

|                                  |                                                              |
|----------------------------------|--------------------------------------------------------------|
| <b>Project title:</b>            | Generation of genomic resources for apple rootstocks project |
| <b>Project number:</b>           | TF 212                                                       |
| <b>Project leader:</b>           | Dr Nicola Harrison, East Malling Research                    |
| <b>Report:</b>                   | Final Report, April 2014                                     |
| <b>Previous report:</b>          | None                                                         |
| <b>Key staff:</b>                | Dr Nicola Harrison                                           |
| <b>Location of project:</b>      | East Malling Research                                        |
| <b>Industry Representatives:</b> | Dr Jim Quinlan                                               |
| <b>Date project commenced:</b>   | 1 February 2013                                              |
| <b>Date project completed:</b>   | 30 April 2014                                                |

## **DISCLAIMER**

*AHDB, operating through its HDC division seeks to ensure that the information contained within this document is accurate at the time of printing. No warranty is given in respect thereof and, to the maximum extent permitted by law the Agriculture and Horticulture Development Board accepts no liability for loss, damage or injury howsoever caused (including that caused by negligence) or suffered directly or indirectly in relation to information and opinions contained in or omitted from this document.*

*Copyright, Agriculture and Horticulture Development Board 2014. All rights reserved.*

*No part of this publication may be reproduced in any material form (including by photocopy or storage in any medium by electronic means) or any copy or adaptation stored, published or distributed (by physical, electronic or other means) without the prior permission in writing of the Agriculture and Horticulture Development Board, other than by reproduction in an unmodified form for the sole purpose of use as an information resource when the Agriculture and Horticulture Development Board or HDC is clearly acknowledged as the source, or in accordance with the provisions of the Copyright, Designs and Patents Act 1988. All rights reserved.*

*AHDB (logo) is a registered trademark of the Agriculture and Horticulture Development Board.*

*HDC is a registered trademark of the Agriculture and Horticulture Development Board, for use by its HDC division.*

*All other trademarks, logos and brand names contained in this publication are the trademarks of their respective holders. No rights are granted without the prior written permission of the relevant owners.*

*The results and conclusions in this report are based on an investigation conducted over a one-year period. The conditions under which the experiments were carried out and the results have been reported in detail and with accuracy. However, because of the biological nature of the work it must be borne in mind that different circumstances and conditions could produce different results. Therefore, care must be taken with interpretation of the results, especially if they are used as the basis for commercial product recommendations.*

## ***AUTHENTICATION***

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

Dr Nicola Harrison  
Research scientist and project leader  
East Malling Research

Signature:  Date: 28/03/2014

### **Report authorised by:**

Dr Mark A. Else  
RECP Programme Leader  
East Malling Research

Signature:  Date: 4 April 2014

## **CONTENTS**

|                                                   |   |
|---------------------------------------------------|---|
| Grower Summary.....                               | 1 |
| Headline.....                                     | 1 |
| Background and expected deliverables .....        | 1 |
| Summary of the project and main conclusions ..... | 1 |
| Financial benefits.....                           | 2 |
| Action points for growers .....                   | 2 |
| <br>                                              |   |
| Science Section.....                              | 3 |
| Introduction .....                                | 3 |
| Materials and methods .....                       | 3 |
| Results.....                                      | 4 |
| Discussion .....                                  | 6 |
| Conclusions .....                                 | 6 |
| References .....                                  | 7 |

## GROWER SUMMARY

### Headline

- The genomic sequence data generated for apple rootstock accessions will improve the scientific understanding of rootstocks and aid marker-assisted breeding for important rootstock traits in the East Malling Rootstock Club

### Background and expected deliverables

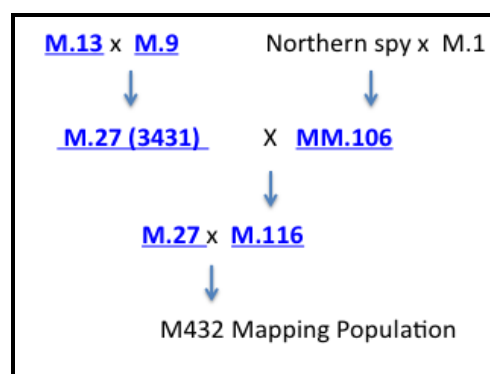
The advent of new technology in the form of next generation sequencing is now changing many aspects of research in horticulture. Of most significance is the increasing ability to transfer knowledge from model organisms, such as the brassica *Arabidopsis*, into non-model crops and growing systems. Such developments include the ability to identify and develop new crop varieties with increased resistance to biotic and abiotic stresses along with an understanding of how these processes work in plant systems.

A major cross-cutting theme in EMR's science strategy is to develop an understanding of how rootstocks control scion architecture and the generation of genomic resources for rootstocks is pivotal to this research. These resources will provide an essential platform that will feed into existing and new funding proposals. Additionally, generation of the sequence data will not only raise the profile of both the HDC and EMR nationally, but will also support the development of collaborative links with scientific and horticultural communities overseas.

The generation of genetic resources for apple rootstocks will enable a more informed approach to rootstock breeding through the development and deployment of marker assisted breeding technologies.

### Summary of the project and main conclusions

Apple accessions, which have been sequenced, include germplasm that is currently used as parental material in the East Malling Rootstock Club (EMRC) breeding program, namely 'M.27', 'M.116' and 'M.M.106' (Figure 1). These rootstocks exhibit a number of interesting traits such as dwarfing, precocity and disease resistance, which are highly valued by the industry. The selected rootstocks



**Figure 1.** Apple rootstocks selected for genome sequencing (blue) and their relevance to the M432 mapping population. NB: 'M.9' sequencing already completed prior to TF 214.

have also been incorporated into a unique seedling progeny known as the mapping population 'M432', which is the basis of a number of research projects that are developing molecular markers for important rootstock traits. In order to gain a better scientific understanding of how these traits are bestowed, genomic sequence data is needed.

### **Financial benefits**

The data generated in this study will inform on-going HDC-funded work by the EMRC by providing essential data for the development of new marker-assisted breeding strategies in apple rootstocks. The potential to reduce the time from initial seedling screening to final selection of new commercial rootstocks with increased tolerance to biotic and abiotic stresses would improve greatly the economic and environmental sustainability of the UK tree fruit industry.

### **Action points for growers**

There are no specific action points from growers arising from genetic research of this nature. In future, the information developed from this project will aid fruit breeders in developing improved genetic material for apple rootstocks.

## **SCIENCE SECTION**

### **Introduction**

The advent of new technology in the form of next generation sequencing is now changing many aspects of research in horticulture. Of most significance is the increasing ability to transfer knowledge from model organisms, such as the brassica *Arabidopsis*, into non-model crops and growing systems. Such developments include the ability to identify and develop new crop varieties with increased resistance to biotic and abiotic stresses along with an understanding of how these processes work in plant systems.

One of the major themes of the new EMR science strategy will be to focus on developing an understanding of how rootstocks control scion architecture and the generation of genomic resources for rootstocks will be pivotal in the success of this research. These resources will provide an essential platform that will feed into new and existing research projects. Additionally, generation of genomic resources will not only raise the profile of both the HDC and EMR nationally, but will also promote collaborative links with scientific and horticultural communities on an international basis.

This HDC-funded project has focussed on sequencing the genomes of several apple rootstocks, namely M27, MM106, M116 and M13. The rootstocks were chosen owing to their importance both commercially and scientifically. A Defra-funded project (WU0115) is currently developing genomic tools for rootstock breeding using a unique mapping population (M432) that segregates for vigour control (dwarfing) and other agronomically important traits. The sequencing of these rootstock accessions will enhance and contribute to the understanding of rootstock science, as well as advance molecular marker development, which will feed into the EMRC breeding program.

### **Materials and methods**

#### ***DNA extraction***

DNA extraction was performed for four rootstock accessions: M27, MM106, M116 and M13. Plant material was snap frozen in liquid nitrogen and high quality DNA was extracted using the commercial DNeasy plant tissue kit (Qiagen). The DNA was sent to The Genome Analysis Centre (TGAC) for high-throughput genome sequencing, where owing to the predicted large size of the apple genome (750 Mbp), it was sequenced in a cost-effective manner. The high-throughput sequencing machine used at TGAC (HiSeq 2000) is able to sequence more DNA to sufficient depth than the previous sequencing platforms and at a cheaper cost per nucleotide.

## Genome assembly

Rootstock genome datasets are now being mapped to the published Golden Delicious genome reference using the software Bowtie2 through a process known as reference guided assembly (RGA). RGA aligns each individual read to the reference sequence.

## Results

### Genome data

Genome data was generated for apple rootstocks at TGAC. The sequencing data was produced on a HiSeq 2000 sequencing platform with the following specifications: 100 bp paired-end reads with an average insert size of 548 base pairs. Table 1 summarises the relevant information and statistics for the generated genome data.

Table 1: Summary statistics for apple rootstock genome data. This table summarises the amount of genomic data generated for each rootstock, the number of pair-end sequencing reads and the distance between the matching paired reads (Insert size).

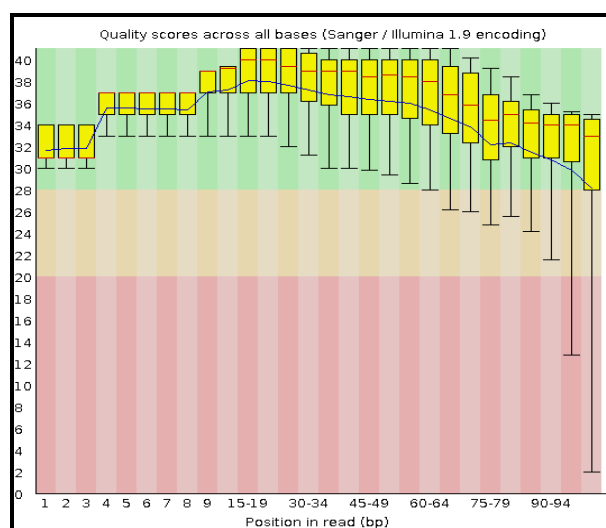
| Rootstock Sample | Insert Size (base pairs) | Paired Reads obtained | Size of Compressed Data/Gigabytes |
|------------------|--------------------------|-----------------------|-----------------------------------|
| M.116            | 553                      | 197,717,206           | 38.6                              |
| M.13             | 543                      | 199,735,550           | 39                                |
| M.27             | 545                      | 220,682,335           | 42.9                              |
| M.M.106          | 551                      | 204,588,960           | 40.2                              |
|                  | <b>Total</b>             | <b>822,724,051</b>    | <b>160.7</b>                      |

### Quality Assessment

All genome datasets were processed for read quality and sequencing cycle quality (Figure 2); where reads were below the quality threshold, they were removed from the datasets.

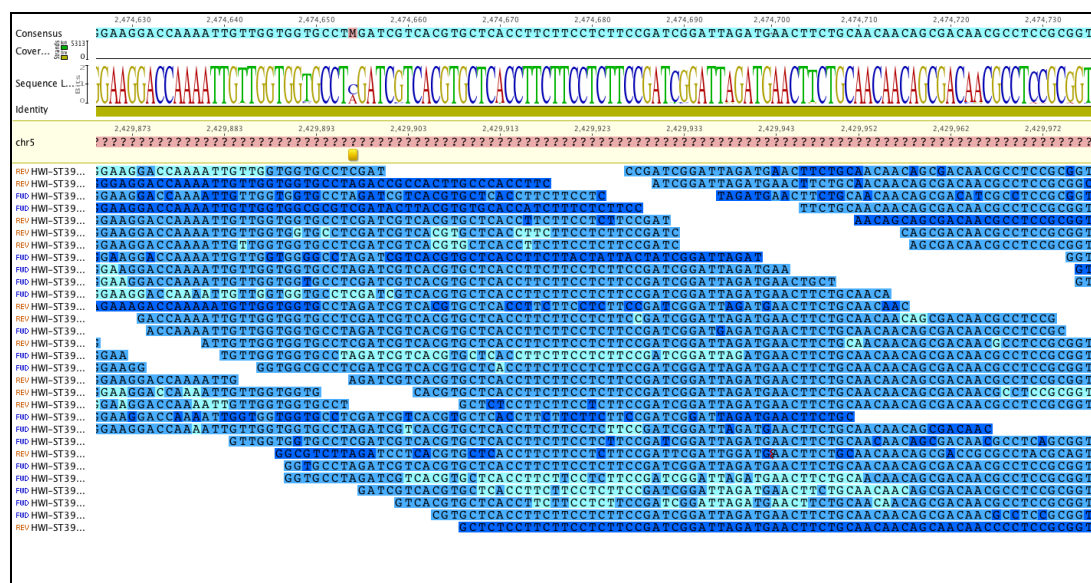
### Reference guided assembly

The genome sequence data was mapped onto the published genome of the apple cultivar Golden Delicious (Velasco *et al*, 2010) using the open source software Bowtie2 (Langmead *et al*. 2009) and visualised in the commercial software



**Figure 2.** An example of sequencing read quality assessment. The quality of each DNA nucleotide from every sequencing read is plotted on a box-plot graph and sequences below a quality threshold are removed.

Geneious (Figure 3). Further comparative analyses of the rootstock genomes is now being undertaken to provide single nucleotide polymorphism (SNP) data that can be used for the generation of molecular markers in rootstock breeding material and progenies.



**Figure 3.** A section of the genome alignment from the M27 reference guided assembly. The figure illustrates how the short DNA sequences are aligned to the reference sequence to build up a draft genome sequence for M27.

## Milestones

**Milestone 1:** DNA extraction from M27, M116 and MM106. - Target Date: 30/02/13

This milestone has been met.

**Milestone 2:** Genome sequencing – Target Date: 30/06/13

This milestone has been met.

**Milestone 3:** First draft genome assemblies and data mining for candidate dwarfing genes - Target Date: 30/01/14

Work towards this milestone is on-going and is expected to be completed by the end of April 2014; we are currently performing reference-guided assemblies on all apple rootstock genome datasets.

## Dissemination of results

The data will be made available to the research community and all datasets will be uploaded to the short read archive in Genbank after publication of results.

## Discussion

This proposal was relevant to HDC TF strategies, namely:

**Objective 3, Target 1b:** The development of these genetic resources will contribute to an understanding of the communication processes between roots and canopy.

**Objective 3, Target 2a:** These resources will feed into and enhance the current breeding, evaluation and development of apple rootstocks.

**Objective 2, Target 1f:** This work will also contribute to new research under development at EMR aimed at understanding the biological processes and interactions involved in replant disease, a key priority in the 2012-13 strategy document for the Tree Fruit panel.

The data generated here has fulfilled the objectives for this project.

The generation of these draft rootstock genomes will provide the basis for further studies, in particular to investigate important rootstock traits such as conferred vigour to the scion. Gene expression is the basic mechanism through which a gene can give rise to a phenotype such as dwarfing. Gene expression analysis can be used in conjunction with the genome data to identify genes that are differentially expressed in different rootstocks and determine any DNA changes between these genes. This leads to a greater understanding of the processes and biochemical pathways that are involved in the manifestation of rootstock traits and will also aid marker-assisted breeding.

## Conclusions

We have successfully assembled draft genome sequences for commercially important apple rootstocks including M9 and M27. Additional genome assemblies for M116, MM106 and M13 are currently underway. The generation of these resources is now providing a framework on which to base further studies, and TF 212 has already added value to Defra project (WU0115), which is developing genomic tools for rootstock breeding including a dense linkage map for apple rootstock breeding and the identification of important markers. These developments will shortly feed into the EMRC, of which HDC is an industrial partner.

### ***Future work at EMR***

Owing to the pedigree of the selection of rootstock accessions that were sequenced (Figure 1), it is now feasible to assemble and more importantly, define the genomic haplotype data for the assembled genomes. This is crucial for the identification and characterisation of key genomic regions such as those involved in vigour control (dwarfing) and the development of molecular markers that can be used to inform rootstock breeding programmes.

Additional work to provide the genome data in an accessible format to the research community is also underway.

### **References**

Geneious version 6.1.2 created by Biomatters. Available from <http://www.geneious.com/>

Langmead B, Trapnell C, Pop M, Salzberg SL. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10:R25.

Velasco et al. (2010). The genome of the domesticated apple (*Malus [times] domestica* Borkh.). *Nature Genetics* 42, 833–839