The ARC Biotechnology Platform: Role in the ARC and elsewhere...

Dirk Swanevelder
SwanevelderD@arc.agric.za
History of the ARC-BTP

• 2010: Task team investigates ARC’s biotechnology capacity across the ARC’s 11 Institutes
  • Recommendation: A new single integrated structure for advanced high throughput, next generation technologies

• 2010: Dr Jasper Rees to implement recommendations

• 2010: ARC’s Biotechnology Platform established
  • Housed on the ARC’s Onderstepoort Veterinary Institute Campus, North of Pretoria
  • Utilize ARC-OVI’s support services

• September 2010: First staff join ARC-BTP

• January 2011: Operational Start
Vision

To create a world class biotechnology platform to lead research in agriculture in Africa
Mission Statement

• To provide excellent biotechnology research and service platforms to enhance food security and promote agriculture in Africa
• To provide biotechnology tools in support of all areas of research and development within the mission of the ARC
• To create a world class research and innovation environment to attract the best researchers and students
• To train the next generation of researchers in cutting edge biotechnologies
Service & Research Facility Model

Unit

Technology Focus of Unit, e.g. Genomics

Core Services

Key technologies implemented as services to research groups

Development Group

Development group responsible for introducing new technology and applications into the core services

Internal Research Teams

Teams lead by BTP scientists, with Postgrad students as the major part of each team. Primarily externally funded with competitive grants

External Research Teams

External research teams
- ARC institutes
- External partners academic partners, e.g. Universities, etc.
- Commercial partners
Technology Focused Units

- **Genomics**: DNA isolation, Next Generation Sequencing, Genetic Mapping, Genotyping, etc.
- **Proteomics**: Sample Preparation, 2D PAGE, MALDI-MS, LC/GC-MS-MS, etc.
- **Bioinformatics**: High-performance computing, Databases, Annotation, Assembly, Expression Analysis, etc.
- **Marker Assisted Breeding**: Robotic platform, High throughput DNA Isolation, Genotyping, Informatics, Mutation Breeding, etc.
- **Plant Phenotyping**: Automated, High throughput imaging and sensing, Mobile and greenhouse based, Informatics, etc.
- **Functional Genomics**: Genome Engineering, Plant Transformation, Tissue Culture, Construct Development, Gene Silencing, Zinc Finger Nucleases, Gene targeting, Male Sterile Technology, Targeted Mutagenesis, Reverse Genetics, Gene identification, Vector Development, etc.
Impact on Agricultural Research and Development

• Faster, better, more selective:
  – Plant Breeding
  – Animal Breeding
  – Cultivar identification (PBR)
  – Pathogen identification
  – Diagnostics
  – Vaccine development
  – Vaccine quality control and safety

• Bringing better crops and animals to farmer at all scales
• Bringing novel, cash generating, sustainable products to the market
• Increasing food and income security
ARC Biotech Platform May 2011
Renovation: 1 000 m²
Occupants: > 40 Researchers
Investment in Scientific Tools

• ARC is investing in advanced, cutting edge Biotech technologies
  • Accepting that this has a short life-span
  • Continual investments required for fast changing field

• ARC aims to create a national & international platform that is shared and fully utilized
Current Technologies @ ARC-BTP

- Next Generation Sequencing platforms
- Genotyping platform
- Automation
- Informatics
- Sample preparation
NGS Platforms @ ARC-BTP
NGS & Genotyping Platform: Illumina HiScanSQ

- Dual function system: SNPs & Sequencing
- High throughput SNP arrays
  - Infinium & GoldenGate technology
  - >1k SNP chips
- Next Generation Sequencing
  - 150 Gb of DNA sequencing per run
  - Paired-end sequencing
  - 1.5 Billion filtered PE clusters per FC
  - 8 lanes per FC (flow cell)
  - Multiplexing up to 96 samples per lane
  - 17.5 Gb per day (100bp PE run)
  - >80% bases @ Q30 for 100 PE run
  - Small genomes: >50X coverage of 96 samples of 4MB Bacterial genomes in ±5 days on a 50bp PE FC

March 2011
NGS Platform: Illumina MiSeq

- 15 Gb of DNA sequence per run
- 25 Gb in late 2014
- Paired-end sequencing
- 300 bp paired-end data
- 40 M filtered clusters per FC
- Multiplexing up to 384 samples per FC

August 2012
NGS Platform: Illumina HiSeq2500

- Dual function system
  - Rapid mode & Standard mode sequencing

- Rapid Mode
  - 150 Gb of DNA sequence in 27 hours
  - 1000 Gb of DNA sequence in 6 days

- Standard Mode
  - 600 Gb DNA sequence in 11 days
  - Paired-end sequencing
  - 1800 M filtered clusters per FC
  - 8 lanes per FC x 2
  - Multiplexing up to 10 000 samples

- NRF Equipment Grant & ARC investment

Feb 2014
Illumina HiSeq2500 Platform

- Biggest DNA sequencer in Africa
- Data for 10 human genomes every 6 days
- Or Cow, Sheep, Chicken, Sunflower, wheat, maize
Automation @ ARC-BTP
Automation @ ARC-BTP
Automation Setup
Automation Setup

- 96 Pipette Head
- 96 Magnet
- 96 Vacuum
- Washable Needles
- Single Pipette
- Barcode Scanner
Informatics: High Performance Cluster & Storage Array
Informatics: High Performance Cluster & Storage Array

- LIMS Server
- LIMS Test Server
- Storage Array: Expandable to 1.2 Petabytes
- UPSs
- High Performance Cluster ±1 Terabyte RAM 13 Nodes
- Gigabyte Switch to SANREN
- 5X 96Gb RAM XServes 1x 48Gb RAM Xserve
- Storage
Informatics: Bioinformatics @ BTP

- **CLC Bio Enterprise Package: Genome Studio** for general use across ARC
- **NGS LIMS: GenoLogics**
- **High speed network access: SANREN**
Sample Preparation: Laser Capture Micro-Dissection Microscope
Sample Preparation: Laser Capture Micro-Dissection

Microtome
Cryostat
 Stereoscopic Microscope
LCM
Zeiss PALM 4: Laser Capture Micro-Dissection Microscope
Zeiss PALM 4: LCM

- Incubator
- Automated Sample Collector
- Touch Screen for Cutting
- Anti-vibration Table
- Fluorescence Cubes
- Screen Pen
What can we do with these technologies?
Genomics Applications

**Sequencing applications**
- Genomes
  - *De novo* and re-sequencing
  - SNPs and CNV calling
- Transcriptomes
  - *De novo* and re-sequencing
  - SNPs and splicing variation
  - Expression profiling
- Small RNA
  - discovery and expression analysis

**SNP applications**
- GWAS
- Association Genetics
- Cultivar and Breed ID
- Genomic Selection
- SNP validation
- Candidate Genes
- Diversity studies
- Methylation
ARC Genomics for Plant Breeding

• Development of new genomic resources
  – Amaranthus, apple, guava, broad bean

• Use of existing genomes
  – Sorghum, cassava, sunflower, wheat, maize, etc

• Use of SNP chips for genotyping and genetics
  – Apple, potato, maize, bean

• Development of Genotype by sequencing
  – Sorghum, wheat, maize, sweetpotato, potato
Genomics for Plant Pathogens

• Fungi
  – Apple scab, powdery mildews, guava wilt
  – Wheat, maize, sorghum, potato pathogens

• Bacteria
  – Bacterial blight in grape, potato scab

• Viruses
  – Grape, apple, sweet potato, potato, citrus
ARC Genomics Unit – Genotyping

• Major genotyping experiments
  • Apple – (ARC) 550 samples
  • Apple (RosBREED - USA) – 1012 samples
  • Peach (ARC) – 192 samples
  • Cattle (ARC) – 96 samples
  • Chicken (ARC) 300 samples
Genetic Mapping with SNPs: Methodology

DNA Isolation from leaves

10k, 50k, 500k Infinium SNP chip

Scan on HiScanSQ

QTL analysis with MapQTL™ 6.0

Draw genetic maps with JoinMap™ 4.1

Analyse and cluster with Genomestudio™
Some projects @ BTP
Sunflower Genomics

- 10 Genomes re-sequenced at 10x coverage by ARC
- Part of the 288 genomes international consortium
- Transfer of draft sunflower genome to ARC – 2011
- Access to 5-10 x coverage of each of 288 genotypes (>7 Terabases)
- ±1800 sunflower genotypes screened for diversity studies at ARC
- Will provide for diversity analysis and gene variation mining
Single Chromosome Isolation

• Isolation of wheat chromosome 7DS and 7DL translocation events from metaphase spreads
• Performed at ARC-BTP in September 2013
• University of Stellenbosch, Prof A-M Oberholster
Turning Waste into Value: Crude glycerol conversion into poly-lysine by *Streptomyces albulus*: A Genomic Perspective

Amanda Dodd
Dr Karl Rumbold
Dr Dirk Swanevelder

University of the Witwatersrand
School of Molecular and Cell Biology
What do we want to achieve?

• Increase biodiesel derived glycerol uptake and poly-lysine production by *S. albulus*
• No genome sequence
• No transcriptome sequence
• Poly-L production pathway known, rest unsure
• Genome size estimated <10 MB
• Original solution: Real time RT-PCRs of selected genes
Draft Genome Sequence of *Streptomyces albulus* Strain CCRC 11814, an ε-Poly-1-Lysine-Producing Actinomycete

Amanda Dodd, Dirk Swanevelder, Jonathan Featherston, Karl Rumbold

School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, South Africa; Biotechnology Platform, Agricultural Research Council, Onderstepoort, South Africa

Here, we report the draft genome sequence of *Streptomyces albulus* strain CCRC 11814, a soil-dwelling, Gram-positive bacterium. *S. albulus* produces ε-poly-1-lysine, which has diverse antimicrobial activity. The genome is 9.43 Mb in size, with a G+C content of 72.2%, and contains 9,177 protein-coding sequences.

Received 7 August 2013  Accepted 7 August 2013  Published 5 September 2013


Copyright © 2013 Dodd et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to Karl Rumbold, karl.rumbold@wits.ac.za.
Using molecular barcodes to investigate pollinator-plant interactions in long-tongued bees

Annemarie Gous

Supervised by

Dr. Dirk Swanevelder (ARC BTP)
Dr. Connal Eardley (ARC PPRI)
Dr. Sandi Willows-Munro (UKZN)
Pollinator-plant Interactions

- Historically studied by observation
- Molecular methods - Barcode of Life Initiative
  - DNA barcodes identify species
  - $rbcL$ and $matK$ recommended for plants
  - ITS increasingly used in plant identification
- Single molecular study on bee-plant interaction
  - Sanger sequencing
Some Findings

• First time using NGS to determine origin of pollen from bees
•Successfully sequenced DNA from pollen dating back to 1900s
• ITS analyses identified fungi associated with the collection
• Universal ITS could be used for screening fungi and bacteria associated with pollinators
NGS in microbial projects...
Why a NGS Approach?

- Majority of microbial species ≠ culture
- Culturing introduces species biases
  - Carbon sources
  - Species growth rates
- Cost-effective @ <R4 per MB on MiSeq platform
- High-throughput, no species separation before NGS
- Automated
Illumina Sequencing Technology

DNA (0.1-1.0 ug)

Sample preparation

Cluster growth

Sequencing

Image acquisition

Base calling

5’ 3’
What does NGSing look like on an Illumina Platform?
Single, Paired-end & Mate Pairs Reads

Single read

100bp

Paired-end

300bp

300bp

Long-insert (mate pair)

50bp

insert size 2 - 5 kb

50bp

From Illumina
Sequencing Strategies

Sanger Sequencing

- Gene of interest, e.g. 16S rRNA
- Conserve regions
- Variable regions
- ~1.5Kb
- ~700 bp
- X1 species
- Contig

- Longer reads
- Species must be separated for each sequencing reaction (DGGE, cloning, etc)
- Low throughput, time consuming, expensive
- Only forward and reverse reads for each fragment
Illumina 16S Metagenomics Protocol

• Prepare 16S rRNA gene amplicons for MiSeq
• Targets V3-V4 region of 16S rRNA
  • ±460bp region
• 5’ overhangs on gene selective primers for multiplexing
• Dual PCR indexes (Nextera) for multiplexing
Illumina 16S Metagenomics Protocol

16S Forward primer with tail

Gene of interest, e.g. 16S rRNA

~1.5Kb

Reverse Index & adaptor P7

Forward Index & adaptor P5

Selective PCR

Dual indexing

Dual indexed with adapters

Forward Read

NGS, PE 300bp

Reverse Read
Illumina 16S Metagenomics Protocol

• MiSeq 300 bp PE sequencing
  • 96 samples per run
  • >100,000 reads per sample
• Use MiSeq Reporter or BaseSpace for metagenomics analyses
  • Greengenes rRNA reference database
NGS Approach

- High sample multiplexing allows more samples per sample prep
- Deep sequencing allows more IDs
- Automation and cost-effective
- Adaptive to different environments and species
- Subgroups ID with specific gene primers allows testing for specific microbes, e.g. nitrogen fixers
Costs ...

- MiSeq ±R3 700 per Gb
- HiSeq2500 ±R1 700 per Gb
- 16S Metagenomics ±R1 400 per sample

More Information

BTP-Core@arc.agric.za
Acknowledgements

- **WITS**
  - Karl Rumbold
  - Amanda Dodd
- **SU**
  - Anandi Bierman
  - Prof Anna-Maria Oberholster
- **ARC-BTP**
  - Jasper Rees
  - Farai Muchadeyi
  - Jonathan Featherston
  - Marija Kvas
  - Kerry-Anne Pillay
  - Jurgens de Bruin
  - Annie Gous
- **Illumina**
  - Zeiss South Africa
- **Funders**
  - NRF
  - THRIP
  - Industry
“Sequenced Genomes Taste Better!”  
Jasper Rees

Dr Dirk Swanevelder
SwanevelderD@arc.agric.za