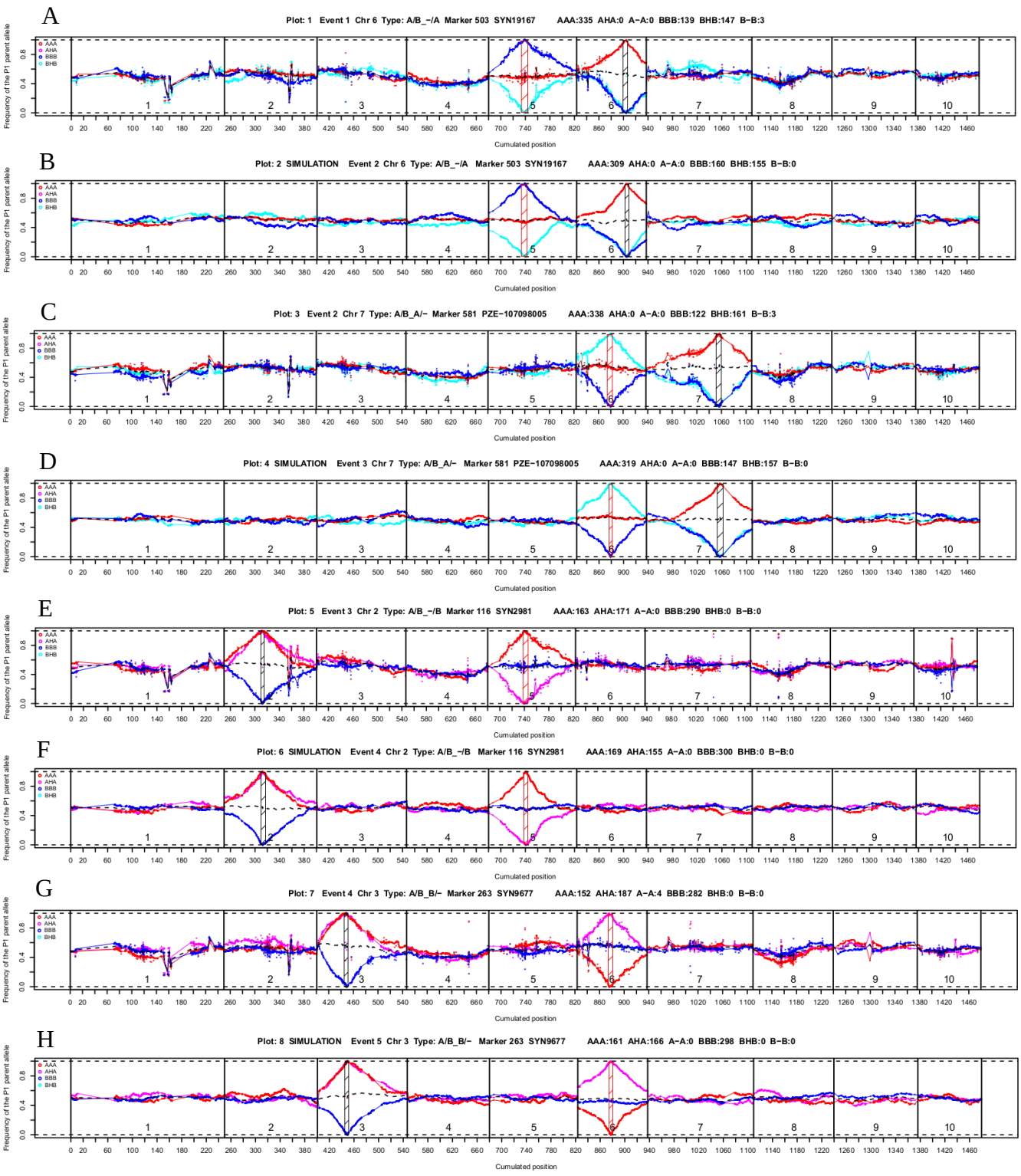

Figure S1. Examples of allele frequency profiles showing characteristic signatures of four different types of 1:2 CNVs

Data are from the Maize doubled-haploid population GABI. Panels A, C, E, G show experimental profiles for CNVs belonging to the respective four categories (type) encoded as: A/B_-/A, A/B_A/-, A/B_-/B, and A/B_B/- as indicated in the header of each panel. The encoding uses a string of 2 groups of

3 characters each, separated by an underscore. The first group is always "A/B" and indicates the reference locus, located at the position where the candidate marker was initially mapped. The second group indicates a second copy of the region targeted by the candidate marker. For both groups (loci), the letters just before and just after the slash represent respectively the haplotypes of the first parent (allele "A") and of the second parent (allele "B"). Panels B, D, F, H show simulation results reproducing the CNV situation inferred from A, C, E and G respectively (see text). The allele of parent 1 is called "A", the allele of parent 2 is called "B", heterozygotes are called "H", and missing data are called "-". Each curve shows the frequency of the allele "A" along the genome (X-axis indicates cumulated genetic positions), when considering different subsets of individuals of the population as follows: pink dots and curve for individuals (denoted "AHA") genotyped "H" at the candidate marker and "A" on both non-candidate flanking markers indicating the allelic context of the region, and similarly cyan for "BHB" individuals, red for "AAA" individuals, dark blue for "BBB" individuals. Curves generated by the software for classes based on missing data (light grey for "A-A" individuals, and black for "B-B" individuals) were hidden here for better clarity of the profiles. Hatched rectangles indicate the estimated confidence intervals on the position of the detected loci involved in the event. They are black for the reference locus (see text) and red for the secondary locus. Dots represent values of individual markers and associated curves show the result of the smoothing procedure used to detect the peaks. Lastly, the black dashed line indicates the frequency of "A" allele based on all individuals of the population. The name of the candidate (non-Mendelian) marker considered is given in the header of each panel, as well as numbers of individuals counted for each three-locus genotype class.