

1 **Supplementary material**

2 **Candidate genes under balancing selection in a plant bacterial pathogen**

3 **José A. Castillo^{1*} & Spiros N. Agathos¹**

4 ¹School of Biological Sciences and Engineering, Yachay Tech University, Hacienda San Jose s/n and Proyecto Yachay, Urcuquí, Ecuador.

5 *Corresponding Author

6 jcastillo@yachaytech.edu.ec

7

8 **Supplementary Table 1.** Summary statistics of nucleotide site frequency spectrum for each phylotype and replicon of RSSC.

Phylotype/ replicon	Tajima's D			θ_w			Fu & Li's D*		
	Highest value	Lowest value	Media value	Highest value	Lowest value	Media value	Highest value	Lowest value	Media value
I/chromosome	3.4633	-2.5060	-0.3099	0.0789	0.0014	0.0056	1.7266	-3.8352	-0.4272
I/megaplasmid	2.5310	-2.5673	-0.3166	0.1395	0.0014	0.0070	1.6620	-4.0969	-0.4285
IIA/chromosome	2.7470	-2.2010	0.8652	0.1027	0.0017	0.0103	1.6397	-2.8045	0.7222
IIA/megaplasmid	2.4308	-1.6949	-0.5583	0.2490	0.0020	0.0162	1.6971	-1.7987	-0.6115
IIB/chromosome	3.0226	-2.3402	0.6499	0.0888	0.0014	0.0082	1.6966	-3.8403	0.3405
IIB/megaplasmid	3.0087	-2.3765	0.6743	0.1409	0.0014	0.0117	1.7475	-4.0058	0.3506
IV/chromosome	2.2454	-1.2638	-0.3198	0.1848	0.0024	0.0110	1.8719	-1.2638	-0.3377
IV/megaplasmid	2.6745	-1.2626	-0.3512	0.2616	0.0024	0.0138	1.7880	-1.2626	-0.3785

9

10 **Supplementary Table 2.** Two-dimensional plot of three summary statistics calculated using sliding window data.

Phylotype/ replicon	R_s		
	θ_w -Tajima's D	Tajima's D-Fu & Li's D*	θ_w -Fu & Li's D*
I/chromosome	-0.013ns	0.809**	-0.028**
I/megaplasmid	0.012ns	0.792**	-0.033*
IIA/chromosome	0.133**	0.738**	0.178**
IIA/megaplasmid	-0.068**	0.964**	-0.099**
IIB/chromosome	0.103**	0.748**	0.108**
IIB/megaplasmid	-0.158**	0.764**	0.127**
IV/chromosome	-0.054**	0.982**	-0.07**
IV/megaplasmid	-0.019ns	0.976**	-0.041*

11 * and ** indicate significant differences at $p < 0.05$ and $p < 0.01$, respectively; ns indicate
 12 non-significant differences, as determined using Spearman's correlation coefficient (R_s)

13

14 **Supplementary Table 3.** List of genes with unknown function and intergenic regions showing highest observed values of three statistics (θ_w ,
 15 Tajima's *D*, and Fu & Li's *D**) in the genome-wide analysis of RSSC phylotypes.

Phylotype/ replicon	Gene ID ^a	Number of significant hits ^b	Summary statistics ^c			Gene description
			θ_w	Tajima's <i>D</i>	Fu & Li's <i>D</i> *	
I/megaplasmid	RSp1676 - RSp1677	1	0.0226	2.3992**	1.5336**	Intergenic region
IIA/chromosome	RCFBP_11263 - RCFBP_11264	1	0.0613**	1.9057**	1.6158**	Intergenic region
IIA/chromosome	RCFBP_10695	1	0.0248**	2.0323**	1.5085**	Conserved hypothetical protein
IIA/chromosome	RCFBP_10153	1	0.0248**	1.9013**	1.5085	Conserved hypothetical protein
IIA/chromosome	RCFBP_21040 - RCFBP_21042	1	0.0331**	2.0892**	1.5509**	Intergenic region
IIA/chromosome	RCFBP_20766	1	0.0265**	2.0659**	1.5188**	Conserved hypothetical protein
IIA/chromosome	RCFBP_11871	1	0.0265**	1.9795**	1.5188**	Conserved hypothetical protein
IIA/chromosome	RCFBP_20933 - RCFBP_20934	1	0.0248**	2.1239**	1.5085**	Intergenic region
IIB/chromosome	RSPO_c00534	1	0.0183**	2.1298**	1.4823**	Hypothetical cytosolic protein
IIB/chromosome	RSPO_c01192	1	0.0324**	3.0226**	1.6101**	Conserved hypothetical protein
IIB/chromosome	RSPO_c01332 - RSPO_c01333	1	0.0451**	2.0604**	1.4903**	Intergenic region
IIB/chromosome	RSPO_c02427	1	0.0183**	2.0777**	1.4823**	Conserved hypothetical protein
IIB/chromosome	RSPO_c02337	1	0.0183**	2.1871**	1.4823**	Conserved hypothetical protein
IIB/chromosome	RSPO_c01923	1	0.0240**	2.7797**	1.5475**	Conserved hypothetical protein
IIB/chromosome	RSPO_c02827	2	0.0352**	2.0961**	1.6253**	Conserved hypothetical protein
IIB/chromosome	RSPO_c02725 - RSPO_c02726	1	0.0211**	2.0978**	1.5182**	Intergenic region
IIB/chromosome	RSPO_c03232 - RSPO_c03231	1	0.0226**	2.1389**	1.5335**	Intergenic region
IIB/megaplasmid	RSPO_m00507 - RSPO_m00508	1	0.0338**	2.7577**	1.3908**	Intergenic region
IIB/megaplasmid	RSPO_m00342 - RSPO_m00343	1	0.0282**	2.1609**	1.5826**	Intergenic region
IIB/megaplasmid	RSPO_m01597	3	0.0465**	2.7299**	1.6703**	Conserved hypothetical protein
IV/chromosome	RPSI07_0472 - RPSI07_0473	1	0.0264**	1.7003**	1.7993**	Intergenic region
IV/megaplasmid	RPSI07_mp0104 - RPSI07_mp0105	1	0.1464**	1.7880**	1.7880**	Intergenic region

16 ^a Systematic gene identifier according to GMI1000, CFBP2957, Po82 or PSI07 strain nomenclature for phylotype I, IIA, IIB or IV respectively.

17 ^b Number of significant windows overlapping described gene.

18 ^c Observed values of statistics for each gene and significance of coalescent simulations using standard neutral model: * $p < 0.1$ and ** $p < 0.05$

19

20