International Coffee Genomics Network (ICGN)
Report 5th Coffee Genomics Workshop held at the XX Plant and Animal Genome (PAG) Meeting
San Diego, California
January 14-19, 2012


ICGN Coffee Genomics Workshop Speakers and Coffee related abstracts presented at PAG (see abstracts of presentations included at the end of this report):

Coffee Genomics Workshop Speakers

1. **Maia, Ivan.** IB -UNESP - Department of Genetics, Botucatu, Brazil (igmaia@ibb.unesp.br). CaPrx, a class III peroxidase gene from *Coffea arabica* specifically induced by root-knot nematode infection.
2. **Qi Sun.** CBSU- Cornell University, Ithaca, New York. (qisun@cornell.edu). Overview of the Genotyping-by-sequencing technology and its applications in diverse organisms.
3. **Lukas Mueller**. Boyce Thompson Institute, Ithaca, New York (lam87@cornell.edu). Sol Genomics Network (http://solgenomics.net/): a resource for coffee genomics.
4. **Clotilde Telling.** Roche Applied Science, Indianapolis, Indiana (clotilde.teiling@roche.com). *De novo* sequencing of complex plant genomes.

Coffee related abstracts presented at other PAG workshops

5. **Josiane Razafinarivo**. Geographycal gradients in genome size variation of Malagasy and African *Coffea* *(this abstract was presented at the Evolution of Genome Size workshop)*

Coffee genomics workshop and ICGN meeting participants

1. ***Francis Quetier***, Director Agence Nationale de la Recherche (ANR), France
2. ***Aleksey Zimin,*** *University of Maryland*
3. ***Alan Andrade,*** *EMBRAPA, Brasil* (alan@cenargen.embrapa.br)
4. ***Juliana Pereira Bravo***, IB -UNESP - Department of Genetics, Botucatu, Brazil.
5. ***Marcela Yepes***, Cornell University (my11@cornell.edu)
6. ***Herb Aldwinckle***, Cornell University

7. ***Chifumi Nagai***, Hawaii Agriculture Research Center (HARC) (cnagai@harc-hspa.com)
8. ***Ray Ming***, University of Illinois at Urbana-Champaign (rming@life.uiuc.edu)
9. ***Alexandre De Kochko***, IRD-CIRAD (dekochko@ird.fr)
10. ***Perla Hamon,*** *IRD-CIRAD* (perla.hamon@ird.fr)
11. ***Alicia Francis,*** *Roche*
12. ***Keithanne Mockatis*****,** Indiana University
13. ***Mohammed Mohiuddin*****,** *Roche*
14. ***Victor Albert****,* State University of New York (vaalbert@buffalo.edu)
15. ***José Goicoechea*****,** Arizona Genomics Institute, University of Arizona
16. ***Kang, Yang Jae*****,** Seoul National University, Korea (kangyungjae@gmail.com)
17. ***Jinguan Li*****,** Max Planck Institute for Plant Breeding Research, Cologne, Germany (jgLi@mpripz.mpg.de)
18. ***Suzy Strickler*****,** Boyce Thompson Institute, Ithaca, NY (srs57@cornell.edu)

Coffee Genomics Workshop at PAG

The Plant and Animal Genome (PAG) meeting is the largest international scientific conference reporting on animal and plant genomics developments in the world, with more than 3,000 participants from 65 countries. The XXI Plant & Animal Genome Conference will be held in San Diego, January 12-16, 2013. For those interested in participating in future meetings see http://www.intlpag.org.

Approximately 50 scientists participated in our 5th coffee genomics workshop held as part of the PAG Meeting in San Diego on January 15, 2012. The co-organizers of the workshop, Marcela Yepes (Cornell University, my11@cornell.edu), Philippe Lashermes (IRD-CIRAD, France, philippe.lashermes@ird.fr), and Rod Wing (University of Arizona) thank the speakers for their participation and contributions. Abstracts of workshop presentations on coffee are included as an appendix at the end of this report. We received very positive feedback from several coffee workshop participants to continue the organization of this event during future PAG meetings (http://www.intl-pag.org/). All ICGN members are invited to participate in the 6th Coffee Genomics Workshop that will be held January 13, 2013 as part of the XXI PAG meeting in San Diego, January 12-16, 2013. Please contact one of the organizers if interested in presenting a talk or poster, or with suggestions for new topics for workshop presentations or for round table discussion at the ICGN meeting. The coffee genomics workshop is an excellent opportunity to present advances in coffee genomics research to the International Plant and Animal Genomics Community and is helping our community explore new collaborations as well as funding opportunities.

ICGN survey and collaboration with the International Coffee Organization

ICGN is conducting a survey to prepare for the next ICGN meeting that will be held in conjunction with the 24th ASIC International Conference on Coffee Science, in Costa Rica (http://www.asic2012costarica.org). The survey is being conducted to help us update our mailing list, identify future priority projects for the community as well as new leadership to help secure funding for new proposals. ICGN members are being asked to

help us contribute to this effort by completing and submitting the survey that is available at our www site (http://www.coffeegenome.org). Please feel free to distribute the survey among colleagues. Results from the survey will be posted at our www site and discussed at the next ICGN meeting.

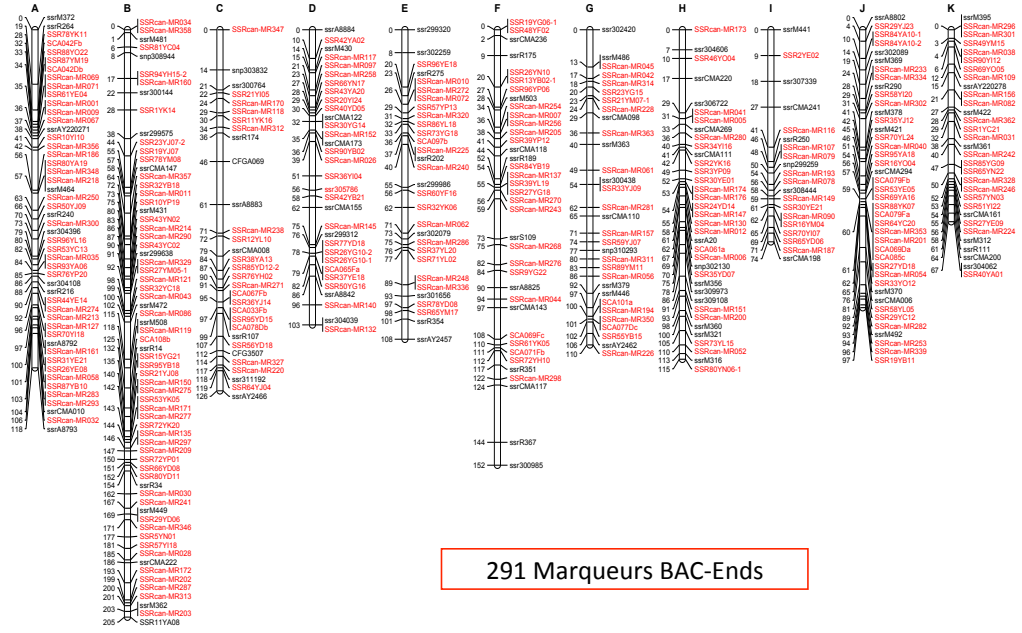
As the first *de novo* coffee genome reference sequences become available in 2012, we would like to take advantage of the momentum to identify new priority projects of interest that ICGN can develop as a community to help mine the data generated and develop innovative tools and advanced resources in coffee genomics to address challenging issues for our community such as climate change adaptation and sustainability that could be accelerated with transforming genomic technologies and strategies. The African Coffee Research Network (ACRN) has joined ICGN as an institutional member, and its Director of Research and Development, Dr. Bayeta Bellachew is now helping us conduct the ICGN survey among ACNR members at several Coffee Research Institutions in Africa. We have already received through ACRN responses from scientists and scientific groups from the following countries: Ethiopia, Kenya, Rwanda, Uganda and Ghana with strong interest to work with ICGN on a global initiative to develop advanced genomic tools to speed up diversity characterization, enhanced utilization and conservation of *Coffea* germplasm in the context of climate change. In addition with support from the International Coffee Organization (ICO), ICO member countries have been contacted to discuss possible interest in developing a global initiative in collaboration with ICGN/ICO aiming at improving conservation and characterization of the world coffee gene pool for varietal development in a world of changing farming systems and climate. Other ICO member countries that have expressed strong interest in working on an ICGN/ICO collaborative proposal include, for Europe: France (IRD-CIRAD); for Latin America: Brazil, Colombia, Guatemala, Costa Rica, Mexico; for Africa: Cote D'Ivoire, Ethiopia, Kenya, Malawi, as well as the InterAfrican Coffee Organization; and for Asia: India and Vietnam.

ICGN is grateful for the invitation by the new ICO Executive Director Dr. Robeiro Oliveira Silva to participate this year as an observer in the upcoming ICO Council meetings in London in 2012, and we are looking forward to working closely with ICO officials on the preparation and submission of a first ICGN/ICO proposal, and to explore potential sources of finance for such joint initiative. Support from ICO will be key for ICGN to secure future funding for diversity conservation efforts in *Coffea* with a broader funding base. The importance of cooperation on innovative coffee research for climate change adaption is key for future sustainability of the world coffee industry and working with ICO will strengthen ICGN efforts to secure funding for such initiative. ICO could help ICGN as a platform to help us share information for the community, help us fund outreach training opportunities in marker-assisted breeding, genomics and bioinformatics, as well as to secure funding for diversity studies and testing of germplasm and crosses generated to benefit breeding programs in multi-location trials for adaptation to climate change. ICO involvement will help us enhance collaborations among coffee research organization in different countries. ICO could help us support regional meetings with researchers interested in coffee genomics to include scientists from developing countries. Progress on coffee genomics research and

the potential of research results for coffee improvement can be discussed, targeting the priority traits for different regions as well as for the coffee industry. Capacity building in developing countries to participate in coffee genomic research can be supported through ICO networking to help us secure funding for those efforts.

Update Status High-Density Mapping of the diploid species *Coffea canephora*

A high density reference genetic map for *Coffea canephora* Pierre is being constructed in collaboration with Nestlé R&D Centre and the Indonesian Coffee and Cocoa Research Institute. The population being mapped is from a cross between two highly heterozygous genotypes, a Congolese group genotype (BP409) and a Congolese-Guinean hybrid parent (Q121). The segregating population is composed of 93 F1 individuals. DNA from the two parental clones and the segregating progeny were distributed to several ICGN members (on request). So far, a high-density genetic map including 1500 markers for which the respective DNA sequences are available has been achieved. This high density map will be used to anchor the scaffolds of the *C. canephora* sequences that are being generated, in order to reconstruct the chromosomal molecules. During the next few months, the mapping efforts will focus on the identification and mapping of SSRs from the *C. canephora* sequence scaffolds that are not or are insufficiently anchored. Both the high-density genetic map and the marker information will be freely available on a dedicated web-site (e.g. SGN web site). Please send updated data or information and comments to Dominique Crouzillat, Nestlé (dominique.crouzillat@rdto.nestle.com) or Philippe Lashermes (philippe.lashermes@ird.fr). Enclosed is a figure showing new SSR positions (in red) on the International Robusta map.



Update on the status of the *Coffea canephora* genome sequencing

With funding from the Agence Nationale de la Recherche (ANR), France, several Institutes (Genoscope-CEA, IRD and CIRAD) are combining their scientific resources and expertise to sequence, assemble, and annotate the entire genome of *C. canephora*. Additional partners include several ICGN members (EMBRAPA/Brazil, ENEA/Italy, University of Trieste/Italy, University of Queensland/Australia, CCRI/India, University of Illinois, Urbana/USA, Hawaii Agriculture Research Center HARC/USA, SUNY Buffalo/USA, University of Ottawa/Canada). A community effort will be needed for genome annotation. The *C. canephora* genome consists of 11 chromosomes, is about 710 Mb in size, and is being sequenced *de novo* with deep coverage using different sequencing platforms to obtain a reference genome for *C. canephora*. Genoscope has lead the sequencing and assembly of the *C. canephora* genome. Patrick Wincker, Head of Sequencing and Coordinator of Eukaryote Annotation and Analysis at Genoscope, presented the sequencing strategy and the status of the project during the 2011 coffee genomics workshop. Their group just completed a first genome assembly.

C. canephora is one of the ancestral progenitors of the widely cultivated, *C. arabica,* a recent allotetraploid species formed from the merger of the diploid species *C. canephora* and *C. eugenioides*. The accession DH200-94, a doubled haploid genotype was selected for sequencing because of its homozygous nature to facilitate genome assembly. *De novo* genome sequencing with deep coverage was performed using both 454 Roche and Illumina technologies. Direct whole genome shotgun (WGS) sequencing and paired-end sequencing of large insert libraries 8kb and 20 kb insert libraries was conducted. Furthermore, clones from two *C. canephora* BAC libraries were BAC-end sequenced using Sanger technology. The overall data generated are as follows:

Roche/454:

WGS 454 sequencing : 28.87 X coverage (assuming a genome size of 710 Mb) including:
– Reads single end : 14.83 X , Mean size: 359 bp
– Long reads single end : 8.24 X , Mean size : 462 bp
Paired-end sequencing of long insert libraries (8 and 20 Kb): 5.8 X (2.2X for 8kb, 3.6X for 20kb), Mean Size : 252 bp.

Sanger:

Two BAC libraries (*Hind III* and *Bam HI*) were constructed in collaboration with Rod Wing at Arizona Genomics Institute.
Sanger BAC-end sequencing: 144,000 BES were generated (72,000 BAC clones x 2 ends 5' and 3'): 0,27X
Mean BAC insert size : 135 Kb, range: 63,2Kb < insert < 253,6 Kb

Illumina sequencing was done at deep coverage (~70X) to correct 454 sequencing errors.

Assembly statistics for the first assembly (as of Jan 2012) (data from Genoscope):

No. of Scaffolds 13,345
Size: 568.6 Mb (80% estimated genome size)
Coverture 28.87 X (454/Sanger) and Illumina 69.7 X
N50: 1.2 Mb (108 scaf)
N80: 65 kb (635 scaf)
Largest scaffold: 9.0 Mb

Update of the *Coffea eugenioides* genome sequencing

This project is funded by the InterAmerican Development Bank (FONTAGRO/SECCI), and sequencing will be started in 2012. The project will be developed collaboratively by the Colombian National Coffee Research Center (CENICAFE), and Cornell University. Funding for this project was secured jointly through a proposal prepared and submitted by Cornell University and CENICAFE and the funded was approved and allocated for the project in 2011.

Once the reference genomes of the diploid species *Coffea canephora* and *Coffea eugenioides (*parental diploid ancestors of the allotetraploid species *Coffea arabica)* become available, they will serve as frames for sequencing and assembly of *C. arabica*, the major cultivated coffee species worldwide. The ICGN goal of sequencing the coffee genome is to build a solid foundation for deciphering the genetic and molecular bases of important biological traits in coffee that are relevant to growers, processors, and consumers. This knowledge is fundamental to allow efficient use and conservation of coffee genetic resources, and for the development of improved cultivars in terms of quality and reduced economic and environmental costs, as well as, to advance efforts to adapt the crop to climate change. To ensure full benefit from the generated coffee genomic sequences and resources by the coffee sector, ICGN will explore through International Funding Agencies, ICO, and the private sector, ways of securing funding for long-term maintenance of the coffee genome databases, and for the development of friendly end-user tools as well as to organize training courses to promote community annotation efforts.

Acknowledgements:

ICGN is particularly grateful to Clotilde Tiling from Roche who kindly accepted our invitation to participate as a speaker in our coffee genomics workshop with a presentation on the advances of the 454 GS FLX+ System that will be used to sequence the *C. eugenioides* genome. We are also grateful to Qi Sun at Cornell University for his contribution with the presentation on the use of genotype by sequencing for diversity studies in plants and its potential application for coffee diversity studies. We also thank Ivan Maia and Lukas Mueller for their presentations in the workshop.

Appendix

Abstracts 5th Coffee Genomics Workshop 2012

Workshop Co-Organizers:

Marcela Yepes, Cornell University (my11@cornell.edu)

Philippe Lashermes, L'Institut de Recherche pour le Développement (IRD), France (philippe.lashermes@ ird.fr)

Rod Wing, University of Arizona (rwing@Ag.arizona.edu)

CaPrx, a class III peroxidase gene from *Coffea arabica* specifically induced by root-knot nematode infection

Fabio E. Severino1, Marcos Brandalise1, Carolina S. Costa1, Mirian P. Maluf2 and **Ivan G. Maia**3, (1)IB - UNESP - Department of Genetics, Botucatu, Brazil, (2)Embrapa Café, Brasília - DF, Brazil, (3)IB -UNESP - Department of Genetics, Botucatu , Brazil

Class III peroxidases (Prx) are enzymes involved in a multitude of physiological and stress-related processes in plants. Here, we report on the characterization of a peroxidase coding gene from *Coffea arabica* (*CaPrx*) that is expressed in early stages of root-knot nematode (RKN) infection. *CaPrx* showed enhanced expression in coffee roots inoculated with RKN (at 12 h post-inoculation), but no difference in expression was observed between susceptible and resistant plants. In contrast, *CaPRx* was not responsive to mechanical injury indicating that the observed up-regulation is not related to a wound response. Assays using transgenic tobacco plants harboring a promoter-Beta-glucuronidase (GUS) fusion revealed that the *CaPRx* promoter was exclusively active in the galls induced by RKN. In cross sections of galls, GUS staining was predominantly localized in giant cells. Up-regulation of GUS expression in roots of transgenic plants following RKN inoculation was observed within 16 h. Altogether, these results point to a putative role of this peroxidase in the general coffee response to RKN infection.

Ivan Maia (Center) with his Research Group at IB -UNESP - Department of Genetics, Botucatu , Brazil **(**igmaia@ibb.unesp.br).

Overview of the Genotyping-by-sequencing technology and its applications in diverse organisms

Qi Sun, Jeff Glaubitz, Robert L. Elshire, Sharon Mitchell, James Harriman, Edward S. Buckler. Cornell University and USDA-ARS-Cornell University

Genotyping by sequencing (GBS) is a simple highly-multiplexed system for constructing reduced representation libraries for the Illumina next-generation sequencing platform developed in Ed Buckler lab of USDA. This presentation will give the overview of this technology, current progress, and its applications in constructing genetic maps, trait mapping and genome contig scaffolding in multiple different species. Prospect will be discussed of applying GBS technology, integrated with whole genome sequencing assembly, to study the genetic diversity in **coffee**, and marker assisted breeding.

Sol Genomics Network (http://solgenomics.net/): a resource for coffee genomics.

Lukas Mueller, Susan R. Strickler, Naama Menda, Aureliano Bombareli Gomez, Jonathan Leto, Robert Buels, and Joseph Goseelin, Boyce Thompson Institute, Ithaca, New York, USA

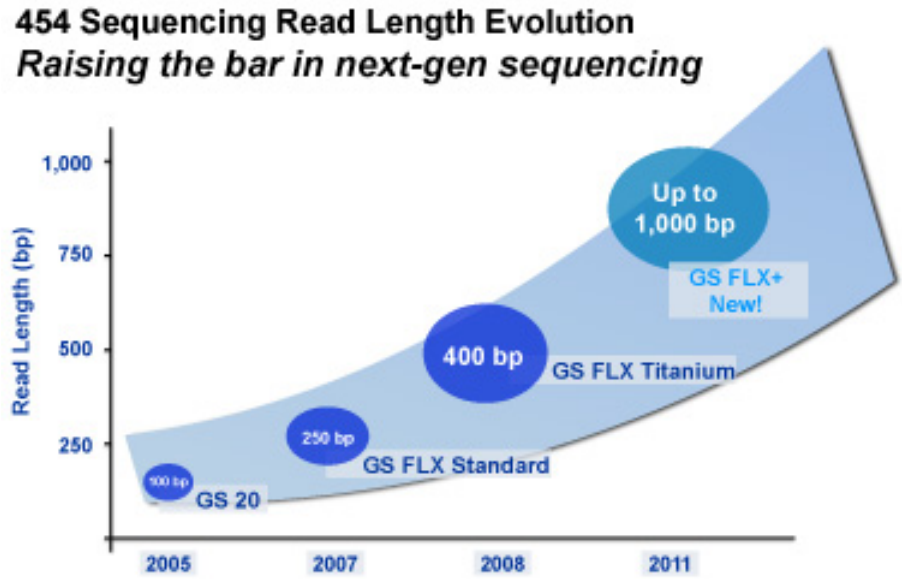
The Sol Genomics Nework (SGN, http://solgenomics.net/) is a clade oriented database for the Solanaceae and closely related Asterids such as Rubiaceae, including coffee. Currently, SGN has several resources available for coffee, such as annotated transcript assemblies, a community curated set of over a hundred loci, and CoffeaCyc, a pathway database based on the Pathway Tools software suite. The SGN database has many further resources that as of yet do not contain coffee data. For example, SGN can store phenotypic and genotypic information for plant accessions. Currently, there are only a handful of coffee accessions in the database, and no genotypic information has been uploaded. Easy to use web interfaces for users to provide these data are available. If both genotypic and phenotypic information are in the database, for certain types of mapping populations, SGN can perform QTL and other analyses directly on the web. SGN also has a number of comparative features based on genetic maps, full genome sequences, and gene family data that we would like to apply to coffee in the near future.

De novo Sequencing of Complex Plant Genomes

Clotilde Teiling[1]
[1]Roche Applied Science, Indianapolis, Indiana

De novo sequencing of complex genomes has been increasingly utilized for research on non-model organisms, unculturable species (metagenomics), or comparative transcriptome analysis.

Roche's GS FLX+ System features the latest innovations in 454 Sequencing reagents and software to support accurate *de novo* sequencing of complex genomes. With read lengths of up to 1,000 base pairs, the GS FLX+ System enables a broad range of applications, including DNA shotgun sequencing, RNA transcript analysis, and targeted resequencing to generate optimal assemblies for accurate assignment of repeat regions. 454 extra-long reads are an ideal tool to bridge through inverted repeats, extensive gene duplications, and multiple inversions found in many plants – enabling variant comparison between strains or within genomes.



The GS FLX+ System provides a powerful combination of features enabling detailed analysis of assembled sequencing data. This presentation will demonstrate the utility of 454 Sanger-like read lengths for accurate assemblies, leading to a better understanding of plant evolution for food or bioenergy research as well as environmental studies.

Roche's group that participated in the coffee genomics workshop including Alicia Francis (center) and Clotilde Teiling (right)

Long 454 GS FLX+ reads provide accurate scaffolds, and co-assemble with short reads for more cost-effective *de novo* assemblies of complex genomes and more comprehensive transcript coverage: Single reads span more exons for more complete transcript reconstruction.

Abstract presented at the workshop on Evolution of genome size:

Geographical gradients in the genome size variation of Malagasy and African *Coffea*

Josiane Razafinarivo[1], Jean Jacques Rakotomalala[1], Spencer Brown[2], Mickael Bourge[2], Serge Hamon[3], Alexandre de Kochko[3], Valerie Poncet[3], Christine Dubreuil-Tranchant[3], Emmanuel Couturon[3], Romain Guyot[3] and Perla Hamon[3], (1)FOFIFA, ANTANANARIVO, Madagascar, (2)CNRS, GIF SUR YVETTE, France, (3)Institut de Recherche pour le Développement, Montpellier cedex 5, France

Abstract Text:
Coffea genus (Rubiaceae), recently increased to 110 species, is divided into two subgenera: *Coffea* subgenus *Coffea* (101 species), native mainly to Africa and the Indian Ocean islands and *Coffea* subgenus *Baracoffea* (nine species), restricted to Madagascar. To date, genome size data are available for only 25 species mainly of African origin. The aim of this study, using flow cytometry on lyophilized material, was to assess the genome size of 44 species natives to the Indian Ocean islands and to investigate its correlation with biogeography, leaf traits and phylogenetic relationships. The mean 2C nuclear DNA content of Mascarocoffea species ranged about 1pg to 1.40 pg. The overall 2C DNA values from African and Indian Ocean islands species were non-linearly distributed. A geographical map according to the putative native origin of the corresponding species showed a gradient in Madagascar and Africa. Genome sizes increased following a North to South-East gradient in Madagascar and an East to Central-West gradient in Africa. No, or only weak, correlations were noted between genome size and leaf parameters. Genetically close species could be highly distinctive in their genome size while divergent species could be similarly sized. The non-random geographic distribution and habitat of species, and the absence of correlation between genome size and genetic relationships, suggested that during *Coffea* genome evolution both DNA content increase and/or decrease occurred independently in Africa and in the Indian Ocean Islands.

Upcoming meetings relevant to our community

- **SOL 2012: 9th Solanaceae Conference, 26-30 August, 2012, Switzerland** *(*http://www.epsoweb.org/event/sol-2012-9th-saolanaceae-conference-26-30-august-2012-switzerland *)*
- **ASIC meeting in Costa Rica November 11-16, 2012 (http://asic2012costarica.org)**
- **PAG meeting in San Diego Jan 12-16, 2013 (http://www.intl-pag.org)**