

### Supplementary material:

**Formulas S1:** Formulas used to calculate three polymorphic information content values, percentage of heterozygous loci and optimum combination of polymorphic markers, UPIC values.

- Polymorphic information content [5], where  $p_i$  is the frequency of the  $i^{\text{th}}$  allele,  $j$  is the  $j^{\text{th}}$  line (DNA sample or taxonomic unit) and  $n$  is the number of alleles for the marker.

$$PIC=1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^n \sum_{j=i+1}^n 2p_i p_j^2 \quad \text{Formula 1.1.1}$$

- For the particular case of cross fertilized organisms that have equiprevalent alleles, the formula can be simplified [6]. Same variable notation as in formula 1.1.1.

$$PIC=1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^n p_i^4 \quad \text{Formula 1.1.2}$$

- In case of self fertilized organisms, or x-chromosome-linked markers in humans, this third term becomes “zero”, and the PIC value is identical to the heterozygosity of the marker [6]. Same variable notation as in formula 1.1.1.

$$PIC=1 - \sum_{i=1}^n p_i^2 \quad \text{Formula 1.1.3}$$

- Percentage of heterozygous loci, where  $H_j$  is the number of heterozygous loci for the  $j^{\text{th}}$  line (DNA sample),  $m$  is the total number of loci or markers, and  $M_j$  is the number of loci for which no alleles were detected (missing data) in the  $j^{\text{th}}$  line.

$$\% \text{ Heterozygous loci} = [H_j / m - M_j] \times 100 \quad \text{Formula 1.2.1}$$

#### UPIC calculation

Allele information of eight polymorphic markers ( $m_1, \dots, m_8$ ) that were run on 12 lines (DNA samples) ( $j_1, \dots, j_{12}$ ) was used in our example to show the mechanics of calculating unique-pattern informative combinations (UPIC). If we assign a letter to each pattern observed for a line (**Table S2b**) and then convert the letters to binary values, where “0” is assigned to an allele pattern present more than once across the lines tested, and “1” is assigned to unique patterns (UP) we obtain **Table S2c**. Let  $UP_{w,j}$  represent the unique pattern indicator for the  $w^{\text{th}}$  marker and the  $j^{\text{th}}$  line shown in **Table S2c**. The number of combinations of markers that can be formed is calculated by the formula below, where  $m$  is the number of polymorphic markers taken in groups of  $k$ , and  $k!$  is  $k$  factorial

$${}^m C_k = \frac{m(m-1)(m-2)\dots \text{to } k \text{ factors}}{k!}$$

Our Perl script that calculate UPIC values only includes possible combinations of  $m$  polymorphic markers taken in groups of  $k$ , for  $k = 2$  to  $k = m/2$ , as  $k = m$  for large values of  $m$  will be computationally intensive.

$${}^m C_k = {}^m C_2 + {}^m C_3 + {}^m C_4 + \dots + {}^m C_{m/2}$$

Let  $N$ =Number of possible combinations as defined above:

$$\text{for } m=8: \\ = {}^8 C_2 + {}^8 C_3 + {}^8 C_4 = \frac{8 \cdot 7}{2!} + \frac{8 \cdot 7 \cdot 6}{3!} + \frac{8 \cdot 7 \cdot 6 \cdot 5}{4!} = 28 + 56 + 70 = 154 \text{ possible combinations } {}^m C_k$$

For each combination of markers, **Table S2d** shows:

- Combined unique patterns ( $CUP_{i,j}$ ) that characterize each line calculated as follows:

$$CUP_{i,j} = \sum_{W^*} C_{w,j}$$

Where  $i = i^{\text{th}}$  set of markers,  $j = j^{\text{th}}$  line,  $W^*$  indicates sum over  $C_{w,j}$  (**Table S2c**) values for all marker in  $i^{\text{th}}$  set of markers.

- Number of Combined Unique Patterns  $NUP_i$

$$NUP_i = \sum_{j=1}^l CUP_{i,j}$$

- Informative combinations ( $IC_i$ ),

$IC_i = \text{True}$  if  $CUP_{i,j} \neq 0$  for all lines( $j$ ), meaning those marker combinations that allow unique identification of each of the lines tested.

Since various IC with different total number of UP can be found, our UPIC script output consists of two columns, one is the total number of UPIC (*i.e.*, 18, **Table S2d**) in the combination, and the other is the marker combination (*i.e.*,  $m_3, m_4, m_8$ ). All UP values of each IC, UPIC plot, for the data in our example are plotted in **Figure 1**.

**Table S1: INPUT FILE for both scripts, UPIC and PIC/Heterozygous loci**

I = Dye/Peak-order; II – Marker; III = Line or DNA sample; IV = amplicon size

I	II	III	IV	I	II	III	IV	I	II	III	IV
B,19	0094_a	Chn1	176	B,18	0015_a	Chn10	206	B,23	0284_a	Chn11	124
B,8	0094_a	Chn10	161	B,30	0015_a	Chn11	206	B,26	0284_a	Chn11	131
B,21	0094_a	Chn11	161	B,19	0015_a	Chn12	206	B,28	0284_a	Chn11	135
B,9	0094_a	Chn12	161	B,18	0015_a	Chn2	201	B,11	0284_a	Chn12	131
B,9	0094_a	Chn2	176	B,19	0015_a	Chn2	206	B,12	0284_a	Chn12	135
B,10	0094_a	Chn3	176	B,17	0015_a	Chn3	201	B,9	0284_a	Chn12	123
B,10	0094_a	Chn4	188	B,18	0015_a	Chn3	206	B,22	0284_a	Chn2	124
B,15	0094_a	Chn5	176	B,15	0015_a	Chn4	201	B,24	0284_a	Chn2	131
B,11	0094_a	Chn6	183	B,16	0015_a	Chn4	207	B,25	0284_a	Chn2	136
B,19	0094_a	Chn8	161	B,17	0015_a	Chn4	212	B,10	0284_a	Chn3	124
B,10	0094_a	Chn9	159	B,25	0015_a	Chn5	201	B,12	0284_a	Chn3	131
B,11	0094_a	Chn9	161	B,26	0015_a	Chn5	206	B,14	0284_a	Chn3	136
B,15	0094_a	Osn7	151	B,14	0015_a	Chn6	201	B,10	0284_a	Chn4	124
B,17	0094_a	Osn7	155	B,15	0015_a	Chn6	207	B,12	0284_a	Chn4	131
B,20	0094_a	Osn7	163	B,24	0015_a	Chn8	123	B,14	0284_a	Chn4	135
B,21	0094_a	Osn7	165	B,33	0015_a	Chn8	206	B,25	0284_a	Chn5	124
B,23	0094_a	Osn7	169	B,17	0015_a	Chn9	206	B,27	0284_a	Chn5	131
B,17	0114_a	Chn1	160	B,20	0015_a	Osn7	114	B,28	0284_a	Chn5	136
B,12	0114_a	Chn10	164	B,25	0015_a	Osn7	166	B,12	0284_a	Chn6	124
B,17	0114_a	Chn11	165	B,32	0015_a	Osn7	207	B,14	0284_a	Chn6	131
B,9	0114_a	Chn12	165	B,23	0076_a	Chn1	118	B,16	0284_a	Chn6	135
B,11	0114_a	Chn2	160	B,28	0076_a	Chn1	133	B,25	0284_a	Chn8	124
B,9	0114_a	Chn3	160	B,14	0076_a	Chn10	112	B,27	0284_a	Chn8	131
B,9	0114_a	Chn4	160	B,33	0076_a	Chn11	112	B,28	0284_a	Chn8	134
B,11	0114_a	Chn5	160	B,14	0076_a	Chn12	112	B,29	0284_a	Chn8	136
B,9	0114_a	Chn6	165	B,13	0076_a	Chn2	124	B,10	0284_a	Chn9	135
B,25	0114_a	Chn8	164	B,14	0076_a	Chn3	127	B,8	0284_a	Chn9	124
B,11	0114_a	Chn9	165	B,16	0076_a	Chn3	133	B,9	0284_a	Chn9	131
B,23	0114_a	Osn7	164	B,16	0076_a	Chn4	136	B,33	0284_a	Osn7	69
B,24	0114_a	Osn7	166	B,19	0076_a	Chn5	124	B,35	0284_a	Osn7	225
B,20	0124_a	Chn1	132	B,16	0076_a	Chn6	162	B,33	0350_a	Chn1	173
B,10	0124_a	Chn10	110	B,17	0076_a	Chn6	165	B,17	0350_a	Chn10	194
B,16	0124_a	Chn10	125	B,21	0076_a	Chn8	112	B,18	0350_a	Chn10	196
B,17	0124_a	Chn10	132	B,22	0076_a	Chn8	115	B,15	0350_a	Chn11	172
B,17	0124_a	Chn11	110	B,14	0076_a	Chn9	112	B,16	0350_a	Chn11	181
B,23	0124_a	Chn11	125	B,15	0076_a	Chn9	115	B,12	0350_a	Chn12	172
B,10	0124_a	Chn12	110	B,18	0076_a	Osn7	105	B,10	0350_a	Chn2	177
B,15	0124_a	Chn12	125	B,34	0194_a	Chn1	167	B,12	0350_a	Chn3	173
B,12	0124_a	Chn2	111	B,12	0194_a	Chn10	158	B,14	0350_a	Chn3	177
B,17	0124_a	Chn2	132	B,34	0194_a	Chn11	158	B,11	0350_a	Chn4	173
B,10	0124_a	Chn3	132	B,16	0194_a	Chn12	158	B,14	0350_a	Chn4	181
B,14	0124_a	Chn4	132	B,12	0194_a	Chn2	167	B,18	0350_a	Chn5	171
B,9	0124_a	Chn4	110	B,23	0194_a	Chn3	167	B,19	0350_a	Chn5	177
B,15	0124_a	Chn5	110	B,15	0194_a	Chn4	167	B,13	0350_a	Chn6	175
B,19	0124_a	Chn5	132	B,16	0194_a	Chn4	172	B,29	0350_a	Chn8	196
B,14	0124_a	Chn6	132	B,14	0194_a	Chn5	167	B,17	0350_a	Chn9	196
B,9	0124_a	Chn6	110	B,35	0194_a	Chn6	167	B,13	0350_a	Osn7	121
B,19	0124_a	Chn8	110	B,32	0194_a	Chn8	158	B,9	0284_a	Chn12	123
B,22	0124_a	Chn8	125	B,13	0194_a	Chn9	158	B,22	0284_a	Chn2	124
B,10	0124_a	Chn9	110	B,21	0194_a	Osn7	168	B,24	0284_a	Chn2	131
B,15	0124_a	Chn9	125	B,22	0194_a	Osn7	175	B,25	0284_a	Chn2	136
B,17	0124_a	Chn9	134	B,23	0284_a	Chn1	124	B,10	0284_a	Chn3	124
B,24	0124_a	Osn7	110	B,25	0284_a	Chn1	131	B,12	0284_a	Chn3	131
B,25	0124_a	Osn7	119	B,26	0284_a	Chn1	136	B,14	0284_a	Chn3	136
B,30	0124_a	Osn7	132	B,10	0284_a	Chn10	131	B,10	0284_a	Chn4	124
B,27	0015_a	Chn1	206	B,11	0284_a	Chn10	135	B,12	0284_a	Chn4	131

**Table S2: UPIC: Calculation of Unique Pattern Informative Combination of polymorphic markers.**

	Amplicon	Line 1	Line 2	Line 3	Line 4	Line 5	Line 6	Line 7	Line 8	Line 9	Line 10	Line 11	Line 12	
a	Marker 15	114											0.333	
		123						0.5						
		166											0.333	
		201		0.5	0.5	0.333	0.5	0.5						
		206	1	0.5	0.5		0.5		0.5	1	1	1	1	
		207				0.333		0.5						0.333
		212				0.333								

MARKERS	Marker order	PATTERNS: identical patterns within each marker are identified by the same letter.												
b	94	1	A	A	A	B	A	C	D	E	D	D	D	F
	114	2	A	A	A	A	A	B	C	B	C	B	B	D
	124	3	A	B	A	C	C	C	D	E	F	D	D	G
	15	4	A	B	B	C	B	D	E	A	A	A	A	F
	76	5	A	B	C	D	B	E	F	F	G	G	G	H
	194	6	A	A	A	B	C	C	D	D	D	D	D	E
	284	7	A	A	A	B	A	B	C	B	D	B	E	F
	350	8	A	B	C	D	E	F	G	G	H	I	J	K

MARKERS	Marker order	Only unique patterns are replaced by # "1", others by "zero"												
c	94	1	0	0	0	1	0	1	0	1	0	0	0	1
	114	2	0	0	0	0	0	0	0	0	0	0	0	1
	124	3	0	1	0	0	0	0	0	1	1	0	0	1
	15	4	0	0	0	1	0	1	1	0	0	0	0	1
	76	5	1	0	1	1	0	1	0	0	0	0	0	1
	194	6	0	0	0	1	0	0	0	0	0	0	0	1
	284	7	0	0	0	0	0	0	1	0	1	0	1	1
	350	8	1	1	1	1	1	1	0	0	1	1	1	1

Number of	UPIC	Marker combination	Examples of combinatorial addition of patterns. Non-zero additions are IC											
d	-	1+2	0	0	0	1	0	1	0	1	0	0	0	2
	-	1+7	0	0	0	1	0	1	1	1	1	0	1	2
	-	2+3	0	1	0	0	0	0	0	1	1	0	0	2
	-	2+6	0	0	0	1	0	0	0	0	0	0	0	2
	18	3+4+8	1	2	1	2	1	2	1	1	2	1	1	3
	23	3+4+5+8	2	2	2	3	1	3	1	1	2	1	1	4
	27	1+3+4+5+8	2	2	2	4	1	4	1	2	2	1	1	5

**Table S3: UPIC values, output of UPIC version 1.0.** The first column corresponds to Unique Patterns (UP) or alleles detected by the marker combination. Combinations of Markers are listed ONLY if they can discriminate ALL DNA samples, we define these combinations as "Informative Combinations" (IC), and their corresponding value UP is the number of patterns or alleles that the combination detects. The user can choose HOW MANY markers to run according to his/her budget, and then finds the combination with the highest UP value. Example of 8 polymorphic markers runs for 12 DNA samples. Note: the combination of markers 350, 15, 76 and 124 can discriminate the 12 DNAs showing 23 different alleles or patterns.

UPIC		M1	M2	M3	M4	M5	M6	M7	M8	
2_Marker_Combinations	18	0015_a	0124_a	0350_a						
	18	0094_a	0015_a	0350_a						
	18	0284_a	0124_a	0350_a						
	18	0284_a	0094_a	0350_a						
3_Marker_Combinations	23	0350_a	0015_a	0076_a	0124_a					
	23	0094_a	0350_a	0284_a	0076_a					
	23	0350_a	0284_a	0076_a	0124_a					
	23	0094_a	0350_a	0015_a	0076_a					
	22	0094_a	0350_a	0015_a	0124_a					
	22	0350_a	0284_a	0015_a	0124_a					
	22	0094_a	0350_a	0284_a	0124_a					
	22	0094_a	0350_a	0284_a	0015_a					
	20	0194_a	0094_a	0350_a	0015_a					
	20	0194_a	0094_a	0350_a	0284_a					
	20	0194_a	0350_a	0015_a	0124_a					
	20	0194_a	0350_a	0284_a	0124_a					
	19	0350_a	0114_a	0015_a	0124_a					
	19	0094_a	0350_a	0114_a	0015_a					
	19	0350_a	0114_a	0284_a	0124_a					
	19	0094_a	0350_a	0114_a	0284_a					
4_Marker_Combinations										
5_Marker_Combinations	27	0094_a	0015_a	0350_a	0076_a	0124_a				
	27	0094_a	0284_a	0350_a	0076_a	0124_a				
	27	0094_a	0284_a	0015_a	0350_a	0076_a				
	27	0284_a	0015_a	0350_a	0076_a	0124_a				
	26	0094_a	0284_a	0015_a	0350_a	0124_a				
	25	0194_a	0015_a	0350_a	0076_a	0124_a				
	25	0194_a	0284_a	0350_a	0076_a	0124_a				
	25	0194_a	0094_a	0284_a	0350_a	0076_a				
	25	0194_a	0094_a	0015_a	0350_a	0076_a				
	24	0015_a	0350_a	0114_a	0076_a	0124_a				
	24	0094_a	0284_a	0350_a	0114_a	0076_a				
	24	0194_a	0094_a	0284_a	0015_a	0350_a				
	24	0284_a	0350_a	0114_a	0076_a	0124_a				
	24	0194_a	0094_a	0284_a	0350_a	0124_a				
	24	0194_a	0284_a	0015_a	0350_a	0124_a				
	24	0194_a	0094_a	0015_a	0350_a	0124_a				
	24	0094_a	0015_a	0350_a	0114_a	0076_a				
	23	0094_a	0284_a	0350_a	0114_a	0124_a				
	23	0094_a	0015_a	0350_a	0114_a	0124_a				
	23	0284_a	0015_a	0350_a	0114_a	0124_a				
	23	0094_a	0284_a	0015_a	0350_a	0114_a				
	21	0194_a	0284_a	0350_a	0114_a	0124_a				
	21	0194_a	0094_a	0284_a	0350_a	0114_a				
	21	0194_a	0094_a	0015_a	0350_a	0114_a				
	21	0194_a	0015_a	0350_a	0114_a	0124_a				
	6_Marker_Combinations	31	0350_a	0284_a	0015_a	0094_a	0076_a	0124_a		
		29	0350_a	0284_a	0015_a	0194_a	0076_a	0124_a		
29		0350_a	0015_a	0194_a	0094_a	0076_a	0124_a			
29		0350_a	0284_a	0194_a	0094_a	0076_a	0124_a			
29		0350_a	0284_a	0015_a	0194_a	0094_a	0076_a			
28		0350_a	0015_a	0094_a	0114_a	0076_a	0124_a			
28		0350_a	0284_a	0015_a	0114_a	0076_a	0124_a			
28		0350_a	0284_a	0015_a	0194_a	0094_a	0124_a			
28		0350_a	0284_a	0015_a	0094_a	0114_a	0076_a			
28		0350_a	0284_a	0094_a	0114_a	0076_a	0124_a			
27		0350_a	0284_a	0015_a	0094_a	0114_a	0124_a			
26		0350_a	0284_a	0194_a	0114_a	0076_a	0124_a			
26		0350_a	0015_a	0194_a	0114_a	0076_a	0124_a			
26		0350_a	0015_a	0194_a	0094_a	0114_a	0076_a			
26		0350_a	0284_a	0194_a	0094_a	0114_a	0076_a			
25		0350_a	0015_a	0194_a	0094_a	0114_a	0124_a			
25		0350_a	0284_a	0015_a	0194_a	0114_a	0124_a			
25	0350_a	0284_a	0194_a	0094_a	0114_a	0124_a				
7_Marker_Combinations	33	0094_a	0350_a	0015_a	0194_a	0076_a	0124_a	0284_a		
	32	0094_a	0350_a	0015_a	0114_a	0076_a	0124_a	0284_a		
	30	0350_a	0015_a	0194_a	0114_a	0076_a	0124_a	0284_a		
	30	0094_a	0350_a	0015_a	0194_a	0114_a	0076_a	0284_a		
	30	0094_a	0350_a	0015_a	0194_a	0114_a	0076_a	0124_a		
	30	0094_a	0350_a	0194_a	0114_a	0076_a	0124_a	0284_a		
	29	0094_a	0350_a	0015_a	0194_a	0114_a	0124_a	0284_a		
8_Marker_Combinations	34	0015_a	0194_a	0114_a	0284_a	0350_a	0094_a	0076_a	0124_a	

**Table S4: Output for PIC values and Percentage of Heterozygous Loci.** The five columns on the bottom correspond to: marker, square of the allele frequency, PIC value for self fertilized organisms, PIC value of equiprequent alleles, and PIC value for cross fertilized organisms with variable allele number.

DNA Sample	% Heterozygous Loci	marker	sum_Fsquare	picval-self_fert	picval equipreq	picval_cross_fertil.
1	0.25	0015_a	0.3191589	0.6808411	0.578978697	0.568978697
2	0.375	0076_a	0.16666666	0.83333334	0.805555564	0.795555564
3	0.5	0094_a	0.26847222	0.73152778	0.659450447	0.649450447
4	0.625	0114_a	0.32986111	0.67013889	0.561330538	0.551330538
5	0.5	0124_a	0.2982253	0.7017747	0.61283637	0.60283637
6	0.5	0194_a	0.3888888	0.61111112	0.459876701	0.449876701
7	0.78	0284_a	0.20476466	0.79523534	0.753306774	0.743306774
8	0.5	0350_a	0.1388888	0.86111112	0.841821101	0.831821101
9	0.5					
10	0.375					
11	0.375					
12	0.25					