



For immediate release

October 2010

First annotation of newly sequenced rubber genome opens doors to further crop improvement

A high-quality draft sequence of the rubber tree (*Hevea brasiliensis*) genome of the RRIM 928 latex/timber clone has been completed in the UK in just nine months. Early studies of leaf and latex transcriptomes indicate that processed gene coding sequences comprise 2 – 4% of the total genome size, a figure consistent with other higher organisms. First annotation of the sequence assembly has found a high similarity of the functional regions of the genome of *Hevea* and other species from the Euphorbiaceae family, which includes cassava and castor bean. Comparison between public domain *Hevea* protein-coding sequences and the *de novo* assembled sequence identified 99.98% of the public domain sequences within the newly assembled genome. Ongoing transcriptome analysis is expected to confirm the completeness of the functional region coverage.

The sequencing, assembly and annotation of the genome marks a significant cornerstone in accelerating the commercial development of this important commodity and in addressing many of the issues facing rubber growers and manufacturers worldwide. The work has been carried out by the Malaysian Rubber Board's research centre in the United Kingdom, the Tun Abdul Razak Research Centre (TARRC), and The Genome Analysis Centre (TGAC), Norwich Research Park, UK.

Hevea brasiliensis is a tree crop cultivated as the primary source of natural rubber, a unique biopolymer that is used widely in industry, transportation, medicine and personal care. Whilst the growing need for natural rubber worldwide secures the future demand for this crop, there are a number of biological and environmental threats that could put its cultivation at risk, such as the effects of climate change and diseases such as South American leaf blight and white root rot. The outcome of this work will be a knowledge-based resource essential for addressing these and other issues affecting the rubber industry now and in the future.

Traditional rubber tree improvement is a lengthy and difficult process. The availability of the annotated *Hevea* genome sequence will enable the development of a large database of molecular markers that can be used to identify genes involved in important traits, providing enormous scope for the improvement of natural rubber properties and of overall crop performance. Novel markers will help make breeding schemes more efficient. For example, traits that may only show up in mature plants will be able to be identified in seedlings with the use of a low cost DNA test. As well as breeding for increased quality, yield and disease resistance, there are many other commercial applications of these breeding tools for latex and timber production.

The TGAC team, specialists in DNA sequencing and bioinformatics, used two next-generation sequencing (NGS) platforms to generate the draft sequence, combining the massive throughput of the Illumina platform with the longer read length of the Roche 454 platform in a dual approach that enabled fast and accurate assembly of an approximately 80-fold sequence coverage of the genome.

The *Hevea* genome is estimated to contain approximately 2 billion bases, making it one of the largest among cultivated crop species and perhaps the largest plant genome sequenced to date. In comparison with cultivated crops, the *Hevea* genome is around 2 ½ times the size of the apple and cassava genomes; 4 to 5 times larger than the genomes of the black cottonwood, cacao, rice and castor bean; and 17 times larger than the *Arabidopsis* genome, which was the first plant genome to be sequenced.

We calculate that around 20% of the genome contains functional genes and, at this early stage of analysis, we predict that the *Hevea* genome may contain in the region of 43,000 gene models, based on initial comparisons of the transcriptome with the genome. This is consistent with predicted gene model counts for black cottonwood, cassava and apple of 45,000, 47,000 and 57,000 respectively. Given that perhaps 80% of the genome may be non-coding, assembly has been particularly challenging.

Work on refining the genome's draft and its annotation with focus on genes linked to commercially important traits is continuing. Whilst the bulk of the sequence generated to date has been through the NGS approach, the next stages will involve targeted sequence improvement through transcriptomics and conventional Sanger sequencing to complete the assembly and further improve the quality of the sequence generated to date using TARRC's newly acquired bioinformatics capability.

For further information please contact: Dr Maria Kolesnikova-Allen, Head, Biotechnology Unit, TARRC, e-mail: mkallen@tarrc.co.uk

For further information about the Genome Analysis Centre please contact Rachael Spooner, email: Rachael.spooner@genome-enterprise.com

Notes for editors

Background information on TARRC

TARRC is the UK-based laboratory of the Malaysian Rubber Board (MRB). It was set up in 1938 and has been at the forefront of key advances made in the fields of polymer science, engineering, testing, analysis and product development. The research centre has contributed to the development and modernisation of the Malaysian rubber industry and has developed a worldwide reputation in the rubber industry as a centre of excellence.

TARRC's Biotechnology Unit was set up in 2008 as part of the MRB's efforts to improve the use of natural rubber as an agricultural and pharmaceutical resource.

TARRC's consultancy arm, Rubber Consultants, was established over twenty-five years ago to allow individual clients to have access to the expertise of TARRC's scientists and extensive laboratory facilities on a confidential basis.

For more information on TARRC visit www.tarrc.co.uk

For more information on Rubber Consultants visit www.rubberconsultants.com

Background information on TGAC

The Genome Analysis Centre (TGAC) is a national genomics and bioinformatics centre, specialising in sequencing and in bioinformatics for analysis and interpretation of genomic data. The centre is located on the Norwich Research Park. It was established by the Biotechnology and Biological Sciences Research Council in partnership with the East of England Development Agency, Norfolk County Council, South Norfolk Council, Norwich City Council and the Greater Norwich Development Partnership.

For more information on TGAC visit www.tgac.bbsrc.ac.uk