Characterizing the *S*-locus of *S₂*-haplotype of self-incompatible *Petunia inflata* by sequence analysis of 17 *S*-locus *F-box* gene-containing bacterial artificial chromosome (BAC) contigs
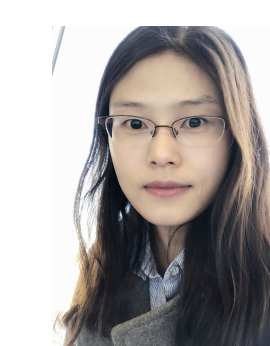
Lihua Wu[*1], Justin S. Williams[2], and Teh-hui Kao[1,2]

[1]Intercollege Graduate Degree Program in Plant Biology, The Pennsylvania State University, University Park, Pennsylvania 16802 USA
[2]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802 USA.

Background Information

Petunia possesses Solanaceae-type self-incompatibility (SI), which allows pistils to reject self-pollen preventing inbreeding, but accept non-self pollen for outcrossing. Self/non-self recognition is regulated by the polymorphic *S*-locus (Fig. 1). A single gene at the *S*-locus, *S-RNase*, encodes the pistil specificity determinant. The first *S-locus-F-box* (*SLF*) gene, named *SLF1*, was identified by sequencing a 328-kb BAC contig containing *S₂-RNase*. Pollen transcriptome analysis revealed 16 additional *SLF* genes linked to the *S₂*-locus (Fig. 1). All 17 *SLF* genes collectively encode the pollen-specificity determinant.

Figure 1. Schematic diagram of *S₂*-locus. Except for the location of *S₂-SLF1*, the locations of the other 16 *SLF* genes (*S₂-SLF2* to *S₂-SLF17*) relative to *S₂-RNase* are yet undetermined.
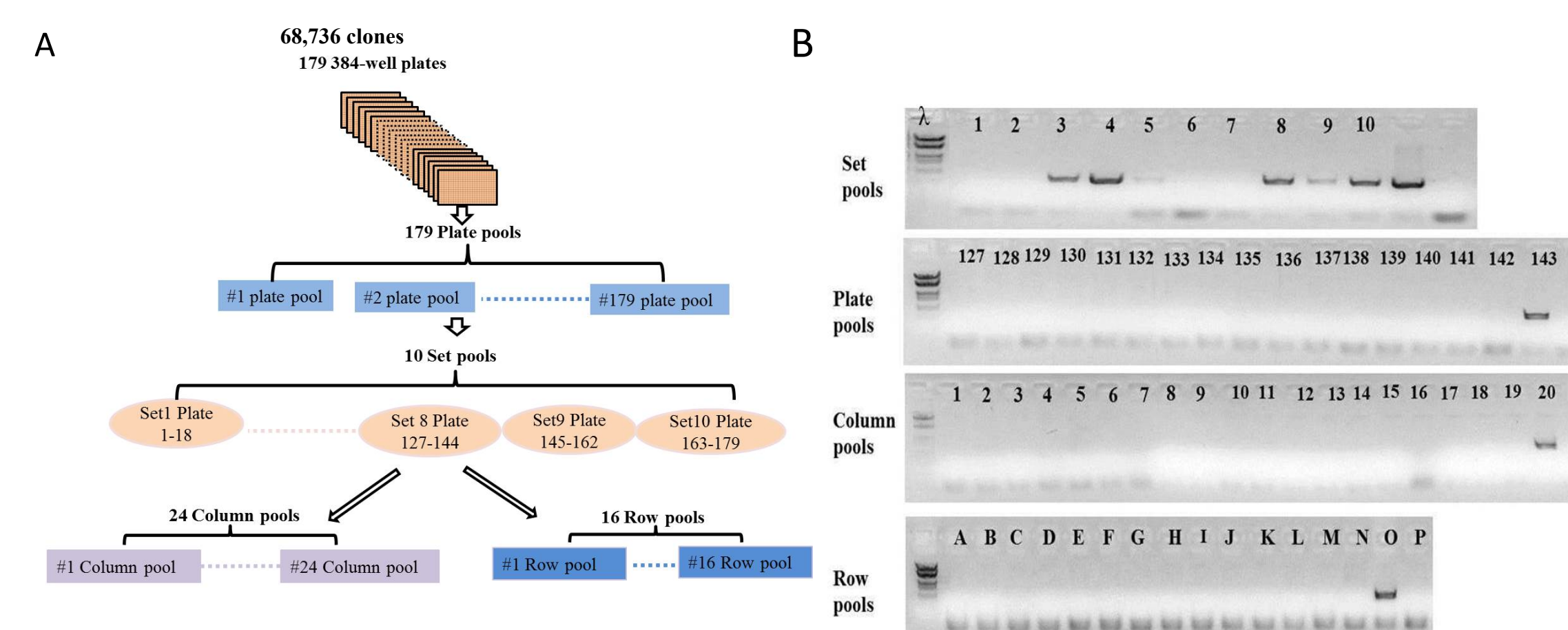

Figure 2. Using gene markers to screen the *S₂S₂* BAC library. (A). Schematics showing the "pooling" strategy used for screening the *S₂S₂* BAC library. (B) An example showing the use of the *S₂-SLF8* specific primers to screen the library, and the identification of a BAC clone, 143O20, containing *S₂-SLF8*.

Methods

We used *SLF2* to *SLF17* as markers to isolate BAC clones from the previously constructed *S₂* library (Fig. 2), and used Illumina Miseq and PacBio SMRT sequencing technology to sequence genomic DNA inserts of these BAC clones, as well as of a previously assembled 881-kb BAC contig containing the 328-kb region. The sequence of each BAC clone was assembled using a combination of Illumina MiSeq and PacBio sequence reads. Illumina read processing and assembly was performed in-house, whereas all PacBIo read quality processing was performed through the SMRT analysis pipeline (v2.3.0) (Fig. 3A).
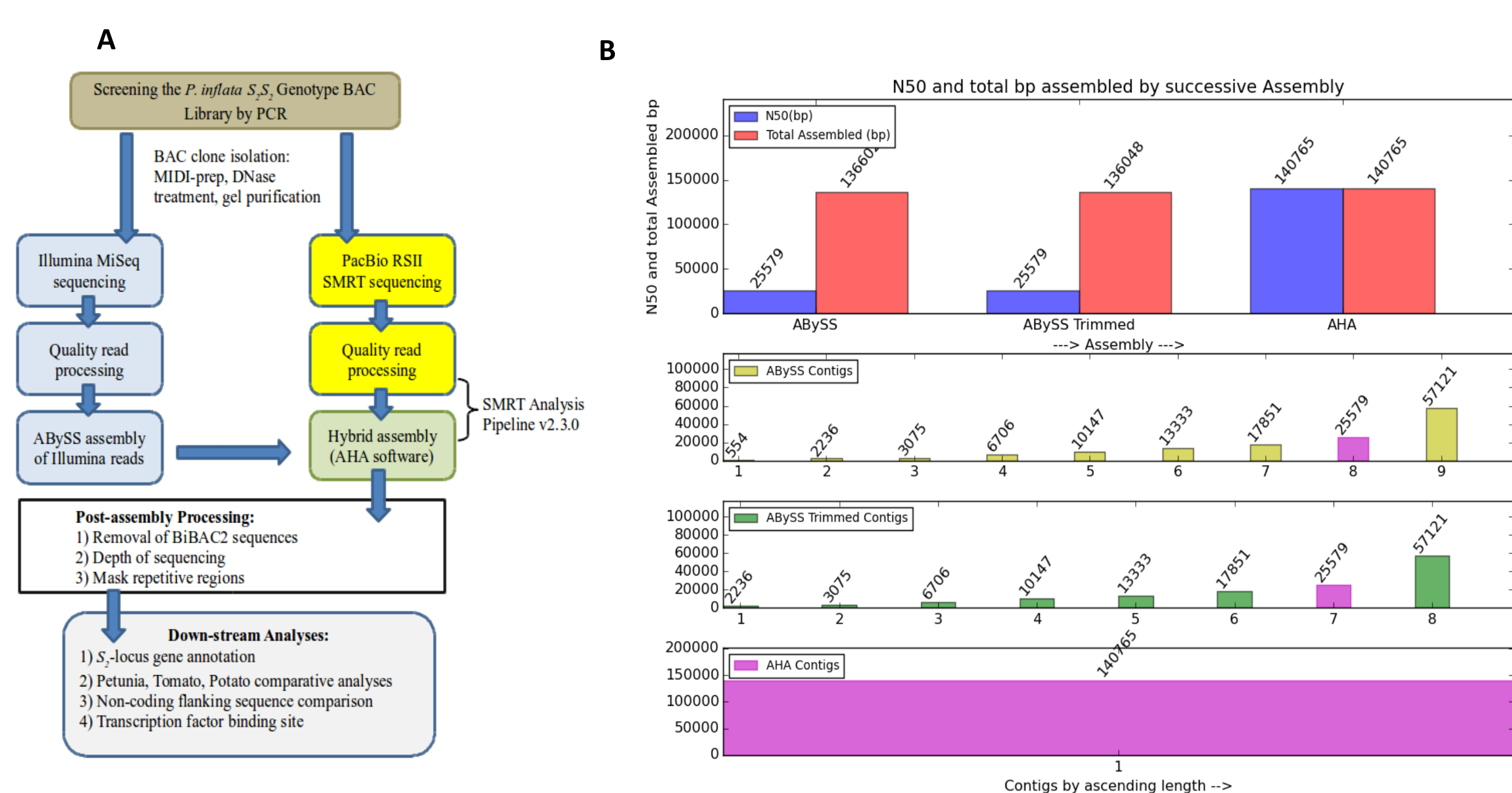

Figure 3: (A). The *S₂*-locus assembly workflow. (B) Scaffold statistics of successive assembly steps of *S₂-SLF12*. The top panel shows the increase in assembly quality with each successive step, as indicated by comparing the N50 value to the total number of base pairs assembled (left to right).

Results

A total of 3.1 Mbp (a single contig each for 13 of the 17 *SLF* genes) were assembled (Table 1), 20.34% of which were repetitive sequences (Fig. 4).

BAC Clone Library	Total Base Pairs Assembled	N50 (bp)	No. of Contigs Assembled	No. of Gaps	Total Length of Gaps
SLF1	152575	152575	1	15	17849
SLF2	103037	103037	1	4	1712
SLF3	129437	129437	1	14	13273
SLF4	19968	19968	1	0	0
SLF5	188970	114450	3	23	34572
SLF6	110096	110096	1	2	889
SLF7	146378	146378	1	4	2963
SLF8	98828	35807	3	1	873
SLF9	180025	180025	1	19	15827
SLF10	119723	72959	3	2	726
SLF11	114660	114660	1	10	15429
SLF12	115416	115416	1	8	4817
SLF13	108701	108701	1	8	4081
SLF14	126512	126512	1	8	7155
SLF15	104282	32263	4	6	5787
SLF16	131162	128026	2	9	9536
SLF17	154395	154395	1	5	2684
SLFlike1	127668	127668	1	7	5570
111P17	167525	99726	5	20	15470
161N9	119846	119846	1	4	4757
164C22	117573	117573	1	10	8675
171B19	232669	96093	4	24	20436
30F20	148556	148556	1	15	7054
48L16	121755	121755	1	7	3818

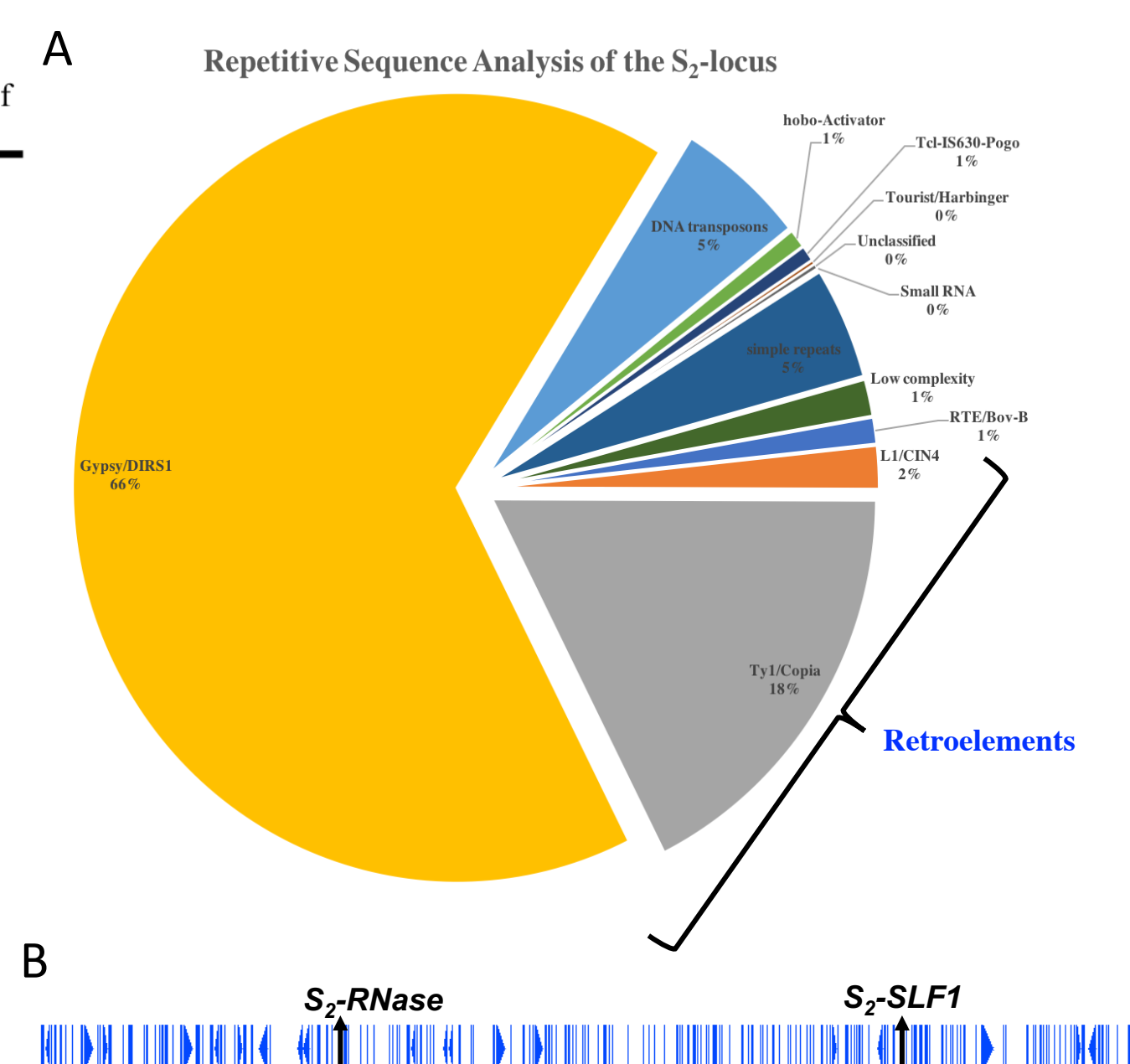Table 1 Post-processed hybrid assembly statistics

Figure 4. Analysis of repetitive sequences. (A) Types of repetitive sequences. (B) An example showing locations of repetitive sequences in the 328-kb contig containing *S₂-RNase* and *S₂-SLF1*.

No additional *SLF* genes were discovered, but 38 additional genes were predicted: 30 of unknown origin and 8 annotated as encoding proteins functioning in pollen germination, pollen tube growth or guidance (Table 2). The sequence of the *S₆ₐ*-locus of *P. inflata*, extracted from the draft genome sequence, contained 29 of these 38 genes, indicating shared characteristics between different *S*-loci of the same species.

Uniprot Accession Number			
CNGC CNGC9_ARATH	cyclic nucleotide-gated ion channel	pollen tube growth	Wang et al., (2013)
BECN1 BECN1_ARATH	Beclin1 like protein	pollen germination; Vacuolar protein	Qin et al., (2007)
DRM DRM1_ARATH	DNA (cytosine-5)-methyltransferase DRM2	DNA methylation	Calarco et al., (2012)
PDI PDI_DATGL	protein disulfide isomerase	pollen tube guidance	Wang et al., (2008)
R27A R27A2_ARATH	60S ribosomal protein L27a-2	protein synthesis	Klinge et al., (2011)
WSD WSD1_ARATH	O-acyltransferase DCAF4	cuticular wax biosynthesis	Li et al., (2008)
CUL4 DCAF4_BOVIN	CULLIN-4	ubiquitination SCF complex member	Seo et al., (2014)
TSS TSS_ARATH	TPR-domain suppressor of STIMPY	cell cycle regulator	Skylar et al., (2011)

Table 2. Eight genes identified in the *S₂*-locus of *Petunia inflata* and annotated as encoding proteins functioning in pollen germination, pollen tube growth, or guidance. .

S-locus remnants on Chromosome 1 of self-compatible tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*) cultivars contained the *S-RNase* remnant and *SLF* remnants in a sub-centromeric region, but did not contain any of the 38 annotated genes, suggesting the unique feature of the S-locus genes involved in SI.
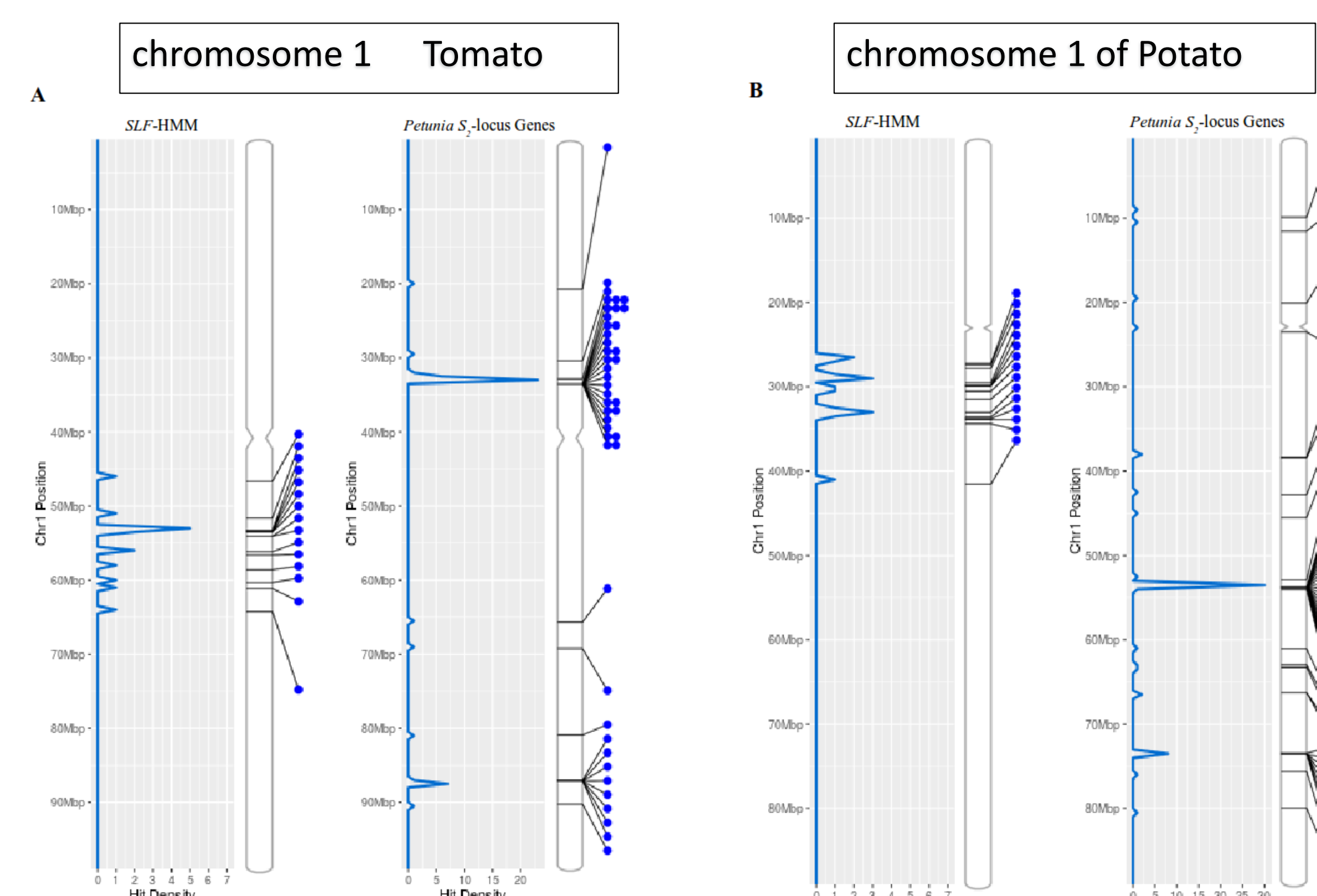

Figure 5. Comparative analyses of the *Petunia S₂*-locus and chromosome 1 of both Potato and Tomato. Panel A shows the locations of *SLF* remnants, as determined by a hidden-markov model, in chromosome 1 of Tomato (left), and the locations of the homologs of the 38 genes identified in the *Petunia S₂*-locus. Panel B shows a similar analysis using chromosome 1 of Potato. In both panels, the hit density (genes per 500 Kbp) and locations of these hits are shown.
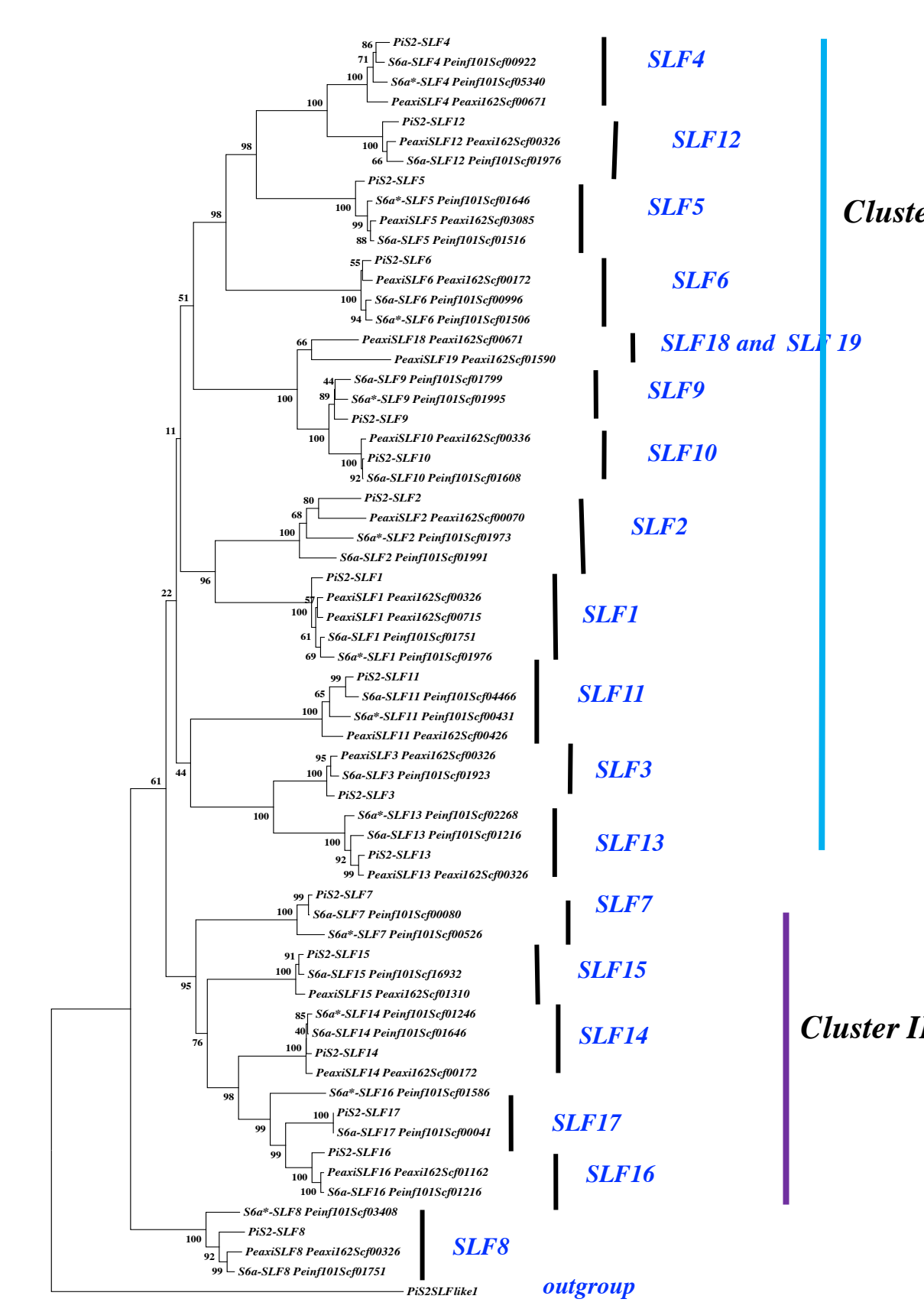

Figure 6. Phylogenetic relationships of *SLF* genes of *S₂*-locus and *S₆ₐ*-locus of *Petunia inflata* and of *Petunia axillaris*.

For both *S₂* and *S₆ₐ* loci, comparison of the upstream and downstream non-coding sequences of different *SLF* genes (Fig. 6 and Fig. 7) revealed that both recombination and retrotransposition might have played a role in the expansion of *SLF* genes.
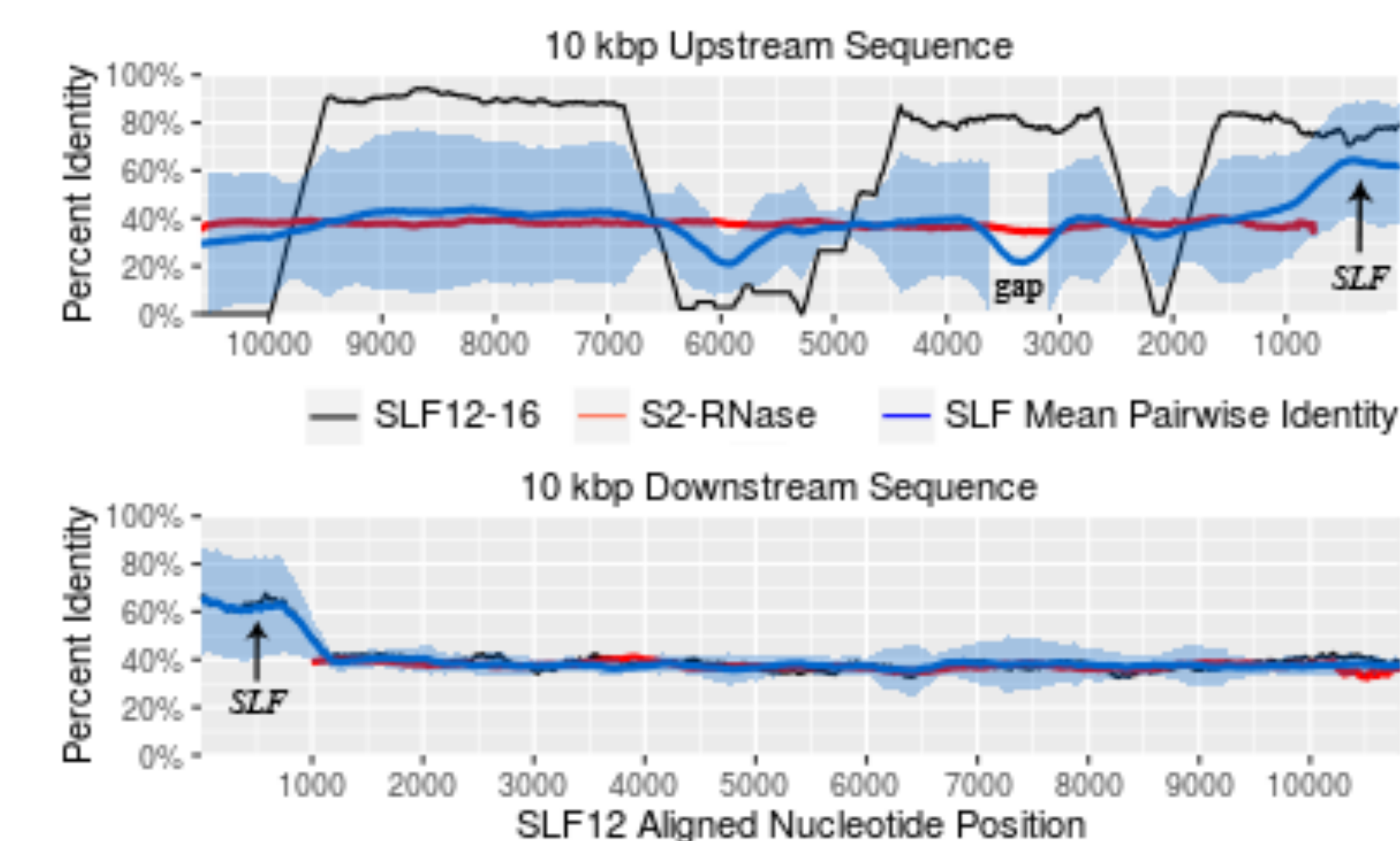

Figure 7. Comparison of non-coding sequences flanking *SLF* genes. Sequence comparison of 10 Kbp upstream (top) and 10 Kbp downstream (bottom) between *S₂-SLF12* and *S₂-SLF16* (black line), and between *S₂-RNase* and all 17 *SLF* genes (red line).

Conclusion

The sequence of the *S₆ₐ*-locus of *P. inflata*, extracted from the draft genome sequence, contained 29 of the 38 genes, indicating shared characteristics between different *S*-loci of the same species. *S*-locus remnants on chromosome 1 of self-compatible tomato (Solanum lycopersicum) and potato (Solanum tuberosum) cultivars contained the *S-RNase* remnant and *SLF* remnants in a sub-centromeric region, but did not contain any of the 38 annotated genes, suggesting the unique feature of the *S*-locus genes involved in SI. For both *S₂* and *S₆ₐ* loci, comparison of the upstream and downstream non-coding sequences of different *SLF* genes revealed that both recombination and retrotransposition might have played a role in the expansion of SLF genes

Acknowledgements

References

1. Wang, Y., Tsukamoto, T., Yi, K. W., Wang, X., Huang, S., McCubbin, A. G., Kao, T.-h. (2004). Chromosome walking in the *Petunia inflata* self-incompatibility (*S-*) locus and gene identification in an 881-kb contig containing *S₂-RNase*. *Plant molecular biology*, 54(5), 727-742.