Linkage group assignment

Based on the genetic linkage maps that were developed using restriction-site associated DNA (RAD) sequencing, we assigned 171 scaffolds with a total length of ~111.9 Mb (~32.5% of the genome sequence) into corresponding linkage groups (LGs), with 25 LGs containing scaffolds >2 Mb. The LG 1 represents the sex (Z) chromosome. There are several reasons for not mapping more scaffolds. RADseq data were relatively short by current standards (46 bp of usable data) and markers were infrequent, around one per 121 kb. The RADseq data from Baxter *et al*. were generated from strains Pearl-Sel and G88 which are different to the genome sequencing strain, Fuzhou-S. Given the high level of sequence variation observed within Fuzhou-S alone, many RADseq markers failed to map to scaffolds using our criteria. In order to assign scaffolds to chromosomes, RADseq data not only had to map to the reference genome but the insect cross also had to contain segregating polymorphism. This relatively small RADseq dataset was able to assign around ~32.5% of the chromosomal sequences, which highlights that very useful genome scaffolding data can be generated with RADseq, even when scaffolding was not the original intention.

Integration of the scaffolds into the *P. xylostella* genetic linkage group. Based on the DBM genetic map generated by RAD sequencing2, 171 scaffolds, with a total length of 111,857,878 bp (~32.5% of the genome sequence), were assigned into the corresponding linkage groups (LGs) and anchored at the genetic loci, except for LGs 2, 11 and 31. Those unassigned scaffolds occurred probably due to the genomic difference between strains for sequencing and linkage mapping. Likewise, orders and orientations of the assigned scaffold were unable to be determined, and the original orientations of them upon assembly were utilized for analysis. N.D. = no data.