3rd INTERNATIONAL CONFERENCE/WORKSHOP

Genomic Impact Of Eukaryotic Transposable Elements

FEBRUARY 24 – 28, 2012

ASILOMAR, PACIFIC GROVE, CALIFORNIA, USA

3rd INTERNATIONAL CONFERENCE/WORKSHOP

Genomic Impact Of Eukaryotic Transposable Elements

Organizer:

Jerzy Jurka

Genetic Information Research Institute
Mountain View, California, USA

ASILOMAR 2012

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.

3rd International Conference and Workshop

Genomic Impact of Eukaryotic Transposable Elements

Organizer: Jerzy Jurka

Friday, February 24, 2012

15:00-18:00	REGISTRATION (Fred Farr Forum)	
18:00-19:00	*Dinner* (Crocker Dining Hall)	
19:30-23:00	Warm-up party/poster previews/preparation of audio-visual (Fred Farr Forum/Kiln)	
	The following equipment will be provided in all sessions: an LCD projector, laser pointer and a microphone. Speakers should load their talks at Fred Farr Forum in the evening preceding their presentations. There will be a limited time for last-minute testing (30 min. before the morning session and during breaks). Due to time constraints, all 10-minute talks should be limited to communication of your specific results only. Please, leave 2-3 minutes from your allowed time for discussion. The projected discussion time for 15-minute presentations is 3-4 minutes, and for 25-minute presentations it is 4-5 minutes. The opening speaker for each session is the chairman of that session.	

Saturday, February 25, 2012

7:30-8:30	*Breakfast* (Crocker Dining Hall)	
8:00-9:00	REGISTRATION (Kiln)	
9:00-9:10	**Jerzy Jurka** – Opening remarks	
9:10-9:35	**David Haussler** – Aspects of the evolutionary impact of retrotransposons on vertebrate genomes	p.1
9:35-10:00	**Norihiro Okada** – Mammalian exaptation burst	p.2
10:00-10:15	**Peter Arndt** – A neutral model to explain fat tails in match length distributions	p.3
10:15-10:45	*Coffee-break* (Fred Farr Forum)/**Group photo**	
10:45-11:10	**Nancy Craig** – A transposase goes to work	p.4
11:10-11:35	**Juergen Brosius** – BC1 RNA, the significance of a tRNA-derived retrogene for the rodent nervous system	p.5
11:35-12:00	**Gill Bejerano** – Cis regulatory co-option in the human genome	p.6
12:00-13:00	*Lunch* (Crocker Dining Hall)	
13:30-13:55	**Mark Batzer** – The primate mobilome	p.7
13:55-14:05	**Lucia Carbone** – Centromeric activation of a novel lineage-specific composite transposable element in the eastern hoolock gibbon (Hoolock leuconedys)	p.8
14:05-14:30	**Juergen Schmitz** – Traces of the past: what retrotransposons tell us about ancient times	p.9
14:30-14:55	**Andrew Shedlock** – Testing alternative models of TE molecular evolution with amniote phylogenomics	p.10

14:55-15:05	**Cesar Martins** – Transposable elements in fish genomes: a chromosome perspective	p.11
15:05-15:20	**Ben Koop** – Repeat families in the rediploidization and speciation of Salmonids	p.12
15:20-15:50	*Coffee-break* (Fred Farr Forum)	
15:50-16:15	**King Jordan** – Human MIRs as chromatin organizing elements	p.13
16:15-16:40	**Matthew Lorincz** – Histone H3K9 writers and readers in ERV silencing: a family affair	p.14
16:40-16:50	**Lucas Gray** – The conserved piggyBac transposase fusion protein CSB-PGBD3 collaborates with AP-1 proteins to regulate nearby genes in primates	p.15
16:50-17:00	**Christoffer Nellaker** – The impact of transposable element variants on mouse genomes and genes	p.16
17:00-17:10	**Nickolai Tchurikov** – Genome-wide profiling of fragmentation sites in Drosophila melanogster chromosomes revealed a strong correlation between fragmentation sites, particular sets of mobile elements and regions of intercalary heterochromatin	p.17
17:10-17:20	**Yanzhu Ji** – Comparative analyses of transposable elements expressed in the transcriptomes of lake sturgeon (Acipenser fulvescens), tiger salamanders (Ambystoma tigrinum), and banner-tailed kangaroo rats (Dipodomys spectabilis)	p.18
17:20-17:30	**Michael Wilson** – Latent regulatory potential of human-specific repetitive elements	p.19
17:30-17:40	**Heather Murton** – Chromatin and the control of LTR retrotransposon silencing and mobilisation in fission yeast	p.20
18:00-19:00	*Dinner* (Woodlands)	
19:20-19:45	**John Moran** – Similarities between the human LINE-1 reverse transcriptase and telomerase	p.21
19:45-19:55	**Fabio Macciardi** – The genetic signatures of transposable elements (TE) in schizophrenia	p.22
19:55-20:05	**Nataša Lindič** – Trying to pin down the mechanism of APOBEC3s inhibition of retrotransposition	p.23
20:05-20:15	**Koichi Ishiguro** – The genome-wide profiling of L1 antisense promoter activity in human cells	p.24
20:15-20:25	**Kyle Upton** – Technical development of Retrotransposon Capture sequencing (RC-seq)	p.25
20:30-23:00	*Happy Hour* / **Poster session** – odd numbers (Kiln)	

Sunday, February 26, 2012

7:30-8:30	*Breakfast* (Crocker Dining Hall)	
9:00-9:25	**Haig Kazazian** – Transcriptome-wide binding of human L1 ORF1 protein reveals its role in Alu retrotransposition and processed pseudogene formation	p.26
9:25-9:50	**Gerald Schumann** – LINE-1 mediated trans-mobilization of human-specific SVA retrotransposons is SVA structure-dependent	p.27
9:50-10:00	**Annette Damert** – Competition for SVAs in gibbons: LAVA and other VNTR containing non-LTR retrotransposons	p.28
10:00-10:25	**Anthony Furano** – Polymer formation and nucleic acid binding properties of the human L1 non-LTR retrotransposon ORF1p protein	p.29
10:25-10:45	*Coffee-break* (Fred Farr Forum)	
10:45-11:10	**Holly Wichman** – An overview of LINE-1 activity in mammals	p.30
11:10-11:35	**Cedric Feschotte** – Genomes without borders – the misfit origins of genetic novelty	p.31
11:35-11:45	**Petra Schwalie** – Waves of repeat driven CTCF binding expansions have shaped mammalian genomes	p.32
11:45-11:55	**Ray Malfavon-Borja** – Ancient and recurrent evolution of the antiviral gene fusion TRIMCyp in primates	p.33
12:00-13:00	*Lunch* (Crocker Dining Hall)	

13:30-13:55	**Prescott Deininger** – Alu/Alu non-allelic homologous recombination	p.34
13:55-14:20	**Scott Devine** – Studying Alu and L1 mutagenesis in human genomes with DNA sequencing technologies	p.35
14:20-14:30	**Adam Ewing** – Large-scale retroelement detection from the Cancer Genome Atlas	p.36
14:30-14:40	**Robyn Leary** – A novel p53 regulated murine endogenous retrovirus	p.37
14:40-14:50	**Krassimira Botcheva** – Distinct genome-wide p53 binding profile in normal and cancer-derived human cells	p.38
14:50-15:00	**Elena Helman** – RetroSeq: a tool to discover somatic insertion of retrotransposons	p.39
15:00-15:25	**David Symer** – Mouse endogenous retroviruses can trigger premature transcriptioal termination at a distance	p.40
15:25-15:55	*Coffee-break* (Fred Farr Forum)	
15:55-16:20	**Dixie Mager** – Complex interactions between endogenous retroviruses and host genes	p.41
16:20-16:30	**Andrea Schorn** – Transposon small RNA expression in the embryonic and trophectoderm lineage	p.42
16:30-16:40	**Justin Blumenstiel** – Dynamics of TE control by the piRNA machinery: The Key Role of Dose	p.43
16:40-16:50	**Erin Kelleher** – Drosophila interspecific hybrids phenocopy piRNA pathway mutants in aberrant piRNA production and TE derepression	p.44
16:50-17:00	**Fernando Rodriguez** – An increase in relative abundance of pi-like RNAs in response to ionizing radiation in the bdelloid rotifer Adineta vaga	p.45
17:00-17:10	**Mireia Jordà** – Epigenetics speaks up for the silent DNA	p.46
17:10-17:35	**Zsuzsanna Izsvak** – Modeling stress signaling and response resulting in transposon activation in human cells	p.47
17:35-17:45	**Hidetaka Ito** – Transgenerational effects and genomic impacts in environmental stress	p.48
19:00-22:30	*Dinner at Monterey Bay Aquarium* (buses depart at 18:45 from Phoebe A. Hearst Social Hall)	

Monday, February 27, 2012

7:30-8:30	*Breakfast* (Crocker Dining Hall)	
9:00-9:25	**Damon Lisch** – We have met the enemy and he is us: co-regulation of transposons and genes during plant development	p.49
9:25-9:40	**Marie-Angèle Grandbastien** – Retroviral-type LTRs as intermediate of the stress response in tobacco	p.50
9:40-9:50	**Darrel Lizamore** – The effect of environmental stress events on the mobility of four LTR retrotransposon families in grapevine (Vitis vinfera) somatic embryo cultures	p.51
9:50-10:05	**Marie-Anne Van Sluys** – Plant LTR-RT study reveals fine-scale individual molecular patterns	p.52
10:05-10:30	**Susan Wessler** – Genome-wide impact of a MITE burst in rice after just 20 generations	p.53
10:30-10:50	*Coffee-break* (Fred Farr Forum)	
10:50-11:15	**Irina Arkhipova** – Rotifer genomes as a source for discovery of novel types of TEs and TE-related genes	p.54
11:15-11:25	**Eugene Gladyshev** – Biochemical properties of NcRVT protein encoded by a reverse transcriptase-like gene from Neurospora crassa	p.55
11:25-11:35	**Ken Kraaijeveld** – Transposon proliferation in an asexual parasitoid	p.56
11:35-11:45	**Jens Bast** – Comparing the transposable element load in sexual and asexual oribatid mites using whole genome information	p.57
11:45-11:55	**Sarah Schaack** – Transposable Element Copy Number in Sexual and Asexual Daphnia	p.58
12:00-13:00	*Lunch* (Crocker Dining Hall)	
13:20-13:45	**Dmitri Petrov** – Population genetics of transposable elements in Drosophila	p.59
13:45-14:10	**Cristina Vieira** – Drosophila endogenous retrovirus regulation in natural populations	p.60

14:10-14:25	**Josef Gonzalez** – Adaptive TE insertions in Drosophila	p.61
14:25-14:35	**Claudia Carareto** – Ancestral polymorphism and re-introduction of transposable elements in Drosophila	p.62
14:35-14:45	**Grace Yuh Chwen Lee** – Long-term and short-term evolutionary impacts of transposable elements on Drosophila	p.63
14:45-14:55	**Anna-Sophie Fiston-Lavier** – Transposaable element discovery and annotation in population using next-gen sequencing data	p.64
14:55-15:05	**Rita Daniela Fernández-Medina –** RNAseq reveals abundant transposable element expression in the mosquito Anopheles funestus	p.65
15:05-15:35	*Coffee-break* (Fred Farr Forum)	
15:35-16:00	**Eugene Koonin** – The multilayer eukaryotic mobilomes: viruses as mobile elements and mobile elements of giant viruses	p.66
16:00-16:25	**Marcella McClure** – The mutualism continuum of retroids	p.67
16:25-16:50	**Vladimir Kapitonov** – "Simple" DNA transposons: scratching the surface	p.68
16:50-17:15	**Arian Smit** – Digging deeper in time; approaches to study ancient transposable elements	p.69
17:15-17:25	**Sarah Oliveira** – Transposable elements: maintenance of centromeric regions and chromosome B origin	p.70
17:25-17:35	**Erez Levanon** – Large scale DNA editing of retrotransposons accelerate mammalian genome evolution	p.71
17:35-17:45	**Julia Arand** – High-throughput DNA methylation analysis of repetitive elements by hairpin bisulfite sequencing	p.72
18:00-19:00	*Dinner* (Woodlands)	
19:30-19:55	**Wojciech Makałowski** – Repetitive sequences: beyond RepBase nd RepeatMasker	p.73
19:55-20:05	**Travis Wheeler** – Improved transposable element detection with profile HMMs in nhmmer	p.74
20:05-20:15	**David Pollock** – Transposable element diversification and evolution	p.75
20:15-20:25	**Casimir Bamberger** – Discovery and analysis of transposable elements with shotgun proteomics	p.76
20:30-23:00	Happy Hour / **Poster session** – even numbers (Kiln)	

Tuesday, February 28, 2012

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

POSTER PRESENTATIONS

Author Index

SPEAKER ABSTRACTS

Aspects of the evolutionary impact of retrotransposons on vertebrate genomes

David Haussler

Center for Biomolecular Science and Engineering, UC Santa Cruz, California, USA

Comparing the genomes of present-day species allows us to computationally reconstruct what most of the DNA bases in the genome of the common ancestor of placental mammals must have looked like approximately 100 million years ago. In many important functional regions, we can reconstruct genome sequences even further back in history. We can then deduce the genetic changes on the evolutionary path from those ancient species to humans. In so doing, we discover how natural selection has shaped us at the molecular level. About five percent of the human genome consists of conserved elements that have remained surprisingly unchanged across millions of years of evolution, suggesting important function. Only one third of these code for protein; the rest are likely to be gene regulatory elements and non-coding RNAs.

In most parts of the genome we find that the protein-coding sequences have changed relatively little during the last ~500 million years of vertebrate evolution, and many have orthologous counterparts in invertebrates. But these represent less than 1.5% of the genome. In contrast, even the “ultraconserved” vertebrate non-coding genomic elements for the most part bear little resemblance to anything we find in invertebrates. Apart from a few highly conserved elements, the non-coding sequences that inhabit introns and intergenic regions between genes in most regions appear to have undergone a complete turnover due to the activity of transposons that add new DNA by inserting into novel locations and facilitate the removal of DNA through the process of non-homologous recombination. We conjecture that most of the non-coding regulatory elements in the genomes of living vertebrates derive from DNA that was put into place by transposons during vertebrate evolution, and hence these vertebrate regulatory elements have no orthologous counterparts in invertebrates.

Examples where ancient transposons have contributed functional regulatory sequences are being discovered at an increasing rate, suggesting that transposons have played a fundamental role in our molecular evolution. In recent work of Craig Lowe, more than 3 million conserved, non-exonic elements (CNEEs) in the vertebrate genomes were used to track evolutionary changes over the full range of vertebrate evolution. This analysis pointed to three periods in the evolution of gene regulatory networks through the introduction of new CNEEs: first a period of innovations in fundamental transcription factors and developmental genes, then an increase in innovations in genes involved in cell-cell communication, and finally, in the last 100 million years or so, an increase in innovations in genes involved in key signaling pathways within cells. These results illustrate new aspects of vertebrate genome evolution revealed by powerful new comparative genomics methodologies.

Exaptation linked to P-T mass extinction

Norihiro Okada

Tokyo Institute of Technology

P(Permian)-T(Triassic) mass extinction is the most extensive extinction we have ever experienced on the earth. This occurred 250 Ma (million years ago). More than 96% living creatures were dead at that time. Since reptiles and mammals diverged about 310 Ma, ancestors of these two lineages struggled to survive after this event. The most extensive environmental change occurred at the boundary of P-T mass extinction is the decrease of oxygen concentration. It is believed that it was up to 30% during the Permian era but after this boundary it decreased to 10%. Ancestors of reptiles and mammals should have adapted to this superanoxia.

For these 3 years, our group has asked what happened on the DNA level for mammals to adapt to this extreme environment. We discovered that more than 100 loci of AmnSINE1 were exapted possibly for this adaptation. We have now three concrete examples. One locus function as an enhancer for fgf8, which is involved in patterning of barrel field in thalamus, one relay station of somatosensory system in mammals. The second functions as an enhancer for Satb2, which is involved in axon formation of corpus callosum. Corpus callosum is specific to mammals, being involved in interhemispheric communication between the left and right cerebral hemispheres. The last example functions as an enhancer for Wint5a, which is involved in closure of the secondary palate of our jaw. Secondary palate is responsible for separation between the nasal and oral cavities, and is assumed to have evolved just after P-T mass extinction for our ancestor to obtain more efficient respiration. I am confident with the notion that this example is one of the most typical ones of exaptation which links to geological events.

A neutral model to explain fat tails in match length distributions

Peter F. Arndt

Max Planck Institute for Molecular Genetics

Over millions of years various evolutionary mechanism left their mark on present day genomes. Today we observe the evolutionary signature of processes like nucleotide mutations, insertion and deletion of sequence segment, the insertion of transposable elements, large scale rearrangements, and whole genome duplications. One aim is to discover and disentangle biochemical processes and evolutionary forces that change genetic information. Recently, an unexpected power law with exponent -3 has been found in the length distribution of exact sequence matches along the human genome itself or in a comparison of closely related genomes. Here we propose a neutral model which is able to explain this phenomenon.

A transposase goes to work

Nancy L Craig*, Peter W Atkinson and Cedric Feschotte

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

Transposable elements are present in virtually all organisms and have tremendous impact on genome structure, function and evolution. Some transposases have been "domesticated" to fufill cellular functions. Notable examples include the RAG recombinase that mediates the assembly of immunoglobulin genes in the immune system, transposase-like proteins that act as transcription factors in regulatory networks and the reverse transcriptases of telomerases that that cap chromosome ends.

We are interested in DNA cut & paste transposases that promote transposition by the introduction of double strand breaks at the transposon ends to excise the element from the donor site and then mediate the joining of the transposon ends of the target DNA.

We have isolated several new high active transposons from the insect Tribolium castaneum insect and have resurrected a mammalian transposon that was a highly active element at likely invaded a number of genomes by horizontal transfer. Our studies on these elements will be discussed.

BC1 RNA, the significance of a tRNA-derived retrogene for the rodent nervous system

Brosius, Juergen

Institute of Experimental Pathology, ZMBE, University of Münster, Münster, Germany

Transfer RNAs can be relatively efficient templates for retroposition, generating as many as several hundred retrocopies in vertebrate genomes. Despite the presence of internal box A and box B promoter elements for RNA polymerase III, in vivo the majority of such copies are transcriptionally silent. Fortuitously, a tRNA-Ala retrocopy must have integrated into a locus that provided immediately or after additional nucleotide changes an upstream promoter element that allowed efficient and cell type specific expression of this copy yielding a gene encoding neuronal BC1 RNA. As the retroposition event and its exaptation occurred after the mammalian radiation but prior to the diversification of rodents, the gene with an age of up to 70-80 million years, is relatively young. As an aside, unlike its tRNA ancestor, BC1 RNA is (along with a few more transcribed ID elements) an extraordinarily efficient template for retroposition yielding tens to about one hundred thousand ID SINEs in rodents. Since the BC1 RNA gene has been under purifying selection for at least 55 million years, what are, apart from previously noticed mild changes in exploratory behavior, the deficits in mice lacking this young non-protein coding RNA gene?

Cis regulatory co-option in the human genome

Gill Bejerano

Stanford University, California, USA

The primate mobilome

*Mark A. Batzer**

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

Primate mobile elements (SINEs, LINEs and SVA elements) 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. The distribution of Alu, L1 and SVA elements throughout various primate genomes makes them useful tools for resolving population genetic relationships and non-human primate phylogenetic relationships. We have also characterized the structural genomic variation associated with the insertion of recently integrated mobile elements in primate genomes along with post insertion recombination based events. These genomic deletions are yet another source of mobile element associated genetic variation within the primate lineage.

Centromeric activation of a novel lineage-specific composite transposable element in the eastern hoolock gibbon (Hoolock leuconedys)

Lucia Carbone (1,2), R. Alan Harris (3), Alan R. Mootnick (4), Aleksandar Milosavljevic (3), Miriam Konkel (5), Mark A. Batzer (5), Mariano Rocchi (6), Nicoletta Archidiacono (6), David IK Martin (1) and Pieter J. de Jong (1)*

1 Children's Hospital Oakland Research Institute, Oakland, CA, USA
2 Oregon Health and Science University, Portland, OR, USA
3 Baylor College of Medicine, Houston, TX, USA
4 Gibbon Conservation Center, Santa Clarita, CA, USA
5 Louisiana State University, Baton Rouge, LA, USA
6 University of Bari, Italy

The small apes or gibbons split from the hominoids common ancestor about 15-18 million years ago, and since then they have undergone rapid chromosomal evolution. The analysis of chromosomal breakpoints from the northern white-cheeked gibbon (Nomascus leucogenys, NLE) revealed an association between the breakpoints and hypomethylated Alu elements. We therefore speculate that the epigenetic repression of transposable elements might have been disrupted during the evolution of the small apes, therefore increasing the rate of chromosome rearrangements in these species (Carbone et al. 2009).

We recently observed another phenomenon which supports this hypothesis. A repeated element expanded in almost all the centromeres of another gibbon species, the eastern hoolock gibbon (Hoolock leuconedys, HLE). The investigation of this phenomenon led us to discover a novel composite transposable element composed by portions of three different repeats recognized by Repeat Masker as L1M5, AluSz6 and SVA. We showed that the new composite element is not present in the genome of human, other great apes and rhesus, representing a gibbon specific evolutionary novelty. A detailed analysis of full length copies of the composite repeat and flanking sequences from fully sequenced gibbon BACs, revealed the presence of direct repeats indicating that this element has been transposing in the gibbon genome. Interestingly, the centromeric repeat expansion is associated with a paucity of alpha-satellite, the main centromeric satellite in primates.

The expansion of L1M5-AluSz6-SVA in the hoolock centromeres closely resembles the massive centromeric expansion of the K-ERV retroelement reported for wallaby (marsupial) interspecific hybrids, which also displayed a globally hypomethylated genome (O'Neill et al. 1998). Both phenomena are an example of chromosome remodeling arising from loss of regulation of transposable elements. Remarkably, in the hoolock gibbon this change has been fixed through evolution and it is not shared by the other genera.

Traces of the past: What retrotransposons tell us about ancient times

Juergen Schmitz

Institute of Experimental Pathology, University of Münster, Germany

Retrotransposons influenced the evolution of genomes throughout the entire eukaryotic tree of life. Ancient waves of retroposon insertion today provide a precise view of the relationships among diverged species. Because of their potentially unlimited number of genomic insertion sites and the complexities of the insertion process, parallel insertions or exact deletions are very rare and, therefore, render retroposon presence/absence data nearly homoplasy-free. On the other hand, apparently conflicting patterns are unmistakable signals of ancestral incomplete lineage sorting or hybridization effects. Reconstructed phylogenies based on retrotransposed insertions revealed that:

- Lemurs invaded Madagaska and Asia from Africa.
- The Koboldmaki share ancestry with higher primates.
- The flying lemur is not a lemur.
- The living rodent “fossil” Laonastes is a relative of the gundi clade in the rodent tree.
- The dog is our closest friend, but not our close relative.
- Marsupials migrated from South America via Gondwana to Australia.
- Early placental orders diverged simultaneously.
- Platypus represents the oldest divergence of mammals.
- Vocal learning in birds evolved during the dinosaur age.

Testing alternative models of TE molecular evolution with amniote phylogenomics

Andrew M. Shedlock

College of Charleston Biology Department and Graduate Program in Genomics

Genomic information from Reptilia, the sister group of mammals, is needed in order to place the many publicly available mammalian and several bird genome sequence assemblies into meaningful molecular evolutionary context. Results of global repeat content and diversity from the recently completed squamate genome assembly for Anolis plus new data for the first turtle genomes as compared to other large-scale reptilian sequences will be discussed in relation to contrasting models of transposable element (TE) impacts on amniote genome structure, function and diversification. Comparative data for these initial reptile assemblies provide a first genome-wide glimpse of the evolutionary dynamics of mobile repeats that remain invisible to mammalian and avian genomic investigations. Goals for integrating studies of amniote TEs with our need to build a predictive theory for the eukaryotic genome, presently still lacking in comparative biology, will be discussed in light of recent discoveries and new technological developments driving the field of evolutionary genomics.

Transposable elements in fish genomes: a chromosome perspective

Cesar Martins

UNESP - São Paulo State University, Institute of Bioscience, Botucatu, SP, Brazil

Transposable elements (TEs) are major components of eukaryotic chromatin and their investigation has been hindered even after the availability of whole sequenced genomes because their complex organization and distribution in the nucleus. The physical mapping of TEs in the chromosomes through molecular cytogenetic tools has proved to be of high value to integrate information from the micro level of the nucleotide sequence to a macro view in the chromosome level. TEs seem to be more abundant in heterochromatin (mostly in the pericentromeric region) in fish genomes probably as a consequence of the (i) low recombination rates occurring in this region, (ii) their elimination from gene-rich regions because of their potential deleterious effects when inserted within genes, or (iii) they could accumulate in heterochromatin as a consequence of their functional involvement in the maintenance of specific genomic areas such as the pericentromeric and telomeric regions. Besides the accumulation of TEs in particular chromosome areas, dispersed non-clustered copies seem to be common in fish genomes. Special attention must be exercised concerning the relationship between the data obtained through the cytogenetic mapping of TEs and the data provided through the complete sequencing of genomes. The dispersed signals of TEs revealed through the classical molecular cytogenetic analysis do not represent dispersed single copies, but small clusters of at least few copies of the DNA element. One interesting feature of the chromosomal distribution of TEs among fishes is that different TEs share a similar general pattern of arrangement over the chromosomes of a specific taxonomic group. It seems that the TEs accumulate in specific genomic regions not as a consequence of the intrinsic nature of the element instead as a consequence of properties of the chromatin surrounding their tag genomic region.

Financial Support: FAPESP – São Paulo Research Foundation, Brazil

Repeat families in the rediploidization and speciation of Salmonids

*David Minkley and Ben F. Koop**

Department of Biology, University of Victoria, Victoria BC

Salmonids are a rapidly evolving species complex (67 species including grayling, ciscos, whitefish, trout, char, and salmon) which are > 92% similar at the noncoding DNA level. Nearly all have significant economic, ecological, conservation, environmental and societal value. The common ancestor of salmon and trout experienced a whole genome duplication (WGD) about 60 MYA, and modern salmonids are considered pseudotetraploid as they are in the process of reverting to a stable diploid state. This makes them ideal organisms for examining some of the early impacts of genome and gene duplications, processes thought to have had pivotal roles in generating genetic diversity, novel functions, and functional specialization. After a WGD, an increased repeat element mobility would likely impact genome stabilization, rediploidization as well as contribute to a salmonid speciation burst. A previous study of 100's of salmon DNA transposons (Tc1 -1.5kb) found waves of recent TE expansion with pairwise identity of ~94-99%, similar to levels found between salmonid species. Additional work now suggests that the Atlantic salmon genome is comprised of over 50% repeats. This causes serious problems for the sequence assembly of an ongoing salmon genome project.

We are currently identifying all of the repeat families along with basic information such as frequencies, terminal repeats, internal structure, length, intra and interfamily similarity values, and consensus sequences, and building a curated salmonid repeat database. Repeat information is required for an assembly of a high-quality genome sequence. It is also be essential for evaluating TE involvement in rediploidization or speciation processes as determined by TE presence at the boundaries of insertions/deletions and rearrangements between homeologues as well as between the diploid pike and the pseudotetrapoloid salmonid genomes. As new salmonid genome sequences arise, a more complete picture of the role of repeats in genome stabilization and speciation will emerge.

Human MIRs as chromatin organizing elements

Jiarong Wang (1), Lluis Montoliu (2), Victoria V. Lunyak (3) and I. King Jordan (1)*

1 School of Biology, Georgia Institute of Technology, Atlanta, GA, USA
2 Department of Molecular and Cellular Biology, Centro Nacional de Biotecnología, Madrid, ES
3 Buck Institute for Research on Aging, Novato, CA, USA

Eukaryotic chromatin is organized into functionally distinct domains by insulator sequences that encode chromatin barrier and enhancer blocking activities. Recent studies indicate that tRNA-derived sequences provide insulator activity in organisms ranging from yeast to human. We developed and applied a bioinformatics screen to search for human insulators that are encoded by the MIR family of transposable elements. MIRs (mammalian-wide interspersed repeats) are an ancient family of tRNA-derived transposable elements that have been implicated as non-coding regulatory elements by virtue of their anomalous levels of sequence conservation. Our feature-based algorithmic screen for the identification of MIR-derived insulators employed a joint analysis of genome sequence, transcription factor binding and histone modification data from CD4+ T cells. Application of this screen uncovered numerous examples of MIRs that bear a combinatoric feature profile consistent with tRNA-related insulators and also partition active versus repressive chromatin domains. The ability of selected MIR-insulators to block regulatory interactions between adjacent chromatin domains was validated using both in vitro and in vivo enhancer blocking assays. The potential functional significance of the predicted MIR-insulators was further confirmed by the analysis gene expression and gene ontology functional annotations. Genes that flank MIR-insulators are differentially regulated in a cell-type specific manner, and genes protected by MIR chromatin barrier activity are enriched in the T-cell receptor signaling pathway. For example, a co-located genomic cluster of three T-cell receptors—CD28, CTLA4 and ICOS—are protected from the spread of repressive chromatin by MIR-insulators to allow for CD4+ T cell specific expression. MIR-insulator chromatin barrier activity at this and numerous other loci is highly tissue-specific. Thus, MIR-insulators provide an example of formerly selfish transposable element sequences that have been domesticated to provide regulatory elements that help to establish cell-type specific gene expression programs.

Histone H3K9 writers and readers in ERV silencing: a family affair

Mohammad Karimi (1), Irina Maksakova (1), Preeti Goyal (1), Misha Bilenky (2), Martin Hirst (3) and Matthew Lorincz (1)*

1 Department of Medical Genetics, The University of British Columbia
2 British Columbia Cancer Agency, Genome Sciences Centre
3 Department of Microbiology and Immunology, The University of British Columbia

DNA methylation plays an important role in silencing of specific genes and retroelements in somatic cells. In mouse embryonic stem cells (mESCs) however, several subclasses of class I and class II endogenous retroviruses (ERVs) are maintained in a silent state by the histone H3 lysine 9 (H3K9) methyltransferase Setdb1/Eset, regardless of the presence of DNA methylation, indicating that these repressive pathways can function independently (1,2). ERVs up-regulated in the SETDB1 KO line concomitantly lose H3K9me3 and catalytically active Setdb1 is required for robust silencing of these elements, implicating a direct role for this mark in transcriptional silencing. Curiously however, we find that the known H3K9me "readers", including HP1α, HP1β, and HP1γ are not required for silencing of these ERVs (3). I will summarize this work and describe our recent studies on the roles of alternative histone modifying enzymes and H3K9 readers in the silencing of Class III ERVs in mESCs. Finally, I will describe an RNAseq-based approach that we employed to identify novel "domesticated" LTR promoters in mESCs and the evolutionary and functional characterization of a subset of these tissue-specific chimeric transcripts.

References:

1. Matsui, T. et al. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature 464, 927-931 (2010).
2. Karimi, M. M. et al. DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. Cell Stem Cell 8, 676-687 (2011).
3. Maksakova, et al. H3K9me3 binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. Epigenetics & Chromatin, 4:12 doi:10.1186/1756-8935-4-12 (2011)

The conserved piggyBac transposase fusion protein CSB-PGBD3 collaborates with AP-1 proteins to regulate nearby genes in primates

Lucas Gray, Tom Pavelitz, Arnold Bailey, Kim Fong, Alan Weiner*

University of Washington Department of Biochemistry

The CSB-PGBD3 fusion protein arose more than 43 Mya when the PGBD3 transposon integrated into intron 5 of the Cockayne syndrome group B gene (CSB) in the common ancestor of simian primates. The PGBD3 transposase ORF is flanked by a 3' splice acceptor site upstream and a polyadenylation site downstream, enabling it to be expressed as an alternative 3' terminal exon. As a result, the CSB genes of all higher primates, from marmoset to human, now generate three proteins by alternative splicing and polyadenylation: full length CSB protein, a CSB-PGBD3 fusion protein that fuses the N-terminus of CSB to a complete piggyBac transposase, and solitary piggyBac transposase driven by a cryptic promoter in CSB exon 5. CSB-PGBD3 is remarkably well conserved, and continues to be expressed in all simian primates from marmoset to human. PGBD3 also gave rise to a family of ~900 internally deleted MITEs known as MER85s, which are dispersed throughout the human genome. To determine what the CSB-PGBD3 fusion protein does in normal cells, why it has been conserved, and whether it affects the severity or nature of Cockayne syndrome, we have used genome-wide ChIP-seq and expression array analysis to determine the role of CSB-PGBD3 in transcription. We found that CSB-PGBD3 binds strongly to MER85 elements as expected, but also found surprising interactions of the CSB-PGBD3 fusion protein with sites bound by the AP-1 family transcription factors Jun and Fos, which are potent proto-oncogenes. These interactions correlate with transcriptional changes induced by CSB-PGBD3 expression, and suggest that CSB-PGBD3 may collaborate with AP-1 proteins in normal cells, or interfere with normal AP-1 function in Cockayne syndrome (CS). Our results have intriguing implications for CS, and provide a clear example of how insertion of transposable elements can generate useful new proteins that reshape gene expression and genomic evolution.

The impact of transposable element variants on mouse genomes and genes

Christoffer Nellåker (1), Thomas M. Keane (2), Binnaz Yalcin (3), Kim Wong (2), Avigail Agam (1,3), T. Grant Belgard (1,4), Jonathan Flint (3), David J. Adams (2), Wayne N. Frankel (5), and Chris P. Ponting (1)*

1 MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK
2 Experimental Cancer Genetics, The Wellcome Trust Sanger Institute, Cambridge, UK
3 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
4 National Human Genome Research Institute, NIH, Bethesda, MD USA
5 The Jackson Laboratory, Bar Harbor, ME, USA

Transposable element-derived (TE) sequence dominates the landscape of mammalian genomes and can modulate gene function by dysregulating transcription and translation. Our current knowledge of TEs in laboratory mouse strains is primarily limited to those present in the C57BL/6J reference genome with virtually all mouse TEs being drawn from three distinct classes, namely short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs) and members of the endogenous retrovirus (ERV) superfamily. Despite their high prevalence, the different genomic and gene properties that govern whether TEs are preferentially purged from, or are retained by genetic drift, in mammalian genomes remain largely unknown. Using whole genome sequencing data from 13 classical laboratory and 4 wild-derived mouse inbred strains, we developed a near complete catalogue of approximately 100,000 polymorphic TE variants (TEVs). SINE variants show tendencies to lie in G+C-rich sequence locations and to be inserted in the flanking regions of genes, whereas LINE or ERV variants occur preferentially in more A+T-rich sequence. In general, TEVs tend to be depleted near to transcriptional start sites, in or near exons, and more particularly, LINE variants are depleted within the introns of transcription factor genes, which we infer to be a consequence of purifying selection of deleterious TEV insertions. Within introns, we find only approximately half the expected number of ERV TEVs that are inserted in the sense transcriptional orientation also reflecting past episodes of negative selection. Our findings demonstrate that past TE insertions have often been highly deleterious, and prioritize surviving TEVs according to their likely contribution to gene expression or phenotype variation.

Genome-wide profiling of fragmentation sites in Drosophila melanogaster chromosomes revealed a strong correlation between fragmentation sites, particular sets of mobile elements and regions of intercalary heterochromatin

Nickolai A . Tchurikov, Dmitrii V. Sosin and Olga V. Kretova*

Engelhardt Institute of Molecular Biology Russian Academy of Sciences, Moscow, Russia

Eukaryotic chromosomes are subjected to non random fragmentation even upon a rapid procedure designated for preparation of high molecular weight DNA samples in the agarose plugs. Fractionation of the DNA samples by the pulsed field gel electrophoresis reveals that 50-200 kb DNA fragments (forum domains) are released during spontaneous degradation of eukaryotic chromosomes. Method for a rapid amplification of forum domains termini (RAFT) was developed. Labeled amplified forum termini (FT) were used for whole genome mapping using Drosophila genomic microarray and in situ hybridization on polytene chromosomes. It was found that FT very often correspond to regions of intercalary heterochromatin and regions of late replication in polytene chromosomes. The largest forum domains correspond to the main clusters of homeotic genes inside BX-C and ANTP-C, cluster of histone genes, and clusters of piRNAs. We also found that about 20% of FT correspond to small chromosomal regions where Ago1, Ago2, small RNAs, and repressive chromatin structures are detected. The results indicate that forum domains correspond to big multi-gene chromosomal units, some of which could be coordinately expressed. The data on the global mapping of forum domains revealed a strong correlation between fragmentation sites in chromosomes, particular sets of mobile elements and regions of intercalary heterochromatin.

Comparative analyses of transposable elements expressed in the transcriptomes of lake sturgeon (Acipenser fulvescens), tiger salamanders (Ambystoma tigrinum), and banner-tailed kangaroo rats (Dipodomys spectabilis)

Yanzhu Ji (1), Soo Hyung Eo(2), Matthew C. Hale(3), Nicholas Marra(1), J. Andrew DeWoody(1)*

1 Department of Forestry and Natural Resources, Purdue University
2 Department of Zoology, University of Wisconsin-Madison
3 Department of Biology, Purdue University

Transposable elements (TEs) have been identified in almost every organism examined by geneticists. TEs are grouped into families based on their biological characteristics, but the distribution of these TE families among phylogenetic lineages is patchy. Some TE families rapidly colonize and expand within a lineage whereas others are broadly distributed among lineages. Furthermore, TE activity varies dramatically across TE families and across branches in the tree of life. We used RNA-seq to characterize the global (genomic) expression of TEs in three disparate vertebrate lineages. We then searched these transcriptomes for known TE families; we also used de novo approaches to identify previously uncharacterized TEs. Our data provide insights into the phylogenetic overlap of expression profiles (or lack thereof) in a myriad of TE families. The identification of TE families that are active in diverse lineages will provide key insight into the colonization/expansion dynamics associated with TE demography and evolution.

Latent regulatory potential of human-specific repetitive elements

Michael D. Wilson (1,2,5), Michelle C. Ward* (1,2), Nuno L. Barbosa-Morais (3), Claudia Kutter (1), Dominic Schmidt (1,2), David Thybert (4), Stephen Watt (1), Sandy Pan (3), Rory Stark (1), Gordon D. Brown (1), Ben Blencowe (3), and Duncan T. Odom (1,2)*

1 Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK
2 Department of Oncology, Hills Road, University of Cambridge, UK
3 Banting and Best Department of Medical Research and Department of Molecular Genetics, Donnelly Centre, University of Toronto, Ontario, CA
4 EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, UK
5 Current address: Department of Molecular Genetics at The Hospital for Sick Children, University of Toronto, Ontario, CA

A significant portion of mammalian genomes are comprised of repetitive elements such as lineage-specific retrotransposons that can be silenced by a variety of mechanisms including species-specific DNA methylation, small RNA deployment, and histone modifications. We used an aneuploid mouse carrying human chromosome 21 to determine whether human-specific repetitive elements are accurately regulated in a heterologous environment. We find that hundreds of locations on human chromosome 21 are aberrantly associated with histones containing activating modifications in both somatic and germline tissues. These regions largely correspond to primate and human lineage-specific repeat elements, and occur at hypomethylated CpG dinucleotides. Furthermore, these activated repeat regions harbour transcription factor binding events and correlate with differences in gene expression. This study reveals hundreds of apparently dormant promoters, transcription factor binding sites and insulators in the human genome and illustrates the species-specificity of mechanisms needed to accurately regulate the repetitive human genome.

Chromatin and the control of LTR retrotransposon silencing and mobilisation in fission yeast

Heather Murton, Josephine Wardle and Simon Whitehall*

Newcastle University, UK

Fission yeast, Schizosaccharomyces pombe is host to a family of 13 Gypsy/Ty3 LTR retrotransposons called Tf2 in addition to ~250 solo LTRs that are the remnants of previous insertion events. In common with many other LTR retrotransposons, expression of Tf2 elements is silenced during normal growth conditions by factors involved in the assembly and maintenance of chromatin however, expression is activated in response to stress conditions, including hypoxia by Sre1 (homologue of human SREBP).

We are currently dissecting the mechanisms that mediate Tf2 transcriptional silencing and have demonstrated that it is dependent upon the HIRA complex, a conserved histone chaperone that mediates replication-independent nucleosome assembly. Inactivation of the HIRA complex dramatically de-represses expression of all 13 Tf2 elements and increases the production of cryptic transcripts from solo LTRs. Tf2 silencing is also dependent upon homologues of the human centromere protein CENP-B, which associate with LTRs and direct the recruitment of multiple histone deacetylases (HDACs). In order to determine whether transcriptional silencing is required to restrict mobilisation, we have established a system to measure the mobilisation frequency of an endogenous Tf2 element (Tf2-12). While loss of the HIRA complex led to a ~50-fold increase in Tf2 mRNA levels, it had minimal impact upon the mobilisation frequency of Tf2-12. In contrast, loss of CENP-B function resulted in a significant increase in mobilisation despite causing only a small (~3 fold) increase in expression. Furthermore, expression of a constitutively active form of Sre1 resulted in a 17-fold increase in mobilisation despite increasing expression to <50% of that observed in the absence of HIRA. As such, transcriptional activation, but not loss of silencing, stimulates element mobilisation demonstrating that Tf2 mRNA levels do not directly correlate with the frequency of mobilisation and indicating that the mobilisation of Tf2 LTR retrotransposons is controlled at multiple levels.

Similarities between the human LINE-1 reverse transcriptase and telomerase

H.C. Kopera (1), J.B. Moldovan (2), T.A. Morrish (1), J.L. Garcia Perez (1,4), and J. V. Moran (1,2,3,5)*

1 Department of Human Genetics
2 Cellular and Molecular Biology Program
3 Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI.
4 Department of Human DNA Variability, Pfizer-University of Granada-Andalusian Government Center For Genomics and Oncological Research, Granada, Spain.
5 Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI

Long INterspersed Element-1 (LINE-1 or L1) retrotransposons have been a major force in shaping the evolution of the human genome, and ongoing L1 retrotransposition contributes to both inter- and intra- individual genetic variation (reviewed in Beck et al., 2011). During the past thirteen years, my laboratory has developed a series of molecular genetic, biochemical, and modern genomic approaches to elucidate mechanistic aspects of L1 retrotransposition. In 2002, we reported an alternative endonuclease-independent (ENi) pathway of L1 retrotransposition in XRCC4- and DNA-PKcs mutant Chinese Hamster Ovary (CHO) cell lines that are deficient in the non-homologous end-joining (NHEJ) DNA repair pathway (Morrish et al., 2002). We subsequently demonstrated that a fraction of ENi retrotransposition events occurred at dysfunctional telomeres in DNA-PKcs mutant CHO cells (Morrish et al., 2007). Here, we will describe recent work, in which we used an in vitro assay (termed the LINE-1 Element Amplification Protocol or LEAP) that mimics the initial integration step in L1 retrotransposition (Kulpa and Moran, 2006). Using LEAP, we demonstrated that LINE-1 ribonucleoprotein (RNP) preparations could utilize oligonucleotide adapter sequences that mimic free telomeric ends as substrates to initiate L1 mRNA reverse transcription (Kopera et al., 2011). Additionally, we found that L1 RNPs are associated with a nuclease activity that can process oligonucleotide adapters prior to being used as substrates in the LEAP reaction, which generally resulted in perfect telomere/L1 junction sequences that are identical to those observed for ENi retrotransposition events in DNA-PKcs mutant CHO cells. Finally, we demonstrated that the L1 ORF1 encoded protein, ORF1p, is not strictly required for telomeric ENi retrotransposition in DNA-PKcs mutant CHO cells. Thus, our data further highlight mechanistic similarities between ENi L1 retrotransposition and the action of telomerase (Kopera et al., 2011).

The genetic signatures of transposable elements (TE) in schizophrenia

Fabio Macciardi (1,2), Emanuele Osimo (2), Marquis Vawter (1), Federica Torri (1,2), Bernard Lerer (4), Steven G Potkin (1), Simona Gaudi (3)*

1 Dept of Psychiatry and Human Behavior, UCI, Irvine, CA, USA
2 Department of Science and Biomedical Technology, University of Milan, Milan, Italy
3 Istituto Superiore di Sanità, Roma, Italy
4 Hadassah Medical Organization-Hebrew University, Jerusalem, Israel

Introduction
Protein-coding genes in the human genome account for 2% of the total sequence, while TEs represent about 45 to 48%, with an increased architectural complexity within and around brain-related genes. TEs have been implicated in playing a key role in dynamically regulating the genome development and function.

Since the majority of genome wide association studies (GWAS) in Schizophrenia (SZ) have mapped disease-associated variants to noncoding regions, with an alleged regulatory role, and most variants fall within well-conserved TEs, we are interested to understand how TEs may be implicated in the etiology of SZ.

Methods
Our aim is to investigate a possible functional relationship between TEs and genes identified as "best" candidates for Schizophrenia, via SNP mapping or CNV analyses. After in-silico analysis of the genomic architecture of TEs and TE-derived Transcriptional Start Sites (TSS) within and around these SZ-associated genes, we have performed a target-resequencing of these genomic regions, to look for differences between subjects who do or do not present the associated risk variants. We then looked for brain-specific expression of SZ risk-genes and associated TEs and TSS in patients and controls.

Results and conclusion
We found preliminary evidence of alternative regulation of tissue-specific gene expression in schizophrenic patients vs controls. Using AHI1 (Torri et al, 2010) as an example, we found a peculiar architecture of Alu, Line1 and other TE elements across the gene region, with a brain-specific expression pattern that looks regulated by TE-derived TSS. Target re-sequencing in subjects with and without the AHI1 risk haplotypes show marked differences in their respective sequences. Our results strongly suggest that any investigation of SZ risk genes must include a detailed analysis of TE distribution and expression, and that it is perhaps only through inclusion of these elements that psychiatric diseases will be fully understood.

Trying to pin down the mechanism of APOBEC3s inhibition of retrotransposition

Nataša Lindič and Nika Lovšin*

Chair of Biochemistry, Department of Chemistry & Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Ljubljana, Slovenia

Endogenous retroviruses and retrotransposons are mobile genetic elements that can cause genomic instability and are considered to be one of the major driving forces in the evolution of eukaryotic genomes. APOBEC3 (A3) proteins are one of the host restriction factors inhibiting retrotransposition. Belonging to the functionally diverse APOBEC protein family, A3 proteins are cytidine deaminases that edit DNA molecules and have been shown to inhibit replication of human L1 and Alu non-LTR retrotransposons. In this study, we investigated the effect of A3 proteins on the retrotransposition of an evolutionary distant relative of L1, fish non-LTR retrotransposon L2. L2 retroelements are widespread and active among vertebrates but have been inactivated in placental mammals and constitute only 2% of the human genome. We show that in ex vivo assay, retrotransposition of eel and zebrafish L2 retrotransposons is most significantly inhibited by human A3A and A3B protein, and to a similar extent as that observed with L1 retrotransposon. Mutational analysis of human A3A and A3B shows that the inhibition of L2 retroelement is partially independent of their enzymatic activity, while the intact nucleic acid binding domain is required for the inhibition. Furthermore, sequencing analysis of the newly integrated L2 elements reveals no DNA editing. Thus, our findings indicate that L2 and L1 retrotransposons share replicative steps and that differences between the two elements do not affect their restriction by A3 proteins. Moreover, we demonstrate that the only APOBECs present in zebrafish, AID, A2a and A2b, have no impact on the retrotransposition of zebrafish L2 retrotransposons. Given that vertical inactivation of L2 retrotransposons coincides with the evolution of mammalian A3 proteins, it is possible that A3 proteins could have silenced L2 elements in mammalian genomes.

The genome-wide profiling of L1 antisense promoter activity in the humnan cells

Ishiguro, Koichi (1), Higashino, Saneyuki (1), Aizawa, Yasunori (2)*

1 Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology
2 Center for Biological Resources and Informatics, Tokyo Institute of Technology

The internal promoter of the human retorotransposon L1 is bidirectional. The promoter on the sense strand contributes to production of the ORF1 and ORF2 proteins through bicistronic transcription from the +1 position of the full-length L1. The antisense promoter (ASP), on the contrary, drives transcription in the opposite direction from the position between +400 and +500 into adjacent sequences, yielding chimeric transcripts with the downstream exons.

Our understanding on the functional impact of the L1ASP transcription remains far from complete. To tackle this question, we have developed a method that allows us to identify transcriptionally active L1ASP loci throughout the human genome, which is named L1ASP Transcriptome System (LATRAS). LATRAS is composed of three steps: (1) L1ASP-specific RT-PCR (2) 454 sequencing (3) sequenced read analysis.

Using LATRAS, we pinpointed active L1ASP positions in three different kinds of human cells, leading to 504 L1ASPs that have transcriptional activity in at least one of the three cell types and 468 of these are not previously reported as active L1ASPs. And, interestingly, 63 of 504 L1ASPs are activated in all the three cell types, some of which were activated in the most of the 20 different human tissues we examined. This implies existence of ubiquitously expressed L1ASP loci. The majority of the L1ASPs are however regulated in cell type-specific manners.

We performed RT-PCR to confirm transcriptional activities of 18 L1ASP loci, which generated many 454 reads in at least one of the cell types and thus seemed to be highly transcriptionally active. And next, methylation status of these L1ASP loci were also assessed, which revealed that most of the active L1ASPs were demethylated. This suggests that LATRAS potentially also can be utilized as a genomewide method to monitor epigenomic status of L1 elements in human cells.

Technical development of Retrotransposon Capture sequencing (RC-seq)

Kyle R. Upton (1), Mark W. Barnett (1), J. Kenneth Baillie (1), Piero Carninci (2), Jeffrey A. Jeddeloh (3) and Geoffrey J. Faulkner (1)*

1 Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, EH25 9RG, UK
2 RIKEN Yokohama Institute, Omics Science Center, 1-7-22 Suehiro-chô, Tsurumi-ku, Yokohama, Kanagawa 230-0045, JP
3 Roche NimbleGen, Inc., 500 S. Rosa Road, Madison, WI 53719, USA

Recent reports of LINE-1 (L1) expression throughout development, and the detection of copy number variation between brain regions, suggest that endogenous L1 mobilization may occur frequently in human somatic cells. Identification of the corresponding integration sites has been hindered by the rare nature of de-novo insertions, and the large number of existing L1 insertions. Our lab has recently overcome these technical barriers by developing a sequence capture next generation sequencing approach termed Retrotransposon Capture sequencing (RC-seq). In our study published recently in Nature we identified roughly twenty thousand somatic L1 and Alu insertions in the hippocampus and caudate nucleus of three healthy donors.

We have recently incorporated a number of technical developments aimed at increasing the efficiency of sequence capture, including multiplexing of hybridization and sequencing, as well as a streamlined informatics pipeline. These developments will allow a reduction of per sample cost, enabling larger patient cohorts to be analyzed. This technology has clear potential to advance studies of somatic retrotransposition by providing a means to link mutation with phenotype.

Transcriptome-wide binding of human L1 ORF1 protein reveals its role in Alu retrotransposition and processed pseudogene formation.

Prabhat K. Mandal (2), Adam D. Ewing (1), Haig H. Kazazian (2)*

1 Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
2 Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA

During the process of retrotransposition, human L1-encoded proteins ORF1p and ORF2p bind with RNAs either in cis or trans to form an RNP, the primary retrotransposition intermediate. It is believed that the ORF1p RNA binding property is critical for RNP formation and subsequent retrotransposition. Here, we have explored the RNA binding property of ORF1p using a recently developed photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation technique (PAR-CLIP) to comprehensively identify ORF1p binding sites in the transcriptome of human cells (HEK293T). By Illumina sequencing, we have analyzed RNAs binding by ORF1p in the context of a functional L1RNP, as indicated by the presence of ORF1p, ORF2p, L1 RNA, and reverse transcriptase (RT) activity.

Our results show that ORF1p binds L1 RNA all across its length but less predominantly in the 5'-UTR, suggesting a reason for strong cis-preference and also a role in protecting L1 RNA from cellular nucleases. We have also seen specific binding for SVA elements at the SINE-R and Alu-like regions, but not in the highly structured VNTR region. Interestingly, ORF1p appears to bind to Alu RNA quite specifically in the single stranded linker region between the highly structured Alu left (AluL) and Alu right (AluR) monomers. An in vitro binding assay confirms that ORF1p binds to the A-rich linker region very strongly but not to the adjacent sequence. Also, ORF1p does not bind AluL and AluR monomers separately without the linker sequence. Mutation of a few nucleotides in the linker region disrupts ORF1p binding, arguing that low concentrations of ORF1p may indeed be required for Alu retrotransposition. Furthermore, sequence analysis revealed that ORF1p binds many cellular transcripts predominantly at the 3'-UTR or last exon. The data also show that processed pseudogenes form disproportionately from those transcripts that bind ORF1p. This study suggests that ORF1p binding may determine why some transcripts form more processed pseudogenes than others. Presently, we are experimentally validating this hypothesis.

LINE-1 mediated trans-mobilization of human-specific SVA retrotransposons is SVA structure-dependent

Julija Raiz (1), Annette Damert (1,3), Sergiu Chira (3), Ulrike Held (1,2), Sabine Klawitter (2), Matthias Hamdorf (2), Johannes Löwer (1), Anton Buzdin (4), Roswitha Löwer (1) and Gerald G. Schumann (1,2)*

1 Section PR2/Retroelements, Paul-Ehrlich-Institut, Langen, Germany
2 Division of Medical Biotechnology, Paul-Ehrlich-Institut, Langen, Germany
3 Institute for Interdisciplinary Research in Bio-Nano-Sciences, Molecular Biology Center, Babes-Bolyai-University, Cluj-Napoca, Romania
4 Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia

SVA elements are non-autonomous, hominid-specific non-LTR retrotransposons and distinguished by their organization as composite mobile elements. They represent the evolutionarily youngest, currently active family of human non-LTR retrotransposons, and sporadically generate disease-causing insertions. Since preexisting, genomic SVA sequences are characterized by structural hallmarks of LINE-1 (L1)-mediated retrotransposition, it has been hypothesized for several years that SVA elements are mobilized by the L1 protein machinery in trans. To test this hypothesis, we developed an SVA retrotransposition reporter assay in cell culture using three different human-specific SVA reporter elements. We demonstrate that SVA elements are mobilized in HeLa cells only in the presence of both L1-encoded proteins, ORF1p and ORF2p. SVA trans-mobilization rates exceeded pseudogene formation frequencies by 12 to 300-fold in HeLa-HA cells, indicating that SVA elements represent a preferred substrate for L1 proteins. Acquisition of an AluSp element increased the trans-mobilization frequency of the SVA reporter element by ÔÅæ25-fold. Deletion of (CCCTCT)n repeats and Alu-like region of a canonical SVA reporter element caused significant attenuation of the SVA trans-mobilization rate. SVA de novo insertions were predominantly full-length, occurred preferentially in G+C-rich regions, and displayed all features of L1-mediated retrotransposition which are also observed in preexisting genomic SVA insertions.

Furthermore, we investigated if the 324-bp sequence of exon1 of the MAST2 gene which is located at the 5' end of the source elements of the human-specific SVAF1 subfamily, playes any role in the exceptional success of this subfamily consisting of ‚â•84 members. We found that the MAST2 sequence acts as a positive transcriptional regulator of SVAF1 elements in germ cells and that the MAST2 gene is cotranscribed with members of the human retrotransposon families L1, SVA and Alu.

Competition for SVAs in gibbons: LAVA and other VNTR containing non-LTR retrotransposons

Annette Damert (1), Robert Persch (2) and Octavian Popescu (1)*

1 Institute for Interdisciplinary Research in Bio-Nano-Sciences, Molecular Biology Center, Babes-Bolyai-University, Cluj-Napoca, Romania
2 Wiesbaden, Germany

Currently SVAs are the only composite non-LTR retrotransposons known in hominoid primates. The sequence of their assembly, however, remains unclear. To address this issue we set out to characterize VNTR-containing composite retrotransposons in the genome of the Northern white-cheeked gibbon. Apart from SVA_A and SVA2 we could identify three as yet unknown families of VNTR-containing composites. The largest one displays, starting from 3', a region combining an L1 and an Alu fragment. This is followed by a VNTR region, an Alu-like domain and 5'CT repeats. Following the nomenclature adopted for SVAs we, therefore, called these elements LAVA (L1-Alu-VNTR-Alu). Based on the analysis of 570 elements retrieved from the GenBank wgs section eight subfamilies were distinguished. Age estimates range between 15 and 4 mya. This implies assembly of the ancestral LAVA element after the split of gibbons and great apes and amplification of the youngest subfamilies after the divergence of Nomascus and Hylobates/Symphalangus. The oldest subfamily displays a full-length SVA_A type Alu-like domain, whereas the younger ones are characterized by specific deletions in this part of the element.

Apart from the LA 3' sequence the 5' part consisting of CT repeats, Alu-like domain and VNTR appears to have acquired two other variant 3' ends: exon 4 and part of intron 4 of the PTGR2 gene and a sequence incorporating part of a FRAM element. Computational prediction indicates that both these variant 3' ends have been added through splicing. The structure of the elements' Alu-like domains resembles that found in SVA_A elements. Taken together, the existence of four different families of composite non-LTR retrotransposons with a common 5' part supports the hypothesis of their sequential assembly starting with acquisition of the Alu-like region and CT repeats by SVA2 and followed by the addition of variant 3' ends.

Polymer formation and nucleic acid binding properties of the human L1 non-LTR retrotransposon ORF1p protein

*K. E. Callahan, A. B. Hickman, C. E. Jones, R. Ghirlando, and A.V. Furano**

NIDDK/National Institutes of Health

The L1 (LINE 1) retrotransposable element encodes two proteins, ORF1p and ORF2p. ORF2p is the L1 replicase, but the role of ORF1p is unknown. Mouse ORF1p, a coiled coil mediated trimer of ~42 kDa monomers, binds nucleic acids and has nucleic acid chaperone activity. We purified human L1 ORF1p expressed in insect cells and made two findings that significantly advance our knowledge of the protein. First, in the absence of nucleic acids, the protein polymerizes under the very conditions (0.05 M NaCl) that are optimal for high (~1nM) affinity nucleic acid binding. The non-coiled coil C-terminal half mediates formation of the polymer, an active conformer that is instantly resolved to trimers, or multimers thereof, by nucleic acid. Second, the protein has a biphasic effect on mismatched double stranded DNA, a proxy chaperone substrate. It protects the duplex from dissociation at 37ºC before eventually melting it when largely polymeric. Therefore, polymerization of ORF1p seemingly affects its interaction with nucleic acids. Additionally, polymerization of ORF1p at its translation site could explain the heretofore-inexplicable phenomenon of cis preference – the favored retrotransposition of the actively translated L1 transcript, which is essential for L1 survival.

An overview of LINE-1 activity in mammals

Holly Wichman

University of Idaho, USA

Complete sequence of a number of mammalian genomes has confirmed that L1s have been a driving force in shaping the mammalian genomes. Although no traces of L1s have been found in monotreme genomes, they are present in all eutherian and marsupial genomes examined to date. Thus they appear to have co-evolved with mammalian genomes for over 150 MY. This unusually long maintenance of transposable element activity has been used to argue that L1s may play a functional role in the genome, but some extinctions of L1s within mammalian lineages have been documented. Using a simple strategy to enrich for markers of recently active L1s, we have examined L1s from over 200 species including representatives of almost all eutherian and marsupial lineages that had arisen by 65 MYA, the time of the mammalian radiation. There are several unexplained features of L1 evolution, including maintenance of active elements in most lineages, the close sequence identity of these active elements within most genomes, and the mechanisms of L1 extinction when it does occur.

Genomes without borders - the misfit origins of genetic novelty

Cedric Feschotte

Department of Biology, University of Texas, Arlington, USA

Horizontal DNA transfer, the passage of genetic material between non-mating species, has long been recognized as a crucial mechanism driving bacterial evolution. In contrast, the evolutionary significance of horizontal transfer in eukaryotes has remained more obscure. We believe this gap in perception is attributable to the disproportionate attention given to the transfer of cellular genes as opposed to that of extragenic or extracellular DNA. In recent years my laboratory and several others have revealed a previously underappreciated incidence of two flows of genetic material horizontally colonizing animal genomes. First, there is growing evidence that many transposons have 'jumped' between widely diverged animals, including invertebrates and vertebrates, to spawn huge waves of lineage-specific transposon invasions. Another form of horizontal transfer common in animals consists in the genomic assimilation (or endogenization) of viral sequences. Although endogenous retroviruses have been known for some time, recent studies reveal that virtually every major type of viruses can be integrated and assimilated by animal genomes. This includes RNA viruses that have no DNA stage in their life cycle, implicating the role of cellular reverse transcriptases, such as those encoded by retrotransposons, in endogenization. In this talk I will argue that the horizontal introduction of transposable elements and viral sequences could be as fundamental to animal evolution as lateral gene transfer is to bacteria to propel genomic variation and innovation.

Waves of repeat driven CTCF binding expansions have shaped mammalian genomes

Petra C. Schwalie (1), Dominic Schmidt* (2,3), Michael D. Wilson (2,3), Benoit Ballester (1), Angela Goncalves (1), Claudia Kutter (2,3), Gordon D. Brown (2,3), Aileen Marshall (2,5), Duncan T. Odom (2,3,4) and Paul Flicek (1,4)*

1 European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
2 Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK
3 Department of Oncology, Hutchison/MRC Research Centre, Hills Road, Cambridge, UK
4 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
5 Cambridge Hepatobiliary Service, Addenbrooke's Hospital, Hills Road, Cambridge, UK

CTCF is a ubiquitous zinc-finger protein that plays essential roles in domain insulation, nuclear architecture and transcription regulation within vertebrates. We experimentally determined the genome-wide occupancy of CTCF in livers of six mammalian species by ChIP-sequencing, finding (1) remarkable conservation of CTCF binding, (2) a larger sequence specificity of CTCF-DNA contacts, and (3) evidence for repeat-mediated expansion of CTCF binding in multiple lineages.

As opposed to canonical transcription factors that show rapid turnover of their genomic binding, we detect that up to 60% of CTCF bound regions are shared along lineages, while as much as 10% are common to all mammalian species studied. Shared regions are evolutionary conserved at the sequence level, and show higher ChIP enrichment, as well stability (as assessed by CTCF RNAi). A subset of binding events contain a second sequence motif, located at a preferred and conserved spacing with respect to the canonical one; this suggests that CTCF contacts the DNA over a larger region than previously described in genome-wide studies. The binding frequency of motif sequences (words) is in general conserved across species, with a few highly bound words corresponding to deeply shared regions of high ChIP enrichment. However, certain motif instances are restrained to only one/two organisms and represent motifs embedded in lineage, respectively species-specific repeat elements. Functionally, we detect no difference between deeply-shared, repeat-associated and regular CTCF binding sites.

Our data suggests a model by which intermitted waves of CTCF motif expansions mediated by repeat elements create a pool of putative bound elements that can take up roles in chromatin structure and transcriptional regulation. As selection acts on the newly created sites, some will be maintained and eventually detected as shared events along a tree branch.

Ancient and recurrent evolution of the antiviral gene fusion TRIMCyp in primates

Ray Malfavon-Borja (1), Lily Wu (2), Michael Emerman (2,3), Harmit S. Malik (1,2)*

1 Department of Genome Sciences, University of Washington, Seattle, WA, USA
2 Division of Basic Sciences, Fred Hutchinson Cancer Research Center, WA, USA
3 Division of Human Biology, Fred Hutchinson Cancer Research Center, WA, USA

TRIM5 is an antiviral factor that inhibits viruses like HIV-1, where inhibition is mediated by a C-terminal B30.2 domain that interacts with viral capsid. Remarkably, soon after the discovery of TRIM5, the TRIM5-CyclophilinA (TRIMCyp) gene fusion was discovered, which also inhibits viruses like HIV-1. TRIMCyp shares structural similarity with TRIM5, except the B30.2 domain is replaced by a capsid interacting CyclophilinA (CypA). TRIMCyp is formed by LINE-mediated retrotransposition of CypA into or downstream of TRIM5, and has been found in the Aotus and Macaque primate lineages. The convergent evolution of antiviral TRIMCyp in two distinct primate lineages suggests the utility of the gene fusion and of CypA retrotranspositions. Thus, the aim of our study was to evaluate the utilization and consequences of CypA retrogenes in primate genomes. An in silico analysis revealed that human, chimpanzee, and rhesus macaque genomes contain 100+ CypA retrogenes. Within the rhesus genome, we located a second CypA retrogene ~14Kb downstream of TRIM5 that we labeled as CypA2. CypA2 is at least 43 million years old and can be found in all extant Old World monkeys (OWM). The majority of extant Hominoids (e.g. human and orangutan) do not encode CypA2 because of a deletion that occurred after the divergence of gibbons. Despite the presence of nonsense mutations in the ORF of present day CypA2 sequences, a subset of OWM still express CypA2 as a gene fusion with TRIM5. Evolutionary reconstruction of the ancestral CypA2 sequence produces a putatively functional CypA and by constructing chimeric proteins with owl monkey and macaque TRIMCyps with CypA2, we have identified residues on the C-terminus of CypA2 that affect interactions with viruses like HIV-1. These findings reveal a third, ancient, yet currently inactive TRIMCyp encoded by OWM, and further demonstrates the utility of LINE-mediated CypA retrotranspositions in primate history.

Alu/Alu non-allelic homologous recombination

Prescott Deininger, Maria E. Morales, and Dale J. Hedges*

Tulane University Cancer Center, New Orleans, LA, USA

Alu elements are the most abundant retroelements in humans, with over 1,000,000 copies. They represent the single most common cause of non-allelic homologous recombination (NAHR), contributing significantly more to genetic instability than their insertional mutagenesis. There are a number of genes that are subject to recurrent Alu/Alu recombination that contributes to disease. These include deletions or duplications within the MLL, MSH2, VHL, and BrCA1 genes contributing to cancer. The reasons for some regions to be more prone to this form of recombination are unknown but may involve the density and relative similarity of Alu elements in these genes, as well as potential broader chromosomal contributions.

We have created a novel cassette-based recombination assay system that allows the evaluation of a broad range of Alu-related sequence factors for their impact on mutagenic recombination. This cassette can be altered and replaced at the same genomic site in specific cells, allowing the testing of multiple variant constructs in the same chromosomal and cellular environment. By expressing the I-SceI endonuclease in these cells, we can create and measure a high level of Alu/Alu recombination. Replacing one of the Alu elements with variations having different percentage and distribution of mismatches demonstrates a dependence on sequence identity, as well as distribution. Analysis of the recombination junctions showed recombination events were spread evenly across the Alu element. Having a single longer region of homology increased the recombination rate, but the recombination events did not map to that longer region. With perfectly matched Alu elements, our system detects only Alu/Alu NAHR events. However, with increasing mismatch, deletions caused by non-homologous end joining (NHEJ) become dominant. Surprisingly, at moderate levels of mismatch, both NAHR and NHEJ in our vector is suppressed. Thus, there is a complicated interplay of stimulation and suppression of local recombination caused by Alu elements.

Studying Alu and L1 mutagenesis in human genomes with DNA sequencing technologies

Scott E. Devine

Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA

Although the hypothesis that L1 mutagenesis might contribute to human cancers has been around for quite some time, it has been difficult to test in humans. However, two studies have now documented that L1 is mobilized in human tumors, providing evidence that L1 may drive tumor formation and/or progression. The first example involved an L1 insertion that disrupted the APC tumor suppressor gene in a case of colorectal cancer (Miki et al. 1992, Cancer Res. 52, 643-645). Given that this insertion disrupted a coding exon (and was absent from normal adjacent tissues), it is likely to have played a causative role in the cancer in which it was discovered. For almost 20 years, this APC insertion was the only known example of a somatic L1 event in a human tumor, suggesting that it might have been a rare anomaly. However, using a new transposon-seq detection technology that was developed in our laboratory, we determined that human lung tumors also support high levels of somatic L1 mobilization (Iskow et al. 2010, Cell 141, 1253-1261). Together, these studies indicate that L1 is mobilized in at least two types of human tumors (colon and lung). Moving forward, we have begun to test the hypothesis that endogenous L1 elements are drivers of human lung tumorigenesis using an approach that is similar to the approach that was used with the Sleeping Beauty transposon in mice (Dupuy et al. 2005, Nature 436, 221-226). Likewise, we are expanding our study to examine somatic retrotransposition in a wide range of human somatic tissues and tumor types, including colorectal, brain, and other tissues. Finally, we are studying the possible relationship between hot L1 copy number and cancer risk. These studies are providing new insights on the link between L1 mobilization and tumorigenesis in humans.

Large-scale retroelement detection from The Cancer Genome Atlas

Adam Ewing, Tracy Ballinger, Jing Zhu, David Haussler*

Center for Biomolecular Science and Engineering, University of California at Santa Cruz

A number of recent advances in algorithms for the detection of structural variants have enabled the discovery of retrotransposon insertions from whole-genome paired-end sequence data on a massive scale. The ability to detect thousands of novel insertion sites is useful not only for improving catalogs of polymorphic variants in human populations but also for detecting cancer-specific insertions where both tumor and normal samples have been sequenced. Here we report on our progress in utilizing several previously developed and novel computational approaches to detect retroelement insertions from The Cancer Genome Atlas (TCGA) and NCI TARGET datasets. These data are unprecedented in scale with the goal of sequencing over 11,000 tumors and matched normal controls across over 25 cancer types. In addition, these projects will allow integration of retroelement detection with functional data from RNA-seq performed on the tumor and normal samples, allowing us to infer which mutations are more likely to have a clinically relevant impact in terms of cancer initiation or progression.

A novel p53 regulated murine endogenous retrovirus

Robyn Leary and Anindya Bagchi*

Department of Genetics, Cell Biology and Development, Masonic Cancer Center and Center for Genome Engineerin, University of Minnesota

With a substantial portion of the human and mouse genomes being comprised of Endogenous Retroviruses (ERV), 8% and 10% respectively, it remains largely unknown whether these sequences currently serve a function in these organisms. Typically, ancient provial integrations are silenced, and their expression is concamitant with diseased states (e.g. cancer, autoimmune disorders, diabetes, lupis). We have identified an ERV in mouse whose expression is dependent on the tumor suppressor, p53. This ERV is expressed in numerous adult and embryonic tissues. Our studies demonstrate that p53 binds the LTR promoter of this ERV and modulates its expression. Single mutations to the p53 response element render the promoter inactive. Interestingly, this ERV has inserted upstream of multiple genes which we demonstrate to be expressed in a p53 dependent manner suggesting that the LTR can act as an enhancer of these genes. Furthermore, these data suggest exaptation of the retroviral sequence. Finally, we hypothesize that the expression of this ERV can play a role in apoptosis. Our data suggest novel function of tumor suppressor p53 in regulating endogenous retroviruses and opens up new avenues to study the role of ERVs in tumor suppression unknown until now.

Distinct genome-wide p53 binding profile in normal and cancer-derived human cells

Krassimira Botcheva (1), Sean R. McCorkle (1), W. Richard McCombie (2), John J. Dunn (1) and Carl W. Anderson (1)*

1 Biology Department, Brookhaven National Laboratory, Upton, NY 11973 USA
2 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724 USA

Given the importance of site-specific binding to DNA for the p53-mediated transcriptional regulation and tumor suppression, extensive efforts have been directed toward experimental identification of p53 binding sites in the human genome. Notably, all de novo genome-wide binding studies reported up to date were done in cancer-derived cell lines. Using a ChIP-seq approach we mapped p53 binding sites for the first time in normal human cells (lung fibroblasts IMR90)1 and by comparing them with sites previously reported by genome-wide studies in cancer cell lines, we showed dramatic differences in their genomic distribution. p53 binding sites mapped in normal, unlike those in cancer cells, are enriched in the immediate vicinity of TSS and at CpG islands. Close proximity of binding sites to TSS has been typically seen in studies of individual p53 target genes, but has not been observed by genome-wide studies before. We found the binding sites in IMR90 are far less enriched at repeats compared to those reported in HCT116 and U2OS cancer cell lines. The observed differences in the p53 binding pattern do not seem to reflect a distinct preference for specific sequences, since the de novo developed p53 motif based on the IMR90 study is similar to those reported previously. More likely the different chromatin landscapes in normal and cancer-derived cells modulate the availability of the binding sites, which is supported by the comparison between the IMR90 ChIP-seq binding data and the IMR90 methylome2 pointing at enrichment of binding sites in hypomethylated DNA.

References:

1. Botcheva K, McCorkle SR, McCombie WR, Dunn JJ and Anderson CW. Cell Cycle 2011; (in press)
2. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Nature 2009; 462:315-22.

RetroSeq : a tool to discover somatic insertion of retrotransposons

Mike Lawrence, Chip Stewart, Gad Getz, Matthew Meyerson

Harvard-MIT Health Sciences & Technology, Broad Institute

Retrotransposons comprise over 40% of the human genome and represent an important class of genetic variation. Though most elements are stationary ancient relics, it has recently been shown that some 100 elements remain highly active, copying and pasting themselves throughout the genome with each generation. Retrotransposon insertions can disrupt gene function, modulate gene expression, and lead to genomic rearrangement. These events have previously been implicated in cancer, including an account of insertional mutagenesis in APC as an early event in colon cancer progression. Due to the inherent difficulty in localizing repeat elements, however, an extensive investigation of retrotransposons' role in cancer has not yet been performed. We developed RetroSeq, a computational framework to identify novel insertions of retrotransposons from paired-end sequencing data. In simulated data, RetroSeq identifies retrotransposon insertions with 99% sensitivity and 99% specificity. We ran RetroSeq on whole-genome sequencing data from 9 colorectal and 22 breast tumors and matched normal samples to determine the extent of somatic retrotransposon activity. We find a broad range of novel retrotransposon insertion events specific to each tumor, with some insertions falling in exonic as well as intronic and intergenic regions. In sum, using RetroSeq we are able to localize novel retrotransposon insertions in paired-end sequencing data and provide evidence for the reactivation of retrotransposons in cancer.

Mouse endogenous retroviruses can trigger premature transcriptional termination at a distance

Jingfeng Li (1), Keiko Akagi (1), Yongjun Hu (4), Anna L. Trivett (5), Christopher J. W. Hlynialuk (1), Deborah A. Swing (6), Natalia Volfovsky (7), Tamara C. Morgan (8), Yelena Golubeva (8), Robert M. Stephens (7), David E. Smith (4), and David E. Symer (1,2,3)*

1 Human Cancer Genetics Program and Dept. of Molecular Virology, Immunology and Medical Genetics,
2 Dept. of Internal Medicine, and
3 Dept. of Biomedical Informatics, The Ohio State University Comprehensive Cancer Center, Columbus, OH
4 Department of Pharmaceutical Sciences, University of Michigan, Ann Arbor, MI
5 Laboratory of Molecular Immunoregulation and
6 Mouse Cancer Genetics Program, National Cancer Institute, Frederick, MD
7 Advanced Biomedical Computing Center, Information Systems Program and
8 Histotechnology Laboratory, SAIC-Frederick, Inc., National Cancer Institute, Frederick, MD

* to whom correspondence should be addressed

Endogenous retrotransposons have caused extensive genomic variation within mammalian species, but the functional implications of such mobilization are mostly unknown. We mapped thousands of endogenous retrovirus (ERV) germline integrants in highly divergent, previously unsequenced mouse lineages, facilitating a comparison of gene expression in the presence or absence of local insertions. Polymorphic ERVs occur relatively infrequently in gene introns. They particularly are depleted from genes involved in embryogenesis or that are highly expressed in embryonic stem cells, and they frequently are oriented antiparallel to genes. Their genomic distribution implies ongoing negative selection due to deleterious effects on gene expression and function. A polymorphic, intronic ERV at Slc15a2 triggers up to 49-fold increases in premature transcriptional termination and up to 39-fold reductions in readthrough transcripts in adult mouse tissues, thereby disrupting protein expression and functional activity. Prematurely truncated transcripts also occur at Polr1a, Spondin-1 and up to ~5% of other genes when intronic ERV polymorphisms are present. Analysis of expression quantitative trait loci (eQTLs) in recombinant BxD mouse strains revealed very strong genetic associations between the polymorphic ERV in cis and disrupted transcript levels. Premature polyadenylation is triggered at genomic distances up to >12.5 kb upstream of the ERV, both in cis and between alleles. The parent of origin of the ERV is associated with variable readthrough transcript levels. This study defines an unexpectedly strong functional impact of ERVs in disrupting gene transcription at a distance, and demonstrates that ongoing retrotransposition can contribute significantly to natural phenotypic diversity.

Complex interactions between endogenous retroviruses and host genes

Dixie L. Mager, Ying Zhang, Katharine Miceli-Royer and Rita Rebollo*

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

Mammalian genomes are littered with sequences related to retroviruses. These endogenous retroviral elements (ERVs) are likely the result of ancient retroviral infections of the germ line that are now inherited as normal constituents of chromosomal DNA. A major focus of our laboratory is to determine the impact that ERVs have on mammalian gene evolution and regulation. The presentation will have two parts. First, our work to understand the epigenetic interplay between genes and nearby ERVs will be presented. We have shown that spreading of epigenetic silencing marks from mouse ERVs to gene promoters is extremely rare, with ERV insertions capable of inducing such spreading likely being subject to strong negative selection. Some ERVs that are allowed to remain near gene promoters are less methylated than average copies, suggesting that the gene's environment can influence the epigenetic state of the ERV. Potential ramifications of this phenomenon will be discussed. In the second part of the presentation, our bioinformatics analyses of the distributions of ERVs and other transposable elements within gene introns will be summarized, as well as strategies to use such distributions to predict which intronic transposable elements may be most likely to disrupt gene transcription.

Transposon small RNA expression in the embryonic and trophectoderm lineage

Andrea Schorn (1), Wendy Dean (2), Doug Baumann (3), Rebecca Doerge (3), Wolf Reik (2), Rob Martienssen (1)*

1 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2 Babraham Institute, Cambridge, UK
3 Department of Statistics, Purdue University, West Lafayette, IN

Epigenetic reprogramming is essential to regain pluripotency during reproduction but also results in transient release of transposable elements (TE) from heterochromatin repression. Small RNAs complementary to transposons are able to suppress transposon activity and guide reassembly of heterochromatin. In plants, transposon release in neighboring tissues triggers the production of mobile small RNAs that are capable of inducing silencing in the generative cells (Slotkin et al. 2009). We wondered if, similarly, nurse tissues in mammals express transposon-targeting small RNAs that could ensure genome integrity in the offspring?

We profiled small RNA expression from mouse embryonic (ES) and trophectoderm stem (TS) cells to assess the potential contribution of nurse tissues to transposon silencing in the embryonic lineage in mice. We found TEs that produce small RNAs in ES or TS cells are on average younger than the mean transposon age genome-wide. Also, ERVs (endogenous retroviruses) and eutherian-specific hAT-Tip100 DNA transposons are particularly over expressed in TS cells. Interestingly, ERV elements are still active in mice and have contributed to placenta evolution. Therefore, small RNA expression in the trophectoderm lineage might reflect a trade-off between the benefits of transposon domestication during eutherian evolution and the necessity of protecting the embryo from active transposition. We propose an in vivo model to test for small RNA signaling between placenta tissues and the embryo proper.

References:

1. Slotkin RK, Vaughn M, Borges F, Tanurdzič M, Becker JD, Feijó JA, Martienssen RA. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell. 2009

Dynamics of TE control by the piRNA machinery: the key role of dose

Justin P. Blumenstiel, Dean M. Castillo, Christine Yoder, Chris C. Harrison, Michelle L. Wickersheim and Jianwen Fang*

University of Kansas

Transposable elements (TEs) are generally harmful genetic parasites that can cause mutation, shape genomes, and contribute to the architecture of gene expression networks. Historically, natural selection has been considered to play the key role in limiting TE proliferation in populations. However, recent studies have demonstrated that an adaptive system of genome defense by piRNA also limits TE proliferation. Using the Drosophila genus as a model, I will discuss our research on how epigenetic inheritance of piRNA protects against TE proliferation across generations. This work demonstrates the key role that dosage plays both in the induction of TE mediated hybrid sterility and in maternally provisioned protection against it. In light of the key role that dosage plays, I will also present work on the molecular evolution of the piRNA machinery that suggests a complex co-evolutionary dynamic between TEs and the machinery of genome defense. In particular, the dominant evolutionary response to increasing TE burden across the Drosophila genus seems to be improved translational efficiency in the piRNA machinery, not an increased rate of evolution.

Drosophila interspecific hybrids phenocopy piRNA pathway mutants in aberrant piRNA production and TE derepression

Erin S. Kelleher and Daniel A. Barbash*

Cornell University

Germline transposition can result in deleterious mutations, extreme chromosomal damage, and gonadal atrophy. Recent research on the Piwi-interacting RNA (piRNA) pathway has made the exciting discovery that these proteins act in concert with TE-derived small RNAs (piRNAs) to silence active TEs in male and female germlines. piRNAs and their associated proteins are maternally deposited, furthermore, thus priming the silencing process in the offspring's primordial gonads.

Eukaryotic genomes evolve rapidly in terms of the activity and distribution of individual TE classes, and current models posulate that the piRNA pathway responds to the dynamic pool of TEs by modulating the composition of maternally deposited piRNAs. We tested this hypothesis by comparing piRNA pathway function between D. melanogaster, D. simulans, and their interspecific hybrids. We show that piRNA abundance from individual TE classes is dramatically different between D. melanogaster and D. simulans, but that maternal representation is rarely predictive of piRNA abundance or TE expression levels in hybrid offspring. Rather, interspecific hybrids phenocopy piRNA pathway mutants in terms of aberrant piRNA protein localization, reduced transcription of precursor piRNAs, disrupted piRNA processing, and global TE derepression. We suggest that our results reflect incompatibilities between rapidly-evolving D. melanogaster and D. simulans piRNA proteins, indicating the extraordinary selection pressure that germline TE activity imposes on eukaryotic genomes.

An increase in relative abundance of pi-like RNAs in response to ionizing radiation in the bdelloid rotifer Adineta vaga

Fernando Rodriguez and Irina Arkhipova*

Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543, USA

Bdelloid rotifers are freshwater invertebrates extraordinarily resistant to ionizing radiation and desiccation. We studied the effects of ionizing radiation on the bdelloid species Adineta vaga in order to investigate whether small RNA-mediated genome defense system that includes PIWI-interacting RNAs (piRNAs) could be induced under conditions when rotifers are undergoing repair of massive DNA double strand breakage incurred upon radiation. For this purpose, small RNA libraries were constructed from non-irradiated wild type Adineta vaga, as well as from A. vaga irradiated with 137-Cs source delivering a dose of 500 Gy.

Initial characteristics of piRNAs were compared between the non-irradiated and irradiated samples. We report that the difference pattern for pi-like RNAs (25-33 nt long, 5' uridine bias) is evident comparing the two biological samples, indicating a substantial increase in relative abundance of pi-like RNA in bdelloid rotifers subjected to a dose of ionizing radiation that provokes DNA damage.

All potential piRNAs were mapped to the initial A. vaga genome assembly composed of 454 reads, to A. vaga transcriptome, and to previously annotated A. vaga transposable elements (TEs). To measure the enrichment/depletion of reads in each assembled contig, a metric calculation of RPM (number of reads per million) was set between the two samples. In agreement with our expectations, the initial annotations meet the correlation between TEs and piRNA, in accordance with the role of piRNAs as central players in transposon silencing. In addition, we found unexpected connections with a number of cellular genes that may be responding to massive DNA damage.

Epigenetics speaks up for the silent DNA

Mireia Jordà, Sergi Lois, Anna Díez, Berta Martín, Víctor Barrera, Miguel A. Peinado*

Institute of Predictive and Personalized Medicine of Cancer (IMPPC), Badalona, Spain

DNA methylation is a key component of the epigenome architecture that plays a crucial role in numerous physiological and pathological processes. Most genome-wide studies analyzing this epigenetic mark have focused the analysis to high density CpG islands and gene promoters, leaving out most of the genome, an especially repeat elements. A haploid human genome contains more than one million Alu repeats, but despite their overwhelming presence only a small fraction is likely to have a specific functional role. Most of them are methylated and silenced, nevertheless, a fraction of Alus remains unmethylated in normal cells and this proportion is significantly increased in cancer cells. The grounds for this "atypical" epigenetic state of unmethylated Alus as their functional implications are unknown.

Although different approaches have been used to make bulk estimates of DNA methylation in repeat elements, there is still a lack of screening strategies that specifically allow a feasible identification of methylation in repetitive sequences. We have developed a new approach in combination with next-generation sequencing that specifically targets unmethylated Alus and allows the generation of a comprehensive maps at the genome scale. Using this approach has been applied to detect unmehylated Alus in normal tissues, primary tumor cells, colon cancer cell lines and embryonic stem cells. Data reveal the existence of distinctive profiles of Alus with differential methylation associated with cell differentiation and tumorigenesis as well as a subset of Alus whose default state is unmethylation. Furthermore, these unmethylated Alus present specific features, including high evolutional conservation and association with active chromatin domains. Our studies support a direct role of unmethylated Alus in cell biology and underscore the need to detect active repeat elements as an essential compartment of the genome that contributes to regulation and complexity of biological process.

Modeling stress signaling and response resulting in transposon activation in human cells

Zsuzsanna Izsvak, Dawid Grzela, Zoltán Ivics*

Max Delbrueck Ctr. for Molecular Medicine, Berlin, Germany

Cellular mechanisms that are directly involved in development or stress-response have crucial role in establishing stable host-transposon co-existence. Our recent study sheds light on a heat shock response-associated mechanism of regulating transposition. Our recent data suggest that Sleeping Beauty (SB) transposition is regulated by heat shock (HS) on different levels. First, the SB transposon has evolutionarily conserved transcription factor binding sites for the HS factors 1 and 2 (Hsf1/2), and the transposase gene is transcriptionally activated by induced Hsf1/2, indicating that HS factor binding sites can be distributed in the host genome by SB. Our data show that the SB transposase protein is a relatively stable protein, but is sensitive to misfolding, arguing for a potential regulation via cellular trafficking and/or folding. In addition to transcriptional regulation, I propose that SB transposition might piggyback the HS response pathway that normally reactivates heat-aggregated, nuclear proteins. Thus, transposons might exist in a "latent" form in the genome and are able to sense developmental and environmental changes and manipulate stress signaling.

Transgenerational effects and genomic impacts in environmental stress

Hidetaka Ito

Faculty of Science, Hokkaido University, Japan

Whether the environment has only a passive role in evolution by selecting the fittest or it may have an active role in generating genetic and epigenetic diversity for selection? To answer the question, I focused on a transposable element or a transposon. An environmental stress-induced transposon changes the structure of the host genome and influences the gene expression. A heat-activated copia-type retrotransposon ONSEN ("hot spring" in Japanese) was activated with heat stress in Arabidopsis thaliana. ONSEN became not only transcriptional active but also synthesize the extrachromosomal DNA copies. We found high frequency of retrotransposition in the progeny of stressed siRNA mutant plants indicating that siRNAs were required to control the transposition of an active transposon. It was surprising that the memory of stress applied to seedlings has been maintained throughout the entire plant development allowing ONSEN to transpose during change of plant generations. Insertion patterns between siblings suggested that the transgenerational retrotransposition occurred in early embryo. Our result suggested the new mechanism to maintain an active retrotransposon and also demonstrated the epigenetic control of retrotransposition as post-transcriptional machinery. Interestingly, an activated transposon inserted on adjacent genes has generated stress-tolerant plants. We demonstrated that transposon inserts conferred heat stress inducibility on adjacent genes and also generated stress-tolerant plants.

We have met the enemy and he is us: co-regulation of transposons and genes during plant development

*Hong Li and Damon Lisch**

Department of Plant and Microbial Biology, U.C. Berkeley

It has become increasingly apparent that epigenetic regulation of host genes and TEs are deeply intertwined processes. Although some mechanisms are thought of as "for" one or the other, nearly all silencing pathways, from microRNAs to RNA-directed DNA methylation, can and do impact both TEs and host genes. Our work focuses on epigenetic silencing of MuDR, an autonomous TE in maize. Silencing can be initiated by genetically combining MuDR with an inverted repeat variant of this element called Muk. Muk induces rapid methylation and heritable silencing of MuDR elements in F1 plants. Remarkably, however, we find that both methylation and transcriptional silencing are absent in leaves during the transition from juvenile to adult growth. The loss of transcriptional silencing in these leaves is associated with down-regulation of SGS3, a key component of a silencing pathway required for the production of trans-acting siRNAs. In addition to hypomethylation of MuDR, the loss of SGS3 is associated with up-regulation of both a target of the tasi-RNA pathway and of natural antisense gene pairs. These data suggest that maize experiences global changes in epigenetic regulation in leaves just prior to reproductively competent growth. Finally, we show that maize genes can be co-opted into this silencing pathway due to the insertion of TEs. Given the vast number and high mobility of TEs in maize, we suggest that this process may have had a significant impact on the evolution of gene regulation in this species.

Retroviral-type LTRs as intermediate of the stress response in tobacco

Bui, Q.T. (1), Anca, I. (2), Deloger, M. (1), Fromentin, J. (2), Maunoury, N. (1), Tran, J. (1), Dorlhac de Borne, F. (3), Daniel, J. (1), Mhiri, C. (1), Quesneville, H. (4), Van Sluys, M-A. (5), Simon-Plas, F. (2) and Grandbastien, M-A. (1)*

1 Institut Jean-Pierre Bourgin (IJPB), INRA, Versailles, France
2 Unité Plante-Microbe-Environnement (UPME), INRA, Dijon, France
3 Tobacco Institute, Imperial Tobacco Group, Bergerac, France
4 Unité de Recherche Génomique Info RGI, INRA, Versailles, France
5 Genomics and Transposable Elements Laboratory (GaTE), IBUSP, São Paulo, Brazil

The expression plant retrotransposons, such as Tnt1 of tobacco (Nicotiana tabacum), is linked to stress pathways and may be involved in host global stress response. The LTRs (Long Terminal Repeat) of retrotransposons can drive the production of chimeric co-transcripts and thus modulate adjacent gene expression in response to their own specific expression patterns. Multiple such co-transcripts produced from the 3'LTRs of tobacco retrotransposons, and extending into downstream adjacent genic sequences, were detected in tobacco EST collections produced in stress conditions.

Our experimental analysis confirm that 3'LTRs can drive the synchronous expression of these co-transcripts in conditions where retrotransposons are transcriptionally activated, such as microbial elicitors or wounding. The retrotransposon response to stress and associated co-transcript production varies depending on the element. In parallel, we are extending our analysis to a larger range of retrotransposons reconstructed from tobacco shotgun genomic data, and are developping a 454 RNA-Seq analysis of retrotransposon expression and chimeric co-transcripts production in stress conditions.

Our current hypothesis is that plant retrotransposons may act as intermediates of stress stimuli, redirecting messages towards cellular functions in stress conditions. This may be of key significance for plant adaptability to environmental changes, as plants can not move to avoid them and have evolved complex and highly coordinated responses to biotic and abiotic challenges.

The effect of environmental stress events on the mobility of four LTR retrotransposon families in grapevine (Vitis vinifera) somatic embryo cultures

Darrell K. Lizamore, Chris S. Winefield*

Dept. of Wine, Food and Biomolecular Sciences, Lincoln University, Lincoln, Canterbury, NZ

The long terminal repeat (LTR) regions of class I transposable elements in plants contain regulatory elements common to the promoters of stress-response genes. Accordingly, expression levels of several transposon families have previously been shown to increase in response to environmental stresses in model species. In this study we analyze the genomic variation resulting from the stimulation of endogenous transposons in grapevine. Somatic embryo cultures of Vitis vinifera cultivars Chardonnay and Pinot noir were regenerated after exposure to a variety of environmental stresses (temperature / UV / osmotic shock and pathogen exposure). Genomic DNA was extracted from young regenerated plantlets and analyzed by sequence-specific amplified polymorphism (S-SAP) transposon display. Simultaneous screening of four endogenous transposon families was achieved by labeling each with distinct fluorescent dyes during amplification. Insertion profiles were then recorded by multiplex capillary electrophoresis. We compare the relative mobility of the transposon families following stress events and plantlet regeneration. The differential activation of endogenous transposon families demonstrates the potential of environmental forces to stimulate transposon mutagenesis in an important agricultural crop.

Plant LTR-RT study reveal fine-scale individual molecular patterns

Douglas S. Domingues (1,4), Guilherme M.Q. Cruz* (1), Cushla J. Metcalfe (1), Fabio T.S. Nogueira (2), Renato Vicentini (3), Cristiane de S. Alves (2), Marie-Anne Van Sluys (1)*

1 Genomes and Transposable Elements Laboratory, Depto. de Botânica, Inst. de Biociências, Universidade de São Paulo, Rua do Matão, 277, 05508-090, São Paulo, Brazil
2 Depto. de Genética, Inst. de Biociências, Universidade Estadual Paulista, campus de Botucatu, Distrito de Rubião Jr., s/n, 18618-000, Botucatu, Brazil
3 Systems Biology Laboratory, Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Av. Cândido Rondon, 400, 13083-875, Campinas, BR
4 Present address: Plant Biotechnology Laboratory, Instituto Agronômico do Paraná, Rod. Celso Garcia Cid (PR-445), km375, 86047-902, Londrina, Brazil

Long Terminal Repeat retrotransposons (LTR-RTs) are the single largest components of most plant genomes and can substantially impact the genome in many ways. It is therefore crucial to understand their contribution to the genome and transcriptome, however a detailed study of LTR-RTs in sugarcane has not been previously carried out. Sixty complete LTR-RT elements were classified into 35 families within four Copia and three Gypsy lineages. Structurally, within lineages elements were similar, between lineages there were large size differences. FISH analysis resulted in the expected pattern of Gypsy/heterochromatin, Copia/euchromatin, but in two lineages there was localised clustering on some chromosomes. Analysis of related ESTs and RT-PCR showed transcriptional variation between tissues and families. Four distinct patterns were observed in sRNA mapping, the most unusual of which was that of Ale1, with very large numbers of 24nt sRNAs in the coding region. The results presented support that distinct small RNA-regulated pathways in sugarcane target the lineages of LTR-RT elements. Individual LTR-RT sugarcane families have distinct structures, and transcriptional and regulatory signatures. Our results indicate that in sugarcane individual LTR-RT families have distinct behaviours and can potentially impact the genome in diverse ways. For instance, these transposable elements may affect nearby genes by generating a diverse set of small RNAs that trigger gene silencing mechanisms. There is also some evidence that ancestral genomes contribute significantly different element numbers from particular LTR-RT lineages to the modern sugarcane cultivar genome.

Genome-wide impact of a MITE burst in rice after just 20 generations

Susan Wessler (1), Jason Stajich (1), Lulu (1), Yuki Monden (1), James Burnette (1), Sofia Robb (1), Yaowu Yuan (2), Yutaka Okumoto (3)*

1 University of California, Riverside
2 University of Washington,
3 Kyoto University

The miniature inverted repeat transposable element (MITE) mPing is actively transposing in four landraces of rice (Oryza sativa). In one of these strains, EG4, we have documented over 25 new mPing insertions per plant per generation (1,2). About 20 years ago a single seed of EG4 was planted and propagated independently as strain HEG4. We have used nextgen technology (Illumina and 454) to sequence both HEG4 and EG4 and compare the two genomes to each other and to the reference genome Nipponbare. Rice is an ideal eukaryote to assess the genome-wide impact of the mPing burst because it has a small genome (~350 Mb) and it can be propagated indefinitely by self-pollination. In my very short talk I will summarize the results of this comparative analysis including the quality and quantity of insertion site polymorphism generated by mPing in EG4 vs. HEG4 and how this compares with the accumulation of SNPS over 20 generations. Comparative analysis is also being used to determine whether the burst is confined to the mPing element and/or whether other rearrangements have occurred.

References:

1. Nato, K. et al. (2006) Proc. Natl. Acad. Sci. 103: 17620-17625.
2. Naito, K. et al. (2009) Nature 461: 131-134.

Rotifer genomes as a source for discovery of novel types of TEs and TE-related genes

Irina R. Arkhipova

Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA

Members of the phylum Rotifera are common microscopic invertebrates living in aquatic habitats. Rotifers of the class Bdelloidea are known for the apparent absence of males and meiosis and the ability to survive desiccation at any life stage, while rotifers of the sister class Monogononta are facultatively sexual. I will summarize several ongoing lines of investigation initiated by analysis of rotifer genomic sequences, which subsequently led to discovery of similar genetic elements in other taxonomic groups.

Previously, we reported that telomeres of bdelloid rotifers, as well as selected fungi, plants and protists, harbor an unusual type of retroelements belonging to the Penelope-like (PLE) class. While conventional PLEs are composed of a reverse transcriptase (RT) and GIY-YIG endonuclease (EN) domains, telomere-associated PLEs lack the EN domain altogether. Recently, we found telomere-associated PLEs in monogonont rotifers and in numerous sequenced fungal and protist genomes. Monogononts harbor both conventional (EN-containing) and telomere-associated (EN-deficient) PLEs within the same genome, while fungi and protists possess only telomeric PLEs. Telomere association is inferred from the presence of the flanking species-specific telomeric repeats, and is invariably associated with the lack of EN domain.

These studies are facilitated by the advent of genome sequencing technologies, which allow us to generate draft genomic sequences from bdelloids and monogononts, and to obtain a preliminary overview of TE diversity and composition in rotifer species.

All investigated bdelloid, but not monogonont, genomes contain members of a novel class of RT-related cellular genes (rvt) that do not belong to any TEs, evolve under purifying selection, and occur in syntenic regions. We found rvt genes in all major taxonomic groups, including protists, fungi, animals, plants, and bacteria, although their distribution is quite patchy. I will also discuss regulation of rvt genes by small RNAs and conditions under which their expression rises dramatically.

Biochemical properties of NcRVT protein encoded by a reverse transcriptase-like gene from Neurospora crassa

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

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

NcRVT belongs to a recently discovered class of reverse transcriptase-related cellular genes of fungi, animals, plants and bacteria. NcRVT protein can be purified from vegetative mycelia of wild-type Neurospora strains grown in the presence of 0.1 μg/ml blasticidin, an antibiotic that blocks protein synthesis. An optional ammonium fractionation step may be used to concentrate NcRVT protein prior to centrifugation in sucrose density gradients. Further purification is achieved by ion-exchange chromatography on DEAE Sepharose, where the protein can be eluted in a nearly pure form. In the presence of manganese, purified NcRVT protein has a potent terminal transferase activity with a pronounced preference for ribo- over deoxyribonucleoside triphosphates. Site-directed mutagenesis of a catalytic aspartate residue confirmed its essential role in this activity, ruling out the presence of any other co-purified polymerases. NcRVT apparently cannot initiate polymerization de novo, instead it extends 11-13 nt endogenous primers that remain non-covalently associated with the protein throughout all purification steps. NcRVT does not readily accept exogenous primers or common primer-template combinations, however extension of co-purified tRNA, ribosomal RNA, and small nucleolar RNA fragments has been observed. We are currently investigating the nature of endogenous primers as well as of extension products in vivo. The unusual biochemical properties of a reverse transcriptase-like protein reinforce the connections between diverse RNA-dependent polymerases and challenge the current views on the role of RNA-templated synthesis in living cells.

Transposon proliferation in an asexual parasitoid

Ken Kraaijeveld, Brechtje Zwanenburg, Benjamin Hubert, Cristina Vieira, Sylvia de Pater, Jacques J. M. van Alphen, Peter de Knijff and Johan den Dunnen*

Leiden University Medical Center, The Netherlands

The widespread occurrence of sex is one of the most elusive problems in evolutionary biology. Theory predicts that inefficiency of purifying selection in asexual lineages can lead to uncontrolled proliferation of vertically transmitted transposable elements (TEs), which eventually could cause their demise. To test this prediction, we sequenced the genomes of a sexual (arrhenotokous) and an asexual (thelytokous) lineage of the parasitoid wasp Leptopilina clavipes using next-generation sequencing. We identified transposons of most major classes (including DNA transposons, LTR and LINE-like retrotransposable elements) in both lineages. Coverage depth showed that copy numbers in the thelytokous lineage exceeded those in the arrhenotokous lineage for DNA transposons, but not LTR and LINE-like elements. However, a small number of gypsy-like LTR elements exhibited a four-fold higher coverage in the thelytokous lineage. Quantitative PCR for this gypsy-like TE showed that high loads of this TE were characteristic for 11 genetically distinct thelytokous wasp lineages. Bisulfite sequencing revealed no methylation of the gypsy-like TE in either lineage. These results indicate that Wolbachia-induced parthenogenesis may have differential effects on different types of TEs. Continued accumulation of TE copies may eventually create a significant genetic load.

Comparing the transposable element load in sexual and asexual oribatid mites using whole genome information

Bast, Jens (1); Schaefer, I. (1); Maraun, M. (1); Scheu, S. (1); Kraaijeveld, K. (2)*

1 Georg-August-University Göttingen, J.F. Blumenbach Institute of Zoology and Anthropology, Göttingen, Germany
2 Leiden University Medical Center, Department of Human Genetics, Leiden, Netherlands

The activity of transposable elements (TEs) is likely to vary with the reproductive mode of the host. Sexual reproduction promotes the spread of transposable elements whereas in parthenogenetic lineages TE proliferation is restricted to vertical transmission. Therefore, TEs are expected to be stuck in parthenogenetic lineages and to decline in activity and abundance over time. Loss of TEs may help to explain the long-term persistence of parthenogenetic species, i.e. “ancient asexuals”, which include several taxa of oribatid mites (Oribatida, Acari). As oribatid mites also include species-rich sexual taxa they offer the unique opportunity for analyzing the fate of TEs in sexual and long-term asexual lineages. We compared genome-wide TE abundance and activity bursts in two sexual (Achipteria coleoptrata, Steganacarus magnus) and two parthenogenetic (Hypochthonius rufulus, Platynothrus peltifer) oribatid mite species. We conducted Illumina sequencing, mapped the reads to TE libraries and compared the distances of extant TE copies to the reconstructed ancestral sequences of TE families. We hypothesize that the parthenogenetic species have reduced TE load compared with the sexual species and that bursts of TE activity are more pronounced in the sexual species.

Transposable element copy number in sexual and asexual Daphnia

Sarah Schaack (1), Abe Tucker (2) and Michael Lynch (2)*

1 Reed College, Portland, OR, USA
2 Indiana University, Bloomington, IN, USA

For decades, biologists have been interested in the interaction between recombination rate and transposable element (TE) copy number. Theoretically, recombination rates can lead to increases or decreases in TE copy number, e.g. via the generation of double-stranded breaks, homologue-dependent DNA repair after excision, or non-homologous recombination. In addition, TE copy number changes could lead to changes in rates of recombination. Our work has aimed to tease apart these dueling factors using the Daphnia system, an emerging model for ecological and evolutionary genomics. Daphnia are cyclical parthenogens, meaning most individuals can reproduce sexually or asexually depending on conditions. In some lineages, the ability to reproduce sexually has been lost. Further, the frequency of sex can be manipulated under laboratory conditions. By combining molecular and bioinformatic datasets across lab-reared lineages and natural populations, we will present data on how TEs and sex appear to be influencing each other in this dynamic system. These results have bearing not only on our understanding of TE evolution, but on the evolution of sex, recombination, and mutation rate itself.

Population genetics of transposable elements in Drosophilia

Dmitri A. Petrov

Stanford University, CA, USA

Drosophila endogenous retrovirus regulation in natural populations

Cristina Vieira, Abdou Akkouche, Marie Fablet*

Biometry and Biological Evolution Laboratory, UMR CNRS 5558, University Lyon 1, France

How transposable elements (TE) are controlled in species and populations is a fundamental question for population geneticists that has been the object of an extensive amount of theoretical and empirical work. The recent advances in molecular biology have shown that TE silencing is tightly linked to the production of a particular type of small interfering RNAs, called piRNAs (for piwi interacting RNAs). Natural populations of Drosophila display a wide variability in the patterns of insertions and copy numbers of TEs. They, therefore, are the ideal material to decipher the mechanisms underlying TE dynamics and regulation in a genome. In addition, piRNA-related genes also present a high level of polymorphism in natural populations. By putting these data side by side, and using the example of an endogenous retrovirus of Drosophila, which displays sequence variability in natural populations, I will show how we can take another look at intra-species variation in TE silencing.

Adaptive TE insertions in Drosophila

*Lidia Mateo and Josefa Gonzalez**

Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain.

In previous works, we have shown that TE insertions in Drosophila melanogaster are a considerable source of recently adaptive mutations. We identified 18 recent putatively adaptive TE insertions and for a subset of them we showed that they are associated with signatures of positive selection and with patterns of population differentiation consistent with adaptation to temperate environments. Most of the adaptive TEs identified are located in introns, UTRs and intergenic regions suggesting that they could be affecting the expression of nearby genes. One of these TEs, FBti0019627 is inserted in the 3' UTR of kmn1 gene. Kmn1 is involved in chromosome segregation and neurogenesis. The annotated transcript of kmn1 includes the TE insertion. However, there are some ESTs that do not contain the insertion or contain only a fragment of the TE. Furthermore, the gene located downstream of kmn1, CG11699, although being annotated as a non-overlapping gene, has one EST that overlaps with kmn1 transcript and seems to skip the TE insertion. Interestingly, FBti0019627 has piRNA binding sites and as such those transcripts that incorporate the TE would potentially be able to interact with piRNAs. To investigate the changes associated with FBti0019627 insertion, we perform 3'RACE and quantitative RT-PCR experiments both for kmn1 and CG11699 genes in flies that contain and do not contain the insertion to try to elucidate (i) how many transcripts are present and whether the TE insertion is included in them or not; and (ii) whether the presence of the insertion affect the expression of these two genes. Understanding how the TE insertion affects the expression of nearby genes will help connect the adaptive mutation to its phenotypic and fitness effects.

Ancestral polymorphism and re-introduction of transposable elements in Drosophila

*Elaine Silva Dias and Claudia Marcia Aparecida Carareto**

UNESP – São Paulo State University, Dept. of Biology, São José do Rio Preto, São Paulo, Brazil

In the evolutionary history of the transposable elements processes as ancestral polymorphism and lateral transfer of sequences between species can originate phylogenetic incongruences. We investigated the evolutionary history of the transposon Bari and the retrotransposon 412 in the sequenced genomes of species of melanogaster group and in natural populations of the sibling species D. melanogaster and D. simulans, with traditional phylogenetic and network approaches. The sequences described in D. melanogaster deposited in Repbase were used for searching the elements in this species. For the other species, the sequences deposited were utilized to the identify reference sequences, which represent the most complete and conserved sequences of each element in each species. These sequences were used for searching complete and incomplete copies in each genome. The complete copies were used in phylogenetic (ML and median joining network) and evolutionary (divergence estimates using K2p and the rate of synonymous substitution and age of the copies) analyses. The transposon Bari found in D. ananassae is closer to the elements of the melanogaster complex than to the sequences of D. erecta, an unexpected relationship when compared with the species phylogeny. For understanding non-resolved relationships observed within the melanogaster complex, the network approach was used. The relationships resolved and the values of divergence of the elements between the species suggest transfers of the both Bari and 412 elements between these species. We showed that ancestral polymorphism and transfers between species, that could have occurred by introgression or horizontal transfers, are process present in the history of the Bari and 412 elements in species of the melanogaster group of Drosophila. Here we propose where and when these transfers would have occurred - in Africa during the Pleistocene and before to the worldwide expansion of the D. melanogaster and D. simulans.

Financial support: FAPESP

Long-term and short-term evolutionary impacts of transposable elements on Drosophila

Yuh Chwen G. Lee and Charles H. Langley*

Department of Evolution and Ecology, University of California, Davis, CA

Transposable elements (TEs) are considered genomic parasites and their interactions with hosts have long been likened to the coevolution between host and other nongenomic, horizontally transferred pathogens. TE families, however, are vertically inherited as integral segments of the nuclear genome. This transmission strategy has been theoretically suggested to weaken the selective benefits of host alleles that can repress TE transposition. On the other hand, the elevated rates of TE transposition and high incidences of deleterious mutations observed during the rare cases of horizontal transfers of TEs families between species could create at least a transient process analogous to the influence of horizontally transmitted pathogens. Here, we address this possibility and demonstrate that the long-term evolution of TE-interacting host genes, which can be influenced by both vertically inheritance and occasional horizontal transfer of TEs families, actually have more prevalent evidence of adaptive evolution than immunity genes interacting with nongenomic pathogens. We developed an analytical model and showed that the strict vertical inheritance of new TE families after the invasion can only lead to a limited time period in which host alleles that repress TE transposition enjoy large fitness advantages. Supporting our quantitative conclusion, we used the recently invaded P element in Drosophila melanogaster as a model system by comparing before and after P element invasion populations and did not find unequivocal evidence supporting strong selective impacts imposed by P elements on host TE-interacting genes. Our result suggested that the selective pressure imposed by horizontally transferred of TE families should be limited. The more prevalent and constant interaction with multiple vertically transmitted TE families may be the main force driving the fast evolution of TE-interacting genes, though with a fundamentally different mechanisms from that of the host-pathogen coevolution.

Transposable element discovery and annotation in population using next-gen sequencing data

Fiston-Lavier, A.S. (1), Gonzalez, J. (1,2), and Petrov, D.A. (1)*

1 Department of Biology, Stanford University, US
2 Institute of Evolutionary Biology, (CSIC-UPF), Barcelona, ES

The availability of next-gen sequencing (NGS) data from multiple strains offers an unprecedented opportunity to annotate transposable element (TE) insertions in populations. Here, we present an integrated package called “T-lex” (version 2.0) that (i) ascertains the presence/absence of annotated TE insertions in the sequenced genome(s), (ii) estimates frequencies of TE insertions in multiple strains that are sequenced individually or as a pool, (iii) identifies putatively misannotated TE insertions defining their target site, and (iv) discovers and annotates “new” TE insertions not annotated or absent in the reference genome.

We experimentally validated each T-lex module starting with the TE annotations for 1,826 known TE insertions (Release 5.39) from the reference genome of Drosophila melanogaster in two re-sequenced single strains (W1 and CantonS) and pooled sequence library of 92 strains from the Drosophila Genetic Reference Panel (http://www.hgsc.bcm.tmc.edu/project-species-i-Drosophila_genRefPanel.hgsc). T-lex estimates the population frequency of these TE insertions with high sensitivity (100%) and specificity (97%). T-lex allowed the re-annotation of 148 TE insertions, mostly non-LTR elements that had misannotated poly(A) tails. As expected, we observed conserved insertion sites for DNA and LTR elements. T-lex also inferred with accuracy the target site duplications for more than 705 TE insertions for which the genomic sequence prior the TE insertion were detected. After removing the TE sequences on the chromosome 2L, all the TE insertions known to be present in the W1 strains were re-discovered and correctly annotated.

T-lex, an efficient, accurate and cost-saving tool for TE insertion annotation in NGS population genomic data, offers the opportunity to perform genome-wide population dynamics analyses of TEs and assess their impact on genome dynamics and evolution. T-lex is available for download at: http://petrov.stanford.edu/cgi-bin/Tlex_manual.html.

RNAseq reveals abundant Transposable element expression in the mosquito Anopheles funestus.

Rita Daniela Fernández-Medina (1), Cláudio José Struchiner (1), José M C Ribeiro (2)*

1 Scientific Computational Program PROCC/FIOCRUZ Rio de Janeiro, Brazil
2 Section of Vector Biology, Laboratory of Malaria and Vector Research, NIAID, NIH, Rockville, MD 20852, USA.

Anopheles funestus is, together with An. gambiae, the most important vector of malaria worldwide, among other things, due to their extreme antropophilic habits.

Much is known about the genetics of An. gambiae since its genome was sequenced almost a decade ago. However, little is known about the basic biology, genetics or evolution of An. funestus whose genome has not yet been sequenced. The lack of knowledge about the full sequenced genome of this species makes it difficult to relate the gene content and other genome characteristics to other Anopheline species. As a way to overcome this, Crawford et al (1) sequenced the transcriptome of an adult female of An. funestus, allowing the study of the complete set of expressed genes in this mosquito.

As a first attempt to characterize the transposable elements present in the genome of An. funestus, we have assembled the tanscriptome sequences obtained by Crawford et al and searched for TE derived sequences. We used BlastN and TBlastX on a series of databases to allow the identification and characterization of expressed TEs and presented the results in the form of an Excel spreadsheet with hyperlinks to the obtained results as we have previously done with the genomic TE sequences in An. gambiae (2).

This approach permitted the identification of several sequences belonging to class I and II, including several full length elements. Here, we will show the expression patterns of TEs in An. funestus in a comparative perspective with An. gambiae.

References:

1. Crawford JE, et al. De novo transcriptome sequencing in Anopheles funestus using Illumina RNA-seq technology. PLoS One. 2010 Dec 2;5(12).
2. Fernández-Medina RD,et al. Novel transposable elements from Anopheles gambiae. BMC Genomics. 2011 May 23;12:260.

The multilayer eukaryotic mobilomes: viruses as mobile elements and mobile elements of giant viruses

Eugene V. Koonin

National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD

The rapid advances of comparative genomics make it increasingly clear that genomes of cellular life forms, viruses and mobile elements form a continuum of interacting and recombining genetic elements without any impenetrable barriers. For many groups of viruses infecting eukaryotes, the extremely common retrotranscribing genetic elements in particular, integration into host genomes is an essential stage of the reproduction cycle. However, data is now accumulating to the effect that even viruses for which integration is not essential often insert fragments of their genomes into the host genomes where these sequences can be employed for antiviral defense or recruited for other roles. Many viruses show specific evolutionary relationships with mobile elements, the prime example being the apparent origin of eukaryotic single-strand DNA viruses (e.g. geminiviruses) from plasmids replicating by the rolling circle mechanism. Giant nucleo-cytoplasmic viruses of eukaryotes such as the mimivirus and marseillevirus harbor their own mobilomes which include recently discovered, unusual genetic elements such as virophages and transpovirons. These elements may serve as vehicles of gene transfer between viruses and their eukaryotic hosts. Together these findings attest to the extremely complex, hierarchical organization and dynamics of eukaryotic mobilomes.

The mutualism continuum of retroids

Marcella A. McClure

Dept. of Microbiology and the Center for Computational Biology, Montana State University, Bozeman, MT 59715, USA

Retroids play far reaching roles in shaping genomes. At one end of the mutualism continuum lie the beneficial roles of Retroids in reproduction and development, in contrast the other end is signified by deleterious disease implications.

Retroids provide vital cellular functions from LTRs upregulating developmental genes and altering transcription of pluripotent stem cells, to providing encoded proteins with new functions (e.g., syncytin). ERV expression is necessary for the development from oocyte to embryo, and for plasticity of the mouse genome. While mobilization of some ERVs in the mouse alter the transcriptome, the integrase protein of another ERV plays a critical role in maternal mitochondrial inheritance. The evolution and intracellular dynamics among and between Retroid, and host genes requires new tools.

Various results from the Genome Parsing Suite, (GPS), which identifies and classifies all Retroids down to the smallest statistically significant signal in any sequenced genome are presented. While most ERVs have accumulated both stop-codons and frame-shift mutations, many are still active in the cell. The extent to which stop codon and frame-shift mutations can be overcome by translational recoding, and gene complementation is an area ripe for computational exploration.

The GPS is being transformed into RASCAL, (short for “a system for Retroid Agent diSCovery and anALysis). RASCAL will be an open source, web-based research platform to study the co-evolution of Eukaryotic and Retroid genomes. RASCAL provides all Retroid information from all sequenced genomes, and a variety of tools to test models of viability and mechanisms of interactions among and between these agents and the genome. RASCAL will include new methods for assessing the potential for gene complementation among populations of Retroids, and new algorithms for detecting intra- and inter-genic recombination. RASCAL will allow users to study both the long-term mutualism that has co-evolved between Eukara and Retroids as well as more recent species-specific relationships.

"Simple" DNA transposons: scratching the surface

Kapitonov, V.V., Bao, W. and Jurka, J.*

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

DNA transposons in eukaryotes are commonly viewed as simple entities. Known autonomous transposons encode typically one transposase protein and sometimes a second protein (e.g. Myb in Harbingers, unclassified proteins in MuDR and En/Spm, and RPA in Helitrons). The only exclusion from such simplicity is present in self-synthesizing Polintons, which code for several proteins. It was even suggested that Polintons have evolved from DNA viruses.

Here, we will present two novel groups of DNA transposons, called Inton and Enton that code for 5-6 conserved proteins. Inton and Enton are present in fungi and they belong to the IS3EU and En/Spm superfamilies respectively, which usually code for only one or two proteins. We will consider Inton and Enton transposons as first examples of complex virus-like objects evolved from "simple" DNA transposons.

We will describe also numerous families of DNA transposons that hijacked protein from their hosts. These proteins are Fen1, lambda and DEDDh nucleases, RecQ and SNF2 helicases, HECT, H3 histone, alpha kinase. Moreover, the same proteins have been independently captured by transposons that belong to different superfamilies: H3 histone in IS3EU and hAT, SNF2 helicase in Inton and Enton; DEDDh exonuclease in P and piggyBac.

We will discuss finding of the conserved host proteins in DNA transposons as powerful hallmarks related to badly understood aspects of transposition mechanism and regulation of transpositions by the host proteins.

Digging deeper in time; approaches to study ancient transposable elements

Arian Smit, Robert Hubley, Juan Caballero, Travis Wheeler, Tom Jones and Sean Eddy*

Institute for Systems Biology, Seattle, WA, USA

The origin of merely half of mammalian genomes and just a quarter of avian genomes can currently be precisely explained, mostly through the expansion of self-replicating fragments of DNA. As the transposed copies of these elements tend to wither away at a neutral rate, they become hard to detect at very old age, and our community has long realized that the great majority of what is left unexplained in amniote genomes must have its origin in (truly ancient) transposable elements as well. We'll describe the nature of transposable elements in the proto-eutherian genome and discuss our efforts to unveil and describe even older elements. Among our methods is the use of reconstructed ancestral genomes to both discover novel ancient elements and define the location of ancient copies with RepeatMasker. Superposition of these annotations on extant species can dramatically expand repeat coverage of these genomes. We further are increasing the sensitivity of RepeatMasker in a variety of ways, most promising of which is the use of HMMER as a search engine to allow use of profile HMMs of interspersed repeats.

Transposable elements: maintenance of centromeric regions and chromosome B origin

Sarah G Oliveira (1), Rita C. Moura (2) and Cesar Martins (1)*

1 Departamento de Morfologia, Instituto de Biociência, Universidade Estadual Paulista/UNESP, Botucatu, São Paulo, Brazil
2 Departamento de Biologia, Instituto de Ciências Biológicas/ICB, Universidade de Pernambuco/UPE, Recife, Pernambuco, Brazil

Based on C-banding and fluorescent in situ hybridization (FISH) of C0t-1 DNA in Coprophanaeus cyanescens (Coleoptera, Scarabaeidae) it was observed an accumulation of repeated DNAs in the pericentromeric region of all chromosomes, including the X and the B chromosomes. Besides the centromeric enrichment of repeated DNAs, the A and the X chromosomes also contained repeated DNAs covering the entire long arms of the chromosomes. Considering previous data indicating that the B chromosome accumulated several repeated DNA classes, including transposable elements; cytogenetic mapping of two LINE retrotransposon probes isolated from the C. cyanescens genome was performed and showed blocks in the centromeric regions of autosomal pairs and B chromosome, indicating the exchange of genetic material between the A and B chromosomes. This analysis indicates an evolutionary relationship between the A complement and the B chromosome at least in the centromeric area. These transposable elements may have been maintained in the genome of C. cyanescens due to the functional role they play in the maintenance of the centromeric regions. The absence of transposable elements in the sex chromosomes suggests that the sex differentiation occurred before the dispersion of these transposable elements into the genome. Subsequently, the suppression of the recombination could produces the differences observed in the distribution of transposable elements between the A complement and the sex chromosomes. These results imply that these transposable elements are involved in the maintenance of the centromeric regions and contributed to the B chromosome origin.

Large scale DNA editing of retrotransposons accelerates mammalian genome evolution

Shai Carmi (1), George M. Church (2), Erez Y. Levanon (1)*

1 Bar-Ilan University, Israel
2 Harvard Medical School, Boston, MA

Genomes are thought to evolve by gradual accumulation of mutations leading to acquisition of novel functions. One of the major sources of mutation is the retrotransposition of mobile elements; their ability to amplify DNA fragments throughout the genome had an important role in evolution and led to formation of new exons, genes, promoters, and to rewiring of gene regulation networks1-6. However, it is unclear how such a large repertoire of functions emerged from a relatively limited number of source sequences. Here, we show that DNA editing, an antiviral mechanism, accelerates the evolution of mammalian genomes by large-scale modification of their retrotransposon sequences. We found numerous pairs of retrotransposons containing long clusters of G-to-A mutations that cannot be attributed to random mutagenesis. These clusters are the hallmark of APOBEC3 activity, a potent antiretroviral protein family with cytidine deamination function. We detected DNA editing in thousands of elements across different mammalian genomes and in various retrotransposon families. Since DNA editing simultaneously generates a large number of mutations, each affected element begins its evolutionary trajectory from a unique starting point, thereby increasing the probability of developing a novel function. Our findings suggest a potential mechanism for domestication of retrotransposons and change our understanding of the processes that shaped mammalian genome evolution.

High-throughput DNA methylation analysis of repetitive elements by hairpin bisulfite sequencing

Julia Arand (1), David Spieler (2), Tommy Karius (1), Daniela Meilinger (3), Alexander Meissner (4), Thomas Jenuwein (5), Guoliang Xu (6), Heinrich Leonhardt (3), Verena Wolf (2) and Jörn Walter (1)*

1 University of Saarland, Biological Sciences, Genetics / Epigenetics, Saarbrücken, DE
2 University of Saarland, Computer Science Department, Saarbrücken, DE
3 LMU München, Biozentrum Martinsried, Dept. Bio II., Martinsried, DE
4 Broad Institute, 7 Divinity Avenue, Rm 207, Cambridge, MA, 02138, USA
5 MPI for Immunology and Epigenetics, Stübeweg 51, 79108 Freiburg, DE
6 Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, CN

Repetitive elements comprise a large portion of the mammalian genome and they are major targets for DNA methylation mediated epigenetic silencing. DNA methylation in mammals is predominantly located at CpG positions and mediated by three different DNA methyltransferases (Dnmts). 80% of these methylated CpG positions are lying at sequences of repetitive elements. Due to high rate of deaminations at methylated cytosines the CpG to TpG mutation rate at repetitive elements is rather high.

In our analysis we monitored the methylation of CpGs at IAP-LTR, L1-5'UTR, B1 elements and major satellites on both DNA strands simultaneously using the hairpin bisulfite method. With high throughput 454 pyrosequencing we analysed more than 100.000 complementary CpGs of repetitive elements at random individual chromosomal loci of different pluripotent cell lines and somatic cells.

Our study comprises quantitative and qualitative novel insights into the distribution of symmetrical CpG methylation across repetitive elements. Excluding all mutated CpG positions, which appear at single strand bisulfite sequencing as unmethylated CpG position, we could revise the global methylation level for B1 elements from former single strand bisulfite data, which show only minor methylation, to a general high methylation rate of 60 to 85% in all analysed cells. Moreover, we show differences in the abundance of hemimethylated CpG positions - element and cell type specific. Furthermore, by analysing ESCs with KOs of single Dnmts followed by Hidden Markov Modelling, we show different efficiencies of the Dnmts at different repetitive elements. Finally, we found next to CpG methylation CpA methylation only in pluripotent cells and only at specific CpA positions at major satellites, not at other repetitive elements and not in differentiated cells.

Repetitive sequences: beyond RepBase and RepeatMasker

Wojciech Makałowski

Institute of Bioinformatics, University of Münster, Münster, DE

With the increased diversity of the sequenced genome, the number of described repetitive elements is growing exponentially. RepBase presently consists of more than twenty-four thousands sequencing from over seven hundred species. However, too often the analyses of newly sequenced genomes are limited to simple run of RepeatMasker on RepBase reference data. At the same time many excellent tools for the repetitive elements and especially transposable elements analyses have been developed. I will overview the whole specktrum of different programs and analytical systems including ab initio discovery, genomic pipe lines and meta analysis approaches.

Improved transposable element detection with profile HMMs in nhmmer

Travis Wheeler (1), Sean Eddy (1), Rob Finn (1), Robert Hubley (2), Tom Jones (1) and Arian Smit (2)*

1 HHMI, Janelia Farm Research Campus, Ashburn, VA 20147, USA
2 Institute for Systems Biology, Seattle, WA 98109, USA

State-of-the-art TE detection is homology based [1], with individual TEs usually identified based on similarity to a consensus sequence constructed to represent a known TE family [2]. While current TE detection methods cover substantial portions of many eukaryotic genomes (e.g. ~45% of human [3]), older TE instances often endure random mutatation to the extent that they are not recognizable using single-sequence homology search, resulting in underestimated masking. Profile methods are known to improve sensitivity over single sequence search [4], with profile HMMs in particular leveraging the additional information content in position-specific residue and indel variability.

nhmmer, a new tool in the HMMER3 suite, brings the power of profile HMMs to DNA homology search, with speed faster than cross_match and only 5-10x slower blastn with sensitive settings. The acceleration over past profile HMM implementations is due to a combination of a heuristic filtering pipeline and an efficient vector-parallel implementation [5], and has negligible impact on sensitivity.

Tests were performed using a benchmark made up of more than 14,000 "ghost" repeat instances, not found in default RepeatMasker search but supported by comparative genomics. While keeping a low false-discovery rate, we find a several percent increase in coverage over both blastn and cross_match through both (a) identification of TE instances not found by single-sequence homology methods and (b) increase in hit lengths.

References:

1. A Smit, R Hubley, P Green, RepeatMasker Open-3.0 1996-2004. http://www.repeatmasker.org
2. J Jurka, et al, Repbase Update, a database of eukaryotic repetitive elements. Cytogentic and Genome Research. 110, 462-467 (2005).
3. E Lander, et al, Initial sequencing and analysis of the human genome. Nature. 409, 879–888 (2001).
4. J Park, et al, Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. Journal of Molecular Biology 284, 1201–1210 (1998).
5. Eddy S, http://hmmer.org.

Transposable element diversification and evolution

David Pollock, Vijetha Vemulapalli, Jason de Koning*

University of Colorado School of Medicine

Transposable elements are often considered to diversify according to a "master element" model, whereby only one or a few types of elements are active at any one time. Recent experimental evidence, however, indicates that human SINES and LINES have far more than just a few active elements. This newer concept of transposable element evolution has not yet been incorporated into analytical models. Here, we will present research on the utility of what we call the MOM, or "Moran Oligarchy Model". These studies use Bayesian approaches to obtain probabilistic statements about family membership and evolutionary history from an ancestrally duplicated sequence. We show that they can provide improved inference on the timing and diversification of subfamilies, on the substitution process after duplication, and on the evolutionary history of ancestrally duplicated sequences. We apply these techniques to study the evolution of Alus in primates and SINE1 in opossums. We show that we obtain sensible agreement between subfamily diversification and diversification times of different species, and that we can discover previously hidden subfamilies and patterns of transposable element expansion. We also use these novel analytical methods and understanding of subfamily structure to analyze substitution processes along the human genome.

Discovery and analysis of transposable elements with shotgun proteomics

C. Bamberger, S. Pankow, and J.R. Yates, III*

The Scripps Research Institute, La Jolla, CA

DNA sequences of complete and partial transposable elements constitute a large proportion of the genome in most multicellular eukaryotes, however transposable element-derived proteins represent only a small fraction of the proteome. The high sequence variability and large number of possible genomic origins of transposon transcripts and low protein abundance further complicate the analysis of transposon-specific proteins. Here we present a new mass spectrometry-based proteomic approach to detect transposable element-derived proteins. The approach allowed us to identify and quantify transposon proteins (1) in the germ cell compartment of the starlet sea anemone Nematostella vectensis and (2) in different murine and human primary cells and cell lines. We show that exposure of adult N. vectensis polyps to UV-light, a natural inducer of mobile element transposition, deregulates transposon protein levels. In addition, mass spectrometry-based quantification allowed reliable measurement for example of L1-derived proteins in murine and human cell lines upon various perturbations. Thus the method enables the direct, high-throughput and unbiased analysis of the translated mobilome and therefore may open up new possibilities to address the biology of transposable elements.

POSTER ABSTRACTS

(1)

From Paracoccidioides genome annotation to species concepts

Anna Muszewska, Marta Hoffman-Sommer, Marcin Grynberg*

1 Laboratory of Bioinformatics and Systems Biology, ICM, University of Warsaw, Zwirki i Wigury 93, 02-089 Warsaw, Poland
2 Institute of Biochemistry and Biophysics, Pawinskiego 5A, 02-106 Warsaw, Poland

Paracoccioides is one of the dimorphic fungi causing systemic infections, even in otherwise healthy individuals. Currently three strains classified as Paracoccioides brasiliensis were sequenced by The Broad Institute. Our identification and classification of transposons in these genomes resulted in a collection of mobile elements belonging to all basic transposable element (TE) types. Transposons constitute approximately 8-9% of the P. brasiliensis Pb03 and Pb18 genomes and two times more in the P. brasiliensis Pb01 genome [1]. The differences are not only in the ratio of each group of elements in the total TE content but also in the abundance of each major type of element. This result is in concordance with the recent separation of P. basiliensis and P. lutzii sibling species upon genomic and physiological analyses [2].

The subject of species concept is one of the most discussed in theoretical biology [3]. In my opinion, the aforementioned example of P. brasiliensis is an argument in favor of considering TEs as components of the definition of species. Of course, the degree of divergence among strains of a species of interest has to be individually defined taking into account data from related taxa. This criterion has the same inconvenience as other sequence based definitions. The advantage is that it incorporates data from almost all genomes, without previous annotation and thorough assembly. The overall pattern of TEs in a species seems to be somehow fixed with single elements being activated and widespread.

References:

1. Desjardins CA et al. (2011) Comparative Genomic Analysis of Human Fungal pathogens causing Paracoccidioidomycosis. PLoS Genetics.
2. Teixeira MM et al. (2009) Phylogenetic analysis reveals a high level of speciation in the Paracoccidioides genus. Mol Phylogenet Evol
3. Taylor JW et al. (2000) Phylogenetic species recognition and species concepts in fungi. Fungal Genet Biol.

(2)

The expression of Alu increases in the acquisition of drug resistance in human leukemic cells

Shirou Yamada, Akira Aizawa, Takashi Kanazawa, Yasuko Kobayashi, Takumi Takizawa, Hirokazu Arakawa*

Department of Pediatrics, Graduate School of Medicine, Gunma University, Gunma-ken, Japan

Development of cellular chemotherapy resistance is a major problem of cancer treatment. We have previously reported that resistance against anticancer drugs, nucleotide analogues (NAs) such as cytosine arabinoside (Ara-C) in leukemic cell-lines was readily introduced by culturing them with the drugs. The irreversible drug resistance was developed by culturing a human leukemia cell line, Jurkat cells with NAs at LD50 concentrations during the first week and keeping them in normal media for another week. The development of the drug resistance with this procedure is unlikely to be survival of pre-existing drug-resistant cells or occasional mutations in genes related to susceptibility to the NAs, because most cells were not killed and almost simultaneously became refractory to the NAs. DNA microarray analysis showed no significant changes in the gene expression profiles before and after the development of drug resistance. So we are interested in the genomic methylation state of drug resistant cells, and the behavure of alu. To evaluate the whole genomic methylation status, we determined the expression amount of total Alu and AluYa5 RNA by quantitative RT-PCR. The total Alu includes AluJ, AluS and AluY all of which are amplified with a signle set of consensus sequence primers, and AluYa5 with specific primers. The expression levels of LINE-1 RNA were also determined. The expression of the total Alu and AluYa5 increased 24 hours after the treatment with Ara-C. The expression of total Alu is over 100 times higher than AluYa5, but the increment of AluYa5 by the culture with Ara-C is significantly larger than that of consensus Alu (by 5.3 vs. 1.4 times). LINE-1 expression did not change by Ara-C.

Consequently these results suggest that the mechanisms of the development of the drug resistance to NAs in leukemia cells may include methylation status of Alu repeats especially young AluYa5.

(3)

Ty1/Copia retrotransposon, LIRE, in Lilium

Nam-Soo Kim (1,2), Jae-Han-Son (1), Sung-Il Lee (1,2), Yeong-Sup Lee (1,2) and Kyong-Cheul Park (3)*

1 BK21 National Program, Korea
2 Department of Molecular Biosciences, Kangwon National University, Korea
3 Institute of Bioscience and Biotechnology, Kangwon National University, Korea

Novel Ty1/Copia retrotransposons, LIREs (Lily Retrotransposon), were isolated from a fosmid library derived from Lilium lancifolium. The lengths of three LIREs, LIRE-1, LIRE-2, and LIRE-3, were 5325, 4270, and 6972 bp, respectively. The LTR sequences varied in length from 205 to 214 and had 2-bp inverted repeat (5'-TG ...CA-3') at the ends. Primer binding site (PBS) was flanked in 3-end of the left LTR and polypurine tract was flanked in the 5'-end of the LTR. Sequence in the Gag gene was too decayed to deduce protein function. The Pol ORF contained the INT, RT, and RH domains in the order of Ty1/Copia element. The INT had DDE motif that was most similar to the transposase of bacterial IS21. The LIREs were highly abundant in the genus Lilium, but were absent in other species in the Liliaceae. Chromosomal distributions of the LIREs were different between L. lancifolium and L. longiflorum. In phylogentic analysis, LIREs showed close relationship with Tgmr from soybean and Tst1 from potato.

(4)

Identification, characterization and influence of transposable elements in the flax (Linum usitatissimum L.) genome

Leonardo M. Galindo, Michael K. Deyholos*

Department of Biological Sciences, University of Alberta, AB, Canada

Transposable elements (TEs) are powerful engines for gene evolution and regulation. An unmasked, whole-genome shotgun assembly of flax (www.linum.ca) was searched to identify transposable elements. Similarity and de-novo analyses showed that complete and partial transposable elements covered more than 20% of the full genome. The LTR Copia and Gypsy elements were the most prevalent elements while DNA transposons constituted a smaller proportion. Selected predicted agronomical important genes that contained TEs either as insertions, chimeras or in close proximity to them, were further characterized and confirmed by PCR and sequencing. The analyzed regions showed that TEs inside genes are mostly located in introns. Several genome sections showed evidence of TEs related to gene families like disease resistance genes, and also of TE clustering. From the most abundant and diverse LTR retrotransposons many were highly fragmented, and others seemed to carry unrelated TE regions between the LTRs and resemble Large Retrotransposon Derivates (LARDs). Some elements were found to be recently active as shown by their high LTR similarity, conserved coding regions and their presence on EST libraries.

(5)

Automatic classification of plant LTR retrotransposons in sequence data obtained by next generation sequencing

Pavel Neumann, Petr Novak, Jaroslav Steinhaisl, and Jiri Macas*

Biology Centre ASCR

Next generation sequencing represents very powerful tool to explore repetitive fraction of large and complex genomes which are characteristic of many plant species. However, enormous number of reads along with their short length hamper analyzes of the sequence data using traditional software tools. We have recently developed a pipeline for clustering of the reads into clusters representing whole repetitive elements or their fragments. At the final step of the pipeline, the sequences belonging to individual clusters are assembled into contigs which are convenient for subsequent processing using various bioinformatic tools. Here we show that the contig sequences can be used for fast and reliable identification of LTR retrotransposons and their classification into major lineages. Simple comparison of nucleotide sequences between the contigs and previously identified elements is not sensitive enough because their divergence is often too high. On the other hand, the protein sequences of individual polyprotein domains are sufficiently conserved to be detected even in very diverged families of LTR retrotransposons. Thus, contigs possessing polyprotein-coding sequences can be identified by FASTY program which compares DNA sequences of the contigs with the databases of polyprotein domains. In order to build these databases, we selected representatives of all major lineages of both Ty3/gypsy and Ty1/copia elements from previously identified and classified sequences. The output of FASTY is parsed to extract data relevant for classification of the contigs into appropriate retrotransposon lineages. Contigs showing high similarity to the database sequences can be reliably classified simply based on the best hits. However, to verify the best hit-based classification and to classify more divergent sequences, the extracted polyprotein domains are always subjected to phylogenetic analysis. The scripts and databases used in the clustering and classification pipeline have been implemented into a GALAXY platform accessible at our server (http://w3lamc.umbr.cas.cz/lamc/resources.php).

(6)

Transposable elements in banana species (Musaceae family)

Petr Novak (1), Eva Hribova (2), Pavel Neumann (1), Andrea Koblizkova (1), Susanne Konig (3), Nils Stein (3), Stefan Taudien (4), Jaroslav Dolezel (2), Jiri Macas (1)*

1 Biology Centre ASCR, Institute of Plant Molecular Biology, Budejovice, Czech Republic
2 C.R. Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Olomouc, Czech Republic
3 Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany
4 Genome Analysis, Leibniz Institute for Age Research, Fritz Lipmann Institute (FLI), Jena, DE

Bananas and plantains are perennial giant herbs grown in the tropical and subtropical countries. Bananas belong to genus Musa, which has traditionally been divided into four sections based on morphological similarities and basic chromosome number: Australimusa (2n = 20), Callimusa (2n = 20) including M. beccarii (2n = 18), Eumusa (2n = 22) and Rhodochlamys (2n = 22) and genus Ensete (2n=18).

To elucidate the genome composition and identify the interspecific differences we explored the genomes in six representatives of Musaceae family using the 454 sequencing. The coverage above 10% enabled characterization identification of repeats which form 30-45 % of the Musaceae genomes. Remaining part of the genome consist of unidentified repeats (~30%) and low or single copy sequences (26-45%).

The major repeats found in Musaceae genomes were LTR-retrotransposons. The most abundant Ty1/copia represented between 13-30% of the genomes while less frequent Ty3/gypsy represented 6-9% of the genomes. Identification of individual lineages of transposable elements was carried out using similarity search against the custom protein database. Identified Ty1/copia sequences were classified as members of Maximus/SIRE, Angela, Tork, Ale, Ivana and Tork with average genomic propotions 15.7, 6.2, 1.4, 0.5, 0.4 and 0.1% respectivelly. The major Ty3/gypsy elements were members of Chromovirus and Tat with genomic proportions 7.5 and 0.4 % respectively.

Phylogenetic analysis based on the reverse transcriptase domain of LTR retrotransposons revealed amplification events which contributed to diversification of Musaceae genomes.

(7)

Folding and structural analysis of mobile element proteins

György Abrusán (1), András Szilágyi (2), Balázs Papp (1), Yang Zhang (3)*

1 Synthetic and Systems Biology Unit, Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary
2 Institute of Enzymology, Hungarian Academy of Sciences, Budapest, Hungary
3 Center for Computational Medicine and Bioinformatics, Department of Biological Chemistry, Medical School, University of Michigan, Ann Arbor, MI, USA

Despite the considerable amount of research on mobile elements, no large-scale structural analyses of the TE proteome have been performed so far, due to limitations of the available structure prediction tools. Such analyses become feasible due to recently developed ab initio protein modeling methods, which, in combination with template based modeling, are capable of predicting reliable structures for certain proteins even with very little similarity to known structures. We folded 840 protein fragments from a representative set of DNA and non-LTR transposable elements, and provide a basic description of their characteristics such as SCOP domain composition and predicted function. We find that DNA transposon proteins tend to have a lower contact order than randomly selected reference proteins. Low contact order is a signature of rapid protein folding, thus our finding indicates that in DNA transposons, folding speed is under selection. Since rapid folding evolves to avoid aggregation of highly expressed proteins, our finding indicates that protein aggregation is a significant selective force on DNA transposons. Additionally, we tested for the presence of disordered regions in more than 5000 TE protein sequences, and find that 'gag' proteins of retrotransposons contain strikingly high amounts of disordered regions. Disorder is a characteristic of chaperones, and the TEs where such proteins have been analyzed experimentally (e.g. human L1 repeats) show that they do have chaperone function. Thus, despite the fact that the sequences of gag proteins show very little conservation across various retrotransposon families, the presence of disordered regions and possibly their chaperone function is likely to be conserved.

(8)

The human transcription factor DUX4 recognizes many LTR elements and can activate their transcription.

Janet M. Young, Linda N. Geng, Zizhen Yao and Stephen J. Tapscott*

Fred Hutchinson Cancer Research Center, Seattle, WA, USA

The human double-homeodomain retrogene DUX4 is normally expressed in the testis. When aberrantly expressed in muscle it causes facioscapulohumeral muscular dystrophy (FSHD), although the mechanism of pathogenesis is unknown. We recently performed ChIP-seq in myoblasts over-expressing DUX4, finding at least 62,000 bound locations in the human genome; these contain a strongly enriched sequence motif. Experimental evidence suggests DUX4 is a positive regulator of transcription and microarray expression analyses show that DUX4 induces germline genes.

Even using standard methods that consider only uniquely-mapped sequence reads, it was clear that ~1/3 of DUX4's binding sites occur in MaLR-LTR elements. We further investigated repetitive element enrichment using methods that are less affected by mapability and found that numerous LTR types were enriched, including many from the MaLR family as well as some ERVL and ERVK types.

We performed RNA-seq on myoblasts over-expressing DUX4, on muscle tissue from an FSHD patient and a control individual, and on a normal testis sample. Preliminary analysis revealed several LTR-initiated transcripts induced by DUX4, and also suggested that satellite transcripts are greatly enriched in DUX4-containing samples. These findings suggest a role for repetitive element transcripts in muscle disease and a role for DUX4 in the biology of normal testis.

Intriguingly, DUX4 itself is known to be a recently evolved multi-copy gene present in variable numbers in different primate genomes and we are investigating whether its evolution parallels that of the repetitive element classes it recognizes. DUX4 also has a number of close paralogs whose function is largely unknown: these might also be excellent candidate retroelement-binding proteins.

This study was supported by NIAMS R01AR045203, NINDS P01NS069539, and Friends of FSH Research; Janet Young is also supported by an NSF Career Award to Harmit Malik.

(9)

Are significant differences in retroid content discernible between individuals?

Melissa S. Robertson, Marcella A. McClure*

Montana State University, USA

The studies presented here focus on detailed analysis of the Retroid content of the primates to provide a context to determine if significant differences between the human composite reference genome and the Venter genome can be discerned.

The Retroid content of these genomes was identified and classified using the Genome Parsing Suite (GPS). This method is unique as it identifies Retroid genomes by finding statistically significant Reverse Transcriptase (RT) protein sequences and then determines the Retroid genome in an “RT outward” manner. The GPS produces user-friendly files that facilitate the comparison of Retroid content within and among genomes. It is predicted that each individual has a unique LINE-1 that contributes to human individuality. Are there any unique Retroids in the single individual genome compared to the composite genome? Can a Venter specific LINE-1 be identified? The full Retroid content of the genomes for Homo sapiens (human composite and Venter genomes), Pan troglodytes (Chimpanzee), Gorilla gorilla (Ape) Macaca mulatta (Rhesus macaque), Callithrix jacchus (Marmoset) and Pongo abelii (Orangutan) are are assessed in detail by the GPS in this study.

(10)

The MOV10 helicase inhibits LINE-1 mobility

*Xiaoyu Li, Jianyong Zhang, Rui Jia, Vicky Cheng, and Chen Liang**

Lady Davis Institute-Jewish General Hospital, Department of Medicine, McGill University, Montreal, Quebec, Canada

Long interspersed element 1 (LINE-1) is an autonomous non-LTR retrotransposon. Its replication poses a great threat to the integrity of the host genome. Host cells therefore have evolved mechanisms to control LINE-1 mobility. Here, we report that a helicase named MOV10 effectively suppresses LINE-1 transposition in culture cells. The helicase motifs GKT (Walker A) and DEAG (Walker B) are indispensible for this function of MOV10. Further studies show that MOV10 diminishes the level of LINE-1 RNA by acting at a post-transcriptional stage. These data suggest that in addition to inhibit retroviruses including human immunodeficiency virus type 1, MOV10 also contributes to the cellular control of LINE-1 replication.

(11)

Exploring endogenous retroviruses in the crocodilian genome

Amanda Yoon-Yee Chong (1), Sally Isberg (1,2), Lorna Melville (3), David A. Ray (4), Travis C. Glenn (5,6), and Jaime Gongora (1)*

1 Faculty of Veterinary Science, University of Sydney, Sydney, NSW, Australia
2 Porosus Pty Ltd, Palmerston, NT, Australia;
3 OIC Berrimah Veterinary Laboratories Department of Resources, Darwin, NT, Australia;
4 Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS, USA;
5 Department of Environmental Health Science, University of Georgia, Athens, GA, USA;
6 Georgia Genomics Facility, University of Georgia, Athens, GA, USA

The saltwater crocodile (Crocodylus porosus) is one of two species of crocodile found in Australia and the only one that is commercially farmed. Runtism is one of the major causes of juvenile mortality [1], and is suspected to be related to inherent genetic factors, of which endogenous retroviruses might be a possibility. Endogenous retroviruses (ERVs) are inherited copies or remnants of exogenous retroviruses that have been integrated into germline cells and therefore passed on to subsequent generations. The broader scale of this project is to identify and characterise endogenous retroviruses in the saltwater crocodile and asses their significance for diseases in farmed populations. Here we will be discussing the screening and sequencing of full length ERVs from a saltwater crocodile genomic library.

DNA fragments from the retroviral pro-pol gene region were used as probes to screen a saltwater crocodile bacterial artificial chromosome (BAC) library. A BAC library from the gharial (Gavialis gangeticus), a sister subfamily to crocodiles, was also screened using the same probes as a control for comparison purposes. A selection of positive clones was sequenced on the Roche GC FLX 454 sequencing platform. DNA sequence contigs will be compared to known ERV sequences to identify full length ERVs and their insertion sites for further analysis. Previous investigations into the ERV complement of crocodilians have revealed two major lineages of ERVs, one of which is unique to the family Crocodylidae (CERV1), and another which is common among a crocodilian species (CERV2 [2]. The data available to date supports the notion that ERVs are present throughout the crocodile genome in many copies of both ancient and recent ERV insertions.

References:

1. Isberg, S., Shilton, C. & Thomson, P. (2009) RIRDC, Union Offset Printing, Canberra
2. Jaratlerdsiri, W., et al. (2009) Journal of Virology, 83, 10305-8.

(12)

Transposable elements: maintenance of centromeric regions and chromosome B origin

Sarah G Oliveira (1), Rita C. Moura (2) and Cesar Martins (1)*

1 Departamento de Morfologia, Instituto de Biociência, Universidade Estadual Paulista/UNESP, Botucatu, São Paulo, Brazil
2 Departamento de Biologia, Instituto de Ciências Biológicas/ICB, Universidade de Pernambuco/UPE, Recife, Pernambuco, Brazil

Based on C-banding and fluorescent in situ hybridization (FISH) of C0t-1 DNA in Coprophanaeus cyanescens (Coleoptera, Scarabaeidae) it was observed an accumulation of repeated DNAs in the pericentromeric region of all chromosomes, including the X and the B chromosomes. Besides the centromeric enrichment of repeated DNAs, the A and the X chromosomes also contained repeated DNAs covering the entire long arms of the chromosomes. Considering previous data indicating that the B chromosome accumulated several repeated DNA classes, including transposable elements; cytogenetic mapping of two LINE retrotransposon probes isolated from the C. cyanescens genome was performed and showed blocks in the centromeric regions of autosomal pairs and B chromosome, indicating the exchange of genetic material between the A and B chromosomes. This analysis indicates an evolutionary relationship between the A complement and the B chromosome at least in the centromeric area. These transposable elements may have been maintained in the genome of C. cyanescens due to the functional role they play in the maintenance of the centromeric regions. The absence of transposable elements in the sex chromosomes suggests that the sex differentiation occurred before the dispersion of these transposable elements into the genome. Subsequently, the suppression of the recombination could produces the differences observed in the distribution of transposable elements between the A complement and the sex chromosomes. These results imply that these transposable elements are involved in the maintenance of the centromeric regions and contributed to the B chromosome origin.

(13)

The dying mobilome: analysis of the deteriorated transposable elements in the genome of An. gambiae.

Rita Daniela Fernández-Medina (1), José M C Ribeiro (2), Claudia Aparecida Carareto (3), Cláudio José Struchiner (1)*

1 Scientific Computational Program PROCC Rio de Janeiro, Brazil
2 Section of Vector Biology, Laboratory of Malaria and Vector Research, NIAID, NIH, Rockville, MD 20852, USA
3 Laboratory of Molecular Evolution, Department of Biology, UNESP, São Paulo State University, 15054-000 São José do Rio Preto, SP, Brazil

Together with their ability to mobilize, an almost universal feature of transposable elements in nowadays eukaryotic genomes is their inability to transpose by themselves, mainly as the result of sequence degradation (either by mutations or deletions). Most of the elements are then, either inactive or non-autonomous. Considering that the bulk of some eukaryotic genomes derive from TEs, they have been conceived as TEs graveyards. It has been shown that once an element has been inactivated, it progressively accumulates mutations and deletions at neutral rates, until completely losing its identity or be lost from the host genome. However, it has also been shown that these “neutral sequences” might serve as raw material for processes of exaptation or domestication by host genomes. We have previously analyzed the TE families present in AnoTExcel, an online Anopheles gambiae-TE specific database containing the general features of the TE landscape in the malaria mosquito genome. Now, we have analyzed the sequence structural variations, the nucleotide divergence and the pattern of insertions and deletions of several superfamilies of TEs belonging to both class I (LTRs and NLTRs) and II in the genome of An. gambiae, aiming at describing the landscape of deterioration of these elements in this particular genome.

Our results describe a great diversity of patterns of deterioration, indicating lineage-specific differences including the presence of Solo-LTRs (43% of the total LTR amount found in AnoTExcel), 5´-deleted NLTRs (up to 99% of the NLTRs constituted by deteriorated sequences) and several non-autonomous and MITEs class II families (84% of which corresponds to truncated sequences). Interestingly, we found fragments of NLTRs corresponding to the RT domain which preserves high identity among them suggesting a possible remaining genomic role for these domains.

(14)

Mechanisms that limit the impact of non-LTR retrotransposons in mosquitoes

Joseph Yeoh and Colin Malcolm*

Queen Mary, University of London, Mile End Road, London, E1 4NS, UK

The persistence of high copy number non-LTR retrotransposons continues to intrigue geneticists. The question of how and why they became abundant is still unsatisfactorily unanswered. For example, L1s contribute as much as 17% to the human genome and active L1 elements have been known to cause diseases when they retrotranspose (Moran and Gilbert, 2002). In the dengue fever mosquito, Aedes aegypti, JuanA makes up to 3% of the total genome, and it is estimated that there are 378 full length copies (Mouches et al, 1992; Biedler and Tu, 2007). With such a high copy number of both active and inactive elements, it would have had an impact on the Ae. aegypti genome, such as generating mutations and causing unequal crossing over.

My work focuses on characterising the impact of introducing an active element into different naive genomes. A few of the questions that I am exploring are the activity of the newly introduced element, the effect on the host genome and the survival of the host in general. In addition to using JuanA, I am also working on closely related elements for comparison. JuanC is 66% identical at the nucleotide level with JuanA and is highly abundant in Culex pipiens sensu lato (Agarwal et al, 1993). Culex pipiens 1 is a low copy number element but is related to the Juan elements and is present in the Culex mosquitoes

(Crainey et al, 2005).

(15)

The role of adipose-derived stem cell aging in autoimmunity

Regina Brunauer, Glenn Geesman, Sirkka Liisa Hostikka, Michelle Atallah, Victoria Lunyak*

Buck Institute for Research on Aging, Novato, CA, USA

Multipotent mesenchymal stromal cells (MSC) are a ubiquitous and versatile type of adult stem cells. They are derived from a variety of tissues such as bone, bone marrow, fat, cartilage and cord blood, and were reported to regenerate mesodermal tissues such as bone, cartilage and muscle. Moreover, they secrete trophic factors supportive for wound healing, control hematopoiesis, and are able to influence the immune response. These properties rendered MSC ideal candidates for stem cell therapies, since allogenic MSC transplants are not rejected, and for treatments of afflictions in which the immune system is involved, such as graft-versus-host disease and autoimmune diseases. These properties render MSC a great tool for cell therapies, but also raise the question whether MSC aging is the reason for the increased incidence of autoimmune diseases in the elderly.

We found that aged adipose-derived stem cells (ADSC), a subtype of MSC, have increased levels of Alu and L1 retrotransposon transcripts, which might be related to either increased transcriptional activity or lack processing by piwi or DICER complexes (1). Retroelements can induce an aberrant antiviral response, i.e. autoimmunity (2). We have demonstrated that upon accumulation of retrotranscripts in senescent ADSC, cells become prompted to secrete proinflammatory cytokines such as IL6 and IL8. We are working out the hypothesis that retroelements are responsible for the pro-inflammatory phenotype of ADSC, focusing on unraveling molecular links to known pathways triggering autoimmunity. Given the special role of MSC in the regulation of the immune response, aging of MSC could have an important role in the development of autoimmune diseases.

References:

1. Wang J et al, Cell Cycle 10:17, 1-15, 2011
2. Stetson DB et al, Cell 134, 587-598, 2008

(16)

High-throughput DNA methylation analysis of repetitive elements by hairpin bisulfite sequencing

Julia Arand (1), David Spieler (2), Tommy Karius (1), Daniela Meilinger (3), Alexander Meissner (4), Thomas Jenuwein (5), Guoliang Xu (6), Heinrich Leonhardt (3), Verena Wolf (2) and Jörn Walter (1)*

1 University of Saarland, Biological Sciences, Genetics / Epigenetics, Saarbrücken, DE
2 University of Saarland, Computer Science Department, Saarbrücken, DE
3 LMU München, Biozentrum Martinsried, Dept. Bio II., Martinsried, DE
4 Broad Institute, 7 Divinity Avenue, Rm 207, Cambridge, MA, 02138, USA
5 MPI for Immunology and Epigenetics, Stübeweg 51, 79108 Freiburg, DE
6 Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, CN

Repetitive elements comprise a large portion of the mammalian genome and they are major targets for DNA methylation mediated epigenetic silencing. DNA methylation in mammals is predominantly located at CpG positions and mediated by three different DNA methyltransferases (Dnmts). 80% of these methylated CpG positions are lying at sequences of repetitive elements. Due to high rate of deaminations at methylated cytosines the CpG to TpG mutation rate at repetitive elements is rather high.

In our analysis we monitored the methylation of CpGs at IAP-LTR, L1-5'UTR, B1 elements and major satellites on both DNA strands simultaneously using the hairpin bisulfite method. With high throughput 454 pyrosequencing we analysed more than 100.000 complementary CpGs of repetitive elements at random individual chromosomal loci of different pluripotent cell lines and somatic cells.

Our study comprises quantitative and qualitative novel insights into the distribution of symmetrical CpG methylation across repetitive elements. Excluding all mutated CpG positions, which appear at single strand bisulfite sequencing as unmethylated CpG position, we could revise the global methylation level for B1 elements from former single strand bisulfite data, which show only minor methylation, to a general high methylation rate of 60 to 85% in all analysed cells. Moreover, we show differences in the abundance of hemimethylated CpG positions - element and cell type specific. Furthermore, by analysing ESCs with KOs of single Dnmts followed by Hidden Markov Modelling, we show different efficiencies of the Dnmts at different repetitive elements. Finally, we found next to CpG methylation CpA methylation only in pluripotent cells and only at specific CpA positions at major satellites, not at other repetitive elements and not in differentiated cells.

(17)

Protein interactions with piALU RNA indicates putative participation of retroRNA in the cell cycle, DNA repair and chromatin assembly

Benjamin J. Blackwell (1), Mary F. Lopez (2), Jianrong Wang (4), Bryan Krastins (2), David Sarracino (2), James R. Tollervey (1), Marek Dobke (3), I. King Jordan (4,5) and Victoria V. Lunyak (1)*

1 Buck Institute for Research on Aging, Novato, CA, USA
2 BRIMS ThermoFisher, Boston, MA, USA
3 Division of Plastic Surgery, University of California San Diego, La Jolla, CA, USA
4 School of Biology, Georgia Institute of Technology, Atlanta, GA, USA
5 PanAmerican Bioinformatics Institute, Santa Marta, Magdalena, Columbia

Recent analyses suggest that transposable element (TE) derived transcripts are processed to yield various small RNA species that play critical roles in gene regulation and chromatin organization as well as genome stability and maintenance. A current study of transposable element-derived transcripts indicates that they are processed through the cellular RNAi machinery1 and the dis-regulation of their processing is associated with diseases such as cancer2, macular degeneration3, autoimmunity4,5 and human adult stem cell aging6. We hypothesize that the assembly of protein complexes onto retrotransposal RNAs might be nucleotide-specific, implying that RNA-binding proteins may undergo conformational and/or functional changes dependent on the RNA length, sequence, the presence of ssRNA, dsRNA, and DNA-RNA hybrids, as well as the frame of the processed intermediates. Using an unbiased RNA affinity assay coupled with mass spectrometry, we provide evidence for the composition of nuclear molecular complexes assembled on a piRNA homologous sequence derived from Alu RNA transcripts. Our data implicate a number of pathways and molecular processes through which processed intermediates of Alu RNA may participate in a multitude of nuclear processes within human adult stem cells. Our data point to potential roles for piALU RNAs in DNA repair, cell cycle, and chromatin regulations.

References

1. Kuramochi-Miyagawa S et al. MVH in piRNA processing and gene silencing of retrotransposons. Genes Dev 2010.
2. Lin C et al. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell 2009.
3. Kaneko H et al. DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. Nature 2011.
4. Bhoj VG, Chen ZJ. Linking retroelements to autoimmunity. Cell 2008.
5. Stetson DB et al. Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell 2008.
6. Wang J et al. Inhibition of activated pericentromeric SINE/Alu repeat transcription in senescent human adult stem cells reinstates self-renewal. Cell Cycle 2011.

(18)

Transposable element analysis within the Oryza Genome Evolution project

Dario Copetti (1), Angelina Angelova (1), Olivier Panaud (2) and Rod A. Wing (1)*

1 Arizona Genomics Institute and School of Plant Sciences, Univ. of Arizona, Tucson, AZ, USA
2 Laboratoire Génome et Développement des Plantes, UMR CNRS-IRD-UPVD, Université de Perpignan, Perpignan, France

The genus Oryza spans 24 species, 10 genome types, diploid and tetraploid genomes. Given the average small genome size and its wide distribution and importance, rice represents an ideal framework to develop accurate comparative genomics tools to study both fundamental and applied aspects of rice biology. The Oryza Genome Evolution project (OGE) follows the Oryza Map Alignment Project (OMAP) developing and analyzing a wider sample of rice species and genome types, as well as providing new sets of data to unravel broader aspects of Oryza evolution such as structural variation, phylogenomics, population genomics, new gene identification, and the role of Transposable Elements (TEs) in genome evolution.

For this project are being generated both whole-genome and chromosome arm-specific sequences, and an accurate TE annotation is required to assist the analysis of the other molecular components. A plant-centered TE library has been created, and a baseline TE annotation is available for all the genomes. De novo repeats detection strategies are being deployed to uncover new or divergent TE sequences in not yet annotated species, also by means of semi-automated methods. Comparisons between species will detect orthologous elements and describe events that shape genomes like proliferation and removal, gene capture and amplification, repeat-mediated recombination, and TE domestication. From the resequencing of large populations, the intra-specific insertion polymorphisms detected will be useful for the exploitation of natural variation in applied aspects such as breeding programs, assigning gene function, and characterization of ecotypes.

The development of this platform is so far new in plants, and the potential outcome relapses out of the analysis of the genus itself, establishing a new dimension for comparative genome studies. The analysis of TE role in genome evolution will benefit of a new and wider point of view.

(19)

Examining the role of Alu retrotransposition in neurological complexity

Jennifer A. Erwin and Fred H. Gage*

Salk Institute for Biological Sciences, CA, USA

The central nervous system is one of the most complex systems within the body, with approximately 1015 synapses and a diverse set of cell types. The remarkable complexity of the CNS suggests a number of possible combinatorial mechanisms to generate this complexity. Recent work has revealed that retrotransposition may contribute to neuronal complexity, diversity and even pathology[1, 2]. Long interspersed nucleotide element-1 (L1) actively retrotransposes as rat and human neurons develop[3, 4]. In a model for Rett syndrome, an autism spectrum disorder, excess L1 retrotransposition can be observed during neurogenesis[5]. Upon insertion, these mobile elements have the potential to alter gene expression or inactivate genes[6]. Many questions remain regarding the regulation of retrotransposition, the full characterization of other mobile elements and the functional significance of retrotransposition during neurogenesis. Using a human embryonic stem cell model of neural development, we find a developmentally specific time when endogenous L1 proteins support retrotransposition of an engineered Alu retrotransposition reporter. Future efforts will focus on characterizing endogenous Alu retrotransposition events in the human brain and exploring of the functional significance of somatic retrotransposition in the brain.

References:

1. Singer, T., et al., LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes?, in Trends in neurosciences. 2010.
2. Baillie, J.K., et al., Somatic retrotransposition alters the genetic landscape of the human brain, in Nature. 2011. p. 1-4.
3. Muotri, A.R., et al., Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition, in Nature. 2005. p. 903-10.
4. Coufal, N.G., et al., L1 retrotransposition in human neural progenitor cells, in Nature. 2009. p. 1127-31.
5. Muotri, A.R., et al., L1 retrotransposition in neurons is modulated by MeCP2, in Nature. 2010. p. 443-446.
6. Ponicsan, S.L., J.F. Kugel, and J.A. Goodrich, Genomic gems: SINE RNAs regulate mRNA production, in Curr Opin Genet Dev. 2010. p. 149-55.

(20)

SINE/Alu derived Piwi-interacting RNA as a tool for pluripotent stem cell generation

James Tollervey (1), Glenn Geesman (1), Cristina Vicente Garcia (2) and Victoria Lunyak (1)*

1 Buck Institute for Research on Aging, Novato, California
2 Centro Nacional De Biotecnologia, Madrid, Spain

Transcriptional activity from retrotransposons has been linked to a wide variety of cellular pathways, including the nuclear organization of chromatin, heterochromatinization of cetromeres and telomeres, X-chromosome inactivation, and genomic rearrangements. Recent evidence suggests that cytotoxicity caused by SINE/Alu retrotransposal RNA accumulation may be an important factor triggering macular degeneration and adult stem cell aging. DNA damage accumulation is biased towards the retrotransposal portion of the human genome, resulting in alterations of transcriptional and signaling pathways, which impact on tissue and organ deterioration in aging, ultimately causing cellular senescence.

Recent work in our lab has found that accumulation of full-length SINE/Alu transcripts drives establishment of an ex vivo senescence phenotype in human adipose-derived stem cells (hADSCs). Increased transcriptional activity from genomic areas bearing retrotransposons interferes with the recruitment of cohesin and condensin complexes, and impedes efficient DNA repair, thus triggering persistent DNA damage responses. Suppression of SINE/Alu transcription through the stable expression of its processed form, recently identified as a mature Piwi-interacting RNA (piRNA), results in the reversal of the senescence phenotype ex vivo. Surprisingly, prior senescent cells revert to a sustainable self-renewing phenotype and exhibit a dramatic upregulation of master pluripotency factors such as Oct4 and Nanog. Alu-derived piRNA expressing hADSCs are capable of embryiod body formation and robust multi-lineage transdifferentiation.

Currently we are exploring the possibility that, similar to the overexpression of master pluripotency factors, the use of Alu-derived piRNA expression can drive transformation of somatic cell lines to pluripotent stem cell like states (iPSCs).

(21)

A modular framework to characterize sequence insertions using paired-end whole genome sequencing data.

Tracy Ballinger, Adam Ewing, Jing Zhu, David Haussler*

Jack Baskin School of Engineering, UC Santa Cruz, Center for Biomolecular Science and Engineering, UC Santa Cruz, Howard Hughes Medical Institute

There are a number of programs available to detect retrotransposon insertions from whole genome high throughput sequencing using paired-end reads, and other programs assemble novel sequence from these data. However, we currently lack an efficient framework to not only detect breakpoints of structural variants, but also to characterize novel insertions of sequence not present in the reference genome assembly. This type of analysis is especially important in disease studies, such as TCGA, where viral sequence and gene duplication can play a role in disease progression. We have built a modular pipeline capable of finding various types of insertions, including novel sequences, retrotransposons, and pseudogenes. This framework can also extend to include applications for detecting deletions, inversions, and translocations. Our method is more efficient than what is currently available because it minimizes any remapping of reads, and because the entire .bam file is only read once. By initially segregating the data set into unmapped reads, discordant read pairs, and split reads, downstream analyses can run more efficiently because they only need to run on the subset of data that are useful. Our application can characterize not only the insertion site, but also the insertion sequence. For example, it can determine the length and most likely subfamily of a new retrotransposon insertion, and, in some cases, the most likely copy in the reference genome from which it originated. An efficient computational pipeline to detect all types of sequence insertions will be essential in large scale genome sequencing projects such as 1000 Genomes and The Cancer Genome Atlas.

(22)

Distinct genome-wide p53 binding profile in normal and cancer-derived human cells

Krassimira Botcheva (1), Sean R. McCorkle (1), W. Richard McCombie (2), John J. Dunn (1) and Carl W. Anderson (1)*

1 Biology Department, Brookhaven National Laboratory, Upton, NY 11973 USA
2 Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724 USA

Given the importance of site-specific binding to DNA for the p53-mediated transcriptional regulation and tumor suppression, extensive efforts have been directed toward experimental identification of p53 binding sites in the human genome. Notably, all de novo genome-wide binding studies reported up to date were done in cancer-derived cell lines. Using a ChIP-seq approach we mapped p53 binding sites for the first time in normal human cells (lung fibroblasts IMR90)1 and by comparing them with sites previously reported by genome-wide studies in cancer cell lines, we showed dramatic differences in their genomic distribution. p53 binding sites mapped in normal, unlike those in cancer cells, are enriched in the immediate vicinity of TSS and at CpG islands. Close proximity of binding sites to TSS has been typically seen in studies of individual p53 target genes, but has not been observed by genome-wide studies before. We found the binding sites in IMR90 are far less enriched at repeats compared to those reported in HCT116 and U2OS cancer cell lines. The observed differences in the p53 binding pattern do not seem to reflect a distinct preference for specific sequences, since the de novo developed p53 motif based on the IMR90 study is similar to those reported previously. More likely the different chromatin landscapes in normal and cancer-derived cells modulate the availability of the binding sites, which is supported by the comparison between the IMR90 ChIP-seq binding data and the IMR90 methylome2 pointing at enrichment of binding sites in hypomethylated DNA.

References:

1. Botcheva K, McCorkle SR, McCombie WR, Dunn JJ and Anderson CW. Cell Cycle 2011; (in press)
2. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Nature 2009; 462:315-22.

(23)

L1 Retrotransposon-initiated fusion transcripts are regulated by tissue specificity and alternative splicing in normal and malignant cells

Rangwala, Sanjida H., Hammell, N., Symer, David E.

The Ohio State University Comprehensive Cancer Center

Over half of the human genome is composed of repetitive elements, nearly 20% from the L1 sub-family alone. When active, these elements can have dramatic effects on the functioning of the genome including, in extreme cases, inserting into and disrupting genes. While most L1 elements are quiescent, increasingly evidence suggests that a number of sites may produce RNA. I have isolated and characterized 1000 cDNA sequences corresponding to L1-initiated transcripts in a panel of human tissues, both normal and malignant (chronic lymphocytic leukemia B cells). Splicing from the L1 into flanking gene regions is common, sometimes resulting in an alternate open reading frame of that gene. Multiple transcript isoforms are often detected originating from the same L1 locus. While some L1-initiated transcripts are expressed in two or more tissues, different tissues also show unique expression patterns. In some cases, these expression patterns mirror those of the gene; however, in others, expression of the RIFT appears to be regulated distinct from that of the host genes. This study indicates an unexpected level of complexity in the complement of transcripts originating from L1 elements, suggesting that regulation of these transcripts involves interplay between both locus- and tissue-specific factors.

(24)

Hermes transposase structure and function.

Joshua Knapp (1), Nancy L. Craig (2) and Peter W. Atkinson (1,3)*

1 Department of Biochemistry & Molecular Biology, University of California, Riverside, CA, USA
2 Department of Genetics and Molecular Biology, Howard Hughes Medical Institute, Johns Hopkins School of Medicine, Baltimore, USA
3 Institute for Integrative Genome Biology and Department of Entomology, University of California, Riverside, CA, USA

The Hermes transposon from the housefly, Musca domestica, is a member of the hAT superfamily of eukaryote class II transposons and is active in a number of organisms. Transposition of Hermes results in an 8 bp target site duplication the sequence of which is characteristic of members of the Ac subfamily of transposons within the hAT superfamily. We are interested in locating and defining those parts of the transposase that may interact with the target site DNA. Based on the published crystal structure of the Hermes transposase, we focused our attention on a positively charged trench located on the “rim” of the Hermes oligomer that is in close proximity to the active site of the transposase. This trench is defined by two alpha helices from adjacent monomers in which reside multiple lysine and arginine residues. We report on the in vitro and in vivo behavior of Hermes transpoases containing mutations in these residues.

(25)

Identification of active Tc1/mariner elements and their associated Stowaway MITEs from the mosquito Aedes aegypti

Kun Liu and Susan R. Wessler*

Department of Botany & Plant Sciences, UC Riverside

Miniature inverted repeat transposable elements (MITEs) are an abundant component of the genomes of plants and animal species where they are preferentially associated with host genes. MITEs comprise over 17% of the genome of the mosquito Aedes aegypti, which is a vector of human disease including dengue and yellow fever. To understand why MITEs are so successful in this species, we took advantage of the availability of a complete genome sequence to identify potentially active autonomous elements and related MITEs and test for activity in yeast. We focused on the Tc1/mariner superfamily whose members are known to mobilize Stowaway-like MITEs in plants. Tc1/mariner elements are abundant in A. aegypti with ~2000 coding elements and ~60 families of Stowaway-like MITEs containing 50 – 3000 copies each. Of these, we mined two candidate active coding elements as well as their associated MITE families. The transposases from both elements were expressed in yeast where they successfully catalyzeed the excision from a reporter construct of the associated MITEs in addition to an artificial nonautonomous element. Features of transposition in yeast include the production of transposon footprints upon excision and reinsertion in the yeast genome at TA dinucleotides. Finally, experiments are underway to identify active candidates from several of the 17 class 2 DNA superfamilies in the A.aegypti genome and assay for transposition in yeast.

(26)

Transposition ex vitro – in vivo

Maryia Trubitsyna, Julia M Richardson, David J Finnegan*

The University of Edinburgh

Class II Transposable Elements (TEs) are discrete DNA sequences that excise from one place in the genome and insert into another using a specific enzyme – transposase. This ability gives TEs a potential to be used in transgenic and gene therapy approaches [1]. In order to understand the mechanisms of transposition at the molecular level we are studying the functional and structural properties of Mos1 and Mboumar-9 transposases from Drosophila mauritiana [2] and from Messor bouviery [3], respectively. Based on these studies we have generated artificial transposons, which display superior transposition efficiency. Moreover, we have established a novel method for the in vivo transposition of both transposons into the bacterial genomic DNA, which leads to the random insertions of the desired sequences. We demonstrate that the insertion loci can be precisely mapped by the bidirectional sequencing. Our method can be widely used for plasmid-free integration of transgenes as well as creating libraries of the random knock-outs. Moreover, our preliminary data shows that this method could be used in other systems, such as yeast (S.pombe).

References:

1. Delauriere, L., et al., Mariner transposons as genetic tools in vertebrate cells. Genetica, 2009. 137(1): p. 9-17.
2. Richardson, J.M., et al., Mechanism of Mos1 transposition: insights from structural analysis. Embo Journal, 2006. 25(6): p. 1324-1334.
3. Munoz-Lopez, M., et al., Transposition of Mboumar-9: Identification of a new naturally active mariner-family transposon. Journal of Molecular Biology, 2008. 382(3): p. 567-572.

(27)

Transposable elements as the structural component of meiotic chromosomes in human and mouse

Spangenberg, V., Dadashev, S., Kolomiets, O., Bogdanov, Yu.*

Vavilov Institute of General Genetics of the Russian Academy of Sciences

Synaptonemal complexes (SCs) are the intranuclear structures which are assembled between homologous chromosomes in the course of meiotic prophase I and facilitate chromosome synapsis and the process of crossing over. The idea of current investigation is to detect the distribution of three types of repeated DNA sequences in the meiotic chromosome structure. –°ombined FISH and immunochemical analysis were performed on human spermatocytes microspreads. Using FISH we had visualized and examined the arrangement of repeated DNA sequences in relation to pachytene SCs stained with fluorescent SYCP3 antibodies. It was found that probes containing Alu and L1 repeats formed “cases” (multiple signals) around SCs, while simple repeat (GT)22 signals were spread over the chromatin. Alu- and L1-repeats were mainly located nearby SCs, however intensity and width of these regions were non-homogeneous along the chromosomes. This possibly can be explained by different chromatin compaction and its genetic characteristics. In our previous investigation on mouse spermatocytes we had found that the B1(Alu)-repeats were colocalizated with the SCs in the similar way as in human. We assume that these DNA repeats are very close to the “anchoring” segments of DNA in both mammalian species and may be involved in the attaching chromatin to SCs.

The work was supported by RBRF grant 10-04-00666-a.

(28)

A new set of markers for human identification based on 32 polymorphic Alu insertions

I.Z. Mamedov, A.A Kurnosov, S.V. Ustyugova, D.A. Shagin, Y.B. Lebedev*

Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Russia, Moscow, Miklukho-Maklaya 16/10, 117997; Imamedov@mx.ibch.ru

A number of genetic systems for human genetic identification based on STRs or SNPs are widely used for crime detection, kinship studies and in analysis of victims of mass disasters. We have developed a new set of 32 molecular genetic markers for human genetic identification based on polymorphic retroelement insertions. Allele frequencies were determined in a group of about 100 unrelated individuals from four genetically distant populations of the Russian Federation. The mean match probability and probability of paternal exclusion, calculated based on population data, were 5.53x10 14 and 99.784% respectively. The developed system is cheap and easy to use as compared to all previously published methods. The application of fluorescence based methods for allele discrimination allows to use the human genetic identification set in automatic and high-throughput formats.

(29)

Repeat associated small RNAs vary among parents and following hybridization in maize

Wesley T. Barber (1), Wei Zhang (1), Hlaing Win (1), Kranthi K. Varala (1), Jane E. Dorweiler (2), Matthew E. Hudson (1), Stephen P. Moose (1)*

1 Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, USA
2 Department of Biological Sciences, Marquette University, Milwaukee, WI, USA

Maize is characterized by high degrees of genomic complexity, genetic diversity, and hybrid vigor. Differences in meristem activity are believed to contribute to the enhanced growth displayed following hybridization. We used Illumina deep sequencing to assess how small RNA (sRNA) populations within the seedling shoot apex and the developing ear vary between two inbred lines (B73, Mo17) and their hybrid. We found that B73 and Mo17 differ in their populations of distinct 21, 22, and 24-nt sRNAs, which primarily originate from repeat regions. For example, the parents differ in their production of 21-nt and 22-nt small interfering RNAs (siRNAs) from specific retrotransposon families, suggesting that the retrotransposon portion of the maize genome may serve as source for genetic variation in post-transcriptional regulation. Hybridization passes on these parental differences, resulting in an individual that differs from both parents in its complexity of sRNAs. Larger deviations from mid-parent sRNA abundance levels were observed in the ear than the shoot apex. The 24-nt siRNAs that differed greatly between the parents tended to accumulate at levels below the mid-parent in the hybrid. Loss of the RNA-dependent RNA polymerase (RDR2) encoded by the modifier of paramutation 1 (MOP1) locus does not suppress hybrid vigor for B73xMo17, which questions the importance of MOP1 dependent changes in 24-nt siRNAs to the hybrid vigor displayed by this cross.

(30)

Combinatorial interaction of Transcription Factors is mediated by Transposable Elements

Testori, Alessandro; Corà, Davide; Caizzi, Livia; Cutrupi, Santina; Friard, Olivier; De Bortoli, Michele; Caselle, Michele*

Institute for Cancer Research and Treatment (IRCC), Italy

Transcription Factors (TFs) exert their regulatory function often in a combinatorial way. An interesting problem is to understand the evolutionary mechanisms which led to clustering of binding sequences of the "right" TFs. While it is easy to create by point mutations a single binding sequence, the combination of several binding motifs in an extended region of DNA is much less easy to achieve in this way. A possible solution is to assume that a suitable template for the sought combination of motifs already exists in the genome: Transposable Elements (TEs in the following) are the natural candidates to play this role. We shall concentrate on the network of cooperative interactions centered on Estrogen Receptor alpha (ERα).

We discuss in particular two sets of Chip-Seq data obtained with and without estrogen stimulation respectively and show that ERα targets, as derived from the Chip-Seq data, are preferentially located within particular classes of TEs. We also show that these TEs contain the binding sequences of a few other TFs which are thus expected to preferentially interact with ERα on a genome wide scale. This turns out to be true both in presence and in absence of estrogen stimulation but different levels of estrogen stimulation correspond to different classes of TEs and thus to different combinations of cooperating TFs. In particular, in the case of the estrogen stimulated data we find a strong enrichment of targets within MIR (Mammalian Interspersed Repeats) Transposable Elements.

We conjecture that the special affinity of ERα for the MIR class of TEs could be at the origin of the important role which ERα assumed in mammalians. Besides MIRs, a considerable portion of ERα targets occurs within younger ERV-like TEs. These targets suggest that a relevant fraction of the ERα regulatory network underwent a massive rewiring in recent times.

(31)

Characterization of a novel smallest LTR retrotransposon in grass

Dongying Gao (1), Mingsheng Chen (2), Blake C. Meyers (3) and Scott Jackson (1)*

1 Center for Applied Genetic Technologies and Institute for Plant Breeding Genetics and Genomics, University of Georgia, Athens, GA, USA
2 State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, CN
3 Department of Plant & Soil Sciences, and Delaware Biotechnology Institute, University of Delaware, Newark, Delaware, USA

LTR retrotransposons are the most abundant fractions in plants. They are large elements (> 4 kb) and usually reside in heterochromatic regions. Here we present a novel retrotransposon named FRetro129 which is only 292 bp and represents the smallest LTR retrotransposon identified so far. The element is found in rice, maize, sorghum and other genomes of grass and suggested that it was present in the ancestor of the grass and originated at least 50-80 MYA. However, this element may still be active in some genomes of grass based on the sequence comparison and mRNA data. Unlike other LTR retrotransposons, the small retrotransposons (SMARTs) are distributed throughout the genomes and are mainly located within or near genes. Insertions of SMARTs not only alter gene structures but only affect the gene expression. Interestingly, qRT-PCR results indicated that an intergenic insertion of SMART increased 4 and 12 times of gene expression for both flanking sides. Additionally, the SMART-specific small RNAs (sRNAs) also identified which are likely involved in the gene regulation. Overall, SMARTs have played an important role in grass genome evolution and genic innovation and also provide a valuable resource for grass gene tagging systems.

(32)

Revealing somatic LINE-1 retrotransposition in the brain

Tatjana Singer, Iryna Gallina, Fred H. Gage*

Salk Institute for Biological Studies, 10010 North Torrey Pines Rd, La Jolla, CA 92037

Recent evidence from our laboratory and others suggests that within an individual, LINE-1 retrotransposons can actively 'jump' during neuronal differentiation and, upon reinsertion, create somatic copy number variations in neuronal genomes. Thus, events related to LINE-1 retrotransposition may have an impact on the transcriptome and diversify neuronal genotypes and phenotypes. To uncover novel LINE-1 insertion sites we have developed a targeted LINE-1 sequencing approach amenable for high-throughput sequencing on the Illumina HiSeq2000 platform. A variety of library preparation protocols and validation procedures has been developed and will be discussed. As a benchmark to validate our methodologies and assess somatic LINE-1 mobility against a fully sequenced human genome we utilize the HuRef genome (C. Venter DNA) as a 'gold standard'.

(33)

Reconstructing evolutionary dynamics of retrotransposable elements

T.E. Kijima and Hideki Innan*

The Graduate University for Advanced Studies, Japan

Transposable elements (TEs) are ubiquitous in prokaryote and eukaryote genomes. Although the number of TEs in a genome varies between species and even between individuals in a single species, there is no doubt that TEs are one of the major components in the genomes of many organisms. The theoretical model for TEs was made in 1980's. Then models were further extended to explore the potential roles of other factors, such as migration and breeding system etc. However these models focus only on the copy number dynamics and didn't consider the sequences of TEs. We here extend the model of with a few modifications. First, the genome is assigned to functionally important regions and junk regions. It is assumed that insertion of TEs in the former regions is deleterious, but not in the latter. Second, we use a more realistic function of the relationship between the transposition rate and point mutations according to a recent data of Alu elements. Then, we simultaneously investigate the evolutionary behavior of the copy number of TEs and their sequence evolution.

(34)

Recombination drives vertebrate genome contraction

Kiwoong Nam, Hans Ellegren*

Uppsala University, Sweden

Selective and/or neutral processes may govern variation in DNA content and, ultimately, genome size. The observation in several organisms of a negative correlation between recombination rate and intron size has been given a selectionist interpretation based on Hill-Robertson interference and the higher efficiency of selection in regions of high recombination. However, the observation could also be compatible with a neutral model in which recombination is mutagenic for length changes throughout the genome. We used whole-genome data on small insertions and deletions within transposable elements from chicken and zebra finch, to demonstrate clear links between recombination rate and a number of attributes of reduced DNA content. Recombination rate was negatively correlated with the length of introns, transposable elements and intergenic spacer, and the rate of short insertions. Moreover, it was positively correlated with gene density, the rate of short deletions, the deletion bias, and the net change in sequence length. All these observations point at a pattern of more condensed genome structure in regions of high recombination. Based on the observed rates of small insertions and deletions, we estimate that the genome of the most recent common ancestor of birds and lizards have lost 17% of its DNA content up till present. As the activity of transposable elements has been low in the avian lineage, the deletion bias is likely to have had a significant effect on genome size evolution in dinosaurs and birds, contributing to the maintenance of a small genome. We also demonstrate that most of the observed correlations between recombination rate and genome contraction parameters are seen using data from the human genome. Recombination may thus be a general force to drive vertebrate genome size evolution.

(35)

REPET v2 : TEs detection and classification improvements

Sandie Arnoux, Véronique Jamilloux, Timothée Chaumier, Claire Hoede, Olivier Inizan, Hadi Quesneville*

Unité de Recherches en Génomique-Info (URGI), INRA, Centre de recherche de Versailles, Route de Saint Cyr, 78026 Versailles, FR

The REPET package (Flutre et al, 2011) integrates two pipelines, which are constantly improving: TEdenovo and TEannot for transposable elements (TEs) detection and annotation. In the last release TEdenovo changed significantly.

The TEdenovo pipeline strategy is to find as much as possible potential TEs, and then to classify putative TEs in order to filter out false positives. The pipeline starts by the detection of repeated sequences comparing by alignments the genome with itself. These alignments are independently clustered according to different tools (RECON, GROUPER, PILER). Then, it builds multiple alignments from the clusters, from which a consensus sequence is derived. These consensus are classified according to TE features and redundancy is removed. Finally, there is the possibility to remove false-positives according to the classification (SSR, host genes, rDNA and under-represented unclassified consensus).

Two steps have been improved in REPET v2:

1) A structural TE detection approach is now implemented : LTRharvest (Ellinghaus et al, 2008) is used to search for LTR retrotransposons, using structural features of this TE category. It allows catching TEs present in only one or two copies in the genome. Potential TEs thus detected and all other derived consensus are put together before the classification and redundancy removal step.
2) Classification benefits from improvement too with the integration of PASTEC, a new classifier that we have developed. It tests all TE classifications and each result is weighted according to the evidences found. In addition to similarities to known TEs in Repbase Update and the search for repeated structures, it also uses HMM profiles, which are interesting to classify TEs and to detect host genes. PASTEC gives precisions about completeness and indicates if TEs are potential chimeras.

For illustrations of these two pipelines and associated tools, in the frame of genome annotation, see Véronique Jamilloux's poster.

(36)

An iterative process for TEs annotation in large genomes

Véronique Jamilloux, Sandie Arnoux, Olivier Inizan, Hadi Quesneville*

URGI – Research unit in Genomics-Info, INRA, Centre de recherche de Versailles, Route de Saint Cyr, 78026 Versailles, FR

The recent successes of new sequencing technologies allow today the sequencing of very large genomes at reasonable costs. Transposable elements (TEs) constitute the most structurally dynamic components and the largest portion of their nuclear genomes, e.g. 85% of the maize genome (Schnable et al. 2009), and 88% of the wheat genome (Choulet et al. 2010). Therefore, TE annotation should be considered as a major task in genome projects, but currently not be obtained automatiquely. This crucial step is now a bottleneck for many genome analyses.

In this context, we improve, in the v2.0 release—see poster “REPET v2” (Arnoux et al. 2012) in this conference—the REPET package (Flutre et al. 2011). REPET, gathers two pipelines: TEdenovo build a TEs library and TEannot annotate TE copies in the genome. Thus, we test a new strategy dedicated to annotate very large genomes, the iterative approach:

1) Detection of young TEs with stringent parameters able to find quickly only the less degenerate ones to build a first TE library.
2) TE annotation and extraction of the corresponding sequences from the initial contigs. We obtain a reduced genome sequence.
3) Detection of the other TEs with sensitive parameters on the reduced genome sequence to build a second TE library.
4) Annotation of the original contigs with the concatenation of the two TE libraries.

The rational is that these large genomes are made of mostly few TE families that recently invade. They will be detected in the first step and this will allow reducing the genome by an important factor. We tested this approach on A.thaliana and we will present the benchmarks obtained. We will also present preliminary results on the 3B chromosome of wheat (293890 contigs, 975Mbp), the first fully sequenced chromosome of an ~17Gbp allohexaploid genome.

(37)

Locus- and domain-dependent control of DNA methylation at mouse SINEs during male germ cell development.

Kenji Ichiyanagi, Yungfeng Li, Tomoko Ichiyanagi, Keigo Hirafuku, and Hiroyuki Sasaki*

Division of Epigenomics, Medical Institute of Bioregulation, Kyushu University, Japan

In mammals, germ cells undergo striking dynamic changes in DNA methylation during their development. However, the dynamics and mode of methylation are poorly understood for SINEs dispersed throughout the genome. We investigated the DNA methylation status of mouse B1 and B2 SINEs in male germ cells at different developmental stages. These SINE elements showed a large locus-to-locus variation in methylation; loci close to Pol II promoters were hypomethylated, while most others were hypermethylated. Interestingly, a mutation that eliminates Piwi-interacting RNAs (piRNAs), which are involved in methylation of LINEs, did not affect the level of SINE methylation, implying a piRNA-independent mechanism. Methylation at B1 loci in SINE-poor genomic domains showed a higher dependency on the de novo DNA methyltransferase DNMT3A but not on DNMT3B suggesting that DNMT3A plays a major role in methylation of these domains. We also found that many genes specifically expressed in the testis possess SINE elements in their promoters, suggesting the involvement of SINE methylation in transcriptional regulation. Taken altogether, our results not only reveal the dynamics and mode of SINE methylation but also suggest how the DNA methylation profile is created in the germline by a pair of DNA methyltransferases.

(38)

Analysis of the expression of transposable elements by HT-seq

Bogdan Tanasa (1), Hong Sook Kim (2), Wenbo Li (2), Michael Rosenfeld (2,3)*

1 The Scripps Research Institute
2 Howard Hughes Medical Institute
3 University of California, San Diego

In the past decades, many studies have proposed the role of transposable and repetitive elements in transcription (many transposable elements contain binding sites for transcription factors that regulate promoters, enhancers, silencers, transcription attenuation or termination), splicing, translation, replication and recombination, nucleosome positioning and nuclear architecture, and in genome evolution. Several reports have shown that the repetitive elements are actively transcribed in human and mouse cells. Here we used high-throughput sequencing technologies (RNA-seq and global run-on assays, namely GRO-seq) to map the expression of repetitive elements in cancer cell lines in response to hormones.

(39)

Global annotation and molecular evolutionary analysis of genomic repeats in the Painted Turtle, Chrysemys picta.

Christopher Botka (1), Quan Fang (1) and Andrew M. Shedlock (2)*

1 Research Computing Group, Harvard Medical School
2 Department of Biology and Graduate Program in Marine Genomics, College of Charleston

The analysis of genome-scale information from the sister group of mammals, Reptilia, is presently closing a large gap in the study of vertebrate biology. Turtles are of central importance to this effort in that they have been difficult to place accurately in the amniote tree of life, they exhibit a number of unique phenotypic and life history traits among vertebrates, and are threatened globally by a variety of human impacts on the environment. In particular, annotating the abundance and age of different classes of mobile DNA repeats inferred from sequence alignment and DNA divergence values provides a basis for testing alternative models of molecular evolution which describe forces that have shaped our own genomes uniquely as compared to other amniote species.

We annotated global repeat content for the full set of pre-published genomic sequence reads generated for the painted turtle, C. picta, the first turtle species now having its complete genome assembled under NIH sponsorship. The Repeatmasker software program, Perl scripting and Excel spreadsheets were used to extract, sort, and chart comparative DNA sequence alignment statistics for specific genomic repeat classes.

Comparison of sequence divergence values based on original turtle query sequence aligned against a comprehensive reference database for vertebrate repeats reveals differential age distributions among major repeat subfamilies. A broad range of values was recovered for all types examined with lower values indicating the presence of relatively young active elements still driving turtle genomic diversity and distributions skewed toward higher mean values suggesting numerous distinct lineages of older inactive repeats that can be viewed as molecular fossils embedded in turtle chromosomes. The present study suggests that alternative modes of evolution are likely operating in different turtle repeat lineages whose proliferation and persistence cannot be easily explained by a strict vertical inheritance of single source genes, such as seen for the dominant form of interspersed repeats in the human genome and that contrasts with the largely inactive compact repetitive DNA landscape of birds.

(40)

Widespread incorporation of transposable element sequences in human genes

Ahsan Huda and Pierre R Bushel

National Institutes of Health, National Institute of Environmental Health Sciences (NIH/NIEHS)

Transposable elements have long been regarded as selfish or junk DNA having little or no role in the regulation or functioning of the human genome. However, over the past several years this view came to be challenged as several studies provided anecdotal as well as global evidence for the contribution of transposable elements to the regulatory and coding needs of human genes. In this study, we explored the incorporation and regulation of coding sequences donated by transposable elements using RNA-seq, Methyl-seq, ChIP-seq, CAGE, and DNase1 HS data in two human hematopoietic cell-lines characterized by the ENCODE project: GM12878 and K562. In each cell-line we found several thousand instances of transposable elements donating coding sequences to human genes. Using de novo transcriptome assembly of the RNA-seq data, we also report that the percentage of genes that incorporate transposable elements in their coding sequences is significantly greater than that obtained from the reference transcriptome assemblies such as Gencode and Refseq. We also analyzed the differential regulation and expression of these transcripts between the two cell types and evaluated their role in conferring cell type specific gene expression. Our data suggests that transposable elements have a significant role in the regulation and cell type specific expression of human gene transcripts in the two cell-lines.

(41)

The mysterious role of TnpB_ IS200/IS605 in transposition: distant eukaryotic homologs are captured in diverse transposable elements

Weidong Bao and Jerzy Jurka*

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

In addition to the transposases (TnpA), bacterial insertion sequences (IS) of IS200/IS605 and IS607 family frequently encode a functionally unknown protein, TnpB. Here we report two group of TnpB-like proteins (Fanzor1 and Fanzor2) encoded in various transposable elements in eukaryotes, as well as in some large double-stranded DNA (dsDNA) viruses infecting eukaryotes. Fanzor and TnpB proteins show the same signature in the C-terminal half region: D-X(125,275)-[TS]-[TS]-X-X-[C4 zinc finger]-X(5,50)-RD, but are highly variable in the N-terminal regions. Fanzor1 proteins are frequently captured in various types of DNA transposable elements: Helitron, Mariner, IS4-like, Sola2, MuDr, etc. In contrast, Fanzor2 proteins seem to be only associated with IS607-type element. The phylogeny and distribution of the TnpB/Fanzor protein indicate that viruses may play extensive roles in lateral gene transfer. Based on the parallelism between 3 Fanzor elements from viruses, we proposed that TnpB/Fanzor could be a unknown methyltransferase. In addition, two new phenomena were revealed in characterizing these diverse transposons, including one-ended Helitrons flanked by variable length of duplicated target site (TSD), a new Helitron subgroup (Helitron2), containing hair pin structure on both ends.

(42)

Dada – a new superfamily of DNA transposons interrupting small RNA genes

Kenji K. Kojima and Jerzy Jurka*

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

We report a new superfamily of DNA transposons from various eukaryotes including animals, fungi, plants and protists. We named this superfamily Dada, since we first identified them from Danio rerio (zebrafish) and Daphnia pulex (water flea). Dada transposases showed a weak similarity to DDE-transposases of other superfamilies, and the alignment revealed they share three catalytic residues and some other conserved residues indicating their clustering with the superfamilies P, MuDR, Rehavkus, hAT and Kolobok. One interesting observation is that they are frequently nested in small RNA genes such as U6 small nuclear RNA genes and tRNA genes. There is some resemblance between Dada and sequence-specific non-LTR retrotransposons.

(43)

The conserved piggyBac transposase fusion protein CSB-PGBD3 collaborates with AP-1 proteins to regulate nearby genes in primates

Lucas Gray, Tom Pavelitz, Arnold Bailey, Kim Fong, Alan Weiner*

University of Washington Department of Biochemistry

The CSB-PGBD3 fusion protein arose more than 43 Mya when the PGBD3 transposon integrated into intron 5 of the Cockayne syndrome group B gene (CSB) in the common ancestor of simian primates. The PGBD3 transposase ORF is flanked by a 3' splice acceptor site upstream and a polyadenylation site downstream, enabling it to be expressed as an alternative 3' terminal exon. As a result, the CSB genes of all higher primates, from marmoset to human, now generate three proteins by alternative splicing and polyadenylation: full length CSB protein, a CSB-PGBD3 fusion protein that fuses the N-terminus of CSB to a complete piggyBac transposase, and solitary piggyBac transposase driven by a cryptic promoter in CSB exon 5. CSB-PGBD3 is remarkably well conserved, and continues to be expressed in all simian primates from marmoset to human. PGBD3 also gave rise to a family of ~900 internally deleted MITEs known as MER85s, which are dispersed throughout the human genome. To determine what the CSB-PGBD3 fusion protein does in normal cells, why it has been conserved, and whether it affects the severity or nature of Cockayne syndrome, we have used genome-wide ChIP-seq and expression array analysis to determine the role of CSB-PGBD3 in transcription. We found that CSB-PGBD3 binds strongly to MER85 elements as expected, but also found surprising interactions of the CSB-PGBD3 fusion protein with sites bound by the AP-1 family transcription factors Jun and Fos, which are potent proto-oncogenes. These interactions correlate with transcriptional changes induced by CSB-PGBD3 expression, and suggest that CSB-PGBD3 may collaborate with AP-1 proteins in normal cells, or interfere with normal AP-1 function in Cockayne syndrome (CS). Our results have intriguing implications for CS, and provide a clear example of how insertion of transposable elements can generate useful new proteins that reshape gene expression and genomic evolution.

(44)

Trying to pin down the mechanism of APOBEC3s inhibition of retrotransposition

Nataša Lindič and Nika Lovšin*

Chair of Biochemistry, Department of Chemistry & Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Ljubljana, Slovenia

Endogenous retroviruses and retrotransposons are mobile genetic elements that can cause genomic instability and are considered to be one of the major driving forces in the evolution of eukaryotic genomes. APOBEC3 (A3) proteins are one of the host restriction factors inhibiting retrotransposition. Belonging to the functionally diverse APOBEC protein family, A3 proteins are cytidine deaminases that edit DNA molecules and have been shown to inhibit replication of human L1 and Alu non-LTR retrotransposons. In this study, we investigated the effect of A3 proteins on the retrotransposition of an evolutionary distant relative of L1, fish non-LTR retrotransposon L2. L2 retroelements are widespread and active among vertebrates but have been inactivated in placental mammals and constitute only 2% of the human genome. We show that in ex vivo assay, retrotransposition of eel and zebrafish L2 retrotransposons is most significantly inhibited by human A3A and A3B protein, and to a similar extent as that observed with L1 retrotransposon. Mutational analysis of human A3A and A3B shows that the inhibition of L2 retroelement is partially independent of their enzymatic activity, while the intact nucleic acid binding domain is required for the inhibition. Furthermore, sequencing analysis of the newly integrated L2 elements reveals no DNA editing. Thus, our findings indicate that L2 and L1 retrotransposons share replicative steps and that differences between the two elements do not affect their restriction by A3 proteins. Moreover, we demonstrate that the only APOBECs present in zebrafish, AID, A2a and A2b, have no impact on the retrotransposition of zebrafish L2 retrotransposons. Given that vertical inactivation of L2 retrotransposons coincides with the evolution of mammalian A3 proteins, it is possible that A3 proteins could have silenced L2 elements in mammalian genomes.

(45)

Waves of repeat driven CTCF binding expansions have shaped mammalian genomes

Petra C. Schwalie (1), Dominic Schmidt* (2,3), Michael D. Wilson (2,3), Benoit Ballester (1), Angela Goncalves (1), Claudia Kutter (2,3), Gordon D. Brown (2,3), Aileen Marshall (2,5), Duncan T. Odom (2,3,4) and Paul Flicek (1,4)*

1 European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
2 Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK
3 Department of Oncology, Hutchison/MRC Research Centre, Hills Road, Cambridge, UK
4 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK
5 Cambridge Hepatobiliary Service, Addenbrooke's Hospital, Hills Road, Cambridge, UK

CTCF is a ubiquitous zinc-finger protein that plays essential roles in domain insulation, nuclear architecture and transcription regulation within vertebrates. We experimentally determined the genome-wide occupancy of CTCF in livers of six mammalian species by ChIP-sequencing, finding (1) remarkable conservation of CTCF binding, (2) a larger sequence specificity of CTCF-DNA contacts, and (3) evidence for repeat-mediated expansion of CTCF binding in multiple lineages.

As opposed to canonical transcription factors that show rapid turnover of their genomic binding, we detect that up to 60% of CTCF bound regions are shared along lineages, while as much as 10% are common to all mammalian species studied. Shared regions are evolutionary conserved at the sequence level, and show higher ChIP enrichment, as well stability (as assessed by CTCF RNAi). A subset of binding events contain a second sequence motif, located at a preferred and conserved spacing with respect to the canonical one; this suggests that CTCF contacts the DNA over a larger region than previously described in genome-wide studies. The binding frequency of motif sequences (words) is in general conserved across species, with a few highly bound words corresponding to deeply shared regions of high ChIP enrichment. However, certain motif instances are restrained to only one/two organisms and represent motifs embedded in lineage, respectively species-specific repeat elements. Functionally, we detect no difference between deeply-shared, repeat-associated and regular CTCF binding sites.

Our data suggests a model by which intermitted waves of CTCF motif expansions mediated by repeat elements create a pool of putative bound elements that can take up roles in chromatin structure and transcriptional regulation. As selection acts on the newly created sites, some will be maintained and eventually detected as shared events along a tree branch.

(46)

Epigenetics speaks up for the silent DNA

Mireia Jordà, Sergi Lois, Anna Díez, Berta Martín, Víctor Barrera, Miguel A. Peinado*

Institute of Predictive and Personalized Medicine of Cancer (IMPPC), Badalona, Spain

DNA methylation is a key component of the epigenome architecture that plays a crucial role in numerous physiological and pathological processes. Most genome-wide studies analyzing this epigenetic mark have focused the analysis to high density CpG islands and gene promoters, leaving out most of the genome, an especially repeat elements. A haploid human genome contains more than one million Alu repeats, but despite their overwhelming presence only a small fraction is likely to have a specific functional role. Most of them are methylated and silenced, nevertheless, a fraction of Alus remains unmethylated in normal cells and this proportion is significantly increased in cancer cells. The grounds for this "atypical" epigenetic state of unmethylated Alus as their functional implications are unknown.

Although different approaches have been used to make bulk estimates of DNA methylation in repeat elements, there is still a lack of screening strategies that specifically allow a feasible identification of methylation in repetitive sequences. We have developed a new approach in combination with next-generation sequencing that specifically targets unmethylated Alus and allows the generation of a comprehensive maps at the genome scale. Using this approach has been applied to detect unmehylated Alus in normal tissues, primary tumor cells, colon cancer cell lines and embryonic stem cells. Data reveal the existence of distinctive profiles of Alus with differential methylation associated with cell differentiation and tumorigenesis as well as a subset of Alus whose default state is unmethylation. Furthermore, these unmethylated Alus present specific features, including high evolutional conservation and association with active chromatin domains. Our studies support a direct role of unmethylated Alus in cell biology and underscore the need to detect active repeat elements as an essential compartment of the genome that contributes to regulation and complexity of biological process.

(47)

Improved transposable element detection with profile HMMs in nhmmer

Travis Wheeler (1), Sean Eddy (1), Rob Finn (1), Robert Hubley (2), Tom Jones (1) and Arian Smit (2)*

1 HHMI, Janelia Farm Research Campus, Ashburn, VA 20147, USA
2 Institute for Systems Biology, Seattle, WA 98109, USA

State-of-the-art TE detection is homology based [1], with individual TEs usually identified based on similarity to a consensus sequence constructed to represent a known TE family [2]. While current TE detection methods cover substantial portions of many eukaryotic genomes (e.g. ~45% of human [3]), older TE instances often endure random mutatation to the extent that they are not recognizable using single-sequence homology search, resulting in underestimated masking. Profile methods are known to improve sensitivity over single sequence search [4], with profile HMMs in particular leveraging the additional information content in position-specific residue and indel variability.

nhmmer, a new tool in the HMMER3 suite, brings the power of profile HMMs to DNA homology search, with speed faster than cross_match and only 5-10x slower blastn with sensitive settings. The acceleration over past profile HMM implementations is due to a combination of a heuristic filtering pipeline and an efficient vector-parallel implementation [5], and has negligible impact on sensitivity.

Tests were performed using a benchmark made up of more than 14,000 "ghost" repeat instances, not found in default RepeatMasker search but supported by comparative genomics. While keeping a low false-discovery rate, we find a several percent increase in coverage over both blastn and cross_match through both (a) identification of TE instances not found by single-sequence homology methods and (b) increase in hit lengths.

References:

1. A Smit, R Hubley, P Green, RepeatMasker Open-3.0 1996-2004. http://www.repeatmasker.org
2. J Jurka, et al, Repbase Update, a database of eukaryotic repetitive elements. Cytogentic and Genome Research. 110, 462-467 (2005).
3. E Lander, et al, Initial sequencing and analysis of the human genome. Nature. 409, 879–888 (2001).
4. J Park, et al, Sequence comparisons using multiple sequences detect three times as many remote homologues as pairwise methods. Journal of Molecular Biology 284, 1201–1210 (1998).
5. Eddy S, http://hmmer.org

(48)

Reviving the Dead: retrotransposition of a reconstructed L1

Lei Yang, John Brunsfeld and Holly Wichman*

University of Idaho, USA

Our long-term survey of L1 (Long Interspersed Nuclear Element -1: LINE-1) activity in over 200 species of mammals has shown them to be active to varying degrees in all orders of placental mammals and the five marsupial orders for which we were able to acquire DNA. However, two independent extinctions of L1 activity were found in speciose groups within two different orders: the megabat family Pteropodidae (24 MYA) and most rodents from the subfamily Sigmodontinae (8 MYA). These two groups comprise ~9% of all mammalian species. We have resurrected the last active L1s of the megabat Pteropus vampyrus with consensus-based reconstruction. The reconstructed element has the expected amino acid at most conserved sites in both open reading frames (ORFs), but contains a 445-bp inter-ORF spacer, which is abnormally long compared to those of L1-active species. Retrotransposition rates of the reconstructed L1 and chimeras of it and human L1rp were tested in HeLa tissue culture with the G418 reporter plasmid. Both megabat ORFs are able to support retrotransposition, however, the long inter-ORF spacer appears to inhibit the retrotransposition, especially combined with either bat ORF. This is in line with the hypothesis that the gradual elongation of the inter-ORF sequence dampened transposition rates to below the replacement rate needed to maintaining an active lineage, leading to its eventual extinction.

(49)

Transposable element discovery and annotation in population using next-gen sequencing data

Fiston-Lavier, A.S. (1), Gonzalez, J. (1,2), and Petrov, D.A. (1)*

3 Department of Biology, Stanford University, US
4 Institute of Evolutionary Biology, (CSIC-UPF), Barcelona, ES

The availability of next-gen sequencing (NGS) data from multiple strains offers an unprecedented opportunity to annotate transposable element (TE) insertions in populations. Here, we present an integrated package called “T-lex” (version 2.0) that (i) ascertains the presence/absence of annotated TE insertions in the sequenced genome(s), (ii) estimates frequencies of TE insertions in multiple strains that are sequenced individually or as a pool, (iii) identifies putatively misannotated TE insertions defining their target site, and (iv) discovers and annotates “new” TE insertions not annotated or absent in the reference genome.

We experimentally validated each T-lex module starting with the TE annotations for 1,826 known TE insertions (Release 5.39) from the reference genome of Drosophila melanogaster in two re-sequenced single strains (W1 and CantonS) and pooled sequence library of 92 strains from the Drosophila Genetic Reference Panel (http://www.hgsc.bcm.tmc.edu/project-species-i-Drosophila_genRefPanel.hgsc). T-lex estimates the population frequency of these TE insertions with high sensitivity (100%) and specificity (97%). T-lex allowed the re-annotation of 148 TE insertions, mostly non-LTR elements that had misannotated poly(A) tails. As expected, we observed conserved insertion sites for DNA and LTR elements. T-lex also inferred with accuracy the target site duplications for more than 705 TE insertions for which the genomic sequence prior the TE insertion were detected. After removing the TE sequences on the chromosome 2L, all the TE insertions known to be present in the W1 strains were re-discovered and correctly annotated.

T-lex, an efficient, accurate and cost-saving tool for TE insertion annotation in NGS population genomic data, offers the opportunity to perform genome-wide population dynamics analyses of TEs and assess their impact on genome dynamics and evolution. T-lex is available for download at: http://petrov.stanford.edu/cgi-bin/Tlex_manual.html.

(50)

Regulation of Small Non-coding RNAs During Spermatogenesis

Xin Zhiguo Li(1), Christian K. Roy (1), Jia Xu (2), Bo Han (1), Ewelina Bolcun-Filas (3), Zhiping Weng (2), John C. Schimenti (3,4), Melissa J. Moore (1), Phillip D. Zamore (1)*

1 Howard Hughes Medical Institute and Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA
2 Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA
3 Cornell University College of Veterinary Medicine, Department of Biomedical Sciences, and Center for Vertebrate Genomics, Ithaca, NY 14850, USA
4 The Jackson Laboratory, Bar Harbor, ME 04609, USA

Three classes of small non-coding RNAs have been identified in animals: microRNAs (miRNAs), small interfering RNAs (siRNAs), and Piwi-interacting RNAs (piRNAs). In mammals siRNAs and piRNAs are detected in germ cells, protecting germ-line genome from transposon activating. In mammals, transposon-silencing piRNAs accumulated early in spermatogenesis, whereas pachytene piRNAs are produced later in sperm development and account for >95% of all piRNAs in adult mouse testes. Without pachytene piRNAs, mice fail to produce functional sperm, but neither the molecular function nor the trigger for pachytene piRNA production is known. Here, we show that production of pachytene piRNAs is induced by the transcription factor A-Myb. A-Myb drives transcription of both pachytene piRNA precursor RNAs and the mRNA encoding Miwi, the Argonaute protein through which pachytene piRNAs function, creating a coherent feed-forward loop.