



# Revealing the complexities of genetically modified plant genomes

Following their successful sequencing and assembly of a highly contiguous and accurate *Arabidopsis thaliana* genome in just 4 days using a single MinION Flow Cell<sup>1</sup>, the team at the J. Craig Venter Institute, together with researchers from the Salk Institute, turned their attention to using long nanopore sequencing reads to characterise transformed, genetically modified lines<sup>2,3</sup>.

One of the most common methods for introducing new genetic material into plant cells is through the use of the bacterium *Agrobacterium tumefaciens*. This plant pathogen randomly inserts DNA contained within its plasmid to double-strand breaks within the host genome. For both scientific and regulatory reasons, it is important to characterise the insertion sites and copy number of this transfer DNA (T-DNA); however, traditional analysis techniques, such as Southern blotting, can be laborious and lack resolution.

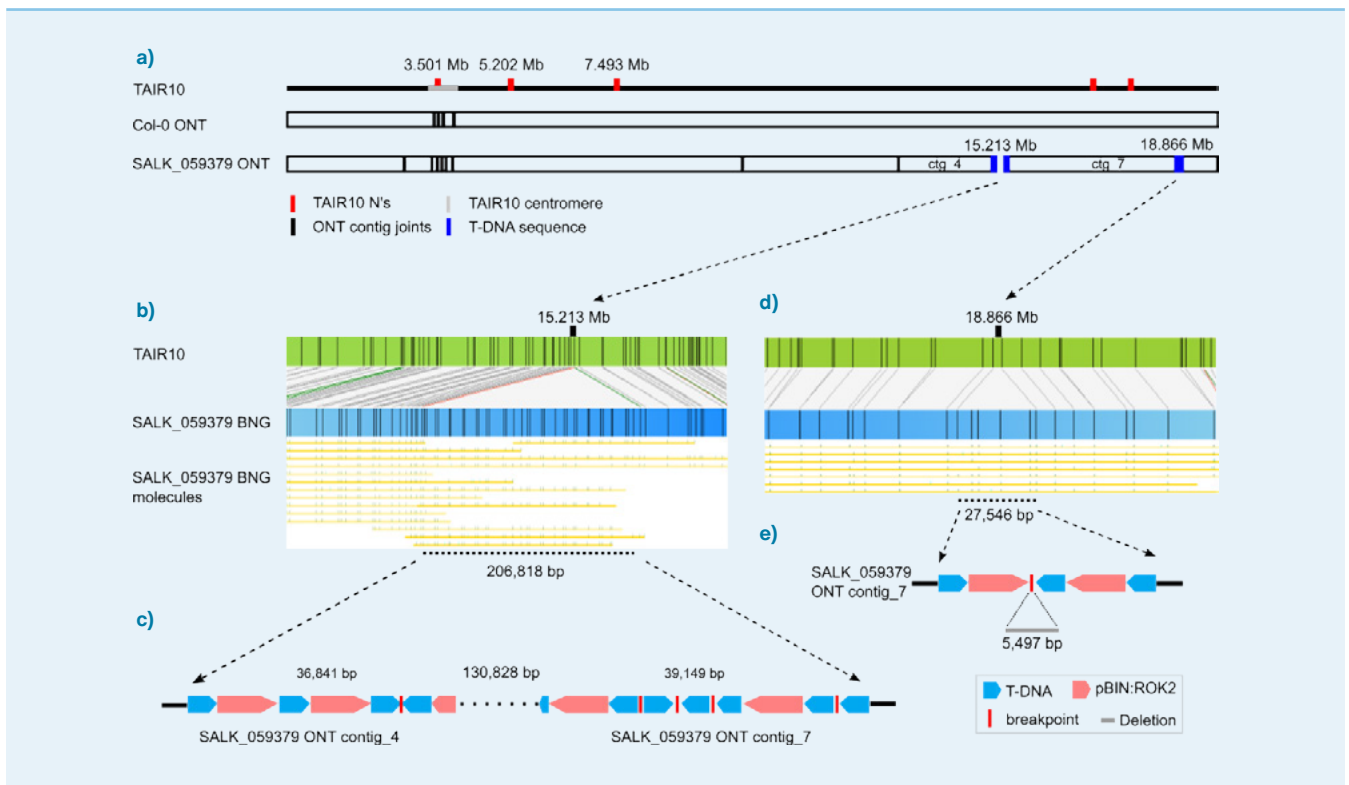
To address this challenge, the research team utilised both optical mapping and long nanopore sequencing reads to examine the genome of two transformed and one reference line of *A. thaliana* — with each line being fully sequenced on single MinION Flow Cells. Highly contiguous genomes were assembled with complete chromosome arms being contained within just one or two contigs. Genome analysis allowed the team to resolve T-DNA structures up to 36 kb in length and revealed large-scale T-DNA associated translocations and exchange of chromosome arm ends.

Moreover, sequence contigs for the two transgenic lines (SAIL\_232 and SALK\_059379) captured up to 39 kb of assembled T-DNA insertion sequence, providing sufficient information to better understand the complexity of such *Agrobacterium*-mediated transgene insertions (e.g. rearrangements, insertions, deletions, etc.) and the effect of these insertions on proximal genes.

**Nanopore sequencing allowed the resolution of T-DNA structures up to 36 kb in length<sup>3</sup>.**

This study provides new insights into the structural impact of engineering plant genomes and demonstrates the utility of state-of-the-art, long-range sequencing technologies to rapidly identify unanticipated genomic changes. The team now plans to utilise the nanopore sequencing data to identify the methylation status of their transformed genomes, with a view to the potential replacement of separate bisulfite sequencing-based methylation analysis. Commenting on this research, Professor Todd Michael remarked: *'It has been known that [T-DNA] insertions have variable length and that there are many of them but, up until Oxford Nanopore technology with long reads, it was really impossible see what those insertions looked like<sup>2</sup>.*

Recently, researchers at University of Nebraska-Lincoln combined a target enrichment strategy with long-read nanopore sequencing to cost-effectively characterise the insertion sites of the 1,166 bp maize *Dissociation (Ds)* transposable element in transgenic soybean lines<sup>4</sup>. *Ds* element DNA was captured and enriched in 51 soybean lines using a biotinylated oligo probe, prior to barcoding and running all samples on a single MinION Flow Cell. The team reported



**Figure 1**

Nanopore sequencing using the MinION resolves assembly errors in the short-read *Arabidopsis thaliana* reference genome and the T-DNA insertion sites in the transformed lines. (a) Chromosome 2 of: TAIR10 wild-type reference line, the nanopore-sequenced Col-0 wild-type line, and SALK\_059379 transformed line. Both chromosome arms are present in single Col-0 contigs and misassemblies in the TAIR10 reference are corrected. Blue boxes in the SALK line represent T-DNA insertions, which were apparent in the (b,d) optical maps. (c,e) The nanopore sequencing data allowed complete resolution of the 27,546 bp insertion and partial resolution of the 206,818 bp insertion. Figure courtesy of Professor Todd Michael, J. Craig Venter Institute, US.

complete concordance with previously obtained results for each of these well-characterised plant lines and insertion sites. Importantly, the total sequencing cost for all 51 samples was \$1,360 (\$27/sample) and all results were generated within just one week. Furthermore, the on-target coverage, which for one cell line was as high as 3,555x, could be further reduced. This, coupled with the potential to increase enrichment efficiency, suggests that more

samples could be analysed in a single sequencing run, enabling further cost savings. In addition, the methodology also allowed accurate mapping of multiple transgene insertions within individual soybean lines. According to the researchers: ‘These results demonstrate that this nanopore-based sequencing method is rapid, convenient, reliable, cost efficient, and high throughput’<sup>14</sup>.

**Find out more about plant genome sequencing at [www.nanoporetech.com/applications](http://www.nanoporetech.com/applications)**

## References

1. Michael, T.P. et al. High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat. Commun.* 9(1):541 (2018).
2. Michael, T. The complex architecture of plant T-DNA transgene insertions. Presentation. Available at: <https://nanoporetech.com/resource-centre/complex-architecture-plant-t-dna-transgene-insertions> [Accessed: 10 December 2019]
3. Jupe, F. et al. The complex architecture and epigenomic impact of plant T-DNA insertions. *PLoS Genet.* 15(1): e1007819 (2019).
4. Li, S. et al. Mapping of transgenic alleles in soybean using a nanopore-based sequencing strategy. *J. Exp. Bot.* 7:70(15):3825-3833 (2019).