High-throughput BAC Fingerprinting For Constructing Physical Maps

Ming-Cheng Luo

Department of Plant Sciences University of California, Davis

Why still think about fingerprinting when sequencing is easily accessible?

Vast majority of species will not be sequenced in the near future.

For those species, contigs of large-insert genomic clones anchored to genetic maps represent a low-cost alternative to genome sequencing that would greatly enhance the accessibility of their genomes for biological research.

For large genomes, a physical map is prerequisite for genome-wide sequencing.

By population of restriction fragments:

- One 6-cutter (Olson et al. 1986)
- One 6-cutter + one 4-cutter (Coulson et al. 1986; Klein et al. 2000)
- Multiplexing one 6-cutter + one 4-cutter (Ding et al. 1999)
- One type IIS restriction followed by determine of the nucleotide sequence at the cleavage site (Brenner and Livak 1989; Ding et al. 2000)
- Four 6-cutter + one 4-cutter (Luo et al. 2003)

By detection method:

- **Staining** (Olson et al. 1986)
- Radioisotope labeling (Coulson et al. 1986)
- Fluorescence dye labeling (Ding et al. 1999)

By electrophoresis type

- Agarose gel (Olson et al. 1986)
- Polyacrylamide gels (Coulson et al. 1986)
- Polyacrylamide gel based sequencer (Gregory et al. 1997)
- Capillary sequencer (Luo et al. 2003)

Multiple digestion

Multi-color fluorescence labeling

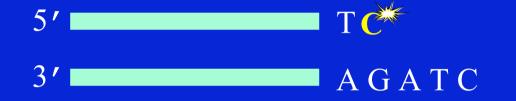
Capillary electrophoresis

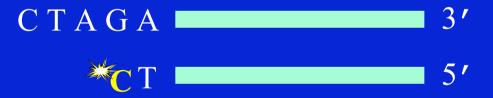
SNaPshot BAC Fingerprinting

Restriction cleavage



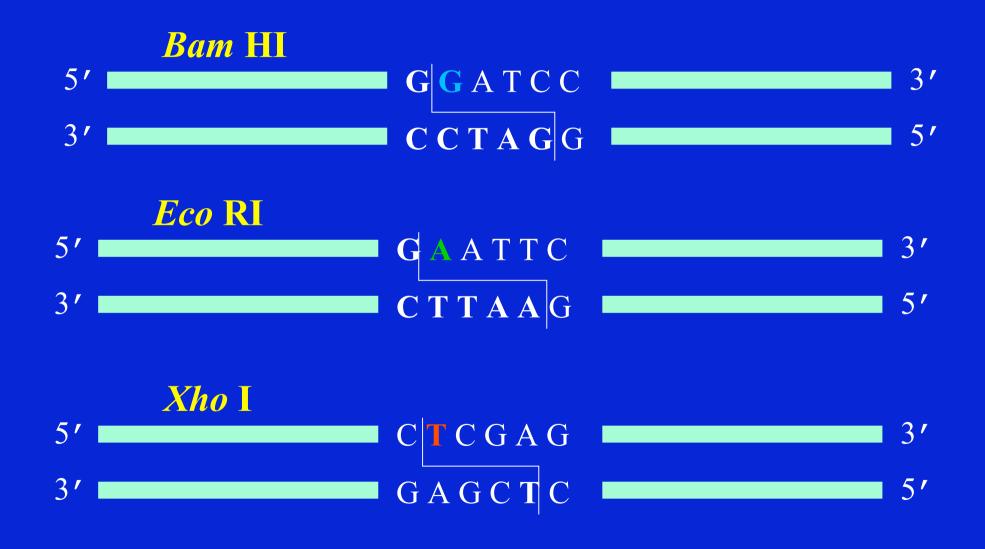
Fluorescent labeling





Xba I

Restriction cleavage and fluorescent labeling



Characteristics of restriction sites and labeling of fragments

Restriction endonuclease	Restriction site	ddNTP	Fluorescent dye	Color of fragment
<i>Eco</i> RI	G^AATTC	Α	dR6G	Green
BamHI	G^GATTC	G	dR110	Blue
XbaI	T^CTAGA	С	dTAMRA	Yellow
XhoI	C^TCGAG	Т	dROX	Red
HaeIII	GG^CC	none		

Other enzyme combination?

Considerations:

- Generate 5' overhang
- AGCT only
- 6-bp cut ?
- Buffer compatibility

Cheap: 7 More expensive: 40 Don't care: 55

Α	G	С	Τ
<i>Eco</i> RI (\$212) *	BamHI (\$212)	<i>Xba</i> I (\$840)	Sall (\$1060)
<i>Hin</i> dIII (\$212)	BglII (\$1060)		Xhol (\$504)

* Prices are based on NEB's list price of 50,000 units at available largest package

Selection of enzyme combination

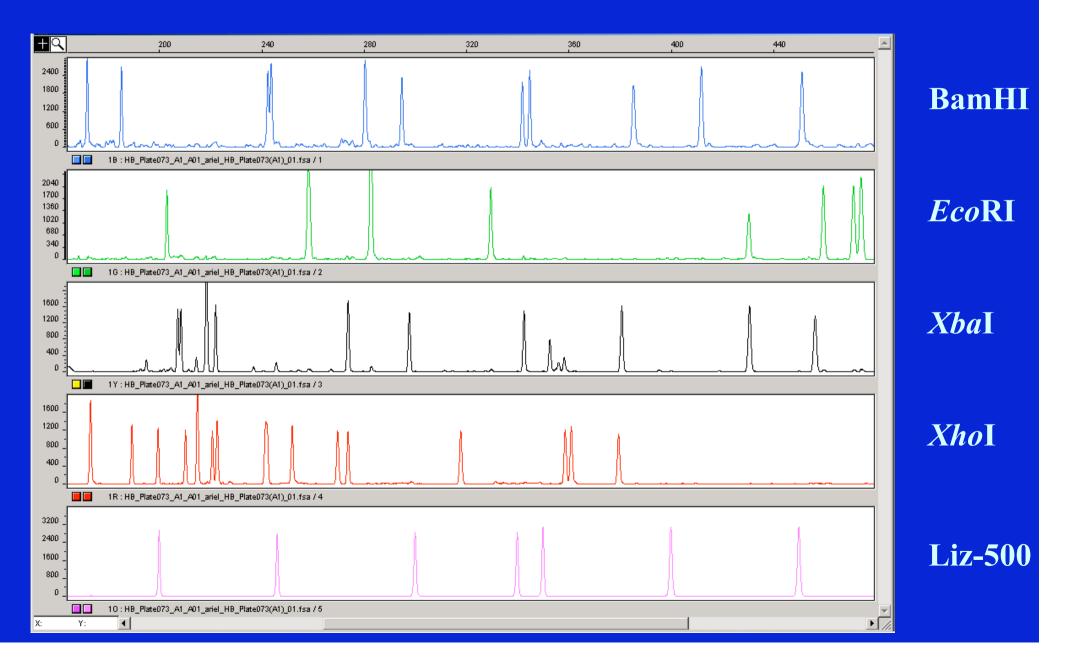
Predicted numbers of restriction fragments in SNaPshot fingerprints of two *Triticum monococcum* and two *T. turgidum* BACs in the range of 50-500 bp

Enzyme	116F2	115G1	BAC1	BAC2	Total
	(107.3 kb)	(128.6 kb)	(173.4 kb)	(147.6 kb)	(556.9 kb)
EcoRI	31	38	32	32	133
<i>Bam</i> HI	21	36	53	32	141
XbaI	31	47	38	41	157
XhoI	26	30	46	23	125
Total	108	151	168	128	
HindIII	43	51	68	77	239

Fragment sizing w/ ABI3730

- Any5Dye
- Denatured condition (Hi-Di, 95 °C, 5')
- 36 cm capillary array / 50 cm capillary array
- Liz-500 Size Standard / longer range

Portion of multi-color fingerprinting profile of a BAC clone



Fingerprinting throughput per single sequencer

Instrument	Daily	Weekly (7 days)	Monthly (28 days)	Annually (330 days)
ABI 3100	480	3,360	13,440	158,400
ABI 3730	2,160	15,120	60,480	712,800
ABI 3730XL	4,320	30,240	120,960	1,424,000

Success rate of the SNaPshot based fingerprinting procedure (ABI 3100)

Library	No. clones tried	No. clones succeeded	Success rate
BamHI	58,261	56,187	96.44%
BamHI BiBAC	19,730	19,455	98.61%
<i>Eco</i> RI	52,914	51,217	96.79%
HindIII	58,615	56,803	96.91%
HindIII BiBAC	26,125	25,907	99.17%
Total	215,645	209,569	97.18%

Fingerprints edit and management

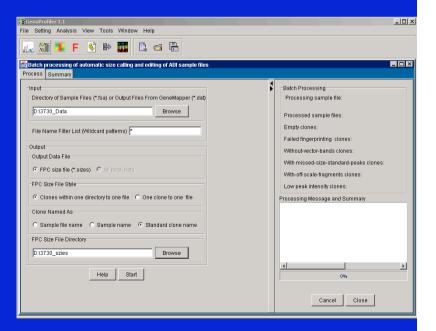
- Fragment size-calling
- **True fragments vs. background noises**
- Batch fingerprint editing
- BAC cross-contamination check and removal
- Fragment frequency analysis
- Dataset management (clones and fragments)
 - **BAC-marker hybridization data conversion**

😰 FP Pipeliner		🛞 GeneMapper - *N	IX_Pla	te049(B2)_2004-05-(4_DVORAK37	30_0344 -	gm is Lo	
File Help		File Edit Analysis Vi	ew To	ols Helj	0				
			P 00				Table Set	ting:	
		🖃 🚠 Project	Sample	es Gen	otypes				
		⊞ DMX_Plate04		Status	Sample File	Sample Name	Comments	Sample Type	Analy
			50		MX_Plate049_H24	MX_Plate049_H24	None	Sample	Analy
			51		MX_Plate049_J02	MX_Plate049_J02	None	Sample	Analy
			52	<u> </u>	MX_Plate049_J04	MX_Plate049_J04	None	Sample	Analy
			53	İ –	MX_Plate049_J06	MX_Plate049_J06	None	Sample	Analy
User name: gm			54	<u> </u>	MX_Plate049_J08	MX_Plate049_J08	None	Sample	Analy
Password: ******			55		MX_Plate049_J10	MX_Plate049_J10	None	Sample	Analy
			56	<u> </u>	MX_Plate049_J12	MX_Plate049_J12	None	Sample	Analy
Analysis method: Analysis1			57		MX_Plate049_J16	MX_Plate049_J16	None	Sample	Analy
Size standard: FP3730			58		MX_Plate049_J18	MX_Plate049_J18	None	Sample	Analy
	. 📕		59		MX_Plate049_J20	MX_Plate049_J20	None	Sample	Analy
Sample file directories: E:\data_1 Add			60		MX_Plate049_J22	MX_Plate049_J22	None	Sample	Analy
			61		MX_Plate049_J24	MX_Plate049_J24	None	Sample	Analy
Delete			62		MX_Plate049_L02	MX_Plate049_L02	None	Sample	Analy
			63		MX_Plate049_L04	MX_Plate049_L04	None	Sample	Analy
			64		MX_Plate049_L06	MX_Plate049_L06	None	Sample	Analy
			65		MX_Plate049_L08	MX_Plate049_L08	None	Sample	Analy
Export contents: Export plot sample table			66		MX_Plate049_L10	MX_Plate049_L10	None	Sample	Analy
								Sample	Analy
Export file directory: E:\size_1 Browse			68	1		MX_Plate049_L14		Sample	Analy
Export file extension: 🔘 .gm 💿 .dat 🔘 .txt			69	<u> </u>	MX_Plate049_L16	MX_Plate049_L16	None	Sample	Analy
			70	<u> </u>		MX_Plate049_L18		Sample	Analy
(Note: Red items are required.)			71	N	ļ	MX_Plate049_L20		Sample	Analy
			72	N		MX_Plate049_L22		Sample	Analy
Progress:			73	1		MX_Plate049_L24		Sample	Analy
			74	<u> </u>	ļ	MX_Plate049_N02		Sample	Analy
			75	<u></u>		MX_Plate049_N04	<u> </u>	Sample	Analy
			76	N		MX_Plate049_N06		Sample	Analy
			77	1		MX_Plate049_N08		Sample	Analy
			78	1	ļ	MX_Plate049_N10		Sample	Analy
			79	<u></u>		MX_Plate049_N12		Sample	Analy
			80	1	ļ	MX_Plate049_N14		Sample	Analy
		< >	81	()) <	MX_Plate049_N16	MX_Plate049_N16	None	Sample	Analy 🗡
Start Exit Help						_	700		
		Analyzing Samples					70%		Stop

FP Pipeliner (www.bioinforsoft.com) GeneMapper (www.appliedbiosystems.com)

Fingerprints Edit

GenoProfiler (wheatdb.ucdavis.edu:8080/wheatdb



e Analysis Manual Edit	ng Pipeline View Tools Help		
3	(4) 🚯 📾 👬 🕨 🛄 🔜 📰 🔜 👘 🚱 🛞 🔞 🕲 🚯		
Projects	🗐 Samples 🏄 Manual Editing 🕼 Fragment Statistics 🔛 Fragment Frequency 🔢 Fingerprint Patterns		
💼 22 🛅 AT01 A	zoom: 📉 🖪 🖸 🖌 🖪 💿 📖 🖽 🖽 🖽 🗐 🛨 🎕 🇇 🗐	Selected SMQS 0.94 FEQS 218.9	
	X490		
		420 450 480 510 540 570 600	
	2800 -		
	2400 -		
		· · · · · · · · · · · · · · · · · · ·	
		- <u> </u>	
	o فارتجا المالية المالية المراجعة المالية المالية المنظمية المنظمة المنطقة ا	March	
	Processed traces Raw traces		
	All Fragments Selected Fragments	No. of all/selected fragments	
	Dye/Peak Size Data Point Peak Height Peak Area Area/Height Selected Index		
	1 31 344.11 5668 128 1391.0 10.87 131	549	
	Image: 132 348.34 5694 66 551.0 8.35 132 Image: 133 347.87 5712 1592 17659.0 11.09 ✓ 133	Total	
	134 351.15 5750 71 1043.0 14.69 134	47	Fingerprints Rdi
	1 36 354.69 5790 26 222.0 8.54 136		Fingerprints Edi
	Image: 137 367.01 5816 94 1086.0 11.55 137 Image: 138 359.96 5849 46 511.0 11.11 138	173	
	■ 139 368.2 5940 47 484.0 9.87 139	Blue	
	Image: 140 371.95 5981 192 2715.0 14.14 140 Image: 141 381.3 6082 159 2153.0 13.54 141		
	142 382.51 6095 140 1395.0 9.96 142		
	Image: 143 384.57 6117 83 1197.0 14.42 143 Image: 144 386.83 6141 78 1090.0 13.97 144	157	
	145 388.72 6161 34 331.0 9.74 146	Green 11	
	Image: 146 390.05 6175 42 394.0 9.38 146 Image: 147 393.47 6211 55 1206.0 21.93 147		
	148 396.82 6248 29 418.0 14.41 148		
	■ 143 399.9 6278 1875 22487.0 11.99 ✓ 149 ■ 150 408.84 6371 80 921.0 11.51 150	Yellow	
	151 412.41 6408 93 1449.0 15.58 151	18	
	Image: 152 415.41 6439 30 238.0 7.93 152 Image: 153 417.06 6466 55 894.0 16.25 153	P	
	■ 154 421.91 6506 2193 36702.0 16.74 V 154	102	
	155 427.37 6662 286 5086.0 19.19 155 156 435.2 6642 56 1046.0 18.68 156	Red	
	1 57 438.87 6659 150 1799.0 11.99 157	11	
	Image: 158 439.03 6681 45 569.0 12.42 158 Image: 159 440.91 6700 85 905.0 10.65 159		
	160 443.27 6724 43 476.0 11.07 160	16 All fragments	
	161 445,35 6745 54 867.0 12.35 161 162 447.62 6768 32 372.0 11.62 162	Orange	FP Miner
		14 Selected fragments	

http://www.bioinforsoft.com/

GenoProfiler

To edit 100,000 BAC fingerprints:

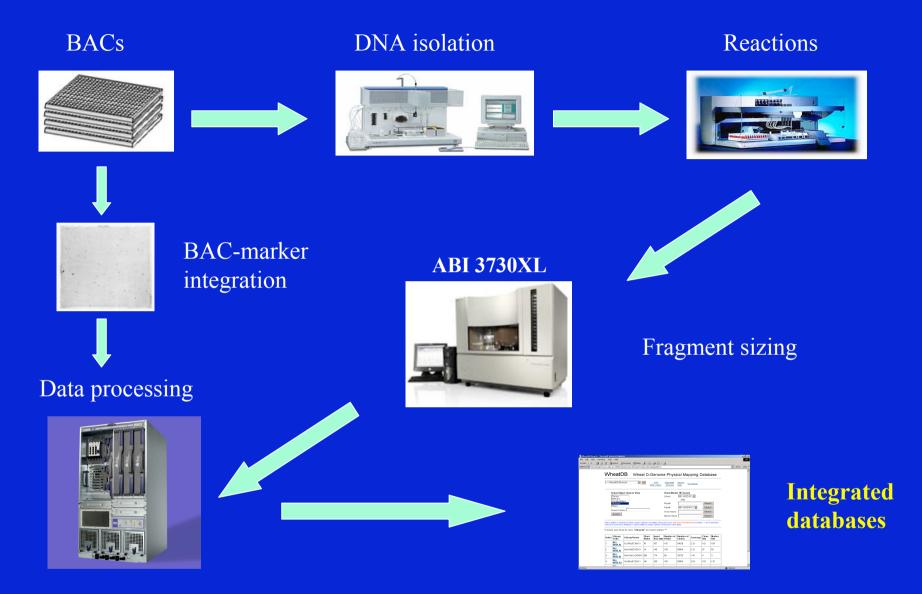
- input -- GeneMapper exported text file
 ≈ 5 min.
- input -- 3100/3700 samples files (.fsa)
 ~ 4.5 hr.

Computer: 3.2 GHz CPU, 1.0 GB RAM

http://wheatdb.ucdavis.edu:8080/wheatdb/

A sample contig (≈ 3 Mb)

\times	FPC Ctg6339 D_V7_2004			
Fil	e Edit Analysis Highlight Add-track Layout	Size options		Help
- Ctę	Zoom 2.3 Whole 6339 of D_V7_2004 Ves		CB Unit Range Con 0 to 1540	tig stats Clones: 208 (43 buried) Markers: 4 Sequenced: 0
				Length: 1541 CB units
• * *	BF201181	BE424164		BG262193
		HD101E24*		HD081C2*
	<u>HB121A2</u>			HB027B4
		HB073I21		124H6
		<u></u>		55P16
		<u></u>	<u>RI095B22</u> HB089J1 <u>9</u>	
		HB017H11 [*]	HI125015	
		HD147B7	HI153G16	-
		<u>3011</u>	HD142H3	
	HI150K20	_	RI120B22	
	<u>BB079N16</u>		HD052B14 HD145D3*	-
	<u></u>		HD143D3 HD081I10	
	HI113K14		HD124D7	
	HB080J7*	_	HD123I16	
	HD003J23		HB092B15	HI136G12
	BB044L20		RI061N21	HD055E2
	<u>HI126J1</u> HB089O3	—	RI044I22 RI012F18*	RI066D17 HI070O9
	HI058J8		HB041N1	
	RI011E7	BIO	70E15	HD030B2
	HD066M11		04N8	HI074C20*
	<u>HI014A1</u>	<u>HI089E</u>	9	<u>HD135K11*</u>
	HI028I15*	HI083E21		<u>RI039J9</u>
		,	Merge 4e-24 RI044I22 Merge 4e-24 HD124D7	
	Merge 4e-23 HI058J8 Merge 4e-23 HI126J1			▼
			750 800 <u>.</u> . 8	50
4		111		*



Contig assembling

http://wheatdb.ucdavis.edu: 8080/wheatdb Luo *et al. Genomics* 82:378-389, 2003

Current Applications

Wheat	
	Catfish
Rice	Rainbow Trout
Barley	Kambow Hout
Soybean	•••
Soybean	
Citrus	
Sorghum	
Brassica	
Grape	
Tomato	

•••

Summary

- High information contents
- Accurate fragment sizing
- Simple procedure, high reproducibility
- Relatively inexpensive
- Automated pipeline, high-throughput

Acknowledgements

Jan Dvorak Carolyn Thomas Patrick E. McGuire

Olin D. Anderson Frank M. You

C. Robin Buell Shu Ouyang Jo Hsiao



NSF/Plant Genome Research Program

"Mapping and sequencing large genome: Let's get physical!"

_ Meyers et al. (2004) Nat. Rev. Genet. 5: 578