active elements. We have conducted the first comprehensive screen for L1 activity across Mammalia. We show that recently active L1s are present in monotremes, marsupials and all orders of placental mammals. Previous evidence for an L1 extinction event in sigmodontine rodents demonstrates that species can lose L1 activity. We uncovered a second L1 extinction event in megabats. We use L1 extinct rodents to examine the proposal of Lyon that L1s facilitate X chromosome inactivation by serving as way stations to move the inactivation signal along the X chromosome.

Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity

Maria del Carmen Seleme[1], *Melissa R.Vetter*[1], *Richard Cordaux*[2], *Laurel Bastone*[1], *Mark A. Batzer*[2], *and Haig H. Kazazian, Jr.* *[1]

[1] Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104,USA
[2] Department of Biological Sciences, Biological Computation and Visualization Center, Center for Bio-Modular Microsystems, Louisiana State University, Baton Rouge, Louisiana, USA

Despite being scarce in the human genome, active L1 retrotransposons continue to play a significant role in its evolution. Because of their recent expansion, many L1s are not fixed in humans and, when present, their mobilization potential can vary among individuals. Previously we showed that the great majority of retrotransposition events in humans are caused by highly active, or hot, L1s. In four populations of diverse geographic origins (160 haploid genomes), we investigated the degree of sequence polymorphism of three highly active, hot L1s and the extent of individual variation in mobilization capability of their allelic variants. For each locus, we found 1 new allele in every 3-5 genomes, including some with nonsense and insertion/deletion mutations. Single or multiple nucleotide substitutions drastically affected retrotransposition efficiency of some alleles. One-third of elements were no longer hot, and these so-called cool alleles increase substantially the range of individual susceptibility to retrotransposition events. Adding the activity of the three elements in each individual resulted in a surprising degree of variation in mobilization capability, ranging from 0% to 390% of a reference L1. Thus, our data suggest that individual variation in retrotransposition potential makes an important contribution to human genetic diversity.

Mobile elements: a novel source of primate genomic variation

*Mark A. Batzer**

Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, 202 Life Sciences Building, Baton Rouge, Louisiana 70803

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. These loci have proven useful for elucidating human population relationships. Some of the other 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 level of gene conversion between Alu elements suggests that it may have an influence on the single nucleotide polymorphism within Alu elements in the genome. We have also identified several genomic deletions associated with the retrotransposition and insertion of recently integrated mobile elements in primate genomes. This type of retrotransposition mediated genomic deletion is a novel source of genetic variation within primates. The distribution of Alu, L1 and SVA elements throughout various primate genomes makes them useful tools for resolving non-human primate systematic relationships.

Mistaken identity – how repetitive elements only indirectly related to retroposons move around the genome

Carsten Raabe[1]*, Valer Gotea*[2]*, Gennady Churakov*[1]*, Wojciech Makalowski*[2]*, Juergen Brosius**[1]

[1] Inst. Exp. Pathol., University of Muenster, D-48149 Muenster, Germany
[2] Inst. Mol. Evol. Genet., Penn State University, University Park, PA 16802, U.S.A.
brosius@uni-muenster.de

We detected a repetitive sequence that is related to a SINE, yet its mode of dispersal appears not to be consistent with a retroposon: Thus far, we could not identify an RNA intermediate. Like lesSINEs, the elements lack A-tails but, in contrast, are also devoid of direct repeats. The elements are restricted to the rodent clade Myodonta. The majority of repeats are encountered on two chromosomes only. Most empty loci contain a SINE that exhibits sequence similarity to the novel repeat. Insertion appears to involve recombination of an extrachromosomal sequence. Once inserted into a locus, they continue to amplify tandemly. It seems that nucleotide sequence conversion leads to homogenization of the elements within a locus.

Highly conserved SINEs in vertebrates

Norihiro Okada

Tokyo Institute of Technology,Tokyo, Japan

Here, we newly characterized an ancient human SINE family that has intrinsic function. We succeeded in constructing a consensus sequence from about 700 copies of this SINE existing in the human genome and found that the SINE (designated AmnSINE1) is a chimera of 5S rRNA and a tRNA-derived SINE. Also in the chicken genome, the same SINE family, whose consensus is identical to that of human, was characterized. Moreover, we found that AmnSINE1 is quite similar to a known zebrafish SINE family (SINE3). We characterized six novel SINE families from the genomes of the coelacanth, rainbow trout, dogfish shark, hagfish, and amphioxus. All these SINE families have a common central domain that is also shared by zebrafish SINE3 and sea urchin SINE2-3_SP, and we collectively name them DeuSINEs (Deuterostomia SINEs). Surprisingly, sequences of the central domain of many ancient DeuSINEs represent non-protein-coding repetitive sequences highly conserved among mammalian genomes, indicating a significant function. The intriguing possibility exist that DeuSINEs were generated before the origin of vertebrates, and that their core domain has an intrinsic function maintained in vertebrate genomes during these 600 million years. The AmnSINE copies are the first examples of ultra conserved non-protein-coding sequences derived from transposable elements.

Alu repeats: From junk to function

*Carol Rubin**

University of California at Davis

"…[A] family of DNA sequences which includes 300,000 highly conserved members interspersed throughout much of the mammalian genome, must have an important function."

With that ultimate sentence in the 1980 Nature publication of the nucleotide sequence of the Alu family, Carl Schmid flung his laboratory into the middle of a genetic philosophical controversy: Should we regard the overwhelming amount of noncoding DNA in the genomes of higher organisms from an existentialist (it's merely "junk" (Ohno) or "selfish" (Orgel and Crick)) or from a rationalist (its ubiquity, persistence, and abundance indicate a benefit to the organism) viewpoint?

In 1998 the Schmid lab reported that human and rodent SINE transcript overexpression transiently stimulates the expression of a cotransfected reporter gene. The same effect had previously been reported for the small adenoviral RNA VAI. VAI deletion mutants do not propagate well, but their growth can be restored if the VAI RNA is supplied.

Most recently we have shown that SINE transcripts, like their VAI counterpart, act to facilitate translational initiation of newly transcribed messages only (1). This activity is of obvious importance to a virus, which must hijack the cellular protein synthetic machinery. By analogy, SINE overexpression would aid a cell in swiftly readjusting its suite of proteins to respond to a stressful condition and would be necessary for normal cell function.

New-found functions for repeated DNA have not been restricted to SINEs. Shapiro and von Sternberg (2) have recently devoted 24 pages to arguing "Why repetitive DNA is essential to genome function". Twenty-five years after Orgel and Crick's "selfish DNA" contention, repeats are now respectable members of the genome.

1. Rubin et al 2002 Nucl Acids Res, 30, 3253-61.
2. Shapiro and von Sernberg 2005 Biol Rev, 80, 227-50.

Role of retrotransposons in the maintenance of heterochromatin

*Nikolai V. Tomilin**

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

Maintenance of constitutive heterochromatin (CH) containing long tandem repeats is known to be promoted by RNA-induced transcription silencing (RITS) complexes, but mechanism of epigenetic formation and maintenance of facultative heterochromatin (FH) remains unclear. Dispersed retrotransposons LINE-1 (L1s) were implicated in chromosome X inactivation in mammals and it was suggested that L1s serves as booster sequences for Xist RNA-containing nucleoprotein facilitating the spread of FH. However, this model cannot explain formation of interstitial FH in other mammalian chromosomes. Another complex (E2F6.com-1) silencing expression of the Myc- and E2F-dependent genes was identified and found to contain protein subunits homologous to the Drosophila silencing complex PcG. DNA-binding subunit of the PcG complex (encoded by the gene PHO) has strong homology to the mammalian ubiquitous transcription repressor/activator YY1 which potential binding sites (consensus CCATNTT) are widely distributed in mammalian genomes. In human genome significant fraction of YY1 sites is associated with Alu retroposons and purified human YY1 binds consensus Alu in a sequence-specific manner (ref. 1). Clusters of the YY1 binding sites recruiting mammalian PcG complex to protein-coding genes can serve as nucleation points for spreading of FH over interstitial chromosome bands like protosilencers induce FH in the yeast subtelomeric regions (ref. 2). Mammalian YY1 is essential for embryonal development (ref. 3) which is consistent with its important role in regulation of FH and tissue-specific gene expression.

1., Oei, S.L., Babich, V.S., Kazakov, V.I., Usmanova, N.M., Kropotov, A.V., Tomilin, N.V. (2004) *Genomics* 83, 873-882.
2., Fourel, G., Lebrun, E., Gilson, E. (2002) *Bioessays* 24, 828-835.
3., Donohoe, M,E,, Zhang, X., McGinnis, L., Biggers, J., Li, E., Shi, Y. (1999) *Mol. Cell Biol*. 19, 7237-7244.

SINE EXPRESSION

*Carl Schmid**

University of Caliornia, Davis

Designing a convincing test of the slippery proposition that SINEs are selfish DNA remains difficult: Demonstrating that SINEs do not have a defined function presents the logical difficulty of proving a negative result. Although finding a function for SINEs would disprove the proposition, proponents of self-fish DNA warn that a direct attempt to search for function is a “sterile exercise”. However, biochemists naturally presume that any protein, which is expressed, must have a function. Recognizing that many RNAs have noncoding functions, this presumption may arguably be extended to other transcription units suggesting an alternative approach toward addressing this question.

Learning the regulated transcription of SINEs will reveal why some are more successful retrotranspositional source genes than others and indirectly test the proposition that SINEs are selfish DNA. Primarily using human Alus as the model, we find that both their internal sequences and unique external flanking sequences serve as cis acting elements for their transcription in response to specific trans acting factors available in particular cell lines. DNA methylation and chromatin condensation globally regulate the entire family although local differences in methylation and condensation may result in differential expression of particular loci. SINEs in mouse and silkworm are expressed in a tissue specific manner and SINE transcription in cultured cells and animals is increased by various stresses such as heat shock and viral infection. The rapid, stress induced increase in SINE transcription correlates with a global decondensation of their chromatin organization and upon completion of the cell stress response, there is a subsequent decrease in their transcription and recondensation of chromatin. Although the function of SINE RNA is unknown, no evidence shows that SINEs are self-fish genes. Rather their tissue specific expression and programmed stress response provide indirect evidence that SINE RNA serves a function.

Computational reconstruction of transposable elements

Andrew J. Gentles

Genetic Information Research Institute, Mountain View, CA

This will be an overview of the steps performed in computationally reconstructing consensus sequences for Repbase. The starting point for such a procedure may be a putative tranposable element copy, or a complete genome sequence for annotation. Many copies of transposons are partial or mutated, and information from them must be combined to reconstruct the ancestral repeat. I will outline some of the approaches we use, some of the problems which can occur; and discuss publicly available tools such as MAFFT (multiple alignment), EMBOSS (consensus building), Censor/RepeatMasker (frontends to BLAST).

Methods for de novo identification of repetitive sequences in newly sequenced genomes

Gábor Tóth

Institute of Genetics, Agricultural Biotechnology Center, Gödöllő, Hungary

The increasing number of genome sequencing projects poses a challenge to repeat identification and repeat masking. While methods based on hand-curated repeat databases provide undoubtedly the most sensitive and precise way to identify dispersed repeats, manual curation of repeat families and their possibly diverged subfamilies cannot keep pace with the emergence of new genomic sequence data. Since repeat families are largely species-specific, the coverage of repeat library-based methods is necessarily insufficient until all repeat families of a genome are identified and classified.
Meanwhile, the presence of mobile elements and other repeats may negatively influence the progress of genome assembly and annotation: correct sequence assembly is hampered by the presence of large interspersed repeats and coding regions in transposable elements can mislead gene prediction methods. On the other hand, incompletely assembled genomic sequence data are increasingly used not only to locate genes targeted by functional genomics but also for large-scale comparative genomic studies. To avoid repeat-induced false local alignments, most sequence comparison and analysis methods require repetitive sequences to be masked beforehand.
To address this need, preliminary repeat libraries have to be built for each genome simultaneously as sequencing progresses, and researchers must also be provided with software tools to identify potential repetitive elements in the sequence regions of their interest.
Several methods have been developed recently for the de novo and automated identification of repetitive sequences in large genomes. They have been implemented in programs like RECON [1], PILER [2], RepeatScout [3] etc. These tools will be surveyed and their utility and limitations will be discussed.

[1] Bao, Z and Eddy, SR (2002) Genome Res 12:1269-1276.
[2] Edgar, RC and Myers, EW (2005) Bioinformatics 21:i152-i158.
[3] Price, AL, Jones, NC and Pevzner, PA (2005) Bioinformatics 21:i351-i358.

RepeatMasker, FEAST and other tools for analyzing and exploiting repetitive DNA

*Arian Smit *, Robert Hubley, Gustavo Glusman, & Lee Hood*

Institute for Systems Biology, Seattle, WA 98103

Our group has created several software programs to analyze interspersed repeats.
RepeatMasker is widely used to mask and annotate repetitive DNA. The core of the program is a stepwise comparison of the query against a database of consensus sequences. We will discuss the optimizations in alignment parameters and the post-alignment processing that contribute to the program\'s high sensitivity and selectivity. The program constructs appropriate libraries for any species, based on the phylogenetic distribution of repeats, which aspect has also been used to help establish species phylogenies and greatly improve interspecies genomic alignments. Amongst the services on the RepeatMasker web site are instant access to complete RepeatMasker annotations and alignments for all well-repeat-characterized assembled genomes and repeat masking for uncharacterized species based on comparison of the query against our transposon-protein database.
FEAST is a program that predicts the extend and orientation of transcribed regions in the genome, based on detecting the genomic signatures of transcription, accumulated over evolutionary time. Two of the four methods employed are based on the uneven distribution of interspersed repeats with respect to genes. FEAST is particularly apt at detecting genes with long introns and lacking sequence conservation and therefore nicely complements existing gene prediction methods,
Time permitting, and dependent on our progress, we will discuss our current efforts focused on the creation of a software suite that should allow non-experts to create RepeatMasker-ready repeat libraries.

Transposons, telomeres, and rotifers

Irina R. Arkhipova

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

Penelope-like elements (PLEs) are eukaryotic retroelements containing an endonuclease (EN) domain of the GIY-YIG type and a reverse transcriptase (RT) domain that is most closely related to the eukaryotic telomerase reverse transcriptases (TERTs) [1,2]. Bioinformatic analysis reveals that, in addition to the seven highly conserved core RT motifs, homology between PLE RT and TERTs extends to the N-terminal as well as the C-terminal (thumb) domains. The special relationship between PLEs and TERTs is also manifested in association of certain PLEs with telomeric repeats, which is observed in two PLE clades, from fungi (Coprina) and from rotifers (Athena). In a combined phylogeny of PLEs and TERTs, the fungal and rotifer PLEs represent the earliest-branching PLE clades, and the EN domain in all members of these two clades does not contain any of the conserved motifs normally required for function of endonucleases of this type. It may be hypothesized that loss of the EN activity, and therefore of the ability to insert at internal chromosomal sites, is associated with the ability of these elements to prime reverse transcription with the available 3\' OH at the chromosome termini. Moreover, the Athena elements from bdelloid rotifers are organized into what appears to be a pair of specialized retroelements, one of which has the potential to provide the RT function, and the other has a non-functional RT domain but codes for a gag-like ORF1 product that may have the potential for telomeric targeting, similar to the pair of telomere-associated retroelements performing chromosome end maintenance function in Drosophila. It is possible that in early eukaryotic evolution similar processes could have given rise to telomerase reverse transcriptases and the corresponding accessory proteins facilitating interaction of telomerases with the chromosome termini.

1. Arkhipova, I.R., Pyatkov, K.I., Meselson, M., and Evgen'ev, M.B. (2003). Retroelements containing introns in diverse invertebrate taxa. Nat. Genet. 33:123-124.

2. Evgen'ev, M.B., and Arkhipova, I.R. (2005). Penelope-like elements - a new class of retroelements: Distribution, function, and possible evolutionary significance. In: Retrotransposable elements and genome evolution, J.-N. Volff, ed., Karger AG: Basel. Cytogenet. Genome Res. 110: 510-521.

Origins of ultraconservation and distal cis-regulation in vertebrates

Gill Bejerano

University of California Santa Cruz

A combined evolutionary and functional proof is presented that a human distal enhancer of a neuro-developmental gene, as well as an ultraconserved exon involved in post-transcriptional regulation, are both derived from an ancient retroposon.

Single Amino Acid Substitutions in L1 ORF1p with Dramatic Effects on Nucleic Acid Chaperone Activity and L1 Retrotransposition

Sandy Martin[*1,2], *Ann Walker*[1], *Dan Branciforte*[1], *Patrick Wai-Lun Li*[1,3], *Fei Wang*[4] *and Mark Williams*[4,5]

[1] Department of Cell and Developmental Biology
[2] Molecular Biology and [3] Human Medical Genetics Programs, University of Colorado School of Medicine, Aurora, CO
[4] Department of Physics and [4,5] Center for Interdisciplinary Research on Complex Systems, Northeastern University, Boston, MA

L1 achieved its high copy number in mammalian genomes by duplicative transposition through an RNA intermediate, i.e., retrotransposition. L1 RNA is the template for translation of the two, cis-acting, L1-encoded proteins and for reverse transcription by target-primed reverse transcription (TPRT). The ORF2 protein encodes endonuclease and reverse transcriptase, both required for TPRT. These activities were predicted by homology. Similar analyses of the ORF1 sequence failed to reveal homology to protein(s) of known function, hence the role of ORF1p has been elusive. Sequence comparisons of ORF1s from mammalian and fish L1s revealed a coiled-coil domain and a conserved domain. The basic region of the conserved domain interacts strongly with nucleic acids, with highest affinity for RNA. The coiled-coil domain is responsible for homotrimerization of the mouse ORF1p. In addition to its binding activities, ORF1p from mouse L1 is a potent nucleic acid chaperone. Here, we describe a natural polymorphic variant of the TF class of mouse L1 that shows elevated retrotransposition frequency in cultured cells, TFC. 20nt substitutions distinguish TFC from TFspa, an element that jumped recently in mice. We isolated the retrotransposition pheonotype to a single amino acid substitution in ORF1, and compared the biochemical activities of purified TFC and TFspa proteins using several in vitro assays. The RNA binding properties of these proteins did not differ significantly, rather, the effect on retrotransposition appears to be due to subtle, but significant differences in their nucleic acid chaperone activities. Specifically, the TFC protein is more effective as a chaperone because it reduces the cooperativity of the helix-coil transition. This finding, combined with previous results showing that single amino acid substitutions in the conserved region of ORF1p reduce or abolish nucleic acid chaperone activity and retrotransposition, further underscores the importance of the nucleic acid chaperone activity of ORF1p to L1 retrotransposition.

Retrotransposon target specificity and genome organization

Xiang Gao [1], *Hou Yi* [1], *Hirotaka Ebina* [2], *Henry Levin* [2], *and *Daniel F. Voytas ** [1]

[1] Genetics, Development and Cell Biology Deparment, Iowa State University, Ames, IA 50010
[2] Building 6B, Rm 2B-220, National Institute of Child Health and Human Development, Bethesda, MD 20892

Recent evidence suggests that retrotransposons contribute to heterochromatin formation through the siRNA pathway and to centromere function through interactions with centromeric proteins. We hypothesize that retrotransposons became integral components of heterochromatin by actively targeting integration to heterochromatic domains. This hypothesis emerges from our work with the yeast Ty5 retrotransposon, which targets to heterochromatin through an interaction between the C-terminus of Ty5 integrase and the heterochromatin protein Sir4p. Here we report families of retrotransposons that have chromodomain-like motifs in the C-termini of their integrases. Many of these retrotransposons are exclusively located within centromeric heterochromatin. The chromodomains fall into three classes, ranging from those that are highly similar to HP1 (Class I) to those that are highly divergent and share only a few amino acid residues with other chromodomains (Class III). Chromodomain-YFP fusions appear as distinct foci within the nucleus that co-localize with CFP fusions to the Arabidopsis HP1 homologue (LHP1). Mutations in conserved residues in the chromodomain diminish the sub-nuclear localization. Pull-down assays indicate that the Maggy chromodomain (Class I) interacts with histone H3 that is dimethylated on Lys9. Collectively, these data support the hypothesis that the novel chromodomains mediate retrotransposon target specificity by recognizing specific chromatin features.

Human syncytins – an extreme example of transposable element domestication

J. Hejnar[*1], *M. Matouskova*[1], *J. Blazkova*[1], *A. Pavlicek* [1,2], *L. Fricova*[1].

[1] Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague
[2] Genetic Information Research Institute, Mountain View, CA

Some transposable elements within the human genome remain intact open reading frames and display a huge translational potential. Up to 8% of the human genome is of retroviral origin, comprising numerous endogenous retroviruses incorporated at some point during evolution. Endogenous retroviral gag and env genes are often expressed in various human tissues and provide e.g. restriction factors interfering with exogenous virus infection and superantigens. Class W and FRD human endogenous retroviruses are most intimately involved in early human development, providing fusogenic env proteins, syncytins, necessarry for trophoblast differentiation in placenta. We present here several genomic and functional features of this retrovirus domestication:
i. There is an inverse correlation between CpG methylation of syncytin-1 long terminal repeat and its expression. CpG methylation plays a principal role in the transcriptional suppression of syncytin-1 in non-placental tissues, where the inadvertent cell fusion might be dangerous for tissue organization and integrity. In contrast, demethylation of the syncytin-1 promoter in trophoblast is a prerequisite for its expression and differentiation of multinucleated syncytiotrophoblast. ii. Correct splicing of subgenomic env mRNA is needed for translation of both syncytins and tissue-specific splicing factors are involved in tight control of syncytin expression. iii. Being expressed (in non-spliced form) in testicular tissues, syncytin-1 provirus serves as a good template for L1-mediated retrotransposition and numerous processed pseudogenes were detected within the HERV-W family.

Enormous impact of retroelements on the genome structure and evolution in land vertebrates

Kordiš, Dušan

Department of Biochemistry and Molecular Biology, Josef Stefan Institute, Ljubljana, Slovenia.
dusan.kordis@ijs.si

Retroelement repertoires have experienced extensive changes during the evolution of land vertebrates. In contrast to the mammals and birds, teleost fishes contain very diverse retroelement repertoires. Major evolutionary changes in retroelement repertoires have therefore taken place in the land vertebrates. Teleost fishes offer a unique advantage for comparative genomics as an outgroup for tetrapods and can provide information about the ancestral state and distinguishing trends in retroelement evolution within the land vertebrates. To investigate the tempo and mode of retroelement evolution a comprehensive survey of retroelement repertoires was conducted using the massive amount of genome sequence data from diverse tetrapod lineages, amphibians and diverse reptilian lineages, to the most ancestral mammalian lineages (monotremes and marsupials). Retroelement diversity in Tetrapoda, as shown by phylogenetic analyses, was found to be much greater than previously expected. Deciphering the ancestral state of retroelement repertoires in diverse amphibian and reptilian lineages, as well as in the most ancestral mammalian lineages, is pivotal to understanding their evolution in land vertebrates. Recognition of the retroelement repertoires in amphibians and reptiles, as well as in monotremes and marsupials, is very important, since it shows for the first time the very diverse retroelement repertoires in both ancestral tetrapod lineages and enables description of their evolutionary history and dynamics in Tetrapoda on the evolutionary time scale of 400 My. The first global insight into the evolutionary genomics of retroelements in the genomes of Tetrapoda will be presented.

Effects of LTR elements on mammalian genes

Dixie L. Mager, Mark T. Romanish, Louie N. van de Lagemaat, Daphne Reiss and Irina A. Maksakova*

Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, BC, Canada and
Deptartment of Medical Genetics, University of British Columbia

Endogenous retroviruses (ERVs) and LTR retrotransposons comprise nearly 10% of mammalian genomes. While retrotranspositional activity of these elements in humans is very low, many ERVs/LTRs still contain active regulatory motifs with the potential to affect cellular genes. Our laboratory has employed several approaches to investigate the impact of ERV LTRs on mammalian gene expression. We have observed that LTRs can serve as one of several alternative gene promoters or, more rarely, can act as the gene's dominant or only known promoter. One intriguing example will be presented in which different LTRs have assumed roles in expression of the orthologous NAIP anti-apoptotic genes in human and mouse. We have also shown that HERV-E LTRs which serve as alternative gene promoters in placenta are much less methylated in that tissue compared to highly related, randomly chosen HERV-E LTRs, suggesting that epigenetic context plays an important role in the probability of LTRs contributing to gene transcription. In other work on epigenetic regulation, we have investigated the actively retrotransposing MusD/ETn endogenous retroviral elements in the mouse and have uncovered evidence for differential methylation of these LTRs, which depends on the adjacent interior sequence. The significance of this finding in explaining the current activity of ETn elements will be discussed.
Finally, in a bioinformatics study, we have analyzed human ERV distributions within gene introns. Density profiles of ERVs across transcriptional units differ for individual ERV families but suggest that LTR polyadenylation signals are the major target for negative selection for most families, resulting in the previously described general antisense bias of LTRs within introns. Furthermore, we detected a unique intronic density profile for ERV9 LTRs, likely reflecting the complex structure of these LTRs. These findings suggest that different ERV families have distinct effects on genes and therefore are subject to differing levels of purifying selection.

Parasites and parasites of parasites: Plant retrotransposons and their genomic impact

François Sabot[1], *Ruslan Kalendar*[1], *Wei Chang*[1], *Jaakko Tanskanen*[1,2], *Marko Jääskeläinen* [1], *and Alan Schulman* [1,2] *

[1] MTT/BI Plant Genomics Laboratory, Institute of Biotechnology, Viikki Biocenter, University of Helsinki, P.O. Box 56, Viikinkaari 4, FIN-00014 Helsinki, Finland
[2] Plant Breeding Biotechnology, MTT Agrifood Research Finland, Myllytie 10, FIN-31600 Jokioinen, Finland

Class I LTR retrotransposons comprise up to 85% in the Triticeae (wheat, barley) genomes and serve as major evolutionary agents. By their sheer numbers and activity, they have a great impact on genome structure and genetic diversity, providing new regulatory patterns and responses to environmental challenges. Autonomous elements encode the main proteins needed for reverse transcription, packaging, and integration. Many copies of a given retrotransposon, which are found in the genome, contain stop codons or frameshifts affecting the encoded polyprotein. However, two types of retroelements, the TRIMs (Terminal Repeats In Miniature) (Witte et al., 2001; Antonius-Klemola et al. 2006) and LARDs (LArge Retrotransposon Derivatives; Kalendar et al. 2004) have been recently identified that code for no proteins, although they possess conserved structures. These non-autonomous elements are probably able to retrotranspose using the enzymatic system of one or more active partners. We are currently investigating the life cycle of LARDs and TRIMs.

Kalendar R, Vicient CM, Peleg O, Anamthawat-Jonsson K, Bolshoy A, Schulman AH (2004) LARD retroelements: Conserved, non-autonomous components of barley and related genomes. Genetics 166: 1437-1450

Witte CP, Le QH, Bureau T, Kumar A (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci USA 98: 13778-13783

Antonius-Klemola K, Kalendar R, Schulman AH (2006) TRIM retrotransposons occur in apple and are polymorphic between varieties but not sports. Theor Appl Genet 113: in press

The contribution of LTR retrotransposons to gene evolution: a tale of three genomes.

John McDonald[1]*, Eric Ganko*[2]*, Jeremy DeBarry*[3] *and Nalini Polavarapu*[1]

[1] School of Biology, Georgia Institute of Technology, Atlanta, GA, USA
[2] Dept Biology, University of North Carolina, Chapel Hill, NC, USA
[3] Dept Genetics, University of Georgia, Athens, GA, USA

Studies on the evolutionary significance of transposable elements (TEs) focuses on two not mutually exclusive phenomenoa: the evolution of TEs themselves and the impact TEs have had on the evolution of the host genomes in which they reside. In this presentation, the results of studies on the evolution and impact of LTR retrotransposons in Drosophila, mice and chimps/humans will be compared and contracted. Evolutionary patterns emerging from these studies will be discussed.

Contrasting patterns of sequence turnover of LTR retrotransposons in different eukaryotes

López-Sánchez [1]*, P., J. M. C. Tubío* [3]*, J. C. Costas* [2]*, and H. F. Naveira*[1]*

[1] Dep. Bioloxía Celular e Molecular, Univ. A Coruña, Spain
[2] Unidade de Mediciña Molecular, INGO, Complexo Hospitalario Univ. de Santiago de Compostela, Spain
[3] Dep. Xenética, Univ. Santiago de Compostela, Spain

The recent availability of the genomes of different species of higher eukaryotes offers an extraordinary opportunity for comparative studies of diversity and evolutionary dynamics of TEs. To begin with, the fraction of the genome occupied by this kind of repeats varies widely among species, from >70% in Zea mais to roughly 2% in Caenorhabditis elegans. In our own species, 45% of the euchromatic genome corresponds to TEs, and their density in some chromosome regions is impressive. Contrary to Drosophila, which shows very short coalescence times for TE sequences, our genome is plagued with “fossil” remnants of mobilization periods that ceased long ago. Among all the different kinds of TEs, we have chosen to carry out an interspecific comparative study of LTR retrotransposons. Thus, we have made detailed phylogenetic reconstructions and estimated the relative abundance of the nine different lineages of the Ty3/gypsy group in Drosophila melanogaster and Anopheles gambiae. In this way, conspicuous differences in relative evolutionary success of the different lineages within each genome were made apparent. Also, anlyses of variation within families of the Ty3/gypsy group, including estimates of soliltary LTR insertions (solos), indicate that sequence turnover is probably much less intense in mosquitoes than in Drosophila. Low turnover rates of TE sequences are also most likely to have produced the observed pattern of insertions in primate genomes. Again, we have restricted our attention to LTR retrotransposons, namely to several families of human endogenous retroviruses (HERVs) that became extinct a long time ago. In this communication we offer some hints of the mechanisms that led to these extinctions, which apparently took place rather “abruptly”, and not as a consequence of a more or less slow progressive reduction of its transposition rate. Possible explanations and consequences of these findings are discussed.

A genomic approach to depict transcriptionally active transposable elements in sugarcane

*Rossi M, Costa APP, de Jesus EM, Domingues DS, Saccaro-Junior N, Kajihara D, Van Sluys MA**

GaTE Laboratory, Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, Brazil. (mavsluys@usp.br)

Eukaryotic genome structure and biological functions evolved, among other strategies, through the gain of nuclear interspersed sequences either as introns within coding genes or as junk DNA between coding units. Among the players in that scene, transposable elements (TEs) have been associated as generators of variability and genome instability due to their inherent mobilization attributes. In order to study the genome impact of TEs in eukaryotic genome evolution, we have focused on sugarcane, a highly poliploid genome among grasses, that is a recent hybrid generated by crosses between Saccharum officinarum and Saccharum spontaneum. Its 1C genome encompasses at least 3000Mbp, which renders gene cloning and genome sequencing difficult tasks. Advantage was taken from the SUCEST project to circumvent this handicap and a study was made in the EST database was composed of 237,954, mostly 5' cDNA, sequences. A total of 276 individual TE sequences were initially identified and, after full length cDNA sequence, annotation and manual characterization of readthrough transcripts, an array was set up to study TE expression patterns. Macroarray studies on four distinct plant tissues (callus, apical meristem, leaf roll and flower) identified callus as the tissue where most TE families are expressed (Araujo and Rossi et al; Plant J, 44:707, 2005). Interesting to note, is that these results also pointed to the fact that within a given somatic tissue of a plant, cells are expressing different TE families concomitantly. Hopscotch-like retrotransposon clones represent the major retroelement being expressed and Mutator-like clones represents the most expressed classical DNA transposons. Further molecular analyses, such as expression pattern and genomic distribution among sugarcane cultivars and parental lines, are being carried to better characterize Mutator-like and hAT-like transposons as well as SURE and Hopscotch-like retroelements.

Financial support: FAPESP and CNPq (Brazil).

Retroposition of processed pseudogenes: the impact of RNA stability and translational control

Adam Pavlicek[1]*, *Andrew J. Gentles*[1], *Jan Paces*[2], *Vaclav Paces*[2] *and Jerzy Jurka*[1]

[1] Genetic Information Research Institute, 1925 Landings Drive, Mountain View, CA 94043, USA
[2] Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo 2, Prague CZ-16637, Czech Republic

Human processed pseudogenes are copies of various cellular RNAs reverse transcribed and inserted into the nuclear genome by the enzymatic machinery of L1/LINE1 non-LTR retrotransposons. While it is generally accepted that germline expression is crucial for heritable transposition of cellular mRNAs, almost nothing is known about the influences of RNA stability, mRNA quality control, and compartmentalization of translation on retroposition of processed pseudogenes. We investigate the role of RNA stability and translation on the retroposition of cellular mRNAs by combining data on processed pseudogenes found in the human genome with published microarray studies on mRNA stability, and predictions of protein localization. We found that frequently retroposed mRNAs are derived from stable transcripts with translation-competent functional reading frames resistant to nonsense-mediated RNA decay (NMD). They are preferentially translated on free cytoplasmic ribosomes and encode soluble proteins. On the other hand, unstable transcripts, mRNAs with premature termination codons, and transcripts encoding membrane and secreted proteins are retroposed at much lower frequencies. This is the first study linking several processes on the RNA level to retroposition of processed pseudogenes in the human genome. Our results strongly indicate that interactions between mRNAs and L1 proteins occur at or near free cytoplasmic ribosomes.

Validation of diverged repetitive elements using phylogenetic analysis and comparative genomics approach

Wojciech Makalowski

Institute of Molecular Evolutionary Genetics and Department of Biology, Center for Comparative Genomics and Bioinformatics, The Pennsylvania State University, University Park, PA 16802, USA

Several studies demonstrated a large number of the TE-derived sequences present in vertebrate mRNAs. Their presence in the open reading frames is especially intriguing because of the potential direct impact on vertebrate proteins\' evolution. However, our recent study showed that many of these cassettes may not translate into proteins but be rather involved in the regulation of a cognate gene expression. Nevertheless, we observe some TE-cassettes that seem to be exapted for protein coding. Interestingly, they all originate in old TEs, e.g. L3, L4, MIR, and their respective alignments to a consensus sequences fall into a gray zone of statistical significance. This poses a challenge how to distinguish a false positive alignment of some random sequences from a real TE-cassette, since a statistical significance does not always imply biological significance. It appears that a phylogenetic analysis and/or comparative genomic approach might be very helpful in this process. I will discuss several techniques used in our lab to validate TE-cassettes along with the discussion of both, real and false positive, TE-cassettes found in human proteins.

Analysis of the relative chronological age of human transposable elements by defragmentation and insertional analysis

Peter E. Warburton[*1], *Yongchao Ge*[1], *Joti Giordano*[1], *Yevgeniy Gelfand*[2], *Gary Benson*[2]

[1] Mount Sinai School of Medicine, New York, USA
[2] Boston University, Boston, USA

We describe a novel evolutionary analysis of the chronological age of interspersed transposable elements (TEs) found in the human genome. The genome has been constantly bombarded by different TEs over millions of years, which now comprise ~45% of the total DNA sequence. We have performed a genome-wide defragmentation analysis of TEs to identify insertions of one TE into another. Arrangements of TEs not representing independent transposition events were excluded, including several arrays of tandemly duplicated TE fragments. This defragmentation analysis has provided a large and powerful data set of hundreds of thousands of transpositions of TEs into other TEs. The relative age of these TEs is implicit in their arrangement because newer elements transpose into older elements that are already present. A matrix of the number of interruptions of each TE into each other TE was constructed, and an objective function was developed that orders these transposons by chronological age. The relative age of ~325 human TEs from all four classes spanning ~100 million years was determined, which is in agreement with the limited phylogenetic analyses of a few families of TEs. This method provides a statistical estimate of the active lifespan of each TE and which TEs were contemporary with each other over evolutionary history. The results are further supported by independent analysis of 5 other mammalian genomes, which confirm that transposons shared between genomes are older than species-specific transposons, which has been used to derive mammalian phylogenetic trees. This is the first analysis of molecular evolution that is not dependent on the assumption of a constant molecular clock.

Automated palaeontology of repetitive DNA with REannotate

Vini Pereira *

University of Sussex, UK

Annotation of transposable elements (TEs) is essential for genome biology. Genome annotation is mostly done by running RepeatMasker with a library of Repbase reference sequences. However, where TEs overlap or have been fragmented, this annotation has little information on the genesis and evolution of TE sequences. Here I present a computational tool for re-annotation and evolutionary analysis of repetitive DNA, REannotate, which takes RepeatMasker annotation as input. REannotate outputs three layers of automated evolutionary analysis for inference of i) common origin of repetitive DNA segments; ii) the temporal order of their origin, if repeats overlap; and iii) the age of the insertion events, for long terminal repeat (LTR)-elements.

The method achieved over 93% accuracy when compared to manual evolutionary analysis of nested TEs in wheat and maize. Re-annotation of human chromosomes provided evidence for a recent expansion of satellite repeats on the Y chromosome, and for a higher rate of evolution on the Y relative to the autosomes. Detection of DNA rearrangements in TE clusters in Drosophila and Arabidopsis suggests that clusters may expand via mechanisms in addition to transposition.

I used the age distributions of retrotransposons from the re-annotation of the Arabidopsis thaliana genome to show that two major groups of LTR-elements have accumulated in the heterochromatic, pericentromeric regions via distinct mechanisms. Metaviridae elements have evolved a targeting preference for these heterochromatic regions; in contrast, Pseudoviridae elements integrate randomly but have been under constant turnover and elimination from euchromatic regions by purifying selection. I estimate the half-life of complete Pseudoviridae elements in A. thaliana euchromatin as 470,000 +/- 50,000 years. A decline in Metaviridae activity and the elimination of Pseudoviridae insertions have both limited the retrotransposon contribution to genome size expansion in Arabidopsis.

Comparative genomics analysis of alu gene conversions

Degui Zhi, Mark Chaisson

Bioinformatics Program, University of California, San Diego, USA

Although the existence of alu gene conversion (AGC) events has been established for more than 10 years, we do not yet have a good understanding of AGC events in human genome. With the human genome sequence, Roy et al 2000 indirectly inferred AGC events via the identification of putative mosaic aluYa5 elements containing diagnostic positions that violates the sequential order based on phylogenic data.

With the recent available chimpanzee genome sequence, we conducted a genome-scale analysis of AGC events based on a direct comparison of human and chimpanzee genome sequences. We examine alus in orthologous locations, and identify sequence features from which evolutionary history of alus could be inferred, including signatures for an AGC event.

AGC substitutes portion of an alu into the corresponding portion from another independently inserted alu. If the two alus has different sequence, one would expect a large number of mismatches localized in a narrow region between the orthologous alus from human and chimpanzee.

We found that among 14833 orthologous alu pairs between human chr22 and chimpanzee chr23, 1.9% (285) pairs have >11 mutations. A significant portion of them contain a narrow region (>80bp) with > 10% sequence divergence. Especially, we found that 109 (5.3%!) out of 2060 aluY pairs have >11 mutations. In addition, we found 2568 out of the 14833 orthologous alu pairs contain gaps. Many of the gaps are in the A-rich tail. In summary, we demonstrate that comparative genomics approach could provide interesting clues for the revelation of AGC events.

Reference:
Roy et al. Genome Res. Vol. 10, Issue 10, 1485-1495, 2000

Dynamics of transposable elements: first steps of invasion and long term evolution.

*A. Le Rouzic et P. Capy**

Laboratoire Populations, Génétique, Evolution, CNRS, 91198 Gif/Yvette Cedex, France.

The evolutionary dynamics of transposable elements are complex, due to the interaction between their intrinsic amplification capacity, selection at the host level, transposition regulation, and genetic drift, raising several questions about the relevant factors. To address these questions approaches based on modelling and/or simulations can be carried out. Here, we present an analysis of their dynamics in different parts of their life cycle, i.e. just after a horizontal transfer and during the long-term co-evolution with the genomes.

It has been shown that when a single copy is present in the genome of one individual, it is almost impossible to maintain the element with a transposition rate in agreement with those observed in natural populations. However, elements whose transposition rate is regulated are able successfully to invade populations, thanks to an initial transposition burst, followed by a strong limitation of their activity. Self-regulation or hybrid dysgenesis may thus represent some genome-invasion parasitic strategies.

The behaviour of a transposable element family (including autonomous and non-autonomous copies) was next followed over several thousand generations. Several parameter sets were applied to the model, which is based on a quantitative approach to transposition activity and selective impact on the host genome. According to the parameter sets chosen, five characteristic dynamics have been identified and described. These include loss, equilibrium state, uncontrolled invasion, cyclic dynamics and domestication. Most of them have been observed in different organisms and sometimes within the same genome. Moreover, it is interesting to stress that the cyclic dynamics observed between autonomous and non-autonomous copies is similar to that described for host/parasite relationships in ecology, suggesting that ecological models could be used at the genome level.

Conflict, compromise or cooperation - Different ways for transposons and genomes to coexist

Andy Flavell[*1], *Runchun Jing*[1], *Naeem Syed*[1], *Noel Ellis*[2], *Gary Muehlbauer*[3], *Boulos Chalhoub*[4]

[1] Plant research Unit, University of Dundee, UK
[2] John Innes Centre, Norwich, UK
[3] University of Minnesota, USA
[3] URGV-INRA, Evry, FRANCE

Transposable elements are ubiquitous genomic parasites with an ancient history of coexistence with their hosts. Many plant transposons have been tolerated by their hosts to the point where they have made substantial changes to genome structure and size. We will present the results of studies on pea and cereal retrotransposons and barley MITEs that illustrate the different ways that a transposon can survive, flourish or perish in its host. A few cases have emerged recently where transposons have been recruited for normal function in the host organism. We will also present data for a cereal grass Ac superfamily element which represents such a case.

Population dynamics of a comprehensive set of transposable elements in the D. melanogaster genome

Petrov, D.A., Lipatov, M., and Lenkov, K.*

Stanford University, Department of Biological Sciences, Stanford, CA 94305

Transposable elements (TEs) are a major and extremely active component of eukaryotic genomes. TEs constitute over 90 % of DNA in some eukaryotes. They are also responsible for a large proportion of visible mutations (~50% in D. melanogaster) and an unknown proportion of advantageous mutations. Many TEs are located within or near genes and may play important roles in gene regulation and function. Beyond being burdensome parasites, TEs are sometimes co-opted into functioning of eukaryotic genomes (e.g. they constitute telomeres in Drosophila) and have been known to cause adaptive mutations in a number of well-established cases. The overall role that TEs play in adaptation remains unclear.

The sequencing of the euchromatic portion of the D. melanogaster genome provided an unprecedented opportunity to study evolutionary forces affecting the maintenance and function of TEs. To begin this endeavor, we have attempted to collect population frequency of all TEs found in the sequenced D. melanogaster genome. We will present data on the frequency distribution of ~950 individual identified TE (out of the total ~1200 TEs) in 72 naturally isolated D. melanogaster strains. We will present evidence supporting the model that ectopic recombination among dispersed TE families is a major force limiting the spread of TEs in the D. melanogaster genome. We will report estimates of the strength of natural selection acting against individual TE families in D. melanogaster. Strength of natural selection varies widely among different families and we will discuss underlying causes of this variation. We will also argue that most (80-90%) of TE-gene associations are deleterious and that most of the ones that we observe in the genome do not generate appreciable selective effects (either deleterious or adaptive) due to their associations with genes. Finally, we will also describe the discovery of a number of putatively adaptive TE insertions and will use these data to discuss the importance of TEs in molecular adaptation.

The organization of the human genome: from chromosomal bands to isochores

Giorgio Bernardi[1]*, *Maria Costantini*[1], *Oliver Clay*[1], *Salvo Saccone*[2], *Fabio Auletta*[1]

[1] Laboratorio di Evoluzione Molecolare, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy
[2] Dipartimento di Biologia Animale "M. La Greca", University of Catania, Via Androne 81, 95124 Catania, Italy

The compositional mosaic of human chromosomes has been defined at its three levels of resolution: low (400-band), high (850 band) and highest (~3200 isochores). At all three levels band borders have been mapped at a 100 kb resolution, a value well below that provided by cytogenetics (1 Mb or more). Below 100 kb, the GC profile becomes turbulent owing to the contribution of specific sequences (exons, introns, interspersed repeats, etc.)
The major results were 1) finding the "rules" that underlie the organization of chromosomal bands; and 2) mapping band borders on compositional grounds alone. The first point provides the best evidence for a precise order in the structuring of chromosomal bands. The second point is tantamount to explaining the molecular basis for chromosomal banding. Both points are relevant for our understanding of the impact of transposable elements on genome organization and evolution.

References

Bernardi G. (2004). Structural and Evolutionary Genomics. Natural Selection in Genome Evolution. Elsevier, Amsterdam. (Reprinted in 2005).
Costantini M., Clay O., Auletta F., Bernardi G. (2006). An isochore map of human chromosomes. Genome Res. (in press).
Costantini et al. (2006). Bands and isochores in human chromosomes. (In preparation).

Substitution pattern of mammalian transposable elements - element specific, regional, and evolutionary aspects.

*Peter F Arndt **

Max Planck Institute for Molecular Genetics, Berlin, Germany

The human genome evolved over millions of years towards its current form. Despite this long time its base composition did not reach a stationary state yet. Instead long range correlated fluctuations of the Guanine-Cytosine (GC) are observed. This so called isochore structure of the genome is found in almost all warm-blooded vertebrates. To learn more about the driving force responsible for this pattern we analyzed regional differences in substitutional biases. Our analysis reveals regional differences of substitution patterns in human Alu elements [1]. We observe substitutional hot spots and a GC enriching substitution pattern at telomeres. We further discuss differences of substitution pattern in transposable elements and unique intergenic sequence segments.

References:

[1] Peter F Arndt, Terence Hwa, Dmitri A Petrov Substantial Regional Variation in Substitution Rates in the Human Genome: Importance of GC Content, Gene Density, and Telomere-Specific Effects. J Mol Evol 60 (2005) 748-63

Life after death: reincarnation of DNA transposons into genetic networks. A case study in the human genome

Cedric Feschotte

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

There is a growing body of evidence that DNA transposons and their encoded transposase have been a frequent source of protein domains for the assembly of new genes during evolution. One possible explanation for the recurrent use of transposase domains for the assembly of new genes is the ability of these enzymes to bind specifically DNA at many unlinked chromosomal sites in the genome within the cognate interspersed transposons. We hypothesize that the recruitment of a transposase DNA-binding domain as part of a new host protein opens the door for natural selection to rapidly assemble a network of DNA targets among multiple binding sites dispersed throughout the genome. In order to test this model we have initiated a detailed study of SETMAR, a human gene of unknown function that originates from the transcriptional fusion of a SET domain with histone methyltransferase activity to a mariner transposase. We show that SETMAR has emerged between 58 and 40 myr ago through an intricate stepwise process involving transposition, genomic deletion and the creation of a new intron. The transposase domain of SETMAR has been subject to purifying selection in all the extant lineages of anthropoid primates, suggesting that the addition of a transposase domain to the pre-existing SET domain led to the advent of a beneficial new function in primates. The signature of purifying selection is particularly intense on the N-terminal region of the transposase, while the catalytic domain is evolving essentially neutrally. We present biochemical and computational data that support a model whereby the specific DNA-binding activity of the ancestral transposase has been retained and now provides a means to target the SET domain to multiple sites within the human genome where it can act to modulate the structure of the surrounding chromatin.

The impact of Mutator-like elements on genome evolution

Ning Jiang[1], Zhirong Bao[2], Xiaoyu Zhang[3], Sean R. Eddy[4] and Susan R. Wessler[5]*

[1] Department of Horticulture, Michigan State Univerisity, East lansing, Michigan 48824
[2] Department of Genome Sciences, University of Washington, Seattle, Washington 98195
[3] Department of Molecular, Cell, and Developmental Biology, UCLA, California 90095
[4] Department of Genetics, Washington University, St Louis, Missouri 63108
[5] Department of Plant Biology, University of Georgia, Athens, Georgia 30602.

Mutator-like transposable elements (MULEs) are found in many eukaryotic genomes and are especially prevalent in higher plants. In maize, rice and Arabidopsis a few MULEs were shown to carry fragments of cellular genes. These chimeric elements are referred as Pack-MULEs in this study. Here we report that there are over 3000 Pack-MULEs in rice containing fragments derived from more than 1000 cellular genes. Pack-MULEs frequently contain fragments from multiple chromosomal loci that are fused to form new ORFs, some of which are expressed as chimeric transcripts. About 5% of the Pack-MULEs are represented in cDNA collections. Functional analysis of amino acid sequences and proteomic data suggest that some captured gene fragments may be functional. Comparison of the cellular genes and Pack-MULE counterparts indicates that fragments of genomic DNA have been captured, rearranged and amplified over millions of years. In addition, an analysis of 32 Mb genomic sequences from Lotus. japonicus, a dicot plant, indicates that the Pack-MULEs in L. japonicus are similar to those in rice in terms of copy number, expression level, and distribution. Given the abundance of Pack-MULEs in both rice and L. japonicus, as well as the widespread occurrence of MULEs in all characterized plant genomes, gene fragment acquisition by Pack-MULEs may represent an important new mechanism for the evolution of genes in higher plants.

Helitrons and the Evolution of DNA Sequence Diversity in Maize

Antoni Rafalski[1]*, Stephan Brunner*[1]*, Kevin Fengler*[1]*, Michele Morgante*[2] *and Scott V. Tingey*[1]

[1] DuPont Crop Genetics, P.O.Box 80353, Wilmington, DE19880-0353
[2] Dipartimento di Produzione Vegetale e Tecnologie Agrarie, Universita' di Udine, Via delle Scienze 208, Polo Scientifico Rizzi – 33100 Udine, Italy
email: j-antoni.rafalski@cgr.dupont.com

We have analyzed the DNA sequence diversity of maize at multiple levels of resolution: Whole chromosome, gene organization level (100-300 kb range) and individual gene polymorphism. At all levels of resolution, intraspecific genetic diversity of maize is extremely high. As has been noticed by Fu and Dooner, maize inbred lines McC and B73 not only differ extensively in the repetitive DNA segments, but also in some cases genic sequences present in one allele are found missing in the other allele (Fu H. and Dooner H.K., Proc Natl Acad Sci USA 2002, 99: 9573-9578). We have extended these observations to other genetic loci and surveyed the occurrence of genic and intergenic non-colinearities in maize gene pool. Four genomic segments of ca 250 kb each have been sequenced from two inbred lines, B73 and Mo17. Intergenic non-homologies were common and frequently consisted or relatively recent insertions of retrotransposon-derived sequences. The differences between the two alleles ranged from 20% to nearly 50%, depending on the locus. Genic non-homologies were relatively rare and were usually pseudogenes, even though some of them are transcribed. Most of these occur as clusters of exonic insertions which were captured and transposed by Helitron transposable elements. Some helitron-generated clusters of exons are transcribed and may give rise to novel functionality over the course of evolution. We predict that more than 10,000 such genic polymorphisms exist in the maize genome, most of them carried on Helitron insertions (Morgante et al. Nature Genetics in press). We have analyzed in detail a family of related Helitrons in maize and teosinte germplasm, and propose their evolutionary relationship. Candidates for the autonomous Helitron elements, as well as their transcripts, have also been identified. The consequences of these findings for evolutionary history of maize will be discussed.

Transposable elements in filamentous fungi

Marie-Josée Daboussi

Institut de Génétique et Microbiologie, Université Paris-Sud, F-91405 Orsay cedex, France;
email :marie-jose.daboussi@igmors.u-psud.fr

In recent years, great progress has been made in the characterization of transposable elements (TEs) that inhabit the fungal genome and a clearer picture of their genomic organization as well as their evolution is emerging. All eukaryotic types described are found, with an extraordinary prevalence of active members of the pogo family. The role of TEs in mutation and genome organization began to be well documented, leading to significant advances in our perception of the mechanisms underlying genetic changes in these organisms. TE-mediated changes, associated with transposition and recombination, have been shown to provide a broad range of genetic variation, useful for natural populations in their adaptation to environmental constraints, especially for those lacking the sexual stage. Interestingly, some fungal species have evolved distinct silencing mechanisms which are regarded as host defense systems against TEs. The examination of forces acting on the evolutionary dynamics of TEs should provide important insights into the interactions between TEs and the fungal genome. Another issue of major significance is the practical applications of TEs in gene tagging and population analysis, which will undoubtly facilitate research in systematic biology and functional genomics.

BAC-enabled comparative investigation of CR1 evolutionary dynamics in the reptilian genomic landscape

Andrew Shedlock

Harvard University, Cambridge, MA, USA

Genome scanning methods that utilize paired end-sequences of clones from large DNA insert genomic libraries (i.e., BAC, cosmid, plasmid) are particularly well-suited to help bridge deep phylogenetic gaps separating whole-genome assemblies presently available among major vertebrate clades. Evolutionary interpretations of chicken-mammal comparisons, for instance, remain difficult in the near complete absence of detailed information for non-avian reptile genome structure. Using BACs to explore the repetitive landscape in the sister lineage to birds can facilitate more accurate reconstruction of ancestral genomic states that may have existed in the amniote common ancestor and help infer what molecular processes likely led to the structural diversity seen today in mammalian genomes and those of their sister group, Reptilia.

Results are presented from comparative analysis of mobile and tandem repeats detected in ~7 Mb of paired sequence reads from genomic libraries of three non-avian target species representing distinct reptilian lineages: American Alligator, Painted Turtle, and Green Anole. In addition, preliminary results from a similar survey of genomic libraries of three higher land bird species, namely the Hoatzin, Rock Dove and Yellow-billed Cuckoo are reviewed relative to that of the complete chicken genome model and a large body of zebra finch ESTs in the public database. The profile of more than 1000 interspersed repeats detected in non-avian reptile and neo-avian genomes is considerably more complex than would be expected from only chicken-mammal comparisons, with both ancient CR1-like elements and lineage-specific LINEs and MIR-like SINEs present that appear to be derived from a distinct clade of retrotransposons. In addition to providing a clearer picture of the complex evolutionary dynamics of CR1 LINEs in Reptilia, it is anticipated that results of these genome scans will yield a large number of previously unavailable powerful phylogenetic markers for advancing both avian and non-avian reptilian systematics.

Towards a unified nomenclature and classification of eukaryotic transposable elements.

Vladimir V. Kapitonov

Genetic Information Research Institute, Mountain View, CA

A standard nomenclature forms a vital core of any biological field dealing with numerous specific objects. The rapid accumulation of sequence data and its analysis by many different groups has led to the use of multiple names and classification systems, making it very difficult to follow the literature and to name newly characterized transposable elements (TEs). For instance, just one simple group of gypsy LTR retrotransposons is represented by hundreds different unrelated names reported in literature and GenBank. The unified nomenclature proposed here is basically implemented in Repbase and was developed in a last few years during large-scale identification and analysis of TEs in different genomes. This nomenclature relies on basic elements of our knowledge that would likely stay invariant in a future. Inevitably, this unified nomenclature depends on a standard classification of TEs. All known TEs belong to three classes: DNA transposons, LTR and non-LTR retrotransposons. In addition, some of these classes can be split further into a small number of subclasses. Despite an enormous diversity of TEs, each class/subclass is composed of a relatively small number of superfamilies characterized by unique names. For instance, a subclass of “cut and paste” DNA transposons is composed of only 12 superfamilies (hAT, Mariner, MuDR, P, En/Spm, piggyBac, Harbinger, Merlin, Transib, Novosib, Academ, and Rehavkus). Given that, every newly identified DNA transposon that belongs to one of these superfamilies should be named in accordance with a core-name, which is the corresponding superfamily name. Analogously, any non-LTR or LTR retrotransposon should also be named accordingly to a name of a clade/superfamily it belongs to. Based on this approach, new names should be introduced for novel superfamilies or subclasses only. Main problems and perspectives of the standard classification of autonomous and nonautonomous DNA transposons and retrotransposons will also be discussed.

Do transposable elements participate in combinatorial epigenetics?

Emile Zuckerkandl

Department of Biological Sciences, Stanford University, Stanford, California 94305; and Institute of Molecular Medical Sciences, P.O. Box 20452, Stanford, California 94309
e-mail: emilez@stanford.edu

At certain evolutionary junctures, two or more mutations participating in the build-up of a new complex function may be required to become available simultaneously in the same individuals. How could this happen in higher organisms whose populations are small compared to those of microbes, and in which chances of combined nearly simultaneous highly specific favorable mutations are correspondingly low? The question can in principle be answered for regulatory evolution, one of the basic processes of evolutionary change, through a resetting of transcription rates of genes. It is proposed that, in eukaryotes, changes in epigenetic trends could arise frequently enough so as to render probable particular conjunctions of changed transcription rates. Such conjunctions could involve mutational changes with low specificity requirements in gene-associated regions of non-protein-coding sequences. The effects of such mutations, notably when they involve histone methylation and acetylation, are expected to be among those that can migrate along chromatin and that are often cellularly inheritable over at least a limited number of generations of cells, and of individuals if the germ line is involved. SINEs and LINEs (Alu and L1 sequences in the case of humans}, which have been considered “junk DNA”, are among the repeat sequences that would appear liable to have repercussions on the function of a nearby promoter, through changes in their numbers and/or perhaps in their distribution. Also present may be preexisting unstably inheritable epigenetic trends leading to cellular diversity, trends that are endemic in a cell population on the basis of already established DNA sequences. Either way, the epigenetically conditioned regulatory trends may display only limited penetrance. The chances for two or more particular epigenetically determined regulatory trends to occur together in a cell are increased thanks to the proposed low specificity requirements for most of the sequence changes in intergenic and intronic DNA that have regulatory impact. Inheritable epigenetic changes (“epigenetic transmutations”) with effects at a distance would then provide time for “assimilation” of the several regulatory novelties through the occurrence and selection of specific classical mutations. These would have effects similar to the epigenetic effects, yet would provide stability and penetrance. The described epigenetic/genetic partnership might in fact be indispensable for certain complex new functions to arise. Thus, the presence of “junk DNA”, besides providing the substratum for zones of different higher-order structures of chromatin, may also be required for making possible the evolution specifically of the most complex organisms. The manifest flexibility organisms have had in building up highly complex new functions is no doubt mostly accounted for by other circumstances; yet, thanks to inheritable combinatorial epigenetic changes occurring at and acting over a distance in chromatin, organisms may, as an additional potentially critical resource, have the capacity of generating regulatory innovations that, although (accidentally) coadapted, can be simultaneous.

New Regulatory Regions of Drosophila 412 Retrotransposable Element Generated by Recombination

*Nathalie Mugnier, Christian Biémont, and Cristina Vieira**

UMR CNRS 5558. Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, 69622 Villeurbanne cedex, France
vieira@biomserv.univ-lyon1.fr

There are no doubts that transposable elements (TEs) have greatly influenced genomes evolution. They have, however, evolved in different ways throughout mammals, plants and invertebrates. In mammals they have been shown to be widely present but with low transposition activity; in plants they are responsible for large increase in genome size. In Drosophila, despite their low amount, transposition seems to be higher. Therefore, to understand how these elements have evolved in different genomes and what are the ways that host genomes have proposed to compel them, are major questions on genomes evolution. We analyzed sequences of the retrotransposable elements 412 in natural populations of the Drosophila simulans and D. melanogaster species that greatly differ in their amount of TEs. We identified new subfamilies of this element that were the result of mutation or insertion deletion process, but also of interfamily recombinations. These new elements were well conserved in the D. simulans natural populations. The new regulatory regions produced by recombination could give rise to new elements able to overcome host control of transposition and thus become potential genome invaders

The Drosophila Heterochromatin Genome Project (DHGP): Identifying Repeats & Using Comparative Sequence Analysis to Follow Heterochromatin Evolution

Christopher Smith, Mark Yandell, Robert Edgar, ShengQiang Shu and Gary Karpen*

Department of Genome Biology, Lawrence Berkeley National Lab, Berkeley CA, USA.

Heterochromatin makes up 30% or more of most metazoan genome sequence, and is comprised of satellite repeats, tandem duplications, and complex 'nests' of transposable elements (TEs). Despite its reputation as 'junk' DNA, heterochromatin also contains essential genes and functional elements that are required for genome stability. The DHGP has finished and annotated the ~14Mb of D. melanogaster heterochromatin and undertaken an extensive comparative analysis to 10 Dipteran genomes. A complete inventory of genes and repetitive sequence for newly sequenced genomes is essential for improving the quality of their annotations and interpreting large-scale datasets. Furthermore, comparative analysis of the genes and repeats between Dipterans can be used to study changes in heterochromatin genes and architecture during evolution.

Using PILER-DF, which finds TEs having 3 or more intact copies in the genome sequence, we have identified ~800 species-specific TEs in Dipterans, including ~100 novel elements. We used these species-specific libraries with the algorithm RepeatRunner to aggressively mask repeats in all 10 genomes. We then identified heterochromatin gene orthologs in the 10 species and compared their local repeat densities. Compared to their euchromatin counterparts, heterochromatin genes have longer introns and are often surrounded by repeat-rich sequence. By integrating results from orthology analysis, comparing conserved intron lengths, and analyzing the repeat content around gene orthologs, we've predicted regions that may have moved into and out of heterochromatin over evolution and are in the process of confirming our predictions with FISH experiments in other species. This method has proven useful to infer the genome location for scaffolds in newly sequenced genomes that have not been completely assembled and promises to elucidate the relationship between repeat content and genome architecture.

The combinatorics of helitron termini in A. thaliana genome revealed strongly structured superfamilies.

Sébastien Tempel[1,2]*,*, Jacques Nicolas*[1]*, Ivan Couée*[2] *and Abdelhak El Amrani*[2]

[1] IRISA-INRIA, Campus de Beaulieu Bât 12 and
[2] CNRS, Université de Rennes 1, UMR 6553 Ecobio, Campus de Beaulieu Bât 14A, 35042 Rennes cedex, France

Helitron is one of the most prolific families of transposable elements in A. thaliana genome. Although 37 families were identified, no classification is available.
Helicase, the protein which transposes helitron, recognizes helitrons extremities. We propose thus to formalize a model of helitron based on these two extremities, and sequence (or gap) of specified size separating them.
We assume the pattern of a terminus has length 36 nucleotides: a preliminary study showed this length to be a good compromise between specificity and sensitivity, given an error threshold of 25% with respect to Repbase's consensus [Jurka et al, 2005]. The consensus sequences of helitrons do not exceed 18000 bp for autonomous and 3000 bp for nonautonomous (AtREP). We chose a maximal gap of 20000 bp for the models of helitrons that contain at least one non AtREP terminus and a maximal gap of 3000 for others models.
Instead of developing a specific program, we used the SVG (String Variable Grammars) language which is able to model biological patterns such as repeats, palindromes or gaps [Searls, 2002]. STAN (Suffix Tree ANalyzer) [Nicolas et al. 2005] is a generic parser able to find all occurrences of SVG models in whole genomes. Note that this approach strongly differs from the usual \\\"pipeline\\\" approach that combines various tools to predict the presence of elements. Here, the model and the parser are clearly separated, allowing checking many models without changing algorithms. We run this way a systematic analysis of the combinatorics of 5' and 3' termini of each helitron in Arabidopsis thaliana. Denote I the set of 5' end of size 36 and J the set of 3' end of size 36 extracted from Repbase, grammars were written as follows for each possible pair of termini $(i,j) \in I \times J$:
i:9 errors - x(0,20000) - j:9 errors
or
i:9 errors - x(0,3000) - j:9 errors
A matrix of occurrences of these models has been produced on I x J. A cell (i,j) of this matrix corresponds to occurrences of the combination of the 5\\\' terminus i and the 3\\\' extremity j.
We have first discovered many chimera families of helitrons, where we called chimera a family composed of termini from two distinct known families [Jurka et al. 2001].
We have then discovered the matrix to be strongly structured. Using notions from concept lattice theory [Ganter, Wille1999], we have defined a robust characterization of superfamily of helitrons. A superfamily is a close set of families characterized by a dense number of occurrences for the subset of I x J, while having very few occurrences outside this subset.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 110:462-7.
Nicolas J, Durand P, Ranchy G, Tempel S, Valin AS. 2005. Suffix-tree analyser (STAN): looking for nucleotidic and peptidic patterns in chromosomes. Bioinformatics. 21:4408-10.
Kapitonov VV, Jurka J. 2001. Rolling-circle transposons in eukaryotes. PNAS. 98:8714-9.
Ganter B. and Wille R., 1999. Formal Concept Analysis. Mathematical Foundations, Springer, Berlin.

The tumor suppressor protein p53 binding sites in the human genome: How are they related to transposons?

*Sirotin M., Cui F. and Zhurkin V.B.**

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

The tumor suppressor protein p53 is an omnipresent transcription factor, modulating expression of thousands of human genes. In addition, p53 is involved in regulation of replication. Our genome contains ~2,000,000 potential p53 binding sites, ~50,000 of which have been detected experimentally. The relationship between the p53 sites and transposable elements has not been discussed in literature, however. Here, we report that more than 50% of these sites are apparently incorporated in transposons.

A typical p53 binding site comprises two decamers RRRCWWGYYY, separated by a spacer, varying from S=0 to 14 bp in the known functional sites. We demonstrate that the genome-wide distribution of the spacer lengths has strong peaks (e.g., for S=0 and 3 bp), exceeding the average background two-fold.

The sites with S=0 occur mostly in the vicinity of genes up-regulated by p53, while the sites with S=3 bp are found close to the down-regulated genes. This observation is important for understanding the genome evolution, because 60-70% of all the sites with S=0 and 3 bp are embedded in transposons (LTR10B, MER61D/E and MLT1H for S=0; LTR-14A, 45, 45B, 52 and Charlie6 for S=3 bp). We suggest a structural model for the p53-DNA interactions, accounting for the mentioned difference between the up- and down-regulated genes.

Our findings for the potential p53 sites are consistent with published data: in 25% of those 40 genes, whose p53 regulation is unambiguously established, the experimentally localized p53 sites are inserted in various repeats, such as Alu, ERV1, MIR. Among these genes are well-known BAX, CASP6, Cyclin G, 14-3-3 Sigma and Killer/DR5, critical for induction of apoptosis and cell cycle arrest.

Hairpin database: why and how?

Clark Jeffries

School of Pharmacy and Renaissance Computing Institute
University of North Carolina at Chapel Hill

About 326 known human microRNAs (miRNAs) now provide examples of RNA hairpins with roles in developmental processes. From bioinformatics and RNA chemistry, one can hypothesize that many other hairpins with double stranded RNA (dsRNA) stems must form routinely in the lives of cells, including hairpins from repeats and their nearby reverse complements. Such dsRNAs might provide additional feedstock for the miRNA pathway. This raises the possibility of a vast and potent system of gene expression control using \"repeat associated short interfering RNA\" molecules. Furthermore, the conventional miRNA pathway includes hybridization with sequences in the 3\' UTR of targeted mRNAs; a few lines of research consider 5\' UTR targets as well. But some genes are known to be controlled in part by intron hybridization with small RNAs. Thus vast is the number of theoretical hybridizations of 3\' UTR, 5\' UTR, and intron mRNA sequences with hairpins, especially hairpins with repeats in their stems. Needed are basic experiments that determine the extent and nature of general hairpin formation and gene mediation. Anticipating that many types of hairpin regulatory mechanisms will be discovered, this talk will provide a basic wish list of hairpin database search capabilities

POSTER ABSTRACTS

(1)

Distribution of LINEs and SINEs in the chicken genome

György Abrusán and Hans-Jürgen Krambeck*

Max Planck Institute of Limnology, Department of Ecophysiology
August Thienemann Str. 2 24306 Plön, Germany

In mammals LINEs and SINEs insert into gene poor regions of the genome, but show a gradual enrichment in GC (and gene) rich regions over time. Here we test whether in the more compact chicken genome, CR1 and MIR elements show a similar pattern. We show, that repeats on the chicken autosomes behave qualitatively similarly to mammalian autosomes; insert into gene poor and accumulate in gene rich regions, but unlike in mammals there is no accumulation on the Z chromosome. In contrast to human Alus, MIRs are most abundant in regions of intermediate GC content. Additionally, we show that similarly to mammalian L1 elements, CR1 insertions are longest on the Z and W chromosome, even after standardization with their local GC content.

(2)

First analysis of the presence of transposable elements in Bos taurus coding genes

Almeida, LM[*1], Silva, IT[2], Silva, WA[2], Carareto, CM[1], Amaral, MEJ[1]*

[1] UNESP – São Paulo State University, Depto. Biologia, , IBILCE, São Jose Rio Preto, SP, Brazil
[2]Faculty of Medicine Ribeirão Preto, University of São Paulo, Brazil
almeidalm@hotmail.com

Currently, with the bovine genome sequence available, the scientific community needs to determine the location, structure, function and expression of genes affecting health, reproduction, production and product quality in cattle. Up to this date 23,000 genes have been identified and characterized in the Bos taurus genome. The activity of transposable elements (TEs) can create a diverse set of genomic changes, providing an enormous source of variability that can be used to create novel genes or modify genetic functions. In an effort to verify the possible contribution of TEs to the bovine genome evolution, we searched the abundance, distribution and insertion orientation of TEs in exons, introns and Untranslated Leader Regions (ULR) in a set of 18 bovine autosomes using the RepeatMasker program. From the 8,579 genes analyzed, 7,263 (84.66%) presented insertions of TE, with an average of 18.80 insertion/gene. From the 136,532 TEs detected 130,132 (95.31%) were classified as retrotransposons without LTRs, being 92,253 (67.57%) short interspersed sequences (SINEs) and 37,879 (27.75%) long interspersed sequences (LINEs), 4,073 (2.98%) as DNA transposon and 2,327 (1.70%) as long terminal repeats retrotransposons (LTR-elements). Sense (49.5%) and antisense (50.5%) orientation frequencies were approximately equal. As expected, the insertion frequencies in exons and ULR regions were very low, 166 (0.12%) and 857 (0.63%) respectively, while intron insertions (135,509 - 99.25%) were more abundant. Our analyses will be extended to all chromosomes, but these preliminary results already show a high number of TEs within bovine coding genes, which may contribute to select candidate TE insertions related with different patterns of gene expression.

Financial support: FAPESP

(3)

Sequence heterogeneity and phylogenetic relationships between the copia retrotransposon in Drosophila species of the repleta and melanogaster groups.

*Luciane M. de Almeida and Claudia M.A. Carareto**

Universidade Estadual Paulista (UNESP), Departamento de Biologia, 15054-000 São José do Rio Preto, SP, Brazil, e-mail: carareto@ibilce.unesp.br.

The retrotransposon copia has been studied in the melanogaster group of Drosophila species, however, very little is known about copia dynamism and evolution in other species of the genus Drosophila. Aiming to broaden our knowledge about the evolutionary history of this element, we analyzed the distribution, the number of insertions in the genome, the 5´LTR-ULR sequence heterogeneity and phylogenetic relationships in 24 species of the Drosophila group repleta. PCR and Southern blot analyses showed that copia occurs in 18 of the 24 species evaluated, with a low number of insertions into the genomes. Sequencing was possible in only eight species. The sequences showed low nucleotide diversity, which suggests selective constraints maintaining this regulatory region over evolutionary time. In contrast, inter-group analyses showed a very high nucleotide divergence between repleta copia sequences, and other copia sequences described in the group melanogaster, D. willistoni and Zaprionus tuberculatus. On the other hand, a very low nucleotide divergence and a closer phylogenetic relationships between D. willistoni / Zaprionus tuberculatus / melanogaster species subgroup suggest the occurrence of copia horizontal transfers between these species. Sixteen transcription factor binding sites were identified in the LTR-ULR repleta and melanogaster consensus sequences, and a similar repetition number of motifs was found between the repleta and melanogaster copia families. However, these motifs are not homologous, neither according to their position in the LTR-ULR sequences, nor according to their sequences. Taken together, the low motif homologies, the phylogenetic relationship and the high nucleotide divergence between the melanogaster and repleta copia sequences reinforce the occurrence of two copia families.

Financial Support: FAPESP

(4)

Transposon display supports transpositional activity of P element in species of the saltans group of Drosophila

*N. Setta[1], A.P.P. Costa[2], F.R. Lopes[1], M.A. Van Sluys[2], C.M.A. Carareto[1]**

[1] Departamento de Biologia, Universidade Estadual Paulista – UNESP, São José do Rio Preto, SP, Brazil
[2] Departamento de Botânica, Universidade de São Paulo – USP, São Paulo, SP, Brazil. e-mail: carareto@ibilce.unesp.br.

Transposon Display was used to evaluate the number of genomic insertions and to infer the capacity of P element (canonical and O-type subfamilies) mobilization in D. sturtevanti (nine strains) and D. saltans (five strains). Pairwise distances between strains regarding to the genomic localization of the insertions were estimated using a 0-1 matrix derived of the pattern of insertions of both subfamilies in each strain. The distances were used to infer mobilization. Copy numbers were highly variable (D. sturtevanti, canonical: 20.11 ± 7.56, O-type: 9.00 ± 6.10; D. saltans, canonical: 16.4 ± 4.28, O-type: 12.60 ± 9.50 insertions in average). The larger values obtained by Transposon Display compared with our previous data by Southern blot using the same strains indicates the superiority the first method over the second for estimating TE copy numbers. The higher copy numbers of the canonical P element in D. sturtevanti and the larger divergence regarding to its genomic locations (23%) when compared to the O-type (14%) in this species as well as the larger divergence regarding to the insertion sites of canonical P element between the strains of D. sturtevanti (25%) than of D. saltans (18%) indicate that the canonical element occupy more differentiated sites in the genome of D. sturtevanti. These data corroborate the hypothesis that the O-type is the oldest subfamily in the saltans group and suggest that the canonical P element is or has been transpositionally active until more recently in the D. sturtevanti genome.

Financial Support: FAPESP

(5)

SURE, a Ty1/copia-like retrotransposon of sugarcane: genomic distribution and expression analysis.

Domingues, DS[1], Jesus, EM[1], Rossi, M[1], Costa, APP[1] , Van Sluys, MA[1] **

[1] GaTE lab (Genomic and Transposable Elements Laboratory), Depto. de Botanica, Instituto de Biociencias, Universidade de Sao Paulo, Brazil.
*corresponding author: doug@ib.usp.br

A transcriptionally active retrotransposon - named SURE, for SUgarcane REtrotransposon - was recently isolated from sugarcane genome using a PCR-based reconstruction based on available data from Sugarcane EST Project. Comparison with other fully-characterized elements revealed that this element has extensive similarity with Tnt1, indicating that SURE is a Ty1/copia-like retrotransposon. Genomic and transcriptional aspects of SURE were analyzed using different strategies. The integrity and copy number of SURE were studied through screening a sugarcane BAC library using probes derived from different coding domains and LTR of SURE. We also evaluated the potential of SURE as a molecular marker by sequence-specific amplified polymorphism (S-SAP). Transcriptional analyses were made by RT-PCR and quantitative real-time RT-PCR. Our results suggest that SURE is a low-copy number retroelement, and most of its copies seems to be complete. SURE can be also used as a molecular marker, as we identified polymorphism among Brazilian cultivars of sugarcane using PCR primers derived from SURE LTR. Transcriptional analyses detected that SURE is expressed in roots, leaves and calli – tissues with no EST related to SURE in Sugarcane EST database. All tissues analyzed displayed a low level of transcription when compared to internal controls. More detailed genomic and transcriptional features of SURE are still needed to be investigated, but our data pointed that SURE has a potential use in sugarcane biotechnology. Financial support: FAPESP and CNPq (Brazil).

(6)

Retrosol, a Tnt1-like retrotransposon: identification and clustering analysis

Manetti, ME; Costa, APP; Rossi, M; Van Sluys, MA*

GaTE Laboratory, Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, Brazil

Tnt1 was the first active plant retrotransposon described from tobacco, and its superfamily comprises elements from Nicotiana (Tnt1) and Lycopersicon (Retrolyc1) species. This study was conducted to characterize Tnt1-related sequences in 21 wild species of Solanum and five cultivars of Solanum tuberosum, representing the Germoplasm collection from INTA-Balcarce, Argentina. Total genomic DNA was prepared from in vivo germinated tubercles or in vitro grown seeds and partial RT coding domain, linker and 5\'portion of the 3\'LTR was amplified by PCR. One hundred and ninety seven distinct sequences among the species were aligned (ClustalX) and grouped using distance method (PAUP). Multiple alignments characterized three groups which differ mainly in the U3 region. These sequences homologous to Tnt1 in their RT domain but differing in the U3 region were named as Retrosol. Phylogenic relationships of the RT coding domain and the U3 promoter region were examined separately using a median-joining network approach. For these analysis 197 sequences reported in this study, 140 Tnt1 sequences from GenBank and 107 Retrolyc1 sequences previously described by the group (Araujo et al., 2001) were used. The resulting RT network showed that sequences group according their genus. Inside each genus, 20% of the sequences are identical indicating being part of the active copy. In contrast, U3 network showed a higher level of diversity whereby an active U3 copy could not be identified. Genomic distribution of the three groups among Solanum was analyzed by Southern blot characterizing an inter-specific polymorphism for all groups. The results presented here suggest that Tnt1-superfamily was present early in the evolution of Solanaceae. Moreover, RT region evolved inside each genus being possible to identify the active sequence, while the highly diverse U3 region presents inter-specific differentiation.

(7)

Retrolyc1 activity revealed by expression analysis and recent polymorphic insertion sites within Lycopersicon genome

Ana Paula Pimentel Costa[*1,2], Regina Y. Hashimoto-Miura[1], Juliane K. Ishida[1] and Marie-Anne Van Sluys[1]*

[1] GaTE Laboratory, Departamento de Botânica, IB-USP, Rua do Matão 277, São Paulo, CEP 055080-090, SP, Brazil
[2] FCBEE, Universidade Presbiteriana Mackenzie Rua da Consolação, 930 - CEP 01302-907, São Paulo – SP, Brazil

Retrolyc1 was originally isolated from the L. peruvianum genome and was shown to be present in other Lycopersicon species. Retrolyc1 displays a Ty1/copia basic structural organization and extensive homology with the tobacco Tnt1 retrotransposon except in the LTR U3 region, located upstream from the transcriptional start. In addition, Retrolyc1 populations are composed of two subfamilies, Retrolyc1A and Retrolyc1B, based on the U3 region sequence. The U3 region directs the expression of the element and is controlled by signals dependent on the host organism and environmental factors. Analysis of Retrolyc1 distribution in different accessions of all 9 Lycopersicon species and expression profile is studied. Copy number and subfamilies relative abundance vary between accessions within a given species. These results characterize a particular amplification history for each subfamily particular to each accession. Retrolyc1-A transcriptional activity was evaluated through transient and permanent expression assays using U3Apromoter:GUS fusions, RT-PCR and Northern blot hybridization. Results from this different experiments support that wounding modulates Retrolyc1A expression. Altogether our results demonstrate that Retrolyc1 is an actively expressed element in both wild and domesticated species that may have contributed to Lycopersicon diversification.

(8)

hAT-LIKE TRANSPOSASES TRANSCRIPTIONALLY ACTIVE IN SUGARCANE GENOME

Jesus, E.M., Rossi, M., Van Sluys, M.A.*[1]

[1] GaTE – Genomes and Transposable Elements Laboratory – Departamento de Botânica – Instituto de Biociências – Universidade de São Paulo – Brasil

Corresponding author – erika_jesus@yahoo.com.br
Key words – sugarcane, transposases, hAT superfamily, expression pattern

hAT superfamily of transposases comprises the first transposable element described in living organisms through the pioneer work of Barbara McClintock in 1946. Sixty years later, genome sequencing projects uncover the presence of large repetitive families in most eukaryotic cells. Through data-mining search we have previously identified twenty one clones with significant similarity to hAT superfamily transposons in the sugarcane EST sequencing project (SUCEST) (Araujo and Rossi et al, 2005; Plant Journal 44:707-717). After full-length cDNA sequencing and clustering analysis of two conserved domains, characteristic of hAT superfamily, two groups were revealed. Southern Blot hybridization assays confirmed different patterns of genomic distribution for the two groups. While group one is present in the genome of all grasses analyzed (sugarcane hybrids and parentals, maize and rice), group two is present only in Saccharum, and is composed of more heterogeneous sequences, that presents similarity with the original elements that constitutes the hAT superfamily: hobo, Ac and Tam3. In this work, we report a comparative analysis of the expression of tree clones chosen among the twenty one hAT homologues (TE074,TE191 and TE257). RT-PCR and quantitative RT-PCR assays were carried using total RNA extracted from leaves and roots of the hybrid sugarcane variety SP-80-3280. RT-PCR assays demonstrated that both groups are transcriptionally active in the tissues analysed. Quantitative RT-PCR assays results point out a higher expression of TE074 in leaves when compared to the other two clones suggesting that the genomic sequences may have distinct expression profiles. Cloning of the terminal inverted repeats (TIRs) are under development to identify the 5' and 3' non-transcribed regions in order to evaluate the promoter differences among these clones.

Financial support – FAPESP and CNPq (Brazil)

(9)

GENOMIC CHARACTERIZATION OF Mutator TRANSPOSONS IN SUGARCANE

Rossi M., Saccaro-Junior N.L., Nakabashi Myna and Van Sluys M.A.*

Departamento de Botânica, Instituto de Biociências, Universidade de São Paulo, Brazil

The maize Mutator (Mu) system has been described as the most active and mutagenic plant transposon so far discovered. In sugarcane, comparative analyses of Mu transposases suggest the existence of four classes of Mu systems (1), which were present in angiosperms prior to the divergence of monocots and eudicots. Sequence studies revealed Class IV and I as the least and the most diverse, respectively. This observation negatively correlates with the number of putative active transposases found for each Class. In order to evaluate the impact of Mu system in the sugarcane hybrid genome, we decided to screen a genomic BAC library from R570 cultivar, and get the genomic full-length sequence of some representatives for each Class. The library screening displayed that copy number greatly differs between classes probably reflecting a difference in transpositional activity. Twelve clones for Class I, II and IV, and the six clones identified for Class III were selected for further studies. Comparative fingerprinting analyses based on restriction fragment polymorphism pattern between BAC clones and the parental species (Saccharum officinarum and Saccharum spontaneum) enabled the identification of the Mu copy origin in the hybrid cultivar genome. Two BAC clones of each Class were selected for sequencing, one from S. officinarum and one from S. spontaneum. Shotgun sequencing is currently being performed. Genomic sequences will allow us to characterize sugarcane Mu system structurally and functionally and, ultimately, add arguments to the understanding of transposable element role in the functional and evolutionary dynamics of plant genomes.

1) Rossi M. et al. (2004) Comparative analysis of sugarcane Mutator-like transposases. *Molecular Genetics and Genomics* 272:194-203.

(10)

A model of Segmental Duplications formation in Drosophila genome: the implication of transposable elements

Fiston A-S[1], Anxolabehere D[2], Quesneville H[1]*

[1] Laboratory Bioinformatics and Genomics, Institut Jacques Monod, Paris, France
[2] Laboratory Genome Dynamics and Evolution, Institut Jacques Monod, Paris, France.

Segmental duplication (SD) is one of the most important genome dynamics actor. Using a computational pipeline, we have detected SDs in the Drosophila melanogaster genome. They represent 2.8% of the genome, with lengths ranging from 346bp to 139kbp. Our analysis indicates an enrichment in transposable elements (TEs) and a high density in pericentromeric and subtelomeric regions. We propose here a model for SD formation based on the \"Synthesis-Dependent-Strand-Annealing\" model (SDSA; Nassif N. et al., Mol. Cell. Biol. 14:1613-1625, 1994): a double-strand breaks (DSB) homologous repair model. Our model that we call: \"Duplication-Dependent-Strand-Annealing\" (DDSA) predicts, after a DSB, the search for an ectopic homologous region to initiate the repair. A repeat near one of the single-strand tails generated by the DSB could find a homologous repeat at an ectopic site. The sequence contiguous to the ectopic repeat is then used as a template for the synthesis. At the end, the template is duplicated in the gap created by the DSB. According to our model, we expect to find repeat regions at the SD ends. Indeed, SD ends are enriched in TEs compared to random sequences located in the same genomic environment. Moreover, DDSA as for SDSA model also predicts an unstable synthesis and we propose that dissociations of the strand during synthesis may occur. If re-annealing follows the dissociation, the synthesis could be continued. The marks of this process that we have observed suggest a containment of the dissociated strand in the repair complex. Our study suggests that TEs are involved in SD formation and thus contribute actively to Drosophila genome architecture.

(11)

AeBuster, a new putatively mobile transposable element from Aedes aegypti

Peter Arensburger[1], David A. O'Brochta[2], and Peter W. Atkinson[1]*

[1] Department of Entomology, University of California,, Riverside, CA 92521-0001
[2] Center for Biosystems Research, University of Maryland Biotechnology Institute, College Park, MD 20742

The Aedes aegypti genome was searched in silico for transposable elements belonging to the hAT superfamilly. We report the discovery of AeBuster, a putatively functional transposable element in the Ae. aegypti genome. The sequence of AeBuster contains several conserved hAT element motifs, nearly perfect inverted repeats and possible transposase binding sites. Four copies of AeBuster were found with nearly identical nucleotide sequences and flanked by different 8 bp. target site duplications, characteristic of hAT element transposition. The transposase sequence of AeBuster appeared to have similarities to sequences from higher organisms. We discuss the nature of these similarities and the importance of AeBuster in the larger context of hAT elements.

(12)

The Regulation of Insect hAT Elements

Peter W. Atkinson[1], Lisa M. Friedli[1], Ala Perumalsamy[1], Thomas A. Laver[1], Robert H. Hice[1], Stephanie A. Russell[1] & David A. O'Brochta[2]

[1] Department of Entomology, Univeristy of California, Riverside, CA 92521
[2] Center for Biosystems Research, University of Maryland Biotechnology Institute, College Park, MD 20742

Members of the hAT superfamily of transposons are found in plants, fungi and animals and are the most abundant class II transposons found in the human genome. Little is known about how the animal members of this superfamily are regulated. We study several of the insect hAT elements including the hobo element from Drosophila, the Hermes element from housefly, and the Herves element from anopheline mosquitoes and are particularly interested in how these elements function when placed into new host species. We will present data summarizing our findings on the role that host factors, methylation and transcriptional control may play in the regulation of these elements both in vitro and in vivo.

(13)

Detection and characterisation of HERV LTRs using Hidden Markov Models and artificial Neural Networks

Farid Benachenhou[1], Patric Jern[1], Göran Sperber[2], Panu Somervuo[3], Merja Oja[3], Samuel Kaski[3] and Jonas Blomberg[1]*

[1] Section of Virology, dpt of Medical Sciences, and
[2] Section of Physiology, dpt of Neuroscience, University of Uppsala, Sweden. 3Department of Computer Science, University of Helsinki, Finland

Retroviral LTRs are important but dynamic structures which have few absolute landmarks useful for detection. They are easiest recognised because of repetitiveness and pairwise occurrence in complete proviruses. However, many exist mainly as single LTRs. Recognition of low copy number single LTRs in large genomic sequence collections is a bioinformatic challenge. We have used a collection of known exo- and endogenous retroviral LTRs to train artificial neural networks (ANNs) and Hidden Markov Models (HMMs) for recognition of single LTRs of beta- and gammaretroviruslike HERVs. In both methods, the R-U5 portion stood out as being most conserved, and most suitable for detection. In Viterbi alignments the polyadenylation signal, and the surrounding gt-rich area were the most conserved. Transcriptional start, inr, and several transcription factor binding sites were less clearly conserved, but were sometimes observed in the alignments. A U5 ANN was relatively fast (around 200 seconds per million base pairs of human genomic sequence [mbp]) and had a relatively high signal-to-noise ratio. It could be used as an adjunct in LTR definition in our retroviral sequence detection system, RetroTector©. HMMs of beta- (1 group) and gammaretrovirus-like (4 groups) LTRs were deduced. HMMs detected 80-90% of corresponding LTRs detected by RepeatMasker (RM) in human chromosome 19. The HMMs also detected around 1% of sequences not detected by RM. Many of the latter probably are relevant LTRs, but a false positive background due to simple repeats was sometimes observed. The HMMs tended to be relatively slow (3600 seconds per mbp). An HMM run of chromosome 19 took 2-3 days on a standard laptop. Although also U3 contained conserved areas, they did not contribute much to the detection, and slowed down the detection process.
We conclude that HMMs can be used for retroviral LTR detection and structural characterisation, but that they are computationally demanding.

(14)

Re-defining transposable element reference sequences for genome annotation.

Nicolas Buisine[*1], *Hadi Quesneville*[2] *and Vincent Colot*[1]

[1] Unité de Recherche en Génomique Végétale, 2 Rue Gaston Crémieux, 91057 Evry Cedex, France
[2] Institut Jacques Monod, 2 place Jussieu, 75251 Paris Cedex 05, FRANCE

The most accurate and sensitive methods to identify transposable elements (TEs) within genome sequences rely on similarity-based comparisons using sets of reference sequences established from already known TEs. This approach, however, is sensitive to the quality of the reference sequences, which are often built by different people, each making different implicit assumptions. For instance, one might want to optimise the TE coding potential, thus introducing a bias towards recently active elements and making divergent copies more difficult to detect. In other cases, reference sequences have been built with the aim of best representing the exhaustive sequence repertoire of a TE family, by constructing a chimera assembled from the longest genomic copies. This introduces large insertions and mostly ignores the coding potential, thus making functional copies more difficult to detect.

To address precisely how this “reference sequence syndrome“ affects TE detection, we have generated three distinct sets of reference sequences that optimise the coding potential the length, or a combination of both, respectively. As a test case, we have used the well-characterized Arabidopsis thaliana TEs to build these three sets. Using a novel TE annotation pipeline (1), we have found that these three sets are complementary to RepbaseUpdate (RU), which is mostly build from consensus sequences. Altogether, our three new sets allow the detection of 10-20% more TEs than when using RU alone.

(1) Quesneville et al., 2005. *Plos Comp. Biol*., 1:166-175.

(15)

The role of RNAi in the silencing of mammalian repetitive elements

J. Mauro Calabrese[1], Amy C. Seila[2], Phillip A. Sharp[1,2]

[1] Department of Biology and [2] Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02144

Genetic studies in Schizosaccharomyces pombe, Caenorhabditis elegans, Drosophila melanogaster, and Arabidopsis thaliana implicate the RNAi pathway in the establishment of heterochromatin at repetitive genomic elements, such as centromeric repeats and retrotransposons. Given its conserved role in these diverse organisms, we hypothesize that RNAi functions to establish heterochromatin at repetitive loci within the mammalian genome. Epigenetic reprogramming of the genome, during which DNA and chromatin modifications in the developing embryo are removed and then re-established, occurs early in mouse development. During this reprogramming process, many of the repetitive regions of the genome are transcriptionally silenced. It is not understood how these regions are targeted for silencing, but precedent from other organisms suggests that RNAi may be involved in this process.
Mouse embryonic stem cells are derived from the inner cell mass (ICM) of the blastocyst at the onset of the epigenetic reprogramming that occurs during early development, suggesting these cells may posses the machinery needed for RNAi-mediated transcriptional silencing. Using a conditional deletion strategy we have isolated ES cells that lack Dicer, the key processing enzyme of RNAi, and are using these cells as a platform to explore the role of RNAi in the silencing of repetitive elements.

(16)

In silico evolutionary analysis of the Tc1-like family of DNA transposons

Claudio Casola

Laboratorio di Biologia Cellulare e dello Sviluppo, Dipartimento di Fisiologia e Biochimica, UniversitÃ di Pisa, Italy

The IS630/Tc1/mariner (ITm) is a widespread superfamily of Class II transposable elements members of which share the TA dinucleotide target duplicated site, and encode a single protein with the typical DDE/D catalytic triad of transposases and integrases. Tc1-like elements (TLEs), initially discovered in the Caenorhabditis elegans genome, form one of the major family within the ITm group, and are characterized by the transposase DD34E catalytic signature. Different aspects regarding the biochemistry of TLE transposase and impact of these DNA transposons, especially on invertebrate genomes, have been investigated in the last twenty years. However, the evolutionary dynamics and the distribution of the Tc1 family is far from being completely depicted, especially when compared to the cognate mariner transposons. In the present study, several thousands TLE sequences annotated or simply deposited in databanks were retrieved to perform an in silico evolutionary analysis of the Tc1 family. On the basis of DNA and aminoacidic similarity and reconstructed phylogenetic trees, collected TLEs were classified in eighteen clades (subfamilies). The highest TLE copy and subfamily number was observed in Anamnia vertebrates, although they were underrepresented in the pufferfish compact genomes compared to zebrafish and the amphibian Xenopus tropicalis, wherein putative active copies were detected Horizontal transmission of Tc1-transposons seemed to be occurred repeatedly in vertebrates, and it was documented also between salmonids and the Platyhelminthes human parasite Schistosoma japonicum, as already reported for other salmonid genomic sequences. The maximum-likelihood analysis of one TLE subfamily evolutionary dynamics in four vertebrate species suggest that these elements evolve neutrally within one host genome whereas horizontally transmitted copies are under purifying selection, similarly to other DNA transposons.

(17)

Retroposition mechanism for gene amplification in 5S rRNA genes

Anat Caspi[1] and Lior Pachter[2]*

[1] UCSF/UCBerkeley Bioengineering
[2] Department of Mathematics, UC Berkeley.

The eukaryotic ribosomal components are encoded by multicopy nuclear ribosomal RNA (rRNA) genes: 28/26S, 18S, 5.8S, and 5S. Of these, the 5S rRNA genes are uniquely transcribed by a different polymerase than the others (RNA-polymerase III), and are dispersed throughout the genome in separate large tandem gene arrays. It has been previously conjectured that 5S genes are capable of retrotransposition[2,3]. In this study we find strong evidence that tandem repeats of 5S rRNA multigene families retrotransposed and that dispersal of these multigene families occured on both ancestral branches (hence supporting interspecies clustering of some 5S sRNA multigene families) and on lineage-specific branches (giving rise to some genome-specific 5S rRNA multigene families).

With the availability of sequence data of closely related genomes, we are now able to identify the phylogenetic branch on which a transposition occurred in the genome, thereby defining temporal bounds on the time of insertion[1]. We examined the temporal pattern of insertions of 5S rRNA gene sequences identified systematically in the full sequences of four closely related Drosophila genomes. Analysis of these sequences revealed that
(i) 5S gene types that are shared among species correspond to insertions that occured before their lineages diverged
(ii) 5S gene types that are genome-specific correspond to lineage-specific insertions
(iii) some 5S rRNA pseudogenes result from ancient insertions that did not fix in the ancient branch.
This evidence supports the hypothesis that a retroposition mechanism is responsible for 5S rRNA gene dispersal and integration into the genome.

[1] Caspi, A and Pachter, L (2006) Genome Res, in print.
[2] Drouin, G and deSa, MM (1995) Mol. Biol. Evol. 12, 481-493.
[3] Rooney, AP and Ward, TJ (2005) PNAS 102:14, 5084-5089.

(18)

Repeat-Induced Point Mutation (RIP) of transposable elements in the fungus Aspergillus nidulans

John Clutterbuck[1], Jerzy Jurka[2] and Vladimir Kapitonov[2]*

[1] Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G11 6NU, Scotland, UK
[2] Genetic Information Research Institute, 1925 Landings Drive, Mountain View, California 94043, USA.

Repeat-Induced Point Mutation was first observed in Neurospora crassa, where it caused numerous C-T transitions in virtually all repeated sequences, and is thus an effective inactivator of transposable elements in this fungus. While there is as yet no experimental evidence of RIP in Aspergillus nidulans, its genome includes representatives of at least 21 transposable element families, many of which show both intact and RIP-affected members. It therefore appears that RIP is sporadic in this species. Analysis of the distribution of RIP-affected members of some families of elements suggests that RIP most commonly occurs near telomeres, but this is not universal. Many affected elements are both clustered and fragmented, and this may part be due to the choice of RIP-affected elements as sites for further insertion. Fragmentation of clustered elements may then be due to faulty transposition involving multiple elements.

(19)

Genomic distribution of recently integrated human Alu retrotransposons

Richard Cordaux, Jungnam Lee, Liv Dinoso and Mark A. Batzer*

Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA

Alu elements represent the largest family of human mobile elements in copy number. They have considerably impacted the architecture of the human genome and their current expansion is responsible for various genetic disorders. A controversial issue with implications for both Alu biology and human genome evolution is whether selective pressures are affecting Alu elements on a large scale. To address this issue, we analyzed the genomic distribution of the three youngest known human Alu subfamilies (Ya5a2, Ya8 and Yb9) in conjunction with their insertion polymorphism status in the human population, since selection can only act on polymorphic elements. We find that: (i) the genomic distribution of polymorphic and fixed recently integrated Alu elements with respect to GC content are statistically indistinguishable, and (ii) recently integrated Alu elements are inserted randomly with respect to GC content of the surrounding genomic DNA. These results provide strong evidence that recently integrated Alu elements are not subject to positive or negative selection on a large scale. Therefore, young Alu elements can be regarded as essentially neutral residents of the human genome. These results also imply that selective processes specifically targeting Alu elements can be ruled out as explanations for the accumulation of Alu elements in GC-rich regions of the human genome.

(20)

Alu-Associated Enhancement of Single Nucleotide Polymorphisms in the Human Genome

*Siu-Kin Ng and Hong Xue**

Department of Biochemistry and Applied Genomics Laboratory, Hong Kong University of Science & Technology, Clear Water Bay, Hong Kong, China

Identifying features shaping the architecture of sequence variations is important for understanding genome evolution and mapping disease loci. In this study, high-resolution scanning of Alu-centered alignments of the human genome sequences has revealed a striking elevation of the frequency of single nucleotide polymorphisms (SNP) in the body and tail of Alu sequences compared to flanking regions. This enhancement in SNP density is evident for all twenty-four chromosomes, and in both the Alu-body and Alu-tail, which together may be referred to as the Alu-SNPs. Reduced levels of Alu-SNPs in the sex chromosomes, especially in the non-recombining NRY region of the Y chromosome, are consistent with recombination events playing an important role in the enhancement. The Alu elements are unstable recombination-mutation hotspots in the human genome, and it is suggested that the Alu-SNPs represent a key manifestation of this instability. Variations in Alu-SNPs among the HapMap populations of northern and western European ancestry (CEU), Han Chinese from Beijing (CHB), Japanese from Tokyo (JPT), and Yoruba from Ibadan, Nigeria (YRI) indicate that the Alu-SNPs provide useful sequence markers, in addition to the Alu-insertion polymorphisms themselves, for the delineation of human genome evolution. That Alu-SNP levels are highest in the youngest Alu-Y, intermediate in the Alu-S of intermediate age, and lowest in the oldest Alu-J is consistent with the occurrence of not only genetic drift but also natural selection on the Alu-SNPs. Such evolutionary selection in turn suggests that Alu-SNPs might include potential sites of disease association, and therefore deserve detailed investigation.

(21)

Transposable elements explain differences and similarities in the correlation structure of Eukaryotic genomes

Manuel Dehnert[*1], Heike Hameister[1], Werner E. Helm[2], Marc-Thorsten Huett[1]*

[1] Bioinformatics Group, Darmstadt University of Technology, Darmstadt, Germany
[2] Mathematics and Science Faculty, University of Applied Sciences, Darmstadt, Germany

Attempts to characterize a species on the basis of statistical properties of its DNA sequence have been formulated for several decades. The most prominent of such genome signatures rely on neighborhood correlations (e.g., dinucleotide frequencies) and, consequently, attribute species identification to mechanisms operating on the dinucleotide level (e.g., neighbor-dependent mutations).

In two recent studies [1,2] we investigated longer-range correlations (up to distances of a few tens of nucleotides) in DNA sequences with methods from information theory. We find that these correlation profiles, when analyzed for a variety of eukaryotic species, display a high degree of intra-species similarity and systematic inter-species differences. Remarkably, these inter-species differences increase with evolutionary distance, i.e. a cluster tree based upon distances of the correlation profile sorts all chromosomes involved into species clusters and approximates the corresponding phylogenetic tree of these species [1]. When comparing closely related species we furthermore find that species distinction on the basis of correlation profiles increases with an increased range of correlations taken into account (up to a few hundreds of nucleotides) [3].

Here we show that the patterning of eukaryotic genomes by repetitive elements explains most (but not all) of these statistical features. We use Repbase data to mask different classes of repetitive DNA in chromosome sequences of different Eukaryotes and observe, how the resulting correlation patterns change. We relate these changes with known differences in properties of mobile genomic elements between the species. Understanding these statistical features as signatures of repetitive elements provides a link to processes of genome evolution, particularly retrotransposition, which pattern the genome on an evolutionary time scale.

By linking the elementary process-oriented properties with genome-wide statistical patterns, this view of the correlation profile as a process signature points towards a system biology treatment of genome evolution.

[1] Dehnert M., Plaumann R., Helm W.E. and HŸtt M.-Th. (2005) Genome phylogeny based on short-range correlations in DNA sequences, *Journal of Computational Biology*, 12, 545-553.
[2] Dehnert, M., Helm, W. E. and HŸtt, M.-Th. (2005) Information theory reveals large-scale synchronisation of statistical correlations in Eukaryote genomes, *Gene* 345, 81-90.
[3]Dehnert, M., Helm, W. E. and HŸtt, M.-Th. (2006) The informational structure of two closely related eukaryotic genomes, *Physical Review E*, submitted.

(22)

The overproduction of SINE RNA is associated to severe developmental defects in Arabidopsis thaliana.

*T. Pélissier[1], M.N. Pélissier[1], T. Elmayan[2], H. Vaucheret[2], C. Bousquet-Antonelli[1] and J.M. Deragon[1]**

[1]CNRS UMR6547–Université Blaise Pascal, BIOMOVE, 24 Av des Landais, 63177 Aubière Cedex France
[2]INRA Laboratoire de Biologie Cellulaire, Route de St-Cyr, 78026 Versailles Cedex France

In mammals and insects, the induction of different cellular stresses (such as heat shock or viral infection) leads to a specific increase in SINE RNAs. In mammals, SINE RNAs have been shown to regulate translation by interacting with the PKR kinase (an enzyme that phosphorylates EIF2α and downregulates translation) or to regulate transcription by binding to the RBP1 subunit of the RNA polymerase II complex. In this context, SINE can be considered as \"stress responsive-genes\" capable of modulating the level of mRNAs and proteins in different stress situations. Nothing is known on the impact of SINE RNAs in plants. We observed that transgenic Arabidopsis thaliana lines overproducing a founder Brassica SINE RNA (under the control of its natural enhancers or enhancers from the 7SL locus) present severe developmental defects including delayed growth and flowering time, abortive siliques, partial sterility, reduction of leaf size, leaf serration associated with a downward curvature, and partial lost of apical dominance. These phenotypes are similar to the ones observed for some arabidopsis mutants impaired in miRNA/tasiRNA production suggesting a cross talk between SINE and miRNA/tasiRNA pathways. In support of this hypothesis, we observed that SINE RNA could bind specifically to HYL1, a dsRBM protein involved in binding miRNA precursors and producing mature mi/tasiRNAs. Also, the SINE-induced phenotype is reverse to wild type in an rdr6 background, a mutation affecting tasiRNA formation. Although these results do not exclude a role for plant SINE RNAs in stress response, they suggest that SINE RNA could play a role in regulating arabidopsis development by having an impact on the miRNA/tasiRNA pathway.

(23)

Mammalian RNAi directed transposon silencing?

Louise Docherty

University of Glasgow, Scotland, UK

It is generally accepted that transposon DNA is found in a silent state in somatic cells, however the initiation of this silencing is poorly understood. RNA targeted methylation and silencing have been reported in several organisms along with the identification of dsRNA and siRNA homologous to transposon sequences. Mutational studies have also linked loss of RNAi function to transposon reactivation. Leading to the hypothesis that the RNAi pathway may be involved in the initiation of transposon silencing.

However endogenous mammalian siRNA remain to be identified. Given the evolutionary conserved nature of the RNAi pathway, it is unlikely that mammals would be completely devoid of endogenous siRNA. Therefore mammalian genomes with their highly repetitive content would appear a logical target for the identification of short functional RNA. With the addition of their identification lending credence to the hypothesis of an RNAi mediated transposon silencing mechanism.

1. Sijen, T. and R.H. Plasterk, Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. *Nature*, 2003. 426(6964): p. 310-4.
2. Tabara, H., et al., The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell*, 1999. 99(2): p. 123-32.
3. Wassenegger, M., et al., RNA-directed de novo methylation of genomic sequences in plants. *Cell*, 1994. 76(3): p. 567-76.

(24)

Repetitive sequence environment distinguishes housekeeping genes

C. Daniel Eller[1], Moira Regelson[1], Barry Merriman[1], Stan Nelson[1], Steve Horvath[1,2,+], York Marahrens[1,+]*

[1] UCLA Department of Human Genetics David Geffen School of Medicine, Gonda Center, 695 E. Young Drive South, Los Angeles, California 90095-7088, USA
[2] UCLA Department of Biostatistics, School of Public Health

* Presenting author
+ Co-directed this work

Housekeeping genes are expressed at high levels and across a wide variety of tissues. In view of reports of repetitive sequences influencing gene expression, we investigated whether an association exists between the breadth and magnitude of housekeeping gene expression and their repetitive sequence context. We show that Alu elements are concentrated around housekeeping genes while several other repetitive sequences, including LINE-1 elements, are excluded from these regions. These properties, in combination with other previously published sequence properties of housekeeping genes, were used to develop a method of predicting housekeeping genes on the basis of DNA sequence alone. Using expression across tissue types as a measure of success, we demonstrate that repetitive sequence environment is by far the most important sequence feature identified to date for distinguishing housekeeping genes. We also examined whether relationships exist between repetitive sequence environment and the breadth and magnitude of expression throughout the entire genome. We find that Alu abundance correlates with breadth of gene expression. Alu elements were also progressively more abundant as one considered more and more highly expressed genes while the densities of LINE-1 elements and several other repeats were negatively correlated with magnitude of gene expression. We propose that Alu elements constitute a cellular defense that protects genes against the repressive effects of other types of repeats.

(25)

Dynamics and evolution of tirant in Drosophila

Marie Fablet[*1], *John McDonald*[2], *Christian Biemont*[1], *Cristina Vieira*[1]

[1] UMR CNRS 5558, Biometrie et Biologie Evolutive, Universite Claude Bernard Lyon1, France
[2] School of Biology, Georgia Institute of Technology, Atlanta, GA, USA

It is clear that transposable elements (TE) make up a high proportion of most genomes, however, the way they invade or are eliminated from species is still a burning topic. While it is known that invasion of a TE into a new species can occur rapidly, we have less information on the dynamics of the loss of copies and the observation of such a phenomenon still ongoing in a species is seldom reported. We approached this question by comparing the behavior of a LTR retrotransposon, tirant, in the close species of the melanogaster subgroup.
Tirant presents in average 11 insertion sites in Drosophila melanogaster flies from natural populations worldwide. In the sibling species Drosophila simulans, only African populations harbor a few insertion sites (1 to 5) on the chromosome arms, although sequences of tirant are present in the heterochromatin of most populations. This distribution in D. simulans either reflects a recent genomic invasion of a new copy of tirant in African populations, or a loss of tirant from the entire species with some sequence relics still present in Africa. In an effort to answer these questions, we focused on the LTR-UTR region of tirant copies from various populations of both D. melanogaster and D. simulans, and found two distinct types of regulatory regions: one, present in both species, and the other, transcriptionally inactive in gonads, present only in D. simulans. We propose that tirant invaded the genome of D. simulans in the past, but that this element is now in the process of disappearing from this species.

(26)

Retroelements and heterochromatin: the role of integrase chromodomains in target specificity

Xiang Gao[1], Hou Yi[1], Hirotaka Ebina[2], Henry Levin[2], and Daniel F. Voytas[1]*

[1] Genetics, Development and Cell Biology Deparment, Iowa State University, Ames, IA 50010
[2] Building 6B, Rm 2B-220 NICHD (National Institute of Child Health and Human Development), Bethesda, MD 20892

Retrotransposons are abundant in eukaryotic genomes and have profound impacts on chromosome organization and gene expression. Recent evidence indicates that retrotransposons contribute to heterochromatin formation through the siRNA pathway and to centromere function through interactions with centromeric proteins. We hypothesize that retrotransposons became integral components of heterochromatin by actively targeting integration to heterochromatic domains. This hypothesis emerges from our work with the yeast Ty5 retrotransposon, which targets to heterochromatin through an interaction between the C-terminus of Ty5 integrase and the heterochromatin protein Sir4p. Here we report families of retrotransposons that have chromodomain-like motifs in the C-termini of their integrases. Many of these retrotransposons are exclusively located within centromeric heterochromatin. The chromodomains fall into three classes, ranging from those that are highly similar to HP1 (Class I) to those that are highly divergent and share only a few amino acid residues with other chromodomains (Class III). For each class, chromodomain-YFP fusions appear as distinct foci within the nucleus that co-localize with CFP fusions to the Arabidopsis HP1 homologue (LHP1). Mutations in conserved residues in the chromodomain diminish the sub-nuclear localization. Pull-down assays indicate that the Maggy chromodomain (Class I) interacts with histone H3 that is dimethylated on Lys9. After Maggy chromodomain is fused to the integrase c-terminus of Tf1 retroelement, the fusion element can integrate into heterochromatin, the histone H3 K9 methylation rich region. Collectively, these data support the hypothesis that the novel chromodomains mediate retrotransposon target specificity by recognizing specific chromatin features.

(27)

Transposable Elements in Gene Coding Regions

Valer Gotea and Wojciech Makalowski*

Institute for Molecular Evolutionary Genetics and Department of Biology; Center for Comparative Genomics and Bioinformatics, The Pennsylvania State University, University Park, PA 16802

Abstract: It became widely accepted that the initial classification of transposable elements (TEs) as \"junk\" and \"useless\" pieces of DNA is inaccurate. Many studies have shown that TEs have complex regulatory functions, and they contribute to the coding regions of many genes. We recently showed that the fraction of TE-containing functional proteins is much smaller (~0.2%) than observations at transcript level would suggest (~4%). In all cases, TE-cassettes in functional proteins were derived from old TEs, whereas young ones, such as Alu and LINE1 elements, are more likely to disrupt the function of host proteins when inserted into their coding regions. We addressed the latter aspect through protein homology modeling and show why that is the case. A detailed analysis of the Alu-cassettes in protein coding regions revealed not only that Alu elements are not likely to be functional if/when translated, but also other surprising potential ways to impact the evolution of the human genome. Among these, we can mention that, in few cases, Alu elements appear to promote incorporation of selenocysteine residues into proteins, as well as they appear to be promote the formation of aberrant transcripts.

Selected references:
1) Thornburg, B.G, V. Gotea, and W. Makalowski (2005) Transposable elements as a significant source of transcription regulating signals. *Gene*, in press.
2) Gotea, V., and W. Makalowski (2005) Do transposable elements really contribute to proteomes? *Trends in Genetics*, submitted

(28)

Mapping polymorphic DcMaster insertion sites in the carrot (Daucus carota L.) genome

Dariusz Grzebelus[1], Barbara Jagosz[1], Philipp W. Simon[2]

[1] Dept. of Genetics, Plant Breeding and Seed Science, Agricultural University of Krakow, Poland
[2] Vegetable Research Crop Unit, USDA-ARS, and Dept. of Horticulture, University of Wisconsin, Madison, WI 53706, USA

DcMaster is a family of PIF-like class II transposable elements identified in carrot. Transposon Display (TD) molecular marker system allows identification of genomic regions containing predefined transposable elements. Individual copies are recognized on the basis of the size polymorphism of regions between the end of the transposon and the adjacent restriction site. To identify polymorphic insertion sites in a QAL x B493 (wild x cultivated) carrot F2 population we used a modified TD protocol: (1) genomic DNA was digested with MseI, (2) adaptors were ligated to the restriction fragments, (3) preamplification was performed with a primer specific to the 5\' subterminal region of DcMaster and an adaptor-specific primer, and (4) nested transposon- and adaptor-specific (+3 selective nucleotides) primers were used for the final amplification. We used 15 primer combinations and identified 106 segregating amplicons. We sequenced 60 products and confirmed presence of the DcMaster terminal sequence on one end of 58 of them (96%), proving that they originated from regions containing a copy of the DcMaster-like element. Segregating amplicons were used to saturate a molecular linkage map of carrot. Sixty markers were attributed to QAL, 41 originated from B493, and the remaining five markers were unlinked. In QAL, they were distributed over all nine linkage groups (1 to 13 markers per linkage group). In B493, one linkage group did not contain any TD markers, the remaining eight groups contained from 2 to 10 markers. The results suggest that insertions of DcMaster-like elements may not be uniformly distributed throughout the carrot genome.

(29)

Annotation of Repeat Elements in Plant Genomes

Heidrun Gundlach, Sindy Neumann, Klaus Mayer*

Munich Information Center for Protein Sequences (MIPS), Institute for Bioinformatics, GSF Research Center for Environment and Health, Neuherberg, Germany

Until recently genome projects focused on the annotation of protein coding genes. The vast amounts of repeat elements (REs) were seen as bothersome junk or parasitic DNA. But as more and more evidence accumulated, that besides their deleterious effects, REs also contribute to useful genetic variety the paradigm changed from “junk” to “necessary evolutionary task forces”. In plants the content of REs ranges from 10 % in the small Arabidopsis genome (125 Mb) to 80 % and more in large grass genomes like maize (2.3 Gb) or wheat (15 Gb). A detailed analysis of structural, compositional and sequence conservation related features of REs allows to ask questions about the evolutionary past of the genomes and to uncover the subtle driving forces underlying the evolution of plant genomes.

As a basis for a consistent and automated RE annotation an analytical framework has been established, facilitating comparative data extraction at different levels of detail. It includes a compilation of known REs to a nonredundant plant RE database and a detailed generic classification schema (ontology). The RE classification has an hierarchical tree structure with three main groups: 1. Simple Sequence Repeats (e.g. micro-, minisatellites and satellites), 2. Mobile Elements (with retrotroelements and DNA transposons) and 3. High Copy Number Genes (e.g. RNA genes, histones). An additional fourth category permits the assignment of general features, like replication type, chromosome location, sequence completeness or subelements. It is also intended to describe fragmented and nested structures of REs. The mobile elements constitute the most important and largest group, encompassing five tree levels.

Our current workflow for a comprehensive RE detection on plant genomic sequences follows three complementary approaches: homology to known REs, detection of typical intrinsic RE features (e.g. LTRs, RE protein signatures) and de novo repeat detection.

(30)

Large-scale deletion in the human genome is mediated by recombination between Alu elements

Shurjo K. Sen[1], Kyudong Han[1], Jianxin Wang[2], Jungnam Lee[1], Hui Wang[1], Pauline A. Callinan[1], Richard Cordaux[1], Ping Liang[2] and Mark A. Batzer[1]*

[1] Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-Scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA
[2] Department of Cancer Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA

Alu elements are the most successful non-LTR retrotransposons in the human genome, with a copy number of over 1.2 million. Numerous analyses implicate recombination between these elements as a cause of human genomic deletions associated with genetic diseases. In this study, we analyze the reference human and chimpanzee genome sequences to determine the evolutionary impact of the Alu recombination-mediated deletion process on the human genome after the human-chimpanzee divergence approximately six million years ago. Using a combination of whole-genome computational data mining and experimental verification, we have characterized 494 human genome-specific deletions (totaling ~400 Kb) that can be directly attributed to this process. This corresponds to a genomic deletion rate of 80 Kb/Myrs, making this process a significant component in the insertion/deletion spectrum of the human genome. 229 of the deletions coincide with known or predicted genes (including three that delete functional exons from the chimpanzee genome), implicating this process as a putative factor for promoting the genetic divergence of the human and chimpanzee lineages. Overall, our data indicate that genomic deletion caused by recombination between Alu elements has been active at a large scale not reflected by the isolated events previously identified, and that this process continues to influence the dynamics of human genome evolution.

(31)

Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium

Jennifer S. Hawkins[*1], HyeRan Kim[2], John D. Nason[1], Rod A. Wing[2], and Jonathan F. Wendel[1]*

[1] Iowa State University, Department of Ecology, Evolution and Organismal Biology, Ames, IA, USA 50011
[2] University of Arizona Genomics Institute and Computational Laboratory, Tucson, AZ, USA 85721

DNA content in eukaryotic nuclei (C-value) varies approximately 200,000-fold, but there is only an approximate 20-fold variation in the number of protein coding genes. Hence, most C-value variation is ascribed to the repetitive fraction, although relatively little is known about the evolutionary dynamics of the specific components that lead to genome size variation. To understand the modes and mechanisms that underlie variation in genome composition, we generated sequence data from genomic shotgun libraries for three representative diploid (n =13) members of Gossypium that vary in genome size from 841 to 2778 Mb (1C) and from a phylogenetic outgroup, Gossypioides kirkii, with an estimated genome size of 590 Mb. Differential lineage-specific accumulation of various families of transposable elements is observed among the different plant lineages. Furthermore, not only do different families of repetitive sequences accumulate at different rates, but various groups within families also accumulate at different rates. Indeed, the major fraction of genome size variation observed in Gossypium is largely due to recent, lineage-specific amplification of one particular group of gypsy-type retrotransposon sequences, Gorge3 (Gossypium retrotransposable gypsy-like element), within the larger genome Gossypium species. Gorge3 sequences have proliferated to the extent that they have achieved not only higher copy numbers in Gossypium species with larger genomes, but higher densities as well. Nevertheless, cumulative copy number estimates including all dispersed repetitive sequences suggest that each genome contains approximately the same proportion of total transposable elements (40-65%).

(32)

Alu-based identification of anonymous primate DNAs

Scott W. Herke[1], Richard Cordaux[1], David A. Ray[2], Jinchuan Xing[1], Jacqueline Zimmerman[1], and Mark A. Batzer[1]*

[1] Department of Biological Sciences, 202 Life Sciences Bldg., Louisiana State University, Baton Rouge LA 70803, USA
[2] Department of Biology, 53 Campus Dr., West Virginia University, Morgantown WV 26506, USA

Modern primate genomes contain lineage-specific, population-specific and individual-specific insertions of Alu elements. In previous studies, three different sets of loci were used to ascertain various relationships within the three anthropoid primate groups: Hominids (H), New World Monkeys (NWM), and Old World Monkeys (OWM). Here, we focused on using these elements to identify anonymous DNA samples from anthropoid primates. Thus, for each node of a combined phylogenetic tree, we tested primers for their ability to amplify loci (~400 in total) from 30 species (n = 6 [H]; 9 [NWM]; 15 [OWM]). An ideal primer pair amplified the locus in most species such that a clean, distinct band of the appropriate size for an “empty” or a “filled” site was visible in a 2% agarose gel. We usually discarded loci that either generated unexpected band sizes or amplified only in a restricted set of species. The phylogenetic position of an anonymous primate DNA sample is best ascertained by using the tree as a dichotomous key for primer selection, with the starting point determined by the level of initial uncertainty regarding the sample’s identity. At terminal branches, conclusions should be based on multiple loci because the within-species polymorphism status of each locus is currently unknown. Similarly, if the DNA could be from a non-analyzed lineage, identifications at the terminal branch level are provisional; by contrast, when the possibilities are limited (e.g., when ensuring the purity of previously authenticated cell lines), such identifications may be considered definitive.

(33)

Retand: A gypsy-like retrotransposons harboring an amplified tandem repeat

Eduard Kejnovsky[*1], Zdenek Kubat[1], Roman Hobza[1], Jiri Macas[2] and Boris Vyskot[1]*

[1] Laboratory of Plant Developmental Genetics, Institute of Biophysics, Academy of Sciences of the Czech Republic, CZ-612 65 Brno, Czech Republic
[2] Laboratory of Molecular Cytogenetics, Institute of Plant Molecular Biology, Academy of Sciences of the Czech Republic, CZ-37005 Ceske Budejovice, Czech Republic

Abstract. Here we describe a novel family of Ty3/gypsy retrotransposons isolated from the dioecious plant Silene latifolia. It includes a non-autonomous element Retand-1 (3.7 kb) and its autonomous partner Retand-2 (11.1 kb). An interesting feature of these retrotransposons is the presence of an array of tandem repeats which is more amplified in the non-autonomous element. Similar arrays of tandem repeats have been already found in several mobile elements including Grande1, Cinful, micropia, Dasheng and ZLRS. The function of these sequences is unknown, although their conserved location among some Ty3/gypsy elements indicates their potential structural and functional significance. Moreover, an autonomous element Retand-2 contains two additional open reading frames in antisense orientation localized between the pol gene and right LTR. RT-PCR revealed that Retand is transcriptionally active in all organs tested (leaves, flower buds and roots), which together with the high sequence similarity of LTRs in individual elements, indicates their recent transpositional activity. The autonomous elements are similarly abundant (2,700 copies) as the non-autonomous ones (2,100 copies) in S. latifolia genome thus representing another example of staggering success of non-autonomous elements. The Retand elements are also present in other Silene species, mostly in subtelomeric heterochromatin regions of all chromosomes.

(34)

Testing models of spliceosomal intron insertion in C. elegans

Min Kyung Kim, Vivek Gopalan and Arlin Stoltzfus*

Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive
Rockville, Md 20850

The mechanistic basis for evolutionary gain of spliceosomal introns is not understood. One possible mechanism is that the donor intron is inserted, in RNA form, into the germ line expressed mRNA recipient by reversal of the splicing reaction (as observed for group I introns in vivo), followed by reverse-transcription and incorporation of the cDNA using host recombination/repair mechanisms. Another possible mechanism is direct reverse-splicing of donor intron RNA into the recipient DNA as observed in the reverse splicing of group II introns. In either case, intron insertion most likely will require complex factors involved in reactions such as RNA maturation by RNA maturase, DNA target site cleavage by endo-nuclease, and cDNA synthesis by reverse transcriptase that are supplied to the intron element by retrotransposons.

To distinguish the two different modes of intron RNA integration into a target, 81 novel intron insertions into 79 genes of Caenorhabditis elegans identified by Coghlan and Wolfe (2004) were used to gather specific characteristics of host genes that are recipients of intron insertion events. Microarray expression data of C. elegans were examined for expression patterns of the recipient genes that might indicate one or the other mechanism of intron gain. We confirm the original result of Coghlan and Wolfe that the recipient genes are not enriched for genes classified as “germ line”, suggesting that the insertion of novel introns is probably mediated at the level of DNA rather than RNA. Nevertheless, the mean quantitative level of expression in C. elegans is higher for the intron recipient genes, both in adult worms and L3 juveniles who are only undergoing germ-line development.

Because the number of samples used in this study is too small, we will investigate this effect further for a much larger set of recipient genes.

1) Coghlan, A. and Wolfe, K. (2004) Origins of recently gained introns in Caenorhabditis. *PNAS* 101: 11362-11367.
2) Dickson, L., Huang, H., Liu, L., Matsuura, M., Lambowitz, A. and Perlman, P. (2001) Retrotransposition of yeast group II intron occurs by reverse splicing directly into ectopic DNA sites. *PNAS* 98: 13207-13212.
3) Roman, J. and Woodson, S. (1998) Integration of the Tetrahymena group I intron bacterial rRNA by reverse splicing in vivo. *PNAS* 95: 2134-2139.

(35)

Whole genome evaluation of LINE-1 insertion polymorphism through draft genome assembly comparison

Miriam K. Konkel[*1], *Jianxin Wang*[2], *Ping Liang*[2], *and Mark A. Batzer*[1]

[1] Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-Scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA
[2] Department of Cancer Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA

Mobile elements represent a relatively new group of markers for the study of human evolution. Long interspersed elements (LINEs) belong to a group of mobile elements comprising approximately 15% of the human genome. LINEs have amplified to a copy number of roughly 100,000 over the last 100 million years of mammalian evolution. Young LINE-1 (L1) elements that have integrated more recently into the human genome have the potential for being polymorphic in different populations at particular chromosomal locations. To identify candidate putative L1 insertion polymorphisms, we computationally compared two draft assemblies of the whole human genome (GenBank and Celera). We found 72 potential polymorphic L1 loci. Based on further manual analysis we selected 34 loci for further experimental studies. Using PCR based assays, we successfully analyzed the 34 loci in 80 unrelated individuals from four diverse human populations: African American, Asian, Caucasian, and South American. Each amplified fragment was verified through automated DNA sequence analysis. All but two of the selected loci appeared to be polymorphic. For one locus we could not detect a filled or L1 occupied site, indicating that the allele frequency is most likely very low or perhaps that this is a retrotransposon unique to a single individual. Another locus was heterozygous for all individuals tested, even though the flanking sequence appeared to be unique in the database suggesting that the mobile element resides in some type of previously unidentified repeat sequence, or that it is subject to paralogous amplification. The remaining 32 L1 insertion polymorphisms were confirmed across the two draft genomes. These elements represent a valuable source of mobile element based genomic polymorphism for the study of human populations.

(36)

The use of the Repbase sequences, RepeatMasker and Censor to reconstruct the duplication history of the transplantation class I genes within the Major Histocompatibility Complex genomic region of primates

Jerzy K Kulski

Centre for Bioinformatics and Biological Computing, School for Information Technology, Murdoch University, Murdoch, Western Australia, Australia; Division of Molecular Life Science, Department of Genetic Information, Tokai University, Isehara, Kanagawa, Japan.

The Major Histocompatibility Complex (MHC) class I genes within the MHC genomic region of primates are variable in copy number and may be paralogous and/or orthologous depending on the species comparison. The transplantation class I genes have in the main coevolved with various retrotransposon and DNA transposon families by duplication, deletion and rearrangement within the Major Histocompatibility Complex genomic region during primate evolution. We have previously used different retrotransposon and DNA transposon family subtypes (Repbase) and the computing tools RepeatMasker and/or Censor to analyse and reconstruct the duplication history of the transplantation genes within humans. In this presentation, I will show how the annotated Alu, L1, LTR, ERV, Charlie and other repeat elements can be used to classify duplicated genomic structures and provide insights into the duplication history of the 62 class I genes of the rhesus macaque and the 17 class I genes of the human and chimpanzee.

1) Kulski JK, Anzai T, Inoko H. ERVK9, transposons and the evolution of MHC class I duplicons within the alpha-block of the human and chimpanzee. *Cytogenet Genome Res*. 2005;110(1-4):181-92
2) Kulski JK, Gaudieri S, Dawkins RL. Using alu J elements as molecular clocks to trace the evolutionary relationships between duplicated HLA class I genomic segments. *J Mol Evol*. 2000 Jun;50(6):510-9
3) Kulski JK, Gaudieri S, Martin A, Dawkins RL. Coevolution of PERB11 (MIC) and HLA class I genes with HERV-16 and retroelements by extended genomic duplication. *J Mol Evol*. 1999 Jul;49(1):84-97.

(37)

The association between non-melanoma skin cancer and young dimorphic Alu elements within the Major Histocompatibility Complex class I genomic region

Jerzy K Kulski[1, 2], David S Dunn[1], Hidetoshi Inoko[2]

[1] Centre for Bioinformatics and Biological Computing, School for Information Technology, Murdoch University, Murdoch, Western Australia, Australia
[2] Division of Molecular Life Science, Department of Genetic Information, Tokai University, Isehara, Kanagawa, Japan.

Non-melanoma skin cancer (NMSC) has increased dramatically worldwide with the highest skin cancer rates reported in Australia. Genes within the Major Histocompatibility Complex (MHC) have been implicated in the development of NMSC and a NMSC susceptibility region within the MHC class I region was previously identified telomeric of the HLA-C gene using high-density microsatellite markers. Here, we report on the relationship between five polymorphic Alu insertions (POALINs) within the MHC class I region and their dimorphic association with NMSC in a study of 154 NMSC patients and 213 normal controls recruited from the town of Busselton in Australia. The allele frequencies of the AluyMICB insertion and the AluyTF insertion that are located respectively centromeric and telomeric of the HLA-C gene were significantly increased ($p<0.05$) in NMSC patients. There was no difference ($p>0.05$) between the cancer patients and controls for the allele frequencies of the other three POALINs that are located within the genomic region of the HLA-A, -G and –F gene cluster. This study confirms our previous study using microsatellite markers (Oka et al 2003 TA 61, 203) that the MHC class I POALINs located within the NMSC susceptibility region and near the HLA-C gene were strongly associated with NMSC. At least 11 genes including HCG22 and C6orf205 that are located within the NMSC susceptibility region or in close proximity to the AluyTF insertion have a potential role in NMSC.

(38)

Sireviruses in plant genomes

Laten, Howard[1], Gouvas, Eftychia[1], Villasenor, Deany[1], Badal, Edward[1], Havecker, Ericka[2], Winfrey, Ron[3], Wright, David[3], and Voytas, Dan[2]*

[1] Loyola University Chicago, Chicago, IL 60626
[2] Iowa State University, Ames, IA 50011
[3] Phytodyne, Inc. Ames, IA 50010

We are investigating the distribution and expression of an unusual class of Ty1-copia-like retroelements that contain a preserved ORF beyond pol. This ORF has optimistically been called envelope-like by us and others based on patterns of predicted membrane-spanning domains and coiled-coils similar to those in mammalian retrovirus envelope proteins, suggesting SIRE1 may be infectious. Copies of SIRE1 are present in Glycine max and the absence of nonsense and frameshift mutations in four of the eight full-length or slightly truncated copies previously sequenced suggests that many members of this 1,000-copy family are functional. Despite its copy number, the only other representatives of SIRE1 from soybean in Genbank are presently in the GSS and EST databases. In contrast, SIRE1 copies are frequently encountered in sequences available from two model legumes: Lotus corniculatus and Medicago truncatula. Some copies in L. corniculatus appear to be as youthful (<100,000 years) as those in G. max. Copies in M. truncatula appear to be considerably older and highly degenerate. We have retrieved, annotated, and analyzed several of these sequences in order to deduce the nature of changes that led to their demise or preservation. Degenerate Sireviruses are also present in pea, maize, Arabidopsis, and tomato. Understanding the function of the env-like protein has been a daunting task for a number of reasons, not the least of which is the fact that no homologs have been found. Sequence conservation of this coding region is equal to that of the structural gag protein, so we believe it plays an important role in element function. We are employing a yeast two-hybrid screen in the hope of detecting interacting proteins from G. max, and will now expand the screen to L. corniculatus. Evaluations of SIRE1-induction by elicitors of plant defense responses using LTR-GUS constructs in tobacco will also be presented.

(39)

Transposon insertion polymorphisms and their impact on human gene activity.

Yuri Lebedev, Ilgar Mamedov, Svetlana Ustyugova, Anna Amosova, Eugene Sverdlov*

Shemyakin-Ovchinnikov Institute Russ.Acad.Sci., Moscow, Russia

Several groups of primate-specific transposable elements, including AluY, L1 subsets and the HERV-K family, amplified in the ancestral genome at the time of human and chimpanzee divergence and during hominid radiation. Following a suggested impact of these retroelements (REs) on human genome evolution, functioning, and diversity, we target the research at comprehensive identification of polymorphic RE insertions and their interplay with the expression of the surrounding human genes.
We developed an experimental technique for simultaneous detection of polymorphic REs in many genomes. The technique exploits DNA subtractive hybridization and allows one to reveal RE integrations that discriminate genomes under comparison. Complementing bioinformatic approaches, it was successfully applied to identification and mapping of polymorphic AluYa5/Ya8 insertions in a mixed sample of 30 individual genomes. 40% of the identified in this work insertions were absent from available databases. All the polymorphisms were verified by testing a panel of genomic DNAs representing 20 Eurasian populations. Using known RE polymorphisms, we constructed a set of 47 genetic markers corresponding to introns of the human genes with dimorphic L1-Ta and AluY insertions. Genotyping of 10 human cell lines of various origins revealed characteristic fingerprints for each cell line. We developed an approach for pairwise comparison of heterozygous gene alleles expression and studied transcription of these genes in a number of cell lines. An association of particular RE insertions with a decreased content of the corresponding hnRNAs was demonstrated. Some characteristics of this effect including tissue specificity, intragene position and orientation of retroelements are discussed. Summarizing current results, the developed techniques provide a reliable tool for further studies on a role of human retroposons in human genome functioning.

(40)

Tissue-specific regulation of HERV-L LTRs

Stephan Weinhardt [1,2], Ulrike Schön[1], Volker Erfle[1], Christine Leib-Mösch[1,3]*

[1] GSF-National Research Center for Environment and Health, Institute of Molecular Virology, Neuherberg, Germany
[2]Insitute of Virology, Technical University of Munich, Germany
[3]Medical Clinic III, Faculty of Clinical Medicine Mannheim, University of Heidelberg, Germany

The human genome contains more than 450,000 endogenous retroviral LTR sequences that can be regarded as mobile regulatory modules for gene expression. We have analyzed over 100 randomly selected human endogenous retroviral (HERV) LTRs in a transient transfection assay using luciferase as a reporter gene. About two thirds of these sequences were still active and able to promote the expression of any gene. Most of these LTRs showed differential activities depending on the cell type. For example, proviral and solitary LTRs of class III elements (HERV-L) proved to be specifically active in human skin keratinocytes (HaCaT, NHEK), HeLa cells, and/or uterus cells (KLE). To define the regulatory sequences involved in tissue specificity, a number of HERV-L LTR sequences with different specificities were compared and potential transcription factor binding sites identified. Computer analysis using the program GEMS launcher (Genomatix) suggested the presence of partially overlapping binding sites for the AP1, Sp1, Egr-1, AP4R/NEUR, SMAD, PAX8 and HOX families of transcription factors. Electrophoretic mobility shift assays (EMSA), supershift assays with specific antibodies, and site directed mutagenesis revealed the importance of Sp1, AP4R, PAX8 and HOX binding in KLE and HaCaT cells, whereas AP1, Egr-1 and SMAD binding sites, in addition, appear to be essential for the transcriptional activity of HERV-L LTRs in keratinocytes.

(41)

Expression of a truncated calbindin protein from a HERV-H LTR in human prostate carcinoma cells

Eva Gebefügi[1], Reinhard Brunmeir[2], Volker Erfle[1], Christine Leib-Mösch[1,3]*

[1] GSF-National Research Center for Environment and Health, Institute of Molecular Virology, Neuherberg, Germany
[2] Institute of Medical Biochemistry, Medical University of Vienna, Max Perutz Laboratories, Vienna, Austria
[3] Medical Clinic III, Faculty of Clinical Medicine Mannheim, University of Heidelberg, Germany

Over 8% of the human genome is composed of endogenous retroviral elements (HERV), many of which still possess active regulatory sequences within their LTRs that may influence the transcription of adjacent genes. In a previous study (Liu and Abraham 1991), a chimeric cDNA fragment containing HERV-H LTR sequences fused to the human calbindin gene was isolated by subtractive cloning from a cell line (PC3) derived from a human prostate carcinoma metastasis. We identified a 5.6 kb proviral HERV-H element that is located in the sense direction about 7 kb upstream of the human calbindin gene (CALB1). Using RT-PCR with HERV-H- and calbindin-specific primers we detected three different aberrant calbindin transcripts in PC3 cells. In the major transcript the HERV-H element splices into the second exon of CALB-1. Western blot experiments confirmed that all transcripts lead to the expression of a truncated 23.6 kDa protein lacking the first of 4 EF-hand motives required for calcium binding. Investigation of the DNA methylation status and chromatin immunoprecipitation (ChIP) analysis suggest that the HERV-H LTR promoter is specifically activated in PC3 cells by demethylation of the DNA and chromatin modification.

(42)

Mobile elements and the birth and death of DNA palindromes in primates

Susanna M. Lewis[*1,2] and Roseanne Richard*[1]

[1] Program in Genetics and Genomic Biology, Hospital for Sick Children and
[2] Department of Molecular and Medical Genetics University of Toronto, Toronto, ON

The presence of perfect DNA palindromes in primate genomes is coming to the fore and the extent to which palindromes cause DNA dysfunction is an area of active exploration. A related issue, which has not had much attention to date, is how purely palindromic sequence motifs are created.

A basic question is whether palindromes are generated from pre-existing inverted repeats (ie. by replication errors) or whether perfect inverted repeats derive, via “center-break” –type decay from perfect palindromes (leaving the origin of the palindromes themselves unexplained). Experimental evidence exists for each scenario, but the mechanisms that have the greatest impact in the human genome are unknown. The distinction is of fundamental importance for understanding the root causes of pathological DNA rearrangement and ultra-rapid evolutionary change.

In this study, we have investigated the relationship between perfect palindromes and perfect inverted repeats in the human genome. All palindromes greater than 90bp in length, and all perfect inverted repeats with arms greater than 90bp in length with ≤ 100bp central spacers were compared. We further narrowed the focus of the investigation to inverted repeats and palindromes that incorporate repetitive elements. By exmining repeats that overlap identified mobile elements it is possible to infer some of the generative processes leading to palindrome creation. The imprint of both de-novo and derivative pathways of palindrome formation were evident, as will be presented

(43)

Ogre retrotransposons: a distinct group of gypsy-like elements with significant impact on genome evolution of legume plants

*Jiri Macas *, Pavel Neumann, Andrea Koblizkova, Alice Navratilova*

Institute of Plant Molecular Biology, Ceske Budejovice, Czech Republic

Ogre elements were first discovered during sequencing of repetitive DNA in pea (Pisum sativum) where they make up 5% of its nuclear DNA. The order of domains coding for the retroelement proteins found within the Ogre sequences is typical for Ty3/gypsy retrotransposons. However, there are several features that make Ogre elements a distinct group: (i) the exceptional size of the elements between 22-25 kb, including LTRs up to 6 kb long; (ii) the occurrence of multiple stop codons in the region between prot and rt/rh domains. It has been proved experimentally that these stop codons are located within an intron that is spliced out from a portion of the Ogre transcripts; (iii) the presence of an extra ORF coding for protein of unknown function upstream of the gag-pol region. Computer-based and experimental approaches revealed that Ogre elements are present within several genera of legumes, including the model legume Medicago truncatula. It was also found that they played a key role in genome evolution in the genus Vicia, where their recent amplification caused up to 50% expansion of the genome size in some species. It is of particular interest that the elements are highly conserved in sequence and transcribed, suggesting they may still be transpositionally active. Thus, they may provide a good experimental model to study retroelement interaction and co-evolution with their host genomes.

(44)

Differential epigenetic control of the highly related MusD and ETn families of active mouse retroelements

Irina A. Maksakova and Dixie L. Mager*

Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, BC, Canada and Dept. of Medical Genetics, Univ. of British Columbia

The LTRs of endogenous retroviruses (ERVs) carry powerful regulatory signals and, therefore, ERV transcription is thought to be suppressed in adult tissues via epigenetic silencing. Here we have examined epigenetic regulation of mouse Early Transposon (ETn) LTR elements and their relatives, MusD ERVs. Suppressed in differentiated tissues, these elements are mainly expressed in embryonic carcinoma (EC) cells, ES cell lines and early embryos. ETns are highly active mouse mutagens - having caused over 20 germ-line mutations. These elements, which lack coding potential, are likely derived via recombination from MusD ERVs that provide proteins for ETn retrotransposition. Despite nearly identical LTRs, transcripts from the more abundant MusD elements are present in embryos at much lower levels than ETn transcripts. Using genomic Southern analysis, we have now found that ETn/MusD LTRs are highly methylated in NIH3T3 fibroblast cells. Moreover, in EC and ES cells, we found that identical CpG sites within ETn/MusD LTRs are much less methylated when LTRs are located in the context of ETn elements compared to MusD elements. Thus, identical CpG sites are differentially methylated depending on genomic context. We hypothesize that this differential LTR methylation results from differences in interior sequences adjacent to the LTRs. MusD elements have a CpG-rich gag gene whereas ETn elements possess a relatively CpG-poor 5' interior region of unknown origin. We suggest that methylation spreading from the MusD gag region results in higher levels of LTR methylation and lower levels of transcription compared to ETns. We are currently conducting experiments to directly assess the effect of the different ETn/MusD internal regions on methylation spreading and transcription from linked LTRs. We propose that, due to their internal structure, ETn elements are less likely to be transcriptionally silenced by the host, contributing to their continual mutagenic activity in the mouse germ line.

(45)

Co-amplification of retroposons in the histone gene clusters of mosquitoes

C.A. Malcolm, T. Adams and P.W. Grosvenor*

School of Biological and Chemical Sciences Queen Mary, University of London, Mile End Road, London E1 4NS, UK

c.a.malcolm@qmul.ac.uk

Retroposons, also termed non-long terminal repeat retroposons, contain a major sub-group distinguished by the presence of a domain coding for an enzyme similar to an apurinic-apyrimidinic endonuclease. These elements normally insert almost non-specifically, but some exceptions have been found in ribosomal RNA genes and other tandemly arrayed repetitive DNA. We examined the possibility that tandem arrays of histone gene cluster repeats also harbour site specific retroposons in two mosquitoes, Anopheles stephensi and An. gambiae. The equivalent of two and half tandem repeats (~14 kb) of the An. stephensi H2a, H2b, H3 and H4 histone gene cluster was cloned and sequenced. The central repeat was flanked by repeats containing an identical 5' truncated element belonging to the Jockey clade, which almost exactly replaces the H4 and H2a genes. Further analysis based on Southern blots and PCR indicates that the element is present in a significant proportion of the histone gene repeats, but that the elements fall into ten or more categories according to the extent to which they are 5' truncated. In An. gambiae the structure of a typical histone cluster repeat is similar, but longer than in An. stephensi. A 5' truncated retroposon from the CR1 clade was found inserted between the H4 and H2a genes. In both species it is clear that the processes promoting concerted evolution of histone gene cluster repeats have co-amplified the regions containing original retroposon insertions. Whether, or not, the insertions were in each case a single event, or multiple and site specific, is still being investigated.

(46)

Comparing repeat libraries

Degui Zhi

Bioinformatics Program, University of California, San Diego, USA

As more genome sequences of related species become available, the question of comparing repeat contents across related species is of emerging importance. Due to various evolutionary processes, the direct identification of orthologous transposon insertions is often not possible, one can gain insights into repeat-related genome evolution via the comparison of repeat libraries.

Comparing repeat libraries is a non-trivial task. For instance, Stein et al 2003 conducted pairwise alignments between repeat family sequences from C. elegans and C. briggsae repeat libraries, but they were unable to identify ortholog pairs between these two species. This is due the fact that different repeat families typically share subsequences, rather than their entire sequences. These shared subsequences, called repeat domains, are of particular interests since they may represent domains that are important for the replication of the repeat families. Thus, our goal of comparing repeat libraries is not to identify ortholog pairs (such as in comparing genes across different species), but to compare the repeat domains.

We developed a repeat domain graph representation of the complex mosaic structure of repeat family sequences. Using this representation, we are able to systematically identify all shared repeat domains across different species. In particular, we found that a 34 bp repeat domain is conserved across 9 repeat families in C. elegans and C. briggsae. A detailed investigation reveals that this repeat domain represents a conserved core for diverged subfamilies of the CEREP5 element. This 34 bp domain preferentially locates to upstream of splicing acceptor sites, with a fixed orientation, indicating it may be under positive selection.

As more genome sequences are available, and repeat libraries are typically constructed by automatic tools, the utility of our approach will multiply.

Reference:

Stein, L.D., et al, *PLoS Biol*, 2003. 1(2): p. E45.

(47)

ICluster: A program for clustering taxa from phylogenetic trees

Daniel Svenback, Anders Kvist and Patrik Medstrand

Dept Experimental Medical Science, Lund University, Lund, Sweden

An important task for biologists is to classify relationships between biological sequences.
To address this problem, we present a novel method implemented in thprogram ICluster, which categorize biological sequences into coherent groups (so called clusters) based on evolutionary relations. ICluster divides a phylogenetic tree into well-separated and coherent subtrees or clusters based on the topology and evolutionary distances represented by the original tree.Sometimes there is not an obvious partition of the tree. In these cases, ICluster gives alternative clustering solutions. ICluster can be used either unsupervised or semi-supervised where the user can interact with the program. An advantage with ICluster is that the number of clusters not has to be specified in advance as is the case for k-means clustering. Furthermore, ICluster consider that taxa can be classified into more than one cluster in terms of cluster overlapping. The method was validated on various data sets with manually curated clusters as well as artificial test cases. We found the partitioning produced by ICluster is in good agreement with manually curated clusters. We discuss how ICluster can be used to sub-divide large families of retroelements into coherent groups which may allow improved unbiased classification of retroelement families.

(48)

Epigenetic regulation of mobile genetic elements in the mouse system

Wolfgang J. Miller[1], Reinhard Brunmeir[2], Sabine Lagger[2], Christian Seiser[2]

[1] Laboratories of Genome Dynamics, Center of Anatomy and Cell Biology, Medical University of Vienna, Austria; [2] Max Perutz Laboratories, Vienna Biocenter, Austria

Similar to the human genome almost half of the mouse genome is composed of sequences derived from parasitic DNAs. The vast majority of these transposable elements (TEs) can be regarded as genomic fossils, i.e. having eroded by random mutations and thus representing remnants of ancient TE-bombardments shaping the mammalian genome in many ways during the evolutionary past. In contrast to humans a significant fraction of mobile DNAs in mice is transpositionally competent causing approximately 10% of all spontaneous phenotypic mouse mutations that are mainly caused by the activity of one specific LTR retrotransposon family, i.e. Intracistral A-particles (IAPs).
The two main epigenetic regulatory silencing mechanisms, acting at the transcriptional (DNA-methylation) and post-transcriptional level (RNA interference), both efficiently repress TE-mobility in various biological systems, are currently under extensive studies. Due to the fact that (i) in the course of early mammalian embryogenesis genomic DNA of mobile elements is present in a demethylated state, (ii) TE silencing is retained in the absence of murine DNA methyltransferse enzymes and (iii) many organisms like invertebrates are in general able to silence mobile DNAs in the absence of DNA-methylation it has been assumed that a third and perhaps more ancestral, chromatin-based TE-silencing mechanism might exist.
Hence we started to survey the potential role of a specific class of histone-modifying enzymes, i.e. the histone deacetylases (HDACs), in IAP regulation in the mouse system. Our preliminary data support our model that HDAC 1 and other HDACs are important components of the transcriptional TE-repression system presumably expanding our view of the molecular repertoire of host-encoded TE-silencing strategies.

(49)

PCR-based detection of Pol III-transcribed transposon RNA

*Max Myakishev *[2], Valentina Kulichkova [1], Oksana Polesskaya [3], Larissa Gause [4] and Irina Konstantinova [1]*

[1] Institute of Cytology, St. Petersburg, Russia
[2] National Cancer Institute, NIH, Bethesda, MD
[3] George Mason University, Manassas, VA
[4] Koltzov Institute of Developmental Biology, Moscow, Russia

BACKGROUND. RNA of repetitive elements is abundant in mammalian cells mostly as read-through transcripts i.e. intronic or untranslated parts of ordinary genes transcribed by Pol II. In addition, transposons are transcribed by Pol III, in particular, during embryogenesis or in response to stress. Until now, Pol III transcripts were detected by primer extension with reverse transcriptase followed by gel electrophoresis.

METHOD. Here we report a PCR-based method of detection of Pol III-transcribed RNA of a specific transposable element. The method is based on the fact that the sequences of 5\' end of the Pol II and Pol III transcripts are different. In brief, the cDNA is synthesized using an oligonucleotide located downstream of the transposon\'s transcription start. A sticky-ended adapter specific to the 5\' end of the Pol III-transcribed transposon RNA is ligated to the cDNA. The Pol II-transcripts do not ligate at this step as their ending sequences are different. The resulted sequence is amplified using a pair of primers hybridizing to the adaptor and the transposon. Only Pol III transcripts amplify at this step.

RESULTS. The method was applied to detection of Pol III transcripts of transposon B1 (a rodent analog of Alu) in rat and mouse cells. The content of Pol III-transcribed B1 was found to be affected by cell stimulation and different in different cell compartments.

ACKNOWLEDGEMENTS. The authors thank Dr. Carl Schmid for the central for this method idea of ligation of an adaptor as means for distinguishing Pol II and Pol III transcripts. The work was done in Institute of Cytology, St. Petersburg, Russia and funded by grant 050449606 from the Russian Fund for Basic Research.

(50)

P element and MITE relatives in the whole genome sequence of Anopheles gambiae

Nouaud D, Quesneville H. and D. Anxolabéhère*

Dynamique du Génome et Evolution, Institut J. Monod , CNRS-Universités P.M. Curie and D. Diderot. 2 place jussieu 75552 Paris (France)

Miniature Inverted Terminal Repeat Elements (MITEs), which are particular class-II transposable elements (TEs), play an important role in genome evolution, because of their very high copy numbers and their recurrent bursts of transposition. The 5' and 3' subterminal regions of a given MITE family often show a high sequence similarity with the corresponding regions of a Class-II TE family. However, the sustained presence over a prolonged evolutionary time of MITEs and TE master copies able to promote their mobility has been rarely reported within the same genome, and this raises fascinating evolutionary questions. We report here for the first time the presence of P transposable elements with related MITE families over a prolonged evolutionary time in the Anopheles gambiae genome. Using a TE annotation pipeline we have identified and analyzed all the P sequences in the sequenced A. gambiae PEST strain genome. More than 0.49% of the genome consists of P elements and derivates. P elements can be divided into 9 different subfamilies, separated by more than 30%. Seven of them are present in both full length and deleted copies. Ten MITE subfamilies are associated with 6 out of the 9 P subfamilies. Southern blot and PCR analysis performed on five natural populations show that these P subfamilies are still active. Comparing their intra nucleotide diversities and their structures allows us to propose the putative dynamics of their emergence. In particular, one MITE family which has a hybrid structure, with ends each of which is related to a different P-subfamily, suggests a new mechanism for their emergence and their mobility.

(51)

The evolutionary history of human DNA transposons: evidence for intense activity during the primate radiation

John K. Pace, II[1], Cedric Feschotte, PhD*[1]

[1] University of Texas at Arlington

DNA transposons are a substantial component of the human genome, comprising approximately 3% of our genomic material. However, unlike for L1 and Alu retrotransposons, the tempo and evolutionary history of human DNA transposons has not been comprehensively analyzed. The prevailing method for dating TE insertions is by calculating the percent divergence of individual copies from their family consensus sequence. Because of differences in CpG content, genomic location, and difficulty in correctly calibrating the molecular clock, this method makes it difficult to uniformly and accurately date relatively ancient repeats. Here we combined this approach with two separate computational methods to infer the evolutionary history of human DNA transposons, both of which do not rely on sequence divergence. First, we analyzed the pattern and timing of all nested insertions of DNA elements into L1 and Alu subfamilies for which a rich paleogenomic record is available. Second, we assessed the presence/absence of ~450 human DNA transposon loci at orthologous positions in eight primate species using data from the ENCODE project. Based on this combined data, 35 out of 125 DNA transposon families recognizable in the human genome were found to be either anthropoid- or primate-specific. About 76,000 individual elements inserted at this time remain in the human. We found no evidence for the activity of any families after the split of new world monkeys, about 40 myr ago. Together this data suggests there was an intense and steady amplification of DNA transposons during the first ~35 myr of primate evolution. Given the distinct potential of DNA transposons in genome restructuring, these elements were likely important players in the early evolution of the primate genome. This period of intense proliferation was apparently followed by a strong bottleneck of activity leading to the massive extinction of virtually all DNA transposons in an anthropoid ancestor.

(52)

RetroposonBase: A Dynamic Web Resource for Bioinformatic Analysis of Retroposons

A.L.Collinson, M.R.Pancholi*, Y.F.Alonge and C.A.Malcolm*

School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End Road, London, E1 4NS, United Kingdom

c.a.malcolm@qmul.ac.uk

A publicly available web resource, RetroposonBase (http://www.retroposonbase.com), has been developed for retroposon (non-LTR retrotransposable element) research. The website consolidates various bioinformatic tools into one highly functional and dynamic arena for post-sequencing analysis.

The website houses RetroposonBank, a database containing thousands of retroposon sequences, accessible from a diverse catalogue of organisms. Full length elements, retroposon sequence tags (RSTs), specific domains, flanking regions and chromosome co-ordinates are also currently available. Multiple sequences are available as nucleotide or protein alignments of retroposon families. The user can query the database by host, clade, family, RetroposonBank (RB) accession number or by a BLAST search. Data can be uploaded by the user via an online submission form.

At present there is no universal classification system for retroposons; RetroposonTree is being developed to combat this issue. The dynamic phylogenetic tree is being assembled using the highly conserved reverse transcriptase domain from RetroposonBank and will be automatically modified upon submission of novel sequences.

Klassify is a program written using Bioperl, it aids in the classification system problem by characterising retroposons into families based on sequence identity.

RetroposonFinder is designed to automate the identification and retrieval of known and highly related retroposon sequences from BLAST databases. Written using Bioperl it uses a set of reverse transcriptase domains representative of the major families identified in our tree to interrogate the database for whole and truncated retroposon sequences. The sequences are characterised and then added to retroposon bank. This program is in the debugging stages of development and will be available for download shortly.

(53)

Retrotransposons in mouse oocytes and cleavage-stage embryos.

Anne E. Peaston[*1], Keith W. Hutchison[1,2], Barbara B. Knowles[1].*

[1] The Jackson Laboratory, Bar Harbor, Maine
[2] Department of Biochemistry, Microbiology and Molecular Biology, The University of Maine, Orono, Maine

The events of very early mammalian development depend on the union of the genomes of the oocyte and sperm to form the genome of the totipotent embryo. This necessitates extensive changes in genomic function, usually described as genomic reprogramming, and generally understood to be mediated by serial epigenetic modifications of nuclear DNA. We observed sequential, developmentally regulated activation and silencing of retrotransposons, in particular Class III endogenous retroviral (ERV) elements, in mouse oocytes and preimplantation embryos. Additionally, TEs acted as alternative promoters and first exons for diverse genes, synchronising their expression. Endogenous retroviruses are usually epigenetically silenced, and marked by cytosine methylation of CpG dinucleotides in their long terminal repeat (LTR) promoters, and so their activation and silencing indicates underlying epigenetic change. We propose that differential transposable element expression triggers sequential reprogramming of the embryonic genome during the oocyte-to-embryo transition and in preimplantation embryos. ERVs differ in the CpG content of their LTRs, and their expression does not perfectly correlate with global changes to cytosine methylation in oocytes and preimplantation embryos. In ongoing work, we are investigating the methylation status of individual unique genomic ERV loci to determine whether DNA methylation status is correlated with expression data. In addition we are using computational and experimental approaches to determine whether full-length ERVs are expressed from multiple dispersed loci or only from a specific subset. Our results address questions regarding changes in chromatin structure in the early embryo.

(54)

Structure and dispersion of two truncated Tvv1 grapevine retrotransposons resulting from illegitimate homologous recombination

*Sophie Blanc and Frédérique Pelsy**

Genetics and Plant Breeding Laboratory , INRA-Colmar, 28 rue de Herrlisheim, Colmar, F-68021, France

Tvv1, a Ty1 copia-like element, was originally reconstituted by a chromosome walking procedure from the grapevine genome. Elements belonging to this family share a highly conserved ORF and LTR sized between 149 and 157 pb. We later identified a full length Tvv1 copy from a BAC clone, named Tvv1-VB. The structural characteristics of Tvv1-VB are two perfect LTR 149-pb long and an ORF that any stop-codon interrupts. Moreover, two deleted elements, Tvv1-750 and Tvv1-1300, were isolated. Compared to Tvv1-VB, Tvv1-750 and Tvv1-1300 display deletions sized 3,460 bp and 3,001 bp, respectively. The deletion breakpoints were identified by boxes 11 and 13-pb long, respectively. In Tvv1-VB full-length element, each box corresponds to two short repeats nearly identical that flank the sequence corresponding to the deletion.
The distribution of these deleted elements within the Vitaceae has been studied. Most of the Vitis species display Tvv1-750, but only 11 V. vinifera varieties display Tvv1-1300. As it is present, Tvv1-1300 segregate in the progeny as a unique copy whereas Tvv1-750 never segregate.
Sequence analysis reveals that the Tvv1-1300 LTRs, 150-pb long, are very well conserved and very homologous to the Tvv1-VB LTRs. At least two copies of Tvv1-750 have been characterized. They display LTRs sized 166 and 170-pb, quasi-identical except the addition of 4 nucleotides, that are 77 % homologous to the Tvv1-VB LTRs.
These data lead us to believe that each deletion could result from an illegitimate homologous recombination event of an active copy. Tvv1-1300 could result from a more recent event than Tvv1-750 that is widely distributed and present in a larger number of copies.

(55)

Regulatory signals on Alu

Paz Polak, Eytan Domany*

Weizmann Institute of Science, Rehovot, ISRAEL

A large fraction of the human genomic \"dark matter \" is composed of ALU repeats, which were previously demonstrated, by several experimental studies, to act as cis-regulatory elements that bind several nuclear factors. Our computational studies showed that ALU harbors putative binding sites of many new proteins which were not previously identified as binding to ALU. A large portion of these proteins are developmental transcription factors. Four of these new putative binding sites we found were already verified, by previous experiments, that showed that the developmental TFs PITX2, GFI-1, NKX2.5 and LUN-1 bind to ALU sequences and, in some cases, regulate gene expression via these binding sites.. These four genes are involved in several development processes; GFI-1 regulates megakaryocyte differentiation; NKX2.5 , PITX2 are early markers of heart differentiation and LUN1 is involved in lung development. The case of LUN1 is particularly striking since all its known binding sites (50 examples) reside on ALU sequences. Our bioinformatic studies on 5k bp upstream of the transcriptional start site of 14000 genes revealed more than 100000 potential binding sites which reside on Alu repeats. Moreover, there are several transcription factors (PITX2, GFI-1, LUN-1 ,LXRE and others), for which over 95% of their respective putative upstream binding sites are located in some ALU sequence. In addition we predict a new list of DNA binding proteins that can bind to Alu.
Previous works observed that the distribution of ALU elements among genes that belong to various biological processes is not even.. We observed that promoters of genes that were related to protein biosynthesis were enriched with ALU sequences and, as consequence, they were rich with ALU mediated putative binding sites. In contrast, promoters of genes that were related to development were poor in ALU sequences and, as a result, they binding sites of several developmental TFs were poorly represented. This led us to propose a novel mechanism that take advantage on the developmental transcription factors binding sites that reside on ALU, to regulate differentiation and proliferation.

(56)

Maverick, a novel class of eukaryotic transposable elements related to double-stranded DNA viruses

Ellen J. Pritham and Cedric Feschotte*

The University of Texas at Arlington; Department of Biology

We recently identified a group of atypical mobile elements designated Mavericks from the nematodes, Caenorhabditis elegans and C. briggsae and the zebrafish, Danio rerio. Here we present the results of computer-assisted searches, which expand the distribution of these elements to a wide range of eukaryotes and indicate that their propagation involves an unprecedented transposition mechanism. Searches of the genome sequences available reveal that Mavericks are widespread in invertebrates and non-mammalian vertebrates but show a patchy distribution in non-animal species, being present in the fungus Glomus intraradices and in several single-celled eukaryotes such as the ciliate Tetrahymena thermophila, the stramenopile Phytophthora infestans and the trichomonad Trichomonas vaginalis, but not detectable in plants. This distribution and phylogenetic analysis of Maverick-encoded proteins is suggestive of an ancient origin of these elements followed by their frequent loss and/or horizontal transmission. Consistent with this later hypothesis, we identified a likely case of recent lateral acquisition of a Maverick into the Cotesia congreta virus genome from an insect host. Sequence analysis confirm that most Mavericks encode a retroviral-like integrase, but lack other open reading frames (ORFs) typically found in retroelements. Nevertheless, the length and conservation of the target site duplication created upon Maverick insertion (5 or 6 bp) is consistent with a role of the integrase-like protein in the integration of a double-stranded DNA transposition intermediate. Mavericks also display long TIRs but do not contain ORFs similar to proteins encoded by DNA transposons. Instead, Mavericks encode a conserved set of 5 to 9 genes (in addition to the integrase) predicted to encode proteins with homology to various proteins encoded by some bacteriophages and diverse eukaryotic double-stranded DNA viruses. Based on these and other structural similarities, we discuss a model for Maverick transposition and speculate that they represent a missing link between these disparate invasive DNA elements.

(57)

A combined evidence framework for the detailed annotation of transposable elements in genome sequences.

Hadi Quesneville[1], Casey Bergman[2], Nicolas Buisine[3], Olivier Andrieu[1], Delphine Autard[1], Danielle Nouaud[1], Christopher Smith[4], Gary Karpen[1], Vincent Colot[3], Michael Ashburner[5], Dominique Anxolabéhère[1]*

[1] Institut Jacques Monod, Paris, France
[2] University of Manchester, Manchester, UK
[3] URGV, Evry, France
[4] LBNL, Berkeley, USA
[5] University of Cambridge, Cambridge, UK

* e-mail: hq@ccr.jussieu.fr

We have developed a combined evidence-model transposable element (TE) annotation pipeline, analogous to systems used for gene annotation, for the TE annotation of D. melanogaster (Quesneville et al. 2005 PLoS Comput Biol. 2005 Jul;1(2):166-75). Whereas most TE annotation efforts rely on a single computational method, we combine evidence from multiple homology-based and de novo TE identification methods to promote a \"TE model\". Tests on the Release 3 D. melanogaster genomic sequence shows that a substantially larger fraction of the genome sequence is composed of TEs than previously estimated. The pipeline allows rapid and thorough annotation of even the most complex TE models, including highly deleted and/or nested elements such as those found in heterochromatic sequences. Our system is designed for use with the Apollo genome annotation tool, allowing automatic results to be manually curated to produce a reliable annotation. Following on the annotation of the euchromatic transposable elements in D. melanogaster genome (Release 4, www.flybase.org), we are now engaged in international collaborations to annotate TEs in the Arabidposis thaliana genome and the Drosophila peri-centromeric heterochromatin, in which TEs are a major component and may play a functional role. Our pipeline will produce the first detailed annotation of heterochromatic sequences in any organism. This will allow us to understand the relationships between TEs, repeats and heterochromatin from a functional and an evolutionary perspective.

(58)

Tracking Alu Evolution in New World Primates

David A. Ray [*1], *Dale J. Hedges* [2], *Erin W. Barnes* [2], *Cheney H. Huang* [2], *Justin D. Fowlkes* [2], *Mark A. Batzer*[2]

[1] Department of Biology, PO Box 6057, West Virginia University, Morgantown, WV 26505
[2] Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803

SINEs are excellent markers for use in phylogenetic, population and forensic studies. Identifying lineage specific mobile element subfamilies is an important first step in harnessing their power. Unfortunately, because the most well characterized Alu subfamilies are from the human genome, these have been the most often targeted for these types of analyses. Thus, most applications have been limited to humans and their nearest relatives, chimpanzees. To expand our understanding of Alu sequence evolution and to increase the applicability of these markers to non-human primate biology, we have analyzed Alu insertions specific to platyrrhine (New World) primates. Patterns along an Alu sequence phylogeny indicate three major classes of platyrrhine-specific Alu sequences. Sequence comparisons further reveal at least three New World primate-specific subfamilies; AluTa5, AluTa10, and AluTa15. Two of these subfamilies appear to be derived from a gene conversion event that produced a recently active fusion of AluSc- and AluSp-type elements. This is a novel mode of origin for new Alu subfamilies. We have also analyzed selected loci from the lineage leading to the marmosets (Callithrix). Initial results indicate that nearly all of the most recent insertions in the lineage are members of these newly described families. The characterization of these Alu subfamilies and confirmation of their recent activity not only increases our understanding of Alu sequence evolution in primates, but also opens the door to the application of these SINE markers to population genetic and phylogenetic questions outside the hominid lineage.

(59)

Methylation of HERV-E LTRs in placenta: comparison between co-opted and not co-opted members

Daphne Reiss and Dixie L. Mager*

Terry Fox Laboratory, BC Cancer Agency, Vancouver, BC, Canada and Dept. of Medical Genetics, University of British Columbia.

It is generally assumed that transposable elements are transcriptionally silenced by DNA methylation in mammalian somatic tissues. For example, in human, LINEs and SINEs are heavily methylated in somatic tissues, their demethylation resulting in activation of their transcriptional activity. However, several studies have shown that Long Terminal Repeats (LTRs) of Human Endogenous Retroviruses (HERVs) can act as alternative promoters for host genes in somatic cells, indicating that some LTRs are transcriptionally active, often in a tissue-specific manner. Strikingly, a significant portion of such LTR-derived gene promoters are active in placenta. Are the LTR-derived promoters hypomethylated specifically in tissues where they are active? If this is the case, is hypomethylation a general characteristic of HERV sequences in placenta? We are examining these questions by determining the methylation level of three HERV-E LTR-derived promoters that are active in placenta, namely those that act as promoters for the Midline1, Endothelin B Receptor and Pleiotrophin genes. We compared their methylation level in placenta to that in blood, where these LTRs are not active. Our results show that these three LTRs have very low levels of methylation in placenta and are heavily methylated in blood, as predicted according to their expression pattern. We also compared their methylation status with that of five closely related HERV-E LTRs with no known gene promoter activity. Although these randomly chosen LTRs present a variable level of methylation in placenta, they are more methylated than the three active LTRs. We are currently investigating the overall methylation level of HERVs in placenta and blood tissues. Our results indicate that methylation of individual HERV LTRs is highly variable, likely due to genomic context and tissue type. Moreover, these findings suggest that low levels of methylation increase the probability of LTRs being co-opted as alternative gene promoters.

(60)

Analysis of transposons from the Tc1-like family in fish genomes

A. Pocwierz-Kotus[1], A. Burzynski[1], R. Wenne[2]*

[1] Institute of Oceanology, PAS, Powstancow Warszawy 55, 81-712 Sopot, Poland
[2] Institute of Biology, University of Gdansk, Kladki 24, 80-822 Gdansk, Poland

A few pairs of primers were designed that enabled PCR amplification of fragment of transposase gene and whole Tc1 transposon sequence. Additionally, Southern hybridisation was performed between digested genomic DNA and transposase gene as molecular probe. These procedures enabled us to detect Tc1 in flounder, plaice, salmon, pike, land locked trout, round goby and perch genomes. We cloned Tc1-like sequence coming from genomes of some species: two individuals of flounder from Baltic Sea and one from North Sea, one plaice and one turbot from Baltic Sea, one pike from Zarnowieckie Lake, north Poland. PCR was performed for all species above. Products of this reaction were obtained from primers complementary to inverted repeats and afterwards were ligated to the SmaI site of pGEM3Zf(+) vector and transformed into Escherichia coli DH5□. Positive clones were identified on basis of □-complementation. Putative recombinant plasmids were screened among many colonies by two pairs of specific PCR primers. Then structures of these selected clones were verified by restriction analysis of plasmid DNA, performing single digestion with HindIII and double digestion with HindIII and EcoRI restriction enzymes. DNA of Tc1-like from selected positive clones was sequenced. Preliminary analysis of these sequences indicates that they are highly diverged both within particular species and between species. We observed that they were grouped into three main clads including in first one Tc1-like elements from plaice, flounder, turbot and in another one from pike and some sequences from turbot. Last clad contains the sequences from pike only. More detailed analysis of the sequence polymorphism of Tc1 is in progress.

(61)

Intra-Genomic Conflict and Evolution of Gene Silencing

Paul Schliekelman[*1] *and John F. McDonald*[2]

[1]Department of Statistics, University of Georgia, Athens, GA; [2]School of Biology, Georgia Institute of Technology, Atlanta, GA

It has been hypothesized that gene silencing originally evolved as a mechanism for defense of the genome against transposable elements (TEs) and other parasitic DNA. Although there is strong circumstantial evidence for the host defense hypothesis, little is known about the evolutionary dynamics of the interaction between transposable elements (TEs) and silencers. We have formulated a model of early genome evolution in order to explore these dynamics. In the model, genomes begin as short lengths of sequence with a few TEs and genes. TE replication increases genome length and when TEs insert in genes or other TEs, the target is made non-functional. Fitness is dependent on the number of functional genes. We show that pressure from TE replication creates moderate strength selection (selection coefficients on the order of a few percent) in favor of TE silencers. If TEs can occasionally escape from silencing, then cycles of TE activity readily develop due to an “arms race” between TEs and TE silencers. We also show that the host defense hypothesis provides a mechanism for large variation in genome length: because silencers appear only rarely, there is a large variance in the waiting time and therefore a large variance in the amount of genome expansion due to TE replication.

(62)

Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages

Shurjo K. Sen[*1], Kyudong Han[1], Jianxin Wang[2], Pauline A. Callinan[1], Jungnam Lee[1], Richard Cordaux[1], Ping Liang[2] and Mark A. Batzer[1]*

[1] Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-Scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA
[2] Department of Cancer Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA

Long INterspersed Elements (LINE-1s or L1s) are abundant non-LTR retrotransposons in mammalian genomes that are capable of insertional mutagenesis. They have been associated with target site deletions upon insertion in cell culture studies of retrotransposition. Here we report 50 deletion events in human and chimpanzee genomes directly linked to the insertion of L1 elements, resulting in the loss of ~18 Kb of sequence from the human genome and ~15 Kb from the chimpanzee genome. Our data suggest that during the primate radiation, L1 insertions may have deleted up to 7.5 Mb of target genomic sequences. While the results of our in vivo analysis differ from those of previous cell culture assays of L1 insertion-mediated deletions in terms of the size and rate of sequence deletion, evolutionary factors can reconcile the differences. We report a pattern of genomic deletion sizes similar to those created during the retrotransposition of Alu elements. Our study provides support for the existence of different mechanisms for small and large L1-mediated deletions, and we present a model for the correlation of L1 element size and the corresponding deletion size. In addition, we show that internal rearrangements can modify L1 structure during retrotransposition events associated with large deletions.

(63)

Transposable elements in the tammar wallaby genome

Katherine Thompson[1], Edda Koina[1], Matthew Wakefield[2] and Jennifer A. Marshall Graves[1]*

[1] The ARC Centre for Kangaroo Genomics, Research School of Biological Sciences, The Australian National University, Canberra, Australia, [2] Walter and Eliza Hall Institute, Melbourne, Australia

Transposable elements (TEs) are a major feature of mammalian genomes, representing at least 50% of the content in species investigated to date. Initially considered 'junk DNA', TEs have been proposed to serve regulatory functions such as X-Chromosome inactivation (XCI). Lyon (1998) proposed that LINE1 elements on the X chromosome in eutherians act to enable the spread and maintenance of XCI. A good correlation of LINE1 distribution and activity on the X chromosome in mice and humans supports this theory, although it does not necessarily follow that LINE1s cause XCI or maintain it. It might be that both reflect the time that the region has been isolated from recombination with the Y chromosome.

TEs have been well studied in a number of eutherians, but not marsupials. The benefit of studying the molecular mechanism of XCI in marsupials is that, while XCI occurs, it is different at the phenotypic and molecular levels from eutherian mammals, being paternal rather than random, incomplete and tissue-specific and does not involve methylation. It would therefore be interesting to examine the distribution of LINEs in marsupials.

We have undertaken an initial analysis of the TE content of the tammar wallaby (Macropus eugenii) genome. The sequence from all tammar BACs publicly available (15MB) has been acquired and analysed using Repeat masker. Results were also analysed using a script that produces a PDF output of the repeat content of a given sequence. The distribution of TEs on the tammar X-chromosome, examined by DNA FISH, showed a high concentration on the short arm and centromere, but a paucity on the distal region of the long arm.

1) Lyon, M.F.: X-Chromosome inactivation: a repeat hypothesis. *Cytogenetics and Cell Genetics* 80 (1998) 133 - 137.

(64)

Isolating short sequence length polymorphism of Alu 3' flanking sequences

Hans G. Thormar[1,2], Bjarki Gudmundsson[3], Gudmundur H. Gunnarsson[1,3],Magnus M. Halldorsson[4], Ymir Vigfusson[4], Haukur Thorgeirsson[4], Jon J. Jonsson[1,2].*

[1]Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Iceland, Reykjavik; Iceland. [2] Department of Genetics and Molecular Medicine, Landspitali–University Hospital, Reykjavík, Iceland. [3] BioCule, Reykjavik, Iceland. [4] Department of Computer Science, University of Iceland, Reykjavik, Iceland.

Approximately 1.2 million Alu repeats are distributed over the human genome. Small fraction contains short sequence length polymorphisms (SSLPs) in their adenine rich 3´ flanking sequences. We describe method for extracting Alu 3' flanking SSLPs based on two-dimensional conformation-dependent gel electrophoresis (2D–CDE) (1). The Alu 3´ flanking sequences were amplified using genetic material from ten individuals. The amplicons were pooled, denatured and renatured. Then mismatched heteroduplexes were isolated from perfectly matched homo- and heteroduplexes using 2D-CDE. Using this methodology, up to 80% of Alu 3´ flanking sequences extracted contained SSLPs when genotyped. We are designing microarray platform to use for large-scale studies of SSLPs in the Alu 3' flanking sequences. To achieve this we have designed virtual PCR web server CATTAGAT (http://genome.cs.hi.is). This web server can perform virtual PCR reactions on the entire human genome and is available under the terms of the GNU General Public License. This web server can simulate the complex Alu 3' flank amplification. Further software has been created allowing automatic microarray probe design based on sequences extracted using GATTAGAT. The methodology described here to extract and analyze SSLPs can probably be applied to other parts of the human genome as well as other genomes.

1) Gunnarsson, G.H., Thormar, H.G., Gudmundsson, B., Akesson, L. and Jonsson, J.J. (*2004) Nucleic Acids Res*, 32, e23.

(65)

Rice sines as molecular markers

Jian-Hong Xu, Chaoyang Cheng, Marcia Yuri Kondo, Suguru Tsuchimoto, Isaku Osawa, Eiichi Ohtsubo, and Hisako Ohtsubo*

IMCB, University of Tokyo

p-SINE1, p-SINE2 and p-SINE3 are SINE families of the cultivated rice (Oryza sativa) and other species of the Oryza genus. Consensus sequences of each families are highly homologous in their 5' end regions, but not in the 3' end regions. Their members are distributed on chromosomes in multiple copies. We identified a young p-SINE1 subfamily (named RA; Recently Amplified) whose members showed insertion polymorphism among O. sativa strains. We used the insertion polymorphism of the RA subfamily members as molecular markers to examine the phylogenetic relationship among strains of O. sativa and the wild species Oryza rufipogon. The strains were classified into several ecotypes by the markers, and the polyphyletic origin of O. sativa was strongly suggested. Some other p-SINE1 members and p-SINE3 members did not show insertion polymorphism within a species, but were polymorphic among seven Oryza species with the AA genome type. We showed that these species could be easily distinguished each other by the polymorphic pattern of these members. On the other hand, p-SINE2 members did not show insertion polymorphism within a genome type, but were polymorphic among different genome types. The genome types were distinguished each other by the polymorphic pattern of these members. These show that rice SINE members constitute a powerful tool for studying the classification and relationship of rice strains in ranks of the ecotype, the species, and the genome type, even when one has limited knowledge of morphological and physiological traits of rice strains.

References:

Cheng, et. al. (2003) *Mol. Biol. Evol*. 20: 67-75
Ohtsubo, et. al. (2004) *Breeding Sci*. 58: 1-11
Xu, et. al. (2005) *Genes Genet. syst*. 80:161-171

(66)

Transposons insert preferentially into heat-shock promoters of Drosophila melanogaster and contribute to adaptive regulatory variation

Jean-Claude Walser[1], Bing Chen[1] and Martin E. Feder[1,2]*

[1] Department of Organismal Biology & Anatomy, [2]The Committees on Evolutionary Biology, Genetics, and Molecular Medicine, The College, The University of Chicago, 1027 E. 57th Street, Chicago, IL 60637, USA

Expression of heat-inducible molecular chaperone-encoding genes, or heat-shock genes, is a universal, ancient, rapid, and massive response to proteotoxic stress. Indeed, in complex eukaryotes, the heat-shock genes exhibit a distinctive chromatin configuration and mode of regulation that poises them for rapid and massive expression. Previously, based on observations of one heat-shock gene (hsp70 of Drosophila), we hypothesized that these distinctive features made heat-shock genes especially susceptible to transposon insertion into their proximal promoters, and that this transposition is a source of variation on which natural selection could act. Here we test this hypothesis by screening for transposons in the proximal-promoter regions of hsp70, 18 additional heat-shock protein/chaperone-encoding genes, 18 non-hsp genes resembling hsp70 in transcriptional regulation and inducibility, and 18 randomly-selected non-hsp-like genes in a worldwide survey of 48 Drosophila populations. The screen revealed at least 162 discrete transposon insertions in the proximal promoters of the 55 genes surveyed. Strikingly, as a group and on a worldwide basis, heat-shock genes clearly exceed other genes in transpositions into the proximal promoter, as hypothesized. Re-analysis of the Berkeley Drosophila Gene Disruption Project\'s genome-wide transposition essentially recapitulates this finding. Unexpectedly, out of the 120+ transposon families in D. melanogaster, almost all naturally-occurring transposons in heat-shock promoters are P elements. Our study indicates that P elements are still remarkably active in the wild and that they produce regulatory variation, which contribute to the adaptive modulation of heat-shock gene expression in natural populations of D. melanogaster. Supported by NSF grant 03-16627.