

Site type	1	2	3	4	5
A. thaliana	C	C	C	C	C
A. arenosa 1	T	T	C	T	C
A. arenosa 2	T	C	C	C	C
A. arenosa 3	T	T	C	T	C
A. suecica 1	Y	Y	Y	Y	Y
A. suecica 2	Y	Y	Y	C	C
A. suecica 3	Y	Y	Y	Y	Y
A. suecica 4	Y	Y	Y	C	C
A. suecica 5	Y	Y	Y	Y	Y

Figure 1-

Diagrammatic illustration of different site types observed in the sequence dataset. Y indicates a case where both C and T nucleotides are observed in the sequence; in A. suecica this would represent a SNP difference between homeologues. Sites where all suecica accessions have Y (site types 1,2,3) would be a ‘fixed heterozygote’, representing a fixed difference between the paralogues. Cases where some A. suecica accessions have a Y while others have a C would represent a polymorphic site in one of the two paralogous loci (sites 4 and 5). Site type 1 represents a fixed difference between arenosa and thaliana, which is observed as a ‘fixed heterozygosity’ in A. suecica samples. In site type 2, the site is polymorphic in A. arenosa and a fixed paralogous difference in A. suecica. Site type 3 represents a fixed difference between paralogues in A. suecica that is not traceable to a SNP in the parental samples. Out of 243 fixed heterozygous sites, 70% were equivalent to site type 1, 25% were equivalent to site type 2, and only 5% were equivalent to site type 3. Thus, the vast majority of fixed differences between homeologues were traceable to the source parents in this dataset. As discussed in the text, only three loci showed any segregating site types resembling 4 and 5.

With site type 4, there is a shared polymorphism between A. arenosa and A. suecica. Site type 5 represents a unique polymorphism in A. suecica that is not observed in the parents. Only three loci showed evidence for site types 4 and 5; of these loci 20 out of 21 loci were of type 4 (i.e. shared with A. arenosa). This suggests that the vast majority of sequence variation in A. suecica was sampled from diversity in A. arenosa.