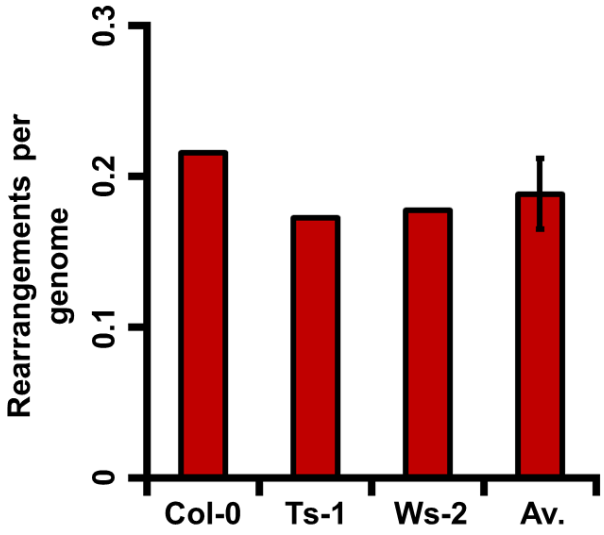
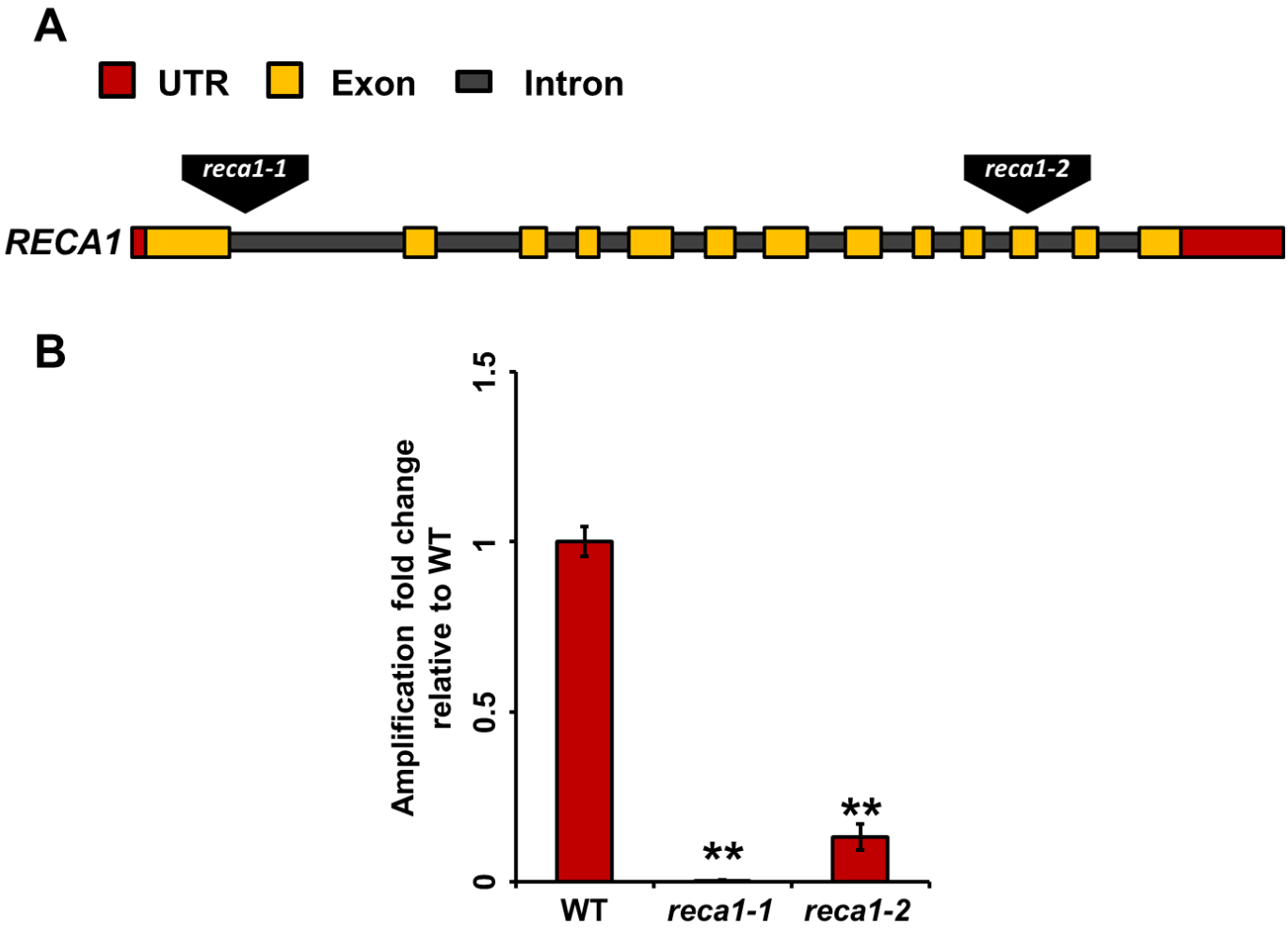
SUPPLEMENTAL MATERIAL

A next-generation sequencing approach reveals U-turn-like inversions as a major source of genomic instability in organelles of *Arabidopsis* and humans

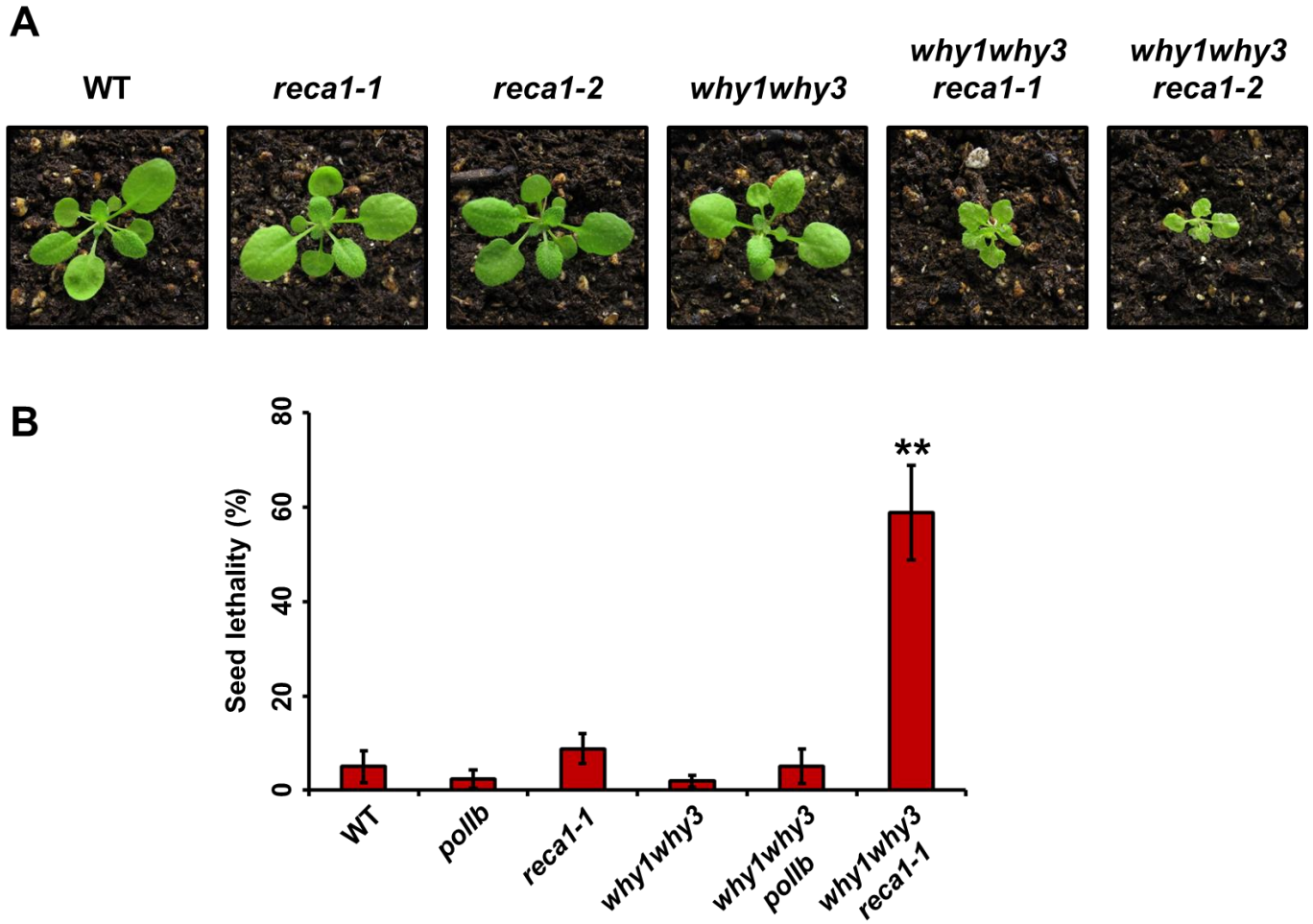
Éric Zampini, Étienne Lepage, Samuel Tremblay-Belzile, Sébastien Truche and Normand Brisson



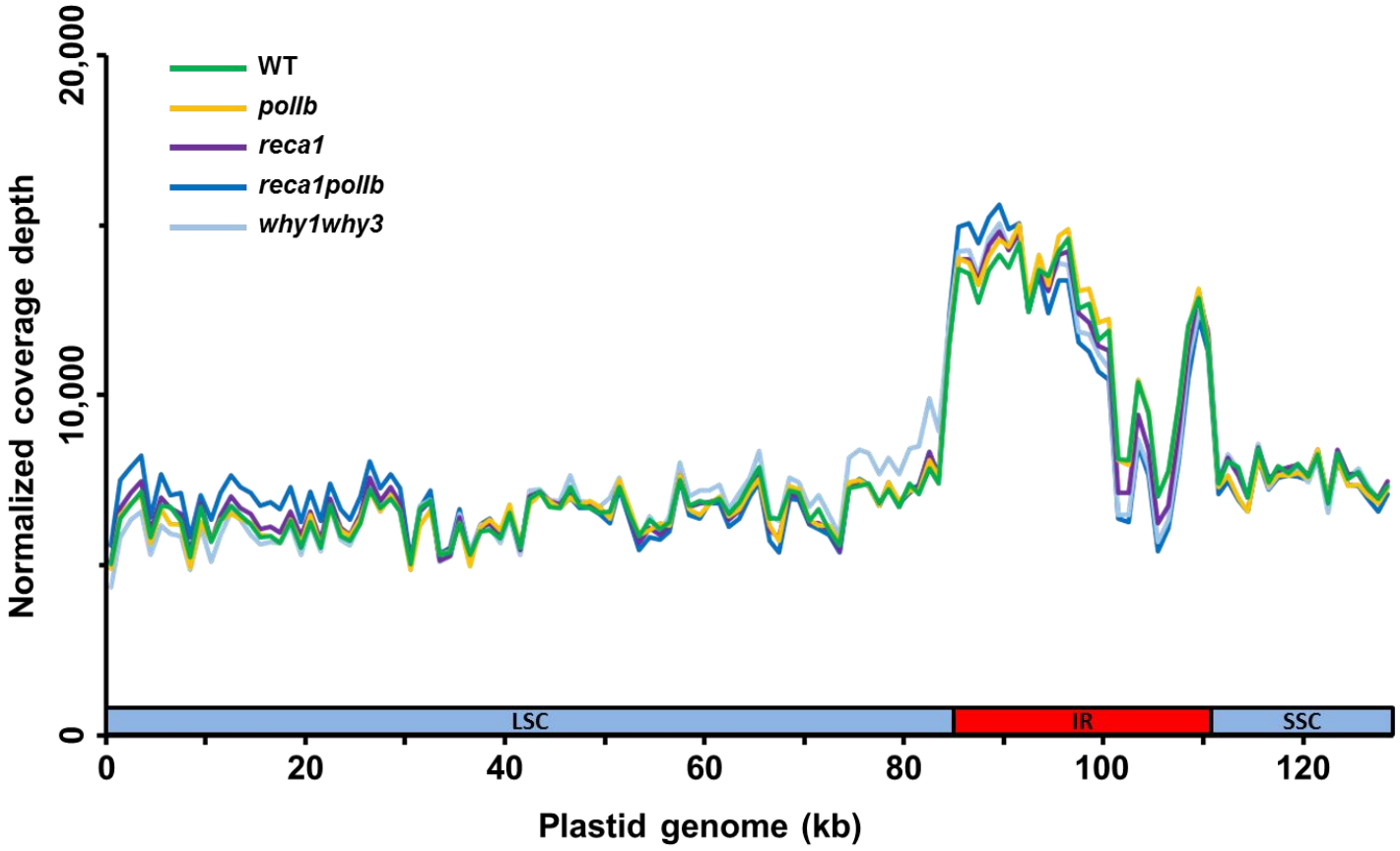
Supplemental Figure S1. Level of plastid DNA rearrangements in three *Arabidopsis* ecotypes. Total number of rearrangements identified for each of the Col-0, Ts-1 and Ws-2 ecotypes. Y axis represents the number of rearrangements per 1,000,000 plastid reads. Av.: average of the three ecotypes. The error bar represents the standard deviation.



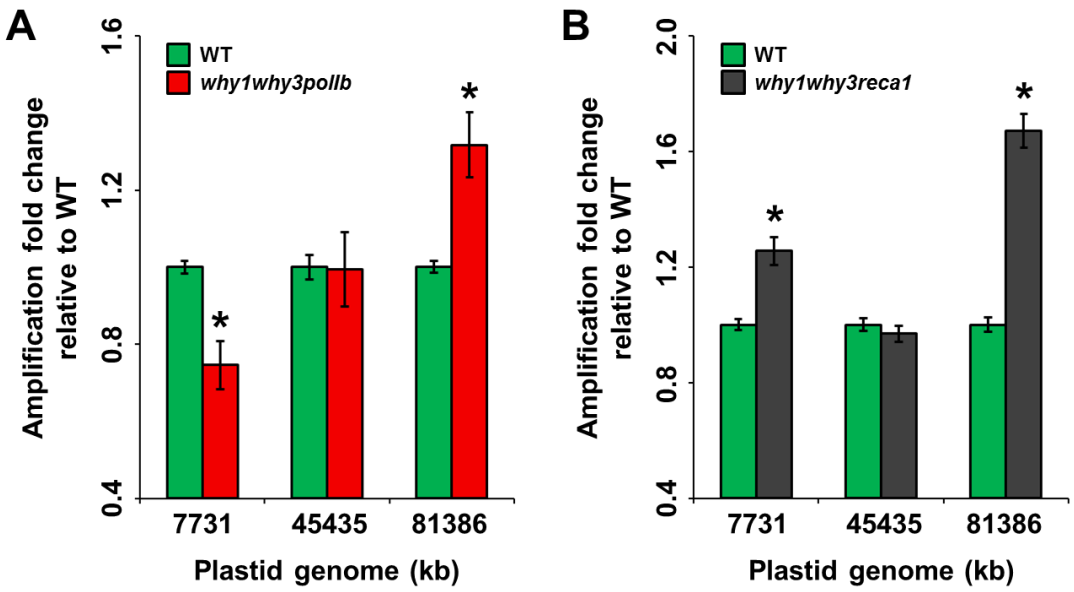
Supplemental Figure S2. Characterization of the *reca1-1* and *reca1-2* T-DNA insertion mutant lines. (*A*) Schematic representation of the *RECA1* gene and insertion positions of the *reca1-1* and *reca1-2* insertions. (*B*) Quantitative PCR measurement of the *RECA1* expression levels for the *reca1-1* and *reca1-2* mutant lines relative to WT plants. Error bars represent the standard error of the mean of three biological replicates. Asterisks indicate a significant difference of a Student's *t* test *p*-value ≤ 0.01 with the WT.



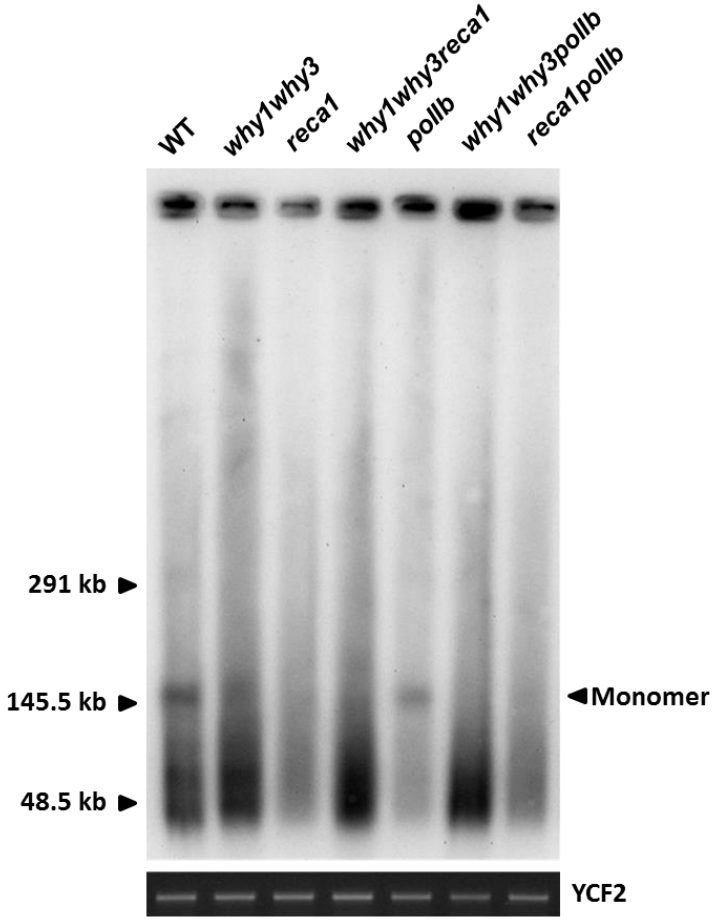
Supplemental Figure S3. Characterization of the *why1why3reca1* phenotype (*A*) Representative photographs of 21-day-old WT, *reca1-1*, *reca1-2*, *why1why3*, *why1why3reca1-1* and *why1why3reca1-2 Arabidopsis* mutant plants. (*B*) Proportion of non-germinated seeds of WT, *pollb*, *reca1-1*, *why1why3*, *why1why3pollb* and *why1why3reca1-1*, four days after vernalization. Two asterisks indicate a significant difference of a Student's *t* test *p*-value ≤ 0.01 with the WT.



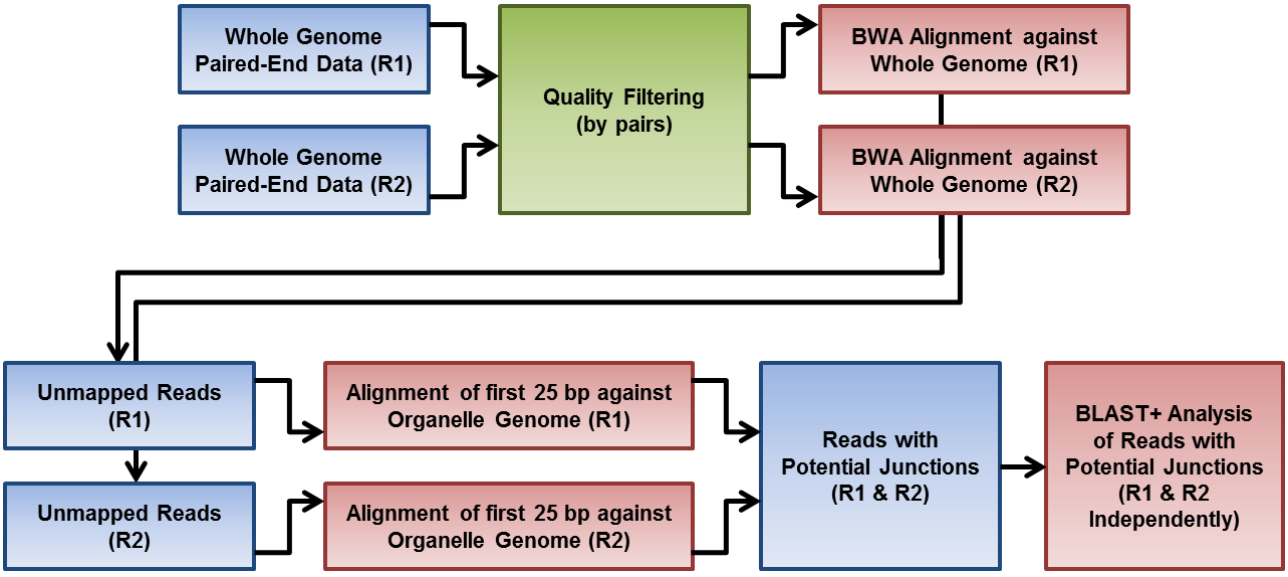
Supplemental Figure S4. Plastid DNA sequencing coverage curves for WT, *pollb*, *reca1*, *reca1pollb* and *why1why3* *Arabidopsis* mutant plants. Plastid sequencing coverage of pools of 14-day-old *Arabidopsis* seedlings of the indicated genotypes. Positions were rounded down to 1 kb. All reads mapping to the plastid large inverted repeats (IRs) were only assigned to the first IR. Y axis represents the number of reads per 1,000,000 total plastid reads. The plastid large-single copy region (LSC), the first IR, and the small-single copy region (SSC) are depicted as a long blue bar, a red bar and a short blue bar, respectively.



Supplemental Figure S5. Plastid DNA quantification at three locations of the genome in WT, *why1why3pollb* and *why1why3reca1* plants. (*A*) Quantitative PCR measurement of ptDNA levels at three sites of the LSC in *why1why3pollb* relative to WT plants. (*B*) Quantitative PCR measurement of ptDNA levels at three sites of the LSC in *why1why3reca1* relative to WT plants. Error bars represent the standard error of the mean of three biological replicates. Asterisks indicate a significant difference of a Student's *t* test *p*-value ≤ 0.05 with the WT.



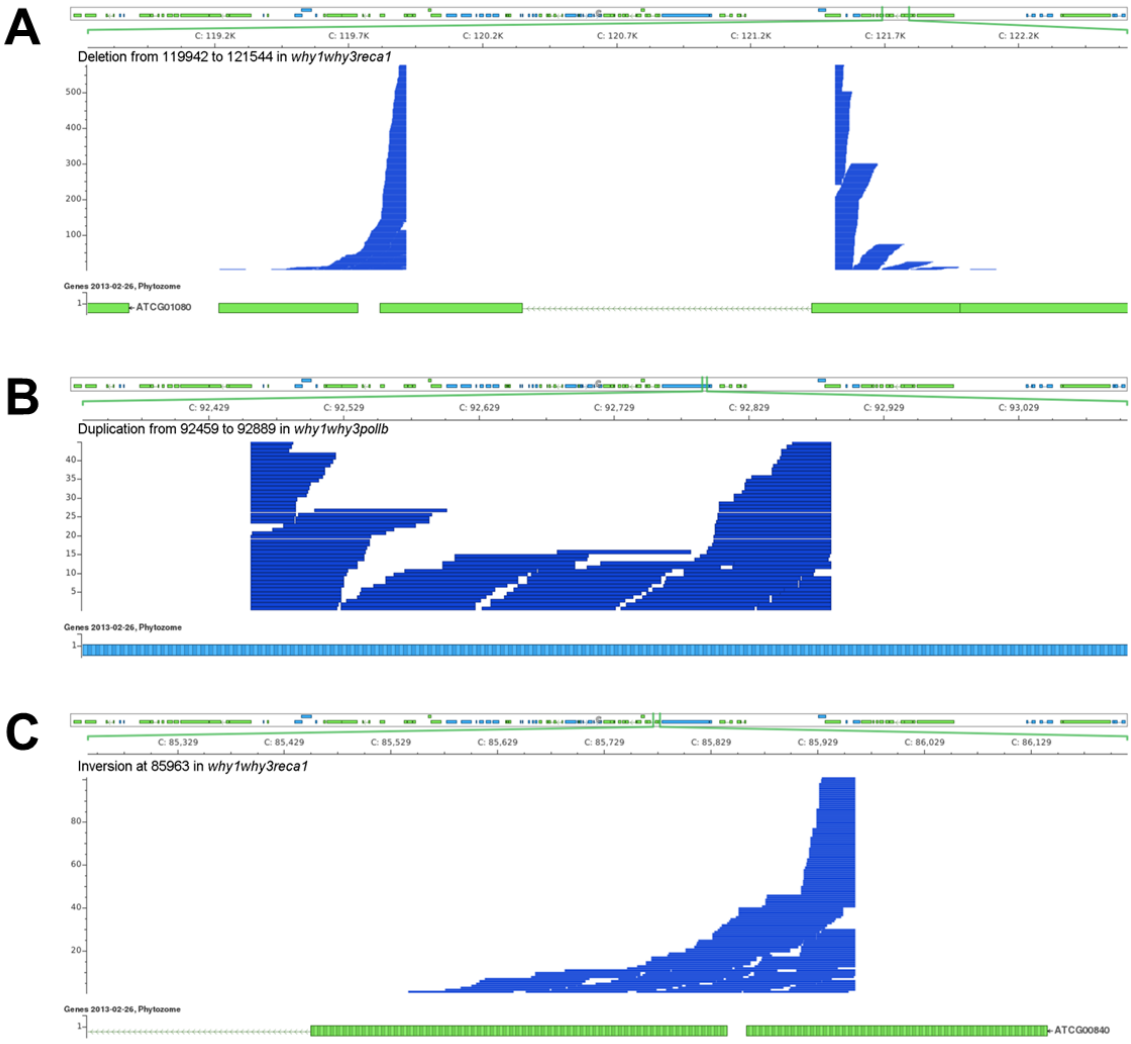
Supplemental Figure S6. Visualization of the distribution of the various forms of the plastid genome. Pulsed-Field Gel Electrophoresis (PFGE) analysis of ptDNA in the indicated genotypes. The amount of DNA loaded in each well was equilibrated relative to the amplification of a *YCF2* fragment.



Supplemental Figure S7. Schematic representation of the analysis workflow. Blue background boxes represent text manipulation steps, while the green and red backgrounds stand for quality filtering and mapping, respectively. Sequencing was performed on both ends of DNA fragments (R1 and R2). BWA: Burrows-Wheeler Aligner. R1 & R2: Paired-end sequencing read1 and read2.

	PCR and Southern Blot	Reporter Systems	Paired-End Analysis	Junction Analysis
Genome-wide	**No** *Primers and probe are specific to a single region*	**No** *Only the inserted region is observed*	**Yes**	**Yes**
Can be performed without genetic modification	**Yes**	**No** *Requires insertion of exogenous DNA into an organism's genome*	**Yes**	**Yes**
Allows comparison of samples	**Yes**	**Yes**	**No** *Detection is heavily dependent on DNA fragment length, which varies between samples*	**Yes**
Provides information about mechanism	**Yes** *Individual rearrangements need to be cloned and sequenced to obtain sequence*	**Yes** *Individual parameters can be varied to assess their importance*	**No**	**Yes** *Exact junction sequence provides some information about mechanism*
Provides a view of both short- and long-range rearrangements	**No** *Primers and probe are specific to a single region*	**No** *A single mechanism is observed at any time*	**No** *Bias toward long-range rearrangements*	**Yes**
Allows detection of short-ranged rearrangements	**Yes**	**Yes**	**No** *Detection of short-range rearrangements is limited by DNA fragment size*	**Yes**

Supplemental Figure S8. Comparison of the techniques used to detect genome rearrangements. Details are provided when relevant.



Supplemental Figure S9. Local mapping of read pairs associated to specific genome rearrangements. The top box in each panel shows a map of the complete plastid genome on which genes on the forward strand are presented in blue and those on the reverse strand in green. The position encompassed by the zoom on the genome are presented below. The mapping of each read is presented as a blue rectangle. The bottom box shows a map of genes in the zoom. A) Representative local mapping for a deletion. This deletion was observed 537 times in *why1why3reca1*. B) Representative local mapping for a duplication. This duplication was observed 77 times in *why1why3pollb*. C) Representative local mapping for an inversion. This inversion was observed 81 times in *why1why3reca1*.

	Col-0	*pollb*	*reca1*	*reca1pollb*	*why1why3*	*why1why3 pollb*	*why1why3 reca1*
Read Pairs	31241719	32810928	30409865	31132957	31438219	32941816	31699029
Pairs with average quality >20	28771745	31343328	28967978	29738528	29944699	31308538	30156212
Plastid Pairs (%)	17,69	16,77	19,51	13,35	20,35	19,16	20,58

Supplemental Table S23 Workflow Statistics for *Arabidopsis* plastid DNA rearrangements.

	Ts-1 (SRX145018)	Ws-2 (SRX145037)
Read Pairs	30012578	45521342
Pairs with average quality >20	26774195	40165795
Plastid Pairs (%)	9,89	9,54

Supplemental Table S24 Workflow Statistics for *Arabidopsis* ecotypes Ts-1 and Ws-2 plastid DNA rearrangements.

	Col-0
Read Pairs	31241719
Pairs with average quality >20	28771745
Mitochondria Pairs (%)	1,27

Supplemental Table S25 Workflow Statistics for *Arabidopsis* mitochondria DNA rearrangements.

	erx385572	erx385573	erx385574	erx385575
Read Pairs	78908272	79738231	54905866	53255234
Pairs with average quality >20	54109701	54783744	52966448	51137017
Mitochondria Pairs (%)	0,55	0,55	0,61	0,64

Supplemental Table S26 Workflow Statistics for human brain mitochondria DNA rearrangements.

	erx385576	erx385577	erx385578	erx385579
Read Pairs	74027979	75446145	47333976	46047382
Pairs with average quality >20	51040536	52250953	45679517	44520994
Mitochondria Pairs (%)	0,45	0,49	0,52	0,53

Supplemental Table S27 Workflow Statistics for human liver mitochondria DNA rearrangements.

	SRX154301	SRX154337	SRX154338	SRX154342
Read Pairs	2611112	2611112	2611112	2611112
Pairs with average quality >20	2611099	2611096	2611098	2611089
E. coli Pairs (%)	99,85	99,83	99,85	99,84

Supplemental Table S28 Workflow Statistics for *E. coli* DNA rearrangements.