FILE S3

Equations for calculating the expected gametic and allelic frequencies under Model 1

The expected frequencies F_c of the different 1n-2c configurations can be derived from the recombination frequencies between the three S_I genes and between the S_I complex locus and the centromere (see schematic representation 1 below), and are given in Table S3-1.

The expected frequencies of each gamete that belong to a specific 1n-2c configuration are equal to $F_c/4$, where F_c is the expected frequency of the specific 1n-2c configuration.

Let *D* be the sum of the frequencies of the viable gametes. Under model 1, *D* simplifies to:

$$D = (1-2r_5)\left(\frac{r_2+1}{2}\right) + 2r_5\left[\frac{1}{2}+\frac{r_1}{2}-r_1r_2\right]$$

The expected frequencies of *viable* gametes are obtained by dividing their corresponding expected frequency by *D*.

Thus, the final frequency of gametes bearing the $S_1^{\,s}$ allele, $f\left(S_1^{\,s}\right)$, is obtained by summing the individual frequencies of viable gamete that bear this allele. Under Model 1, this simplifies to:

$$f\left(S_1^{\,s}\right) = \left[\frac{r_2}{2} + r_5\left(r_1 - r_2 - 2r_1r_2\right)\right] \Big/ D$$

Similar equations can be derived for Models 2 to 4, with numerical examples based on observed data given in File "S1 - Genetic models.xlsx".

Under all models, the estimated for recombination fractions r_1 and r_2 are naturally given by dividing the observed recombination fractions, $r_1$max and $r_2$max, by 2*D*.

$$\hat{r}_1 = \frac{r_1 \max}{2D};\ \hat{r}_2 = \frac{r_2 \max}{2D}.$$

In the case of the presence of the additional factor S_IC (see schematic representation 2 below), the effect of S_IC is applied after the initial selection due to the S_IA- S_I- S_IB locus action.

As S_IC only affects the $S_I^{\,s}$ gametes when a recombination event occurs between S_IB and S_IC, the expected frequencies of these gametes are obtained by multiplying them by $1-r_3$, where r_3 is the recombination fraction between S_IB and S_IC, and dividing them by the relative sum of all gamete frequencies, $1-r_3 f\left(S_1^{\,s}\right)$. The frequencies of other viable gametes are obtained by dividing them by the relative sum of all gamete frequencies, $1-r_3 f\left(S_1^{\,s}\right)$.

TABLE S3-1

Expected frequencies F_c of the different 1n-2c configuration expressed in function of the recombination frequencies between the three S_1 genes, and associated gametic frequencies under no selection and Model 1 selection

1n-2c configuration S_1A	S_1	S_1B	*Expected configuration frequency without selection* (F_c)	*Expected gamete frequency without selection* (F_g)	*Survival under Model 1*	*Expected gamete frequency under selection (Model 1)*
s	s	s		$F_c^1/4$	Aborted	
s	s	s		$F_c^1/4$	Aborted	
			$F_c^{\,1} = (1-2r_5)\left[1-2(r_1+r_2-2r_1r_2)\right]$			
g	g	g		$F_c^1/4$	Viable	F_g/D
g	g	g		$F_c^1/4$	Viable	F_g/D
s	s	s		$F_c^2/4$	Aborted	
g	g	g		$F_c^2/4$	Viable	F_g/D
			$F_c^{\,2} = 2r_5\left[1-2(r_1+r_2-2r_1r_2)\right]$			
s	s	s		$F_c^2/4$	Aborted	
g	g	g		$F_c^2/4$	Viable	F_g/D
s	s	s		$F_c^3/4$	Viable	F_g/D
g	g	s		$F_c^3/4$	Viable	F_g/D
	X		$F_c^{\,3} = 2r_2(1-2r_1)(1-2r_5)$			
s	s	g		$F_c^3/4$	Aborted	0
g	g	g		$F_c^3/4$	Viable	F_g/D
s	s	s		$F_c^4/4$	Aborted	0
g	s	s		$F_c^4/4$	Aborted	0
X			$F_c^{\,4} = 2r_1(1-2r_2)(1-2r_5)$			
s	g	g		$F_c^4/4$	Viable	F_g/D
g	g	g		$F_c^4/4$	Viable	F_g/D
s	s	s		$F_c^5/4$	Viable	F_g/D
s	g	s		$F_c^5/4$	Viable	F_g/D
X	X		$F_c^{\,5} = 4r_1r_2(1-2r_5)$			
g	s	g		$F_c^5/4$	Aborted	0
g	g	g		$F_c^5/4$	Viable	F_g/D

s s s		$F_c^{6}/4$	Aborted	0
s s g		$F_c^{6}/4$	Aborted	0
X	$F_c^{6} = 4r_2r_5(1-2r_1)$			
g g s		$F_c^{6}/4$	Viable	F_g/D
g g g		$F_c^{6}/4$	Viable	F_g/D
s s s		$F_c^{7}/4$	Viable	F_g/D
s g g		$F_c^{7}/4$	Viable	F_g/D
X	$F_c^{7} = 4r_1r_5(1-2r_2)$			
g s s		$F_c^{7}/4$	Aborted	0
g g g		$F_c^{7}/4$	Viable	F_g/D
s s s		$F_c^{8}/4$	Aborted	0
g s g		$F_c^{8}/4$	Aborted	0
X X	$F_c^{8} = 8r_1r_2r_5$			
s g s		$F_c^{8}/4$	Viable	F_g/D
g g g		$F_c^{8}/4$	Viable	F_g/D

The sum of the frequencies of the viable gametes thus simplifies to:

$$D' = D \Big/ \left(1 - f\left(S_1^s\right)\right) r_3,$$

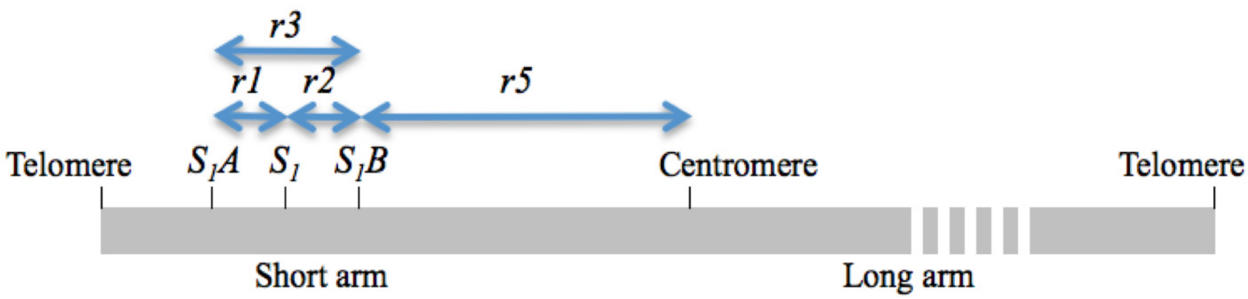
while the final frequency of gametes bearing the S_1^s allele converts to:

$$f'\left(S_1^s\right) = \left(1 - r_3\right) f\left(S_1^s\right) \Big/ \left(1 - f\left(S_1^s\right) r_3\right),$$

and the estimates for r_1 and r_2 convert to:

$$\hat{r}_1 = \frac{r_1 \max}{2D'};\ \hat{r}_2 = \frac{r_2 \max}{2D'}.$$

Schematic representation 1: positions of the three S_1 genes on rice chromosome 6, expressed as recombination franctions between each other and between the S_1 locus and the centromere.



Schematic representation 2: positions of the three S_1 genes and the S_1C gene on rice chromosome 6, expressed as recombination fractions between each other and between the S_1 locus and the centromere.