Supplementary Information

Supplementary Methods, Supplementary References, Supplementary Figures 1-4, Supplementary Tables 1-6, Supplementary Sequence Analyses 1-3

Supplementary Methods

Genome assembly and annotation

The long homostyle assembly (LH_v2) was generated using SOAPdenovo v2.04[1] to assemble contigs, then scaffolded incrementally using three long mate-pair (LMP) libraries (5, 7 and 9 kb). A *k*-mer length of 81 was used to assemble paired-end (PE) reads (-K 81) and a *k*-mer length of 41 was used to scaffold contigs (-k 41). Prior to assembly, adapters were trimmed from the LMP reads using NextClip[2]. The SOAPdenovo GapCloser tool was used to fill gaps in the scaffolds. The assembly was screened for contamination and contaminated sequences were removed. Additional paired-end read assemblies were generated using ABySS v1.3.4[3]. For the thrum parent version 1 assembly (TP_v1), PE reads were assembled using a *k*-mer length of 71 (k=71). TP_v1.1 was generated by scaffolding the TP_v1 assembly with the 9 kb thrum parent LMP reads using SOAPdenovo v1.05[4] (prepare, -K 71). TP_v2 was generated by scaffolding the TP_v1 contigs with the 9 kb thrum parent LMP reads using SOAPdenovo v2.04 (prepare -K 71, map -k 71). The short homostyle assembly (SH_v2) was generated by assembling the PE reads (k=85), then using SOAPdenovo v2.04 to scaffold with the 9 kb thrum parent LMP reads (prepare -K 85, map -k 63). The pin parent assembly (PP_v2) was generated by assembling the PE reads (k=71), then using SOAPdenovo v2.04 to scaffold with the 9 kb thrum parent LMP reads (prepare -K 71, map -k 71). Sequences under 200 bp were removed from all assemblies before further analysis.

Only the long homostyle genome assembly was annotated. RepeatModeler Open v1.0.7 (http://www.repeatmasker.org) was used to identify *de novo* repetitive sequences in the scaffolds; repeats were annotated using the repeat library with a local installation of RepeatMasker Open v4.0.1 (http://www.repeatmasker.org). The *ab initio* annotation software AUGUSTUS[5] was trained with a set of full-length transcripts assembled using TopHat v2.0.11[6] and Cufflinks v2.1.1[6], then used to predict protein-coding genes using repeats, protein alignments from related species, and RNA-Seq transcript models as additional evidence. PASA[7] was used to correct the final gene models. The assembled long homostyle genome comprises 411.2 Mb assembled into 67,619 contigs over 200 bp, with 24,622 predicted genes.

Read alignment over *S* locus

A modified long homostyle reference genome file was generated by removing contigs LH_v2_0002458, LH_v2_0067593, LH_v2_0003915 and LH_v2_0000241 from the whole genome sequence assembly and adding the manually curated *S* locus contig (Supplementary Figure 1). Pin, thrum, short homostyle and long homostyle reads were aligned to the reference using BWA v0.7.7[8]. Reads aligning to the *S* locus contig with a mapping quality > 30 were extracted from the resulting BAM file using SAMtools v0.1.19[9]. Read coverage over the 455,881 bp *S* locus contig was generated using the genomeCoverageBed function of BEDTools v2.17.0[10] and the average coverage in 5 kb windows across the *S* locus contig was plotted.

In silico differential gene expression analysis

RNA was isolated in biological replicates from 15-20 mm buds of four wild-type pin plants and four wild-type thrum plants for RNA-Seq with Illumina HiSeq2000 (Supplementary Table 1a); reads were screened for rRNA removal using SortMeRNA v1.9[11], then adapter- and quality-trimmed with trim galore v0.3.3 (Q20) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore). The

RNA-Seq reads were aligned to the long-homostyle (LH_v2) genome assembly with TopHat v2.0.13[6] and differential expression carried out between the four pin- and thrum-replicate libraries using Cuffdiff[6]; this was guided by LH_v2 gene model annotations after manual curation of all *S* locus genes. The number of fragments per kilobase of transcript per million fragments mapped +1 (FPKM+1) ($\log_{10}$-transformed) is reported for genes at the *S* locus in Fig. 4a.

Analysis of thrum-specific genome regions

We crossed the individual pin and thrum plants used for genome sequencing to generate segregating pin and thrum progeny which were then sequenced in separate pools; the thrum progeny pool was not used in this analysis. RNA-Seq reads from the four pin, and four thrum, replicate libraries (Supplementary Table 1a) were aligned to the thrum parent (TP_v1) genome assembly using TopHat v2.0.13[6]; transcripts were assembled and merged with Cufflinks and Cuffmerge v2.2.1[6], and differential expression carried out with Cuffdiff[6]. Transcripts showing thrum-specific or near to thrum-specific expression (cut-off < 0.1 FPKM for pin flower) were identified from the differential expression results. Pin-progeny genomic reads were aligned to the TP_v1 assembly using BWA v0.6.2[8], and the per-base depth of read coverage for each contig calculated with the SAMtools v0.1.18[12] “depth” tool. The per-base depth of read coverage was then used to determine the mean depth and breadth of pin-progeny read coverage across each transcript region in genomic contigs identified by a thrum-specific transcript. The transcripts were classified into two groups using the *k*-means algorithm implemented in the *scikit-learn* package for Python (n_clusters=2); with the mean breadth and $\log_{10}$-transformed depth of pin-progeny read coverage across each transcript region as input variables; matplotlib was used for plotting in Python (Fig. 2b)[13,14]. Gene identities of thrum genome-specific transcripts were determined by alignment to the LH_v2 assembly and gene model annotations using Exonerate v2.2.0[15]. The number of thrum-specific (391) and pin-specific (270) genes identified in this analysis was based on a < 0.1 FPKM cut-off (for pin flower) using contigs >= 200 bp.

Bayesian relaxed-clock analysis

Multiple sequence alignment of full-length nucleotide coding sequences for *DEFICIENS (DEF)*, *GLOBOSA (GLO)* and *GLO*^{T} was carried out with MUSCLE in MEGA6[16]; species and accession numbers are listed in Supplementary Table 6a. DAMBE v6.3.3 was used (default option; fully resolved sites only) to inspect the above alignment (i) and *Primula GLO* and *GLO*^{T} sequences (ii) for sequence saturation[17]; the index of substitution saturation (Iss) (i=0.3870, or 0.4351 with 0.11 proportion of invariant sites (see below), ii=0.1187) was significantly lower than the critical value (Iss.c) (i=0.7243, ii=0.7318) ($p < 0.0001$) indicating low saturation. PAML v4.9 (yn00)[18] was used to calculate the mean number of synonymous substitutions per synonymous site (Ks) for i= 1.5593 and ii=0.443109.

Bayesian age estimation was implemented in BEAST v2.1.2[19] with a Yule tree prior and an uncorrelated lognormal relaxed clock. The GTR + I + Γ substitution model was selected based on the AIC result from jModelTest v2.1.7[20] with two gamma categories and an estimated proportion of invariant sites (initial value, 0.11); the estimate option was selected for the shape, rates and frequencies (initial values, default). Normal distribution priors with mean (±SD) based on age estimates from previous studies were used as calibration points for the divergence of *DEF-GLO* = 274.750 (±37.237) million years ago (MYA)[21-23], and the most recent common ancestors of *Arabidopsis thaliana-A. lyrata* = 12.950 (±3.009) MYA[24]; Lamiales-Solanales = 90.250 (±7.447) MYA[25,26]; Rosids-Asterids = 118.750 (±4.712) MYA[25,26] and the Asterids = 110.000 (±5.472) MYA[25,26] (Supplementary Table 6b); monophyly was enforced for the nodes used for calibration, and the *Primula GLO-GLO*^{T} clade. Nine independent Markov Chain Monte Carlo (MCMC) runs with 1 x 10^{8} generations and a sample frequency of 5,000 were combined using LogCombiner v1.7.5[27] (10% burn-in). The maximum clade credibility tree (Fig. 5) was generated with TreeAnnotator v1.7.5[27] and visualised in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) was used to assess the effective sample size (ESS) of all estimated parameters, as well as mixing and convergence of the MCMC to stationarity. The mean (5–95%Highest Posterior Density) coefficient of variation of the combined runs was 0.35 (0.14-0.59), which indicates rate heterogeneity among branches and supports the selection of a relaxed clock.

Detection of recombination in *S* locus flanking regions

Genomic paired-end reads from pin and thrum parental plants (see Supplementary Table 1) were aligned to the long homostyle (LH_v2) genome with BWA v0.6.2[8]. SAMtools v0.1.18[12] was used to remove PCR-duplicates (over-amplified fragments) with the “rmdup” tool, and for variant calling between the two read libraries and LH_v2. The genotype (GT) sub-field in the resulting Variant Call Format (VCF) files was used to determine the genotype for pin and thrum at each nucleotide position; two analyses were then carried out: firstly, a phased analysis using only heterozygous sites in thrum and secondly, using heterozygous sites in thrum as well as homozygous sites in thrum where at least one of the alleles in pin was different to thrum at that site. Sites were excluded with depth (DP) < 10, genotype quality < 30, or mapping quality (MQ) < 20 for heterozygous thrum sites (first and second analysis), and in either pin or thrum for homozygous thrum sites (second analysis). The signal of recombination was analysed following the approach used in Hybrid-Check[28]. In brief, the cumulative binomial probability was calculated for the *S* locus left- and right-flanking sequences using a sliding window of 5,000 bp and an overlap (step size) of 1,000 bp to test if the observed frequency of variant sites in each window was significantly lower than expected given the total number of variant sites in each flanking sequence; this was performed using variant sites in both (i) and (ii) above. In cases where ambiguous bases (Ns) were present, the total size of the window, or flanking sequence as a whole, was reduced by the number of Ns in that window or flanking sequence, respectively, with windows comprising solely of Ns being excluded from the analysis; sites excluded from the genotyping analysis above based on depth and quality cut-offs were omitted in the same manner.

Three analyses were carried out for both left- and right-flanking sequences: (i) including all variant sites, (ii) with variant sites in coding sequences excluded, (iii) with variants sites in genic regions (including introns, exons, 3’- and 5’-untranslated regions) excluded, based on LH_v2 gene annotations. This combination of analyses eliminated the possibility that functional sequence conservation within coding regions under purifying selection might be a constraint on sequence divergence, leading to regions of reduced polymorphism. The $-\log_{10}$(cumulative binomial probability) and total number of single nucleotide polymorphisms (SNPs) in each window was plotted in R, the uppermost dashed horizontal line indicates $-\log_{10}(p=0.05)$, and the lower dashed line the $-\log_{10}(p=0.05)$ with Bonferroni correction based on the total number of windows analysed in each flanking region.

Supplementary References

1 Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18-18 (2012).

2 Leggett, R. M., Clavijo, B. J., Clissold, L., Clark, M. D. & Caccamo, M. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* **30**, 566-568 (2014).

3 Simpson, J. T. *et al.* ABySS: A parallel assembler for short read sequence data. *Genome Research* **19**, 1117-1123 (2009).

4 Li, R. *et al.* De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research* **20**, 265-272 (2010).

5 Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research* **33**, W465-W467 (2005).

6 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**, 562-578 (2012).

7 Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* **31**, 5654-5666 (2003).

8 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).

9 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993 (2011).

10 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

11 Kopylova, E., Noe, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211-3217 (2012).

12 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).

13 Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in science and engineering* **9**, 90-95 (2007).

14 Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* **12**, 2825-2830 (2011).

15 Slater, G. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).

16 Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729 (2013).

17 Xia, X. DAMBE5: A Comprehensive Software Package for Data Analysis in Molecular Biology and Evolution. *Molecular Biology and Evolution* **30**, 1720-1728 (2013).

18 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and Evolution* **24**, 1586-1591 (2007).

19 Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol* **10**, e1003537 (2014).

20 Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Meth* **9**, 772-772 (2012).

21 Aoki, S., Uehara, K., Imafuku, M., Hasebe, M. & Ito, M. Phylogeny and divergence of basal angiosperms inferred from *APETALA3*- and *PISTILLATA*-like MADS-box genes. *J Plant Res* **117**, 229-244 (2004).

22 Hernández-Hernández, T., Martínez-Castilla, L. P. & Alvarez-Buylla, E. R. Functional Diversification of B MADS-Box Homeotic Regulators of Flower Development: Adaptive

Evolution in Protein–Protein Interaction Domains after Major Gene Duplication Events. *Molecular Biology and Evolution* **24**, 465-481 (2007).

23 Kim, S. *et al.* Phylogeny and diversification of B-function MADS-box genes in angiosperms: evolutionary and functional implications of a 260-million-year-old duplication. *American Journal of Botany* **91**, 2102-2118 (2004).

24 Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R. & Mathews, S. Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **107**, 18724-18728 (2010).

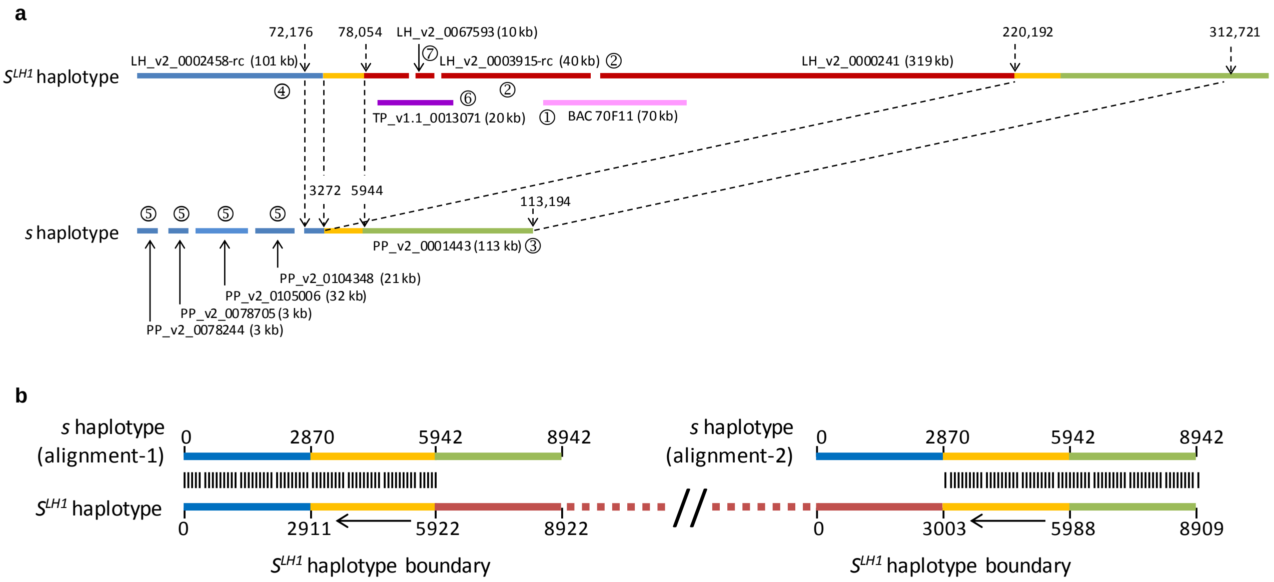
25 Bell, C. D., Soltis, D. E. & Soltis, P. S. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* **97**, 1296-1303 (2010).

26 Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. L. & Hernández-Hernández, T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, 10.1111/nph.13264 (2015).

27 Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**, 1969-1973 (2012).

28 Ward, B. J. & Oosterhout, C. Hybridcheck: software for the rapid detection, visualization and dating of recombinant regions in genome sequence data. *Molecular Ecology Resources* **16**, 543-539 (2016).

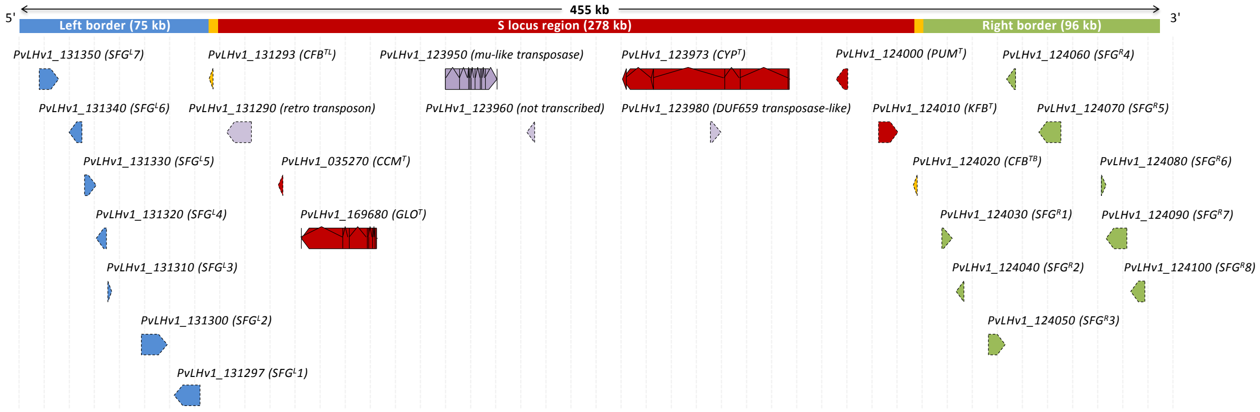
Supplementary Figures



Supplementary Figure 1 | Sequential assembly of the *P. vulgaris S* locus

a, Long homostyle S^{LH1} haplotype assembly was initiated with BAC70F11 (pink) from a BAC library[38] screened with the GLO^T cDNA[36]. Sequence contigs from *de novo* genome assemblies (Supplementary Table 1b), long homostyle (LH_v2), pin parent (PP_v2) and thrum parent (TP_v1.1), were incorporated into the S^{LH1} and *s* haplotypes using Blastn analysis (97% identity threshold). BAC 70F11 ① identified and linked LH_v2_0003915-rc (reverse complement) ② and LH_v2_0000241 ②. Sequence between 220,192 and 312,721 of LH_v2_0000241 identified PP_v2_0001443 ③ which also aligned to LH_v2_0002458-rc ④; locations within each contig showing regions of homology are shown. LH_v2_0002458_rc identified four contigs from the pin genome assembly genome PP_v2_0078244, PP_v2_0078705, PP_v2_0105006 and PP_v2_0104348 ⑤. Contigs LH_v2_0002458-rc and LH_v2_0003915-rc both identified TP_v1.1_0013071 ⑥ (purple) which bridged these two contigs and enabled placement of LH_v2_0067593 ⑦. Regions of the assemblies are colour coded: the sequence present in the S^{LH1} haplotype and absent from the *s* haplotype (red), the duplicated region flanking the S^{LH1} haplotype (yellow), sequence flanking the *S* locus to the left (blue) and right (green). **b**, Diagram showing two sequence alignments of ~9 kb from *s* haplotype with left and right border regions of the S^{LH1} haplotype. Alignments centred on the single copy *Cyclin-like F box* (*CFB*) sequence (yellow) in the *s* haplotype present as a tandem duplication in the S^{LH1} haplotype. Arrows show direction of *CFB* transcription. Sequences colour coded as in **a**. Regions of homology (97% similarity threshold) (||||), and base numbers of aligned sequences are shown (see Supplementary Sequence Analysis 1).

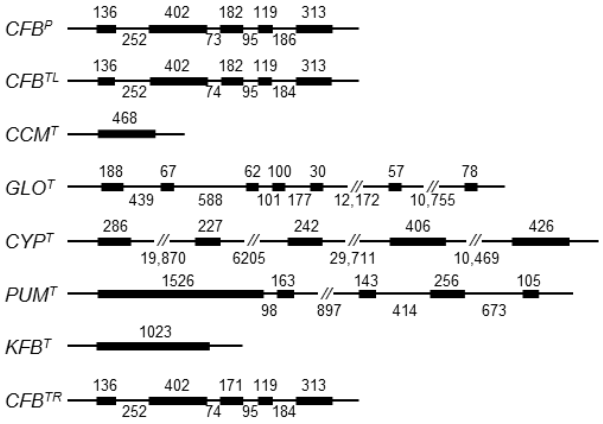
a



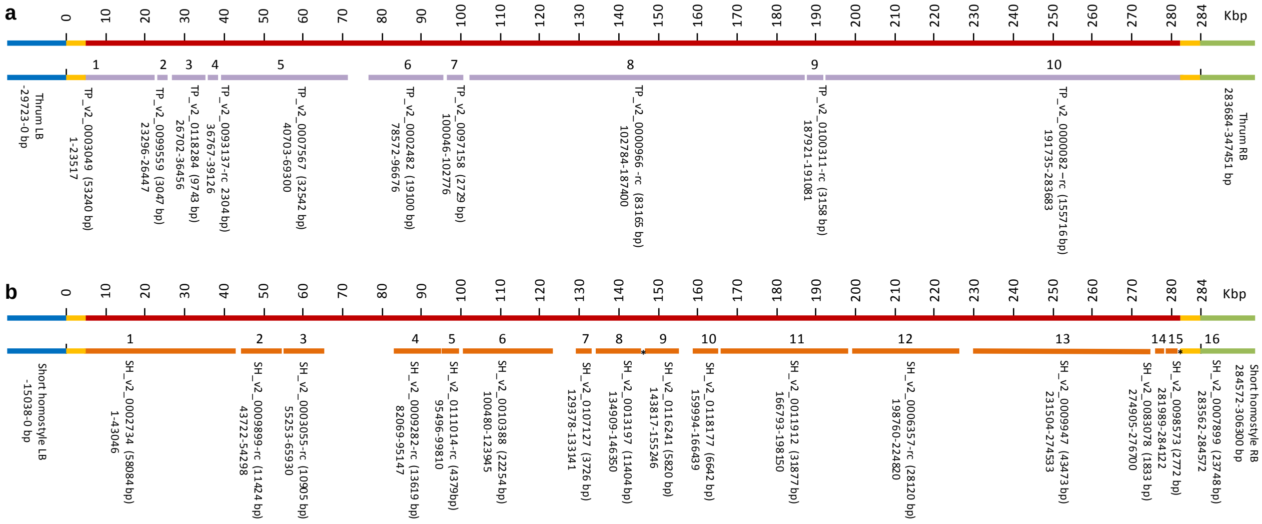
b

Gene model	Reason excluded from further analysis
*PvLHv1*_131290	Retrotransposon reverse transcriptase sequence; gene model not supported by flower transcripts
*PvLHv1*_123950	Mutator-like transposase sequence; gene not expressed in flowers
*PvLHv1*_123960	No similarity to expressed sequence in any species, flower expression is from non-*S* locus copies only
*PvLHv1*_123980	DUF659 transposase-like sequence, not *S* locus-specific

c

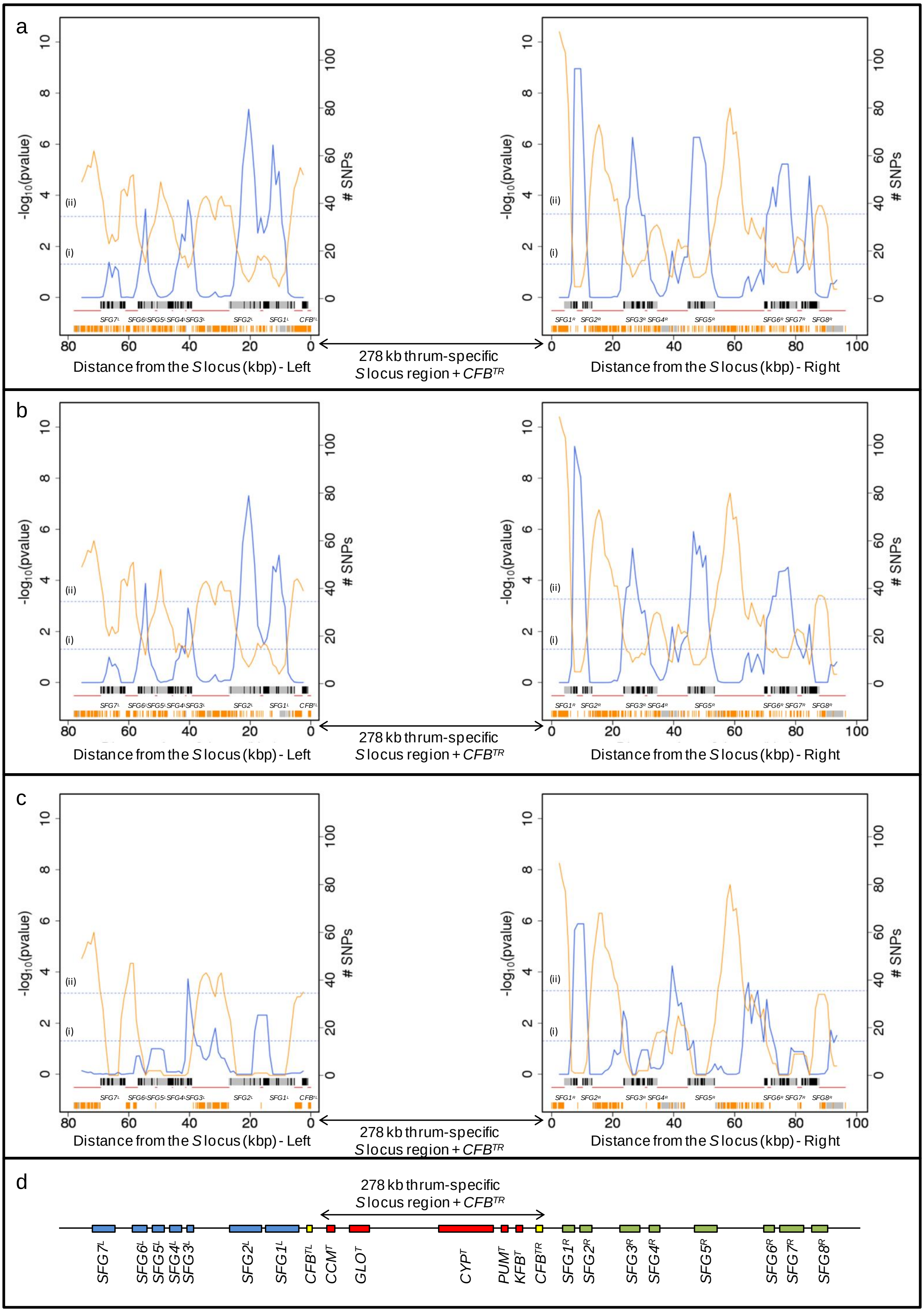


Supplementary Figure 2 | Annotation of gene models within and flanking the *S* locus. **a**, The 278 kb *S* locus region is shown in (red), the 3 kb tandemly duplicated *CFB* loci (yellow), and left (blue) and right (green) flanking sequences. The manually curated 278 kb region contains only 609 unresolved bases in repetitive regions. Automated gene models were manually curated for *CFB* loci and five *S* locus genes *CCM*^{T}, *GLO*^{T}, *CYP*^{T}, *PUM*^{T} and *KFB*^{T} (red) which are predicted only from thrum flower transcript data as thrum-specific. Other gene models are from non-curated automated annotation; predicted genes in purple were excluded from further analysis. Gene models are labelled and colour coded by location, exons shown by vertical lines, introns by linking lines, direction of transcription by arrows. Vertical lines across the 455 kb region represent 10 kb increments. **b**, Table of predicted *S* locus gene models not characterised and the rationale for exclusion. **c**, Manually curated gene models for the five thrum flower-specific *S* locus genes and flanking *CFB* loci from thrum RNA-Seq data aligned to genomic sequence, and the single *CFB*^{P} locus from pin RNA-Seq data; the 11bp deletion in *CFB*^{TR} is in exon 3. Exons (thick lines) and introns (thin lines) shown in base pairs. Long introns, not to scale, are identified by //.

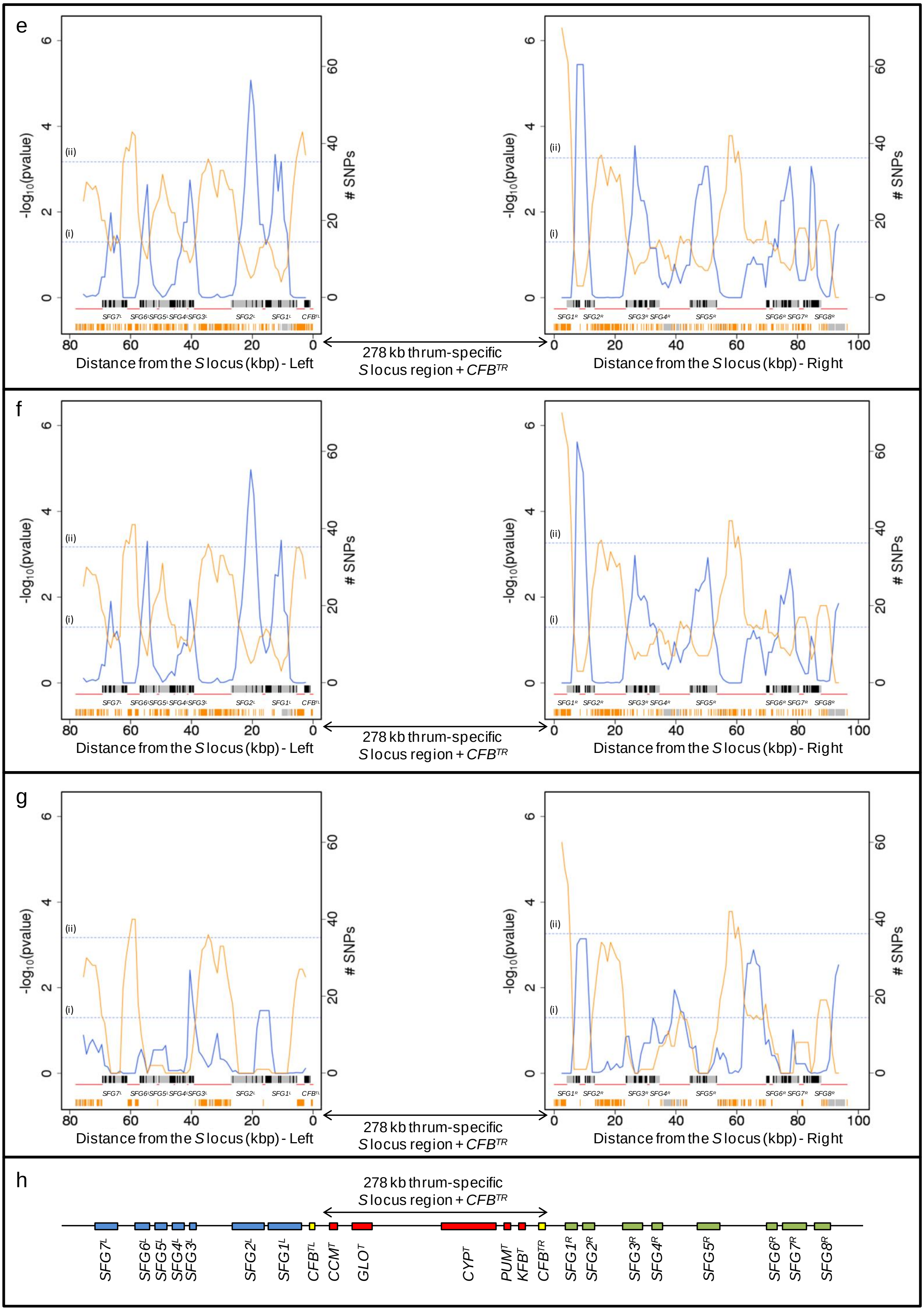


Supplementary Figure 3 | Alignment of *S* and *S*^{SH1} haplotypes to *S*^{LH1}.

S^{LH1} sequence comprising the region absent from the *s* haplotype (red), the 3 kb duplicated *Cyclin-like F Box* genes (yellow), and left (blue) and right (green) flanking sequences are shown as a contiguous line on top. Assembled contigs, some as reverse complement (rc), are shown, each contig is labelled and sizes shown in 10 kbp increments. A gap indicates that contigs do not overlap, overlaps are designated by *. Alignment coordinates for the long homostyle assembly are shown for each contig; similarity threshold 97% identity. **a**, Thrum parent genome assembly (TP_v2) contigs (purple); ten contigs span the region. **b**, Short homostyle genome assembly (SH_v2) contigs (orange); sixteen contigs span the region (see also Supplementary Table 1b).



Supplementary Figure 4a-d



Supplementary Figure 4e-h

Supplementary Figure 4 | Recombination analysis of sequences flanking the *S* locus.
The number of single nucleotide polymorphisms (# SNPs) in 5 kb sliding windows across the *S* locus flanking sequences between pin and thrum and heterozygous sites in thrum (a-c), and heterozygous sites in thrum only (e-g), are plotted (orange line) across the 75 kb to the left (Distance from *S* locus (kbp) - Left) and the 95 kb to the right (Distance from *S* locus (kbp) - Right) of the 278 kb thrum-specific region. The cumulative binomial probability ($-\log_{10}$(p-value)) of observing the number of SNPs shown (or fewer) given the frequency of SNPs in the flanking sequence as a whole is shown for 5 kb sliding windows across the *S* locus flanking sequences (blue line); horizontal dotted lines represent the critical values $-\log_{10}(p=0.05)$ (i) and $-\log_{10}(p=0.05)$ with Bonferroni correction (ii). Peaks in the blue line correspond to genomic regions where pin and thrum sequences are significantly homogenized, consistent with the effect of (recent) recombination. Note that these homogeneous regions include both genic (e.g. *SFG5*^{R}) and intergenic loci (e.g. between *SFG1*^{R} and *SFG2*^{R}) (panel a), illustrating that sequence similarity is not just the result of strong purifying selection against non-synonymous mutations. To show this more formally, SNP distribution analysis of the left and right flanking sequences with exons omitted (b and f), and with both introns and exons omitted (c and g) are also shown, which illustrates that it is recombination rather than selection that homogenised the sequence variation. Individual SNPs and their locations across the left and right *S* locus flanking sequences are shown by vertical orange bars; unresolved bases represented by Ns in the sequence, and sites excluded based on depth and quality cut-offs for SNP calling, were omitted from the 5 kb sliding window and are indicated by vertical grey bars alongside the orange SNP bars. Genes within these left (*SFG1-7*^{L}) and right (*SFG1-8*^{R}) flanking sequences, exons (black bars) and introns (grey bars), are indicated; the intergenic regions are shown by red lines. In some cases introns/exons (grey/black bars) and SNPs/omitted sites (orange/grey lines) in close proximity cannot be distinguished at this resolution. Schematic representation (d and h) of the 278 kb thrum-specific *S* locus region genes (red), left flanking genes *SFG1-7*^{L} (blue), and right flanking genes *SFG1-8*^{R} (green), are shown aligned to the data in parts a-c and e-g, with tandemly duplicated *CFB*^{TL} and *CFB*^{TR} loci which flank the 278 kb thrum-specific region in yellow.

Supplementary Tables

Supplementary Table 1 | Genome sequencing libraries and assemblies.

a

SRA Accession	Material	Type	Insert size (bp)	Read count
ERR929866	Long homostyle	Genomic	522	140114901
ERR929867	Long homostyle	Genomic	4131	40450009
ERR929868	Long homostyle	Genomic	6675	54234033
ERR929869	Long homostyle	Genomic	8818	53977718
ERR929879	Short homostyle	Genomic	464	168866674
ERR929874	Pin parent	Genomic	423	205079089
ERR929884	Thrum parent	Genomic	368	147913358
ERR929885	Thrum parent	Genomic	9780	225676870
ERR929875	Pin progeny pool	Genomic	349	197538574
SRR1578143[*]	Pin flower buds	RNA	223	100768798
SRR1578146[*]	Thrum flower buds	RNA	180	81045610
SRR1578142[*]	Oakleaf flower	RNA	220	24995179
SRR1578144[*]	Pin mature flower	RNA	209	33400153
SRR1578141[*]	Oakleaf leaf	RNA	219	14310589
SRR1578145[*]	Pin leaf	RNA	215	45723021
ERR950225	Thrum mature flower	RNA	214	75716502
ERR929876	Root, pin & thrum	RNA	199	77754060
ERR929877	Fresh seed[†]	RNA	205	63843125
ERR929878	Seedlings[†]	RNA	235	80545495
ERR929870	Pin flower rep. 1	RNA	296	39027602
ERR929871	Pin flower rep. 2	RNA	275	26497129
ERR929872	Pin flower rep. 3	RNA	318	31792278
ERR929873	Pin flower rep. 4	RNA	284	34838987
ERR950226	Thrum flower rep. 1	RNA	296	30318958
ERR929881	Thrum flower rep. 2	RNA	295	32575413
ERR929882	Thrum flower rep. 3	RNA	269	23971121
ERR929883	Thrum flower rep. 4	RNA	286	38550763

Footnotes: Illumina sequence data is available under Bioproject PRJEB9683

[*] Sequences previously submitted under Bioproject accession PRJNA260472[34]

[†] Samples include a mix of pin and thrum plant material

b

Flower form	Assembly description	Contig count	Total (Mb)	N50 (kb)	Assembly prefix
Long homostyle	ERR929866 scaffolded with ERR929867, ERR929868, ERR929869	67619	411	294.8	LH_v2
Short homostyle	ERR929879 scaffolded with ERR929885	125497	553	11.9	SH_v2
Pin	ERR929874 scaffolded with ERR929885	105238	582	14.9	PP_v2
Thrum	ERR929884 not scaffolded	156809	482	8.3	TP_v1
Thrum	ERR929884 scaffolded with ERR929885	159254	615	12.8	TP_v1.1
Thrum	ERR929884 scaffolded with ERR929885	124558	626	13.7	TP_v2

Supplementary Table 2 | Primers used in PCR analysis.

Primer	Sequence 5’ – 3’	Figure
s-F	TTGCTGCTCCGTTGAAAGAG	1c
s-R	CTGTTTAACTGACATACTCATGC	1c
SLB-F	CGAATTGGACTGATTCAGATG	1c
SLB-R	TTATCACATGCATATATAGCTAG	1c
SRB-F	CTACTCTCTTTTAGTTTGGATGAACC	1c
SRB-R	ATACTGTTTAACTGACACTCATGC	1c
GLOT-F	GAGAACAAGAAAGCTAGAGAG	3b, 3c
GLOT-R	GTCTAGCATCCCACAACCTAA	3b, 3c
GLO-F	CGGTATATATGCCCGCTTCCGTCTAA	3b, 3c
GLO-R	GCATGGTGAGTTGGTGACACTAAAATTGCT	3b, 3c

Supplementary Table 3 | Thrum-specific transcripts from *k*-means analysis

Transcript number	Contig assembly	Thrum transcript read			Thrum transcript ID	*S* locus gene
		coverage	depth	log depth		
1	TP_v1_3432270	1.73	0.02	-1.67	TCF_v1_004306	KFB^T
2	TP_v1_3559018	16.76	0.19	-0.72	TCF_v1_016075	GLO^T
3	TP_v1_3534674	20.90	0.21	-0.68	TCF_v1_013133	CYP^T
4	TP_v1_3579940	21.48	0.22	-0.67	TCF_v1_020151	GLO^T
5	TP_v1_3103680	22.30	0.25	-0.60	TCF_v1_001217	KFB^T
6	TP_v1_3432270	37.64	0.62	-0.20	TCF_v1_004305	PUM^T
7	TP_v1_3526291	54.01	0.54	-0.27	TCF_v1_012419	CYP^T
8	TP_v1_3554783	53.98	0.70	-0.15	TCF_v1_015330	GLO^T
9	TP_v1_3103680	61.84	0.90	-0.05	TCF_v1_001218	KFB^T

Supplementary Table 4 | Summary of plants from three-point cross analysis.

Parent phenotype	Genotype
Wild type pin (♀) *	*okl s hih* *okl s hih*
Hose in Hose *Oakleaf* thrum (♂) *	*OKL s hih* *okl S HIH*

Progeny phenotype	Genotype	Recombination event
Oakleaf pin *	*OKL s hih* *okl s hih*	No cross-over, parental alleles
Hose in Hose thrum *	*okl S HIH* *okl s hih*	No cross-over, parental alleles
Wild type pin	*okl s hih* *okl s hih*	Single cross-over between *okl-s*
Hose in Hose *Oakleaf* thrum	*OKL S HIH* *okl s hih*	Single cross-over between *OKL-S*
Hose in Hose *Oakleaf* pin	*OKL s HIH* *okl s hih*	Single cross-over between *s-HIH*
Wild type thrum	*okl S hih* *okl s hih*	Single cross-over between *S-hih*
Hose in Hose pin	*okl s HIH* *okl s hih*	Double cross-over between *okl-s and s-HIH*
Oakleaf thrum *	*OKL S hih* *okl s hih*	Double cross-over between *OKL-S and S-hih*

Footnote: Phenotype and corresponding genotypes of pin and thrum parents plants used in the three-point cross[38], and their progeny with detail of pollen meiotic recombination events that resulted in the observed progeny classes. Plants used for PCR linkage analysis (Fig 3b) indicated by *.

Supplementary Table 5 | Expression analysis of *S* locus and flanking region genes.

Region[*]	Gene name	Gene number[†]	FPKM[‡] Pin	FPKM[‡] Thrum
Left border region	$SFG^{L}7$	*PvLHv1*_131350	0.989	1.481
	$SFG^{L}6$	*PvLHv1*_131340	15.327	34.242
	$SFG^{L}5$	*PvLHv1*_131330	0.756	0.585
	$SFG^{L}4$	*PvLHv1*_313320	14.056	16.252
	$SFG^{L}3$	*PvLHv1*_131310	7.569	7.808
	$SFG^{L}2$	*PvLHv1*_131300	30.037	27.433
	$SFG^{L}1$	*PvLHv1*_131297	0.023	0.111
S locus and tandem repeat region	CFB^{TL}	*PvLHv1*_131293	0.049	0.043
	--- [§]	*PvLHv1*_131290	0.000	0.005
	CCM^{T}	*PvLHv1*_035270	0.000	0.097
	GLO^{T}	*PvLHv1*_169680	0.000	51.616
	--- [§]	*PvLHv1*_123950	0.000	0.000
	--- [§]	*PvLHv1*_123960	1.505	0.666
	CYP^{T}	*PvLHv1*_123970	0.000	1.012
	--- [§‖]	*PvLHv1*_123980	0.019	0.035
	PUM^{T}	*PvLHv1*_124000	0.000	0.600
	KFB^{T}	*PvLHv1*_124010	0.000	0.403
	CFB^{TR}	*PvLHv1*_124020	0.000	0.000
Right border Region	$SFG^{R}1$	*PvLHv1*_124030	58.731	55.394
	$SFG^{R}2$	*PvLHv1*_124040	282.556	260.819
	$SFG^{R}3$	*PvLHv1*_124050	8.649	9.313
	$SFG^{R}4$	*PvLHv1*_124060	31.040	27.490
	$SFG^{R}5$	*PvLHv1*_124070	18.306	13.688
	$SFG^{R}6$	*PvLHv1*_124080	0.000	0.013
	$SFG^{R}7$	*PvLHv1*_124090	10.621	10.675
	$SFG^{R}8$	*PvLHv1*_123100	11.244	11.358

Footnotes: * See Fig. 2 and Supplementary Fig. 1 and 2
[†] See Supplementary Fig. 2
[‡] Fragments per kb of transcript per million fragments mapped.
[§] Gene models not used in Fig. 4a (see Supplementary Fig. 2b)
[‖] 123980 is located within the intron of 123970

Supplementary Table 6 | Bayesian relaxed-clock phylogenetic analysis.

a

Species	Order	Family	Major lineage	Gene name	Clade (*DEF/GLO*)	Accession no. (GenBank)
Antirrhinum majus	Lamiales	Plantaginaceae	Asterids	*AmGLO* *AmDEF*	*GLO* *DEF*	AB516403.1 X52023.1
Arabidopsis thaliana	Brassicales	Brassicaceae	Rosids	*AlPI* *AlAP3*	*GLO* *DEF*	NM_122031.3 NM_115294.5
Arabidopsis lyrata	Brassicales	Brassicaceae	Rosids	*AtPI* *AtAP3*	*GLO* *DEF*	XM_002871885.1 XM_002877924.1
Petunia hybrida	Solanales	Solanaceae	Asterids	*PhFBP1* *PhPMADS1*	*GLO* *DEF*	M91190.1 X69946.1
Primula denticulata	Ericales	Primulaceae	Asterids	*PdGLO* *PdGLO*^{T}	*GLO* *GLO*	KT257671 KT257675
Primula elatior	Ericales	Primulaceae	Asterids	*PeGLO* *PeGLO*^{T}	*GLO* *GLO*	KT257670 KT257677
Primula farinosa	Ericales	Primulaceae	Asterids	*PfGLO* *PfGLO*^{T}	*GLO* *GLO*	KT257673 KT257678
Primula veris	Ericales	Primulaceae	Asterids	*PveGLO* *PveGLO*^{T}	*GLO* *GLO*	KT257669 KT257674
Primula vialii	Ericales	Primulaceae	Asterids	*PviGLO* *PviGLO*^{T}	*GLO* *GLO*	KT257672 KT257676
Primula vulgaris	Ericales	Primulaceae	Asterids	*PvGLO* *PvGLO*^{T} *PvDEF*	*GLO* *GLO* *DEF*	DQ381428.1 KT257666 DQ381427.1

Footnote: Sequences used in Bayesian relaxed-clock phylogenetic analysis

b

Divergence	Age ranges (MYA)	Reference(s)	Mean age applied (SD) *
Arabidopsis thaliana and *A. lyrata*	8.0 - 17.9	Beilstein *et al*. (2010)[72]	12.950 (3.009)
DEFICIENS (DEF) and *GLOBOSA (GLO)*	213.5 - 336.0	Hernández-Hernández *et al*. (2007)[70], Kim *et al*. (2004)[3], Aoki *et al*. (2004)[71]	274.750 (37.237)
Asterids (Ericales, Solanales and Lamiales)	101.0 - 119.0	Bell *et al.* (2010)[46], Magallón *et al*. (2015)[45]	110.000 (5.472)
Lamiales and Solanales	78.0 - 102.5	Bell *et al*. (2010)[46], Magallón *et al*. (2015)[45]	90.250 (7.447)
Rosids and Asterids (Brassicales, Ericales, Solanales and Lamiales)	111.0 - 126.5	Bell *et al*. (2010)[46], Magallón *et al*. (2015)[45]	118.750 (4.712)

Footnote**:** ***** mean age in million years ago (MYA) and standard deviation (SD) are shown to 3 decimal places as applied in BEAST v2.1.2 (with normal distribution priors). Age ranges encompass upper and lower boundaries as reported in the original studies. Divergence times are those generated using lognormal distributions for the fossil priors[46], and the uncorrelated lognormal (UCLN) time-tree[45].

Supplementary Sequence Analysis

Sequence Analysis 1 | Comparison of *S* locus flanking sequences.

PCR products shown in Fig 1c. were sequenced and alignments are shown. **a,** Sequence alignment of PCR products obtained with S^{LH1} haplotype left border primers (SLB-F and SLB-R) from thrum (T), long homostyle (LH) and short homostyle (SH) DNA. The haplotype profile of each plant is shown. **b,** as in part **a** but using S^{LH1} right border primers (SRB-F and SRB-R). **c,** as in part **a** but showing alignment of PCR sequences obtained using *s* haplotype primers (s-F and s-R) with pin (P), thrum (T) and short homostyle (SH) DNA. Identical bases are indicated (*).

a

```
SLH1 LB T (S/s)         GATTCAGATGTTTAACACTTCATATATACTTGTAGCGGGTATACCCACTAATTTAACAAATAGTATTAACTATTTTATTTTAAACTAACCGTAGGTCAAA
SLH1 LB LH (SLH1/SLH1)  GATTCAGATGTTTAACACTTCATATATACTTGTAGCGGGTATACCCACTAATTTAACAAATAGTATTAACTATTTTATTTTAAACTAACCGTAGGTCAAA
SLH1 LB SH (SSH1/s)     GATTCAGATGTTTAACACTTCATATATACTTGTAGCGGGTATACCCACTAATTTAACAAATAGTATTAACTATTTTATTTTAAACTAACCGTAGGTCAAA
                        ****************************************************************************************************

SLH1 LB T (S/s)         CAATCGTTTTAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCAAATATTTAGTCGTGAAAGTGAAACTGTGAACTTGGT
SLH1 LB LH (SLH1/SLH1)  CAATCGTTTTAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCAAATATTTAGTCGTGAAAGTGAAACTGTGAACTTGGT
SLH1 LB SH (SSH1/s)     CAATCGTTTTAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCAAATATTTAGTCGTGAAAGTGAAACTGTGAACTTGGT
                        ****************************************************************************************************

SLH1 LB T (S/s)         CGGTCTTCTTCTCTGTACCATAGATTGCCTTGCATGCGAGATATTAAATAATTTTGGATGATTACATAGATAGATACTTTTGAGTTTTCACACACTAGCT
SLH1 LB LH (SLH1/SLH1)  CGGTCTTCTTCTCTGTACCATAGATTGCCTTGCATGCGAGATATTAAATAATTTTGGATGATTACATAGATAGATACTTTTGAGTTTTCACACACTAGCT
SLH1 LB SH (SSH1/s)     CGGTCTTCTTCTCTGTACCATAGATTGCCTTGCATGCGAGATATTAAATAATTTTGGATGATTACATAGATAGATACTTTTGAGTTTTCACACACTAGCT
                        ****************************************************************************************************

SLH1 LB T (S/s)         ATAAACATATATACTATGACTACGTTCAAAAAAAATTAAAAAAATACTATAACTAGGAAAAAAAATTGAAAAATGATCCGGTATGAATCTGACGCCAGTT
SLH1 LB LH (SLH1/SLH1)  ATAAACATATATACTATGACTACGTTCAAAAAAAATTAAAAAAATACTATAACTAGGAAAAAAAATTGAAAAATGATCCGGTATGAATCTGACGCCAGTT
SLH1 LB SH (SSH1/s)     ATAAACATATATACTATGACTACGTTCAAAAAAAATTAAAAAAATACTATAACTAGGAAAAAAAATTGAAAAATGATCCGGTATGAATCTGACGCCAGTT
                        ****************************************************************************************************

SLH1 LB T (S/s)         GAAACTGATCCTAATACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTGAAATCCGATCATAATACCAAAAACCGAAATCCAAAGAAAATTG
SLH1 LB LH (SLH1/SLH1)  GAAACTGATCCTAATACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTGAAATCCGATCATAATACCAAAAACCGAAATCCAAAGAAAATTG
SLH1 LB SH (SSH1/s)     GAAACTGATCCTAATTCTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTGAAATCCGATCATAATACCAAAAACCGAAATCCAAAGAAAATTG
                        *************** ************************************************************************************

SLH1 LB T (S/s)         GTAAATAGCGAAAA   514
SLH1 LB LH (SLH1/SLH1)  GTAAATAGCGGAAA   514
SLH1 LB SH (SSH1/s)     GTAAATAGTGGAAA   514
                        ******** * ***
```

b

```
SLH1 RB T (S/s)         AACCAGGTCGTAACAAAACCCAACTTTTCCAGTTAATGGATGATAGTAATTTAGTAGCGGTCCTTATTAGTTACCCATTATTTTAAATTTTGATTATTCA
SLH1 RB LH (SLH1/SLH1)  AACCAGGTCGTAACAAAACCCAACTTTTCCAGTTAATGGATGATAGTAATTTAGTAGCGGTCCTTATTAGTTACCCATTATTTTAAATTTTGATTATTCA
SLH1 RB SH (SSH1/s)     AACCAGGTCGTAACAAAACCCAACTTTTCCAGTTAATGGATGATAGTAGTTTAGTAGCGGTCCTTATTAGTTACCCATTATTTTAAATTTTGATTATTCA
                        ************************************************ ***************************************************

SLH1 RB T (S/s)         CTTGATGATCATCTATGTTCCACAATCGATCTTTATCCGATAAATATCTAGTGGACAAAACAGAAGACTACATGTGGAATATTTTCCTACTTTTTTCCTT
SLH1 RB LH (SLH1/SLH1)  CTTGATGATCATCTATGTTCCACAATCGATCTTTATCCGATAAATATCTAGTGGACAAAACAGAAGACTACATGTGGAATATTTTCCTACTTTTTTCCTT
SLH1 RB SH (SSH1/s)     CTTGATGATCATCTATGTTCCACAATCGATCTTTATCCGATAAATATCTAGTGGACAAAACAGAAGAATACATGTGGAATATTTTCCTACTTTTTTCCTT
                        ******************************************************************* ********************************

SLH1 RB T (S/s)         TTTTCTCAAGAAAAGACTCCAGGGTCTCGTTCTGGAAAAATATAATTAATTAGTTTATAATCGAAGTTCATAACTTTATATGATCGAATTGGACAGATTC
SLH1 RB LH (SLH1/SLH1)  TTTTCTCAAGAAAAGACTCCAGGGTCTCGTTCTGGAAAAATATAATTAATTAGTTTATAATCGAAGTTCATAACTTTATATGATCGAATTGGACAGATTC
SLH1 RB SH (SSH1/s)     TTTTCTCAAGAAAAGACTCCAGGGTCTCGTTCTGGAAAAATATAATTAATTAGTTTATAATCGAAGTTCATAACTTTATATGATCGAATTGGACAGATTC
                        ****************************************************************************************************

SLH1 RB T (S/s)         AGATGTTTAACACTTCATACTTGTAGCGGGTATACCCACTAATCTAACAAATAGTATTAACTATTTTATTTTAAACTAACCGTAGGTCAAACAATCGTTT
SLH1 RB LH (SLH1/SLH1)  AGATGTTTAACACTTCATACTTGTAGCGGGTATACCCACTAATCTAACAAATAGTATTAACTATTTTATTTTAAACTAACCGTAGGTCAAACAATCGTTT
SLH1 RB SH (SSH1/s)     AGATGTTTAACACTTCATACTTGTAGCGGGTATACCCACTAATCTAACAAATAGTATTAACTATTTTATTTTAAACTTACCGTAGGTCAAACAATCGTTT
                        ***************************************************************************** **********************

SLH1 RB T (S/s)         TAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCAAATATTTAGTCGTGAAAGTGAAACTGTGAACTTGGTCGGTCTTCT
SLH1 RB LH (SLH1/SLH1)  TAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCAAATATTTAGTCGTGAAAGTGAAACTGTGAACTTGGTCGGTCTTCT
SLH1 RB SH (SSH1/s)     TAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCAAATATTTAGTCGTGAAAGTGAAACTGTGAACTTGGTCGGTCTTCT
                        ****************************************************************************************************

SLH1 RB T (S/s)         TCTCTGTACCATAGATTGCCTTGCATGCGAGATATTAAATAATTTTGGATGATTACATAGATAGATATTTTTGAGTTTTGACACACTAACTATAAACATA
SLH1 RB LH (SLH1/SLH1)  TCTCTGTACCATAGATTGCCTTGCATGCGAGATATTAAATAATTTTGGATGATTACATAGATAGATATTTTTGAGTTTTGACACACTAACTATAAACATA
SLH1 RB SH (SSH1/s)     TTTCTGTACCATAGATTGCCTTGCATGCGAGATATTAAATAATTTTGGATGATTACATAGATAGATATTTTTGAGTTTTGACACACTAACTATAAACATA
                        * **************************************************************************************************

SLH1 RB T (S/s)         TATACTCTGACTACGTTCAAAAAAAATTAAAAAAATATTATAGCTAGGAAAAAAAATTGAAAAATGATCCGGTATGAATCTGACGCCAGTTGAAACTGAT
SLH1 RB LH (SLH1/SLH1)  TATACTCTGACTACGTTCAAAAAAAATTAAAAAAATATTATAGCTAGGAAAAAAAATTGAAAAATGATCCGGTATGAATCTGACGCCAGTTGAAACTGAT
SLH1 RB SH (SSH1/s)     TATACTCTGACTACGTTCAAAAAAATTTAAAAAAATACTATAGCTAGGAAAAAAAATTGAAAAATGATCCGGTATGAATCTGACGCCAGTTGAAACTGAT
                        ************************* *********** **************************************************************

SLH1 RB T (S/s)         CCTAATACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTGTATCAATTTACATTAATTAAATGGCTTTAAACGAGTC   785
SLH1 RB LH (SLH1/SLH1)  CCTAATACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTGTATCAATTTACATTAATTAAATGGCTTTAAACGAGTC   785
SLH1 RB SH (SSH1/s)     CCTAATACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTGTATCAATTTACATTAATTAAATGGCTTTAAACGAGTC   785
                        *************************************************************************************
```

c

```
s P (s/s)         TTGACCAATAAGAATCTCGCACACCTATCCTCTCTCTTGGTTCAAGAGTATTTTGTATACATTCTTGAGGATAATTATCAGTACTGGATGCACTAACCGC
s T (S/s)         TTGACCAATAAGAATCTCGCACACCTATCCTCTCTCTTGGTTCAAGAGTATTTTGTATACATTCTTGAGGATAATTATCAGTACTGGATGCACTAACCGC
s SH (S^SH1/s)    TTGACCAATAAGAATCTCGCACACCTATCCTCTCTCTTGGTTCAAGAGTATTTTGTATACATTCTTGAGGATAATTATCAGTACTGGATGCACTAACCGC
                  ****************************************************************************************************

s P (s/s)         AGTAGATATACAAAATTTTGAATTGGATTTAGCAAGACCTCGGGATAAATATGCGTAAACCTTGTGATCGCTCTCCAGATTTTTTTGTTGCCTAGAATTG
s T (S/s)         AGTAGATATACAAAATTTTGAATTGGATTTAGCAAGACCTCGGGATAAATATGCGTAAACCTTGTGATCGCTCTCCAGATTTTTTTGTTGCCTAGAATTG
s SH (S^SH1/s)    AGTAGATATACAAAATTTTGAATTGGATTTAGCAAGACCTCGGGATAAATATGCGTAAACCTTGTGATCGCTCTCCAGATTTTTTTGTTCCCTAGAATTG
                  ***************************************************************************************** **********

s P (s/s)         AATGACCCCGCTTGTTTGCTTTTCTCTTTGATAATAATCGGGATTTTCTCTAATTCGGGAAAGTTTTTTAGGCATATCTGCTTAGAGAGTCCGTTCCTGA
s T (S/s)         AATGACCCCGCTTGTTTGCTTTTCTCTTTGATAATAATCGGGATTTTCTCTAATTCGGGAAAGTTTTTTAGGCATATCTGCTTAGAGAGTCCGTTCCTGA
s SH (S^SH1/s)    AATGACCCCGCTTGTTTGCTTTTCTCTTTGATAATAATCGGGATTTTCTCTAATTCGGGAAAGTTTTTTAGGCATATCTGCTTAGAGAGTCCGTTCCTGA
                  ****************************************************************************************************

s P (s/s)         TCGCTGCTTAAGATATATTGGTCATCAATAACCATCCATATAATATCCAAAAACCAATACAATAGTCACTAAAATGGAAGACTTGAAATGTCATTTTTTA
s T (S/s)         TCGCTGCTTAAGATATATTGGTCATCAATAACCATCCATATAATATCCAAAAACCAATACAATAGTCACTAAAATGGAAGACTTGAAATGTCATTTTTTA
s SH (S^SH1/s)    TCGCTGCTTAAGATATATTGGTCATCAATAAACATCCATATAATATCCAAAAACCAATACAATAGTCACTAAAATGGAAGACTTGAAATGTCATTTTTTA
                  ******************************* ********************************************************************

s P (s/s)         TTTCATCCAAAAGTAAGTAATGAAACCACCTGCTGACTGCTGTTTATGTAGGTCTCATGAAATAGAGTGTTTTTAAGAAAATCCTTACTAAAAACGGATG
s T (S/s)         TTTCATCCAAAAGTAAGTAATGAAACCACCTGCTGACTGCTGTTTATGTAGGTCTCATGAAATAGAGTGTTTTTAAGAAAATCCTTACTAAAAACGGATG
s SH (S^SH1/s)    TTTCATCCAAAAGTAAGTAATGAAACCACCTGCTGACTGCTGTTTATGTAGGTCTCATGAAATTGAGTGTTTTTAAGAAAATCCTTACTAAAAACGGATG
                  *************************************************************** ************************************

s P (s/s)         AGGGGTATAAGTGAACGATAGTCCAAGAGATGCATGCAAATAAAAATAAACTAACCAAAATCACGCCAAGAGGTTGAAACTGTGCTAAAACGTATAATAT
s T (S/s)         AGGGGTATAAGTGAACGATAGTCCAAGAGATGCATGCAAATAAAAATAAACTAACCAAAATCACGCCAAGAGGTTGAAACTGTGCTAAAACGTATAATAT
s SH (S^SH1/s)    AGGGGTATAAGTGAACGATAGTCCAAGAGATGCATGCAAATAAAAATAAACTAACCAAAATCACGCCAAGAGGTTGAAACTGTGCTAAAACGTATAATAT
                  ****************************************************************************************************

s P (s/s)         CCGACGGATTCTCTAAAGACAATAGGATTCTCAAGGATAAATCATGTCCAAGCCATTGTATAAAGTCTCTGGAACTTGCATGGATTTCCATAGTGGATAT
s T (S/s)         CCGACGGATTCTCAAAAGACAATAGGATTCTCAAGGATAAATCATGTCCAAGCCATTGTATAAAGTCTCTGGAACTTGCATGGATTTCCATAGTGGATAT
s SH (S^SH1/s)    CCGACGGATTCTCAAAAGACAATAGGATTCTCAAGGATAAATCATGTCCAAGCCATTGTATAAAGTCTCTGGAACTTGCATGGATTTCCATAGTGGATAT
                  *************:**************************************************************************************

s P (s/s)         TGAAATAAATTATATGTCGAAGGTCACTGCCTTTCTTAGTTACATCAAAATATAGTGGCACTATATATAGGATAAGAAATTGTTTAGTTCTACCATTAGT
s T (S/s)         TGAAATAAATTATATGTCGAAGGTCACTGCCTTTCTTAGTTACATCAAAATATAGTGGCACTATATATAGGATAAGAAATTGTTTAGTTCTACCATTAGT
s SH (S^SH1/s)    TGAAATAAATAATATGTCGAAGGTCACTGCCTTTCTTAGTTACATCAAAATATAGTGGCACTATATATAGGATAAGAAATTGTTTAGTTCTACCATTAGT
                  **********:*****************************************************************************************

s P (s/s)         TACATCAAAATAGAAACTTTCAGCAAATAAAAGTCATAAAAGTATACCACGTGAGTGACGTTTGGAGCCTAATACTTTTCCTGTAAA   887
s T (S/s)         TACATCAAAATAGAAACTTTCAGCAAATAAAAGTCATAAAAGTATACCACGTGAGTGACGTTTGGAGCCTAATACTTTTCCTGTAAA   887
s SH (S^SH1/s)    TACATCAAAATAGAAACTTTCAGCAAATAAAAGTCATAAAAGTATACCACGTGAGTGACGTTTGGAGCCTAATACTTTTCCTGTAAA   887
                  ***************************************************************************************
```

Sequence Analysis 2 | Alignment of *s* haplotype with S^{LH1} left and right borders.

Annotated alignment of sequences by Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Coloured text represents the left (blue) and right (green) flanking regions, the tandemly duplicated sequences at the boundary (yellow) and sequence absent from the *s* haplotype (red). **a,** Alignment-1 (Supplementary Fig. 2b) showing the *s* haplotype (8942 nucleotides) with the left *S* locus border sequence from the S^{LH1} haplotype (8922 nucleotides). The *CFB* locus start codon (white text, highlighted green) and stop codons (white text, highlighted red) are shown in antisense. **b,** Alignment-2 (Supplementary Fig. 1b) of the *s* haplotype (8942 nucleotides) with the right *S* locus border sequence from the S^{LH1} haplotype (8909 nucleotides); the *CFB* locus start codon (white text, highlighted green) and stop codons (white text, highlighted red) are shown in antisense. The premature stop codon in the S^{LH1} allele of *CFB* (white text, highlighted blue) caused by the 11 base deletion and sequence changes (white text, highlighted grey) are also shown. **c,** Alignment of the left and right border regions from the S^{LH1} haplotype centred around the *CFB* locus, and corresponding region of the single *CFB* locus from the *s* haplotype. Text colours are as defined in **b**. Annotation of start codons, stop codons, in antisense, and the 11 bp deletion are as defined in **c**. The direction of transcription is indicated by an arrow above the antisense start codon. All sequences are shown in bold, with the exception of introns which are non-bold text. The position of primers used to define the left and right border sequences (Fig. 1c) are labelled and shown by underlining.

a

```
s           TATATAGTTTATATTGTACACTATATTATATATGTATACAATGACATGGTAATTTTATCGTACTAATTAAGATTTAAAACCATATGCTAAATGAAACTAA 100
S^LH1 LB    TATATAGTTTATATTGTACACTATATTATATATGTATACAATGACATGGTAATTTTATCGTACTAATTAAGATTTAAAACCATAGGCTAAATGAAACTAA 100
            ************************************************************************************ ***************

s           TTAAAAAATATAATTATGTACATGATTTTAACGGATAAATGAAACTAACCTCTTAAAAAAATGATTTAAAACTACCGAAATAAAGATAGTTTGTGTTTAA 200
S^LH1 LB    TTAAAAAATATAATTATGTACATGATTTTAACGGCTAAATGAAACTAACCTCCTAAAAAAATGATTTAAAACTACCGAAATAAAGATAGTTTGTGTTTAA 200
            ********************************** ***************** ***********************************************

s           TCAACACTAATTTTAATTATTATTTATATTTCTGTAATAACGAAATTTTAAACCCTTCAATGTAATTTAAAAACATACAAAGTATCGATACGTTATTTTT 300
S^LH1 LB    TCAACACTAATTTTAATTATTATTTATATTTCTGTACTAACGAAATTTTAAACCCTTCAATGTAATTTAAAAACATACAAAGTATCGATACGTTATTTTT 300
            ************************************ ***************************************************************

s           ACGGCTAAATAAAACTAACCTACTAAAAATCAATTTAAACTAGTAGAATTATAATAAAAAAAATTATAACAAAATTTAAAAAACTAACCTTACACGGTGC 400
S^LH1 LB    ACGGCTAAATAAAACTAACCTACTAAAAATCAATTTAAACTAGTAGAATTATAATAAAAAAAATTATAACAAAATTTAAAAAACTAACCTTACACGGTGC 400
            ****************************************************************************************************

s           AAAGACCTTTATAAATTTTGGGTAACGAGTAATTCTAATGTACAAATTTTTATAAATAGGTATTATTTTTATATATTTTTTTGCATATGATGGTATTTTT 500
S^LH1 LB    AAAGACCTTTATAAATTTTGGGTAACGAGTAATTCTAATGTACAAATTTTTATAAATAGGTATTATTTTTATATATTTTTTTGCATATGATGGTATTTTT 500
            ****************************************************************************************************

s           TAGTGGTTGGGTTATGTGAAAATGTTGCATATATATATGAGTGAATATTGAATCGAATAGATAAGGAATATGATTGGTGTAAGAAAAAGACATATTTTTG 600
S^LH1 LB    TAGTGGTTGGGTTATGTGAAAATGTTGCATATATATG--AGTGAATATTGAATCGAATAGATAAGGAATATGATTGGTGTAAGAAAAAGACATATTTTTG 598
            ************************************   *************************************************************

s           GATAAGAAAGTAGATTTCATATTTATGAAAAAATAAATAG-AAAAAAATATATATTTCAGTTGGTATATAGTAACTACTAAAATTAGTTACTAAATTCTA 699
S^LH1 LB    GATAAGAAAGTAGATTTCATATTTATGAAAAAATAAATAGAAAAAAAATATATATTTCAGTTGGTATATAGTAACTACTAAAATTAGTTACTAAATTCTA 698
            **************************************** ***********************************************************

s           AGTAAGTTACTAACATAGTTGGTATTTACCTACTAAAATGTGGTAGTTAGTAATCAATTAGTATTTAGCAACCGATAAAATAGTAATAACTAATTAGTTG 799
S^LH1 LB    AGTAAGTTACTAACATAGTTGGTATTTACCTACTAAAATGTGGTAGTTAGTAATCTATTAGTATTTAGCAACCGATAAAATAGTAGTAACTAATTAGTTG 798
            ******************************************************* ***************************** **************

s           GTAAATTACTAACTGCCGAAATTTAAGTTACTATTCAGTTACTACTTACCAACTGAAAAGTGTAAGTTACAGGTTAGTTACTATTTAGTGACTAATAATA 899
S^LH1 LB    GTAAATTACTAACTGCCGAAATTTAAGTTACTATTCAGTTACTACTTACCAACTGAAAAGTGTAAGTTACAGGTTAGTTACTATTTAGTGACTAATAATA 898
            ****************************************************************************************************

s           TGGTAGTTGCAAATTACTTGTTACGTCATAAAAATATGAAATTTTAGTTAGAGATTCGGTTAGTGATTAGTAACTACCAAATTAGTTGGAATTTTAAGTT 999
S^LH1 LB    TGGTAGTTGCAAATTACTTGTTACGTGATAAAAATATGAAATTTTAGTTAGAGATTCGGTTAGTGATTAGTAACTACCAAATTAGTTGGAATTTTAAGTT 998
            ************************** *************************************************************************

s           ACTATGTGGGCATTTTATTGTG-TGTGAAGCGTTGATCGGTATATAATGCGATATTCAAACTCGACACCAATGTATGATCTGGGTTGAAGAAGGCTAAAC 1098
S^LH1 LB    ACTATGTGGGCATTTTATTGTAGTGTGAAGCGTTGATTGGTATATAATGCGATATTCAAACTCGACACCAATGTATGATCTGGGTTGAAGAAGGCTAAAC 1098
            *********************  ************** **************************************************************

s           CACTTTAGCACGCACAAGTGCAGCATATGCTAATCTAAATTTTGAAAAAAATGAAAATTTGCACTTATGTATATTTTTTTAATTTTTTGTTTAGTGCCCC 1198
S^LH1 LB    CACTTTAGCACGCACAAGTGCAGCATATGCTAATCTAAATTTTGAAAAAAATGAAAATTTGCACTTATGTATATTTTTTTAATTTTTTGTTTAGTGCCCC 1198
            ****************************************************************************************************

s           TCCCTGTCGCACGCGTCTAAGCCAACTAGCTCTTTTGGCTTTGGCTAGTGATGACCCTGCATGAGTGGAATTTTTGTAAAAGTTATAACGGCCTATGATT 1298
S^LH1 LB    TCCCTGTCGCACGCGTCTAAGCCAACTAGCTCTTTTGGCTTTGGCTAGTGATGACCTTGCATGAGTGGAATTTTTGTAAAAGTTATAACGGCCTATGATT 1298
            ******************************************************** *******************************************

s           AGATTTTTTTACCATGCATTTGGTTTAAATACCGGTACAGGGTTCGAATCTTACTCATATCCCATGAGTGATATCATTTGATCCCCTTAATAAGATTTTA 1398
S^LH1 LB    AGATTTTTTTACCATGCATTTGGTTTAAATACCGGTACAGGGTTCGAATCTTACTCATATTCCATGAGTGATATCATTTGATCCCCTTAATAAGATTTTA 1398
            ************************************************************ ***************************************
```

```
s          AAAAGGAAAAAAAATGTCTTTTGATACCTTGGAGTATATAGTAAGGAAAATATTAATATTATTCCCCACTTACAAGTCAATCCTAACTCGCCACCAAATA 1498
S^LH1 LB   AAAAGGAAAAAAAATGTCTTTTGATACCTTGGAGTATATAGTAAGAAAAATATTAATATTATTCCCCACTTACAAGTCAATCCTAACTCGCCACCAAATA 1498
           ********************************************* ******************************************************

s          GGCATACACATAGAGAACTATACATATGTGGATTCTAAATATGTTAGCATAAGAAAGTTGTGATTGAAGTGGAAAATAAGTTTAACTCAATGTTTCGAGT 1598
S^LH1 LB   GGCATACACATAGAGAACTATACATATGTGGATTCTAAATATGTTAGCATAAGAAAGTTGTGATTGAAGTGGAAAATAAGTTTAACTCAATGTTTCGAGT 1598
           ****************************************************************************************************

s          AAAGTTGCCATCTAATTATCCTTGGATTGTCAAGTTTCCTTATGTTGAACATCGATTTATAAATTTAATTGATGTTATTATAAAAATCCGGATTTTCGTA 1698
S^LH1 LB   AAAGTTGCCATCTAATTATCCTTGGATTGTCAAGTTTCCTTATGTTGAACATCGATTTATAAATTTAATTGATGTTATTATAAAAATCCGGATTTTCGTA 1698
           ****************************************************************************************************

s          GAAAAATGCATGTGTTAAGAGAAATTATTGTCTCATATCGTCTAGATATGTTAGTACATATATGGTATATAATCTAATAAAATCCCCATTACTATCAACA 1798
S^LH1 LB   GAAAAATGCCTGTGTAAAGAGAAATTATTGTCTCATATCGTCTAGATATGTTAGTACATATATGGTATATAATCTAATAAAATCCCCATTACTATCAACA 1798
           ********* ***** ************************************************************************************

s          ATACTTTTTAAAGTCGTGCGGCCCAACAAGCGAGTATGAAATTGGGCCAAATCGGACAGTATTGGTGGTAGGAAGCAGATGGTGACGGCCTGATCGGGAG 1898
S^LH1 LB   ATACTTTTTAAAGTCGTGCGGCCCAACAAGCGAGTATGAAATTGGGCCAAATCGGACAGTATTGGTGGTAGGAAGCAGATGGTGACGGCCTGATCGGGAG 1898
           ****************************************************************************************************

s          ATCGGGTCGTTACAATTGGTATCAGAACCGCTATCTAGCAGATGGTTTGTGAAACTTCCCCCTCCCCGGTAAGGACGGCGTGTCAGACACGATGACGTCG 1998
S^LH1 LB   ATCGGGTCGTTACAATTGGTATCAGAACCGCTATCTAGCAGATGGTTTGTGAAACTTCCCCCTCCCCGGTAAGGACGGCGTGTCTG--ACGATGACGTCG 1996
           ************************************************************************************ *  ************

s          TGCATGAAGGGGGAAGAGTGTAAAAATCCGGATTTTCATACG-AAATGCCGGTGTAAAGAGAAAGTATTGTCTCACATCGATGACGAAATTGGGCCAAAT 2097
S^LH1 LB   TGCATGAAGGGGGAAGAGTGTAAAAATCCGGATTTTCATACGAAAATGCCGGTGTAAAGAGAAAGTATTGTCTCACATCGATGACGAAATTGGGTCAAAG 2096
           ****************************************** *************************************************** ****

s          TGGACAATATTAGTGGTAGAAAGCAGATAGTGACGGTCTGATCCGAAAACCAAATCGTTACAGTTATCCTCTGTAATCTATAGTACTATAACCTCCATTT 2197
S^LH1 LB   GGGACAATATTAGTGGTAGAAAGCAGATAGTGACGGTCTGGTCCGAAAACCAAGTCGTTACAGTTATCCTCTGTAATCTATAGTACTATAACCTCCATTT 2196
            *************************************** ************ **********************************************

s          TCCTATGAAATGATGTGGTGATTGTACCCGTGAGCAGTGAGAGTCCGTAACTATACTGATGAATGTAATTGTATCGATCGGCATTGGTATAATGTGACGT 2297
S^LH1 LB   TCCTATGAAATGATGTGGAGATTGTACCCGTGAGCAGTGAGAGTCCGTGACTATGCTGATGAATGTAATTGTATCGATCGGCATTGGTATAATGTGACGT 2296
           ****************** ***************************** ***** *********************************************

s          AGAGTCACGAAAAACTATTTGCACATTCACTCTTTGTGACAGAAAATGAGGATAAATATACAGTTTTGAATCTTTTGATCAATGTCGTTTTTGTCTTCCC 2397
S^LH1 LB   AGAGTCACGAAAAACTATTTGCACATTCACTCTTTGTGACAGAAAATGAGGATAAATATACAGTTTTGAATCTTTTGATCAATGTCGTTTTTGTCTTCCC 2396
           ****************************************************************************************************

s          TTCTGTTACAAACAGTGTATGAAAAAGTAAAAAGTGTGTACATATAATTCACAAACAATGTGTGAAAGAGTTAAAACTGTAGAGTCACTGACGCTAATGT 2497
S^LH1 LB   TTATGTTACAAACAGTGTATGAAAAAGTAAAAAGTGTGTACATATAATTCACAAACAATGTGTGAAAGAGTTAAAACTGTAGAGTCACTGACGCTAATGT 2496
           ** *************************************************************************************************

s          TGACCTATAATTCAGGTGGATTCTTTAGATAAAATATCCGAAGTATACCAAGTATTGTCGCCTATCGTTTTAAGTTCATAATTATGAAATGATTGGGTTT 2597
S^LH1 LB   TGACCTATAATTCAGGTGGATTCTTTAGATAAAATATCCGAAGTATACCAAGTATTGTCGCCTATCGTTTTAAGTTCATAATTATGAAATGATTGGGTTG 2596
           ***************************************************************************************************

s          GTCTTAAAAATGGCATGGTTAGAAAGTCTCTTCCAAATAGCATCATATTTAAGTAACTTCTTTACCAATAGATACATATATTGTAAAAAGAAAAACTCAA 2697
S^LH1 LB   GTCTTAAAAATGGCATGGTTAGAAAGTCTCTTCCAAATAGCATCATATTTAAGTAACTTCTTTACCAATAGATACATATATTGTAAAAAGAAAAACTCAA 2696
           ****************************************************************************************************

s          AATACAGAAAAATCAAATAGGCGCGACTACATGAAATATTATTCATATCTCTTAGCAAAATTGGCCGGAGAAAAATTGCATTGACTTTAGCGGTCGATAG 2797
S^LH1 LB   AATACAGAAAAATCAAATAGGCGCGACTACATGAAATATTATTCATATCTCTTAGCAAAATTGGCCGGAGAAAAATTGCATTGACTTTAGCGGTCGATAG 2796
           ****************************************************************************************************

s          ATACTTATTTTATCTTAATTTAAGACCTTAAAGTCAAAAGTTATCGTTGCAACGCATAATTTG------------------------------------- 2860
S^LH1 LB   ATACTTATTTTATCTTAATTTAAGACCTTAAAGTCAAAAGTTATCGTTGCAACGCATAATTTGGAGAGCTATAAGTCAAAAGTTATCGTTGCAACGCATA 2896
           ***************************************************************

s          -----GAGAGCTATAACACGGTTTGCTAAATCATTCTTAACTAGTTAACCTTTTCAATACTTGTTAACAAAGAACATACTGATCCCATGTTTGATTGTAA 2955
S^LH1 LB   ATTTGGAGAGCTATAACACGGTTTGCTAAATCATTCTTAACTAGTTAACCTTTTCAATACTTGTTAACAAAGAACATACTGATCCCATGTTTGATTGTAA 2996
                ***********************************************************************************************

s          ATGCAAACCCGTCCTGCAAGTATAAACCCAAACTGACACCTAAAAACCGCGAGATTAGTCTTAAACAACAAAACAGTCTATTTTCATAAAAAAATAAAAAA 3055
S^LH1 LB   ATGCAAACCCGTCCTGCAAGTATAAACCCAAACTGACACCTAAAAACCGCGAGATTAGTCTTAAACAACAAAACAGTCTATTTTCATAAAAAAAAATAAAA 3096
           ********************************************************************************************* * ****

s          AAATAGTCA-AACTCTACTTCTGATTAATAAAAAACTGATCACCAAGATTATACATAATAACATAAAGTTATCTACTAATCTAATTATGTCACATAGAAG 3154
S^LH1 LB   AAATAGTCAGAACTCTACTTCTGATTAATAAAAAACTGATCACCAAGATTATACATAATAACATAAAGTTATCTACTAATCTAATTATGTCACATAGAAG 3196
           ********* ******************************************************************************************

s          GGAACTTTTGGAAATTTAATGAGAAAGAGGAGGAAAAGGAAGAAAACAATTAGAAATTGGATAACAAAACCCCTTCGTAATTTTGAGTTCGAGTCCTTCA 3254
S^LH1 LB   GGAACTTTTGGAAATTTAATGAGAAAGAGGAGGAAAACGAAGAAAACAATTAGAAATTGGATAACAAAACCCCTTCGTAATTTTGAGTTCGAGTCCTTCA 3296
           ************************************* **************************************************************

s          CTCCTTGAATCACAACCTCTAAGCCTGAGGCTCTTCTAAGACAAGAGTTTCTACTAGTGACCATAATTGAATTTTCATGAGAAAACCGGTAATTCACCCC 3354
S^LH1 LB   CTCCTTGAATCACAACCTCTAAGCCTGAGGCTCTTCTAAGACTGGAGTTTCTACTAGTGACCATAATTGAATTTTCATGAGAAAACCGGTAATTCACCCC 3396
           ******************************************  ********************************************************

s          ATAGTAGTAATCTAGTTTACATCTTCCCGAGCTATCAATAATTTCAACCTCGAATCCCGGGGAGAGGGGTTTCCCAACAACTTGTACGTGTGTGATACTG 3454
S^LH1 LB   ATAGTAGTAATCTAGTTTACATCTTCCCGAGCTATCAATAATTTCAACCTCGAATCCCGGGGAGAGGGGTTTCCCAACAACTTGTACGTGTGTGATACTG 3496
           ****************************************************************************************************

s          CAAATTGAAAG-AGGAAAAAAAATCATTAATCCGAGTTACAAAAAAATATTGTGCAAAAAAATACTAATCATGACAAGCTTGCTACATAAAAAAGATCTGA 3553
S^LH1 LB   CAAATTGAAAGGGAAAAAAAAAATCATTAATCCGAGTTACAAAAAAATATTGTGCAAAAAAATACTAAT---GACAAGCTTGCTACATAAAAAAGATCTGA 3593
           ***********    *****************************************************   *****************************

s          GTGGACATTAGACAAAGCAATATATGAGTGAAGAATGAATTTGTAGTGCTAAAATTTTAAGAAGAAGGAATTTAGCAAACTAACCGTAAATAGTATAAGC 3653
S^LH1 LB   GTGGACATTAGACAAAGCAATATATGAGTGAAGAATGAATTTGTAGTGCTAAAATTTTAAGAAGAAGGAATTTAGCAAACTAACCGTAAATAGTATAAGC 3693
           ****************************************************************************************************

s          CATCAGTTTCGTGGATTTGTACCCGTCCCAATAGTTCCATTTGTAAGATTCCACCAACGCAAAGAACAGGTTCAGGGAGCTTAAATTTTTGCAAGCAGTT 3753
S^LH1 LB   CATCAGTTTCTTGGATTTGTACCCGTCCCAATAGTTCCATTTGTAAGATTCCACCAACGCAAAGAACAGGTTCAGGGAGCTTAAATTTTTGCAAACAGTT 3793
           ********** *********************************************************************************** *****

s          CTCCTAATTTATATTTAAGCAAAAAGTAAAAAGAATCATTATCTTGAATTGATATGAAAGTTACAAAAACTTAAAAGATCACAGTTGGCATGACATACCT 3853
S^LH1 LB   CTCCTAATTTATATTTAAGAAAAAAGTAAAAAGAATCATTATCTTGAATTGATATAAAAGTTACAAAAACTTAAAAGATCACAGTTGGCATGACATACCT 3893
           ******************* *********************************** ********************************************
```

```
S           GAGCCATTCGAAACTCAGGCGAAATATATGTCCATGAGAACGCATTGTCATAATACTTGTGACTAGCTTTACACTTGTACCTTCCATATTTGTACAATTT 3953
S^LH1 LB    GAGCCATTCGAAACTCAGGCGAAATATATGTCCATGAGAACGCATTGTCATAATACTTGTGACTAGCTTTACACTTGTACCTTCCATATTTGTACAATTT 3993
            ****************************************************************************************************

S           TTTTTGGGACTTGGGATAGCCCATCCGAAATCGCACGGCCCTAGCTGAATATATGGAACACCTAGTTTCCAAAGATGGCACTGAGCAAACCCAAAACTTC 4053
S^LH1 LB    TTTTTGAGACTTGGGATAGCCTATCCGAAATCGCACGGCCCTAGCTGAATATATGGAACACCTACTTTCCAAAGATGGCACTGAACAAACCCAAAACTTC 4093
            ****** ************** ****************************************** ******************* ***************

S           CATTAAACAACGGAGTTATGAAAGAATAATTG-AAAAACAGGTCAGAAAATCACCTTGAAAGGGTTGAACGTGAATCTCCGATACAATGCAGATATTGGA 4152
S^LH1 LB    CATTAAACAACGGAGTTATGAAAGAATAATTGAAAAAACAGGTCAGAAAATCACCTTGAAAGGGTTGAACGTGAATCTCCGATACAATGCAGATATTGGA 4193
            ******************************** *******************************************************************

S           AATCAAGTTGTAAACTAGTTTTTGGGGAACTGAGGGATCAGTTTGGCCTGTGCTTGACCAATAAGAATCTCGCACACCTATCCTCTCTCTTGGTTCAAGA 4252
S^LH1 LB    AATCAAGTTGTAGACTAGTTTTTGGGGAACTGAGGGATCAGTTTGGCCTGTGCTTGACCAATAAGAATCTCGCACACCTATCCTCTCTCTTGGTTCAAGA 4293
            ************ ***************************************************************************************

S           GTATTTTGTATACATTCTTGAGGATAATTATCAGTACTGGATGCACTAACCGCAGTAGATATACAAAATTTTGAATTGGATTTAGCAAGACCTCGGGATA 4352
S^LH1 LB    GTATTTTGTATACATTCTTGAGGATAATTATCAGTACTGGATGCACTAACCGCAGTAGATATACACAATTTTGAATTGGATTTAGCAAGACCTCGGGATA 4393
            ***************************************************************** **********************************

S           AATATGCGTAAACCTTGTGATCGCTCTCCAGATTTTTTTGTTGCCTAGAATTGAATGACCCCGCTTGTTTGCTTTTTCTCTTTGATAATAATCGGATTTT 4452
S^LH1 LB    AATATGCGTAAACCTTGTGATCGCTCTCCAGATTTTTTTGTTCCCTAGAATTGAATGACCCCGCTTGTTTGCTTTTTCTCTTTGATAATAATCGGATTTT 4493
            ****************************************** *********************************************************

S           CTCTAATTCGGGAAAGTTTTTTAGGCATATCTGCTTAGAGAGTCCGTTCCTGATCGCTGCTTAAGATATATTGGTCATCAATAACCATCCATATAATATC 4552
S^LH1 LB    CTCTAATTCGGGAAAGTTTTTTAGGCATATCTGCTTAGAGAGTCCGTTCCTGATCGCTGCTTAAGATATATTGGTCATCAATAACCATCCATATAATATC 4593
            ****************************************************************************************************

S           CAAAAACCAATACAATAGTCACTAAAATGGAAGACTTGAAATGTCATTTTTTATTTCATCCAAAAGTAAGTAATGAAACCACCTGCTGACTGCTGTTTAT 4652
S^LH1 LB    CAAAAACCAATACAATAGTCACTAAAATGGAAGACTTGAAATGTCATTTTTTATTTCATCCAAAAGTAAGTAATGAAACCACCTGCTGACTGCTGTTTAT 4693
            ****************************************************************************************************

S           GTAGGTCTCATGAAATAGAGTGTTTTTAAGAAAATCCTTACTAAAAACGGATGAGGGGTATAAGTGAACGATAGTCCAAGAGATGCATGCAAATAAAAAT 4752
S^LH1 LB    GTAGGTCTCATGAAATAGAGTGTTTTTAAGAAAATCCTTACTAAAAACGGATGAGGGGTATAAGTGAACGATAGTCCAAGAGATGCATGCAAATAAAAAT 4793
            ****************************************************************************************************

S           AAACTAACCAAAATCACGCCAAGAGGTTGAAACTGTGCTAAAACGTATAATATCCGACGGATTCTCAAAAGACAATAGGATTCTCAAGGATAAATCATGT 4852
S^LH1 LB    AAACTAACCAAAATCACGCCAAGAGGTTGAAACTGTGCTAAAACGTATAATATCCGACGGATTCGCAAAAGACAATAGGATTCTCAAGGATAAATCATGT 4893
            **************************************************************** ***********************************

S           CCAAGCCATTGTATAAAGTCTCTGGAACTTGCATGGATTTCCATAGTGGATATTGAAATAAATTATA---TGTCGAAGGTCACTGCCTTTCTTAGTTACA 4949
S^LH1 LB    CCAAGCCATTGTATAAAGTCTCTGGAACTTGCATGGATTTCCATAGTGGATATTGAAATAAATAATAGTATGTCGAGAGTCACTGCCTTTCTTAGTTACA 4993
            *************************************************************** ***   ******  **********************

S           TCAAAATATAGTGGCACTATATATAGGATAAGAAATTGTTTAGTTCTACCATTAGTTACATCAAAATAGAAACTTTCAGCAAATAAAAGTCATAAAAGTA 5049
S^LH1 LB    TCAAAATATAGTGGCACTATATATAGGATAAGAAATTGTTTAGTTCTACCATTAGTTACATCAAAATAGAAACTTTCAGCAAATAAAAGTCATAAAAGTA 5093
            ****************************************************************************************************

S           TACCACGTGAGTGACGTTTGGAGCCTAATACTTTTCCTTGTAAAATCTCTATTTTCTACCTCTTTTTGTCTCATTGCTGCTCCGTTGAAAGAGTATGAAA 5149
S^LH1 LB    TACCACGTGAGTGACGTTTGGAGCCTAATACTTTTCCTTGTAAAATCTCTATTTTCTACT---------CTCTTT------------------------- 5159
            ***********************************************************          *** **

S           TAGTTTTACCTTTTTTTTTTTAATTTTCCTTAGTTTGGATGAACCAGGTCGTAACAAAACCCAACTTTTCCAGTTAATGGATGATCGTAATTTAGTAGCG 5249
S^LH1 LB    ------------------------------TAGTTTGGATGAACCAGGTCGTAACAAAACCCAACTTTTCCAGTTAATGGATGATAGTAATTTAGTAGCG 5229
                                          ******************************************************* **************

S           GTCCTTATTAGTTACCCATTATTTTAAATTTTGATTATTCACTTGATGATTATCTATGTTCCACAATCGATCTTTATCCGTTAAATATCTAGTGGACAAA 5349
S^LH1 LB    GTCCTTATTAGTTACCCATTATTTTAAATTTTGATTATTCACTTGATGATCATCTATGTTCCACAATCGATCTTTATCCGATAAATATCTAGTGGACAAA 5329
            ************************************************** ***************************** *******************

S           ACAGAAGAATACATGTGGAATATTTTCCTACTTTTTTCCTTTTTTCTCAAGAAAAGACTCCAGGGTCTCGTTCTGGAAAAATATAATTAATTAGTTTATA 5449
S^LH1 LB    ACAGAAGAATACATGTGGAATATTTTCCTACTTTTTTCCTTTTTTCTCAAGAAAAGACTCCAGGGTCTCGTTCTGGAAAAATATAATTAATTAGTTTATA 5429
            ****************************************************************************************************

S           ATCGAAGTTCATAACGTTATATGATCGAATTGGACTGATTCAGATGTTTAACACTTCATATATACTTGTAGCGGGTATACCCACTAATTTAACAAATAGT 5549
S^LH1 LB    ATTGAAGTTCATAACGTTATATGATCGAATTGGACTGATTCAGATGTTTAACACTTCATATATACTTGTAGCGGGTATACCCACTAATCTAACAAATAGT 5529
            ** ************************************************************************************* ***********

S           ATTAACTATTTTATTTTAATCTAACCGTAGGTCAAACAATCGTTTTAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCA 5649
S^LH1 LB    ATTAACTATTTTATTTTAAACTAACCGTAGGTCAAACAATCGTTTTAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCA 5629
            ******************* ********************************************************************************

S           AATATTTAGTCGTGAAAGTGAAACTGTGAACTTGGTCGGTCTTCTTCTCTGTACCATAGATTGCCTTGCATACGAGATATTAAATAATTTTGGATGATTA 5749
S^LH1 LB    AATATTTAGTCGTGAAAGTGAAACTGTGAACTTGGTCGGTCTTCTTCTCTGTACCATAGATTGCCTTGCATGCGAGATATTAAATAATTTTGGATGATTA 5729
            *********************************************************************** ****************************

S           CATAGATAGATACTTTTGAGTTTTCACACACTAGCTATAAACATATATACTATGACTACGTTCAAAAAAAAATTAAAAAAATACTATAACTAGGAAAAAAA 5849
S^LH1 LB    CATAGATAGATACTTTTGAGTTTTCACACACTAGCTATAAACATATATACTATGACTACGTTCAAAAAAAAATTAAAAAAATACTATAACTAGGAAAAAAA 5829
            ****************************************************************************************************

S           TTTGAAAAATGATCCGGTATGAATCTGACGCCAGTTGAAACTGATCCTAGTACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTGTATCAAT 5949
S^LH1 LB    ATTGAAAAATGATCCGGTATGAATCTGACGCCAGTTGAAACTGATCCTAATACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTG-----AA 5924
             ************************************************ *******************************************     *

S           TTACATTAATTAAATGGCTTTAAACGAGTCCTTGCATGAGTATGTCAGTTAAACAGCATTTACCAATAGTGTTACTCGTAAAATTATTATACTCTTTGTG 6049
S^LH1 LB    ATCCGATCAT--AATACCAAAAACCGAAATCCAAAGAAAATTGGTAAATAGCGGAAAATGTAACATAATTTATAATACTAG----------CTATATATG 6012
             * *  * **  ***  *   ** ***   *       * *  ** * *     *  ** ** **  * *  ** *  **           ** * * **

S           TATCTGCACAAATTCTT-TGGCTCTGTTTGCTTCCCCGGAAAACAATTTCCAGCTAATTTTTTCAAGGAAAACATTTTTATGAAAACTAAT-TGTCTTTC 6147
S^LH1 LB    CATGTGATAATATTCATATGAATATATATTCTTATATG--TTATAACTTCAATTATATATATATAATTGTAAAGTTTATATCATTTCTACTGTATATTAT 6110
            ** **   * **** * **   * * * * ***    *    * ** *** *    ** * *  **    **  *** *** *   *** * * * **

S           ATGATTTTCTTAGTG-ATTGGTTGTGTAGAAAAATTTCATGGAAAATATGAATTATACTTGATAATTTCTCTAACAAACACAAAAATAGATCACATAATA 6246
S^LH1 LB    ATATAATTGTAATTTTAAAGTCAATATACTCTAAATCCAAGTCTTAGATTGTTAATGTTTTCTGATAATTTATTCTTATTCATTTTTTTTAACAACTTGT 6210
            **    ** * * *  *  *    * **    ** * ** *    * **   * **  **  * **   *    *  *  **    *       *

S           GTAAAAAAAATCAACAAACAAATAAAAGTATGAAATTTGGTAGATATTTCCAACATAAAAAGTCATATTTTACCAAACAAAGTCTTTAACAAACAAATAA 6346
S^LH1 LB    TAACGAGCGAACCGCGTCCGCATCAGTGGAAGTGGAAGGCAATTCTTTTAAGTACGAGTCAAACTATCAAGACAAGGTTGAGTTTTTTTTTTTTCATTCA 6310
              *  *   * *  *   *  ** *  * * *      *  *    ***       *   *  *       ** *     *** ***        * * *
```

```
s           AAGGATAATCTGTG--TGAATATAGGATTCTCTTGATATACTGAAAATTAACTGCCCGCGGATATCACGACCACATTCCAATACAAAAAATTGAACAAAG 6444
SLH1 LB     GCTGGTAGTCCAAGAACGCATGCTTCTTTACCATCTTGTTTTGCAAAATACAAATTCATG--TATAAGAACTAATTTTTTTATATATTAATGGTGGGAAA 6408
               * ** **   *   * **      **  * *  * *  ** *** **      *  *  *** *  ** *  **        *  *** *    **

s           AATCAACAATGAATTGTTAATATTTTAAGTAGATGTAAGAGAAAAGATGTTGGAGGAAGGGGGTTAACACCCAAATAAGTTATCACGTTAACACCTTCGA 6544
SLH1 LB     AACCCACTCCTTTATTTATTAATAAAAAGTGGATGAAAAACACCACAAAACAGAACAGAG----TAACAATTGGAAAGATTCTTGAAATTACAG-TACTT 6503
            ** * **       * *    **   **** **** ** * *  * *     **  *  *    *****     * *  ** *     * ***  * *

s           GCTTTTCGAGCTTTTAATCT------CATGTAAATCTAGCCTTTTTTTTTATATCAAATCATAGTATAAATACGCAATTTTAGTTGTAATAACCCGGTAA 6638
SLH1 LB     GATATACTAAAATTAGAACTTTCAGCAATAAAAGAAAAATCTAAAGTTTAAGAACTAAATTGCGTTGCACTTCAGAGTTGAGGGACCAATATTGATATTG 6603
            * * * * *   **  * **       **   **    *  **    *** * * * **     **   * * *  * **   *    ****      *

s           GAGAAGAATATACCAATTGTAAT-AACTGGATTTTCGT--AGAAAATGTCGTGGTGTAGAGAAAAAAGTTAACTCACATCGCCTAGATATTTGAGTATTT 6735
SLH1 LB     ACCAAGAGTCCACCTCTTCCTCTGGTCCACTTCTTCATTCGGTTCAGTTCATCTTCTTTGATAGAGAGAGACAGAGCACAGTCTCTCAATCTTCTTCTTC 6703
               **** *  ***  **     *   *     * *** *    *   *  ** *  * *     * * **  *     **  * **    ** *   * **

s           ATATGGTATAT-TAGTTATGAGTATCATTCTTCTTATCAATAAGGCATTTT------AAAACCGTGCGGCCCAAGAAACAATGACGAAATTGCA------ 6822
SLH1 LB     ATTTCTTAATCGAAGATAAGGAAAACCCTTCTCTTCTAATTTCAAAATATTTCTTAATTTACTGGGCAAATCAATGATGTTTTCTGAGATTAATGTATCA 6803
            ** *  **     ** ** *   * *  *  **** * * *     ** **         ** * **    ***  *    *   ** ***

s           -CCAAACTGGACATTATTGGTAGTAG----GAAACAGATGATGACGGCCTGATCGGAGAACCGAGTCGTTAT--AATTGTTATCAGAGCCGCTATCTAGC 6915
SLH1 LB     TATATACATACCAAATTTACCAGTAGATCCGCACCTGAGTTTGCTTCTCTTTCCACATAGCTTAAAAGATAAGTACCCCTCTGCATCTTCTCTTTTATGC 6903
               * **     **   **    *****    * * * **   **     **   *  * * *  *   * **   *    *   **    * ** *   **

s           CGATTTCACGAGTGATTGAATCCCGAAAGGACGGCATGCTCGGCGAGGACGTCGTGTCTGAACAGGAGAAGAGTGTAACAACCCGT----ATTTTTGTAG 7011
SLH1 LB     TGTTAATTAGCTTTTTAACTTTCC--AAGGACA-CTTACTTTTATCAGACAAGATG-TTAATAAGTTGTAGACAGTCACTTTCTAAGATGATTTTTGTAA 6999
             * *     *  *  *    * **  ******  * * **       ***    **  *  *  **  * ***  ** **   *        *********

s           AAAATGACATGGT-----ATAGAAAATAAAGTTATCTCACCGCGCATATCTATATAAG-----TACATATATAGTATATAACTTAAGAGTATCCTCGTTC 7101
SLH1 LB     AATTTACCACATTTAGTCCTTGTAGAATAGCTCTTGCCTCTGCTTCTATCCATGTATTGCTGGTAATTTTACCATAAAAACCTTGTATAAATTTACCTTC 7099
            **  *  **    *      * * * *  *  *  *  * * **   **** ** **       **  * **   ** * * ***       **   * ***

s           TT-ATCACCAATATTTTTTAAAACTGTGCGGCACGAGAAGGAATGACGAAATTGGCCAAAACGGACAATATTGATGATAAGAAACAGATGGTGGTGGTCT 7200
SLH1 LB     ATGATCTCTTAAAATACCACCAACACCACTTAAATTTATGGTGCTTTTAAAATAGCCATCTGTATTCAATTTCATAAATCCACTCGGAGGATTTT-CCCA 7198
             * *** *  * * *       ***    *    *    * **        *** * ****         *  ** ** *     *  * ** *  *   *

s           GGTCGGAGGACCGGGTCATTACATTAGTTTTCCATATTTTTAGTTTCAAAGTTTTTTTACGTTATTATTATTTTTTTGGATCTGGTTATGGTTTTTTTTA 7300
SLH1 LB     CTTTACAACCCTTGCTTTTTTCCCATTCTTTACTGATTTTAA----CGAACTTATTTCCTCCCAAATCCAATTTTTAGGAAATCTTTTATCTTTTTGAGA 7294
              *   *   *  * *  ** *      *** *  ***** *    * ** ** ***         *     * ***** ***  *  **      *****   *

s           CGTTTTTAAGGGAGAACCAAAGTATGATAGTATAATAGAATACGCATATATATTTTTTTAGTGATACCATGAGGTATAATAGTATATGCACAT----ATA 7396
SLH1 LB     AAAGTTCTGAATTAATCCAATTATTTCCAGAATGATATTTTGAGCATCCATTCTTGTATTATCAAATTTACTAACACAACGAGCTTTCCAAATGGACCAA 7394
                **        * ****    *   ** ** ***   *  ****  **  ** * *   * * *         * **     * * **  **      *

s           GGTAAGTAATAAGCTAGTTATTTTATTATCAAAATTAAGTCAAG--TTATTTTATCATCAAAATTAAGACAAGTTATTTTCCTCTCTTTTCCATAGTCTA 7494
SLH1 LB     ATAATATACACAGGGAGAGCCTTTAAAATCTTTTTTTCCAGATGGAAACTTTTCCTATCATGAATCCACCAAGAAATAAATCTTCCTTTTACTGTTAAAA 7494
               *  **   **  **    ****  ***    **     * *     ****   ****  * *    ****  **      **  *****  *       *

s           ATAAAAATATTTTCTAAAATTCTAATTTATCAAATCAAACAACCATCAGAAAATCTTATTTTTCCGAAATTTATTTTAT---TTCCATGATTTTCCCCCA 7591
SLH1 LB     CATTTTGTTTTTCCAAATACTGAGCAAAGTACTTCCATACTTCCTTAGCAATATCACTTGTACTCAACAAATGTTCTAAGCTTTCTTCTTTTTTCCTACA 7594
                   * *** * ** * *        *     ** **  ** *   ** ***   * *   * * *  * ** **    ***      ****** **

s           ----AAACAAATACAGCCTTAGTTTACTTAAATTCTAT-CTCTGCTACTTATATAAAATTTATATTATATACAGATTTATATATATAGTCATTCCAATTT 7686
SLH1 LB     GCATTCACAACTAGAAGCTAATTGAACACCAATTCTTTGAACAGCAAGGTCCACTGGCACTGCATCATTTAAACACTTTCAAGCAAAATAGGAGACCTTC 7694
                  **** ** *  ** *  **   ****** *   * ** *  *  *       *  ** ** ** * * **  *   * * *       **

s           GAACTAATCTATATTGCAACTTTGACTTTCAAAATTAACATATCTATATTGACGAAATGAAGAAAATTCAAAAGATAATCATACAGTTAAACCTTATAAC 7786
SLH1 LB     AAAGGTATTATTTTATCCCAAATCATCTCATATATCCAGTTGTTTGTATCTTTCTGTCTTATAATATTCCAGGCAGATTTAAATGAAAATTCACCATTAA 7794
             **   **    * *  *     * *  *   * **  *  * * * ***           * ** **** *   * * * * *     *  *   ** *

s           TATATCAACTCAAGATATGCTCATATAAGTTTGAATTGGAACGGAAAAGTAATTATATAAAAATGTACC--TAACATCAAAGCATGCATGTCTATGTCAT 7884
SLH1 LB     TGTCATCCATCCATATAGCTTGATCTGGTTCATTCTTTTGATAGATTAAGTTTTGAGTAAACAGTCACTGTTTATTACTAGGTATGATTGCCTTAATAAC 7894
            * *      ** * ***   * ** *   *     **   *  **  *     **   **** *   **   * *   * * * ***  ** **   * *

s           ACAAAGGAAGTGTCAACCCAAAGTTACTCGGCATTTTAGTCAAGGATGAAGATTTCTACTTTTACACATTTATTGCACCTTATTTACAACAACAATAAAA 7984
SLH1 LB     TTCATAGTTCCACTCACCATTAGAAATTAGTCTTGATAGCAGCAAATTATTATCTATATTATGATCTTTGACTGGCCCATTTCCAACCCAATTATCTAGG 7994
               *  *        ***    **  * * * * *  ***       ** *  ** * ** * * *    *   * ** * **     **    *  *   *

s           TTTGGTTCGCCACCCGAGTCGGGAGGCTGTCAATACCT-TATCTGAAATAAGCATGCCATGTTCTTGTTAAAAGAACCACAGAAACTGCCAACTATAGTT 8083
SLH1 LB     TTCACTGATACATTTCATTTTCCAATCAACCAAAGAATGTATTTTTCAGCAACATCCCTTACTTTT-----------AACATTCTGGACCATGTGTTAGT 8083
            **   *    **    * *    *  *   ***    * *** *   *  * *** ** *  * **          ***       ***  *  *   *

s           TATATATGTCGTACGTACACCCGTTGACGTTTGGTGTGAACTCTCACACCAGACTCATCACA-CTTCTGATACATTGACGACTTCTCATTGGACCATCCG 8182
SLH1 LB     ACCCTTTGTAG------CACCAATCTCTATTGGATGCTTTTTCTTGCAATATCTTGCTCTCATATACTGAGCCCATAATGAATTTT--TTGTTCTAAAC- 8174
                * *** *      ****  *     ** * **     ***  **  *   *  ** **  * ****  *  * * ** ** *  ***  * *  *

s           ATTGTCACAACAGATAAGTGGAAATT--TGCAATAACAAAAAATCATGGTCAGGGACGGATGCCTCATTGTGCAAGGGGGTCAAGTGCCCCCTTTGATCC 8280
SLH1 LB     --TTCCACCATAGTTTGATTGCAAAAGCTTCAGAGATATCTTTAAAACTTCTCAAACCAAGCCCACCTTCAGTAGTAGG----TCTGCAAACTTTACTCC 8268
              *  *** * ** *   * * **    * **   * *       *   **    **  *  ** * **  * *   **      ***   ****  ***

s           GGAGCAAAAATTAAAAAAATATTTTTTTTTCTGTTTTTTCAAGTATCCTATTAATATAAGTCAAGTCAAGAAAAGTTGAGGCTGCCCAATTTTACCCATT 8380
SLH1 LB     -------ATTTAACCCAATGATATTTATTTCTGCCATTAGTATTTCCCCA-------AAGAAAATTTGCTATTATTTGATCTATCTTTCTTATCACCAGT 8354
                   *  * *   **  ** *** ******   **    * *  ** *       ***  ** *    *  * ****    *    ** *  *** *

s           TTTTTCAAAAATACCCCCACTCTATTTTAATAATTACCAATATGCATAATCTAATTAATTAACTATTTGTTCTCACTATATTTTCCTAATTACGGATGTA 8480
SLH1 LB     TTTGGAGGATTCATAGCAGCCATAGTATAAATAGGAATAGTTTGAAGAACATGTCTAGCAAGTAACAATTTGCCTCCCATGTTTAAATACTGAGTTTTCC 8454
            ***     *   *   *  *  ** * ***  *  *  * ** * **  *   **   *   *    **  * *     ***    *  *  *  *

s           CCATAAATTAATTAACAATTCGTTTTAACCTCCATAAGTATAATAAATGATACAAAATCTGAATATTATGTTAAAAATAAAATATATTACGAGATTTATG 8580
SLH1 LB     ACCCTGATAGTTTACTTCTCACTTTGAGGATTAATTCATGGTATAATAAACACTTTTTATGTCCAGAATACAGTTTACACCCTAAATAGGTGAAAGGAGG 8554
             *    **   ***    *   *** *   *  **   *   ****   * **    * **   *   *  **       * *   ** **      *   * *

s           TCATTACGGGTCTCACACGCAACGCGTGGGTAATTTTACTAG-TGTATAAGTATGTAATCCGTCTAAATATATAATAT-TTCATGTAAATATATTCAAAC 8678
SLH1 LB     CTTTTTAATAGCTAAATTCAGAATCTTGAATACTGCTTCTTTCTGTTTATCTGTAAGTCCCTTAGAAATATACATAAAGCTCTTCTCTATATTTACTTCT 8654
               **      ** *      *  * **  ** *  * **   *** **  * *     ** * ******* *  *   ** * *  **** * *

s           T---TAAGATATGAGTAGTGTATAAGTCTTTCTTTTT--AAGTGTTTGAGGTTCTCAGTACAAATCCCTTTACCTTTTGTTTTTTTCTTTTAAATCATTT 8773
SLH1 LB     TGGCCCGAGTCTGTATTGTACTGCTGAAGTACCCTTTGAATATGTTTGATAGAGTTGGTATTAGCACCACTGAAAATAATAATATCATTAGCGAACGCCA 8754
            *        * **  * **      *   * *  ***  *  *******     *  ***  *   **  *     *  *  * *  **    * *
```

```
s          TATCGCTAATACAAAATATGCATTTCACT--TTTACTCCAAT--CGACTTTTCAATGTCTACATCCACGTGTACTTTATTACACAAAACACGTTAAAAAG 8869
S^LH1 LB   AATGGCTAATTATAGGAGTTCCTATAGCTGATTTAAAAGATTTCCAACTCTTCTTGCTGATCAGATTGTTAAGATTCATGGTCAATACTTCTAAACATAG 8854
            ** ******   *    * * * *  **  ****    *  *  * *** ***    *    **      *    ** **     * *   *   * * **

s          ACAAAACTTGGTTACGAATACTATTTTGTTGATCTGGATTGAAGTAACTTATAAGAAATACATGTTCATTTCG 8942
S^LH1 LB   AATAAAAAGACTTGAGGATAGGG----GATCTCCTTGTCT-TAGTCCATTTGTAGGCTTGAAATATCCTGTAG 8922
           *  ***     **  * ***       * *   ** *  *  ***   **   **   *   *   ** * * *
```

b

```
s          TATATAGTTTATATTGTACACTATATTATATATGTATACAATGACATGGTAATTTTATCGTACTAATTAAGATTTAAAACCA--TATGCTAAATGAAACT   99
S^LH1 RB   GATGCAG-----ATTGGGCAGGAAAC-ATAGATG-ATAGAAAGA----GTACTTCTGGTG-GTTGTTTTTATTTTGAAAACAACCTTGTTTCTTG--GTT   86
            **  **     ****  **  * *  *** *** *** *** **    *** ** *   *   *  **     *** *** **     ** *   **    *

s          AATTAAAAAATATAATTATGTA-CATGATTTTAACGGATAAATGAAACTAACCT----CTTAAAA-AAATGATTTAAAACTACCGAAATAAAGATAGTTT  192
S^LH1 RB   CAGTAAAAAGCAAAATTGCATATCTTTATCT--ACTGCTGAGGCAGAATATGTTGCAGCTGGAAGTGGATGCACTCAACTTATTTGGATGAA-ACAG---  180
            * ******  * ****   ** * * ** *  ** * * *   * * **   *    **  **    ***   * **  **     ** ** * **

s          GTGTTTAATCAACACTAATTTTAAT-------TATTATTTATATTTCTGTAATAACGAAATTTTAAACCCTTCAATGTAATTTAAAAACATACAAAGTAT  285
S^LH1 RB   ATGTT-AACTGATTATAATGTCAAACAAGAAATATTAACTTTGTTT-TGTGATAACAAAAGT-----GCTATAGATTTAGCTAAAAATC-----CAGT-T  267
            **** **   *   **** * **         *****  * * *** *** ***** *** *      *  *  ** **  * **** *      *** *

s          CGATACGTTATTTTTACGGCTAAATAAAACTAACCTACTAAAAATCAATTTAAACTAGTAGAATTA--TAATAAAAAAAATTATAACAAAATTTAAAAAA  383
S^LH1 RB   CAACA------TTCCAGAACTAAACACA-TTGATCTA--AGGCATCATTTTA--TTAG-AGAACTAGTTGAAGAAAAGAGCT-TAAT----TTTAGAACA  350
           * * *      **  *   ***** * *  * * ***  *   **** ****   *** **** ** **  *  **** *  * ***       **** ** *

s          C---TAACCTTACA--CGGT--GCAAAGACCTTTATAAA--TTTTGGGTAACGAGTAATTCTAATGTAC-AAATTTTTATAAATAGGTATTATTTTTATA  473
S^LH1 RB   TGTTTCAACTGAAAGACAATTAGCAGATATTTTCACAAAGCCCCTTGATA-CAAACAAGTTTGAATCATTAAGAAGTGACATAGGGGTAT-GTCATTATC  448
               * * ** * *  *  *  *** * *  ** * ***     * * ** * * *  ** * *   *    **    * *  *****  *  ****

s          TATTTTTTTGCATATGATGGT---ATTTTTTAGTGGTTGGGTTATGTGAA-AATGTTGCATATATATATGAGTGAATATTGAATCGAATAGATAAGGAAT  569
S^LH1 RB   AATCTAATTGAATATTCTTGCTAAATATGTCTGTA-TTGAGTACCTTCAATAATTTT---TATTTCTAAGGTTAAGTCTT-ACTTGAGC---TATG--AT  538
            ** *  *** ****  * *    ** * *  **  *** **    * ** *** **   *** * ** *  * * * ** * * **     ** *  **

s          ATGATTGGTGTAAGAAAAAGACATATTTTTGGATAA-GAAAGTAGATTTCATATTTATGAAAAAATAAATAGAAAAAAATAT---ATATTTCAGTTG---  662
S^LH1 RB   GTTTTTGTTTT--GAAAAAGCCGCTTTTTGGTGTCTCAAAAACAGAT---------GTGAGGTAATTTGTCTTAAAACCTCTCTGGTAATAGAGTTGCAA  627
            *  *** * *  ******* *   **** *  *    ***  ****          ***   ***   *   ****  * *    ** *  *****

s          GTATATAGTA-ACTACTAAAAT--TAGTTACTAAATTCTAAGTAAGTTACTAACATAGTTGGTATTTACCTACTAA-AATG-TGGTAGTTAGTAATCAAT  757
S^LH1 RB   GCACAAAGTGCACTTTTTACACAGTGGTTGC-AAGTTC------AGTTGCAACCCCAGTTGAAACC-ACATGCTAATAATACTGGTTGCAAGC--TCAGT  717
           * * * ***  ***  * * *   * *** * ** ***      **** * * *  ******  *   ** * **** ***  **** *  **   *** *

s          TAGTATTTAGCAACCGATAAAATAGTAATAACTAATTAGTTGGTAAATTACTAACTGCCGAAATTTAAGTTACTATTCAGTTACTACTTACCAACTGAAA  857
S^LH1 RB   T--------GCAAC------ATCAGTTGAAACC---TCATACGCATAATACTGA------AAAATTCAGTGATTGTCCGAACACT--CTA--GAATGAAA  790
           *        *****      *  ***   ***    *  *  * * * *** *      *** ** *** * * * *    ***   **    * *****

s          AGTGTAAGTTACAGGTTAGTTACTATTTA-GTGACTAATAATATGGTAGTTGCAAATTACTTGTT-ACGTCATAAAAATATGAAATTTTAGT-TAGAGAT  954
S^LH1 RB   AAGCCAAGACAAAGTCTTGTTATTTTTGACATCACTTATACTCT------------TTCCTTGCTGAGGTCATCAAACTA----GTTTTTGTGTAGAGAT  874
           *    ***  * **  * **** * ** *  * *** *** * *            ** **** * * ***** *** **     **** ** *******

s          TCGGTTAG---TGATTAGT---AACTACCAAA--TTAGTTGGAATTTTAAGTTACTATGTGG---GCATTTTATTGTGTGTGAAGCGTTGATCGGTATAT 1043
S^LH1 RB   AAAAACGGATTTGCTCAATTTAAGCTCCCCAAATTTAATTTTTATTA-GAGTTATTCTTTGATAAACATATTTTCTTGAGTAAA---TTGCTAGAT-TTT  969
                  *   ** * * *   * ** ** **  *** **   ***   ***** * * **     *** ** *  ** ** ** **   *** * * * * *

s          AATGCGATATTCAAACTC-GACACCAATGTA--TGATCTGGGT--TGAAGAAGGCTAAACCACTTTAGCACGCACAAGTGCAGCATATGCTAATCT--AA 1136
S^LH1 RB   TTTGTTCTCTTAAAAGTTAGATATCTGTTTAATTGTTCTTTATCTTGAATCAAGTCAAATTTCCTTAATAAGC-----TTCAACTTTTG--AATTTTCAA 1062
             **   * ** *** *  ** * * *  * **  *** *** *  ****  * *  ***   * ***  * **     *   ** * * **  *** *  **

s          ATTTTGAAAAAAA-----TGAAAATTTGCACTTATGTATATT--TTTTTAATTTTTTGTTTAGTGCC---CCTCCCTGTCGCACGCGTCTAAGCCAACTA 1226
S^LH1 RB   ATTTTGGATGAATCTGTTTAAAAAGGAATGTTTATCTGTATTAACTGCTGGATTTAT-CTTAATGCCAACTGTTTCCTTCGTAT---TTTAAGTTGATTA 1158
           ****** *  **      * ****        **** * ****   *  *   *** *  *** ****     *  *  *** *    * ****   * **

s          GCTCTTTTGGCTTTGGCTAGTGATGACCCTGCATGAGTGGAATTTTTGTAAAAGTTATAACGGCCTATGATTAGATTTTTTTACCATGCATTTGGTTTAA 1326
S^LH1 RB   ATTTTTCAAATTT---CTATTGATAAGTT---ATTTCCTAAATTTTCGCCAAAATC----TGTTTCATAATT-GATTTG-----CATGTATCTCTTTCTA 1242
             * **     **   *** **** *      **      ****** *  *** *      *    ** *** *****      **** ** *  **  *

s          ATACCGGTACAGGGTTCGAATCTTACTCATATCCCATGAGTGATATCATTTGATCCCCTTAATAAGATTTTAAAAAGGAAAAAAAATGTCTTTTGATACC 1426
S^LH1 RB   AT--TGTAAAAAGATTTGGACTTTATTGCTATTT-AT-ACTAAACTC-TTTTATTTTTCAAATCA-ATCACAATTATCAGATTTATTCTCTCTCTCTTTC 1336
           **   *  * * * ** * *  *** *  ***   ** * * *  ** *** * **      *** * **   **  *  * *   * * *** * *   *  *

s          TTGGAGTATATAGTAAGGAAAATATTAATATTAT-----TCCCCAC-------TTACAAGT-CAATCCTAACTCGCCACCAAATAGGCATACACATAGAG 1513
S^LH1 RB   TT------CTTATCAAAGAAACCCTAAGCTTTCTGGAAACCCACACAGCGTAGTTCTAAGTGCAGTTCCAAC-CGTAGTTGAA--ACCATGTCTTCAGAC 1427
           **        **  ** ****   * *   ** *      ** ***        **  ***** ** * * *** **      **    ***      ***

s          A-ACTATACATATGTGGATTCT---AAATATGTTAGCATAAG---AAAGTTGTGATTGAAGTGGAAAAT---AAG-TTTAACTCAATGTTTCGAGTAAAG 1602
S^LH1 RB   ATACTA-GTTTGTCTGGCTTATGACAAATAGGGGGAGATGAGTTTATGGTTCATGTTAATGTTTACAGTTTCAAGCTTTAATTTACAATTGCTGTTTTAA 1526
           * ****    * * *** ** *   ***** *     ** **   *  ***    ** * **  * * *   *** ***** * * *   ** *   *  *

s          TTGCCATCTAATTATCCTTGGATTGTCAAGTTTC---CTTATGT--TG-----AACATCGATTTATA-----AATTTAATTGATGTTATTATAA------ 1681
S^LH1 RB   TTTTAATATGAAACTCTGTGAATGGATACATTCTGTGCTTATATGATGCCTGCAATTCTAACTTGCAGGGGGAGCTAAATAGGTGTTACAAAAGCACTGA 1626
           **   ** * *   **  ** ** *  *  **     ***** *  **     **     * **  *     *  * *** * *****  * *

s          -AAATCCGGATTTTCGTAGAAAAATGCATGTGTTAA---------GAGAAATTATTGTCTCATATCGTCT-AGATA---TGTTAGTAC--AT-ATATGGT 1764
S^LH1 RB   TAAAT--GAAATACAGTAAAAGACTGTAC-TGCTAACTTATTGCAGGTGGATTAGTGTTT-ATATAAACTTAGATGATTTGTCAAAACCAATGACAAAGG 1722
            ****  * * *   *** ** * ** *  ** ***          *    **** *** * ****   ** ****    *** *  **  ** * *  *

s          ATATAATCTAATAAAATCCCCATTACTA-TCAACAATACTTTTTAAAGTCGTGCGGCC--CAACAAGCGAG--TATGAAATTGGGCCAAATCG-GACAG- 1857
S^LH1 RB   GGGAGAT-TGTTAGGA-CCAGATTCCTGGTCTA-AATTCTATTTGTTGTCTAATGGTTTTCACTAAATAATCCTATCTGTTTATTATGAACTAAGACAGG 1819
                ** *  **  * **  ***  ** ** * *** ** ***   ***     **    **  **   *   ***    **      **    *****

s          TATTGGT-GGTAGGAAGCA--GATG--------GTGACGGCCTG-ATCGGGAGAT--CGGGT--CGTTACAA--TTGGT----ATCAGAACCGCTATCTA 1935
S^LH1 RB   TGTTGTTAGGTTTCAAGCAATGTTGCAACCTCAGTTACAACCTATGTTGATAGGTTTCTAGTAATGTTGTAAGCTCAGTTACAACCACAGTTACAACCTG 1919
           * *** * ***   *****  * **        ** **  ***   * *  ** *  *  **   ***  **  *  **    * ** *    * * **

s          GC-AGATGG--TTTGTGA----AACTTCCCCCT----CCCCGGTAAGGACGGCGTGTCAGACACGATGACGTCGTGCATGAAGGGGGAAGAGTGTAAAAA 2024
S^LH1 RB   CTTAGACAGCGTTGGTGGTTTCAAGTACCAGTTGCAACCTCAGTTGCAACCTGCTGTTCAGCAGAATTCTGTC-----TAGTTCTGTAAAACTGCATAAC 2014
              ***  *  ** ***     ** * **   *    ** * **    **    ***    **  **   ***     *      * ** * ** * **
```

```
s          TCCGGATTTTCATACGA--AATGCCGGT------GTAAAGAGAAAGTATTGTCTCACATCGATGACGAA-----ATTGGGC-CAAATTGGACAATATTA- 2109
S^LH1 RB   T---GTTCTTTAT-TGATTAATATTTGTTCTACAGTATGCAGGATATATT--CTCATAT-GATCTTAAAGGTTTAATATGCGTAAATTGAG-AATATCAA 2106
           *   * * ** **  **  ***    **      ***   ** *  ****  **** ** ***    **     * *  **  ******   ***** *

s          -GTGGTAGAAAGCAGATAGTGACGGTCTGATCCGAAAAC-CAAATCGTTACAGTTAT----CCTCTGTAAT-----CTATAGTACTATAACCTCCATTTT 2198
S^LH1 RB   AGTGTTAAGGAA-AGATGTTTTCTAAAAGATATAAAGAGATAAGTTTATAAGGTTATGTTTCCTTTTTAATATGATCTCTATTCTTATGCTTGTCCTTAT 2205
            *** **   *  ****  *  *     ***   ** *   ** *   **  *****    *** * ****     ** ** *  ***      * ** *

s          CCTATGAAATGATGTGGTGATTGT-ACCCGTGAGCAGTGAGAGTCCGTAACTATACTGAT-GAATGTAATTGTATCGATCGGCATTG-GTATA-ATGTG- 2293
S^LH1 RB   CACTTGTTTCAATCTTTTGTTCCTTATCTATATTCAGT-ATAATTCGT--TTAAA--GATTGAATATAAT-ATGGAGTTAGGTGTTTCATATATATATGG 2299
           *   **     ** *  ** *  * * *  *   **** * * * ***   ** *  *** **** ****  *   * * **  **   **** ** **

s          --ACGTAGAGTCACGAAAA---ACTATTTGCACATTCACTCTTTGTGACAGAAAATGAGGATAAATATACAGTTTTGAA----TCTTTTGATCAATGTCG 2384
S^LH1 RB   AAAGATATTATCTCCAAAAGATATTATCTTTATATGT-TTCCTAGTTATAGGAAA---AGACTAACTTCTATTTCCAAGCTAGTTATTTAACAGGTATTA 2395
             *  **   ** * ****   * *** *  * **    ** * ** * ** ***    **  **  *  * **   *     *  *** *    * *

s          TTTT--------TGTCTTCCCTTCTGTTACAAACAG---TG---TATGAAAAAGTAAAAAGTGTGTACATATAATTCACAAACAATGTGTGA-------A 2463
S^LH1 RB   ACTTATGAAGATTGATGTCAATCCTACAAAAGGCAGACGTGCAATATGGAGAGATTGAA-GCCAATCAAAACAAGTCAAAAGCAATCTTCAAGATCCTAA 2494
              **        **   **  * **   * *  ***  **   **** * *  *  ** *    *  * * ** *** ** **** *   *       *

s          AGAGTTAAAACTGT-------------AGAGTCACTGAC---GCT-------AATGT-TGACCTATAATTCAGGTGG---ATTCTTTAGATAAAATATCC 2536
S^LH1 RB   AGAAAGAAGATTATTTTACTTGTATAAATAGTTGATAACCTAGCTTGTAAAAAGTGTGTGTTCTAAATTATAAAGAACTTATTTTCTAGAGAGAATACAC 2594
           ***   ** * * *             * ***   * **   ***       * *** **  *** * *  *        *** * **** * ****  *

s          G--AAGTATACCAAG----TATTGTCGC----CTATCGTTTTA-AGTTCATAATTATGAAAT--GATTGGGTTTGTCTTA---AAAATGGCATGGTTAGA 2620
S^LH1 RB   GTGAGCTTTAAAAAGAAAACAATGGTGGGATGCCATTGATTTATAAGTGATAGTTATAGAGTAGGAGCAATATAGTGATAGGCAGAAAAACCTAGAGGGA 2694
             *  *  * **  ***   * **  *     * ** * **** *  * *** ****  * *  **      * **  **   * **   * * *   **

s          ----AAGTC--TCTTCCAAA-TAGCATCATATTTAAGTAACTTCTT----TACCA--------ATAGATAC-------ATATATTGTAAAAAGAAAAACT 2694
S^LH1 RB   GTCTAAGTTGTTCATTAAGACTTGAGTCATATATTGTAAGCTTTTTATAATATCATAATCTGGGTAAGTGCCCCGGAGACGTAGCCTAGTTAGGTGAACT 2794
               ****   ** *  * * * *  ****** *    * *** **    ** **         **  * *       *  **   **   **   ****

s          C--------AAAATACAGAAAAATCAAAT--AGGCGCGACTACATGAAATATTATTCATATCTCTTAGCAAAATTGG------CCGGAGA---AAAATTG 2775
S^LH1 RB   CCGTTGTTAAATGTGCGTGTGAATTATTTTCAAGTACATTTAATTTTAAAACAATTC-TGTCCATTAATACACTTGGTTCCAACTGCAGTTCCAACATTA 2893
           *        **  * *     *** *  *  * *  *   **  *  ** *  **** * **  ***  * * ****      * * **    ** ***

s          CATTGA-CTTTA-GCGGTCGATAGATACTT--ATTTTATCTTAATTTAAGACCTTAAAGTCAAAAGT-TATC-GTTGCAACGCATA---ATTTGGAGAGC 2866
S^LH1 RB   CTTGGAACCTTATACACTTGGTTCCAACTGCAGTTCCAACCCCACTTGGAACCTTACTG-CAAAACTGCATCTAGAACAGCGGATAGGAATTTCAATCAC 2992
           * * ** * ***  *  * * *    ***    **  * *   * **   ******  * ***** *  ***     ****  ***   ****  *   *

s          T-------ATAACACGGTTTGCTAAATCATTCTTAACTAGTTAACCTTTTCAATACTTGTTAACAAAGAACATACTGATCCCATGTTTGATTGTAAATGC 2959
S^LH1 RB   TTAATTTTAGAACACGGTTTGCTAAATCATTCTTAACTAGTTAACCTTTTCAATACTTCTTAACAAAGAACATACTGATCCCATGTTTGATTGTAAATGC 3092
           *       *  ************************************************ *****************************************

s          AAACCCGTCCTGCAAGTATAAACCCAAACTGACACCTAAAAACCGCGAGATTAGTCTTAAACAACAAAACAGTCTATTTTCATAAAAAATAAAAAAAAAT 3059
S^LH1 RB   AAACCCGTCCTGCAAGTATAAACCCAAACTGACACCTAAAAACCCCGAGATTAGTCTTAAACAACAAAACAGTCTATTTTCA-AAAAAAAAAAAAAAAAA 3191
           ******************************************** ************************************* ****** *********

s          AGTCA-AACTCTACTTCTGATTAATAAAAAACTGATCACCAAGATTATACATAATAACATAAAGTTATCTACTAATCTAATTATGTCACATAGAAGGGAA 3158
S^LH1 RB   AGTCAGAACTCTACTTCTGATTAATAAAAAACTGATCACCAAGATTATACATAATAACATAAAGTTATCCACTAATCTAATTATGTCACATCGAAGGGAA 3291
           ***** *************************************************************** ********************* ********

s          CTTTTGGAAATTTAATGAGAAAGAGGAGGAAAAGGAAGAAAACAATTAGAAATTGGATAACAAAACCCCTTCGTAATTTTGAGTTCGAGTCCTTCACTCC 3258
S^LH1 RB   CTTTTGGAAATTTAATGAGAAAGAGGAGGAAAACGAAGAAAACAATTAGAAATTGGATAACAAAACCCCTTCGTAATTTTGAGTTCGAGTCCTTCACCCC 3391
           ********************************* *************************************************************** **

s          TTGAATCACAACCTCTAAGCCTGAGGCTCTTCTAAGACAAGAGTTTCTACTAGTGACCATAATTGAATTTTCATGAGAAAACCGGTAATTCACCCCATAG 3358
S^LH1 RB   TTAAATCACAACCTCTAAGCCTGAGGCTCTTCTAAGACTGGAATTTCTACTAGTGACCATAATTGAATTTTCATGAGAAAACCGGTAATTCACCCCATAG 3491
           ** ***********************************  ** *********************************************************

s          TAGTAATCTAGTTTACATCTTCCCGAGCTATCAATAATTTCAACCTCGAATCCCGGGGAGAGGGGTTTCCCAACAACTTGTACGTGTGTGATACTGCAAA 3458
S^LH1 RB   TAGTAATCTAGTTTACATCTTCCCGAGCTATCAATAATTTCAACCTCGAATCCCGGGGAGAGGGGTTTCCCAACAACTTGTACGTGTGTGATACTGCAAA 3591
           ****************************************************************************************************

s          TTGAAAGAGGAAAAAAA-TCATTAATCCGAGTTACAAAAAAATATTGTGCAAAAAAATACTAATCATGACAAGCTTGCTACATAAAAAAGATCTGAGTGG 3557
S^LH1 RB   TTGAAAGGGAAAAAAAAATCATTAATCCGAGTTACAAAAAAATATTGTGCAAAAAAATACTA---ATGACAAGCTTGCTACATAAAAAAGATCTGAGTGG 3688
           ******* * ******* ********************************************   ***********************************

s          ACATTAGACAAAGCAATATATGAGTGAAGAATGAATTTGTAGTGCTAAAATTTTAAGAAGAAGGAATTTAGCAAACTAACCGTAAATAGTATAAGCCATC 3657
S^LH1 RB   ACATTAGACAAAGCAATATATGAGTGAAGAATGAATTTGTAGTGCTAAAATTTTAAGAAGAAGGAATTTAGCAAACTAACCGTAAATAGTATAAGCCATC 3788
           ****************************************************************************************************

s          AGTTTCGTGGATTTGTACCCGTCCCAATAGTTCCATTTGTAAGATTCCACCAACGCAAAGAACAGGTTCAGGGAGCTTAAATTTTTGCAAGCAGTTCTCC 3757
S^LH1 RB   AGTTTCTTGGATTTGTACCCGTCCCAATAGTTCCATTTGTAAGATTCCACCAACGCAAAGAACAGGTTCAGGGAGCTTAAATTTTTGCAAGCAGTTCTCC 3888
           ****** *********************************************************************************************

s          TAATTTATATTTAAGCAAAAAGTAAAAAGAATCATTATCTTGAATTGATATGAAAGTTACAAAAACTTAAAAGATCACAGTTGGCATGACATACCTGAGC 3857
S^LH1 RB   TAATTTATATTTAAGCAAAAAGTAAAAAGAATCATTATCTTGAATTGATATGAAAGTTACAAAAACTTAAAAGATCACAGTTGGCATGACATACCTGAGC 3988
           ****************************************************************************************************

s          CATTCGAAACTCAGGCGAAATATATGTCCATGAGAACGCATTGTCATAATACTTGTGACTAGCTTTACACTTGTACCTTCCATATTTGTACAATTTTTTT 3957
S^LH1 RB   CATTTGAA-------CGCATTAT----CCATGAGAACGCATTGTCATAATACTTGTGACTAGCTTTACACTTGTACCTTCCATATTTGTACAATTTTTTT 4077
           **** ***       ** * ***    *************************************************************************

s          TGGGACTTGGGATAGCCCATCCGAAATCGCACGGCCCTAGCTGAATATATGGAACACCTAGTTTCCAAAGATGGCACTGAGCAAACCCAAAACTTCCATT 4057
S^LH1 RB   TGGGACTTGGGATAGCCCATCCGAAATCGCACGGCCCTAGCTGAATATATGGAACACCTACTTTCCAAAGATGGCACTGAACAAACCCAAAACTTCCATT 4177
           ************************************************************ ***************************************

s          AAACAACGGAGTTATGAAAGAATAATTGAAAAA-CAGGTCAGAAAATCACCTTGAAAGGGTTGAACGTGAATCTCCGATACAATGCAGATATTGGAAATC 4156
S^LH1 RB   AAACAACGGAGTTATGAAAGAATAATTGAAAAAACAGGTCAGAAAATCACCTTGAAAGGGTTGAACGTGAATCTCCGATACAATGCAGATATTGGAAATC 4277
           ******************************** ******************************************************************

s          AAGTTGTAAACTAGTTTTTGGGGAACTGAGGGATCAGTTTGGCCTGTGCTTGACCAATAAGAATCTCGCACACCTATCCTCTCTCTTGGTTCAAGAGTAT 4256
S^LH1 RB   AAGTTGTAGACTAGTTTTTGTGGAACTGTGGGATCGGTTTGGCCTGTGCTTGACCAATAAGAATCTCGCACAGCTATCCTCTCTCTTGGTTCAAGAGTAT 4377
           ******** *********** ******* ****** ************************************ ***************************

s          TTTGTATACATTCTTGAGGATAATTATCAGTACTGGATGCACTAACCGCAGTAGATATACAAAATTTTGAATTGGATTTAGCAAGACCTCGGGATAAATA 4356
S^LH1 RB   TTTGTATACATTCTTGAGGATAATTATCAGTACTGAATGCACTAACCGCAGTAGATATAGACAATTTTGAATTGGATTTAGCAAGACCTCGGGATAAATA 4477
           *********************************** *********************** * **************************************
```

```
s           TGCGTAAACCTTGTGATCGCTCTCCAGATTTTTTTGTTGCCTAGAATTGAATGACCCCGCTTGTTTGCTTTTCTCTTTGATAATAATCGGGATTTTCTCT 4456
S^LH1 RB    TGCGTAAACCTTGTGATCGCTCTCCAGATTTTTTTGTTCCCTAGAATTGAATGACCCCGCTTGTTTGCTTTTCTTTTTGATAATAATCGGGATTTTCTCT 4577
            ************************************** *********************************** *************************

s           AATTCGGGAAAGTTTTTTAGGCATATCTGCTTAGAGAGTCCGTTCCTGATCGCTGCTTAAGATATATTGGTCATCAATAACCATCCATATAATATCCAAA 4556
S^LH1 RB    AATTCGGGAAAGTTTTTTAGGCATATCTGCTTAGAGAGTCCGTTCCTGATCGCTGCTTAAGATATATTGGTCATCAATAACCATCCATATAATATCCAAA 4677
            ****************************************************************************************************

s           AACCAATACAATAGTCACTAAAATGGAAGACTTGAAATGTCATTTTTTATTTCATCCAAAAGTAAGTAATGAAACCACCTGCTGACTGCTGTTTATGTAG 4656
S^LH1 RB    AACCAATACAATAGTCACTAAAATGGAAGACTTGAAATGTCATTTTTTATTTCATCCAAAAGTAAGTAATGAAACCACCTGCTGACTGCTGTTTATGTAG 4777
            ****************************************************************************************************

s           GTCTCATGAAATAGAGTGTTTTTAAGAAAATCCTTACTAAAAACGGATGAGGGGTATAAGTGAACGATAGTCCAAGAGATGCATGCAAATAAAAATAAAC 4756
S^LH1 RB    GTCTCATGAAATAGAGTGTTTTTAAGAAAATCCTTACTAAAAACGGATGAGGGGAATAAGTGAACGATAGTCCAAGAGATGCATGCAAATAAAAATAAAC 4877
            ****************************************************** *********************************************

s           TAACCAAAATCACGCCAAGAGGTTGAAACTGTGCTAAAACGTATAATATCCGACGGATTCTCAAAAGACAATAGGATTCTCAAGGATAAATCATGTCCAA 4856
S^LH1 RB    TAACCAAAATGACGCCAAGAGGTTGAAACTGCGCTAAAACGTATAATATCCGACTGATTCTCTAAAGACGATAGGATTCTCAAGGATAAATCATGTCCAA 4977
            ********** ******************** ********************** ******* ****** ******************************

s           GCCATTGTATAAAGTCTCTGGAACTTGCATGGATTTCCATAGTGGATATTGAAATAAATTATA---TGTCGAAGGTCACTGCCTTTCTTAGTTACATCAA 4953
S^LH1 RB    GCCATTGTATAAAGTCTCTGGAACTTGCATGGATTTCCATAGTGGATATTGAAATAAATAATAGTATGTCGAGAGTCACTGCCTTTCTTAGTTACATCAA 5077
            *********************************************************** ***   ******  **************************

s           AATATAGTGGCACTATATATAGGATAAGAAATTGTTTAGTTCTACCATTAGTTACATCAAAATAGAAACTTTCAGCAAATAAAAGTCATAAAAGTATACC 5053
S^LH1 RB    AATATAGTGGCACTATATATAGGATAAGAAATTGTTTAGTTCTACCATTAGTTACATCAAAATAGAAACTTTCAGCAAATAAAAGTCATAAAAGTATACC 5177
            ****************************************************************************************************

s           ACGTGAGTGACGTTTGGAGCCTAATACTTTTCCTTGTAAAATCTCTATTTTCTACCTCTTTTTGTCTCATTGCTGCTCCGTTGAAAGAGTATGAAATAGT 5153
S^LH1 RB    ACGTGAGTGACGTTTGGAGCCTAATACTTTTCCTTGTAAAATCTCTATTTTCTAC-TCTCTTT------------------------------------- 5239
            ******************************************************* *** ***

s           TTTACCTTTTTTTTTTTAATTTTCCTTAGTTTGGATGAACCAGGTCGTAACAAAACCCAACTTTTCCAGTTAATGGATGATCGTAATTTAGTAGCGGTCC 5253
S^LH1 RB    --------------------------TAGTTTGGATGAACCAGGTCGTAACAAAACCCAACTTTTCCAGTTAATGGATGATAGTAATTTAGTAGCGGTCC 5313
                                      ******************************************************* ******************

s           TTATTAGTTACCCATTATTTTAAATTTTGATTATTCACTTGATGATTATCTATGTTCCACAATCGATCTTTATCCGTTAAATATCTAGTGGACAAAACAG 5353
S^LH1 RB    TTATTAGTTACCCATTATTTTAAATTTTGATTATTCACTTGATGATCATCTATGTTCCACAATCGATCTTTATCCGATAAATATCTAGTGGACAAAACAG 5413
            ********************************************** ***************************** ***********************

s           AAGAATACATGTGGAATATTTTCCTACTTTTTTCCTTTTTTCTCAAGAAAAGACTCCAGGGTCTCGTTCTGGAAAAATATAATTAATTAGTTTATAATCG 5453
S^LH1 RB    AAGACTACATGTGGAATATTTTCCTACTTTTTTCCTTTTTTCTCAAGAAAAGACTCCAGGGTCTCGTTCTGGAAAAATATAATTAATTAGTTTATAATCG 5513
            **** ***********************************************************************************************

s           AAGTTCATAACGTTATATGATCGAATTGGACTGATTCAGATGTTTAACACTTCATATATACTTGTAGCGGGTATACCCACTAATTTAACAAATAGTATTA 5553
S^LH1 RB    AAGTTCATAACTTTATATGATCGAATTGGACAGATTCAGATGTTTAACACTTCATA----CTTGTAGCGGGTATACCCACTAATCTAACAAATAGTATTA 5609
            *********** ******************* ************************    ************************ ***************

s           ACTATTTTATTTTAATCTAACCGTAGGTCAAACAATCGTTTTAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCAAATA 5653
S^LH1 RB    ACTATTTTATTTTAAACTAACCGTAGGTCAAACAATCGTTTTAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCAAATA 5709
            *************** ************************************************************************************

s           TTTAGTCGTGAAAGTGAAACTGTGAACTTGGTCGGTCTTCTTCTCTGTACCATAGATTGCCTTGCATACGAGATATTAAATAATTTTGGATGATTACATA 5753
S^LH1 RB    TTTAGTCGTGAAAGTGAAACTGTGAACTTGGTCGGTCTTCTTCTCTGTACCATAGATTGCCTTGCATGCGAGATATTAAATAATTTTGGATGATTACATA 5809
            ******************************************************************* ********************************

s           GATAGATACTTTTGAGTTTTCACACACTAGCTATAAACATATATACTATGACTACGTTCAAAAAAAATTAAAAAAATACTATAACTAGGAAAAAAATTTG 5853
S^LH1 RB    GATAGATATTTTTGAGTTTTGACACACTAACTATAAACATATATACTCTGACTACGTTCAAAAAAAATTAAAAAAATATTATAGCTAGGAAAAAAAATTG 5909
            ******** *********** ******** ***************** ****************************** **** ************ ***

s           AAAAATGATCCGGTATGAATCTGACGCCAGTTGAAACTGATCCTAGTACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTGTATCAATTTAC 5953
S^LH1 RB    AAAAATGATCCGGTATGAATCTGACGCCAGTTGAAACTGATCCTAATACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTGTATCAATTTAC 6009
            ********************************************* ******************************************************

s           ATTAATTAAATGGCTTTAAACGAGTCCTTGCATGAGTATGTCAGTTAAACAGCATTTACCAATAGTGTTACTCGTAAAATTATTATACTCTTTGTGTATC 6053
S^LH1 RB    ATTAATTAAATGGCTTTAAACGAGTCCTTGCATGAG--TGTCAGTTAAACAGTATTTACCAATAGTGTTACTCGTAAAATTATTATACTCTTTGTGTATC 6107
            ************************************  ************** ***********************************************

s           TGCACAAATTCTTTGGCTCTGTTTGCTTCCCCGGAAAACAATTTCCAGCTAATTTTTTCAAGGAAAACATTTTTATG-AAAACTAATTGTCTTTCATGAT 6152
S^LH1 RB    TGCACAAATTCTTTGGCTCTGTTTGCTTCCCCGGAAAACAATTTCCAGCTAATTTTTTCAAGGAAAACATTTTTATGGAAAACTAATTGTCTTTCATGAT 6207
            ***************************************************************************** **********************

s           TTTCTTAGTGATTGGTTGTGTAGAAAAATTTCATGGAAAATATGAATTATACTTGATAATTTCTCTAACAAACACAAAAATAGATCACATAATAGTAAAA 6252
S^LH1 RB    TTTCTTAGTGATTGGTTGTGTAGAAAAATTTCATGGAAAATATGAATTATACTTGATAATTTCTCTAACAAACACAAAAATAGATCACATAATAGTAAAA 6307
            ****************************************************************************************************

s           AAAATCAACAAACAAATAAAAGTATGAAATTTGGTAGATATTTCCAACATAAAAAGTCATATTTTACCAAACAAAGTCTTTAACAAACAAATAAAAGGAT 6352
S^LH1 RB    AAAATCAACAAACAAATAAAAGTATGAAATTTGGTAGATATTTCCAACATAAAAAGTCATATTTTACCAAACAAAGTCTTTAACAAACAAATAAAAGGAT 6407
            ****************************************************************************************************

s           AATCTGTGTGAATATAGGATTCTCTTGATATACTGAAAATTAACTGCCCGCGGATATCACGACCACATTCCAATACAAAAAATTGAACAAAGAATCAACA 6452
S^LH1 RB    AATCTGTGTGAATATAGGATTCTCTTGATATACTGAAAATTAACTGTCCGCGGATATCACGACCACATTCCAATACAAAAAATTGAACAAAGAATCAACA 6507
            ********************************************** *****************************************************

s           ATGAATTGTTAATATTTTAAGTAGATGTAAGAGAAAAGATGTTGGAGGAAGGGGGTTAACACCCAAATAAGTTATCACGTTAACACCTTCGAGCTTTTCG 6552
S^LH1 RB    ATGAATTGTTAATATTTTAAGTAGATGTAAGAGAAAAGATGTTGGAGGAAGGGGGTTAACACCCAAATAAGTTATCACGTTAACACCTTCGAGCTTTTCG 6607
            ****************************************************************************************************

s           AGCTTTTAATCTCATGTAAATCTAGCCTTTTTTTTTATATCAAATCATAGTATAAATACGCAATTTTAGTTGTAATAACCCGGTAAGAGAAGAATATACC 6652
S^LH1 RB    AGTTTTTAATCTCATGTAAATCTAGCCTTTTTTTTTATATCAAATCATAGTATAAATACGCAATTTTAGTTGTAATAACCCGGTAAGAGAAGAATATACC 6707
            ** *************************************************************************************************

s           AATTGTAATAACTGGATTTTCGTAGAAAATGTCGTGGTGTAGAGAAAAAAGTTAACTCACATCGCCTAGATATTTGAGTATTTATATGGTATATTAGTTA 6752
S^LH1 RB    AATTGTAATAACTGGATTTTCGTAGAAAATGTCGTGGTGTAGATAAAAAAGTTAACTCACATCGCCTAGATATTTGAGTATTTATATGGTATATTAGTTA 6807
            ******************************************* ********************************************************

s           TGAGTATCATTCTTCTTATCAATAAGGCATTTTAAAACCGTGCGGCCCAAGAAACAATGACGAAATTGCACCAAACTGGACATTATTGGTAGTAGGAAAC 6852
S^LH1 RB    TGAGTATCATTCTTCTTATCAATAAGGCATTTTAAAACCGTGCGGCCCAAGAAACAATGACGAAATTGCACCAAATTGGACATTATTGGTAGTAGGAAAC 6907
            *************************************************************************** ************************
```

```
s           AGATGATGACGGCCTGATCGGAGAACCGAGTCGTTATAATTGTTATCAGAGCCGCTATCTAGCCGATTTC------------------------------ 6922
SLH1 RB     AGATGATGACGGCCTGATCGGAGGACCGAGTCGTTATAATTGTTATCAGAGCCGCTATCTAGCCGATTTCGGTCACTACAAAAAAATACTCAAATAGCGA 7007
            *********************** **********************************************

s           ---------------------------------------------------------------------------------------------------- 6922
SLH1 RB     CGGGATTAGTGACGCTATTTTTTGTTACGAGCTTATTAACGACGGTTTTTGTGACGGTTTTGTGACAGTTTAGTCAATTTGTGACGGCTTTGTGACAGTT 7101

s           --------------------------------------------------------------------------ACGAGTGATTGAATCCCGAAAGGACG 6948
SLH1 RB     TAGTCAATTTGCGACAGGTTTGTGACAGCTTTTGACTCTCATTAAATCAAACATTATTTTGTAATGGTACTAGCACGAGTGATTGAATCCCGAAAGGACG 7207
                                                                                      **************************

s           GCATGCTCGGCGAGGACGTCGTGTCTGAACAGGAGAAGAGTGTAACAACCCGTATTTTTGTAGAAAATGACATGGTATAGAAAATAAAGTTATCTCACCG 7048
SLH1 RB     GCATGCTCGGCGAGTACGTCGTGTCTGAACAGGAGAAGAATGTAACAACCCGTATTTTTGTAGAAAATGACATGGTATAGAAAATAAAGTTATCTCACCG 7307
            ************** ************************ ************************************************************

s           CGCATATCTATATAAGTACATATATAGTATATAACTTAAGAGTATCCTCGTTCTTATCACCAATATTTTTTAAAACTGTGCGGCACGAGAAGGAATGACG 7148
SLH1 RB     CGCATATCTATATAAGTACATATATAGTATATAACCTAAGAGTATCCTCGTTCTTATCACCAATATTTTTTAAAACTGTGCGGCACGAGAAGGAATGACG 7407
            *********************************** ****************************************************************

s           AAATT-GGCCAAAACGGACAATATTGATGATAAGAAACAGATGGTGGTGGTCTGGTCGGAGGACCGGGTCATTACATTAGTTTTCCATATTTTTAGTTTC 7247
SLH1 RB     AAATTTGGCCAAAACGGACAATATTGATGATAAGAAACAGATGGTGGTGGTCTGGTCGGAGGACCGGGTCATTACATTAGTTTTCCATATTTTTAGTTTC 7507
            ***** **********************************************************************************************

s           AAAGTTTTTTTACGTTATTATTATTTTTTTGGATCTGGTTATGGTTTTTTTTACGTTTTTAAGGGAGAACCAAAGTATGATAGTATAATAGAATACGCAT 7347
SLH1 RB     AAAGTTTTTT-ACGTTATTATTATTTTTTTGGATCTGGTTATGGTTTTTTTTACGTTCTTAAGGGAGAACCAAAGTATGATAGTATAATAGAATACGCAT 7606
            ********** ********************************************** ******************************************

s           ATATATTTTTTTAGTGATACCATGAGGTATAATAGTATATGCACATATAGGTAAGTAATAAGCTAGTTATTTTATTATCAAAATTAAGTCAAGTTATTTT 7447
SLH1 RB     ATATATTTTTTTAGTGATACCATGAGGTATAATAGTATATGCACATATAGGTAAGTAATAAGCTAGTTATTTTATTATCAAAATTAAGTCAAGTTATTTT 7706
            ****************************************************************************************************

s           ATCATCAAAATTAAGACAAGTTATTTTCCTCTCTTTTCCATAGTCTAATAAAAATATTTTCTAAAATTCTAATTTATCAAATCAAACAACCATCAGAAAA 7547
SLH1 RB     ATCATCAAAATTAAGACAAGTTATTTTCCTCTCTTTTCCATAGTCTAATAAAAATATTTTCTAAAATTCTAATTTATCAAATCAAACAACCATCAGAAAA 7806
            ****************************************************************************************************

s           TCTTATTTTTCCGAAATTTATTTTATTTCCATGATTTTCCCCCAAAACAAATACAGCCTTAGTTTACTTAAATTCTATCTCTGCTACTTATATAAAATTT 7647
SLH1 RB     TCTTATTTTTCCGAAATTTATTTTATTTCCATGATTTTCCCCCAAAACAAATACAGCCTTAGTTTACTTAAATTCTATCTCTGCTACTTATATAAAATTT 7906
            ****************************************************************************************************

s           ATATTATATACAGATTTATATATATAGTCATTCCAATTTGAACTAATCTATATTGCAACTTTGACTTTCAAAATTAACATATCTATATTGACGAAATGAA 7757
SLH1 RB     ATATTATATACAGATTTATATATATAGTCATTCCAATTTGAACTAATCTATATTGCAACTTTGACTTTCAAAATTAACATATCTATATTGACGAAATGAA 8006
            ****************************************************************************************************

s           GAAAATTCAAAAGATAATCATACAGTTAAACCTTATAACTATATCAACTCAAGATATGCTCATATAAGTTTGAATTGGAACGGAAAAGTAATTATATAAA 7847
SLH1 RB     GAAAATTCAAAAGATAATCATACAGTTAAACCTTATAACTATATCAACTCAAGATATGCTCATATAAGTTTGAATTGGAACGGAAAAGTAATTATATAAA 8106
            ****************************************************************************************************

s           AATGTACCTAACATCAAAGCATGCATGTCTATGTCATACAAAGGAAGTGTCAACCCAAAGTTACTCGGCATTTTAGTCAAGGATGAAGATTTCTACTTTT 7947
SLH1 RB     AATGTACCTAACATCAAAGCATGCATGTCTATGTCATACAAAGGAAGTGTCAACCCAAAGTTACTCGGCATTTTAGTCAAGGATGAAGATTTCTACTTTT 8206
            ****************************************************************************************************

s           ACACATTTATTGCACCTTATTTACAACAACAATAAAATTTGGTTCGCCACCCGAGTCGGGAGGCTGTCAATACCTTATCTGAAATAAGCATGCCATGTT- 8046
SLH1 RB     ACACATTTATTGCACCTTATTTACAACAACAATAAAATTTGGTTCGTCACCCGAGTCGAGAGGCTGTCAATACGTTATCCGAAATAAGCATGCCATGTTT 8306
            ********************************************** *********** ************** ***** *******************

s           -------------------------CTTGTTAAAAGAACCACAGAAACTGCCAACTATAGTTTATATATGTCGTACGTACACCCGTTGACGTTTGGTGTG 8121
SLH1 RB     ATGAGGCACGATTGAAATAAATGATCTTGTTAAAAGAACCAGAGAAATTGCCAACTATAGTTTATATATGTCGTACGTACACCCGTTGACGTTTGGTGTG 8406
                                     **************** ***** ****************************************************

s           AACTCTCACACCAGACTCATCACACTT-CTGATACATTGACGACTTCTCATTGGACCATCCGATTGTCACAACAGATAAGTGGAAATTTGCAATAACAAA 8220
SLH1 RB     AACTCTCACACCAGACTCGTCACACTTTCTGATACATTGACGACTTCTCATTGGACCATCCGATTGTCACAGCAGATAAGTGGAAATTTACAATAACAAA 8506
            ****************** ******** ******************************************* ***************** **********

s           AAATCATGGTCAGGGACGGATGCCTCATTGTGCAAGGGGGTCAAGTGCCCCCTTTGATCCGGAGCAAAAATTAAAAAAATATTTTTTTTTCTGTTTTTTC 8320
SLH1 RB     AAATCATGGTCAGGGGCGGATGTCTCATTGTGCAGGGGGGTCAAGTCCCCCCTTTGATCCGGAGCAAAAATTAAAAAAAAATATTTTTTTTCAGTTTTTT 8606
            *************** ****** *********** *********** ********************************** *******     *****

s           AAGTATCCTATTAATATAAGTCAAGTCAAGAAAAGTTGAGGCTGCCCAATTTTACCCATTTTTTTCAAAAATACCCCCACTCTATTTTAATAATTACCAA 8420
SLH1 RB     AAGTAT---------------------------------------------------------------------------------------------- 8612
            ******

s           TATGCATAATCTAATTAATTAACTATTTGTTCTCACTATATTTTCCTAATTACGGATGTACCATAAATTAATTAACAATTCGTTTTAACCTCCATAAGTA 8520
SLH1 RB     ---------------------------------------------------------------------------------------------------- 8612

s           TAATAAATGATACAAAATCTGAATATTATGTTAAAAATAAAATATATTACGAGATTTATGTCATTACGGGTCTCACACGCAACGCGTGGGTAATTTTACT 8620
SLH1 RB     ---------------------------------------------------------------------------------------------------- 8612

s           AGTGTATAAGTATGTAATCCGTCTAAATATATAATATTTCATGTAAATATATTCAAACTTAAGATATGAGTAGTGTATAAGTCTTTCTTTTTAAGTGTTT 8720
SLH1 RB     ---GTATAAATATGTAATCCGTCTAAATATATAATATTTCATGTAAATATATTCAAACTAA----------------------CTTCTTTTTAAGTGTTT 8687
               ****** ************************************************* *                       ****************

s           GAGGTTCTCAGTACAAATCCCTTTACCTTTTGTTTTTTTCTTTTAAATCATTTTATCGCTAATACAAAATATGCATTTCACTTTTACTCCAATCGACTTT 8820
SLH1 RB     GAGGTTCTCAGTACAAATCCCTTTACCTTTTGTTTTTTTCTTTTAAATCATTTTATCCCTAATACAAAATATGCATTTCACTTTTACTCCAATCGAATTT 8787
            ********************************************************* ************************************** ***

s           TCAATGTCTACATCCACGTGTACTTTATTACACAAAACACGTTAAAAAGACAAAACTTGGTTACGAATACTATTTTGTTGATCTGGATTGAAGTAACTTA 8920
SLH1 RB     TCAATGTCTACAACCACGTGTACTTTAATACACAAAACACATTAAAAAGACAAAACTTGGTTACGAATACTATTTTGTTGATCTGGATTGAAGTAACTTA 8887
            ************ ************** ************ ***********************************************************

s           TAAGAAATACATGTTCATTTCG 8942
SLH1 RB     TAAGAAATACATGTTCATTTCG 8909
            **********************
```

c

```
CFB^P     ----GGTTTGTCTTAAAAATGGCATGGTTAGAAAGTCT-CTTCCAAATAGCATCATATTTAAGTAACTTCTTTACCAATAGATACATATATTGTAAAAAG 95
CFB^TL    ----GGTTGGTCTTAAAAATGGCATGGTTAGAAAGTCT-CTTCCAAATAGCATCATATTTAAGTAACTTCTTTACCAATAGATACATATATTGTAAAAAG 95
CFB^TR    TATTGTAAGCTTTTTATAATATCATAATCTGGGTAAGTGCCCCGGAGACGTAGCCTAGTTAGGTGAACTCCGTTGTTAAATGTGCGTGTGAATTATTTTC 100
              *     * ** * ***  ***  *  *      * *  *  *   * * * ** *** ** *  **  *    * *  * * * *    **

CFB^P     AAAAACTCAAAATACAGAAAAATCAAATAGGCGCGACTACATGAAATATTATTCATATCTCTTAGCAAAATTGGCCGGAGAAAAATTGCATTGACTTTAG 195
CFB^TL    AAAAACTCAAAATACAGAAAAATCAAATAGGCGCGACTACATGAAATATTATTCATATCTCTTAGCAAAATTGGCCGGAGAAAAATTGCATTGACTTTAG 195
CFB^TR    AAGTACATTTAATTTT---AAAACAATTCTGTCC—ATTAATACACTTGGTTCCAACTGCAGTTCCAACATTACTTGGAACCTTATACACTTGGTTCCAA  195
          **  **    ***      *** *** *  *  *   *  **  * *    * **  *    *  *** ***    ***     **    ***  *  *

CFB^P     CGGTCGATAGATACTTATTT--TATCTTAATTTAAGACCTTAAAGTCAAAAGTT-ATCG----TTGCAACGCATAATTTGGAGAGCTATAACACGGTTTG 288
CFB^TL    CGGTCGATAGATACTTATTT--TATCTTAATTTAAGACCTTAAAGTCAAAAGTT-ATCG----TTGCAACGCATAATTTGGAGAGCTATAACACGGTTTG 288
CFB^TR    CTGCAGTTCCAACCCCACTTGGAACCTTACTGCAAAACTGCATCTAGAACAGCGGATAGGAATTTCAATCACTTAATTT-------TAGAACACGGTTTG 288
          * *  * *  *  *  * **   * **** *  ** **   *     ** **   ** *    **  * * * ******       ** ***********

CFB^P     CTAAATCATTCTTAACTAGTTAACCTTTTCAATACTTGTTAACAAAGAACATACTGATCCCATGTTTGATTGTAAATGCAAACCCGTCCTGCAAGTATAA 388
CFB^TL    CTAAATCATTCTTAACTAGTTAACCTTTTCAATACTTGTTAACAAAGAACATACTGATCCCATGTTTGATTGTAAATGCAAACCCGTCCTGCAAGTATAA 388
CFB^TR    CTAAATCATTCTTAACTAGTTAACCTTTTCAATACTTCTTAACAAAGAACATACTGATCCCATGTTTGATTGTAAATGCAAACCCGTCCTGCAAGTATAA 388
          ************************************* **************************************************************

CFB^P     ACCCAAACTGACACCTAAAAACCGCGAGATTAGTCTTAAACAACAAAACAGTCTATTTTCATAAAAAATAAAAAAAAATAGTCA-AACTCTACTTCTGAT 487
CFB^TL    ACCCAAACTGACACCTAAAAACCGCGAGATTAGTCTTAAACAACAAAACAGTCTATTTTCATAAAAAAAATAAAAAAATAGTCAGAACTCTACTTCTGAT 488
CFB^TR    ACCCAAACTGACACCTAAAAACCCCGAGATTAGTCTTAAACAACAAAACAGTCTATTTTCA-AAAAAAAAAAAAAAAAAAGTCAGAACTCTACTTCTGAT 487
          *********************** ************************************* ****** * ******* ***** ***************

CFB^P     TAATAAAAAACTGATCACCAAGATTATACATAATAACATAAAGTTATCTACTAATCTAATTATGTCACATAGAAGGGAACTTTTGGAAATTTAATGAGAA 587
CFB^TL    TAATAAAAAACTGATCACCAAGATTATACATAATAACATAAAGTTATCTACTAATCTAATTATGTCACATAGAAGGGAACTTTTGGAAATTTAATGAGAA 588
CFB^TR    TAATAAAAAACTGATCACCAAGATTATACATAATAACATAAAGTTATCCACTAATCTAATTATGTCACATCGAAGGGAACTTTTGGAAATTTAATGAGAA 587
          ************************************************ ********************* *****************************

CFB^P     AGAGGAGGAAAAGGAAGAAAACAATTAGAAATTGGATAACAAAACCCCTTCGTAATTTTGAGTTCGAGTCCTTCACTCCTTGAATCACAACCTCTAAGCC 687
CFB^TL    AGAGGAGGAAAACGAAGAAAACAATTAGAAATTGGATAACAAAACCCCTTCGTAATTTTGAGTTCGAGTCCTTCACTCCTTGAATCACAACCTCTAAGCC 688
CFB^TR    AGAGGAGGAAAACGAAGAAAACAATTAGAAATTGGATAACAAAACCCCTTCGTAATTTTGAGTTCGAGTCCTTCACCCCTTAAATCACAACCTCTAAGCC 687
          ************ *************************************************************** **** ******************

CFB^P     TGAGGCTCTTCTAAGACAAGAGTTTCTACTAGTGACCATAATTGAATTTTCATGAGAAAACCGGTAATTCACCCCATAGTAGTAATCTAGTTTACATCTT 787
CFB^TL    TGAGGCTCTTCTAAGACTGGAGTTTCTACTAGTGACCATAATTGAATTTTCATGAGAAAACCGGTAATTCACCCCATAGTAGTAATCTAGTTTACATCTT 788
CFB^TR    TGAGGCTCTTCTAAGACTGGAATTTCTACTAGTGACCATAATTGAATTTTCATGAGAAAACCGGTAATTCACCCCATAGTAGTAATCTAGTTTACATCTT 787
          *****************  ** ******************************************************************************

CFB^P     CCCGAGCTATCAATAATTTCAACCTCGAATCCCGGGGAGAGGGGTTTCCCAACAACTTGTACGTGTGTGATACTGCAAATTGAAAG-AGGAAAAAAATCA 886
CFB^TL    CCCGAGCTATCAATAATTTCAACCTCGAATCCCGGGGAGAGGGGTTTCCCAACAACTTGTACGTGTGTGATACTGCAAATTGAAAGGGAAAAAAAAATCA 888
CFB^TR    CCCGAGCTATCAATAATTTCAACCTCGAATCCCGGGGAGAGGGGTTTCCCAACAACTTGTACGTGTGTGATACTGCAAATTGAAAGGGAAAAAAAAATCA 887
          **************************************************************************************    **********

CFB^P     TTAATCCGAGTTACAAAAAAATATTGTGCAAAAAAATACTAATCATGACAAGCTTGCTACATAAAAAAGATCTGAGTGGACATTAGACAAAGCAATATAT 986
CFB^TL    TTAATCCGAGTTACAAAAAAATATTGTGCAAAAAAATACTAA---TGACAAGCTTGCTACATAAAAAAGATCTGAGTGGACATTAGACAAAGCAATATAT 985
CFB^TR    TTAATCCGAGTTACAAAAAAATATTGTGCAAAAAAATACTAA---TGACAAGCTTGCTACATAAAAAAGATCTGAGTGGACATTAGACAAAGCAATATAT 984
          ******************************************   *******************************************************

CFB^P     GAGTGAAGAATGAATTTGTAGTGCTAAAATTTTAAGAAGAAGGAATTTAGCAAACTAACCGTAAATAGTATAAGCCATCAGTTTCGTGGATTTGTACCCG 1086
CFB^TL    GAGTGAAGAATGAATTTGTAGTGCTAAAATTTTAAGAAGAAGGAATTTAGCAAACTAACCGTAAATAGTATAAGCCATCAGTTTCTTGGATTTGTACCCG 1085
CFB^TR    GAGTGAAGAATGAATTTGTAGTGCTAAAATTTTAAGAAGAAGGAATTTAGCAAACTAACCGTAAATAGTATAAGCCATCAGTTTCTTGGATTTGTACCCG 1084
          ************************************************************************************* **************

CFB^P     TCCCAATAGTTCCATTTGTAAGATTCCACCAACGCAAAGAACAGGTTCAGGGAGCTTAAATTTTTGCAAGCAGTTCTCCTAATTTATATTTAAGCAAAAA 1186
CFB^TL    TCCCAATAGTTCCATTTGTAAGATTCCACCAACGCAAAGAACAGGTTCAGGGAGCTTAAATTTTTGCAAACAGTTCTCCTAATTTATATTTAAGAAAAAA 1185
CFB^TR    TCCCAATAGTTCCATTTGTAAGATTCCACCAACGCAAAGAACAGGTTCAGGGAGCTTAAATTTTTGCAAGCAGTTCTCCTAATTTATATTTAAGCAAAAA 1184
          ********************************************************************* ************************ *****

CFB^P     GTAAAAAGAATCATTATCTTGAATTGATATGAAAGTTACAAAAACTTAAAAGATCACAGTTGGCATGACATACCTGAGCCATTCGAAACTCAGGCGAAAT 1286
CFB^TL    GTAAAAAGAATCATTATCTTGAATTGATATAAAAGTTACAAAAACTTAAAAGATCACAGTTGGCATGACATACCTGAGCCATTCGAAACTCAGGCGAAAT 1285
CFB^TR    GTAAAAAGAATCATTATCTTGAATTGATATGAAAGTTACAAAAACTTAAAAGATCACAGTTGGCATGACATACCTGAGCCATTTG-AACGCAT------- 1276
          ****************************** **************************************************** * *** **

CFB^P     ATATGTCCATGAGAACGCATTGTCATAATACTTGTGACTAGCTTTACACTTGTACCTTCCATATTTGTACAATTTTTTTTGGGACTTGGGATAGCCCATC 1386
CFB^TL    ATATGTCCATGAGAACGCATTGTCATAATACTTGTGACTAGCTTTACACTTGTACCTTCCATATTTGTACAATTTTTTTTGAGACTTGGGATAGCCTATC 1385
CFB^TR    ---TATCCATGAGAACGCATTGTCATAATACTTGTGACTAGCTTTACACTTGTACCTTCCATATTTGTACAATTTTTTTTGGGACTTGGGATAGCCCATC 1373
             * **************************************************************************** ************** ***

CFB^P     CGAAATCGCACGGCCCTAGCTGAATATATGGAACACCTAGTTTCCAAAGATGGCACTGAGCAAACCCAAAACTTCCATTAAACAACGGAGTTATGAAAGA 1486
CFB^TL    CGAAATCGCACGGCCCTAGCTGAATATATGGAACACCTACTTTCCAAAGATGGCACTGAACAAACCCAAAACTTCCATTAAACAACGGAGTTATGAAAGA 1485
CFB^TR    CGAAATCGCACGGCCCTAGCTGAATATATGGAACACCTACTTTCCAAAGATGGCACTGAACAAACCCAAAACTTCCATTAAACAACGGAGTTATGAAAGA 1473
          *************************************** ****************** *****************************************

CFB^P     ATAATTG-AAAAACAGGTCAGAAAATCACCTTGAAAGGGTTGAACGTGAATCTCCGATACAATGCAGATATTGGAAATCAAGTTGTAAACTAGTTTTTGG 1585
CFB^TL    ATAATTGAAAAAACAGGTCAGAAAATCACCTTGAAAGGGTTGAACGTGAATCTCCGATACAATGCAGATATTGGAAATCAAGTTGTAGACTAGTTTTTGG 1585
CFB^TR    ATAATTGAAAAAACAGGTCAGAAAATCACCTTGAAAGGGTTGAACGTGAATCTCCGATACAATGCAGATATTGGAAATCAAGTTGTAGACTAGTTTTTGT 1573
          ******* ******************************************************************************* ***********

CFB^P     GGAACTGAGGGATCAGTTTGGCCTGTGCTTGACCAATAAGAATCTCGCACACCTATCCTCTCTCTTGGTTCAAGAGTATTTTGTATACATTCTTGAGGAT 1685
CFB^TL    GGAACTGAGGGATCAGTTTGGCCTGTGCTTGACCAATAAGAATCTCGCACACCTATCCTCTCTCTTGGTTCAAGAGTATTTTGTATACATTCTTGAGGAT 1685
CFB^TR    GGAACTGTGGGATCGGTTTGGCCTGTGCTTGACCAATAAGAATCTCGCACAGCTATCCTCTCTCTTGGTTCAAGAGTATTTTGTATACATTCTTGAGGAT 1673
          ******* ****** ************************************ ************************************************

CFB^P     AATTATCAGTACTGGATGCACTAACCGCAGTAGATATACAAAATTTTGAATTGGATTTAGCAAGACCTCGGGATAAATATGCGTAAACCTTGTGATCGCT 1785
CFB^TL    AATTATCAGTACTGGATGCACTAACCGCAGTAGATATACACAATTTTGAATTGGATTTAGCAAGACCTCGGGATAAATATGCGTAAACCTTGTGATCGCT 1785
CFB^TR    AATTATCAGTACTGAATGCACTAACCGCAGTAGATATAGACAATTTTGAATTGGATTTAGCAAGACCTCGGGATAAATATGCGTAAACCTTGTGATCGCT 1773
          ************** *********************** * ***********************************************************
```

```
CFB^P     CTCCAGATTTTTTTGTTGCCTAGAATTGAATGACCCCGCTTGTTTGCTTTTCTCTTTGATAATAATCGGGATTTTCTCTAATTCGGGAAAGTTTTTTAGG 1885
CFB^TL    CTCCAGATTTTTTTGTTCCCTAGAATTGAATGACCCCGCTTGTTTGCTTTTCTCTTTGATAATAATCGGGATTTTCTCTAATTCGGGAAAGTTTTTTAGG 1885
CFB^TR    CTCCAGATTTTTTTGTTCCCTAGAATTGAATGACCCCGCTTGTTTGCTTTTCTTTTTGATAATAATCGGGATTTTCTCTAATTCGGGAAAGTTTTTTAGG 1873
          ***************** *********************************** **********************************************

CFB^P     CATATCTGCTTAGAGAGTCCGTTCCTGATCGCTGCTTAAGATATATTGGTCATCAATAACCATCCATATAATATCCAAAAACCAATACAATAGTCACTAA 1985
CFB^TL    CATATCTGCTTAGAGAGTCCGTTCCTGATCGCTGCTTAAGATATATTGGTCATCAATAACCATCCATATAATATCCAAAAACCAATACAATAGTCACTAA 1985
CFB^TR    CATATCTGCTTAGAGAGTCCGTTCCTGATCGCTGCTTAAGATATATTGGTCATCAATAACCATCCATATAATATCCAAAAACCAATACAATAGTCACTAA 1973
          ****************************************************************************************************

CFB^P     AATGGAAGACTTGAAATGTCATTTTTTATTTCATCCAAAAGTAAGTAATGAAACCACCTGCTGACTGCTGTTTATGTAGGTCTCATGAAATAGAGTGTTT 2085
CFB^TL    AATGGAAGACTTGAAATGTCATTTTTTATTTCATCCAAAAGTAAGTAATGAAACCACCTGCTGACTGCTGTTTATGTAGGTCTCATGAAATAGAGTGTTT 2085
CFB^TR    AATGGAAGACTTGAAATGTCATTTTTTATTTCATCCAAAAGTAAGTAATGAAACCACCTGCTGACTGCTGTTTATGTAGGTCTCATGAAATAGAGTGTTT 2073
          ****************************************************************************************************

CFB^P     TTAAGAAAATCCTTACTAAAAACGGATGAGGGGTATAAGTGAACGATAGTCCAAGAGATGCATGCAAATAAAAATAAACTAACCAAAATCACGCCAAGAG 2185
CFB^TL    TTAAGAAAATCCTTACTAAAAACGGATGAGGGGTATAAGTGAACGATAGTCCAAGAGATGCATGCAAATAAAAATAAACTAACCAAAATCACGCCAAGAG 2185
CFB^TR    TTAAGAAAATCCTTACTAAAAACGGATGAGGGGAATAAGTGAACGATAGTCCAAGAGATGCATGCAAATAAAAATAAACTAACCAAAATGACGCCAAGAG 2173
          ********************************* ******************************************************* **********

CFB^P     GTTGAAACTGTGCTAAAACGTATAATATCCGACGGATTCTCAAAAGACAATAGGATTCTCAAGGATAAATCATGTCCAAGCCATTGTATAAAGTCTCTGG 2285
CFB^TL    GTTGAAACTGTGCTAAAACGTATAATATCCGACGGATTCGCAAAAGACAATAGGATTCTCAAGGATAAATCATGTCCAAGCCATTGTATAAAGTCTCTGG 2285
CFB^TR    GTTGAAACTGCGCTAAAACGTATAATATCCGACTGATTCTCTAAAGACGATAGGATTCTCAAGGATAAATCATGTCCAAGCCATTGTATAAAGTCTCTGG 2273
          ********** ********************** ***** * ****** ***************************************************

                         ←
CFB^P     AACTTGCATGGATTTCCATAGTGGATATTGAAATAAATTATA---TGTCGAAGGTCACTGCCTTTCTTAGTTACATCAAAATATAGTGGCACTATATATA 2382
CFB^TL    AACTTGCATGGATTTCCATAGTGGATATTGAAATAAATAATAGTATGTCGAGAGTCACTGCCTTTCTTAGTTACATCAAAATATAGTGGCACTATATATA 2385
CFB^TR    AACTTGCATGGATTTCCATAGTGGATATTGAAATAAATAATAGTATGTCGAGAGTCACTGCCTTTCTTAGTTACATCAAAATATAGTGGCACTATATATA 2373
          ************************************** ***   ******  ***********************************************

CFB^P     GGATAAGAAATTGTTTAGTTCTACCATTAGTTACATCAAAATAGAAACTTTCAGCAAATAAAAGTCATAAAAGTATACCACGTGAGTGACGTTTGGAGCC 2482
CFB^TL    GGATAAGAAATTGTTTAGTTCTACCATTAGTTACATCAAAATAGAAACTTTCAGCAAATAAAAGTCATAAAAGTATACCACGTGAGTGACGTTTGGAGCC 2485
CFB^TR    GGATAAGAAATTGTTTAGTTCTACCATTAGTTACATCAAAATAGAAACTTTCAGCAAATAAAAGTCATAAAAGTATACCACGTGAGTGACGTTTGGAGCC 2473
          ****************************************************************************************************

                                                                      s-F (Pin-specific)
CFB^P     TAATACTTTTCCTTGTAAAATCTCTATTTTCTACCTCTTTTTGTCTCATTGCTGCTCCGTTGAAAGAGTATGAAATAGTTTTACCTTTTTTTTTTTAATT 2582
CFB^TL    TAATACTTTTCCTTGTAAAATCTCTATTTTCTAC---------TCTCTTT-------------------------------------------------- 2526
CFB^TR    TAATACTTTTCCTTGTAAAATCTCTATTTTCTAC---------TCTCTTT-------------------------------------------------- 2514
          **********************************         **** **
                                        SRB-F (Thrum-specific)

CFB^P     TTCCTTAGTTTGGATGAACCAGGTCGTAACAAAACCCAACTTTTCCAGTTAATGGATGATCGTAATTTAGTAGCGGTCCTTATTAGTTACCCATTATTTT 2682
CFB^TL    -----TAGTTTGGATGAACCAGGTCGTAACAAAACCCAACTTTTCCAGTTAATGGATGATAGTAATTTAGTAGCGGTCCTTATTAGTTACCCATTATTTT 2621
CFB^TR    -----TAGTTTGGATGAACCAGGTCGTAACAAAACCCAACTTTTCCAGTTAATGGATGATAGTAATTTAGTAGCGGTCCTTATTAGTTACCCATTATTTT 2609
               ******************************************************* ***************************************
SRB-F continued (Thrum-specific)

CFB^P     AAATTTTGATTATTCACTTGATGATTATCTATGTTCCACAATCGATCTTTATCCGTTAAATATCTAGTGGACAAAACAGAAGAATACATGTGGAATATTT 2782
CFB^TL    AAATTTTGATTATTCACTTGATGATCATCTATGTTCCACAATCGATCTTTATCCGATAAATATCTAGTGGACAAAACAGAAGAATACATGTGGAATATTT 2721
CFB^TR    AAATTTTGATTATTCACTTGATGATCATCTATGTTCCACAATCGATCTTTATCCGATAAATATCTAGTGGACAAAACAGAAGACTACATGTGGAATATTT 2709
          ************************* **************************** *************************** *****************

CFB^P     TCCTACTTTTTTCCTTTTTTCTCAAGAAAAGACTCCAGGGTCTCGTTCTGGAAAAATATAATTAATTAGTTTATAATCGAAGTTCATAACGTTATATGAT 2882
CFB^TL    TCCTACTTTTTTCCTTTTTTCTCAAGAAAAGACTCCAGGGTCTCGTTCTGGAAAAATATAATTAATTAGTTTATAATTGAAGTTCATAACGTTATATGAT 2821
CFB^TR    TCCTACTTTTTTCCTTTTTTCTCAAGAAAAGACTCCAGGGTCTCGTTCTGGAAAAATATAATTAATTAGTTTATAATCGAAGTTCATAACTTTATATGAT 2809
          ***************************************************************************** ************ *********

  SLB-F (All sequences)
CFB^P     CGAATTGGACTGATTCAGATGTTTAACACTTCATATATACTTGTAGCGGGTATACCCACTAATTTAACAAATAGTATTAACTATTTTATTTTAATCTAAC 2982
CFB^TL    CGAATTGGACTGATTCAGATGTTTAACACTTCATATATACTTGTAGCGGGTATACCCACTAATCTAACAAATAGTATTAACTATTTTATTTTAAACTAAC 2921
CFB^TR    CGAATTGGACAGATTCAGATGTTTAACACTTCATA----CTTGTAGCGGGTATACCCACTAATCTAACAAATAGTATTAACTATTTTATTTTAAACTAAC 2905
          ********** ************************    ************************ ****************************** *****

CFB^P     CGTAGGTCAAACAATCGTTTAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCAAATATTTAGTCGTGAAAGTGAAACT 3082
CFB^TL    CGTAGGTCAAACAATCGTTTAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCAAATATTTAGTCGTGAAAGTGAAACT 3021
CFB^TR    CGTAGGTCAAACAATCGTTTAAATCTTTAATTTGTTATGCTTTTATTGGCTGGACTTCCCTATTTTGTCCACCAAATATTTAGTCGTGAAAGTGAAACT 3005
          ****************************************************************************************************

CFB^P     GTGAACTTGGTCGGTCTTCTTCTCTGTACCATAGATTGCCTTGCATACGAGATATTAAATAATTTTGGATGATTACATAGATAGATACTTTTGAGTTTTC 3182
CB^TL     GTGAACTTGGTCGGTCTTCTTCTCTGTACCATAGATTGCCTTGCATGCGAGATATTAAATAATTTTGGATGATTACATAGATAGATACTTTTGAGTTTTC 3121
CFB^TR    GTGAACTTGGTCGGTCTTCTTCTCTGTACCATAGATTGCCTTGCATGCGAGATATTAAATAATTTTGGATGATTACATAGATAGATATTTTTGAGTTTTG 3105
          ********************************************* ***************************************** ***********

CFB^P     ACACACTAGCTATAAACATATATACTATGACTACGTTCAAAAAAAATTAAAAAAATACTATAACTAGGAAAAAAAATTGAAAAATGATCCGGTATGAATC 3282
CFB^TL    ACACACTAGCTATAAACATATATACTATGACTACGTTCAAAAAAAATTAAAAAAATACTATAACTAGGAAAAAAAATTGAAAAATGATCCGGTATGAATC 3221
CFB^TR    ACACACTAACTATAAACATATATACTCTGACTACGTTCAAAAAAAATTAAAAAAATATTATAGCTAGGAAAAAAAATTGAAAAATGATCCGGTATGAATC 3205
          ******** ***************** ****************************** **** ************ ************************
CFB^P     TGACGCCAGTTGAAACTGATCCTAGTACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTG-TATCAATTTACATTAATTAAAT--GGCTTTA 3379
CFB^TL    TGACGCCAGTTGAAACTGATCCTAATACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTGAAATCCGATCATAATACCAAAAACCGAAATCC 3321
CFB^TR    TGACGCCAGTTGAAACTGATCCTAATACTATGGTTTCAGATTGGTTATGGATTTCAAATAAAATTTTG-TATCAATTTACATTAATTAAAT--GGCTTTA 3302
          ************************ *******************************************  ***   * * * **   ***    *   *
                          s-R (Pin-specific)                     SLB-R (Thrum LB-specific)
CFB^P     AACGAGTCCTTGCATGAGTATGTCAGTTAAACAGCATTTACCAATAGTGTTACTCGTAAAATTATTATACT 3450
CFB^TL    AAAGAAAATTGGTAAATAGCGGAAAATGTAACATAATTTATAATACTAGCTATATATGCATGTGATAA--- 3389
CFB^TR    AACGAGTCCTTGCATGAG--TGTCAGTTAAACAGTATTTACCAATAGTGTTACTCGTAAAATTATTATACT 3371
          ** **    * * *       *  * *  ****  *****  *     * **    *  *  *  **
                        SRB-R (Thrum RB-specific)
```

Sequence Analysis 3 | Sequence comparison of *CYP*^{T}, *GLO*^{T}, *CCM*^{T} and *CFB*

a, Nucleotide sequences of *GLO*^{T} cDNA from *S* (from RNA-Seq data) and *S*^{LH1} (predicted from genomic sequence) haplotypes are shown. Start codons (ATG) (white highlighted green), stop codons (TAA) white highlighted in red. Retro-transposon insertion in *GLO*^{T} *S*^{LH1} (RTinsert) (white highlighted grey). Sequence length is shown. Predicted amino acid sequences are presented in single letter code, stop codon as *, and lengths indicated. Common sequence in *S* haplotype *GLO*^{T} and *S*^{LH1} haplotype *GLO*^{T} before the point of divergence by the retrotransposon insertion is underlined. **b,** Predicted cDNA sequences based on manually annotated *Cyclin-like F Box* (*CFB*) genes showing the pin *CFB*^{P} *s* allele long homostyle[1] *CFB*^{TL} and *CFB*^{TR} *S*^{LH1} allele. Start codons (ATG) as white text highlighted green, stop codons (TAA) as white text highlighted red. The 11 base deletion, which causes a premature stop codon (TAA) in the *CFB*^{TR} *S*^{LH1} allele is indicated (Δ11), white text highlighted blue, this and the corresponding region in *CFB*^{P} *s* and *CFB*^{TL} *S*^{LH1} is shown, white highlighted grey. Single nucleotide polymorphisms between different alleles not shown. The length of each sequence is shown. Predicted amino acid sequences are shown in single letter code, stop codon as *, and lengths indicated. Changes arising from coding sequence SNPs are highlighted grey. Amino acid sequence in CFB^{TR} corresponding to the CFB^{P} and CFB^{TL} before changes caused by the 11 bp deletion is underlined. Clustal Omega alignment of predicted CFB polypeptides is shown, identical amino acids (*), amino acids with strongly similar properties (:) and weakly similar properties (.) are indicated. The last two divergent amino acids highlighted in CFB^{P} reflect *CFB*^{P} specific SNPs. **c,** Nucleotide sequences of *CYP*^{T} from the thrum *S* haplotype and long homostyle[1] *S*^{LH1} and *S*^{LH2} haplotypes are shown. The start codons (ATG) are shown white highlighted in green, and stop codons (TAA) white highlighted in red. The position of the single nucleotide insertion (T) in *CYP*^{T} *S*^{LH1} is shown white highlighted in grey; the resulting premature stop codon is shown white highlighted in blue. The G-C transversion in *CYP*^{T} *S*^{LH2} is shown white highlighted in brown. The length of each sequence is shown. *CYP*^{T} *S* allele from RNA-Seq data, *CYP*^{T} *S*^{LH1} allele predicted from genomic sequence and *CYP*^{T} *S*^{LH2} allele from cDNA sequence. Predicted amino acid sequences are presented in single letter code, stop codon as *, and lengths indicted. Common sequence between the full length *S* haplotype CYP^{T} and *S*^{LH1} haplotype CYP^{T} before the point of divergence caused by the base insertion is underlined. The D-H substitution in *CYP*^{T} *S*^{LH2} is shown white highlighted in brown. **d**, Clustal Omega alignment of predicted peptide sequences containing the Conserved Cysteine Motif from *CCM*^{T} compared to the non-*S* locus *P. vulgaris CCM* sequence, and other representative sequences from species as shown. Identical amino acids in red (*), amino acids with strongly similar properties in blue (:) and weakly similar properties in yellow (.) are indicated. The consensus Conserved Cysteine Motif is defined.

a

***GLO*^{T} *S* allele**

```
ATGGGGAGAGGAAAGGTAGAGATAAAGAGGATTGAAAACTCGAATATCAGACAAGTGACGTATTCAAACAGGAGAAATGGGATACTGAAAAAGGCCAAGG
AGATCTCGGTTTTGTGTGATGCTCAGGTCTCCCTTATTATTTTCTCTAGCTCCAGTAAGATGCATGATTACTGCAGTCCAAATTCCTCGTTAATTAACAT
CTTGGATGCATATCAGAAGCAATCTGGGATTAGGTTGTGGGATGCTAGACATGAGAACCTTAGCAATGAAATTGAGAGGGTCAAAAAAGAGAATGACAAT
ATGCAGATTGAGCTCAGATACTTGAAGGGAGAAGATATACAATCTTTGCACCACAAGGAGCTCATGTCTATAGAAGATGCACTCGAAAATGGACTAACTC
GTGTTCGCGAGAGACAGATGGAGATCTACAGAATGGCAAAAGACAATTTCGCTGATAAAGAAAGGCTTCTAGAAGATGAGAACAAGCGCCTTGGCTACAA
ATTTCAGCAAGTGATGGATATGCAGATGCCTTGCTCCTACCGTGTACAGCCGCTTCAACCAAATTTACACGATCAGTTTTAA 582
```

***GLO*^{T} *S*^{SH1} allele**

```
ATGGGGAGAGGAAAGGTAGAGATAAAGAGGATTGAAAACTCGAATATCAGACAAGTGACGTATTCAAACAGGAGAAATGGGATACTGAAAAAGGCCAAGG
AGATCTCGGTTTTGTGTGATGCTCAGGTCTCCCTTATTATTTTCTCTAGCTCCGGTAAGATGCATGATTACTGCAGTCCAAATTCCTCGTTAATTAACAT
CTTGGATGCATATCA(RTinsert)GAAGCAATCTGGGATTAGGTTGTGGGATGCTAGACATGAGAACCTTAGCAATGAAATTGAGAGGGTCAAAAAAGAGAA
TGACAATATGCAGATTGAGCTCAGATACTTGAAGGGAGAAGATATACAATCTTTGCACCACAAGGAGCTCATGTCTATAGAAGATGCACTCGAAAATGGA
CTAACTCGTGTTCGCGAGAGACAGATGGAGATCTACAGAATGGCAAAAGACAATTTCGCTGATAAAGAAAGGCTTCTAGAAGATGAGAACAAGCGTCTTG
GCTACAAATTTCAGCAAGTGATGGATATGCAGATGCCTTGCTCCTACCGTGTACAGCCGCTTCAACCAAATTTACACGATCAGTTTTAA 582
```

GLO^{T} *S* predicted protein

```
MGRGKVEIKRIENSNIRQVTYSNRRNGILKKAKEISVLCDAQVSLIIFSSSSKMHDYCSPNSSLINILDAYQKQSGIRLWDARHENLSNEIERVKKENDN
MQIELRYLKGEDIQSLHHKELMSIEDALENGLTRVRERQMEIYRMAKDNFADKERLLEDENKRLGYKFQQVMDMQMPCSYRVQPLQPNLHDQF* 193
```

GLO^{T} *S*^{SH1} predicted protein

```
MGRGKVEIKRIENSNIRQVTYSNRRNGILKKAKEISVLCDAQVSLIIFSSSGKMHDYCSPNSSLINILDAYHVNYTNWDQNPYLEENEKRYRERKQLFLL
SVLTVNDTDDDYL* 113
```

b

CFB^{P} *s* allele

```
ATGGAAATCCATGCAAGTTCCAGAGACTTTATACAATGGCTTGGACATGATTTATCCTTGAGAATCCTATTGTCTTTTGAGAATCCGTCGGATATTATAC
GTTTTAGCACAGTTTCAACCTCTTGGCGTGATTTTGCGATCAGGAACGGACTCTCTAAGCAGATATGCCTAAAAAACTTTCCCGAATTAGAGAAAATCCC
GATTATTATCAAAGAGAAAAGCAAACAAGCGGGGTCATTCAATTCTAGGCAACAAAAAAATCTGGAGAGCGATCACAAGGTTTACGCATATTTATCCCGA
GGTCTTGCTAAATCCAATTCAAAATTTTGTATATCTACTGCGGTTAGTGCATCCAGTACTGATAATTATCCTCAAGAATGTATACAAAATACTCTTGAAC
CAAGAGAGAGGATAGGTGTGCGAGATTCTTATTGGTCAAGCACAGGCCAAACTGATCCCTCAGTTCCCCAAAAACTAGTTTACAACTTGATTTCCAATAT
CTGCATTGTATCGGAGATTCACGTTCAACCCTTTCAAGTGCCATCTTTGGAAACTAGGTGTTCCATATATTCAGCTAGGGCCGTGCGATTTCGGATGGGC
TATCCCAAGTCCCAAAAAAAAATTGTACAAATATGGAAGGTACAAGTGTAAAGCTAGTCACAAGTATTATGACAATGCGTTCTCATGGACATATATTTCGC
CTGAGTTTCGAATGGCTCAGGAGAACTGCTTGCAAAAATTTAAGCTCCCTGAACCTGTTCTTTGCGTTGGTGGAATCTTACAAATGGAACTATTGGGACG
GGTACAAATCCACGAAACTGATGGCTTATACTATTTACGTATCACACACGTACAAGTTGTTGGGAAACCCCTCTCCCCGGGATTCGAGGTTGAAATTATT
GATAGCTCGGGAAGATGTAAACTAGATTACTACTATGGGGTGAATTACCGGTTTTCTCATGAAAATTCAATTATGGTCACTAGTAGAAACTCTTGTCTTA
GAAGAGCCTCAGGCTTAGAGGTTGTGATTCAAGGAGTGAAGGACTCGAACTCAAAATTACGAAGGGGTTTTGTTATCCAATTTCTAATTGTTTTCTTCCT
TTTCCTCCTCTTTCTCATTAAATTTCCAAAAGTTCCCTTCTATGTGACATAA 1152
```

CFB^{TL} *S*^{LH1} allele

```
ATGGAAATCCATGCAAGTTCCAGAGACTTTATACAATGGCTTGGACATGATTTATCCTTGAGAATCCTATTGTCTTTTGCGAATCCGTCGGATATTATAC
GTTTTAGCACAGTTTCAACCTCTTGGCGTGATTTTGCGATCAGGAACGGACTCTCTAAGCAGATATGCCTAAAAAACTTTCCCGAATTAGAGAAAATCCC
GATTATTATCAAAGAGAAAAGCAAACAAGCGGGGTCATTCAATTCTAGGGAACAAAAAAATCTGGAGAGCGATCACAAGGTTTACGCATATTTATCCCGA
GGTCTTGCTAAATCCAATTCAAAATTGTGTATATCTACTGCGGTTAGTGCATCCAGTACTGATAATTATCCTCAAGAATGTATACAAAATACTCTTGAAC
CAAGAGAGAGGATAGGTGTGCGAGATTCTTATTGGTCAAGCACAGGCCAAACTGATCCCTCAGTTCCCCAAAAACTAGTCTACAACTTGATTTCCAATAT
CTGCATTGTATCGGAGATTCACGTTCAACCCTTTCAAGTGCCATCTTTGGAAAGTAGGTGTTCCATATATTCAGCTAGGGCCGTGCGATTTCGGATAGGC
TATCCCAAGTCTCAAAAAAAATTGTACAAATATGGAAGGTACAAGTGTAAAGCTAGTCACAAGTATTATGACAATGCGTTCTCATGGACATATATTTCGC
CTGAGTTTCGAATGGCTCAGGAGAACTGTTTGCAAAAATTTAAGCTCCCTGAACCTGTTCTTTGCGTTGGTGGAATCTTACAAATGGAACTATTGGGACG
GGTACAAATCCAAGAAACTGATGGCTTATACTATTTACGTATCACACACGTACAAGTTGTTGGGAAACCCCTCTCCCCGGGATTCGAGGTTGAAATTATT
GATAGCTCGGGAAGATGTAAACTAGATTACTACTATGGGGTGAATTACCGGTTTTCTCATGAAAATTCAATTATGGTCACTAGTAGAAACTCCAGTCTTA
GAAGAGCCTCAGGCTTAGAGGTTGTGATTCAAGGAGTGAAGGACTCGAACTCAAAATTACGAAGGGGTTTTGTTATCCAATTTCTAATTGTTTTCTTCGT
TTTCCTCCTCTTTCTCATTAAATTTCCAAAAGTTCCCTTCTATGTGACATAA 1152
```

CFB^{TR} *S*^{LH1} allele

```
ATGGAAATCCATGCAAGTTCCAGAGACTTTATACAATGGCTTGGACATGATTTATCCTTGAGAATCCTATCGTCTTTAGAGAATCAGTCGGATATTATAC
GTTTTAGCGCAGTTTCAACCTCTTGGCGTCATTTTGCGATCAGGAACGGACTCTCTAAGCAGATATGCCTAAAAAACTTTCCCGAATTAGAGAAAATCCC
GATTATTATCAAAAAGAAAAGCAAACAAGCGGGGTCATTCAATTCTAGGGAACAAAAAAATCTGGAGAGCGATCACAAGGTTTACGCATATTTATCCCGA
GGTCTTGCTAAATCCAATTCAAAATTGTCTATATCTACTGCGGTTAGTGCATTCAGTACTGATAATTATCCTCAAGAATGTATACAAAATACTCTTGAAC
CAAGAGAGAGGATAGCTGTGCGAGATTCTTATTGGTCAAGCACAGGCCAAACCGATCCCACAGTTCCACAAAAACTAGTCTACAACTTGATTTCCAATAT
CTGCATTGTATCGGAGATTCACGTTCAACCCTTTCAAGTGCCATCTTTGGAAAGTAGGTGTTCCATATATTCAGCTAGGGCCGTGCGATTTCGGATGGGC
TATCCCAAGTCCCAAAAAAAAATTGTACAAATATGGAAGGTACAAGTGTAAAGCTAGTCACAAGTATTATGACAATGCGTTCTCATGGATA(Δ11)ATGCGT
TCAAATGGCTCAGGAGAACTGCTTGCAAAAATTTAAGCTCCCTGAACCTGTTCTTTGCGTTGGTGGAATCTTACAAATGGAACTATTGGGACGGGTACAA
ATCCAAGAAACTGATGGCTTATACTATTTACGTATCACACACGTACAAGTTGTTGGGAAACCCCTCTCCCCGGGATTCGAGGTTGAAATTATTGATAGCT
CGGGAAGATGTAAACTAGATTACTACTATGGGGTGAATTACCGGTTTTCTCATGAAAATTCAATTATGGTCACTAGTAGAAATTCCAGTCTTAGAAGAGC
CTCAGGCTTAGAGGTTGTGATTTAAGGGGTGAAGGACTCGAACTCAAAATTACGAAGGGGTTTTGTTATCCAATTTCTAATTGTTTTCTTCGTTTTCCTC
CTCTTTCTCATTAAATTTCCAAAAGTTCCCTTCGATGTGACATAA 1141
```

CFB^{P} *s* predicted protein

```
MEIHASSRDFIQWLGHDLSLRILLSFENPSDIIRFSTVSTSWRDFAIRNGLSKQICLKNFPELEKIPIIIKEKSKQAGSFNSRQQKNLESDHKVYAYLSR
GLAKSNSKFCISTAVSASSTDNYPQECIQNTLEPRERIGVRDSYWSSTGQTDPSVPQKLVYNLISNICIVSEIHVQPFQVPSLETRCSIYSARAVRFRMG
YPKSQKKLYKYGRYKCKASHKYYDNAFSWTYISPEFRMAQENCLQKFKLPEPVLCVGGILQMELLGRVQIHETDGLYYLRITHVQVVGKPLSPGFEVEII
DSSGRCKLDYYYGVNYRFSHENSIMVTSRNSCLRRASGLEVVIQGVKDSNSKLRRGFVIQFLIVFFLFLLFLIKFPKVPFYVT* 383
```

CFB^{TL} *S*^{LH1} predicted protein

```
MEIHASSRDFIQWLGHDLSLRILLSFANPSDIIRFSTVSTSWRDFAIRNGLSKQICLKNFPELEKIPIIIKEKSKQAGSFNSREQKNLESDHKVYAYLSR
GLAKSNSKLCISTAVSASSTDNYPQECIQNTLEPRERIGVRDSYWSSTGQTDPSVPQKLVYNLISNICIVSEIHVQPFQVPSLESRCSIYSARAVRFRIG
YPKSQKKLYKYGRYKCKASHKYYDNAFSWTYISPEFRMAQENCLQKFKLPEPVLCVGGILQMELLGRVQIQETDGLYYLRITHVQVVGKPLSPGFEVEII
DSSGRCKLDYYYGVNYRFSHENSIMVTSRNSSLRRASGLEVVIQGVKDSNSKLRRGFVIQFLIVFFVFLLFLIKFPKVPFYVT* 383
```

CFB^{TR} *S*^{LH1} predicted protein

```
MEIHASSRDFIQWLGHDLSLRILSSLENQSDIIRFSAVSTSWRHFAIRNGLSKQICLKNFPELEKIPIIIKKKSKQAGSFNSREQKNLESDHKVYAYLSR
GLAKSNSKLSISTAVSAFSTDNYPQECIQNTLEPRERIAVRDSYWSSTGQTDPTVPQKLVYNLISNICIVSEIHVQPFQVPSLESRCSIYSARAVRFRMG
YPKSQKKLYKYGRYKCKASHKYYDNAFSWIMRSNGSGELLAKI* 243
```

CLUSTAL Omega alignment

```
CFBTR SLH1  MEIHASSRDFIQWLGHDLSLRILSSLENQSDIIRFSAVSTSWRHFAIRNGLSKQICLKNFPELEKIPIIIKKKSKQAGSFNSREQKNLESDHKVYAYLSR
CFBP s      MEIHASSRDFIQWLGHDLSLRILLSFENPSDIIRFSTVSTSWRDFAIRNGLSKQICLKNFPELEKIPIIIKEKSKQAGSFNSRQQKNLESDHKVYAYLSR
CFBTL SLH1  MEIHASSRDFIQWLGHDLSLRILLSFANPSDIIRFSTVSTSWRDFAIRNGLSKQICLKNFPELEKIPIIIKEKSKQAGSFNSREQKNLESDHKVYAYLSR
            *********************** *: * *******:******.***************************:**********:*****************

CFBTR SLH1  GLAKSNSKLSISTAVSAFSTDNYPQECIQNTLEPRERIAVRDSYWSSTGQTDPTVPQKLVYNLISNICIVSEIHVQPFQVPSLESRCSIYSARAVRFRMG
CFBP s      GLAKSNSKFCISTAVSASSTDNYPQECIQNTLEPRERIGVRDSYWSSTGQTDPSVPQKLVYNLISNICIVSEIHVQPFQVPSLETRCSIYSARAVRFRMG
CFBTL SLH1  GLAKSNSKLCISTAVSASSTDNYPQECIQNTLEPRERIGVRDSYWSSTGQTDPSVPQKLVYNLISNICIVSEIHVQPFQVPSLESRCSIYSARAVRFRIG
            ********:.******* ********************.**************:******************************:*************:*

CFBTR SLH1  YPKSQKKLYKYGRYKCKASHKYYDNAFSWIMRSNGSGELLA----KI-----------------------------------------------------
CFBP s      YPKSQKKLYKYGRYKCKASHKYYDNAFSWTYISPEFRMAQENCLQKFKLPEPVLCVGGILQMELLGRVQIHETDGLYYLRITHVQVVGKPLSPGFEVEII
CFBTL SLH1  YPKSQKKLYKYGRYKCKASHKYYDNAFSWTYISPEFRMAQENCLQKFKLPEPVLCVGGILQMELLGRVQIQETDGLYYLRITHVQVVGKPLSPGFEVEII
            *****************************   *        *****::****************************************************

CFBTR SLH1  -----------------------------------------------------------------------------------
CFBP s      DSSGRCKLDYYYGVNYRFSHENSIMVTSRNSCLRRASGLEVVIQGVKDSNSKLRRGFVIQFLIVFFLFLLFLIKFPKVPFYVT
CFBTL SLH1  DSSGRCKLDYYYGVNYRFSHENSIMVTSRNSSLRRASGLEVVIQGVKDSNSKLRRGFVIQFLIVFFVFLLFLIKFPKVPFYVT
            ***********************************************************************************
```

c

CYP^{T} *S* allele

ATGTTTTTCATAAGTTGTAAGCTGGCAAAGAGAGGAAAAATGCAAGTGAGCTTTGTGTTGTGCTTCTTACTCCTATGTTTTTATATATATATATAATTTTTG
CTTTCCTGATTAGAGCGGGATATTATCTGTGGTGGAGGCCCAGGAGAATCCAGCTGCACTTCTCTAAGCATGGTATTAGAGGTCCAAACTACCACGTTCT
TTCAGGCAACACAAAGGAGCTTGTTGATTTGACGGTTAAGGCTGCTTCTCAAACCTTTCCTCGTTCTCCTCATAACATTGTCCCTAAGGTTCTCCCCTTC
TATCACCAATGGAAGAAAATATATGGTGCAACTTTTTTATTGTGGTTTGGACCAGTGGCTAATCTCACGCTGTCTGATCCGGTGCTTATCACAGAGATTT
TGATTAGTAAATCATCGGAATTATTTGAGAAAACTGAGTCACCCCAACATGTTAGAAAAGTTGAAGGTGATGGTTTGATTACTCTTCGAGGAGAGAAATG
GGTTCACCACAGAAAAATCATCACACCGTCCTTCTACATTGATAATCTCAAGCTAATGGTGCCTATTATGGGAAATAGCATGGTAACGATGTTGAACAAA
TGGGTGGAGATTTCCAAAAACAGTACCATCGAAATTGATGTGTCACATTGGTTCCAAGATCTCACTGAGGAGATAATTAGCCACATAGCCTTTGGAAGAA
GCTGTAAAGAGGGAAAACCAATATTTGCGTTACATTCACAAAATATGGCTTATGCAATTGGTTCCTATAACAAGTTCTTCATTTCAGCGTACAGGTTTCT
TCCCACTAAACAGAATAGGCAATTTTGCAAACTGAATAGAGAAATGAAAATATCCTTGACGAAGCTGATAAACCAAAGAATGAAGGACAATAATAATTTT
ATAGGAGCATCTTCAGAACAGTGTCCAGACGATTTATTGGAACTCATGGTTAAAGCCAGCAAAAAGAATGTTCAGACAAAGCAAATGTCAGCACGGGATG
CATTTACAACCTACAACATAATTGAAGAGTGCAAGACCATCCTATTTGCTGGAAAATACACTACCTCGGCCATGATGACGTGGACCACTGTTCTCCTTGC
AATGCACCCTCTGTGGCAGGAACTGGCTCGCAAAGAGGTCCTAAGGGTGTGCATGGATCACGATTTTCCGACAAAAGATGATGTCACGAAACTTAAAACA
CTCAGCATGATATTAAATGAATCACTAAGACTTTACCCACCAGTTGTTGCTTTACTGAGACGGGCCAAGAGTGACATGGAGTTTGGTGGCTGTACGATTT
TACGAGGGACGGAGCTTCTGATCCCAATTGTGGGCATCCATCACGATTTAGAAATATGGAGTCAGGAAGCAACCGAGTTCAATCCGTCAAGGTTCGCTCC
AGGGGTTTCCAAGGCTACAAAGCATCCCACAGCCTTCATGCCATTCGGACTAGGGAATCGTAGATGCGTCGGGCAAAATTTGGCTATCCTCCAGACTAAA
CTGGCCATTGCCATGATCTTAAAACGATTTTCATTCAATCTTGCCCCAAGCTATGAGCACGCACCAACAGTTGTGATGTTCCTTGATCCTCAATACCGTG
CACCCATTACCTTTCATACATTGTAA 1626

CYP^{T} *S*^{LH1} allele

ATGTTTTTCATAAGTTGTAAGCTGGCAAAGAGAGGAAAAATGCAAGTGAGCTTTGTGTTGTGCTTCTTACTCCTATGTTTTTATATATATATATAATTTTTG
CTTTCCTGATTAGAGCGGGATATTATCTGTGGTGGAGGCCCAGGAGAATCCAGCTGCACTTCTCTAAGCATGGTATTAGAGGTCCAAACTACCACGTTCT
TTCAGGCAACACAAAGGAGCTTGTTGATTTGACGGTTAAGGCTGCTTCTCAAACCTTTCCTCGTTCTCCTCATAACATTGTCCCTAAGGTTCTCCCCTTC
TATCACCAATGGAAGAAAATATATGGTGCAACTTTTTTTATTGTGGTTTGGACCAGTGGCTAATCTCACGCTGTCTGATCCGGTGCTTATCACAGAGATT
TTGATTAGTAAATCATCGGAATTATTTGAGAAAACTGAGTCACCCCAACATGTTAGAAAAGTTGAAGGTGATGGTTTGATTACTCTTCGAGGAGAGAAAT
GGGTTCACCACAGAAAAATCATCACACCGTCCTTCTACATTGATAATCTCAAGCTAATGGTGCCTATTATGGGAAATAGCATGGTAACGATGTTGAACAA
ATGGGTGGAGATTTCCAAAAACAGTACCATCGAAATTGATGTGTCACATTGGTTCCAAGATCTCACTGAGGAGATAATTAGCCACATAGCCTTTGGAAGA
AGCTGTAAAGAGGGAAAACCAATATTTGCGTTACATTCACAAAATATGGCTTATGCAATTGGTTCCTATAACAAGTTCTTCATTTCAGCGTACAGGTTTC
TTCCCACTAAACAGAATAGGCAATTTTGCAAACTGAATAGAGAAATGAAAATATCCTTGACGAAGCTGATAAACCAAAGAATGAAGGACAATAATAATTT
TATAGGAGCATCTTCAGAACAGTGTCCAGACGATTTATTGGAACTCATGGTTAAAGCCAGCAAAAAGAATGTTCAGACAAAGCAAATGTCAGCACGGGAT
GCATTTACAACCTACAACATAATTGAAGAGTGCAAGACCATCCTATTTGCTGGAAAATACACTACCTCGGCCATGATGACGTGGACCACTGTTCTCCTTG
CAATGCACCCTCTGTGGCAGGAACTGGCTCGCAAAGAGGTCCTAAGGGTGTGCATGGATCACGATTTTCCGACAAAAGATGATGTCACGAAACTTAAAAC
ACTCAGCATGATATTAAATGAATCACTAAGACTTTACCCACCAGTTGTTGCTTTACTGAGACGGGCCAAGAGTGACATGGAGTTTGGTGGCTGTACGATT
TTACGAGGGACGGAGCTTCTGATCCCAATTGTGGGCATCCATCACGATTTAGAAATATGGAGTCAGGAAGCAACCGAGTTCAATCCGTCAAGGTTCGCTC
CAGGGGTTTCCAAGGCTACAAAGCATCCCACAGCCTTCATGCCATTCGGACTAGGGAATCGTAGATGCGTCGGGCAAAATTTGGCTATCCTCCAGACTAA
ACTGGCCATTGCCATGATCTTAAAACGATTTTCATTCAATCTTGCCCCAAGCTATGAGCACGCACCAACAGTTGTGATGTTCCTTGATCCTCAATACCGT
GCACCCATTACCTTTCATACATTGTAA 1627

CYP^{T} *S*^{LH2} allele

ATGTTTTTCATAAGTTGTAAGCTGGCAAAGAGAGGAAAAATGCAAGTGAGCTTTGTGTTGTGCTTCTTACTCCTATGTTTTTATATATATATATAATTTTTG
CTTTCCTGATTAGAGCGGGATATTATCTGTGGTGGAGGCCCAGGAGAATCCAGCTGCACTTCTCTAAGCATGGTATTAGAGGTCCAAACTACCACGTTCT
TTCAGGCAACACAAAGGAGCTTGTTGATTTGACGGTTAAGGCTGCTTCTCAAACCTTTCCTCGTTCTCCTCATAACATTGTCCCTAAGGTTCTCCCCTTC
TATCACCAATGGAAGAAAATATATGGTGCAACTTTTTTATTGTGGTTTGGACCAGTGGCTAATCTCACGCTGTCTCATCCGGTGCTTATCACAGAGATTT
TGATTAGTAAATCATCGGAATTATTTGAGAAAACTGAGTCACCCCAACATGTTAGAAAAGTTGAAGGTGATGGTTTGATTACTCTTCGAGGAGAGAAATG
GGTTCACCACAGAAAAATCATCACACCGTCCTTCTACATTGATAATCTCAAGCTAATGGTGCCTATTATGGGAAATAGCATGGTAACGATGTTGAACAAA
TGGGTGGAGATTTCCAAAAACAGTACCATCGAAATTGATGTGTCACATTGGTTCCAAGATCTCACTGAGGAGATAATTAGCCACATAGCCTTTGGAAGAA
GCTGTAAAGAGGGAAAACCAATATTTGCGTTACATTCACAAAATATGGCTTATGCAATTGGTTCCTATAACAAGTTCTTCATTTCAGCGTACAGGTTTCT
TCCCACTAAACAGAATAGGCAATTTTGCAAACTGAATAGAGAAATGAAAATATCCTTGACGAAGCTGATAAACCAAAGAATGAAGGACAATAATAATTTT
ATAGGAGCATCTTCAGAACAGTGTCCAGACGATTTATTGGAACTCATGGTTAAAGCCAGCAAAAAGAATGTTCAGACAAAGCAAATGTCAGCACGGGATG
CATTTACAACCTACAACATAATTGAAGAGTGCAAGACCATCCTATTTGCTGGAAAATACACTACCTCGGCCATGATGACGTGGACCACTGTTCTCCTTGC
AATGCACCCTCTGTGGCAGGAACTGGCTCGCAAAGAGGTCCTAAGGGTGTGCATGGATCACGATTTTCCGACAAAAGATGATGTCACGAAACTTAAAACA
CTCAGCATGATATTAAATGAATCACTAAGACTTTACCCACCAGTTGTTGCTTTACTGAGACGGGCCAAGAGTGACATGGAGTTTGGTGGCTGTACGATTT
TACGAGGGACGGAGCTTCTGATCCCAATTGTGGGCATCCATCACGATTTAGAAATATGGAGTCAGGAAGCAACCGAGTTCAATCCGTCAAGGTTCGCTCC
AGGGGTTTCCAAGGCTACAAAGCATCCCACAGCCTTCATGCCATTCGGACTAGGGAATCGTAGATGCGTCGGGCAAAATTTGGCTATCCTCCAGACTAAA
CTGGCCATTGCCATGATCTTAAAACGATTTTCATTCAATCTTGCCCCAAGCTATGAGCACGCACCAACAGTTGTGATGTTCCTTGATCCTCAATACCGTG
CACCCATTACCTTTCATACATTGTAA 1626

CYP^{T} *S* predicted protein

MFFISCKLAKRGKMQVSFVLCFLLLCFYIYIIFAFLIRAGYYLWWRPRRIQLHFSKHGIRGPNYHVLSGNTKELVDLTVKAASQTFPRSPHNIVPKVLPF
YHQWKKIYGATFLLWFGPVANLTLSDPVLITEILISKSSELFEKTESPQHVRKVEGDGLITLRGEKWVHHRKIITPSFYIDNLKLMVPIMGNSMVTMLNK
WVEISKNSTIEIDVSHWFQDLTEEIISHIAFGRSCKEGKPIFALHSQNMAYAIGSYNKFFISAYRFLPTKQNRQFCKLNREMKISLTKLINQRMKDNNNF
IGASSEQCPDDLLELMVKASKKNVQTKQMSARDAFTTYNIIEECKTILFAGKYTTSAMMTWTTVLLAMHPLWQELARKEVLRVCMDHDFPTKDDVTKLKT
LSMILNESLRLYPPVVALLRRAKSDMEFGGCTILRGTELLIPIVGIHHDLEIWSQEATEFNPSRFAPGVSKATKHPTAFMPFGLGNRRCVGQNLAILQTK
LAIAMILKRFSFNLAPSYEHAPTVVMFLDPQYRAPITFHTL* 541

CYP^{T} *S*^{LH1} predicted protein

MFFISCKLAKRGKMQVSFVLCFLLLCFYIYIIFAFLIRAGYYLWWRPRRIQLHFSKHGIRGPNYHVLSGNTKELVDLTVKAASQTFPRSPHNIVPKVLPF
YHQWKKIYGATFFIVVWTSG* 120

CYP^{T} *S*^{LH2} predicted protein

MFFISCKLAKRGKMQVSFVLCFLLLCFYIYIIFAFLIRAGYYLWWRPRRIQLHFSKHGIRGPNYHVLSGNTKELVDLTVKAASQTFPRSPHNIVPKVLPF
YHQWKKIYGATFLLWFGPVANLTLSHPVLITEILISKSSELFEKTESPQHVRKVEGDGLITLRGEKWVHHRKIITPSFYIDNLKLMVPIMGNSMVTMLNK
WVEISKNSTIEIDVSHWFQDLTEEIISHIAFGRSCKEGKPIFALHSQNMAYAIGSYNKFFISAYRFLPTKQNRQFCKLNREMKISLTKLINQRMKDNNNF
IGASSEQCPDDLLELMVKASKKNVQTKQMSARDAFTTYNIIEECKTILFAGKYTTSAMMTWTTVLLAMHPLWQELARKEVLRVCMDHDFPTKDDVTKLKT
LSMILNESLRLYPPVVALLRRAKSDMEFGGCTILRGTELLIPIVGIHHDLEIWSQEATEFNPSRFAPGVSKATKHPTAFMPFGLGNRRCVGQNLAILQTK
LAIAMILKRFSFNLAPSYEHAPTVVMFLDPQYRAPITFHTL* 541

d

Species	***gene***	***amino acid sequence over Conserved Cysteine Motif***
Primula vulgaris	CCM^{T}	QETCCRWLKAVDPECVCGLLLRLPTFLARPAHSYTVAVGDACVFTYPCAS
Primula vulgaris	*CCM*	EEICCRWLKDVDHECVCGLLLRFPTFLARPAHSYTVAVGDACMFTYPCAS
Brachypodium distachyon	XP_014751076.1	YRDCCRWLKEVEPGCVCETLLRLPSFLVKPQHTYTVRVGKTCKLEYRCGG
Oryza sativa	EEE67730.1	NRDCCRWLKEVEPACVCDALLRLPPFLVKPQHKYTVKVGDSCKYTYHCGG
Sorghum bicolor	XP_002463297.1	YRDCCRWLKEVEPDCVCEALLRLPPFLVRPQHKYTVKVGNSCKFSYRC--
Zea mays	NP_001140718.1	YRDCCRWLKEVEPDCVCEALLRLPPFLVKPQHKYTVKVGNSCKFTYRCGG
Nelumbo nucifera	XP_010256204.1	SEECCRWLREVDSMCVCQLFVPMPTFLRRPLHSYTIMVDHTCKVVFKCPQ
Medicago truncatula	XP_013466549.1	EDNCCRWARALDSRCVCEILVRLPPFLIRPLHTYSVVFGESCTVTYSCGG
Malus domestica	XP_008368867.1	EDNCCKWLNQLDNECVCELLVHLPNFLIRPAHAYSVAVGETCTVSYTCGT
Erythranthe guttata	XP_012836621.1	EEECCRWLKNIDSVCVCKLLLYLPPFLTRPHHNYGVDVGGLCDISFNCGS
Cucumis sativus	KGN44871.1	EENCCKWVQQVDSECVCELLSRLPAFLKRPIHNFSVTIGGSCNATYWCGG
Populus euphratica	XP_011007542.1	EDTCCRWLNDVDDECICELLVRLPPFLSRPRHEYTIKIDDSCSVSYTCGF
Petunia hybrida	AGJ98242.1	EEECCRWMKTVDSECVCGLLAQLPPFLSRPLHQYTAVVDASCSVTFMCSS
Solanum lycopersicum	XP_004247316.1	EENCCRWMKQVDNECVCDLLVRLPLFLSRPVHQYTILVDPGCNITFECGS
Nicotiana tomentosiformis	XP_009627635.1	EEECCRWMKQVDSECICDLLVHLPPFLSRPVHQYTVVVDASCNVTFECGS
Brassica rapa	XP_009104413.1	QQECCKWLKQMDNECVCDLLVRLPPLLAKPAHNYTVFVDESCIVTYTCGG
Arabis alpina	KFK35474.1	QQECCKWLKQMDNECVCDLLVRLPPLLAKPAHNYTVFVDESCIVTYTCGG
Arabidopsis thaliana	NP_001078333.1	QQECCKWLKQMDNECVCDLLVRLPPLLAKPIHNYTVFVDESCIVTFVCGG
Vitis vinifera	XP_010652136.1	EEACCRWLKEIDDECVCDLLAHLPLFLTRPSHYYTVSVDPSCSVTFSCGG
Eucalyptus grandis	XP_010039258.1	EQDCCKWLKEVDSECVCDLLVRLPSFLTRPIHVYTVTVPDLCNVTFSCGG
		**:* . :: *:* : :* :* :* * : . * : *

consensus Conserved Cysteine Motif: $\mathrm{CC(^{R}/_{K})Wx_{4}(^{D}/_{E})x_{2}CVCx_{2}Lx_{2}LPx(^{F}/_{L})Lx(^{R}/_{K})PxHxYx_{7}Cx_{3}(^{Y}/_{F})xC}$