

Enhanced Genome Assemblies of French-Bred Dactylis glomerata and Medicago sativa: Achieving High-Quality Tetraploid Genomes

Marie Pegard, Adela Poublan-Couzardot, Philippe Barre, Simon de Givry, Christine Gaspin, Fabrice Legeai, Frédéric Choulet, Christophe Klopp, Bernadette Julier

▶ **To cite this version:**

Marie Pegard, Adela Poublan-Couzardot, Philippe Barre, Simon de Givry, Christine Gaspin, et al.. Enhanced Genome Assemblies of French-Bred Dactylis glomerata and Medicago sativa: Achieving High-Quality Tetraploid Genomes. FCAG 2025 - 36th Meeting of the EUCARPIA Fodder Crops and Amenity Grasses Section, Aberystwyth University, Sep 2025, Aberystwyth, United Kingdom. pp.1-2. ⟨hal-05265025⟩

HAL Id: hal-05265025

https://hal.inrae.fr/hal-05265025v1

Submitted on 17 Sep 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



Breeding and Genetic Improvement for a Net-Zero Future

Abstracts of the 36th Meeting of the EUCARPIA Fodder Crops and Amenity Grasses Section

Aberystwyth

1-5 September, 2025

Edited by Gancho Slavov and Ana Winters

Enhanced genome assemblies of French-bred *Dactylis glomerata* and *Medicago sativa*: achieving high-quality tetraploid genomes

Marie Pegard[1], Christophe Klopp[2], Adela Poublan-Couzardot[2], Philippe Barre[1], Simon de Givry[2], Christine Gaspin[1], Fabrice Legeai[3,4], Frédéric Choulet[5], Bernadette Julier[1]

[1]INRAE, URP3F, Lusignan, France
[2]INRAE, Genotoul-Bioinfo, MIAT, Auzeville-Tolosane, France
[3]INRAE, BIPAA, IGEPP, Le Rheu, France
[4]INRIA, IRISA, GenOuest Core Facility, Rennes, France
[5]INRAE, Université Clermont Auvergne, GDEC, Clermont-Ferrand, France

Having a reference genome for forage species is crucial for several reasons. It enables a deep understanding of the genetic structure and diversity of these species, essential for breeding and improvement programs. It is used to identify genes responsible for important agronomic traits, such as disease resistance, abiotic stress tolerance, and nutritional quality. A reference genome also facilitates the use of marker-assisted selection and genomic prediction, accelerating genetic improvement aimed at releasing robust and environmentally adapted varieties. It supports studies on natural populations to discover adaptive genes, aiding conservation and innovation in sustainable agriculture.

Haplotype-resolved assembly of autopolyploid genomes is challenging due to high sequence similarity and large repetitive regions. Recent evidence shows that combining HiFi (high-fidelity) long reads and Hi-C data achieves better contiguity and phasing accuracy. We present an assembly strategy based on post-processing Hifiasm results, significantly improving the final assembly of the 3.2 Gbp genome of *Medicago sativa* (lucerne) and the 7.49 Gbp genome of *Dactylis glomerata* (cocksfoot), both autotetraploids, using HiFi data and Hi-C reads. The assembly process began with an initial assembly using Hifiasm, which resulted in good quality assembly (For lucerne 3,415 Mbp across 5,696 contigs N50 = 94 Mbp, L50 = 147 and for cocksfoot 7,664 Mbp across 3,918 contigs N50 = 99 Mbp, L50 = 215) but with imbalanced haplotype lengths and inconsistent BUSCO scores. Hi-C scaffolding revealed inversions and misassemblies, requiring manual curation. Contigs were reassigned using Toulbar2, a combinatorial optimization tool, based on shared protein alignments. For lucerne, further refinement involved adjusting constraints by modifying the cost of shared protein and penalizing unlikely haplotype switches while preserving the original Hifiasm structure. Additional steps included the identification and removal of mitochondrial and chloroplast sequences from the assembly. Telomeric repeats were detected, although often present at the chromosome extremities. This method improved haplotype lengths, BUSCO scores, and Hi-C scaffolding accuracy, reducing manual corrections and resulting in more balanced final haplotype sizes. Final haplotype sizes were 759–779 Mbp across 451-753 contigs for lucerne and 1.84-1.92 Gbp across 347-587 contigs for cocksfoot, with BUSCO completeness over 98% for lucerne and 96% for cocksfoot, and duplication rates of 8.2–13.2% and 7-9.5%, respectively.

This study demonstrates the necessity of developing strategies to improve assemblies of challenging polyploid genomes. Having these two high-quality genomes is valuable for further fundamental and applied research in plant genetics, contributing to food security and the sustainability of agricultural systems.