Molecular approaches in the analysis of red clover rhizobium symbiosis

Dinkins, R.D.*; Koch, L. †; Sullivan M.L.†; Zhu, H. †
* USDA-ARS, Forage-Animal Production Research Unit, Lexington KY, USA; † USDA-ARS, US Dairy Forage Research Center, Madison WI, USA; USDA-ARS, US Dairy Forage Research Center, Madison WI, USA; Department of Plant and Soil Sciences, University of Kentucky, Lexington KY, USA

Key words: Trifolium, genome assembly, gene expression, nodulation, nitrogen fixation genes

Abstract
Red clover (*Trifolium pratense* L.) is an important forage and pasture legume grown throughout temperate regions. Because it can form a symbiotic relationship with rhizobium, it is able to fix atmospheric nitrogen. Due to the high cost of nitrogen fertilizers, pasture legumes have been increasingly important in forage production settings. Red clover has not been a model legume primarily due to self-incompatibility and the associated high level of genomic heterozygosity, therefore it has not been a significant contributor in molecular or genetic studies and basic information on red clover legume/rhizobium symbiosis is lacking. Using recently annotated genomic resources, RNA-seq expression analysis and CRISPR/Cas9 mutagenesis, we characterized a number of genes that are expressed only in nodule forming roots. These include genes that encode proteins with homology to nodule-specific cysteine rich proteins (NCRs) and nodule-specific polycystin-1, lipoxygenase, alpha toxic (PLAT) domain proteins (NPDs) that are postulated to be involved in plant rhizobium interactions. Our results indicate that red clover has one of the highest numbers of expressed NCR and ATS3-like/NPD peptides currently known in the inverted-repeat lacking clade (IRLC) of legumes. Knowledge of the expression of these genes and the continued analysis of the genetic variation in red clover should aid in breeding genotypes with increased rhizobium selection specificity and increased nitrogen fixation efficiency.

Introduction
Red clover is an important forage legume crop in temperate regions of the world and a key component of sustainable intensification pasture-based production systems. Such legumes offer an opportunity to reduce nitrogen fertilizer use when interseeded with forage grasses. Red clover's beneficial attributes of high protein forage and reduced need for nitrogen fertilizer input can contribute to reducing the environmental footprint of pasture-based agriculture. For instance, an interseeded grass-clover pasture can contribute more than 200 kg nitrogen per ha/year to the cropping system (Carlsson and Huss-Danell 2003). Red clover is an ideal legume pasture choice for many areas of the world because it tolerates soil acidity and poor drainage, however its limited persistency provides good biomass yield for only two or three seasons (Ball et al. 2007; Taylor 2008). Better understanding of the genetic basis of traits affecting persistence, forage yield and quality, livestock nutrition, and interactions with soil rhizobium is needed to facilitate genetic improvement.

Molecular work on red clover has lagged behind many other cultivated crop species despite its relatively small genome (Vizintin L. 2006). The primary reason for this is that it is an outcrossing species, making genetic gains more difficult (Riday and Krohn 2010). The first available genomic resources were developed by Isobe et al. (Isobe et al. 2003) using restriction fragment length polymorphisms (RFLPs) that provided a linkage map for trait mapping. This effort has been refined by the addition of microsatellite simple sequence repeat (SSR) and amplified fragment length polymorphism (AFLPs) markers, providing a more comprehensive linkage map consisting of roughly 7500 markers (Isobe et al. 2009). As the cost of sequencing has gone down, more tools and transcriptomic resources were developed that allowed gene expression analysis on a genome-wide scale (Chakrabarti et al. 2016; Yates et al. 2014). Subsequently, whole genome resources have been developed that will allow for genome analysis despite the high heterozygocity and heterogeneity of red clover cultivars. This work describes some of these

available resouces, their uses in characterizing gene expression involved in the interaction with rhizobium, as well as highlighting molecular biology approaches being used with red clover.

Methods

The red clover (*Trifolium pratense*) genomes (Accession Nos. GCA_000583005.2, GCA_900079335.1, GCA_900292005.1, GCA_020283565.1); the white clover (*Trifolium repens*) genome (Accession No. GCA_005869975.1); the western clover (*Trifolium occidentale*) genome (Accesstion No. GCA_012979555.1) were downloaded from NCBI; the subterranean clover (*Trifolium subterraneum*) genome was obtained as a pseudo-molecule file via a link from clovergarden.jp: (https://drive.google.com/drive/folders/1sP5UmbIlBGRGsgy1tcghGN8XIjQJ8nu7) - (Accessed 2022_11_07). The red clover annotations and assembly information were downloaded from Legume Information Service (https://data.legumeinfo.org/Trifolium/pratense/) and Phytozome (https://phytozome-next.jgi.doe.gov/info/Tpratense_v2). Minimap2 (Li, 2018) was used to identify pairwise alignments between the Trifolium genome accessions listed above and the resulting .bed files were visualized with the software Circos (Krzywinski et al., 2009). Root harvest from inoculated and uninoculated red clover plants, mRNA isolation, RNA-seq library preparation and sequencing, and mapping to the genome assembly was done as described in Dinkins et al. (2021; 2022).

Results and Discussion

The high heterozygosity in red clover has complicated the generation of chromosome-level assemblies, however four genome assemblies are currently in the NCBI databases. The first, derived the from Tatra cultivar published by researchers at Masaryk University (Istvanek et al. 2014) that can be accessed at GenBank Accession No. GCA_000583005.2 is comprised of 267,372 contigs encompassing 305 Mb that contain gene information but are not assigned to linkage groups or assembled chromosomes. This work was followed shortly after by The Genome Analysis Center (TGAC) and collaborators, with a genome from the cultivar Milvus B, (Accession No. GCA_900079335.1) that consisted of the seven chromosomes and 39,913 unlinked scaffolds (De Vega et al. 2015). The total sequence length is 345 Mb of which 309 Mb are linked to the linkage group scaffolds. The annotated assembled genome is not available at NCBI, however the assembly annotations can be downloaded from Legume Information Service (https://data.legumeinfo.org/Trifolium/pratense/) and Phytozome (https://phytozome-next.jgi.doe.gov/info/Tpratense_v2). The TGAC group followed with a second genome submission (labeled as version 3) in 2018 (GenBank Accession No. GCA_900292005.1).

The most recent genome assembly, ARS_RC_1.1 (GenBank Accession No. GCA_020283565.1), was derived from the HEN17-A07 line in the USDA-ARS collection. This work was carried out by ARS scientists at the U.S. Dairy Forage Center in Madison, WI and the U.S. Meat Animal Research Center in Clay Center, NE using not only short read sequence, but adding both long read sequence and Hi-C proximity ligation leading to a higher quality assembly with 20% more asembled sequence and an error rate three orders of magnitude lower than the previous TGAC assembly (Bickhart et al. 2022). The ARS_RC_1.1 genome is comprised of 414 Mb assembled into seven linkage groups, plus the chloroplast and mitochondrial genomes although it still contains 90 unlinked scaffolds. The chromosome order and linkage of both the TGAC and ARS_RC_1.1 genomes are relatively similar (Figure 1). Other genome assemblies have been described in the literature (Yan et al. 2022), but the lack of public availability of these precludes usage by the red clover research community.

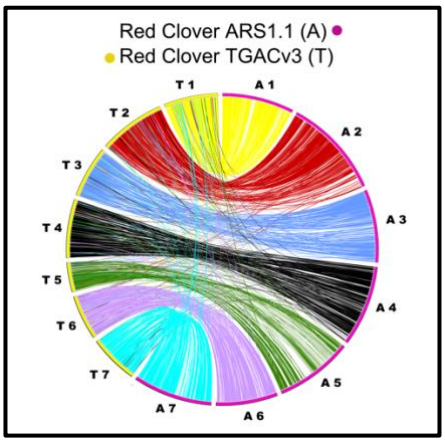

Figure 1. Synteny at the 100 Kb range between the red clover ARS_RC_1.1 genome (in magenta) and the TGAC red clover assembly (in yellow). Linkage Group designations are top right LG1-LG7 clockwise for ARS_RC_1.1 (A1 – A7) and counter-clockwise for the TGAC genome (T1 – T7).

While additional work is warranted to further refine the assemblies, the availability of these red clover assemblies, along with the increasing availability of closely related Trifolium species, such as subterranean clover, western clover and white clover (Figure 2), will aid in the mapping of genes and chromosomal regions important in legume biology.

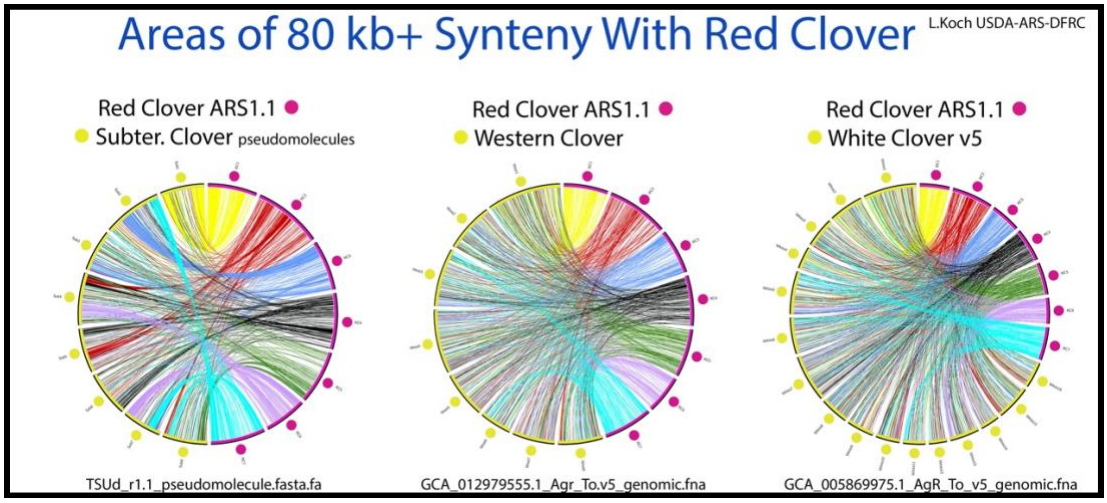

Figure 2. Synteny at the 80 Kb range between the red clover ARS_RC_1.1 genome with the genome assemblies of subterranean, western and white clovers. Linkage Group designations are as presented in Figure 1.

Use of the genome for mapping of gene expression using RNA-seq has allowed for the characterization of several traits associated with red clover. One of the most important contributions of red clover in pasture settings is nitrogen fixation. While red clover has not been one of the species used for modeling the interaction with rhizobium and the symbiosis process, the availability of the red clover genome now makes it possible for red clover to join in this role, while also creating the potential to identify red clover specific genes in this interaction and in nitrogen fixation (Dinkins et al., 2022). Red clover has been shown to have, and express, one of the highest number of nodule-specific cysteine rich proteins (NCRs) and nodule-specific polycystin-1, lipoxygenase, alpha toxic (PLAT) domain proteins (NPDs). Both the red clover NCRs and NPDs genes were found to be expressed in RNA samples derived from roots harboring nodules with no, or minimal, expression in non-inoculated roots (Figures 3 and 4).

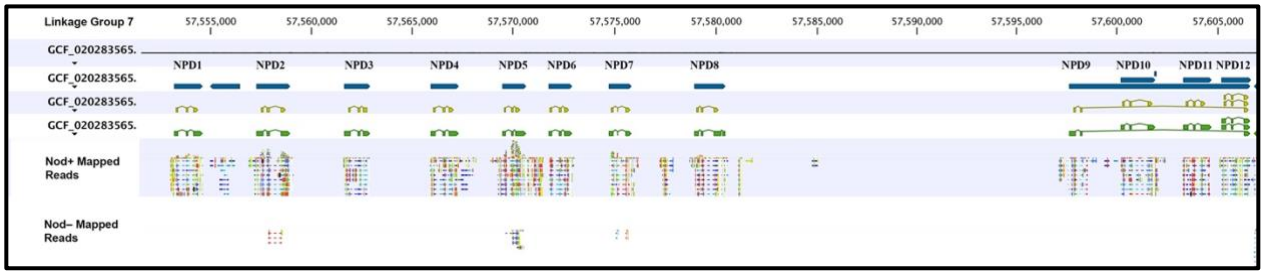

Figure 3. Genomic alignment and expression of nodule-specific polycystin-1, lipoxygenase, alpha toxic (PLAT) domain protein (NPD) genes in red clover. The blue bars denote the location and direction of the genes, the yellow and green broken bars denote the coding sequence and mRNA, respectively, where the thick regions represent the exons linked by the thin lines representing the introns. The expression profiles in the Nod– and Nod+ mapped reads show the number of reads mapping to each region, where each bar represents one sequence read, and the dotted lines indicate alignment to the exon (solid bars) across the intron.

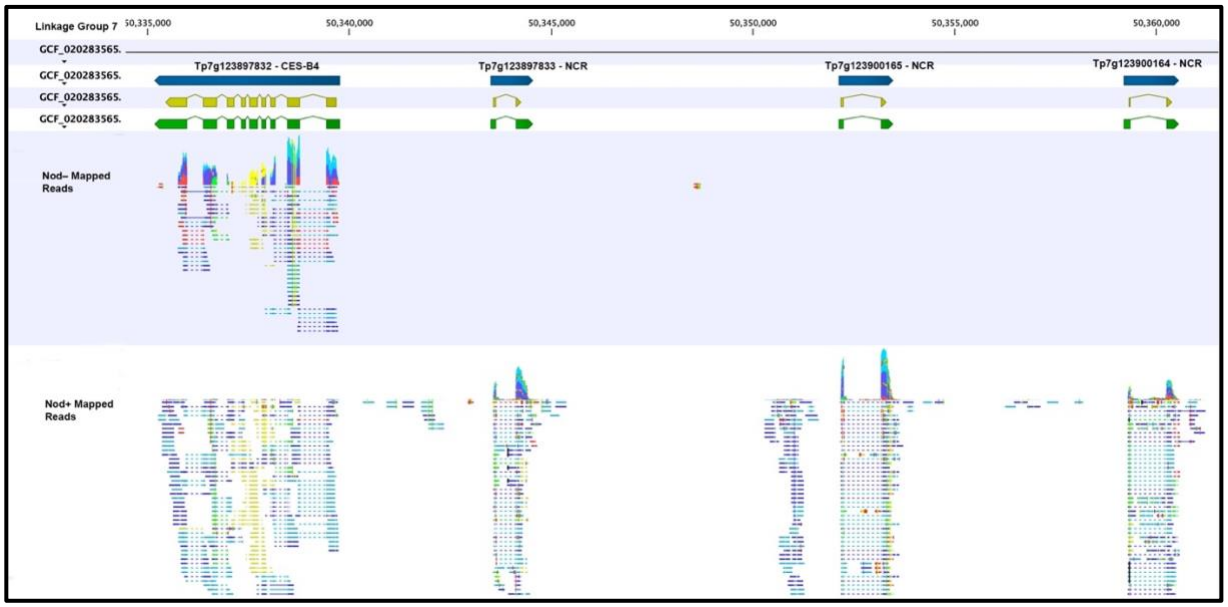

Figure 4. Expression of three closely linked red clover nodule-specific cysteine rich (NCRs) protein genes on Linkage Group 7. Upstream of the NCRs is Tp7g123897832, a putative cellulose synthase gene that is expressed constitutively in both red clover inoculated (Nod+) and uninoculated (Nod–) roots. Descriptions for each lane is as presented in Figure 3.

Conclusions

The availability of red clover genome assemblies has allowed significant progress in characterizing red clover genotype diversity and persistence (Osterman et al. 2021; Jones et al. 2020; Li et al. 2019; Istvanek et al. 2017), mapping of 2,4-D tolerance (Benevenuto et al. 2019), identifying genes involved in isoflavone production (Dinkins et al. 2021; Shi et al. 2021), identifying red clover genotypes differing in their nitrogen fixation potential (Trněný et al. 2019) and nutritional quality (Herbert et al. 2022), and examining gene expression during rhizobium symbiosis (Dinkins et al. 2022). With further refinements of genomic reference sequences and the associated gene annotations, these resources will be useful tools to further our understanding of the genetic basis of traits of interest and physiological processes such as

plant-rhizobium interactions, persistence, and forage yield and quality. Understanding the genetic basis of these will facilitate development of improved red clover cultivars through via breeding or transgenic approaches.

Acknowledgements

The authors acknowledge Troy Bass and Julie Hancock for greenhouse and laboratory work on the Nod+ and Nod– plants. This research was supported by the United States Department of Agriculture USDA–ARS CRIS project 5042-21000-002-00D. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

References

Ball DM, Hoveland CS, Lacefield GD (2007) Southern Forages.332

Benevenuto J, Bhakta M, Lohr DA, Ferrão LFV, Resende MFR, Kirst M, Quesenberry K, Munoz P (2019) Cost-effective detection of genome-wide signatures for 2,4-D herbicide resistance adaptation in red clover. Sci Rep 9 (1). doi:10.1038/s41598-019-55676-9

Bickhart DM, Koch LM, Smith TPL, Riday H, Sullivan ML (2022) Chromosome-scale assembly of the highly heterozygous genome of red clover (*Trifolium pratense* L.), an allogamous forage crop species. Gigabyte 2022:1-13. doi:10.46471/gigabyte.42

Carlsson G, Huss-Danell K (2003) Nitrogen fixation in perennial forage legumes in the field. Plant Soil 253 (2):353-372. doi:10.1023/a:1024847017371

Chakrabarti M, Dinkins RD, Hunt AG (2016) *De novo* Transcriptome Assembly and Dynamic Spatial Gene Expression Analysis in Red Clover. Plant Genome 9 (2). doi:10.3835/plantgenome2015.06.0048

De Vega JJ, Ayling S, Hegarty M, Kudrna D, Goicoechea JL, Ergon A, Rognli OA, Jones C, Swain M, Geurts R, Lang C, Mayer KF, Rossner S, Yates S, Webb KJ, Donnison IS, Oldroyd GE, Wing RA, Caccamo M, Powell W, Abberton MT, Skot L (2015) Red clover (*Trifolium pratense* L.) draft genome provides a platform for trait improvement. Sci Rep 5:17394. doi:10.1038/srep17394

Dinkins RD, Hancock J, Coe BL, May JB, Goodman JP, Bass WT, Liu J, Fan Y, Zheng Q, Zhu H (2021) Isoflavone levels, nodulation and gene expression profiles of a CRISPR/Cas9 deletion mutant in the isoflavone synthase gene of red clover. Plant Cell Rep 40:517-528. doi:10.1007/s00299-020-02647-4

Dinkins RD, Hancock JA, Bickhart DM, Sullivan ML, Zhu H (2022) Expression and Variation of the Genes Involved in Rhizobium Nodulation in Red Clover. Plants 11 (21). doi:10.3390/plants11212888

Evers GW (2011) Forage legumes: Forage quality, fixed nitrogen, or both. Crop Sci 51:403-409

Herbert, DB, Gross, T, Rupp, O., Becker, A. (2021) Transcriptome analysis reveals major transcriptional changes during regrowth after mowing of red clover *(Trifolium pratense*). BMC Plant Biol 21:95. doi:10.1186/s12870-021-02867-0

Isobe S, Klimenko I, Ivashuta S, Gau M, Kozlov NN (2003) First RFLP linkage map of red clover (Trifolium pratense L.) based on cDNA probes and its transferability to other red clover germplasm. Theor Appl Genet 108 (1):105-112. doi:10.1007/s00122-003-1412-z

Isobe S, Kolliker R, Hisano H, Sasamoto S, Wada T, Klimenko I, Okumura K, Tabata S (2009) Construction of a consensus linkage map for red clover (*Trifolium pratense* L.). BMC Plant Biol 9:57. doi:10.1186/1471-2229-9-57

Istvanek J, Dluhosova J, Dluhos P, Patkova L, Nedelnik J, Repkova J (2017) Gene Classification and Mining of Molecular Markers Useful in Red Clover (*Trifolium pratense*) Breeding. Front Plant Sci 8:367. doi:10.3389/fpls.2017.00367

Istvanek J, Jaros M, Krenek A, Repkova J (2014) Genome assembly and annotation for red clover (*Trifolium pratense*; Fabaceae). Am J Bot 101 (2):327-337. doi:10.3732/ajb.1300340

Jones C, De Vega J, Lloyd D, Hegarty M, Ayling S, Powell W, Skøt L (2020) Population structure and genetic diversity in red clover (*Trifolium pratense* L.) germplasm. Sci Rep 10 (1). doi:10.1038/s41598-020-64989-z

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. (2009) 19:1639-1645. doi: 10.1101/gr.092759.109.

Li, H. (2018) Minimap2: Pairwise Alignment for Nucleotide Sequences. Bioinformatics 34(18):3094-3100. doi:10.1093/bioinformatics/bty191

Li W, Riday H, Riehle C, Edwards A, Dinkins R (2019) Identification of Single Nucleotide Polymorphism in Red Clover (*Trifolium pratense* L.) Using Targeted Genomic Amplicon Sequencing and RNA-seq. Front Plant Sci 10. doi:10.3389/fpls.2019.01257

Osterman J, Hammenhag C, Ortiz R, Geleta M (2021) Insights Into the Genetic Diversity of Nordic Red Clover (*Trifolium pratense*) Revealed by SeqSNP-Based Genic Markers. Front Plant Sci 12. doi:10.3389/fpls.2021.748750

Riday H, Krohn AL (2010) Genetic map-based location of the red clover (*Trifolium pratense* L.) gametophytic self-incompatibility locus. Theor Appl Genet 121 (4):761-767. doi:10.1007/s00122-010-1347-0

Shi K, X, Pan X, Liu J, Gong W, Gong P, Cao M, Jia S, Wang Z. (2021) Unveiling the Complexity of Red Clover (*Trifolium pratense* L.) Transcriptome and Transcriptional Regulation of Isoflavonoid Biosynthesis Using Integrated Long and Short-Read RNAseq. Int J Mol Sci 22:12625. doi:10.3390/ijms222312625

Taylor N (2008) A century of clover breeding developments in the United States. Crop Sci 48:1-13. doi:10.2135/cropsci2007.08.0446

Trněný O, Vlk D, Macková E, Matoušková M, Řepková J, Nedělník J, Hofbauer J, Vejražka K, Jakešová H, Jansa J, Piálek L, Knotová D (2019) Allelic Variants for Candidate Nitrogen Fixation Genes Revealed by Sequencing in Red Clover (*Trifolium pratense* L.). Int J Molec Sci 20 (21). doi:10.3390/ijms20215470

Trujillo DI, Silverstein KAT, Young ND (2019) Nodule-specific PLAT domain proteins are expanded in the Medicago lineage and required for nodulation. New Phytol 222 (3):1538-1550. doi:10.1111/nph.15697

Vizintin L. J, B and Bohanec B. (2006) Genetic characterization of selected Trifolium species as revealed by nuclear DNA content and ITS rDNA region analysis. Plant Sci 170:859-866

Yan Z, Sang L, Ma Y, He Y, Sun J, Ma L, Li S, Miao F, Zhang Z, Huang J, Wang Z, Yang G (2022) A de novo assembled high-quality chromosome-scale *Trifolium pratense* genome and fine-scale phylogenetic analysis. BMC Plant Biol 22 (1). doi:10.1186/s12870-022-03707-5

Yates SA, Swain MT, Hegarty MJ, Chernukin I, Lowe M, Allison GG, Ruttink T, Abberton MT, Jenkins G, Skot L (2014) De novo assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. BMC Genomics 15:453. doi:10.1186/1471-2164-15-453