## Supporting File S1

## Assessing the inter-chromosomal pattern when centromere locations are unknown

The inter-chromosomal prediction for offspring heterozygosity under automixis is that all chromosomes in a given offspring should either retain 100% of parental heterozygosity or become fully homozygous at markers in the centromere regions. If centromere locations are unknown, this prediction cannot directly be assessed. However, if mapped markers are available, it is possible to test for specific “segregation patterns” by tabulating, for each marker, the individuals in which the marker becomes homozygous and in which it retains parental heterozygosity. If offspring are produced by central fusion, one would expect to find on each chromosome markers that retain parental heterozygosity in all offspring. Similarly, with pure terminal fusion, one would expect to find on each chromosome markers that become homozygous in all individuals. A sufficient number of markers is needed so that it can be assumed that each chromosome contains at least one marker that is in full linked with the centromere. If some offspring are produced by a terminal fusion and some by central fusion, one would expect to find on each chromosome markers that are heterozygous in a given set of offspring (those produced by central fusion) and homozygous in the rest (those produced by terminal fusion), with the important point being that it should be the same set of individuals that retain heterozygosity for all these markers and each chromosome should contain at least one of these markers.

We illustrate this with using a reduced set of loci with complete information (no missing genotypes) for all eight automictic offspring ($N$ = 1693 loci). With eight offspring, there are $2^8$ = 256 possible segregation patterns, each of which can be represented binary string for offspring1 to offspring8 (zero: homozygous, 1: heterozygous). For instance 00011000 is a marker, which is heterozygous in offspring4 and offspring5, and homozygous in all other offspring. We identified the segregation pattern for each of the 1693 markers and counted how many times and on how many linkage groups each specific segregation pattern occurred. Only one segregation pattern occurred on all ten linkage groups: homozygous in all individuals except individual V04_04. This pattern was shown by a total of 332 loci, with between 12 and 106 loci per linkage group. Moreover, on each linkage group, these markers were located in just one region. The ten other most common segregation patterns (Supporting Table S4) include loci that were heterozygous in all offspring (found on eight linkage groups) and loci homozygous in all offspring (found six linkage groups), but they did not occur in just a single region in these linkage groups and probably contain some error (genotyping error alignment error, etc.). The only other pattern that was observed on more than four linkage groups is a pattern that is very similar to the presumed centromeric one (Supporting Table S4), and indeed was found in many pericentromeric regions. None of the segregation patterns among the 27 self-fertilized offspring occurred on more than three linkage groups (a total of 769 markers were investigated), except for 34 loci distributed across nine linkage groups that were heterozygous in all individuals. Within each linkage group, these loci did not occur in a single region, and show strongly differing segregation patterns compared to adjacent markers, which suggests that they may be explained by alignment errors (e.g., false mapping of paralogous loci to a single position).

Overall these results show that even without information on the centromere locations it is possible to conclusively infer the mode of reproduction, given a sufficient number of mapped markers. Conversely, the results also show that mapping of centromeres can be achieved and even if some offspring are produced by terminal fusion and others by central fusion, and that the proportion of offspring produced by terminal vs. central fusion can be directly estimated from the same data.