

Michiel van Eijk¹, Jan van Oeveren¹, Marcel Prins¹, Marco van Schriek¹, Anker Sørensen¹, Edwin van der Vossen¹, Nathalie van Orsouw¹, René Hogers¹, Diana Rigola¹, Antoine Janssen¹, Hanne Volpin¹, Mark van Haaren², An Michiels², Anne Deslattes-Mays² and Arjen van Tunen¹

KeyGene's Crop Genome Center assembling genomes of the 6F crops: food, feed, fiber, fuel, flowers and fun

¹ Keygene N.V., Agro Business Park 90, 6700 AE, Wageningen, The Netherlands
² Keygene Inc., 155 Gibbs Street, Rockville, MD 20850, United States of America



Abstract

The genomes of the world's most important crop species differ enormously in size, ploidy levels, repeat composition and germplasm diversity, posing a tremendous challenge to utilize them for breeding and trait improvement. KeyGene's recently established Crop Genome Center addresses this challenge by developing and applying sequence-based methodologies to advance genome assemblies, elucidate genetic diversity and clone trait genes of 6F crops. Our recently developed Whole Genome Profiling (WGP) technology has been applied to build sequence-based physical BAC maps of a number of crops with increasing genome complexity, including melon, tomato, strawberry, Brassicas, wheat, tobacco and others. Novel strategies for draft whole genome sequencing (WGS) have been developed based on the next-generation sequencing platforms GS FLX Titanium (Roche Applied Science) and Genome Analyzer (GA) II (Illumina Inc.) in combination with WGP. Using these methods, the melon genome has been sequenced to 20X using GS FLX Titanium and 60X with Illumina GAII, resulting in a high-quality reference genome assembly. Whole genome sequencing for Brassicas is currently ongoing.

WGP technology

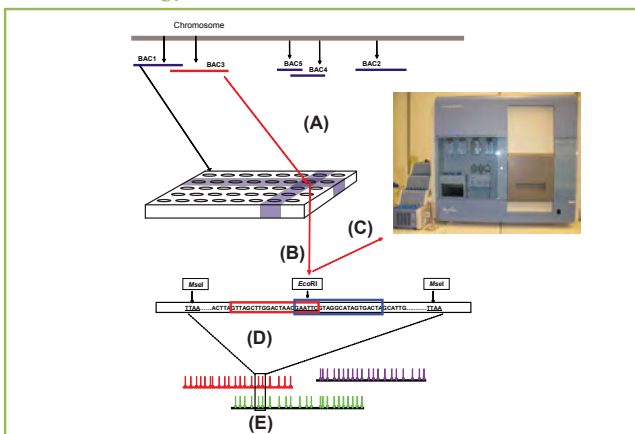


Fig. 1: Overview of the WGP technology
 A) BAC libraries (~10 GE) in 384-well plates are used for DNA isolation
 B) BAC DNAs are pooled in a multidimensional format
 C) Pooled BACs are digested with restriction enzymes, barcoded adaptors are ligated and fragment termini sequenced using the Illumina GAII platform
 D) 30 to 50 WGP sequence tags are assigned to individual BACs by deconvolution
 E) Sequence-based BAC contigs are built from WGP tags using adapted FPC

Integrating WGP and WGS

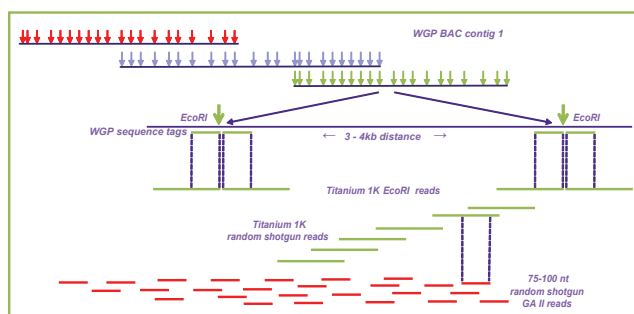


Fig. 2: Genome sequencing using WGP and WGS



Fig. 3: Linking WGP contigs with WGS scaffolds to generate super scaffolds

Table 1: Results high stringency WGP maps

WGP parameter	Arabidopsis	Strawberry	Wheat	Melon	Tomato	Solanaceous crop	B. oleracea	B. rapa	B. napus
genome size (Mbp)	130	170	280	450	950	2600	720	550	1200
genome equivalents BACs tested*	5.9	9.9	7.2	13.2	10.9	10.7	12.5	8.4	11.8
# unique WGP tags (FPC ready)	41,824	66,055	47,895	117,238	261,913	1,106,231	124,961	71,719	175,844
# tagged BACs (FPC ready)	4,303	11,621	13,888	39,195	66,084	203,094	61,249	28,511	95,160
average # WGP tags/BAC	34	32	16	26	33	47	24	25	22
# contigs	255	471	786	1085	2521	9247	1542	779	3069
# BACs in contigs	4,056	10,122	11,002	29,982	52,617	163,166	44,499	21,348	65,293
# singleton BACs	247	1499	2886	9213	13467	39928	16750	7163	29867
average contig size (# BACs)	16	21	14	28	21	18	29	27	21
N50 contig size (# BACs)	25	38	35	54	26	32	62	67	41
average contig size (kbp)*	438	503	414	346	378	341	421	450	384
N50 contig size (kbp)*	702	720	777	546	563	460	695	779	585
total genome coverage (Mbp)*	112	237	326	375	953	3150	649	351	1180
% genome coverage	86%	139%	116%	83%	100%	121%	90%	64%	98%

* assuming 125 kb average BAC insert size

Conclusions

- WGP provides high resolution sequence-based physical maps and is widely applicable in complex genomes of the 6F crops;
- WGP in combination with Whole Genome Sequencing enables generation of draft whole genome sequences with superior assemblies without Sanger sequencing and paired-end sequencing and provides direct access to BAC clones;
- The GS FLX Titanium 1K platform is well suited to complete draft sequences in combination with WGP.

Acknowledgements: We are grateful to 454 Life Sciences / Roche Applied Sciences for providing access to its GS FLX Titanium 1K

KeyGene® is a registered trademark of Keygene N.V.
 Keygene N.V. owns patents and patent applications covering its Whole Genome Profiling technology