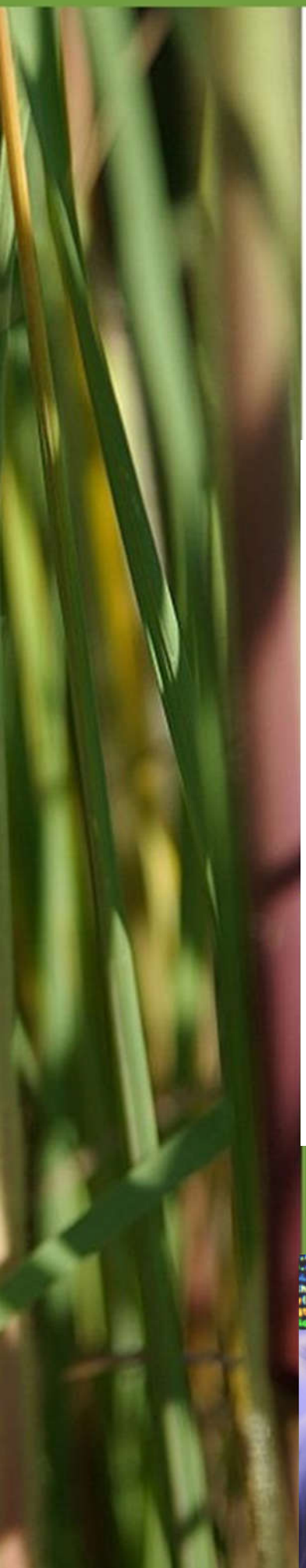
WATBIO

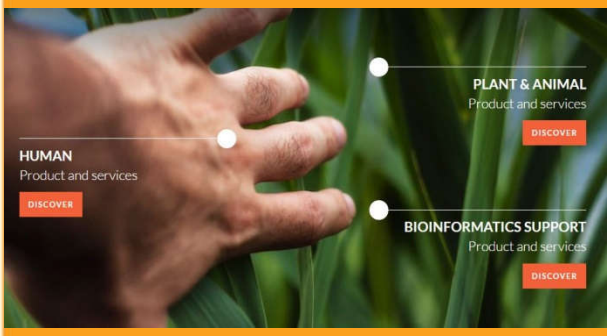
Innovation Note 2

Development of improved perennial non-food and bioproduct crops for water stressed environment

2 May 2017



IGA Technology Services: innovation in sequencing through WATBIO



IGA Technology Services sequencing and bioinformatics offer

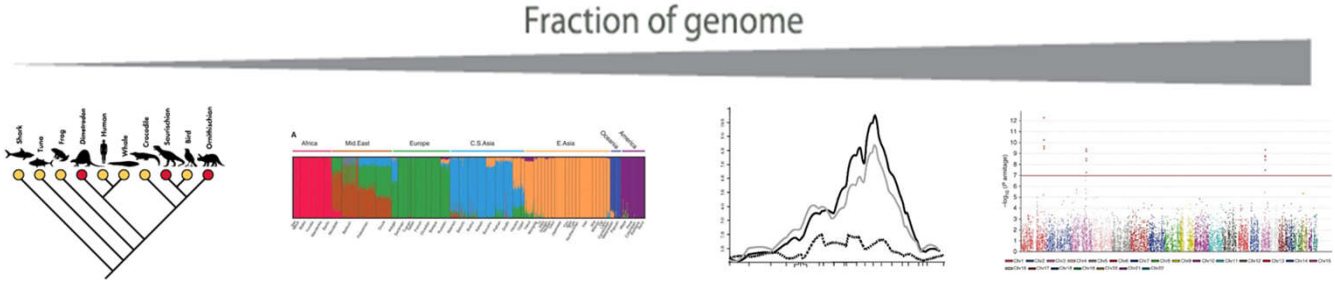
From Davide Scaglione and Simone Scalabrin
Bioinformatics Operation Manager at
IGA Technology Services

IGA Technology Services (IGATS) based in Udine in Italy is the leading SME partner in WATBIO and conducts all of the next-generation sequencing work in the project. A large proportion of the WATBIO DNA and RNA sequencing work is done on poplar, and the WATBIO project has enabled IGATS to drive significant innovation in the type of sequencing involved.

Genome-wide association analysis (GWAS) involves scanning the genome in a wide range of individuals to find associations between variations in genes and observed traits (the phenotype). For example, in a drought experiment we might be interested in what are the genetic variants, usually within individual genes, that are characteristic of trees with a large leaf under drought-stressed and under well-watered conditions. Given the high number of genes in a complex organism, this is like searching for a needle in the haystack. To properly tackle the problem, the whole genome needs to be examined carefully. The challenge presented by WATBIO was to obtain an adequately dense genotypic data set in an ambitious study of black poplar. While a previously developed set of 12,000 variants was available, it did not have enough genetic markers. The lack of an adequate number of markers leads to greater uncertainty about the links between markers. This leads to low statistical power from utilising markers that are not tightly linked to functional genes and unexpected diversity may need to be explored.

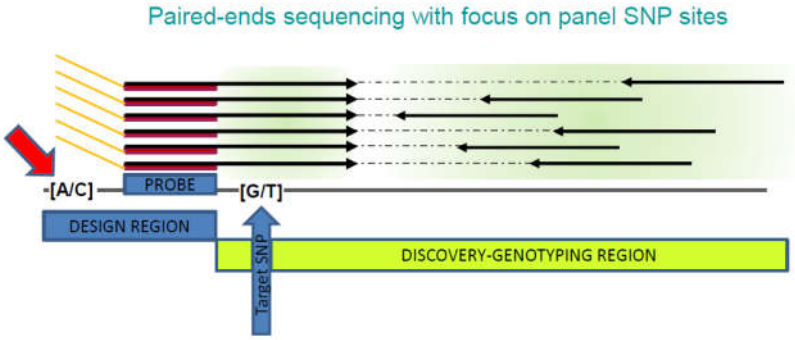
IGATS pioneered an innovation in the use of the enrichment technology developed by NuGEN Technologies Inc. (NuGEN) which utilizes single primer-based extension technology (SPET). SPET delivers a customizable sample preparation product featuring technologies that address common limitations and maximize informative sequencing. With minor modifications to the original protocols, we adopted the standard preparation kit so that data from reads from both ends of DNA fragments are used. Our original target set of polymorphisms (90,000) all resided on or near all the coding genes of black poplar. The consideration of reads from both ends enabled us to call about 500,000 SNPs greatly increasing the efficiency of markers production.

From wet-lab to digital analysis for a large number of samples and *loci*.



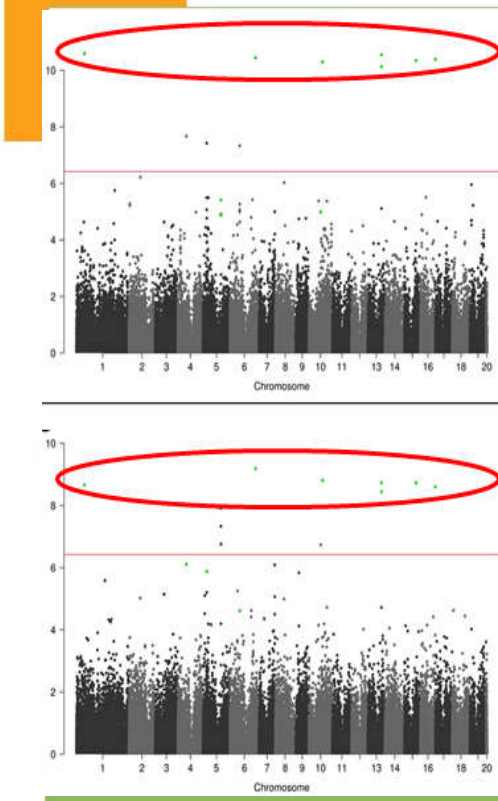
Above: The genome complexity reduction of different approaches varies according to the requested genotyping level. GWAS (Association studies) is one of the most complicated.

For the first time a single primer enrichment system enabled the genotyping of both target and non-target polymorphism on black poplar for more than 500,000 nucleotides: some 66,000 form the target set of 90,000, while the remainder were *de novo* identified by the sequencing information for each locus. Quality of data points was assessed by the successful repeated discovery of markers associated with phenotypic traits under study, especially at the University of Southampton.



Above: The SPET genotyping concept. The IGATS innovation is the development of sequencing that makes use of reads from the right side complementing the standards NuGEN approach of reading from the left side.

From this pilot experiment, NuGEN has decided to promote the utilization of their technology also for massive genotyping by *ad-hoc* optimization of the kit to reduce the sequencing costs for this type of application and give an improved multiplexing capacity to cope with very large cohort of samples. Complementing this, IGATS as an SME member of the WATBIO consortium, is now routinely offering genotyping services using this innovation with reduced costs thanks to the further optimization in chemistry, faster protocol and higher-level of throughput now available from the manufacturer. This method has a high chance of becoming a leading genotyping-by-sequencing solution for medium-large polymorphisms number (5,000-90,000) in medium-large cohort of samples (>500).



Above: The X axis represents the genome space of black poplar distributed over 19 chromosomes. 55 variants (green dots) in 44 genes associated with the key trait leaf size were identified in plants that are irrigated (upper plot) or rainfed (lower plot). The reduced number of visible green dots is because a large proportion of the SNPs responsible for leaf size are very close to each other on the genome.

WATBIO (Development of improved perennial non-food biomass and bioproduct crops for water stressed environments) is funded by the European Union's Seventh Programme for research, technological development and demonstration under the grant agreement FP7-311929