Novel multi-platform next generation assembly methods for mammalian genomes

James Lindsay[1], Jin Zhang[1], Ion Mandoiu[1], Yufeng Wu[1], Marilyn Renfree[2], Andrew Pask[1], Thomas Heider[1], Craig Obergfell[1], and Rachel O'Neill[1]

[1] University of Connecticut
[2] University of Melbourne

The Baylor College of Medicine, Australian Government and University of Connecticut teamed together (under the KanGO consortium) as part of an international effort to generate a de novo genome sequence for marsupial model species, the tammar wallaby (M. eugenii). This important model organism is part of a large group of mammals (Metatherian) that harbor unique life history traits and genome features. Our work involved assembling this 2.7 GB mammalian genome from sequence produced from Sanger sequencing reads (Baylor College of Medicine and RIKEN), Illumina sequencing reads (Australian Genome Research Foundation), ABI SoliD (Baylor) and 454 reads (AGRF and the Center for Applied Genetic Technologies at the University of Connecticut) using unique assembly tools we have developed.

Currently there is no single de novo assembler that is capable of handling all types of sequencing information available for the tammar wallaby. Popular assemblers such as Phrap, Newbler, and Velvet are optimized to work with one or two technologies. Additionally the amount of sequence required to perform de novo assembly also varies according to the platform. Therefore in order to integrate all the data it is necessary to conduct the assembly in a hierarchical manner, utilizing longer and higher confident reads first. Next shorter reads with higher depth are used to reduce or eliminate inconsistencies or gaps, arrange non contiguous sequences into ordered scaffolds and for validation. The initial assembly was performed using the Atlas Genome assembler. This was followed by a novel process of using mate pair information to identify short Illumina, 454 and Solid sequencing reads which can be used to further improve the assembly.