Tools and Technologies to Advance Plant Research

Programme and Abstracts

Exploring the many and new applications of Next Generation Sequencing in Plant Research.

Workshop organized by GARNet and the Arabidopsis Sectional Interest group of the Genetics Society.

Genetics Society Arabidopsis Sectional Interest Group Meeting

Tools and Technologies to Advance Plant Research

University of Liverpool, UK

Organising Committee
Ant Hall (University of Liverpool)
Ruth Bastow (GARNet)
Charis Cook (GARNet)

Acknowledgements
This Genetics Society Sectional Interest Group meeting is funded by The Genetics Society.

About GARNet
GARNet represents UK-based Arabidopsis researchers via a committee of elected members. It aims to make sure the UK Arabidopsis community remains competitive and productive at the national and international level by helping researchers make the best use of available funding, tools and resources.
www.garnetcommunity.org.uk

About the Genetics Society
The Genetics Society was founded by William Bateson in 1919 and is one of the oldest "learned societies" devoted to Genetics in the world. Its activities include organising meetings to promulgate genetics, sponsoring research through fieldwork grants and student bursaries, and promoting the Public Understanding of Genetics.
http://www.genetics.org.uk/

Table of Contents

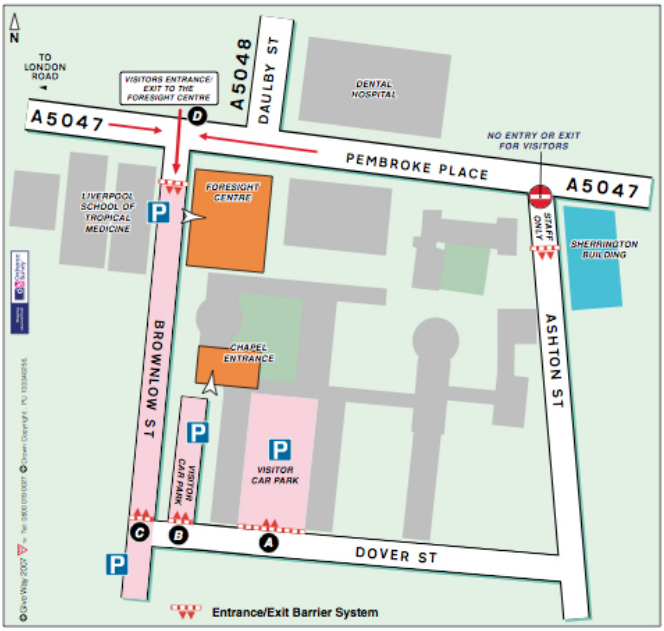
Venue Information

Wifi: Please ask at reception for a password to access the wireless network.

Food: Vegetarian and vegan options are labelled for your convenience. Please note that meat is not halal.

Venue: A map of the immediate area around the venue is provided below. Car parking is available to delegates in the carparks shown. Please take your carpark ticket to the Foresight Centre reception for instructions on how to validate it.

If you are travelling on the train, leave the Lime Street Station by the exit nearest to ticket office. Go through the taxi rank and turn right from the station and take the first left into Pudsey Street. At end of road turn right into London Road and continue on this road, which becomes Pembroke Place. The Foresight Centre can be accessed by pedestrians through gates on Pembroke Place. The entrance to the Foresight Centre (Old Royal Infirmary Hospital) is No1, the first entrance on the left.



Programme

10:00 Registration and coffee

10:30 Welcome and Introduction – Anthony Hall

11:00 ‘The past, present, and future of next generation sequencing: the good, the bad and the ugly.’ Neil Hall (University of Liverpool)

Session 1 Next Generation Sequencing

11:45 ‘1001 Arabidopsis genomes and GWAS.’ Arthur Korte (Gregor Mendel Institute)

12:15 ‘Mutant hunting.’ Anthony Hall (University of Liverpool)

12:45 ‘Mapping complex traits in Arabidopsis thaliana.’ Paula Kover (University of Bath)

13:15 Lunch

Session 2 De–novo and comparative genomics

14:00 ‘Using multiple Brassicaceae genomes as a basis for comparative ‘omics.’ Eva-Maria Willing (Max Planck Institut for Plant Breeding Research)

14:30 ‘Complementary NGS approaches towards cereal crop genomes.’ Klaus Meyer (MIPS)

15:00 Refreshment Break

Session 3 Novel uses of generation sequencing

15:30 ‘An NGS approach to mapping chromatin structure in the Arabidopsis nuclear and plastid genomes.’ Nick Kent (University of Cardiff)

16:00 ‘Alternative splicing by RNA-seq.’ John Brown (University of Dundee)

16:30 ‘Single molecule direct RNA sequencing.’ Gordon Simpson (University of Dundee)

17:00 ‘Genome-wide analysis of cytosine methylation in plant DNA.’ Tom Hardcastle (University of Cambridge)

Finish 17:30

The past, present, and future of next generation sequencing: the good, the bad and the ugly.

Neil Hall

University of Liverpool, UK

High throughput sequencing is becoming a ubiquitous tool in biology as costs rapidly decrease. As well as de novo sequencing and re-sequencing of genomes, DNA sequencing is increasing being applied to transcriptional studies (RNAseq, CAGE), protein DNA interactions (ChIPseq) mutational studies (RIT-seq) (SHOREmap) and community profiling (metagenomics and metagenetics). In this rapidly changing field there is a plethora of new technologies on the market and many that are promised for the near future. Here I will give an overview of the current state-of-the-art technology and applications, and also give a personal view of how the technology may develop in the next few years.

Section 1: Next Generation Sequencing

1001 Arabidopsis genomes and GWAS

Arthur Korte

Gregor Mendel Institute, Austria

The common weed *A. thaliana* is highly selfing and naturally exists as locally adapted inbred lines that can readily be grown in replicate under controlled conditions. This makes it an excellent model for studying the genetics of natural variation, and, indeed, shared inbred lines have been a resource for the Arabidopsis community since its inception. More recently, over 1,300 lines have been genotyped using a 250k SNP-chip to facilitate genome-wide association studies (GWAS) and efforts are now underway to sequence over 1,000 lines leading to the identification of millions of SNPs and structural variants, increasing the 'toolkit' for GWAS dramatically. On the other side, new GWAS methods beyond single trait / single marker analysis have been developed. The combination of increased marker density and knowledge of genome architecture and the improved statistical analysis greatly enlarged our understanding of the genotype-phenotype map in *A. thaliana*.

Mutant hunting

Anthony Hall, Neil Hall, Rachel Brenchley and Laura Gardiner

University of Liverpool, UK

Next generation sequencing technology is making it possible to rapidly map and identify mutations responsible for specific traits or phenotypes in Arabidopsis. These strategies include simultaneous mapping and mutant identification (SHOREmapping) and direct sequencing of mutants. While these approaches are extremely useful for Arabidopsis and are quickly becoming routine, in crop species genome resources are often poor and the genome sizes are huge.

Here, we will describe an alternative strategy, based on the SHOREmapping approach. We will apply this mapping and mutant identification to a complex genome, in this case wheat. Bread wheat is an allohexaploid with a genome size of 17GB. While a draft genome is available it is fragmented. Current sequencing technologies and computational speeds make a direct re-sequencing of a bulk segregating population of an F2 with sufficient sequence depth prohibitively expensive for large genomes. Therefore, our first step in the development of a SHOREmapping approach has been to produce an enrichment array allowing us to sequence just the genic portion of wheat (150 Mb). We have used this in combination with a pseudo wheat genome, constructed based on syntany between Brachypodium and wheat. We will describe how this approach is being used to map a mutant in wheat.

Mapping complex traits in Arabidopsis thaliana

Paula Kover

University of Bath, UK

A major goal of evolutionary genetics is to understand how genetic changes contribute to adaptive evolution. To achieve such an understanding it is necessary to combine knowledge of the genetic basis of traits under selection with knowledge of how selection acts on the genetic variation available to modify phenotype. The fact that most traits of ecological, evolutionary and economical importance are complex (i.e. determined by multiple loci and affected by the environment), has made it more difficult to study the evolutionary process at the genetic level empirically.

I will review different methods to identify genetic factors underlying quantitative variation, including the Multiparent Advanced Genetic InterCross (MAGIC) lines developed by the intercross of 19 accessions of *Arabidopsis thaliana*. In particular, I will present results from mapping efforts to characterize the genetic basis of natural variation in flowering time, including genome-wide searches for loci that respond to selection for early flowering (using an experimental evolution approach). Comparison of these two approaches allows interesting analysis about the genetic basis of adaptive traits, as well as the predictability and repeatability of the adaptive process.

Session 2: De-novo and Comparative Genomics

Using multiple Brassicaceae genomes as a basis for comparative 'omics

Eva-Maria Willing

Max Planck Institute for Plant Breeding Research, Germany

The family of Brassicaceae is economically important family of flowering plants. It contains over 330 genera more than 3,000 species. Profound phylogenetic trees have been attempted, while the genome structures of a broad range of species within different genera have been well characterized. In addition, the presence of *Arabidopsis thaliana* within this family provides a reference system and many tools to test the validity of hypthotheses. These properties make it an ideal system to conduct comparative genomics and transcriptomics in plants. *Arabis alpina,* equipped with a perennial life style, diverged 25-30 million years ago from the annual plant *A. thaliana*. The relative large evolutionary distance compared to the distance between the well-characterized plants *A. thaliana* and *Arabidopsis lyrata* (4-5 million years) is assumed to accelerate the power of comparative genomics. In addition to establishing a first draft genome sequence of *A. alpina*, we deeply sequenced the cycling transcriptome from *A. thaliana* and *A. alpina* grown under the same day length regime. This setup enables us to study expression conservation and differences of cycling genes according to time, amplitude, shape and alternative splicing between the two species, but also links our findings to regulatory sequences by identifying expression specific cis-regulatory modules.

Complementary NGS approaches towards cereal crop genomes

Klaus Meyer

MIPS/IBIS, Helmholtz Center Munich, Germany

Access to cereal genomes is hampered by their enormous size, their high repeat content and (in part) by polyploidy. We developed approaches that seek to circumvent these limitations by making use of different complementary strategies that aim to detect, assemble and position genes along the chromosomes. The approaches are driven by genomic properties of cereal genomes and exploit their level of gene and synthetic conservation.

Genome Zipping makes use of pronounced synthetic relationships among grass genomes and a consensus scaffold/ a combination of synthetic scaffold are used to position and approximate gene ordering in wheat, barley among others. In silico gene traps on the other hand enable capture, order and assembly of orthologous and paralogous genes from WGS data based on homology relationships to reference datasets. For polyploids machine learning based classification to subgenomes and chromosome arm assignment (CarmA) using informative polymorphisms help to assign resulting assemblies to chromosome territories and to feed genes into the genome zipper based positioning.

Comparison of different whole chromosome shotgun assemblies (WCS) against fl-cDNAs and of genome zipper derived scaffolds from 454 derived data and Illumina WCS assemblies indicate persistant shortcomings of current assembly algorithms. For the time being combined and complementary approaches to reduce economic and technological limitations are pragmatic solutions.

Session 3: Novel Uses of Next Generation Sequencing

An NGS approach to mapping chromatin structure in the Arabidopsis nuclear and plastid genomes

Nick Kent

Cardiff University, Wales

Next-generation sequencing technology not only allows the determination of the base sequence of entire genomes, but is also amenable to the analysis of genomic DNA:protein architecture. Digestion of chromosomes in vivo with nucleases such as DNaseI or micrococcal nuclease generates mixtures of DNA fragments representing genomic sequences protected from cleavage by the binding of chromatin proteins. Using paired-end mode NGS it is possible to determine short sequence tags for both ends of individual DNA fragments within such nuclease-digested chromatin samples. When aligned to the original genome these paired sequence reads yield two useful pieces of information. Firstly, their sequences identify the genomic location of a nuclease-protected chromatin particle. Secondly, the distance between the sequences provides information about the size of the protected DNA species and therefore the type of chromatin particle. Analysed in aggregate, the millions of sequence read pairs which derive from a eukaryotic cell culture chromatin sample provide genomic maps of nuclease-resistant complexes ranging from individual transcription factor-bound DNA elements up to polynucleosomes. We have been applying this technology to a wide variety of eukaryotic model systems and will present data comparing chromatin organisation between the Arabidopsis and yeast nuclear genomes. Interestingly, the probe nucleases we employ penetrate not only nuclear membranes in vivo, but also those of chloroplasts and mitochondria. We therefore also demonstrate the possibility of using this technology to probe developmental changes in plastid nucleoid structure.

Alternative splicing in Arabidopsis - RNA-seq and beyond

John W. S. Brown

Division of Plant Sciences, College of Life Sciences, University of Dundee, James Hutton Institute, Dundee, DD2 5DA

Alternative splicing (AS) produces multiple mRNAs from the same gene through variable selection of splice sites during pre-mRNA splicing. It is the main origin of proteome complexity in eukaryotes, plays a key regulatory role in the development of all multi-cellular organisms and modulates gene expression in response to environmental signals. The advent of NGS has revolutionised analysis of AS with the result that the frequency of occurrence of AS has increased substantially over the last five years. In Arabidopsis, more than 60% of intron-containing genes undergo AS and this is likely to increase as different tissues at various developmental stages and growth conditions are analysed. Similar estimates have been obtained in other plant species (e.g. potato, barley). We are using RNA-seq, complemented by high resolution RT-PCR, to address the regulation of AS by splicing factors, how AS affects expression and its function in responses to environmental cues. RNA-seq data is allowing us to address AS conservation and protein isoforms.

Single molecule direct RNA sequencing

Gordon Simpson, Alexander Sherstnev[1], Céline Duc[1], Christian Cole[1], Vasiliki Zacharaki[1], Csaba Hornyik[3], Jennifer Grant[1], Nicholas Schurch[1], Fatih Ozsolak[2], Patrice M. Milos[2], Geoffrey J. Barton[1] and Gordon G. Simpson[1,3]

[1]University of Dundee, Scotland, UK, [2]Helicos BioSciences Corporation, Cambridge, MA, USA, [3]James Hutton Institute, Scotland, UK.

In order to examine the impact of regulated 3' end formation genome-wide we applied direct RNA sequencing (DRS) to A. thaliana. In this true single molecule sequencing procedure the site of RNA cleavage and polyadenylation is defined with an accuracy of ± 2nt in the absence of errors induced by reverse transcriptase, ligation or amplification. Here we show the authentic transcriptome in unprecedented detail and how 3' end formation impacts genome organization. We reveal extreme heterogeneity in RNA 3' ends, discover previously unrecognized non-coding RNAs and propose widespread re-annotation of the genome. We explain the origin of most poly(A)+ antisense RNAs and identify cis-elements that control 3' end formation in different registers. We have used this approach to identify and quantify genome-wide shifts in gene expression and 3' end formation in different mutant backgrounds and environmental conditions. These findings are essential to understand what the genome actually encodes, how it is organized and the impact of regulated 3' end formation on these processes. In addition, they reveal ways in which DRS can be used more widely to refine transcriptome and genome interpretation.

Genome-wide analysis of cytosine methylation in plant DNA

Thomas J. Hardcastle and David Baulcombe

University of Cambridge, UK

Cytosine methylation can be investigated at a genome wide level through high-throughput sequencing of bisulphite treated DNA. Treatment of denatured DNA with sodium bisulphite converts unmethylated cytosines into uracil; sequencing these data allows, in principle, not only the identification of methylation loci but an assessment of the proportion of cells in which methylation takes place and the identification of differential methylation between samples.

In practice, a number of issues must be resolved before such statements can be made. An unbiased mapping of unmethylated and methylated reads back to the reference genome is required. Since unmethylated cytosines are converted to uracil, reads sequenced from unmethylated DNA will map less perfectly to the genome than methylated reads. Standard tools now exist to account for this bias and will be briefly introduced.

Further challenges exist in identifying regions of methylation and identifying differential methylation. In order to find biologically meaningful results, we must account for the natural variation in methylation status between biological replicates. We present a novel set of methods for discovery of methylation loci and differential methylation from replicated data.

List of Participants

Speakers

Neil Hall	University of Liverpool	Neil.hall@liv.ac.uk
Arthur Korte	Gregor Mendel Institute	arthur.korte@gmi.oeaw.ac.at
Ant Hall	University of Liverpool	anthony.hall@liverpool.ac.uk
Paula Kover	University of Bath	p.x.kover@bath.ac.uk
Klaus Meyer	MIPS/IBIS, Helmholtz Center Munich	k.mayer@helmholtz-muenchen.de
Nick Kent	University of Cardiff	KentN@cardiff.ac.uk
John Brown	James Hutton Institute	John.brown@hutton.ac.uk
Gordon Simpson	James Hutton Institute	Gordon.simpson@hutton.ac.uk
Tom Hardcastle	University of Cambridge	david.baulcombe@plantsci.cam.ac.uk
Eva-Maria Willing	Max Planck Institut for Plant Breeding Research	willing@mpipz.mpg.de

Participants

Juan Bai	University of Leeds	fbsjba@leeds.ac.uk
Rhydian Beynon-Davies	Lancaster Environment Centre	r.beynon-davies@lancaster.ac.uk
Mohamad Sulfazli Bin Mohd Sobri	University of Manchester	mohamadzulfazli.binmohd-sobri@postgrad.manchester.ac.uk
Roksana Bonyadi-pour	University of Warwick	r.bonyadipour@warwick.ac.uk
Maurice Bosch	IBERS, Aberystwyth University	mub@aber.ac.uk
Phaitun Bupphada	University of Liverpool	phaitun@liverpool.ac.uk
Mark Caddick	University of Liverpool	caddick@liv.ac.uk
Ka Wai Chan	Source BioScience	kawai.chan@sourcebio-science.com
Xianmin Chang	Liverpool University	changxianmin2002@yahoo.co.uk

James William Cooper	University of Leeds	bs08j2c@leeds.ac.uk
Dr Andrew Charles Cuming	University of Leeds	a.c.cuming@leeds.ac.uk
Linda Damore	University of liverpool	linda.damore@liv.ac.uk
Jack Davies	University of Liverpool, Institute of Integrative Biology	jack.davies@liverpool.ac.uk
Louisa Violet Dever	University of Liverpool	l.dever@liv.ac.uk
Askim Hediye Sekmen Esen	Ege University	hediye.sekmen@ege.edu.tr
Peter Etchells	University of Manchester - Faculty of Life Sciences	Peter.Etchells@manchester.ac.uk
Brian Forde	Lancaster University	b.g.forde@lancaster.ac.uk
Lorenzo Frigerio	University of Warwick	l.frigerio@warwick.ac.uk
Dr. Patrick Gallois	University of Manchester	patrick.gallois@manchester.ac.uk
Laura Gardiner	University of liverpool	lgardine@liv.ac.uk
Neil Graham	University of Nottingham	neil.graham@nottingham.ac.uk
Murray Grant	University of Exeter	M.R.Grant@exeter.ac.uk
Claire Grierson	University of Bristol	claire.grierson@bris.ac.uk
Dr. James Hartwell	University of Liverpool	hartwell@liv.ac.uk
Timothy Hearn	University of Cambridge	tjh70@cam.ac.uk
Lars Hennig	Swedish University of Agricultural Sciences	lars.hennig@slu.se
Christine Hicks	University of Warwick	christine.hicks@warwick.ac.uk
Elaine Howell	University of Birmingham	e.c.howell@bham.ac.uk
Jo Hulsmans	University of Warwick	j.r.m.hulsmans@warwick.ac.uk
Dr Lee Hunt	University of Sheffield	l.hunt@sheffield.ac.uk
Charlotte Hurst	Durham University	c.h.hurst@dur.ac.uk
Stephen Jackson	University of Warwick	stephen.jackson@warwick.ac.uk

Stephanie Johnson	Durham University	stephanie.johnson@durham.ac.uk
Alex Jones	The Sainsbury Laboratory	alex.jones@tsl.ac.uk
Dr Yasuko Kamisugi	Universitry of Leeds	bmbyk@leeds.ac.uk
Lalit L. Kharbikar	Harper Adams University College	00701340@harper-adams.ac.uk
Heather Knight	Durham University	p.h.knight@durham.ac.uk
Professor Marc Knight	Durham University	m.r.knight@durham.ac.uk
Fiona Lahive	University of Reading	f.m.lahive@pgr.reading.ac.uk
Dr Lindsey Leach	University of Birmingham	l.j.leach@bham.ac.uk
Frederikke Malinosky	Copenhagen University, Department of Plant and Environmental Sciences	froxilla@hotmail.com
Galebotse Mathengwane	The University of Manchester	galebotse.mathengwane@postgrad.manchester.ac.uk
Fergus Meade	University of York	fergus.meade@york.ac.uk
Jacqueline Monaghan	The Sainsbury Laboratory	jacqueline.monaghan@tsl.ac.uk
Karen Moore	University of Exeter	kamoore@exeter.ac.uk
Shima Nazrl	University of Nottingham	stxsn11@nottingham.ac.uk
Kim Osman	University of Birmingham	k.osman@bham.ac.uk
Geraint Parry	University of Liverpool	geraint.parry@liverpool.ac.uk
Jon Pittman	University of Manchester	jon.pittman@manchester.ac.uk
Mr. Kanok Preativatanyou	Institute of Integrative Biology, University of Liverpool	K.Preativatanyou@liv.ac.uk
Dr Mags Pullen	University of Durham,	m.l.pullen@durham.ac.uk
Mark Quinton-Tulloch	University of Liverpool	mark.quinton-tulloch@liverpool.ac.uk
Deborah Rathbone	Biorenewables Development Centre	debsorah.rathbone@york.ac.uk

Michael R Roberts	Lancaster University	m.r.roberts@lancaster.ac.uk
Federico Sabbadin	University of York (Department of Biology)	fs525@york.ac.uk
Professor Julie Scholes	University of Sheffield	j.scholes@sheffield.ac.uk
Daniel Shaw	University of Leeds	bs08d2s@leeds.ac.uk
Minu sherry	Coventry University	minusherry@yahoo.com
Dr Hamad Siddiqui	University of Warwick	h.siddiqui@warwick.ac.uk
Ambra De Simone	University of Leeds	bsads@leeds.ac.uk
Sandra Smieszek	RHUL	sandra@cs.rhul.ac.uk
Tom Smith	University of York	T_O_M_85@hotmail.com
Rebecca Stafford	York University	rss516@york.ac.uk
Sean Stevenson	University of Leeds	seenstevo@googlemail.com
Monisha Sundar	University of Manchester	moni.sundar@gmail.com
Heather Sutton	University of Liverpool	bs0u8123@student.liv.ac.uk
Dr Jane E Taylor	Lancaster University	j.e.taylor@lancaster.ac.uk
Dr Elinor Thompson	University of Greenwich	te30@gre.ac.uk
Gabriela Toledo	University of Edinburgh	gtoledo@staffmail.ed.ac.uk
Joseph Vaughan	University of York	jv506@york.ac.uk
Edward Venison	Centre for Plant Integrative Biology / University of Nottingham	edward.venison@nottingham.ac.uk
Dr Rachel Wade	Wiley	rwade@wiley.com
Jade Louise Waller	University of Liverpool	bs0u8178@liv.ac.uk
Ying Wang	The University of Warwick	Y.Wang.2@warwick.ac.uk
Lorraine Williams	University of Southampton	lew@soton.ac.uk
Mark Owen Winfield	University of Bristol	mark.winfield@bristol.ac.uk
Thilo Winzer	University of York	thilo.winzer@york.ac.uk
Jianhua Yang	University of birmingham	j.yang.6@bham.ac.uk