1-ii) Data mining from model genome database for application to functional genetic diversity research: A simplified approach

Kazuo N. Watanabe, Ph.D.
Professor, Gene Research Center
University of Tsukuba, Japan

Synopsis

The databases from the model organisms associated with bioiformatics could facilitate the initiation of research on under-invested and under exploited species. Using the information on orthologous super gene families such as cytochrome P450 as examples, a diverse plant species can be immediately studied on their genetic diversity for the assessment of genetic erosion and potential values in uses.

Background information-1)

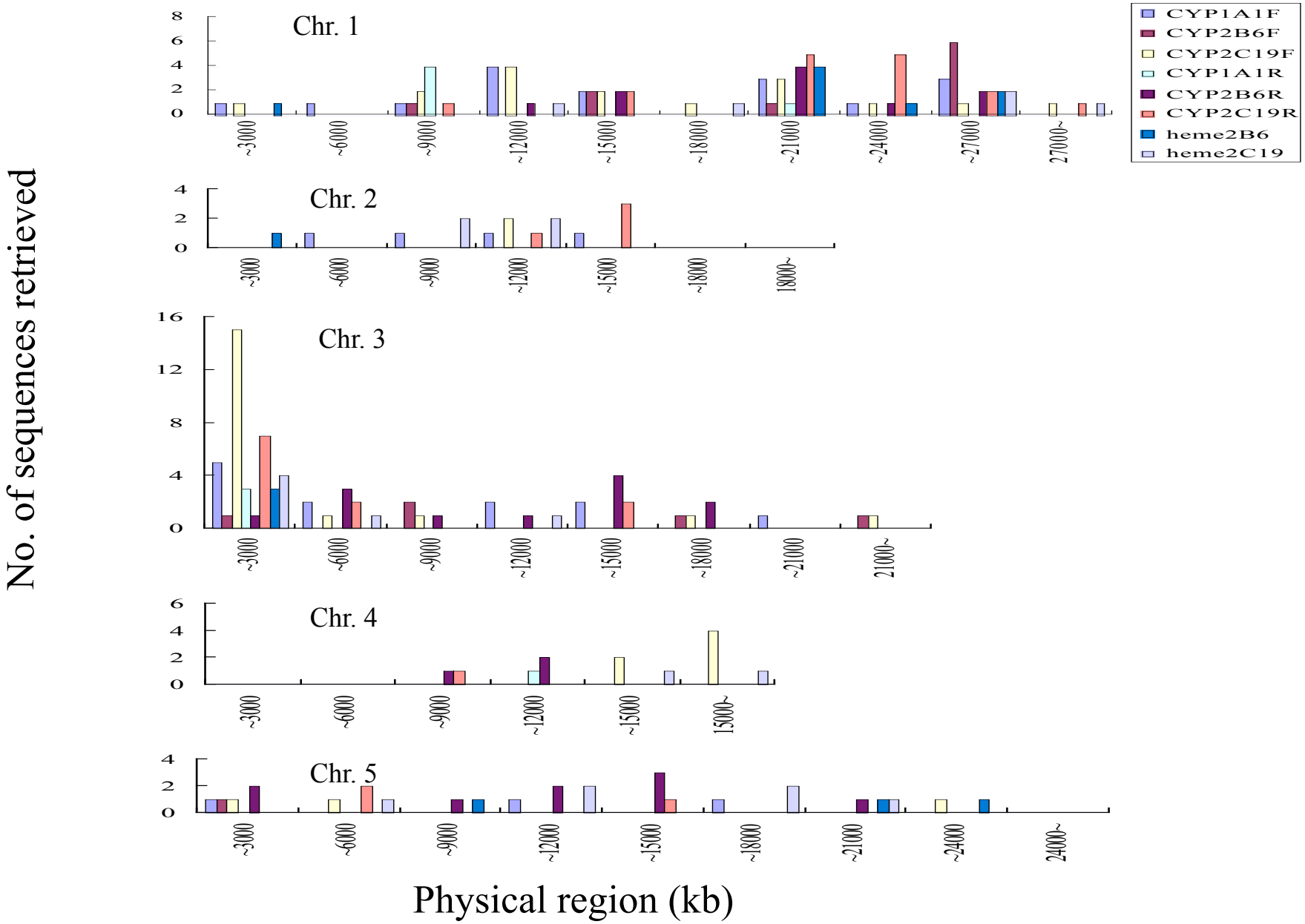
To investigate and develop new genetic tools for assessing genome-wide diversity in higher plant species, polymorphisms of gene analogues of mammalian cytochrome P450 monooxygenases were studied. Data mining of *Arabidopsis thaliana* indicated that a small number of primer-sets derived from P450 genes could provide universal tools for the assessment of genome-wide genetic diversity in diverse plant species that do not have relevant genetic markers or for which there is no prior knowledge of inheritance traits.

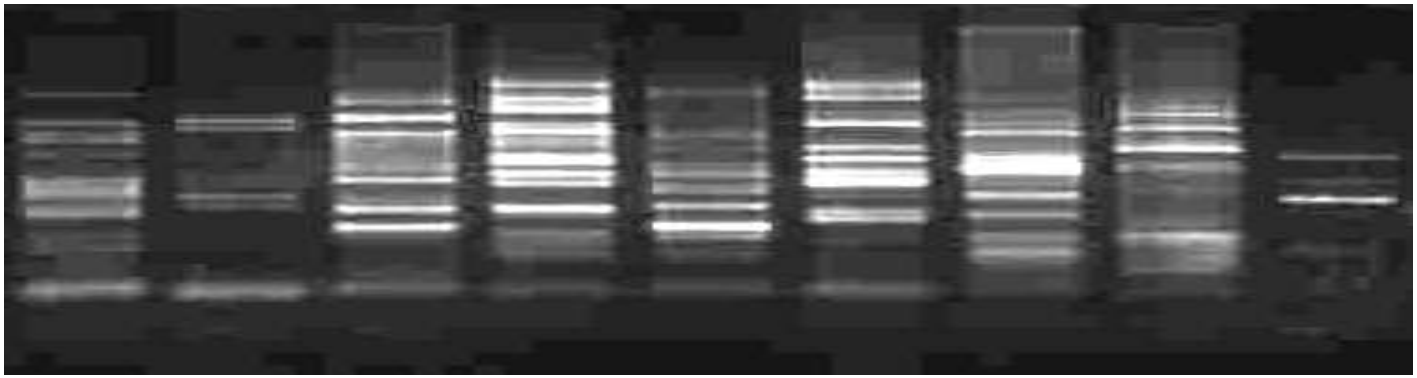
Table 2 Sequence information of PCR primers constructed based on the known P450 genes

Primer	Sequence (5' to 3')	Reference
CYP1A1F	GCC AAG CTT TCT AAC AAT GC	Inui et al. (2000)
CYP2B6F	GAC TCT TGC TAC TCC TGG TT	Inui et al. (2000)
CYP2C19F	TCC TTG TGC TCT GTC TCT CA	Inui et al. (2000)
CYP1A1R	AAG GAC ATG CTC TGA CCA TT	Inui et al. (2000)
CYP2B6R	CGA ATA CAG AGC TGA TGA GT	Inui et al. (2000)
CYP2C19R	CCA TCG ATT CTT GGT GTT CT	Inui et al. (2000)
heme2B6	ACC AAG ACA AAT CCG CTT CCC	Kiyokawa et al. (1997)
heme2C19	TCC CAC ACA AAT CCG TTT TCC	Kiyokawa et al. (1997)

Table 3 Annealing temperatures of each primer-set and total number of fragments scored from 51 species representing 28 families

Primer-set (Forward/Reverse)	Ann. Temp.	No. of fragments
CYP1A1F/CYP1A1R	56.0	63
CYP1A1F/CYP2B6R	52.0	63
CYP1A1F/CYP2C19R	46.5	63
CYP1A1F/heme2B6	56.0	58
CYP1A1F/heme2C19	56.0	42
CYP2B6F/CYP1A1R	52.0	57
CYP2B6F/CYP2B6R	52.0	43
CYP2B6F/CYP2C19R	46.5	41
CYP2B6F/heme2B6	52.0	42
CYP2B6F/heme2C19	52.0	60
CYP2C19F/CYP1A1R	56.0	59
CYP2C19F/CYP2B6R	52.0	68
CYP2C19F/CYP2C19R	46.5	61
CYP2C19F/heme2B6	56.0	58
CYP2C19F/heme2C19	56.0	57


No. of sequences retrieved
Chr. 1
Chr. 2
Chr. 3
Chr. 4
Chr. 5
CYP1A1F
CYP2B6F
CYP2C19F
CYP1A1R
CYP2B6R
CYP2C19R
heme2B6
heme2C19
~3000
~6000
~9000
~12000
~15000
~18000
~21000
~24000
~27000
27000~
18000~
21000~
15000~
24000~
Physical region (kb)



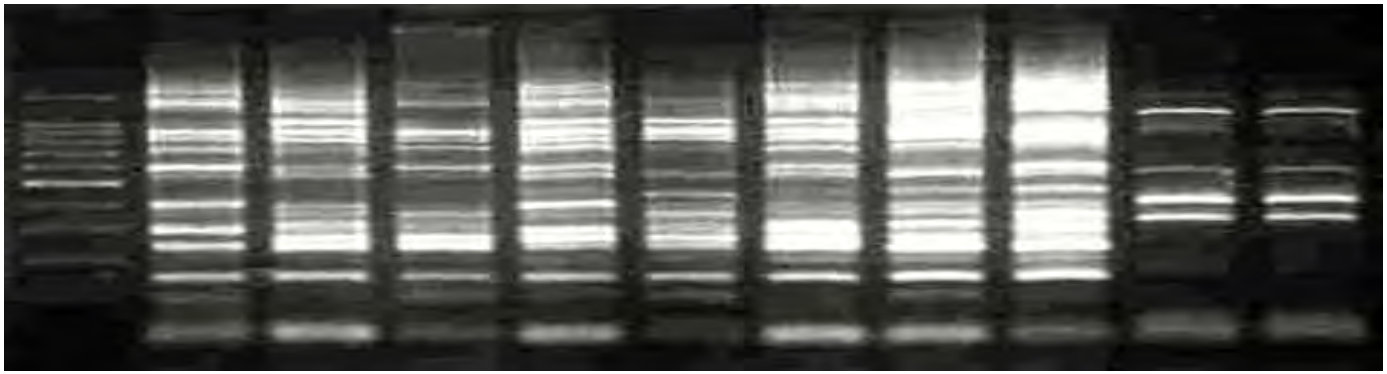
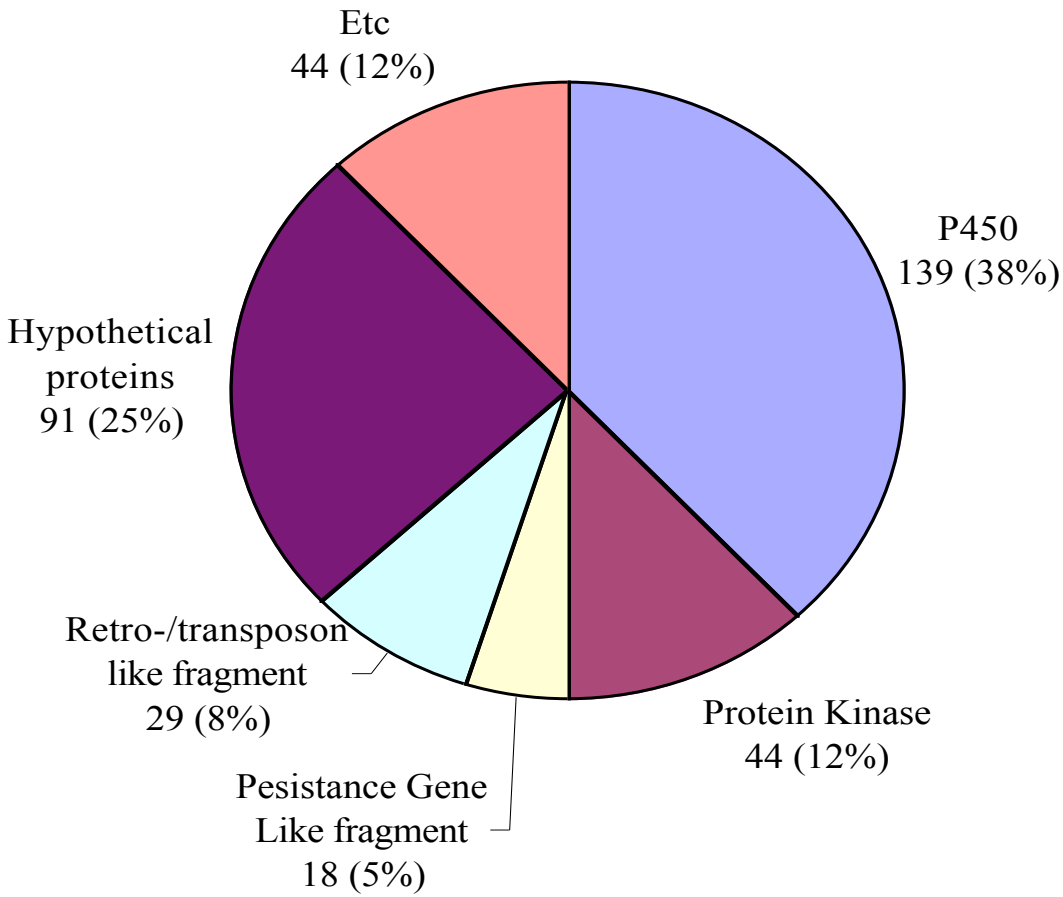


Table 4 Number of PCR fragments observed in *Solanaceae* and frequencies of intra-specific polymorphisms using mammalian P450 associated primers

Primer-set	Total fragments[1] (n=21)	Frequency of polymorphic fragments (%)			
		C. annuum (n=4)	*L. esculentum* (n=6)	*S. tuberosum* (n=8)	*S. melongena* (n=3)
CYP1A1F/CYP1A1R	38	22/29 (76)	22/23 (96)	24/25 (96)	18/21 (86)
CYP1A1F/CYP2B6R	33	16/18 (89)	26/26 (100)	19/19 (100)	8/8 (100)
CYP1A1F/CYP2C19R	35	12/17 (71)	15/17 (88)	22/22 (100)	17/19 (89)
CYP1A1F/heme2B6	45	21/23 (91)	26/26 (100)	38/38 (100)	18/24 (75)
CYP1A1F/heme2C19	32	15/18 (83)	20/20 (100)	14/19 (74)	21/25 (84)
CYP2B6F/CYP1A1R	32	17/19 (89)	22/24 (92)	21/24 (88)	13/18 (72)
CYP2B6F/CYP2B6R	36	23/26 (88)	23/23 (100)	26/26 (100)	10/15 (67)
CYP2B6F/CYP2C19R	26	13/18 (72)	22/22 (100)	15/15 (100)	8/8 (100)
CYP2B6F/heme2B6	37	23/24 (96)	25/25 (100)	15/20 (75)	6/14 (43)
CYP2B6F/heme2C19	36	15/15 (100)	23/23 (100)	22/23 (96)	13/13 (100)
CYP2C19F/CYP1A1R	44	31/32 (97)	32/32 (100)	35/35 (100)	5/11 (45)
CYP2C19F/CYP2B6R	56	26/29 (90)	34/35 (97)	26/27 (96)	17/20 (85)
CYP2C19F/CYP2C19R	48	15/18 (83)	16/18 (89)	27/29 (93)	13/21 (62)
CYP2C19F/heme2B6	37	19/23 (83)	31/31 (100)	21/25 (84)	7/14 (50)
CYP2C19F/heme2C19	36	21/21 (100)	20/20 (100)	29/29 (100)	4/10 (40)


Etc
44 (12%)
P450
139 (38%)
Hypothetical
proteins
91 (25%)
Retro-/transposon
like fragment
29 (8%)
Protein Kinase
44 (12%)
Pesistance Gene
Like fragment
18 (5%)

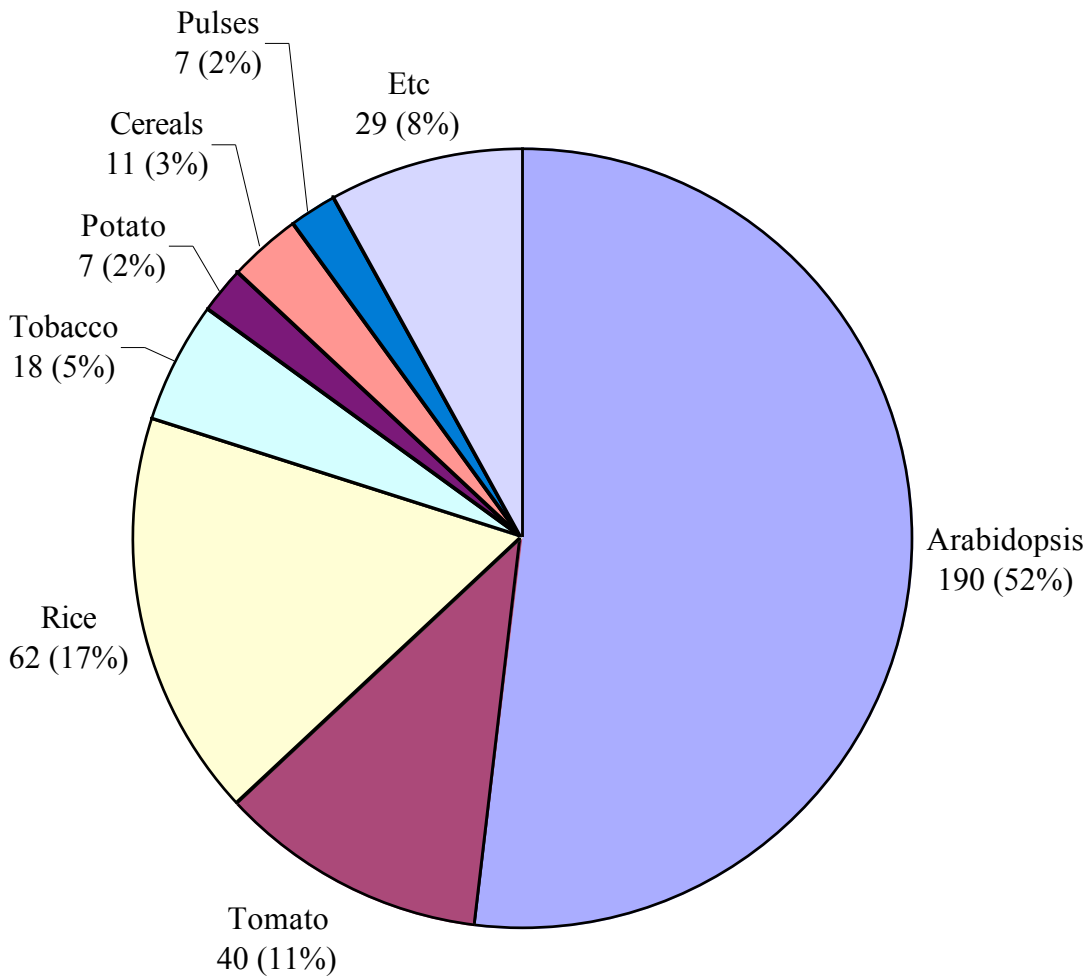

Pulses
7 (2%)
Etc
29 (8%)
Cereals
11 (3%)
Potato
7 (2%)
Tobacco
18 (5%)
Arabidopsis
190 (52%)
Rice
62 (17%)
Tomato
40 (11%)

Table 5. A representative summary of the association survey between tomato PCR products amplified with PBA primer pairs and plant P450 genes in *Arabidopsis* that showed high homology (> 70 %) to the fragments.

ication of nced tomato clone	PBA primer pair that amplified tomato clone in the left column	Plant CYPs Protein deduced from sequence	Putative activity of the plant CYP	Tomato sequence in *Arabidopsis* Chromosome
A1FhB6-1	CYP1A1F/ heme 2B6	CYP51A2	Obtusifoliol 14alpha-demrthlase	I
A1FhC19-3	CYP1A1F/ heme 2C19	CYP86A1	Omega-hydroxylase for saturated and unsaturated C12 to C18 fatty acids	V
A1FhC19-3	CYP1A1F/ heme 2C19	CYP86A8	Omega-hydroxylase for saturated and unsaturated C12 to C18 fatty acids	II
o-B6FB6R-2	CYP2B6F/ CYP2B6R	CYP90A1	23 alpha-hydroxylase for 6-oxo-cathasterone	V
o-B6FB6R-2	CYP2B6F/ CYP2B6R	CYP90B1	22s alpha-hydroxylase for 6-oxo-cathasterone	III
o-B6FhB6-7	CYP2B6F/heme 2B6	79A2	N-hydroxylase for phenylalanine	V
o-B6FhB6-7	CYP2B6F/heme 2B6	79B2	N-hydroxylase for tryptophan, tryptophan analogues	IV
o-B6FhB6-7	CYP2B6F/heme 2B6	79B3	N-hydroxylase for tryptophan	II
36FhC19-1	CYP2B6F/heme 2C19	83A1	N-hydroxylase for indole-3-acetyldoxime	IV
36FhC19-1	CYP2B6F/heme 2C19	83B1	N-hydroxylase for indole-3-acetyldoxime	IV
219FC19R-2	CYP2C19F/CYP2C19R	CYP88A3	Multifunctional entkaurenoic acid oxidase	I
219FC19R-2	CYP2C19F/CYP2C19R	CYP88A4	Multifunctional entkaurenoic acid oxidase	II
219FC19R-2	CYP2C19F/CYP2C19R	CYP701A3	Multifunctional entkaurenoic acid oxidase	V
o-C19FhB6-5	CYP2C19F/heme 2B6	75B1	3'-hydroxylase for narigenin	V
219FhC19-3	CYP2C19F/heme 2C19	CYP72B1	26-hydroxolase for brassinolide	II

Summary

These amplified fragments possessed homologies to other genes and proteins in different plant varieties. We conclude that the sequence diversity of P450 gene-analogues in different plant species reflects the diversity of functional regions in the plant genome and is therefore an effective tool in functional genomic studies of plants.

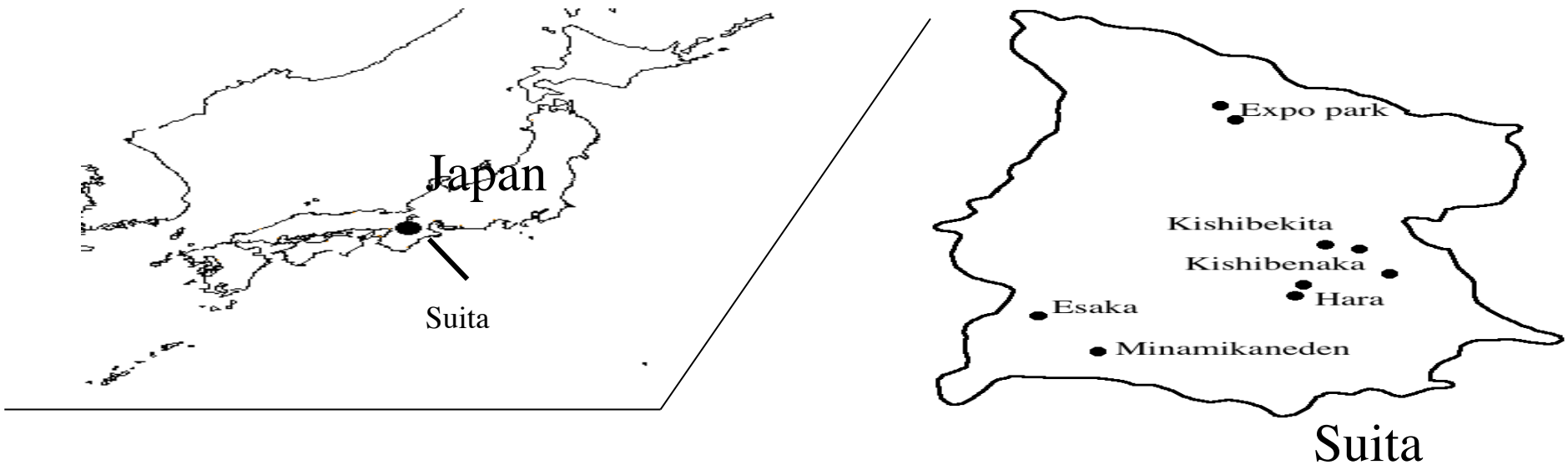


Fig. 1. Sampling sites and accessions of Suita Kuwai, *Sagittaria trifolia* var. *typica* Makino forma *suitensis* Makino

Fig. 2. Mixed samples of Suita Kuwai, *Sagittaria trifolia* L. var.*typica* Makino *suitensis* Makino. Scale in cm

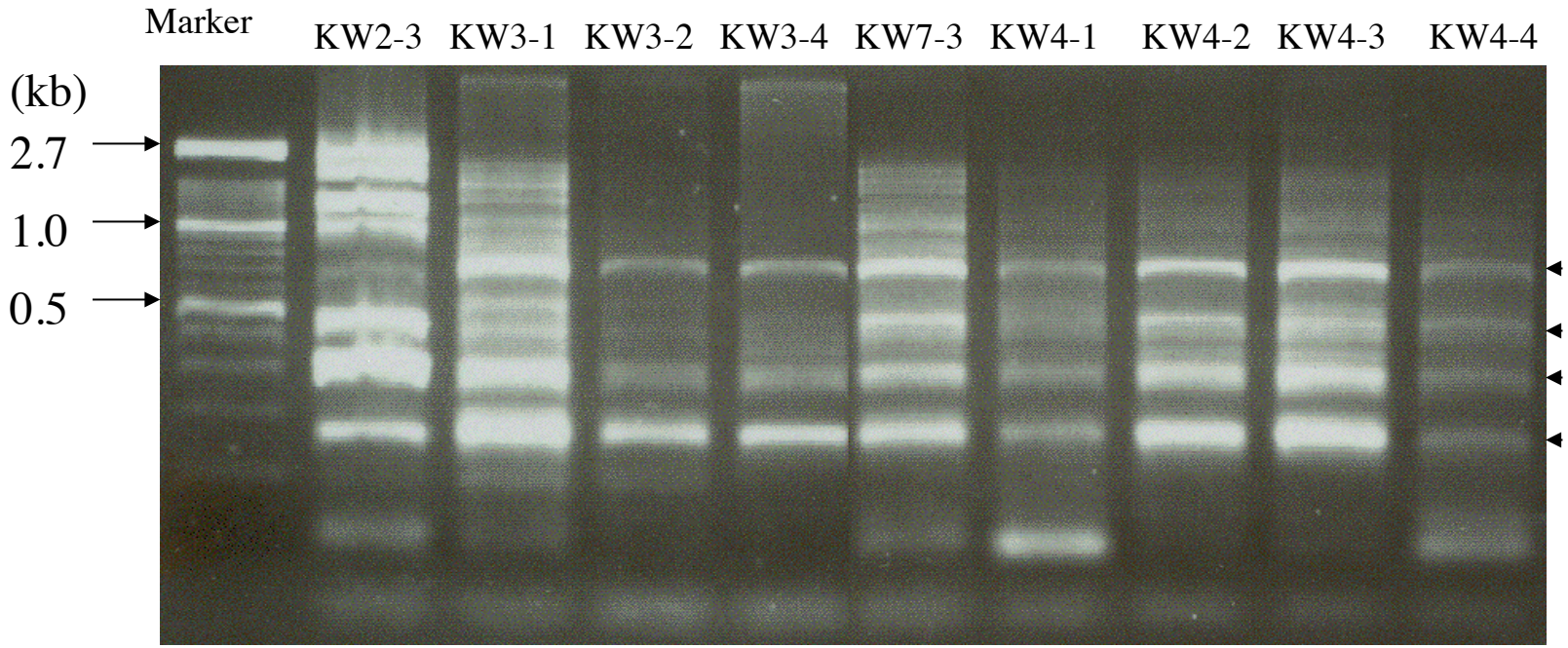


Fig. 3. Common and unique PCR products using cytochrome P450 associated primers (CYP2B6F-heme2B6) on *S. trifolia*. var. *typica* Makino forma *suitensis* Makino. Arrow indicate the common bands for the taxon.

Annealing temperature for each cytochrome P450 analogue (PBA) generating primer set and frequencies of polymorphic bands and accessions on Suita Kuwai, *Sagittaria trifolia* var. *typica*Makino forma*suitensis* Makino

				Polymorphic bands	
Primer pair	Annealing temperature (^{O}C)	Number of accessions w/ polymorphism	Total bands	Within accession (%)	Among accessions (%)
CYP1A1F-CYP1A1R	56.1	2	12	1 (8.3)	11(91.7)
CYP1A1F-CYP2B6R	51.9	1	16	3(18.8)	10(62.5)
CYP1A1F-CYP2C19R	46.5	2	13	2(15.4)	12(92.3)
CYP1A1F-heme2B6	56.3	3	14	3(21.4)	10(71.4)
CYP1A1F-heme2C19	56.3	2	14	2(14.3)	9(64.3)
CYP2B6F-CYP1A1R	51.9	3	27	3(11.1)	23(85.2)
CYP2B6F CYP2B6R	46.5	1	9	1(11.1)	6(66.7)
CYP2B6F-CYP2C19R	52.5	2	14	3(21.4)	12(85.7)
CYP2B6F-heme2B6	52.5	1	20	8(40.0)	16(80.0)
CYP2B6F-heme2C19	52.5	2	16	3(18.1)	14(87.5)
CYP2C19F-CYP1A1R	51.9	2	13	1(7.7)	8(61.5)
CYP2C19F-CYP2B6R	46.5	2	16	2(12.5)	11(68.8)
CYP2C19F-CYP2C19R	56.1	2	14	3(21.4)	10(71.4)
CYP2C19F-heme2B6	56.4	1	12	2(16.7)	11(91.7)
CYP2C19F-heme2C19	56.4	1	11	1(9.1)	6(54.5)
Total			221	32(14.5)	177(80.1)

Use of molecular markers for resistance breeding in potato.XII. Application of PBA and RGL markers. (Presented at 103th Meeting of Jpn Soc. Breed., Sept., 2003)

Yamanaka, S., S. Ikeda, A. Imai, Y.S. Luan, J.A. Watanabe and K.N. Watanabe
Gene Research Center, University of Tsukuba
Japan

Potato genetic map

From the first report of Bonierbale *et al*. (1988), there are several genetic maps reported using different marker systems such as RFLP (Gebhardt *et al*. 1989, Tanksley *et al*. 1992), AFLP (van Eck *et al*. 1995, Meyer *et al*. 1998) and SSR (Milbourne *et al*. 1998).

Due to using different marker systems in different populations, however, integration of above maps have not conducted yet.
It is difficult to utilize these information directly at once.

↓

To bridge among these information, genetic mapping using diverse marker systems will be required.

PBA marker

(P450 based analogues; Yamanaka *et al*. TAG 108: 1-9, 2003)

Newly developed markers based on functional genetic information

Detect polymorphism in diverse plant species

Evaluate genome-wide diversity

PCR-based and small number of primer-sets

RFLP, RAPD
SSR, ISSR, etc.

Biased distribution in the genome

↓

Sometimes, difficult to precise evaluation

Characterization of plant genetic resources
Assessment of genetic diversity

Wide distribution in the genome (eg. *Arabidopsis*, Rice)
Evaluation the diversity from the view point of functional genomics
SNP supports variability of functional regions

PBA, RGL, etc.

Suzuki *et al*. (2001), Tanaka *et al*. (2001), Yamanaka *et al*. (2003)

Objectives

Estimation of the distribution of functional markers such as PBAs, RGLs (Resistant gene-like fragments) in diploid potato genome

Integration of the information based on different marker systems

Are these functional markers representative for the diversity over the potato genome?

Materials and methods

Plant materials

A diploid population of Potato (*Solanum tuberosum*)
152 F_1 individuals derived from crossing between
86.61.26 and 84.194.30 (Watanabe *et al*. 1994)

Mapping

Construction the map based on pseudo-testcross
using MapMaker/EXP

Molecular markers used in this study

PBA	(Yamanaka *et al*. TAG 2003)
RGL	(Leister *et al*. Nat. Genet. 1996, modified by Hämäläinen *et al*. TAG 1998, Watanabe *et al*. Breed Sci. 2003)
SSR	(Milbourne *et al*. MGG 1998)
CAPS	(Chen *et al*. TAG 2001)
RFLP	(Tanksley *et al*. Genetics 1992)
RFLP-STS	(constructed in this study)
RAPD	(Williams *et al*. Nucl. Acid. Res. 1990)
AFLP	(Vos *et al*. Nucl. Acid. Res. 1995)

Table Number of markers used for parental polymorphism survey

Marker	No. tested
PBA	15
RGL	3
SSR	69
CAPS	45
RFLP	33
RFLP (STS)	88
RAPD	94
AFLP	4
Total	352

Table Efficiency for detection of polymorphic loci in different marker systems

Marker	No. of tested markers (A)	Markers with polymorphism	No. of polymorphic loci (B)	Frequency (B/A)
SSR	69	14	14	0.20
RFLP	33	12	12	0.36
RFLP-STS	88	15	15	0.17
CAPS	45	3	3	0.07
RAPD	94	33	42	0.47
AFLP	4	4	6	1.50
PBA	15	15	27	1.80
RGL	3	1	2	0.67
Total	352	97	121	

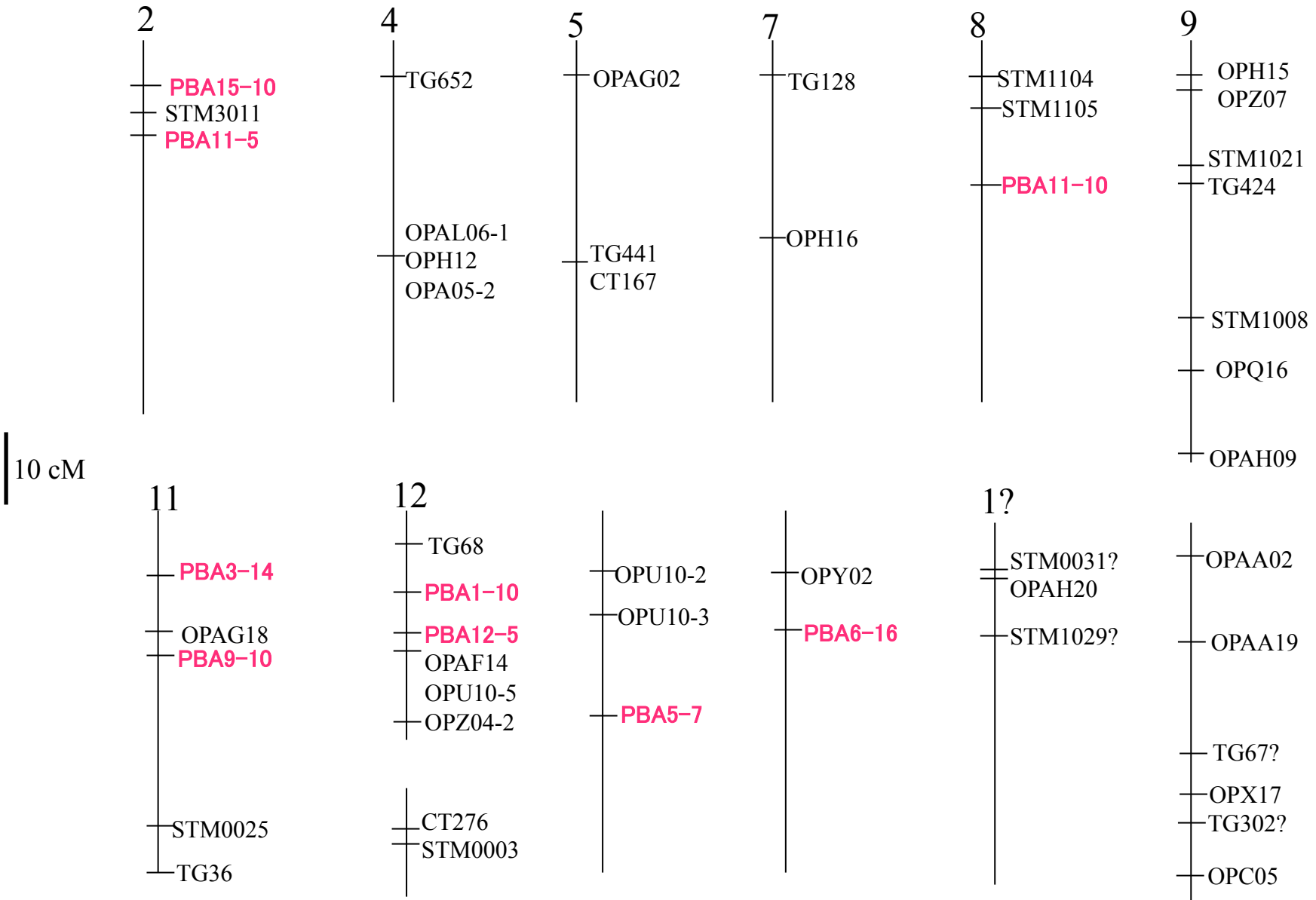


Fig. Linkage map based on polymorphisms in 86.61.26

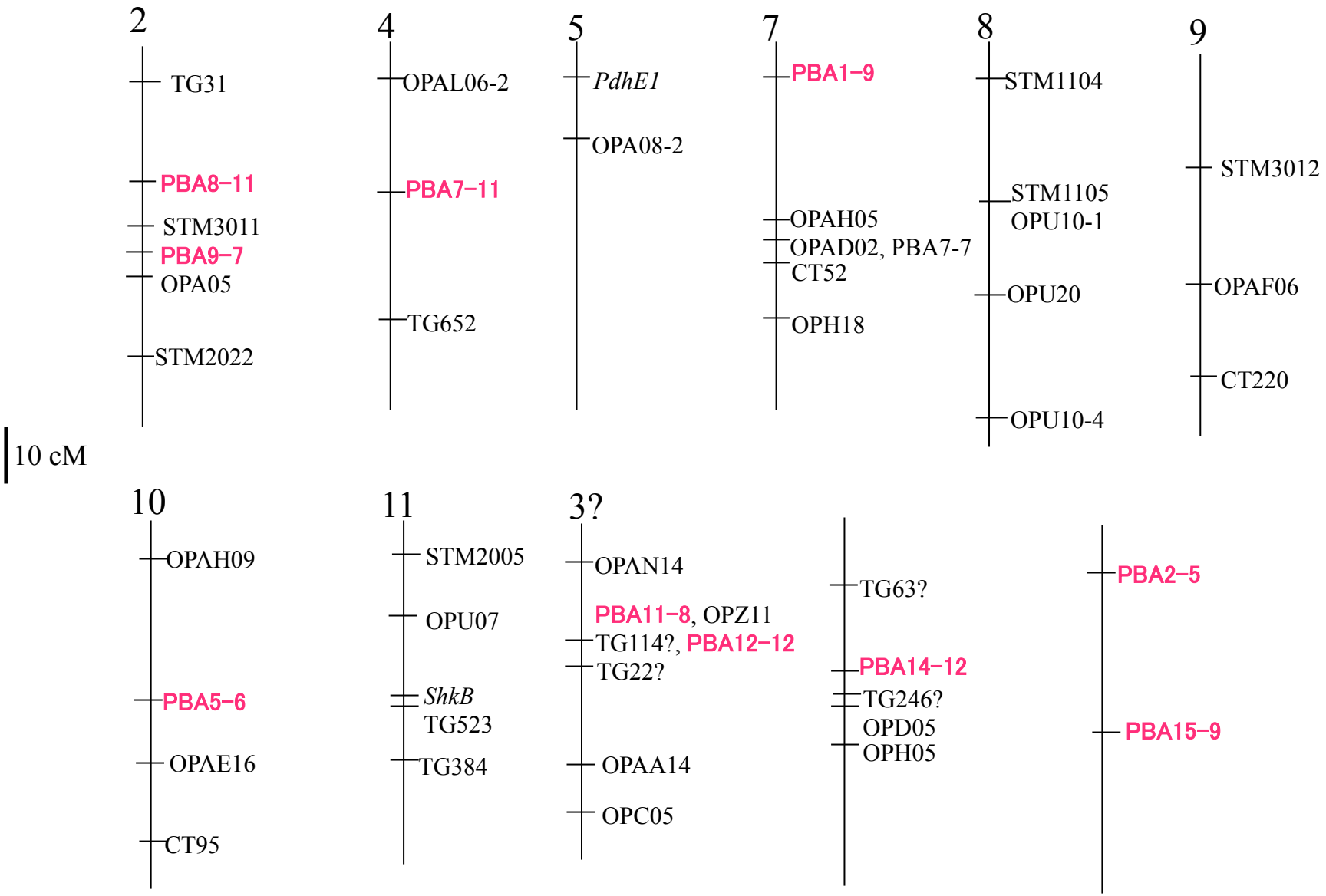


Fig. Linkage map based on polymorphisms in 84.194.30

Summary

PBAs are widely distributed over the potato genome, suggesting that these PBAs could reflect the genome-wide diversity in potato.

To fill the gap of map, resistance gene-like fragment (RGL/RGA) primer sets can facilitate the process.