Figure S9  Construction and analyses of the DA and F344 semi-finished genomes. Genomic DNA from DA and F344 male rats was sequenced using Genome Analyzer. After quality filtering, over one billion high-quality reads for each strain were aligned to the BN rat reference genome (RGSC3.4) using SOAP2.21. 91% of the reads were mapped, covering 98.9% of the BN genome. SOAP software package was used to generate consensus sequences (CNS) and to call single-nucleotide polymorphisms (SNP), short insertion-deletions (Indel), and structural variations (SV), which were further quality filtered, functionally annotated, compared to variations in dbSNP, and validated using Sanger sequencing. Gap-containing regions of the reference genome were covered using the GapCloser tool. Assembly of consensus sequences and contig end extension generated an additional 59-Mb of sequence for either strain, corresponding to 22% of the unknown sequences in the BN genome. Distribution of copy number variation (CNV) candidates and coverage of repetitive elements were also analyzed.