MINUTES OF ICGN MEETING HELD AT THE XVIII PLANT AND ANIMAL GENOME MEETING (PAG), SAN DIEGO, CALIFORNIA JANUARY 11, 2010

ICGN member participants (14 scientists)

Alan Andrade	EMBRAPA, Brazil
Marco Cristancho	CENICAFE, Colombia
Philippe Lashermes	IRD, France
Alexandre de Kochko	IRD, France
Bertrand Benoit	CIRAD, France
Christophe Montagnon	CIRAD, France
Alberto Pallavicini	University of Trieste, Italy
Giovanni Giuliano	ENEA, Italy
Nicolas Roux	Bioversity International, France
Marcela Yepes	Cornell University, USA
Ray Ming	University of Illinois, USA
Chifumi Nagai	Hawaii Agriculture Research Center, USA
Lukas Mueller	Boyce Thompsom Institute for Plant Research, USA
Aureliano Bombarely	Boyce Thompsom Institute for Plant Research, USA

Non-ICGN participants (3 scientists)

David Sankoff	Department of Mathematics and Statistics, University of Ottawa, Canada (field bioinformatics)
Adriana Muñoz	School of Information Technology and Engineering, University of Ottawa, Canada (field bioinformatics)
Nicole Rice	Southern Cross University, SCU, Australia

Agenda ICGN meeting

1. The third coffee genomics workshop was held January 10, 2010 as part of the XVIII PAG meeting (See workshop program at http://www.intl-pag.org/18/18-coffee.html). Abstracts of the oral and poster presentations on coffee at the XVIII PAG meeting are included as appendix at the end of this report.

2. Progress reports presented by the coordinators of the ICGN working groups: Ray Ming for Working group 2, Philippe Lashermes for working group 3, Alex de Kochko and Alan Andrade for Working group 4, and Lukas Mueller for working group 6.

3. Update of progress of working group 3 of the ICGN towards an international initiative for *de novo* sequencing of the coffee genome using multiple next generation sequencing platforms. Project to sequence the genome of the diploid species *Coffea canephora* has been funded by the Agence Nationale de la Recherche (ANR) France, sequencing is currently on going at Genoscope, France.

4. Update on the international effort to generate a reference genetic map for *C. canephora* using high-density sequence-based markers. Nestlé R&D and the Indonesian Coffee and Cocoa Research Institute have recently agreed to share their efforts with ICGN members toward the development of a reference genetic map of *C. canephora*. Countries that have agreed to participate in the effort include Brazil, India, Italy, France and USA. Other groups interested on participating, please contact Philippe Lashermes or Dominique Crouzillat (dominique.crouzillat@rdto.nestle.com).

5. Invite ICGN members to participate in the next Coffee Genomics Workshop at the XIX Plant and Animal Genome Meeting in San Diego to be held January 15-19, 2011 (http://www.intl-pag.org/).
   If interested to attend, please send abstracts to one of the workshop organizers: Philippe Lashermes (Philippe.Lashermes@mpl.ird.fr), Marcela Yepes (my11@cornell.edu) or Rod Wing (rwing@Ag.arizona.edu).

Summary meeting

Third Coffee Genomics Workshop at PAG

Approximately 50 scientists, including 14 ICGN members, participated in the 3rdcoffee genomics workshop held as part of the Plant and Animal Genome Meeting in San Diego January 10, 2010. The co-organizers of the workshop appreciated the participation of the invited speakers and their contributions.

Program and abstracts of the oral presentations and posters on coffee are included as an appendix at the end of this report. We received very positive feedback from the 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 4th Coffee Genomics Workshop that will be held January 16, 2011 as part of the XIX PAG meeting in San Diego, January 15-19, 2011. Please contact one of the organizers if interested on presenting a talk or poster, or with suggestions for new topics for oral presentations or for the round table discussion. The coffee genomics workshop is a good opportunity to present advances on coffee genomics research to the International Plant and Animal Genomics Community and is helping our community explore funding opportunities.

Progress report from the coordinators of the different ICGN working groups present

Working group 2- Genetic Mapping. Ray Ming. Different groups are working on mapping diploid and tetraploid populations. However with the lower cost of sequencing using next generation sequencing technologies, transcription sequencing of individuals of mapping populations should accelerate mapping efforts using sequence based markers such as SSRs and SNPs. A suggested method is to extract the RNA of the two parents, sequence it using Solexa Illumina to define SNPs, and map the SNPs to the progeny

Last year, a pilot project based on shotgun sequencing of *Coffea canephora* using the FLX454 Titanium technology was discussed in San Diego, and several ICGN members expressed interest. Ray Ming (working group 2), Alex de Kochko (working group 4), Georgio Graziosi, and Alan Andrade (working group 4) as well as other ICGN members paid for 12 runs of 454 GS FLX Titanium and one Illumina run. The 454 Sequencing was done at the University of Illinois, USA, and ENEA, Italy, and Illumina PE run at SCU, Australia. The data and additional markers (SSRs) derived from the 454 *C. canephora* WGS runs are expected shortly.

International-Genetic-Coffee-Map

Update by Philippe Lashermes, Chair ICGN Steering Committee. Nestlé R&D and the Indonesian Coffee and Cocoa Research Institute have recently agreed to share their efforts with ICGN members towards the development of a reference genetic map of *Coffea canephora*. An identified common objective is the establishment of a high-density genetic map (about 2000 markers). ICGN

members that would be willing to contribute on a volunteer basis to this mapping effort are welcome to contact Philippe Lashermes (ICGN) or Dominique Crouzillat (Nestlé R&D):

- Upon request, DNA samples from the 93 individual segregating plants of the population BP409 X Q121 and the two parental clones will be sent to the participating ICGN members by R&D Nestlé at Tours (France).

- After genotyping of the whole population, the participating ICGN members will send back the genotype data and the relevant information regarding the analyzed markers (i.e. sequence, primers, sequence accession number).

- The additional sequence-characterized markers will be mapped on the already existing *Coffea canephora* map comprising about 1,200 loci and the new genetic map as well the markers information's will be freely available on a dedicated web-site (e.g. SOL web site).

Ideally for genome assembly, the reference coffee genetic map will need to have 2 sequence based markers per 1 million bp. A high density saturated map with sequence based markers such as SSRs and SNPs will be necessary to anchor scaffolds to the genetic map once the coffee genome sequences are generated. Therefore ICGN will still need to map approximately another 1,000 markers into the Indonesian population. Working group 2 and 4 could also contribute to the effort once the new SSR markers are available from the 454 and Illumina sequencing. Currently IRD/CIRAD, University of Trieste, University of Illinois, Brazil and India have expressed interest on the mapping effort.

ICGN has also interest on having access to RNA of the parents and the population to develop SNPs by *de novo* sequencing. However, Nestlé R&D and Indonesia will need to be officially approached by ICGN with this request.

A proposal for SNP development as a community effort to make it most cost effective was discussed. Ray Ming will follow up on preparing and circulating the proposal among ICGN members. The suggested strategy based on the Indonesian pseudo-test cross population that was generated from crossing two F1 parents, would be to sequence all progeny (93 individuals) to generate markers

Working group 3- Physical Mapping and *de novo* Coffee Genome Sequencing Philippe Lashermes. On behalf of ICGN, working group 3 has been working on developing a cost efficient strategy to *de novo* sequence the coffee genome, as well as to secure interagency funding for the project. The strategy to sequence the coffee genome will involve the adaptation of next generation sequencing technologies (454-Roche and Solexa/Illumina), and will target high coverage *de novo* sequencing to generate reference genomes for both

the two diploid parental species of *C. arabica*: *C. canephora* and *C. eugenioides*. The reference sequences of the diploid ancestors of *C. arabica* will serve as frame-work for sequencing the *C. arabica* genome.

Two proposals submitted to the Agence Nationale de la Recherche ANR (France) on behalf of ICGN were approved for funding. The first phase of the proposal is now completed and allowed to BAC end sequence at Genoscope the two *C. canephora* BAC libraries (*Hin*d III and *Bam*HI) that were constructed in collaboration with Rod Wing at the Arizona Genomics Institute. In 2009, 73,000 *C. canephora* BAC clones were BAC end sequenced using Sanger technology at Genoscope. The second phase of the project associating 3 French Institutes (CIRAD, IRD and Genoscope/CEA) is on going to *de novo* sequence the *C. canephora* genome using multiple sequencing platforms and high depth coverage to obtain a reference genome. The initiative was presented at different ICGN meetings (Campinas, 2008; San Diego, 2009) and in a White-paper that was sent to all ICGN members in March 2008. Full technical details of the project were given by P. Wincker (Genoscope) and discussed during the last PAG meeting. The objective is to establish and annotate the complete sequence of *C. canephora* (acc. 200-94). All information will be fully available to further expert annotation and analyses by the ICGN community.

In parallel and since the last ICGN meeting (San Diego, 2009), the pilot project based on shotgun sequencing using the FLX454 Titanium technology (see Working group 2 activities) has evolved to an international consortium aiming the draft sequencing of *C. canephora* (See abstract presentation by A. de Kochko). Discussions were held during the PAG meeting on ways to integrate the two initiatives to benefit ICGN and avoid waste of resources as well as duplication of efforts.

CENICAFE and Cornell, on behalf of working group 3 of the ICGN, submitted a proposal to the InterAmerican Development Bank (FONTAGRO) in 2008. The proposal was approved in 2009 to construct a BAC library for the diploid maternal parent of *C. arabica* (*C. eugenioides*) and *de novo* sequence the *C. eugenioides* genome. The project will be started in 2010.

Once completed, the high coverage reference sequences for the diploid species *C. canephora* and *C. eugenioides,* using a combination of 454 Titanium and Illumina, should serve as solid frame work for future sequencing and assembly of the genome of the allotetraploid species *C. arabica,* the main cultivated coffee species through out the world.

Working group 4- Transcriptomics Alan Andrade discussed the need to increase the number of ESTs for *C. canephora*. Currently aprox 60,000 ESTs for *C. canephora* are available (20,000 from Brazil and 40,000 from Nestlé). A collaborative project sequencing *C. canephora* using different tissues and specific environments could increase the number. ENEA (Italy) is interested on

contributing to coffee transcriptomics through 454 sequencing of cDNAs from other partners and synthesis of a *Coffea* microarray using its custom combimatrix platform.

Working group 6- Bioinformatics Lukas Mueller

Lukas Mueller updated our group on the preliminary release of the tomato genome by the Solanaceae community. Through the *Solanaceae* genomics network (http://sgn.cornell.edu/) comparative resources of interest to the coffee community are available.

A future concern for ICGN as large amounts of 454 and Illumina data are generated from the coffee genome sequencing projects will be data analysis and long term storage. Also, it will be important to discuss as a community unigene strategy, centralized and mirror builds, as well as a community annotation effort.

The problem of database longevity apparently has not been solved yet even for the *Arabidopsis* and *Solanaceae* communities. Therefore, ICGN will have to promote this effort and explore funding through private companies or other resources. Montpellier/Agropolis is considering the creation of a bioinformatics resource site to follow up with data analysis after primary coffee genome annotation. Community annotation of the coffee genome should be possible with the expressed sequenced tags (ESTs) collections developed by different groups (>350,000 ESTs). Expert annotation within our community would be extremely valuable, as well as, from collaborations with other communities.

Starting in 2011, we will organize in conjunction with the coffee genomics workshop at the PAG meeting in San Diego, a coffee bioinformatics session to discuss coffee resources.

Upcoming meetings of interest to the coffee community

World Coffee Conference
Guatemala City, Guatemala (26 to 28 February 2010)
The next World Coffee Conference will take place in Guatemala City, Guatemala, from 26 to 28 February 2010. The Conference will bring together coffee growers, representatives from government, the private sector and international agencies and will provide a unique opportunity to address the challenges of world coffee supply and demand, coffee development sector and sustainability. Further information about the Conference can be found on the conference website: www.wcc2010guatemala.com

23rd International Conference on Coffee Science - ASIC 2010

Bali - INDONESIA 3 – 7 October 2010
www.asic-cafe.org

APPENDIX PROGRAM AND ABSTRACTS

3rd Coffee Genomics Workshop at the Plant and Animal Genome Meeting,

San Diego, Jan 10, 2010.

Sunday Afternoon, 10 January, 2010 3:50 pm to 6:00 pm

Coffee Genomics Workshop - Pacific Salon 2

Co-Organizers:

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

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

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

Speakers:
Alan ANDRADE, EMBRAPA (Brazil) (alan@cenargen.embrapa.br) **"Drought tolerance in coffee: Identification of candidate genes and study of its natural variation"**

Philippe LASHERMES, IRD (France) (philippe.lashermes@ird.fr) **"*Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleo-hexaploid ancestral genome"**

Marco CRISTANCHO, CENICAFE (Colombia) (Marco.Cristancho@cafedecolombia.com) **"In Silico characterization of gene families present in the coffee genome (*Coffea* sp.)"**

Benoit BERTRAND, CIRAD (France) (benoit.bertrand@cirad.fr) **"Gene Expression divergences between the allopolyploid *Coffea arabica* and its diploid relatives appear environment-dependent"**

Alexandre De KOCHKO, IRD (France) (dekochko@ird.fr) **"Progress on the preliminary sequence of the *Coffea canephora* genome"**

Aleksey ZIMIN, University of Maryland (USA) (alekseyz@ipst.umd.edu) **"*De Novo* Genome Assembly From The Next Generation Sequencing Data"**

Drought Tolerance In Coffee: Identification Of Candidate Genes And Study Of Its Natural Variation

Pierre Marraccini[1,2] , **Luciana P Freire**[1] , **Natalia G Vieira**[1] , **Gabriel S C Alves**[1] , **Felipe Vinecky**[1] , **Thierry Leroy**[2] , **Gustavo C Rodrigues**[3] , **Antônio F Guerra**[3] , **Alan C Andrade**[1]

[1] Embrapa Recursos Genéticos e Biotecnologia, LGM-NTBio, Parque Estação Biológica, 70770-917 Brasília-DF, Brazil
[2] CIRAD UMR DAP, Montpellier SupAgro, 2 place Viala - 34060 Montpellier, France
[3] Embrapa Cerrados, BR 020 Km 18, 73310-970, Planaltina-DF, Brazil

Drought stress significantly affects coffee yield, productivity and quality. Thus, the goal of this study was to investigate the molecular mechanisms underlying the response to drought stress in coffee plants by different approaches. Candidate gene identification was performed by comparing gene expression and protein profile of different genetic materials (tolerant vs. susceptible) as well as, under different conditions of water supply (irrigated vs. non-irrigated). In this work, the genetic materials studied were Conillon clones of *Coffea canephora* and two cultivars of *C. arabica*. The applied water stress (PD=-3,0 MPa) to the conillon plants was achieved under green-house conditions and, in the case of arabica, adult plants cultivated under field conditions were used. Under field conditions, leaves were collected during day and night, and the most pronounced observed water-stress was of PD =-1,7 MPa. After selection of candidate genes by different strategies, the expression was confirmed by qPCR analysis. The natural variation of some selected candidate genes was also performed using a set of different genotypes. The data obtained indicated that several genes displayed decreased expression upon water stress and usually these were encoding-genes of proteins involved in photosynthesis. On the other hand, the applied water stress on coffee plants also induced a set of genes such as RD29, DREBA and NAC, which have already been described in literature as genes involved in plant responses to drought. In addition, this study also revealed the importance of other factors controlling the expression of these genes, such as the circadian clock and the age.

In silico Characterization Of Gene Families Present In The Coffe Genome (*Coffea* sp.)

Marco Cristancho[1] , **Edgar F. Salcedo**[1] , **Alvaro Gaitan**[1] , **Jaime Arcila**[1] , **Herb Aldwinckle**[2] , **Marcela Yepes**[2]

[1] Colombian National Coffee Research Center – CENICAFE, Chinchiná, Caldas, Colombia
[2] Department of Plant Pathology and Plant Microbe Biology, Cornell University, Geneva, New York 14456, USA

Exponential growth of genomic information from different sequencing projects makes essential the deployment of bioinformatics tools for visualization and data analysis. To characterize the main families of genes present in the coffee genome, we compared the predicted protein sequences of coffee with those of the model organisms *Arabidopsis thaliana* and *Populus trichocarpa*. The prediction of proteins was performed from 58,343 public EST sequences from four species of *Coffea*, including 41,985 sequences from *C. arabica*. A pipeline for protein family identification for coffee was standardized, by using several bioinformatics tools such as EST scan, BLASTP and the algorithms TRIBEMCL and OrthoMCL. With this pipeline, we identified 8,588 gene families of coffee that were shared with *Arabidopsis* and *Populus*, and 4,289 families unique to the *Coffea* genus. We are currently analyzing these unique *Coffea* families. We also investigated in detail 417 families identified as genes for resistance to different pathogens. The allocation of each family member was validated by comparing the annotations to InterProScan and Blast. The families identified will also help the search for agronomically important candidate genes associated with genomic regions of interest from the identification of orthologous regions and the comparative genomic analysis of coffee with better-characterized model species. It is hoped that the identification and comparison of orthologous genes and/or paralogs present in other related species will facilitate evolutionary studies of some of the families of most interest in coffee. Details of the bioinformatics pipeline and visualization of the gene families can be found at http://bioinformatics.cenicafe.org .

Coffea (Asterids) And *Vitis* (Rosids) Derive From The Same Paleo-Hexaploid Ancestral Genome

Alberto Cenci , **Marie-Christine Combes** , **Philippe Lashermes**

IRD - Institut de Recherche pour le Développement, UMR RPB (CIRAD, IRD, Université Montpellier II), BP 64501, 34394 Montpellier Cedex 5, FRANCE

Polyploidy or whole genome duplication (WGD) is a widespread phenomenon in plants and is thought to have played a major role in their diversification and adaptation. Subsequent gene loss and rearrangements further affect gene copy numbers and fractionate ancestral gene linkages across multiple chromosomes. Analysis of the complete sequence of *Vitis vinifera* revealed that the rosid clade derives from a paleo-hexaploid ancestor. The *Coffea* genus belongs to the *Rubiaceae* family, one of the largest tropical angiosperm families. *Rubiaceae* as well as *Solanaceae* are included in the Asterid I clade of dicots. To elucidate the genomic history of asterids, the sequence of an 800 kb region of diploid *Coffea* genome was compared to the orthologous regions of *V. vinifera*, *Populus trichocarpa* and *Arabidopsis thaliana*. A very high level of co-linearity between around 80 genes of rosids and *Coffea* was found indicating that the *Coffea* genome (and consequently the ancestral genome of all asterids) and rosids share the same hexaploid ancestor. Moreover, the high level of co-linearity between the *Coffea* and *V. vinifera* genomic regions we analyzed shows that the diploidization process (loss of duplicated and redundant copies from the whole genome duplication) was very advanced in the most recent common ancestor of rosids and asterids. Finally, no additional genome duplication was detected in the *Coffea* lineage. Differences in gene loss rates were detected among the rosid species and linked to the divergence in protein sequences.

Gene Expression Divergences Between The Allopolyploid *Coffea arabica* And Its Diploids Relatives Appear Environment-Dependant

Amélie Bardil[1] , **Marie-Christine Combes**[2] , **Philippe Lashermes**[2] , **Benoit Bertrand**[1]

[1] CIRAD - Centre de Coopération Internationale en Recherche Agronomique pour le Développement, UMR RPB (CIRAD, IRD, Université Montpellier II), BP 64501, 34394 Montpellier Cedex 5, FRANCE
[2] IRD - Institut de Recherche pour le Développement, UMR RPB (CIRAD, IRD, Université Montpellier II), BP 64501, 34394 Montpellier Cedex 5, FRANCE

Polyploidy is widespread among many major crops. In coffee, the main cultivated species, *Coffea arabica*, is an allotetraploid containing two diploid subgenomes which originated from two different diploid species, *C. canephora* and *C. eugenioides*. Here we showed that the gene expression changes between the natural but recent coffee allopolyploid species in its two diploid relatives is environment-specific. Using spotted 70-mer oligo-gene microarrays targeting 15,522 unigenes, leaf gene expression patterns from plants growing in two temperature conditions were compared between the two parental species and *C. arabica*. At the lowest temperature, we observed a massive dominance and transgressive expression in *C. arabica* when compared to its two relatives since 47 to 49 % of unigenes were differentially expressed with the proportions of up- or down-regulation approximately equal (23-24%). Surprisingly at the warmest temperature, we observed a strong disequilibrium. The divergence between *C. arabica* and *C. eugenioides* was rather identical to that observed at the lowest temperature since we observed over 40% of the unigenes differentially expressed, but on the other hand the divergence between *C. arabica* and *C. canephora* were only 9%. These data show that numerous genes in *C. arabica* are non-additively expressed and that divergences in gene expression pattern between allo and diploid genomes are function of the environment conditions. These results reinforce the hypothesis of a better functional plasticity of the allopolyploids in comparison to their related diploids species and consequently the evolutionary advantage of this genome architecture.

Progress On The Preliminary Sequence Of The *Coffea canephora* Genome

Alexandre de Kochko[1] , **Victor Albert**[2] , **Alan C. Andrade**[3] , **Giovanni Giuliano**[4] , **Giorgio Graziosi**[5] , **Robert Henry**[6] , **Ray Ming**[7] , **Chifumi Nagai**[8] , **Steve Rounsley**[9] , **David Sankoff**[10]

[1] UMR DIAPC, IRD - BP34394, 34394 Montpellier Cedex 5, France
[2] Department of Biological Sciences, University at Buffalo (SUNY), Buffalo NY, USA
[3] Laboratory of Molecular Genetics (LGM-NTBio), Embrapa, Recursos Genéticos e Biotecnologia, C.P. 02372, 70770-900 Brasília-DF, Brazil
[4] ENEA Casaccia Research Center- Post Bag 026, 00123 S.M. di Galeria Rome, Italy
[5] Dipartimento di Scienze della Vita, Università di Trieste P.le Valmaura 9 34143 Trieste, Italy
[6] Southern Cross University, PO Box 157, Lismore NSW 2480, Australia
[7] Department of Plant Biology, University of Illinois at Urbana Champaign, Urbana, IL 61801, USA
[8] HARC, 99-193 Aiea Heights Drive, Aiea, HI 96701, USA
[9] BIO5 Institute & School of Plant Sciences, Keating Bioresearch Building, 1657 E. Helen St, Tucson, AZ 85719 USA
[10] Department of Mathematics and Statistics University of Ottawa 585 King Edward Avenue Ottawa, Canada K1N 6N5

An international consortium of 10 laboratories from 6 countries started the sequencing of the diploid species *Coffea canephora* during the second semester of 2009. The genus *Coffea* is a member of the family Rubiaceae, one of the largest among angiosperms, mainly represented in tropical areas, and a member of the Asterid clade. Interestingly, although coffee is the second most valuable commodity exported by developing countries, ours is the first known effort to get a broad view of its genome sequence. Based on deep sequencing technologies, the genome of a double haploid plant was used for performing 12 runs with the Roche Titanium technology and 1 run with the Illumina G2 technology. Available sequences in public data banks were used to help the assembly of the obtained sequences. We expect to obtain an assembly sufficient enough to allow an assessment of the general organization of the genome, to permit comparisons with existing sequenced genomes, and to lead to better understanding of *Coffea* genome evolution. Identification of the great majority of genes should provide insight into specific metabolic and developmental pathways. A dramatic increase in the quantity of genetic markers will also be provided, permitting the establishment of more dense genetic maps for *C. arabica*. Identification of transposable elements and analysis of their distribution will also be greatly facilitated.

De novo Genome Assembly From The Next Generation Sequencing Data.

Aleksey V Zimin , **Steven L Salzberg** , **James A Yorke**

University of Maryland, College Park, MD 20742, USA

Next generation sequencing technologies such as 454 sequencing by Roche, ABI SOLiD or Illumina sequencing provide large amounts of genomic data quickly and at significantly lower cost, compared to traditional Sanger sequencing. As the read lengths increase, one can apply the traditional genome assembly software, such as Celera Assembler to create de novo assemblies of the genomes sequenced utilizing the next generation technologies. We applied the Celera Assembler to the sequence data for domestic turkey, produced by Roche, VBI and USDA to create a *de novo* assembly. The results obtained from the data set that includes 5x coverage by the 454 reads, 30x coverage by Illumina reads, and less than 0.1x coverage by Sanger BAC ends indicate, that the Celera Assembler can be used to create a high quality assembly with scaffold N50 size of over 1.5Mb and contig N50 size of over 12Kb.

Other abstracts related to coffee

Metabolic Pathway Networks For Cereal Plants In The Gramene Database

Palitha Dharmawardhana[1] , **Liya Ren**[2] , **Jim Thomason**[2] , **Doreen Ware**[2,3] , **Pankaj Jaiswal**[1]

[1] Oregon State University, Department. of Botany and Plant Pathology, 2082 Cordley Hall, Corvallis, OR, 97331-2902, USA
[2] Cold Spring Harbor Laboratory Cold Spring Harbor Laboratory, 1 Bungtown Rd, Cold Spring Harbor, NY, 11724, USA
[3] USDA-ARS NAA Plant, Soil & Nutrition Laboratory Research Unit, Cornell University, Ithaca, NY, 14853, USA

The Gramene database (www.gramene.org), a comprehensive comparative plant genomics platform developes and curates RiceCyc and SorghumCyc pathway databases for cereal plants. RiceCyc with 342 known and/or predicted metabolic pathways for *Oryza sativa japonica cv. Nipponbare* has undergone several rounds of data quality enhancement and manual curation whereas SorghumCyc with 328 pathways for *Sorghum bicolor Strain BTX623* is in its initial computational build. The plant metabolic pathways module within *Gramene* mirrors several other species specific pathways such as *Arabidopsis, Medicago,* Tomato, Potato and **Coffee** as well as MetaCyc reference database allowing the user to extract interspecific comparison between pathways and associated genes. The user is also able to download lists of genes associated with each pathway. The database comes with the Omics Viewer data visualization tool. This tool allows users to overlay microarray, transcriptomic, proteomic, and metabolomic datasets with expressed values on pathway maps. The overlaid views allow to visualize the pathways and reactions that are up/down regulated in an experiment or a set of experiments. We have also built an Omics Validator tool to validate user provided expression data files by mapping probe IDs from various microarray platforms to their respective gene IDs.