

**Project number:** 5762  
**Funding source:** DAFF-RSF 07-566

**Date:** Sept 2012  
**Project dates:** Dec 2007 – Dec 2011

## Developing biotechnology-based resources for white clover



### Key external stakeholders:

Plant breeders, forage genetics community

### Practical implications for stakeholders:

Biotechnology-based breeding approaches are increasingly becoming routinely deployed in many of the most significant crop plant species. White clover has complex genetics and poor genome-based resources that currently limit application of such approaches. The main goal of this project was to increase the DNA sequence based resources for the species, and explore advanced genetical approaches for dissecting key traits for the improvement of the species. The results contribute to an increased capacity within the international forage research community to develop biotechnology-based breeding approaches in white clover.

### Main results:

A reference transcript set of approximately 70,000 expressed sequence tags was developed for white clover using 454-sequencing.

Over 200,000 single nucleotide polymorphisms, which can be used as the basis for molecular marker development, were identified in white clover.

Association mapping using a candidate gene approach and half-sib families was validated in the species, and an association between leaf size and an orthologue of the PHANTASTICA gene was identified.

### Opportunity / Benefit:

The research provides a rich DNA sequence based resource for further exploitation in the path towards the deployment of biotechnology-based breeding approaches in white clover. Sequence data will be placed in the public domain to maximize its utility to the scientific community.

### Collaborating Institutions:

Institute of Biological, Environmental and Rural Sciences, UK.  
International Trifolium Network (ITN)

**Teagasc project team:** Dr. Dan Milbourne (Co-PI)  
Dr. Susanne Barth (Co-PI)  
Jeanne Mehenni-Ciz  
Marialaura Destefanis  
Angela Ruggerio

**External collaborators:**

---

### 1. Project background:

Because of its nitrogen fixing capacity, white clover has an important potential role in both animal nutrition and fertilizer input reduction in grassland based production systems. White clover shows the potential to maintain high levels of productivity under low fertilizer N input conditions while reducing animal-based methane emissions and sequestering atmospheric carbon. Thus, white clover is a key component of the drive towards “sustainable intensification” of pasture-based production systems central to the goals of Food Harvest 2020. However, the ability to produce improved varieties of white clover suitable for a range of production systems and environments is central to the long term success of this role. Teagasc is actively involved in the breeding of improved white clover varieties specifically for Irish pastoral production systems.

Biotechnology-based breeding approaches are increasingly becoming routinely deployed in many of the most significant crop plant species. The successful application of these technologies in the breeding of a species is based on a combination of favorable reproductive genetics (inbreeding diploid species are the most optimal systems for deployment of the technology), well developed genome-based resources, and in some cases, an ability to rapidly phenotype very large numbers of genotypes of the target plant for complex traits using simple, single step approaches. White clover is an outbreeding allotetraploid with a breeding system that does not favour the easy application of biotechnology-based breeding approaches. In addition, genome-based resources for the species are very poor.

---

### 2. Questions addressed by the project:

The rationale behind this project was to fill some of the aforementioned knowledge and resource gaps. Little can be done regarding the unfavourable genetics of white clover. Instead this project focused on two main areas. Firstly, we aimed to develop a significant sequence based resource for the species using novel next-generation sequencing approaches. Secondly, we aimed to test the potential in white clover for an approach called “association mapping”, which identifies associations between genes/markers and important phenotypic characteristics – an important step in dissecting the genetics of such traits.

---

### 3. The experimental studies:

In the first part of the study, we addressed the extreme paucity of genome-based resources for white clover through the development of a large scale SNP resource for the species. Radical improvement in next generation sequencing technology experienced over the recent years allowed us to generate several hundred thousand sequences from cDNA of a single white clover inbred line using 454-sequencing. These were assembled into a “reference transcript set”, to which tens of millions of shorter reads generated from multiple further genotypes using the Illumina sequencing platform, were aligned using a bioinformatics pipeline based on programmes designed to handle nucleotide alignments from next generation sequencing experiments. This process allowed nucleotide variation over the white clover genome to be identified, as described in more detail below.

Subsequently, we tested the potential of an approach called “association mapping” for the genetic dissection of traits of interest to breeders in white clover. Association mapping is an approach originally developed in human genetics whereby large populations of unrelated or partially related individuals are characterised for specific traits (phenotyped), their DNA extracted, and the genome of each individual subsequently surveyed for DNA polymorphisms showing a statistically significant association with the traits that were originally surveyed. Occurrences of such associations are indicative of genes (or alleles thereof) that are important in the control of the trait in question, and this is the first step in being able to use molecular marker-based approaches to select for these loci in breeding programmes. This project represented the first attempt to perform this type of analysis in white clover. In addition, the population type used (a half-sib family-based

population) was specifically designed to enable the transfer of the population to multiple research groups worldwide. During the course of the project the traits leaf area (length and width) and digestibility traits; neutral and acid detergent fibre (NDF and ADF) content were measured in the population, the latter using near infrared spectroscopy. A candidate gene approach was used for association analysis whereby nucleotide polymorphisms in genes involved in pathways related to the underlying traits were tested for association with the traits measured in the population. The sequence based resource generated in the first part of the project acted as the source of the candidate genes in question.

#### 4. Main results:

We used the Roche 454-sequencing approach to generate 750,000 EST sequences of a white clover inbred line, assembled these sequences into approximately 70,000 “reference transcript sequences”, and partially annotated these sequences using a BLAST-based similarity approach. To further exploit this resource against a backdrop of sequencing technology improvements, we designed an additional experiment to generate a further 100 million sequences in 2 additional white clover inbred lines and 25 million sequences in a mixed sample of 24 divergent white clover genotypes using the Illumina HiSeq2000 and GAII platforms. Using a strategy originally developed for allotetraploid *Brassica* species as a template, we developed a bioinformatics pipeline for this dataset that allowed the identification of over 200,000 polymorphic single nucleotide polymorphism (SNP) sites in the genome of white clover. Significantly, this cutting-edge approach allowed us to distinguish allelic and homeolog-specific variation in this allotetraploid species, a major limitation for the deployment of SNPs for marker development in white clover. This resulted in a highly useful database of reference sequences and associated SNPs to use as a basis for marker development for the global white clover genetics and breeding community.

An association mapping population of 384 half-sib clover families was planted in a field trial at Oak Park. Plant material was harvested from all plots on four occasions (three in 2010 and one in 2011) for NDF and ADF analysis. ADF and NDF measurements were carried out on three of the four harvests using NIRS, subsequent to chemistry-based calibration. In addition leaf size was measured and recorded for the entire population in both years. Candidate genes for NDF and ADF (from the phenylpropanoid pathway) and leaf size were identified from the literature, and the sequences of these genes in white clover were established by comparing published sequences in other species to the reference transcript assembly of 70,000 sequences developed in the previous part of the study. In total, 4 candidate genes were assayed on the entire population: three from the phenylpropanoid pathway (two representatives of the HCT multigene family, and one from the 4CL family) and a white clover ortholog of the meristem patterning gene PHANTASTICA/AS1, which has been shown to be involved in leaf size in other species as diverse as poplar and maize. Analysis of the phenylpropanoid (PP) pathway gene sequences has shown no strong associations with variation for ADF or NDF in the population. Analysis of the PHANTASTICA ortholog revealed an association of a SNP haplotype with variation in leaf size over two years. Although the lack of association for variation in PP gene sequences is disappointing, the association for leaf size is, to the best of our knowledge, the first demonstration of successful association mapping in white clover to date.

#### 5. Opportunity/Benefit:

The long term impact of the research will enable the deployment of biotechnology-based breeding approaches in white clover.

#### 6. Dissemination:

The technical nature of the project meant that results were disseminated entirely through the scientific literature and appropriate scientific meetings

#### Main publications:

Peer-reviewed publications:

1. Magee AM, Aspinall S, Rice DW, Cusack BP, Sémon M, Perry AS, Stefanović S, Milbourne D, Barth S, Palmer JD, Gray JC, Kavanagh TA, Wolfe KH. (2010)  
**Localized hypermutation and associated gene losses in legume chloroplast genomes**  
Genome Research (12):1700-10.
2. Nagy I, Barth S, Mehenni-Ciz J, Abberton MT, Milbourne D.  
**A hybrid next generation transcript sequencing-based approach to identify allelic and**

---

***homeolog specific single nucleotide polymorphisms in white clover***

Submitted to BMC Genomics

Scientific abstracts or articles including those presented at conferences:

***Hybrid NGS-based transcript sequencing to identify inter- and intra-homeolog single nucleotide polymorphisms in allotetraploid white clover.***

Presented by Dan Milbourne at the 29<sup>th</sup> Triennial Meeting of the EUCARPIA Fodder Crops and Amenity Grasses Section Meeting, Dublin Castle, Ireland , 4<sup>th</sup>-8<sup>th</sup> Sept 2011

***Generation of SNP markers from next generation sequencing technologies for the forage species white clover (*Trifolium repens* L.) towards association mapping***

Poster Presentation by Jeanne Mehenni-Ciz at the Irish Plant Scientist Association Meeting (IPSAM), May 2010.

---

**7. Compiled by:** Dan Milbourne, Susanne Barth.

---