May 2008
Issue Number 19
Editor: Joyce Van Eck

# SOL Newsletter

## Contents of this Issue



## Community News

### The Potato Genome Sequencing Consortium

*Contributed by Christian Bachem*

The Potato Genome Sequencing Consortium (PGSC) currently comprises thirteen groups from different countries across the world and includes all major producers and breeders of potato (Figure 1). The aim of these research groups is to elucidate the complete genomic sequence of this important food crop by the year 2010. The availability of the complete sequence of the 840 Mb of potato DNA will enable potato breeders and related industries to exploit the genetic potential of the fourth most important field crop in the world. The basis of the sequencing project is an ultra-high dense molecular marker map of the diploid potato clone RH89-039-16. This genotype has been used to create and anchor clones from a library of around 75,000 BACs provided by Wageningen University's Plant Breeding Laboratory who are coordinating the project.

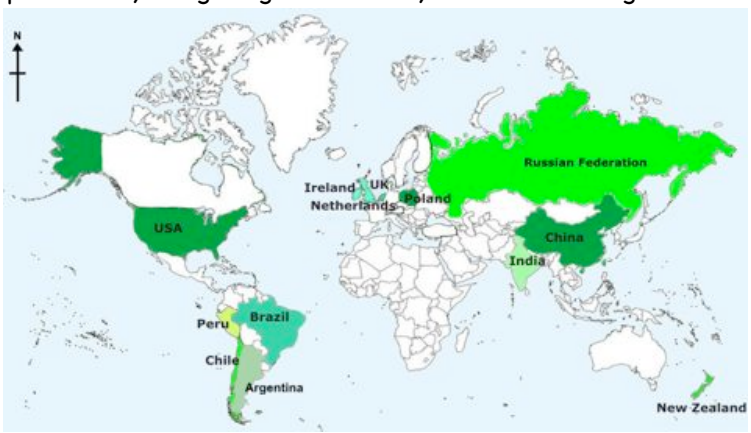


**Figure 1.** Geographical distribution of participants of the PGSC. Contact information for the partners can be found on www.potatogenome.net.

The coordinators estimate that about 9 of the 12 chromosomes of potato are currently fully financed (Chromosomes 1 and 5, Netherlands; 2, India; 4, UK/Ireland; 6, USA; 9, New Zealand; 10, 11, China; 12, Russia) two further chromosomes are financed to at least 50% (chromosomes 3, Latin America and 8, Poland) leaving chromosome 7 currently free for additional participants. Additional grants have been submitted and potential participants have being contacted to fill these gaps.

The core effort of the project is a BAC-by-BAC and chromosome-by-chromosome

sequencing strategy. The majority of this sequencing work is being carried out using traditional "Sanger sequencing", however, an increasing emphasis is being put into the use of new generation sequencing equipment such as the 454 Genome Sequencer for parallel BAC sequencing. In addition, strategies for whole genome shotgun sequencing are also being discussed as an additional resource for speeding up the sequencing effort.

Currently, around 1400 BAC contigs have been anchored on the physical map. This corresponds to at least 2800 BACs that can potentially be used as seed BACs to start sequencing. The availability of BAC-end sequences for the complete library created by our US partner Robin Buell (Michigan State University) allows partners to pick extension BACs for further sequencing. In addition to sequence-based verification of physical chromosomal location of BAC clones, the group in Wageningen has been using fluorescent *in situ* hybridization (FISH) to verify the physical map. 115 potato BACs have now been checked by FISH and only five of these appeared to be located away from the predicted location giving an 88% success rate. Furthermore, using five-color FISH, it will now be possible to color-code all potato chromosomes giving an unequivocal cytogenetic chromosomal assignment (Figure 2).

The individual progress on sequencing chromosome by chromosome is given on the project web site (www.potatogenome.net). At this time, about 1200 BACs have been sequenced or are in the process of being sequenced. The policy of the consortium is to submit BAC sequences simultaneously to the PGSC database and to Genbank with the possibility of putting a six-month hold on release for quality control.

An important part of PGSC activities is also capacity building for countries with less developed infrastructures. Currently, students from Latin America are being trained in Wageningen in bioinformatics and similar schemes are being set up for researchers from other groups. Further integration of the *Solanaceae* bioinformatics efforts will be pursued by a recently formed bioinformatics subcommittee of the PGSC.

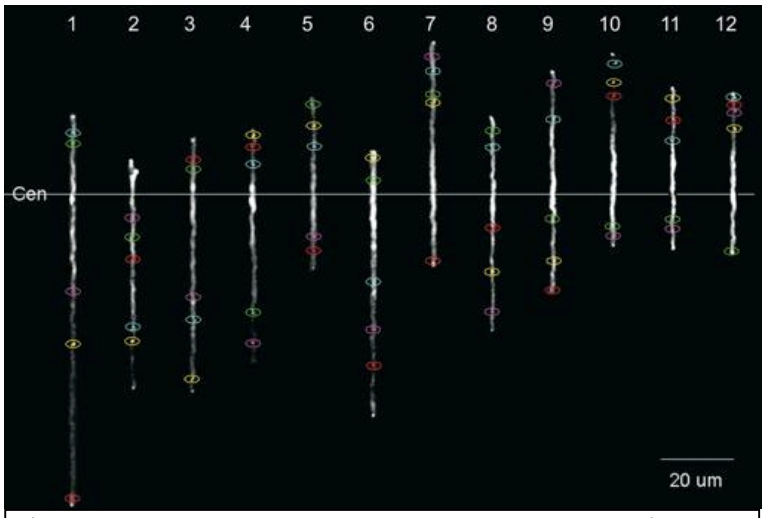


**Figure 2.** DAPI stained pachytene chromosomes of potato from root tips showing a five-color FISH using BACs known to be located at various positions & generating unique color "barcodes" for each chromosome. Figure was reproduced by kind permission of Dr Xiaomin Tang, WUR, Wageningen.

Over the last year it has become clear that there has been a very rapid increase in sequencing volume by the groups involved. With several groups still just at the beginning of their active sequencing phase this productivity is set to increase over the current year. To help with this, both the construction of a new BAC library using sheared DNA and the increasing exploitation of new sequencing technologies are likely to facilitate the timely completion of the targeted end date of the PGSC project in 2010.

## Breeders Toolbox

A Breeders Toolbox feature has been added to the SOL Genomics Network (SGN, www.sgn.cornell.edu). At this time, we are soliciting input from the community on the tools/information to include in this toolbox that would make the results generated by sequencing projects easily accessible and applicable for breeding programs.

Please contact Joyce Van Eck (jv27@cornell.edu) with suggestions on how to improve SGN for breeders.

# Fifty Years of *Solanaceae* Collection at the Radboud University Nijmegen, the Netherlands

*Contributed by Gerard van der Weerden*

In 1958, about a year after the opening of the Faculty of Mathematics and Physics, later Faculty of Science, Mathematics and Computer Science, the first seedlist of the Botanical Garden was published. Prof. Dr. H.F. Linskens initiated to build up a *Solanaceae* collection. In the preface of the first seedlist he wrote:

> "*The Catholic University of Nijmegen (The Netherlands), has adjoined a faculty of Science. With this we have started to build up a Botanical Garden in connection with the Botanical Laboratory. We are going to give our special attention to the family of Solanaceae. Herewith we have the pleasure to send you our first seedlist. We request you to place us on your mailing list for seeds catalogue*".

The seedlist contained 190 accessions. Now, fifty years later, the seedlist is replaced by a website and many more accessions are available for the scientific community (http://www.bgard.science.ru.nl/).

To celebrate this event, there will be a one-day symposium on June 12, 2008. You are welcome to attend the symposium. There are no registration costs, but please register by sending us an e-mail with the subject 'registration' by June 1st.

For the program and additional information see the symposium poster. The poster accompanies the pdf file of this newsletter at http://www.sgn.cornell.edu/solanaceae-project/index.pl#SOL_news or simply follow the link for the newsletter on the SGN homepage (sgn.cornell.edu).

**Contact information:**

Titti Mariani c.mariani@science.ru.nl

Gerard van der Weerden g.vanderweerden@science.ru.nl

## Solanaceae Poetry

### Tomato/Plant Reproduction Haiku

by Ashley Denney

Emasculation

pollination, frustration

where's your anther been?

Ashley is an undergraduate in Patricia Bedinger's lab at Colorado State University.

# Tomato Sequencing Updates

**Chromosomes 1, 10 (US)**

*Contact: Joyce Van Eck (jv27@cornell.edu)*

307,000 sheared fosmid clones (800 384-well plates) have now been ordered and replicated. End sequences for approximately 58,200 fosmid clones are available on SGN.

The total number of BAC clones localized by the US group using FISH now stands at 147. The number of BACs positioned for each of the chromosomes is: Chr. 1 - 24; Chr. 2 - 7; Chr. 3 - 13; Chr. 4 - 16; Chr. 5 - 10; Chr. 6 - 9; Chr. 7 - 13; Chr. 8 - 4; Chr. 9 - 18; Chr. 10 - 15; Chr. 11 - 13; and Chr. 12 - 5. BAC clones located within 0.5 micrometer of the telomere have been identified for fifteen of the twenty-three tomato chromosome arms (neglecting the short arm of chromosome 2). BACs located within 1 micrometer of the telomere have been identified for an additional five arms. BACs located within 1 micrometer of the euchromatin/heterochromatin border (as estimated from recombination nodule data) have been identified for seventeen chromosome arms. The majority (thirty of thirty-seven) of the BACs positioned close to either telomeres or euchromatin/heterochromatin borders are associated with mapped marker sequences.

Figure 3 contains an image of FISH labeling with three BACs in the euchromatin on the long arm of chr 10. BACs LE_HBa0234C10 (purple signal), LE_HBa0256L16 (green signal), and LE_HBa0011E16 (red signal) contain markers mapped at 45, 58.5 and 73 cM, respectively. The FISH signals have been superimposed on the phase contrast image of the SCs in the lower photograph. This labeling experiment was performed by undergraduate Dylan Westfall as part of an Honors Thesis project with Dr. Stephen Stack.

Lukas Mueller has accepted a tenure-track faculty position at the Boyce Thompson Institute for Plant Research (BTI) beginning June 2008. He will continue and expand his program in plant bioinformatics. The physical proximity of BTI on the Cornell campus means minimal impact of this move to both the SGN group and the SGN site.

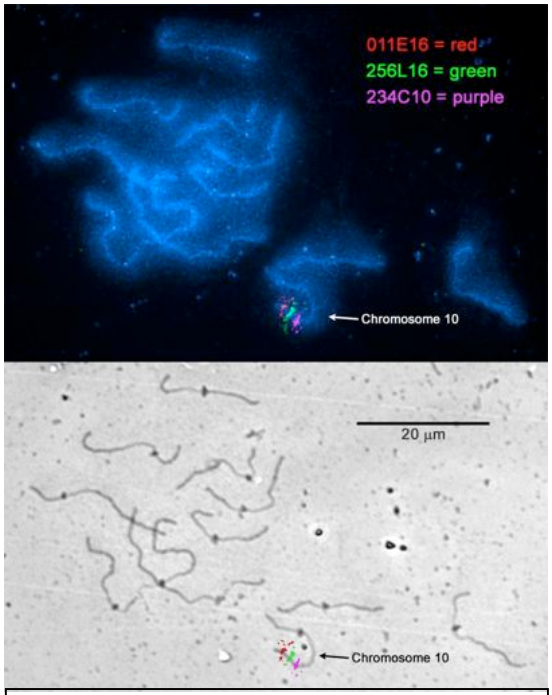


**Figure 3.** FISH of three BACs in the euchromatin on the long arm of chr 10.

**Chromosome 2 (Korea)**

*Contact: Sunghwan Jo (shjo@kribb.re.kr)*

As of today, 153 BAC clones (17,592,982 bp) have been completed as HTGS phase 3. Nine additional BACs are currently in the sequencing pipeline. Of this figure, 12,824,899 bp are unique. In the last period, we have focused on validation of BAC assembly with bridge BACs and IL mapping. We generated twenty-nine contigs that constitute 11,182,936 bp and fifteen singletons that constitute 1,164,963 bp. The longest contig constitutes 1,143 kb having fourteen BACs. Gaps between contigs might be short in that most gaps are 1-3cM apart on the Tomato-Expan 2000. However, we had a difficult time finding candidate BACs to bridge the gaps, so we are using overgo screening from three libraries with BAC ends where no candidates for extending have

been found by BES BLAST. Furthermore, some BACs (for example C02Slm0014P22, C02Slm008E03, C02Slm0108P14, C02Hba0075D08, C02Hba0044O16) represent a different organization with the tomato genome meaning artificially rearranged. We think this is likely due to chimera clones during BAC library construction. We also found transposons in tomato BAC sequences that moved from the *E. coli* genome. Those BAC sequences caused confusion in assembling the contig.

**Chromosome 3 (China)**

*Contact: Chuanyou Li (cyli@genetics.ac.cn)*

Recently, the groups of Professor Jianru Zuo (Institute of Genetics & Developmental Biology, Chinese Academy of Sciences) and Professor Bin Han (National Center for Gene Research , Chinese Academy of Sciences) joined in the chr 3 sequencing efforts. Currently, forty additional BACs are in various stages of the sequencing processes.

**Chromosome 4 (UK)**

*Contact: Karen McLaren (kb1@sanger.ac.uk) or Helen Beasley (hr1@sanger.ac.uk)*

10,606,245 bp of sequence have been generated at the Wellcome Trust Sanger Institute for chr 4 as of April 25th 2008. Of this figure, 10,084,624 bp are unique. The sequence has been produced from 102 BACS originating from the LE_HBa, SL_MboI, and SL_EcoRI libraries. We intend to finish all BACS that will contribute to chr 4 to HTGS phase 3 and currently eighty-seven BACs that correspond to 8,956,260 bp of sequence have been deposited in the public databases at EMBL/GenBank/DDBJ as phase 3. All other chr 4 BACs with EMBL/GenBank/DDBJ accessions are currently active in our sequencing pipeline at HTGS phases 0 to 2.

FISH analysis for additional chr 4 contig confirmation performed by the Hans de Jong Laboratory in Wageningen has identified a total of fifteen sequenced BACs to date that do not localize to chr 4. These BACs have been removed from the chr 4 minimal tilepath and have been moved to the unmapped BAC registry on SGN. The majority of these unmapped BACs appear to be heterochromatic, however, two of these BACs have since been mapped to other chromosomes. Further chr 4 BACs are being verified as potential sequencing candidates following a round of hybridizations for missing markers and we intend to use the IL mapping technique for additional chr 4 confirmation prior to sequencing.

The progress of chr 4 can be viewed through the development of the TPF and AGP files that we upload to SGN. The TPF indicates the expected relative positions of the BACs along the chromosome and the AGP provides assembly information of the finished sequences.

**Chromosome 5 (India)**

*Contact: Akhilesh Tyagi (akhilesh@genomeindia.org)*

At the Indian Initiative on Tomato Genome Sequencing, we have been able to confirm the position of sixty-seven BACs on chr 5. Sequencing is in progress on all these BACs, out of which twenty-one are in phase III, twenty-eight are in phase II and ten are in phase I. The remaining eight BACs are in the early phase of sequencing or library preparation. All the phase II and phase III sequences have been submitted to GenBank and their assembly data uploaded to SGN. A search is on to find new extension BACs by performing overgo hybridizations on the filters available for the three tomato libraries. In addition, new nucleation points are also being identified by developing CAPS markers for the 200 BACs assigned to India for mapping purposes.

**Chromosome 6 (The Netherlands)**

*Contact: Sander Peters (sander.peters@wur.nl)*

We have finished BACs H028D14, M026P18, M074G21, and E005I21 to phase 3. In total, we have completed forty-five BACs to phase 3 and 110 BACs are in the finishing pipeline. Furthermore, we have fingerprinted twelve seeds and 159 candidate BACs with SNaPshot, which are currently processed with FPC to identify new extension BACs.

We identified several domains on the long arm euchromatin, which are poorly covered by seed BACs. The largest domain has a genetic distance of 11 cM on the EXPEN 2000 genetic map. This domain is delimited by BACs, which have been anchored to the genetic map, and which we have used for BAC FISH to determine the physical distance on pachytene chr 6. Furthermore, we calculated the sequence size from overlapping BAC inserts flanking the 11cM interval and this allowed us to estimate a ratio value of 0.6 Mb/micrometer. The distance between anchored BACs delimiting the 11cM interval was 4.28 micrometer corresponding to a sequence distance of approximately 2.5 Mb. Currently, we use pooled-BAC FISH and multi-color FISH to aid the targeting of additional seed BACs.

**Chromosome 7 (France)**
*Contact: Farid Regad (regad@ensat.fr)*
Update pending.

**Chromosome 8 (Japan)**
*Contact: Shusei Sato (ssato@kazusa.or.jp)*

We finished 128 BAC clones to Phase 3 that produced a non-redundant length of 12,562,802 bp. We are continuing the accumulation of Selected BAC Mixture (SBM) shotgun data, which reached to 2.2 million files generating 1.5 Gb of total length. We mapped 712 EST SSRs markers on the map of EXPEN 2000 F2 population. Using the information on sixty-two of these markers mapped on chr 8, we obtained thirty-four additional seed clones for sequencing.

**Chromosome 9 (Spain)**
*Contact: Antonio Granell (agranell@ibmcp.upv.es)*
Update pending.

**Chromosome 11 (China)**
*Contact: Zhonghua Zhang (zhangzhonghua.cass@gmail.com) or Sanwen Huang (huangsanwen@caas.net.cn)*

Seventeen seed BACs have been finished as phase 3, and four BACs are in the sequencing pipeline. Combined with BACs previously sequenced by other chromosome projects, a total of twenty-five BACs for chr11 have been sequenced and submitted to GENBANK and SGN, of which sequences of two BACs (C00HBa00291F09 and C00SLm0121I03) are deposited in Chr00 directory of SGN. In addition, two BACs (LE_HBa0153A06 and LE_HBa0316E10) have been mapped on chr11, but the sequences are not available.

**Chromosome 12 (Italy)**
*Contact: Mara Ercolano (ercolano@unina.it)*

To date, we have sequenced fifty BACs: twenty-one in HTGS1, eleven in HTGS2, and eighteen in HTGS3. FISH analysis for chr 12 contig confirmation performed by the Hans de Jong Laboratory in Wageningen has identified two BACs that do not localize to chr12 (results were confirmed by IL-based verification). One additional BAC was shown to be chimeric and will be annotated. We will continue to use the FISH analysis as a resource for mapping confirmation. BACs are being verified as potential sequencing candidates following de novo IL mapping procedure. Using this method, forty-nine BACs have been mapped on various chromosomes, of which four are on chr 12. The data have been uploaded on SGN. The IL-based mapping method was also useful to verify the map position of eleven BACs previously mapped on chr12 by Syngenta. Of these, eight were confirmed on chr 12, while three mapped to other chromosomes.

## Second International Tomato Finishing Workshop

Participants of the
Second International Tomato Finishing Workshop

Following on from the success of the first workshop held a year ago at the Wellcome Trust Sanger Institute, the Second International Tomato Finishing Workshop was held recently at the WICC in Wageningen (April 24-25, 2008).

The two-day workshop was attended by delegates from the ten countries involved in the sequencing project representing all twelve chromosomes being sequenced by the consortium. Also in attendance were members of the International Tomato Annotation Group (ITAG).

Updates on progress and finishing strategies were heard from all groups with discussions on the tomato finishing standards document being the main focus of day one. The use of new technology data was also introduced as a topic, which needs to be taken into consideration, at a later date, with respect to the finishing standards.

Day two continued with talks on how to submit new technology to the public databases and was followed by a

session on finishing tools used by the various groups including a demonstration of TOPAAS developed by PRI at Wageningen.

Strategies for finishing problematic sequence and highly repetitive clones were covered in more detail before a presentation from the annotation group to highlight some findings from initial gene predictions. The workshop concluded with a summary of recent changes to the SGN web interface including how to access and make full use the clone registry information.

Ongoing discussion points from the workshop will be circulated to the tomato sequencing mailing list including some key issues that require final agreement from all the sequencing partners across the whole consortium. Examples include the use of restriction digest data for assembly confirmation and the use of miscellaneous feature tags on sequence submissions to embl/ddbj/Genbank.

If you wish to subscribe to the tomato sequencing mailing list, please contact Joyce Van Eck (jv27@cornell.edu).

# Announcements

## Publications


Ames M, Spooner DM (2008) DNA from herbarium specimens settles a controversy about origins of the European potato. Am J Bot 95:252-257.

Brumbarova T, Matros A, Mock H-P, Bauer, P (2008) A proteomic study showing differential regulation of stress, redox regulation and peroxidase proteins by iron supply and the transcription factor FER. Plant J 54:321-334.

Spooner DM, Núñez J, Trujillo G, del Rosario Herrera M, Guzmán F, Ghislain M (2007) Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. Proc Natl Acad Sci 104:19398-19403.


## Conferences

Rhine River, Cologne, Germany

**SOL 2008**

**5th Solanaceae Genome Workshop**

October 12 – 16, 2008

Cologne, Germany

http://www.sol2008.org

Early Registration

April 1 – July 14, 2008

Abstract Submission

April 1, 2008 – July 14, 2008

## Solanaceae Recipes

# Curried Potatoes and Peas (Aloo Mutter)

*From http://www.globalgourmet.com*

1 Tbsp ghee or butter
1 small onion, chopped
3 cloves garlic, minced
2 tsp finely chopped ginger
1 tsp ground cumin
1/2 tsp ground turmeric
1/2 tsp ground coriander
15 oz. can crushed tomatoes
2 red potatoes, scrubbed, diced
1 1/2 cups frozen peas
1 tsp garam masala
2 - 4 Tbsp fresh cilantro, chopped

In a large saucepan, saute the onion in the butter over medium-low heat for 5 minutes. Add the garlic and ginger and saute for one more minute. Add the cumin, turmeric, and coriander; stir well.

Add the tomatoes and bring the mixture to a boil. Add the potatoes, cover, and simmer for 10 minutes. Add the peas and cook for 10 more minutes. Stir in the garam masala.

Just before serving, sprinkle the fresh cilantro on top.