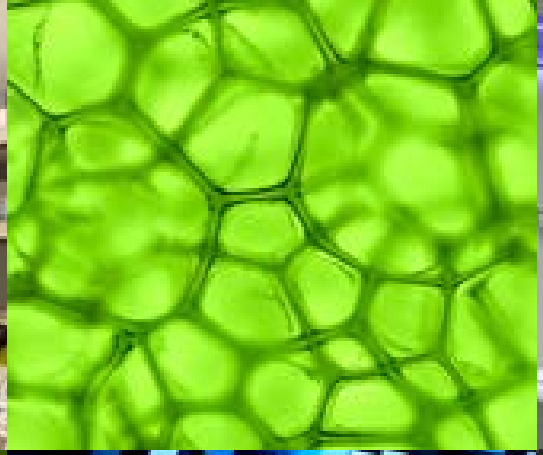
Annual Report for the year ended
31 March 2015

TGAC
The Genome Analysis Centre™
BBSRC
Greater Norwich Development Partnership

norwich research park

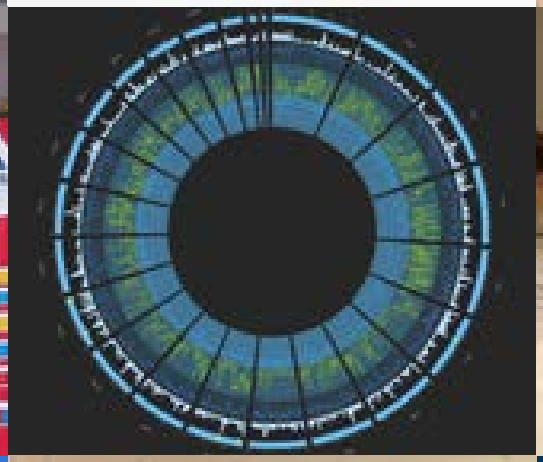
Registered company number: **06855533**

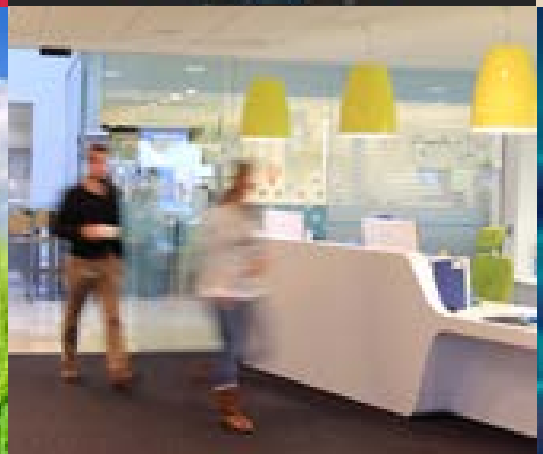










START
FINISH



Contents

The Genome Analysis Centre ("TGAC") is a company limited by guarantee and a registered charity.

The Annual Report provides information on the legal purposes of the charity, the activities it undertakes and its main achievements. The Trustees' Report and Financial Statements that can be found at www.tgac.ac.uk have been prepared in accordance with the Statement of Recommended Practice: Accounting and Reporting by Charities ("SORP 2005"), applicable United Kingdom accounting standards, the Charities Act 2011 and the Companies Act 2006.

Photo Credits

P2: TGAC, BBSRC; P4-5: Norwich Research Park; P6: Bread by veganbaking.net; P7, 50: Ash Tree by Kate Nicol; P27, 57 Black footed ferret by Graham Etherington; P35: Yellow rust, Andy Davis, John Innes Centre; P41: Optalysys; P55: Yellow rust spores, Kim Findlay, John Innes Centre; P57: Rust-spores, Tolga Bozkurt; P59: Albugo candida, Agathe Jouet.

Acknowledgements

Design by TGAC Business Development & Communications Group (marcomms@tgac.ac.uk)
Printed by Swallowtail Print

For a digital copy of this document please visit: http://tgac.ac.uk/annualreport

Chair's Message

This has been another exciting year to chair TGAC's Board of Trustee Directors. The Institute has continued to grow in staff numbers and diversify its science. We have continued to contribute to major national and international initiatives through our work in bioinformatics and genomics, and to provide training to other scientists and science education to the public. We seek to use our science to generate economic benefit directly through our commercial arm, Genome Enterprises Ltd, and indirectly through open publication of our science.

Of course, increasing the size and activity of TGAC gives increased complexity and we have restructured the Institute and made senior appointments accordingly. Sarah Cossey and Professor Federica Di Palma were appointed as Director of Operations and Science respectively during the year. Stuart Catchpole as Head of Business Development & Communications, Dr Tim Stitt as Head of Scientific Computing and Dr Dan Swan as Head of Pipelines and Platforms. Dr Ksenia Krasileva has joined us as Group Leader in wheat genomics jointly with The Sainsbury Laboratory. To help in our training activities, we have opened two state-of-the-art training suites.

In March 2015, we had a small fire in the plant room of one of our main buildings. Everyone involved was tremendously impressed with the way TGAC and the support services on the Norwich Research Park responded to the incident, which tested our business continuity planning and processes. The response demonstrated how well the teams worked together, and the impact on our business was limited to days. We also took this incident as an opportunity to make some upgrades to facilities and support services.

Sir John Sulston came to the end of his term as a Board member in December 2014. We thank John for his commitment to the Institute and for his wise counsel on both scientific and operational matters. His expertise will be missed, but we have been fortunate in welcoming to the Board Professor Dame Janet Thornton, who, until July 2015, was Director of the European Bioinformatics Institute.

At the end of July 2015, we said goodbye to Professor Mario Caccamo, who stepped down as Director. Professor Neil Hall was appointed as the new Director in December 2015 and starts his post in April 2016. We are pleased that Professor Dylan Edwards has joined us as a part-time interim Director during the interregnum. I am confident that TGAC will continue to have a major impact in delivering new research in bioinformatics and genomics.

Professor Nigel L Brown OBE FRSE FRSB FRSC
TGAC Chair of Trustee Directors
15 December 2015

Executive Summary

In 2014-15, we saw the Institute continue its growth from the previous twelve months, with the development of a new science strategy and key appointments being made to ensure TGAC has the right skills and expertise to take us forward. By the end of the year, all senior posts in the Operational Division were appointed, as well as two new Group Leaders in the Science Faculty.

We continued with an ambitious scientific learning and education programme in computational bioinformatics, and much success in knowledge exchange supporting our science groups with budding new industrial partnerships. We also had a number of key public engagement activities, sharing our science and learning from our wider stakeholders.

In line with our plans for growth of the Institute, we developed our facilities with over £1.5M of capital investment into a new laboratory; TGAC entrance and reception; two new training suites; new meeting room facilities; refurbishment of offices; kitchen and toilet facilities. We are ambitious with our plans to make TGAC's working environment for staff and visitors inspiring to match the Institute's vision and culture.

The existing Executive Team of Professor Dylan Edwards (Interim Director), Sarah Cossey (Director of Operations), and Professor Federica Di Palma (Director of Science), together with a widened Senior Management Team will lead the Institute during the transition between Directors in 2015-16.

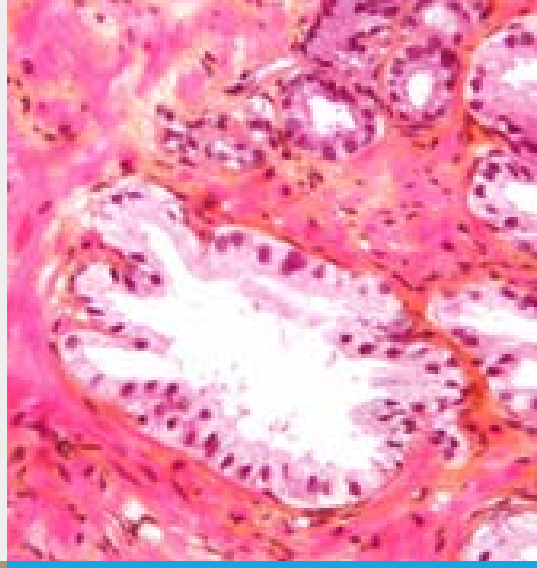
We have some major programmes planned for 2015-16. With the launch of the science strategy and review of our brand and communications, including a new website. We are excited by what more TGAC can deliver in terms of National Capability in Genomics, alongside an ambitious and expanded research programme. We hope you find this report highlighting the new science strategy, our group leaders, and performance in 2014-15 of interest.

Dylan Edwards, Sarah Cossey, Federica Di Palma, Executive team



126

Publications to date

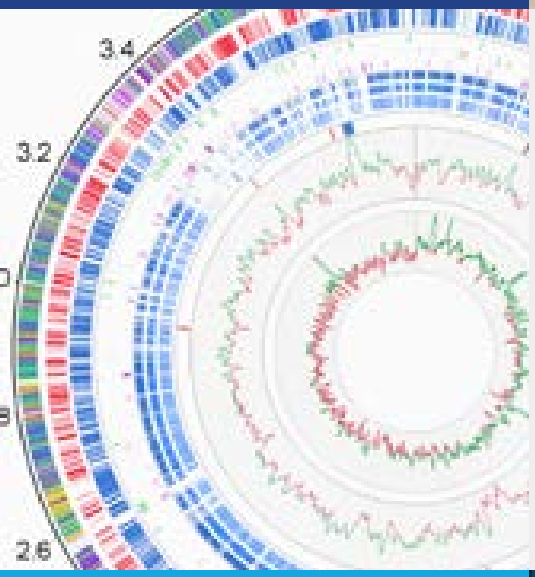


3960+

Citations of TGAC scientific papers to date

18

Visiting workers in year, 4 PhD students, 2 Year in Industry students

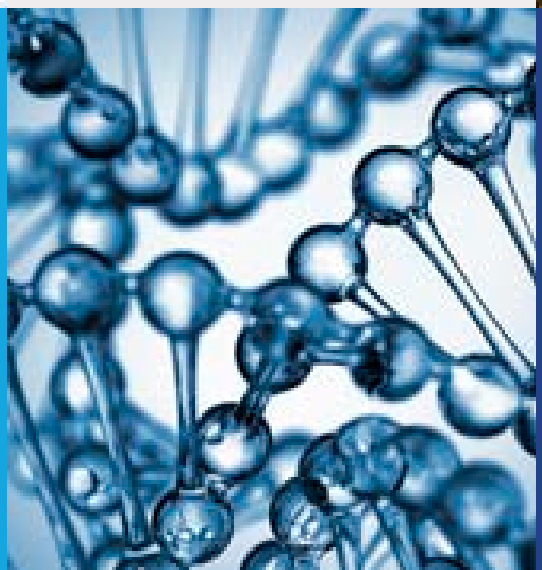


104

Employees at end of March 2015

31

Active Competitive Grants



78%

Of staff said it was a great place to work

260

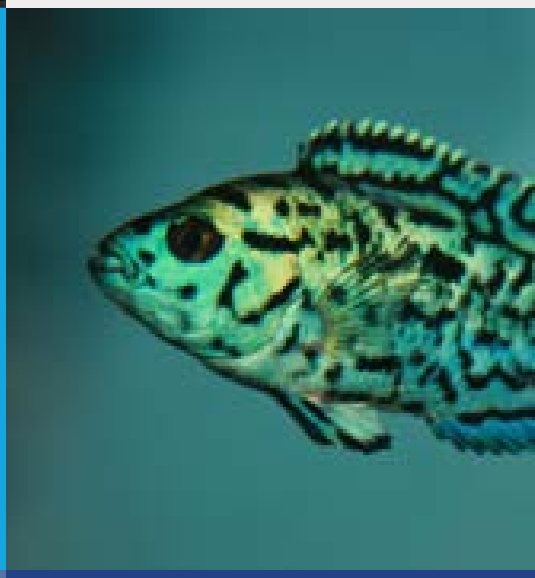
Delegates trained in computational biology

50%

Success rate for grants (per cent of total value of grant applications submitted)

1638

People reached via our outreach programme

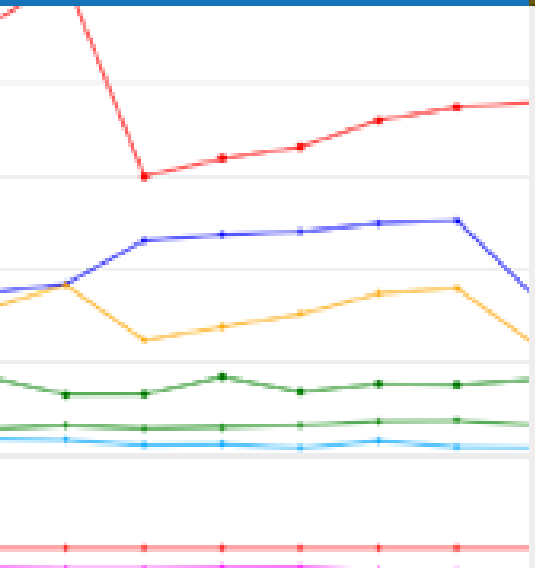


£15.4M

Income in 2014/15

£5.2M

Invested in capital



£12.8M

Expenditure

STRATEGIC REPORT

"It has been a huge thrill for me since working with TGAC, learning about the truly game-changing work that is being done. The Institute has really hit its stride and the results coming through are vastly impressive.

Dylan Edwards
TGAC Interim Director

Mission

TGAC's mission in 2014-15 was to advance bioscience by enabling and developing computational and genomics data-driven approaches in biology. TGAC sought to build excellence in genomics and computational bioscience and to develop as a world-leading centre in bioinformatics and applied biotechnology.

Strategy

TGAC has developed to be one of the leading UK research and innovation research centres specialised in genomics and computational biology applied to plant, animal and microbial science. The strategy has relied on the implementation of novel applications, establishing collaborative work and the development of skills to advance knowledge and promote the growth of the UK bio economy.

Looking Ahead

TGAC's objectives for 2015 -16 are:

i. Deliver our existing BBSRC strategically funded Programme in Bioinformatics, which runs to March 2017, and to establish new scientific programmes that will build collaboratively on the foundation it creates.

ii. Embed a new scientific strategy to; increase our knowledge base in bioscience, improve plant, animal and human health and to enable research to empower both academia, and industry with new technologies and scalable bioinformatics approaches.

iii. Establish the National Capability in Genomics as the UK's leading provider of genome analysis.

iv. Build our Knowledge Exchange and Commercialisation interactions with Industry, and increase the translation of our science into commercial usage.

v. To further develop TGAC's long term science strategy, and operational support reflecting on key stakeholder and public opinion.

vi. Deliver a leading training programme in Computational Biology.

vii. Develop TGAC as an ELIXIR Centre of Excellence.

viii. To establish a National Capability in DNA Synthetic engineering.

Culture

TGAC's development has been founded on a culture defined by the following values:

- Openness – TGAC promotes the dissemination of data and distribution of software code by following data-sharing policies that are embedded in all the research programmes.
- Technical excellence – As a national capability, TGAC is committed to test, evaluate and offer access to cutting-edge technologies, staff training and protocol development.
- Skilled personnel – TGAC pursues excellence at all levels of operation, reflected in its Strategic Human Resources programme.
- Innovation – Access to novel technologies and state-of-the-art hardware platforms provides the foundation for instigating novel solutions and innovative science.

Public Benefit

TGAC aims to advance biological and biotechnological science for the public benefit by undertaking and promoting research relating to genomes and their functions, in particular by carrying out the following activities:

i. Generating and analysing genome sequences and other 'omics datasets, for a whole range of different organisms and depositing them in public databases.

ii. Researching how genomic function and variation may be exploited for the public benefit.

iii. Developing and exploiting technologies to identify and measure genomes and genomic products.

iv. Developing bioinformatics for analysing, interpreting and exploiting large-scale 'omics data.

v. Training those engaged in or associated with genome biology, bioinformatics and related subjects in new technologies and data analysis methods.

vi. Engaging with the general public to educate them about these activities and listening to their views about exploiting advances in genomics.

OUR PERFORMANCE

"This new package of investments will ensure that the UK maintains this leadership position and continues to drive the potential of synthetic biology to contribute to the economy and society.

BBSRC Chief Executive, Jackie Hunter, speaking about £40M investment in UK synthetic biology, TGAC to receive £1.9M.

2014 - 15 Objectives

TGAC's principal objectives for the year ended 31 March 2015 were to:

i. Deliver the BBSRC-funded programmes, including: the Institute Strategic Programme in Data-Driven Science (Bioinformatics); the National Capability in Genomics; Knowledge Exchange & Commercialisation; Strategic Human Resources; and Public Engagement.

ii. To further develop TGAC's long- term science strategy, and operational support reflecting on key stakeholder and public opinion.

iii. Develop the research programme focusing on Health and Vertebrate Biology.

iv. Deliver a leading training programme in Computational Biology.

v. Develop an ELIXIR Centre of Excellence.

vi. To implement new strategies for our Knowledge Exchange & Commercialisation, building relationships with industrial partners to maximise the impact from TGAC's science.

vii. To build a DNA Synthesis Pipeline on the Norwich Research Park.

We made good progress against our objectives and this annual report aims to highlight our key achievements.

Key Statistics

We carried out 388 projects with over 226 other organisations

We collaborated with organisations in over

22 countries

TGAC Structure

TGAC is led by the Interim Director Prof Dylan Edwards, and supported by a Director of Operations (Sarah Cossey) and a Director of Science (Prof Federica Di Palma), who together form the Executive team. In addition, a Senior Management Team comprised of the Head of the 361° Division (Dr Vicky Schneider); Head of Scientific Computing (Dr Tim Stitt); Head of Business Development and Communications (Stuart Catchpole); and Head of Platforms and Pipelines (Dr Daniel Swan), supported the Executive in 2014-15. Research within TGAC is undertaken by the Science Faculty, which comprises of seven group leaders, four project leaders and currently two fellows with a further four fellows under recruitment.

TGAC Governance

The Board of Trustees comprises of a Chair, and six Trustees. TGAC also has two corporate member organisations, the BBSRC, the UEA. Two Patron organisations, the Norfolk Local Authorities, and the Department for Business Innovation and Skills also support TGAC.

The Board has two sub committees, the Finance Resources and Audit Committee and the Remuneration Committee. (http://www.tgac.ac.uk/governance-structure/)

TGAC'S Scientific Advisory Board

TGAC's Scientific Advisory Board is responsible for providing strategic advice to the Director of TGAC and the Board of Trustees on issues relevant to TGAC's Mission and Science programme. The membership is composed of internationally recognised scientists who meet once a year. (http://www.tgac.ac.uk/scientific-advisory-board/)


Research
Science Faculty
Technology Platforms
Operations
BBSRC
20 Years of Pioneering Great British Bioscience
National Capability
TGAC
The Genome Analysis Centre™
BBSRC
Greater Norwich Development Partnership
Training, Skills
361° Division
elixir

RESEARCH STRATEGY

Pursuing innovative approaches to high-impact science in an open, dynamic and cooperative environment.

Decoding Living Systems

TGAC's research brings together a wealth of expertise in biosciences, bioinformatics, high performance computing and statistics to understand complex biological systems in plants and animals and their interaction with the environment.

Our advanced genomics and computational platforms support data-intensive research that embraces and confronts modern scientific challenges arising from data scale and complexity. We develop and implement new technologies and apply computational methods to process, store and interpret data, to enable bioscience research.

Science Faculty

The Faculty collectively conducts three kinds of research activities:

- **Fundamental** research to increase our knowledge base in bioscience
- **Applied** research to improve plant, animal and human health
- **Enabling** research to empower both academia and industry with new technologies and scalable bioinformatics approaches

TGAC has three interdisciplinary programmes under which the Faculty, Fellows, postdocs and students aggregate to consolidate their expertise and to significantly increase our scientific impact.

Scientific Programmes

DB

Digital Biology

We focus on developing computational tools and infrastructure to push the boundaries of data-driven informatics in the life sciences and enable key stages of the data lifecycle: management, integration, and interpretation.

Projects include:

- The development of novel platforms for management of research data and software
- Implementation of computationally intensive algorithms for data quality assessment and assembly
- Design of large-scale data visualisation to improve user experience
- Developing and implementing best practice and training in bioinformatics

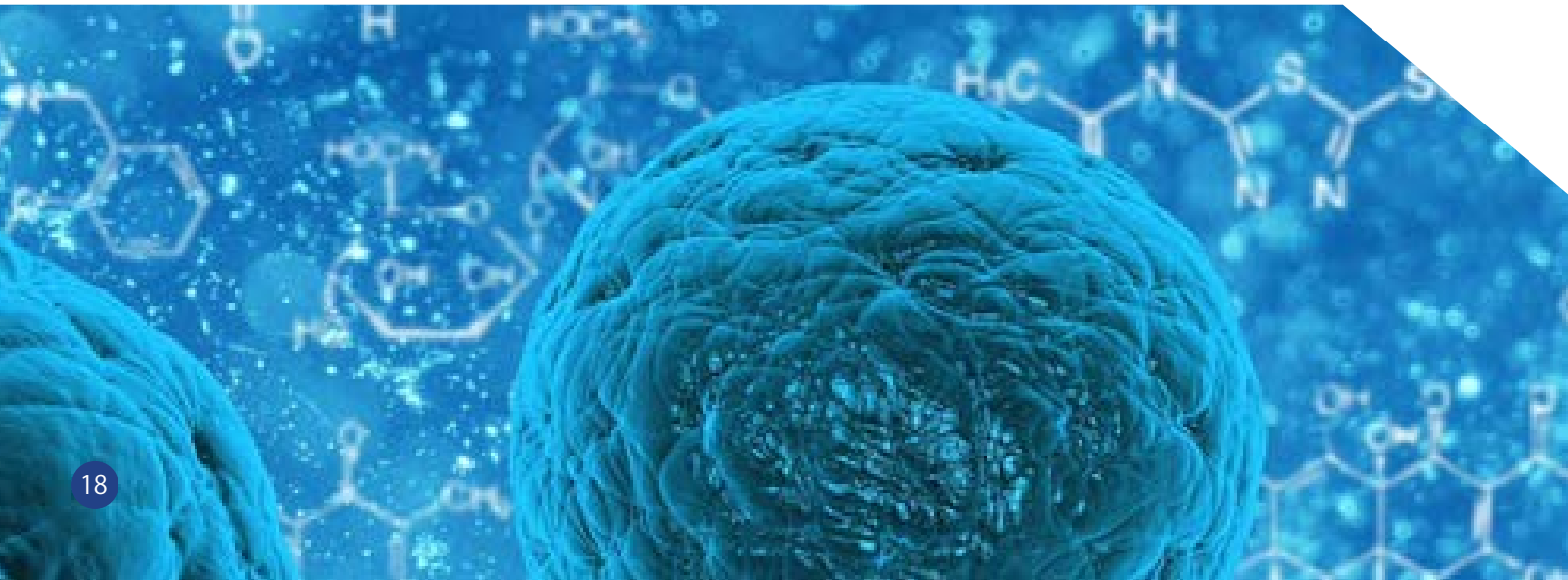

Jurkowski


Stitt


Davey


Schneider

Elixir



O&E

Organisms & Ecosystems

We use multidisciplinary approaches to understand genomes, and how they regulate and control the biological complexities of plant and animal functions.

Projects include:

- Understanding genetic diversity and its impact on traits of agronomic interest and evolution of vertebrate species
- Developing genomic tools and resources to enable plant breeders, research communities and genebanks
- Investigate fundamental principles of plant innate immunity and the interactions between plants and their microbial pathogens
- Developing and implementing multiscale integrative and network approaches to interpret diverse datasets, to identify pathways, to infer interactions between genes, proteins, lipids and metabolites and to understand genotype to phenotype relationships and system heterogeneity
- Developing and implementing effective computational methodologies for the assembly and annotation of animal and large complex plant genomes

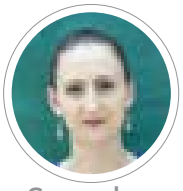

Di Palma


Clark


Swarbreck


Krasileva


Ayling


Saunders (Fellow)


Korcsmáros (Fellow)

EB

Engineering Biology

In this emerging programme our research includes the development of software engineering and machine learning approaches in support of projects in synthetic biology and high-throughput phenotyping.


Swan


Di Palma


Clark


Stitt

AYLING GROUP

Sarah Ayling
PhD in Bioinformatics from the University of Manchester
Started at TGAC in 2011

O&E

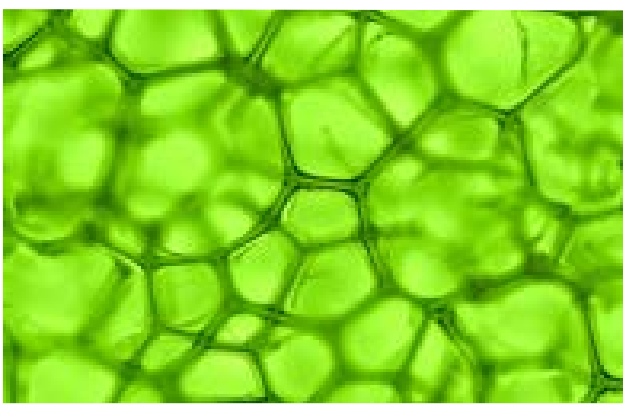
The Ayling Group focuses on supporting crop improvement through the application of genomics approaches. Developing and implementing assembly approaches for complex plant genomes; identifying and characterising genetic diversity and its effects for traits of agronomic interest; and developing genomic tools and resources for plant breeders, researchers and genebanks.

Project Highlights

Wheat TILLING-by-sequencing
As part of the BBSRC sLoLa "Tritceae genomics for sustainable agriculture," we have developed an exome capture for bread wheat in collaboration with JIC and UC Davis. Using this capture, we have sequenced hundreds of accessions from the Cadenza TILLING population developed at JIC/RRES and identified more than six million mutations. This data will be soon be publically available at www.wheat-tilling.com.

Barley BAC sequencing
As part of the BBSRC grant "UK draft barley genome," we have sequenced the Minimum Tile Paths (MTP) of two Morex chromosomes (2H and 0H) and assembled each BAC individually (~18,000 BACs). These assemblies are the UK's contribution to the International Barley Sequencing Consortium's MTP sequencing of the barley genome.

Phenomics
Through collaboration with JIC and IBERS, we are developing approaches to capture and analyse images and other high throughput digital data types in order to measure wheat growth and automatically identify developmental stages.



Selected Publications

I.W.G.S.C. **A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome**. *Science*, 345 (2014)

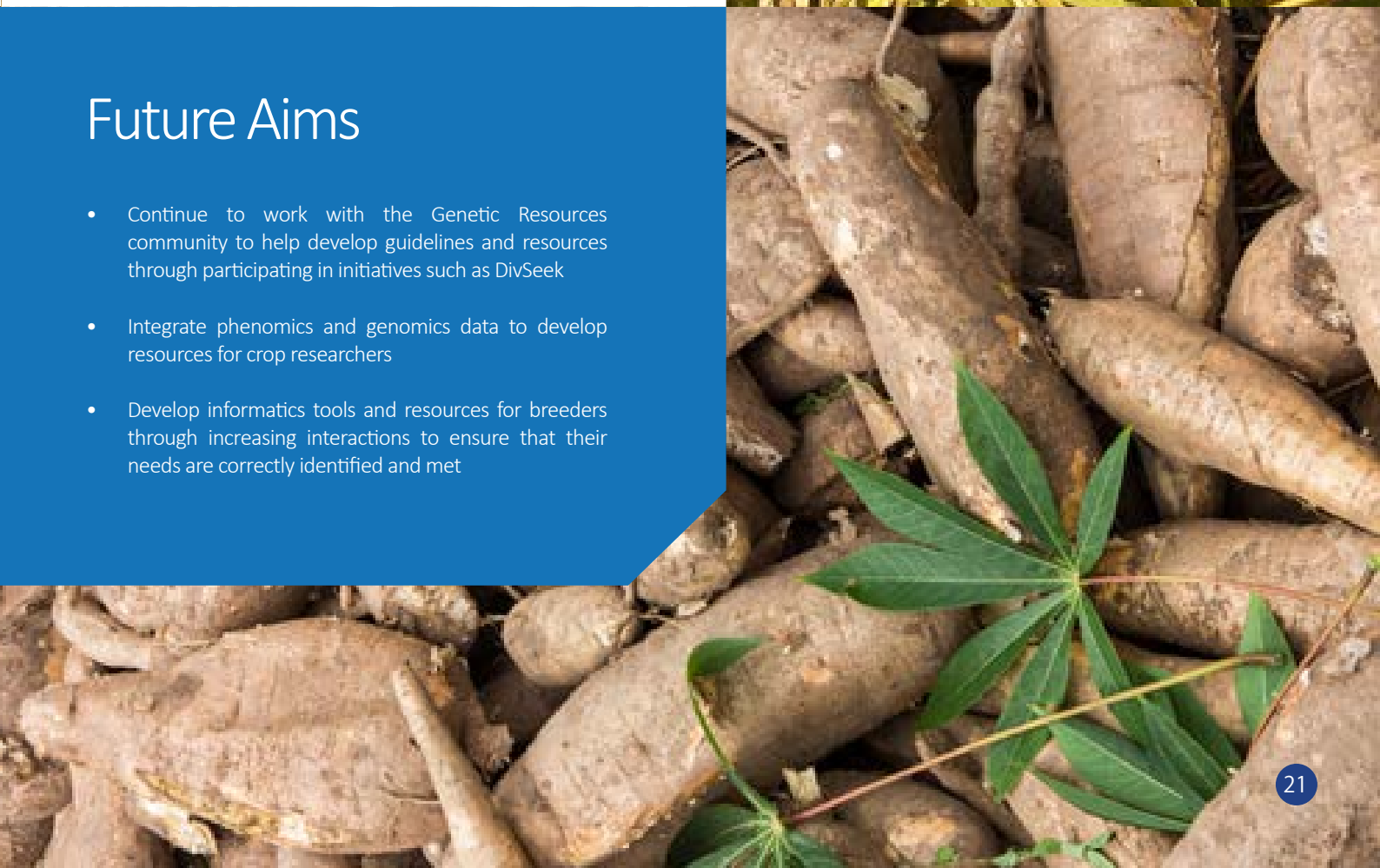
Cocker et al. **Oakleaf: an S locus-linked mutation of Primula vulgaris that affects leaf and flower development**. *New Phytol. (*2015)

Li et al. **Integration of genetic and physical maps of the Primula vulgaris S locus and localization by chromosome in situ hybridization.** *New Phytol.* (2015)

Ramamurthy et al. **Skeletonization of 3D Plant Point Cloud Using a Voxel Based Thinning Algorithm.** Accepted by IEEE EUSIP. EURASIP Journal in *Signal Processing* (2015) (http://www.eurasip.org/Proceedings/Eusipco/Eusipco2015/papers/1570096671.pdf)

Future Aims

- Continue to work with the Genetic Resources community to help develop guidelines and resources through participating in initiatives such as DivSeek
- Integrate phenomics and genomics data to develop resources for crop researchers
- Develop informatics tools and resources for breeders through increasing interactions to ensure that their needs are correctly identified and met



CLARK GROUP

Matt Clark
PhD, Max Planck Institute for Molecular Genetics
Started at TGAC in 2010

O&E DB

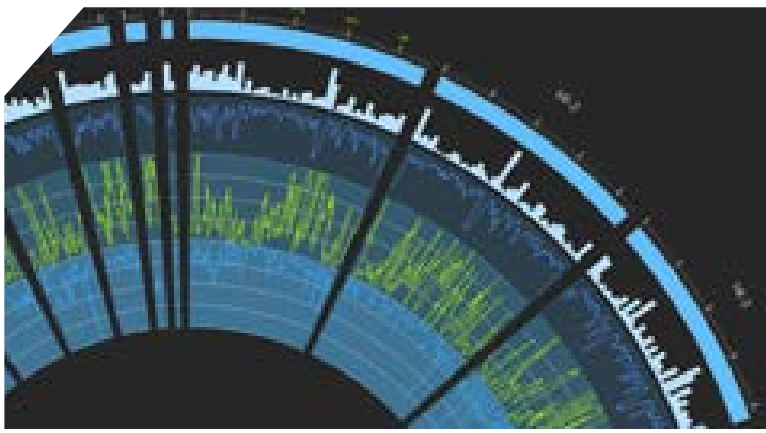
The Clark Group specialises in the application of genomic and genetic techniques to tackle a variety of biological questions in the fields of plant and microbial genomics. The team use a multidisciplinary approach, combining expertise in genomics, genetics, plant biology, cell biology and bioinformatics – with our strong technical background we regularly test and develop approaches based around novel techniques.

Project Highlights

Triticeae Genomics For Sustainable Agriculture
With external collaborators, we are developing genomic tools for Triticeae, especially the most complicated - bread wheat. Sequencing the bread wheat genome, and developing novel data types and tools to analyse it's complexity (it is 17Gb, 80+ per cent repetitive and hexaploid) – our approach will be used to understand, identify and harness natural variation in key cultivars and wild relatives, and contribute to global food security.

Solanaceae genomics
Potatoes and tomatoes are important crops, but they also serve as an excellent model for resistance to a wide range of pathogens – viral, bacterial, fungal, oomycetes, aphid and nematode. We are sequencing a wild relative of the potato *Solanum verrucosum* which is resistant to late blight, and is also a diploid which is self-compatible making it suitable for forward and reverse genetics. We are also using RenSeq (R-gene encrichment sequencing) to mine wild relatives of potato and tomato for novel resistance genes. Working closely with collaborators in The Sainsbury Laboratory, James Hutton Institute and Simplot, we are able to test these genes and generate commercial resistant varieties either by GMO or marker assisted breeding.

Nornex – Ash dieback consortium
TGAC is leading the ash tree genome sequencing and the UK pathogen samples. The Nornex consortium have sequenced the genome of a more resistant ash tree, as well as a range of trees that vary in resistance and have found an association to a resistance factor. By sequencing over fifty isolates from the UK, Europe and Asia, we have highlighted the small population that migrated to Europe and then the UK, and the likely orgins of the pathogen in Asia. We are now examining effector complements in detail to understand their effect on virulence, and genes that affect metabolism of fungicides and toxins.



Selected Publications

Heavens D., Garcia G., Clavijo B., Clark M. **A method to simultaneously construct up to 12 different sized Ilumina Nextera long mate pair libraries with reduced DNA, time and costs**. *Biotechniques* (2015) (in press)

Rallapalli G., Saunders D.G.O., Yoshida K., Edwards A., Lugo CA., Collin S., Clavijo B., Corpas M., Swarbreck D., Clark M., Downie AJ., Kamoun S., Team Cooper, MacLean D., **Lessons from Fraxinus, a crowd-sourced citizen science game in genomics**. *eLife*, 4 (2015)

McMullan M., Gardiner A., Bailey K., Kemen E., Ward BJ., Cevik V., Robert-Seilaniantz A., Schultz-Larsen T., Balmuth A., Holub E., Van Oosterhout C., Jones J DG. **Evidence for suppression of immunity as a driver for genomic introgressions and host range expansion in races of *Albugo candida*, a generalist parasite**. *eLife, 4* (2015)

Future Aims

- Plant-microbial interactions
- Crop Genomics
- Surveillance and diagnostics

DAVEY GROUP

Rob Davey
PhD Computer Science, IFR/JIC (awarded by UEA)
Started at TGAC in 2010

DB

The Davey Group focuses on research into understanding how best to manage, represent and analyse data for open science, as well as exploring new hardware, algorithms and methodologies to develop tools to push the boundaries of data-driven informatics in the life sciences. The team applies their research expertise to develop novel infrastructure platforms for data and software dissemination and publication, assembly algorithms for viral and microbial metagenomics, large-scale data visualisation, and best practice and training in bioinformatics.

Project Highlights

Collaborative Open Plant Omics (COPO)

The development of high-throughput genomics, transcriptomics, proteomics and metabolomics technologies has opened the way to new data-generative approaches to plant research. Alongside these large plant science datasets held in public and private laboratories around the globe, there are a large number of tools to help researchers disseminate, analyse and publish those datasets. However, the disparate nature of the tools, data formats and scientific problems has resulted in a lack of interoperable, production-quality software available for data analysis and dissemination. In collaboration with Warwick University, the Oxford eResearch Centre and the European Bioinformatics Institute, we are designing and developing the Collaborative Open Plant Omics (COPO) framework to address this disparity in interoperability and easy access to important description, deposition and publication services in the 'omics data realm.

Wheat Information System (WheatIS)

Under the umbrella of the Wheat Initiative, the WheatIS Expert Working Group (EWG) was set up in 2013 to build an international wheat information system, called WheatIS, with the aim of supporting the wheat research community. We are currently developing a Europe-wide federated system of data management "nodes" to provide single point-of-access web-based systems to available data resources and bioinformatics tools. It comprises a network of existing wheat data and analysis platforms (TGAC Grassroots Genomics (http://www.tgac.ac.uk/grassroots-genomics/) and Field Pathogenomics yellow rust projects, CerealsDB at Bristol, WheatIS node at INRA, France) that conform to a shared specification for integration to provide the wheat scientific community easy access to a large variety of wheat data and analysis.

iPlant UK

The iPlant Collaborative is a virtual organisation funded by the US National Science Foundation (NSF) to create cyberinfrastructure for the plant sciences. Harnessing the power of some of the world's fastest supercomputers, iPlant provides huge cloud-based storage space and a virtual lab bench, which put global plant science data and online tools in one place. Through BBSRC funding, we will extend iPlant into an international collaboration by building a UK iPlant node at TGAC in Norwich. As well as the existing iPlant US resources, software tools for specific plant science sequencing, systems biology and image analysis projects at the Universities of Warwick, Liverpool and Nottingham will be adapted by a dedicated team of programmers so that they can be integrated into iPlant UK. These will then be made freely and openly available for the wider plant science community to use.

Selected Publications

West C., James SA., Davey RP., Dicks J., Roberts IN. **Ribosomal DNA Sequence Heterogeneity Reflects Intraspecies Phylogenies and Predicts Genome Structure in Two Contrasting Yeast Species**. *Systematic Biology*, 63 (2014)

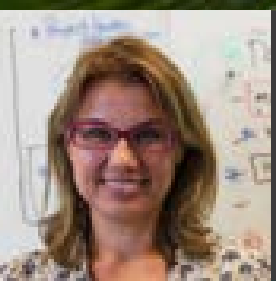
Aleksic J., Alexa A., Attwood T. K., Hong N. C., Dahlo M., Davey R.P., Dinkel H., Forstner K. U., Grigorov I., Heriche J. K., Lahti L., MacLean D., Markie M. L., Molloy J., Schneider M. V., Scott C., Smith-Unna R., Vieira B. M. **The Open Science Peer Review Oath**. *F1000 Research, 3* (2014)

Leggett RM., MacLean D. **Reference-free SNP detection: dealing with the data deluge**. *BMC Genomics,* 15 (2014)

Leggett RM., Heavens D., Caccamo M., Clark MD., and Davey RP. **NanoOK: Multi-reference alignment analysis of nanopore sequencing data, quality and error profiles**. *Bioinformatics* (first published online September 17 2015)

Future Aims

- Information Infrastructure and Semantic Integration
- Visualisation of non-model organism genomic datasets
- Exploiting new technologies for metagenomics sequencing and analysis



DI PALMA GROUP

Federica Di Palma
PhD University of Reading
Professor, School of Biological Sciences, UEA
Started at TGAC in 2014

O&E EB

Through a combination of comparative genomics, new technologies, novel computational approaches and fish and mouse engineering strategies, the Di Palma Group studies the impact of genome diversity (coding and non-coding) on the evolution of phenotypic traits with broad implications for animal and human health. As genome sequence information from vertebrate species increases, non-conventional model systems provide a unique opportunity to study the key molecular mechanisms underlying genotype to phenotype relationship.

Project Highlights

Cichlids – Developing the Cichlid Model for the Study of Vertebrate Genetic and Functional Diversity

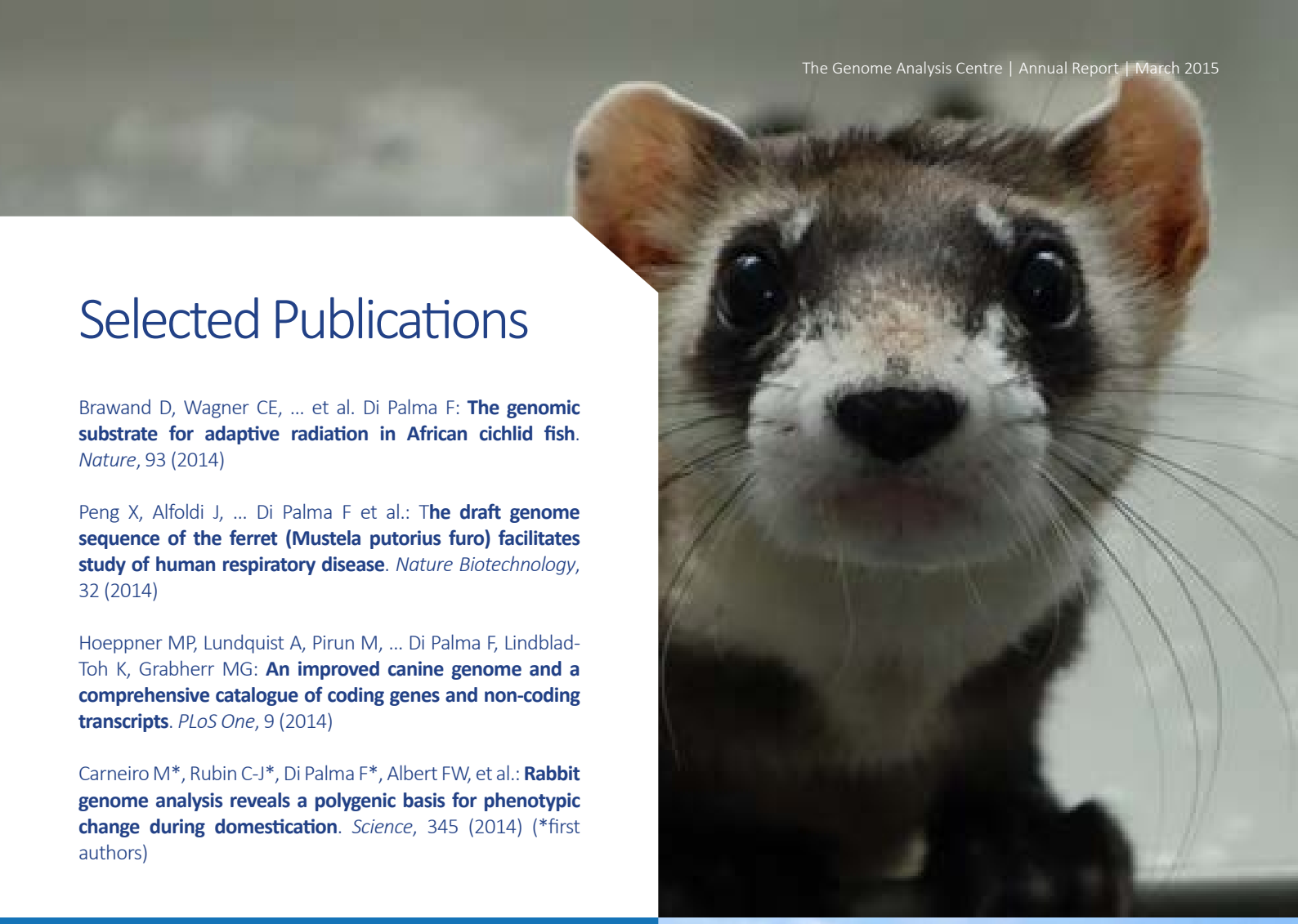
The rapid and extensive diversification of East African cichlids has allowed for the evolution of a tremendous range of phenotypes reflecting minimal genetic differentiation. Such a collection of species can thus be viewed as a collection of mutants, screened by natural selection for adaptive phenotypic differences. We use comparative functional genomics (mRNA levels, chromatin organisation, transcription factor occupancies) to study the evolution of tissue and species specific divergence in these fish. This approach, unravelling the genetic basis of functional diversification, is fundamental for our understanding of the origins of vertebrate diversity and can also have significant implications for animal and human health. One of our approaches is to infer essential drivers of tissue-specific regulation and evolution in cichlids by reconstructing regulatory networks underlying key cellular processes. We then test regulatory networks and their involvement in phenotypic divergence by genetic manipulation strategies e.g. CRISPR.

Swine Flu – Understanding influenza A virus; linking transmission, evolutionary dynamics, pathogenesis and immunity in pigs

Swine influenza viruses (SIV) are one of the major pathogens responsible for respiratory diseases in pig herds on a global scale. TGAC is involved in a BBSRC funded sLoLa project in collaboration with the Pirbright Institute to better understand the dynamics of these viruses, and economic advantages of vaccination for the UK swine industry. The project is examining the conditions and requirements for transmission of the virus between pigs, the effects of vaccination on transmission, and the potential for the development of novel vaccines to block transmission of the virus between pigs.

Functional Annotation of Farm Animal Genomes (FAANG)

Establishing the sequence of a genome is only the first step in understanding the complexities that give rise to phenotypes. Now that many vertebrate genome sequences are complete, we are engaged in improving the annotation of coding sequences, and in identifying the non-coding and regulatory sequences, which harbour most of the genetic variants associated with complex traits. The aim of this project is to enable the coordinated functional annotation of key farmed animal genomes by establishing the necessary data infrastructure. The infrastructure will comprise substantial hardware and compute capacity together with software to enable the use of the hardware for functional annotation of farmed animal genomes. This project brings together scientists in two leading UK institutions (TGAC and The Roslin Institute) with the EMBL- European Bioinformatics Institute.



Selected Publications

Brawand D, Wagner CE, ... et al. Di Palma F: **The genomic substrate for adaptive radiation in African cichlid fish**. *Nature*, 93 (2014)

Peng X, Alfoldi J, ... Di Palma F et al.: T**he draft genome sequence of the ferret (Mustela putorius furo) facilitates study of human respiratory disease**. *Nature Biotechnology*, 32 (2014)

Hoeppner MP, Lundquist A, Pirun M, ... Di Palma F, Lindblad-Toh K, Grabherr MG: **An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts**. *PLoS One*, 9 (2014)

Carneiro M*, Rubin C-J*, Di Palma F*, Albert FW, et al.: **Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication**. *Science*, 345 (2014) (*first authors)

Future Aims

- Determining the genetic and developmental mechanisms generating vertebrate diversity
- Understanding genotype to phenotype relationships to determine genes, pathways and mechanisms involved in animal and human health and disease

JURKOWSKI GROUP

Wiktor Jurkowski
PhD at Jagiellonian University in Krakow at the Faculty of Chemistry
Started at TGAC in 2014

DB

The Integrative Genomic Group led by Dr Wiktor Jurkowski, develops systems biology approaches to untangle the complexity of biological processes at the molecular, organ and organismal level. We focus on the bioactivity of natural compounds, their effects on homeostasis and the understanding of gastrointestinal tract microbiota spatial specificity and host-interactions. We are applying network analytical approaches to integrate and interpret genomics, transcriptomics and metabolomics data.

Method development for integration of transcriptomics and metabolomics data

Knowledge gathered from molecular networks could serve as a mean to break down the data complexity or define weights for predictive modelling. We recently published ONION, a tool that applies molecular interaction network topology analysis for data regularisation. This will improve the ability to rank biologically meaningful genes and metabolites and to discover novel associations between them. Currently, we are focusing on improving metabolite clustering and on network-based weights to define penalties in predictive models on multi-omics data sets.

Effect of diet intervention on health

We are identifying genes, proteins and metabolites to act as biomarkers and to generate new biological hypotheses relating to the effects of bioactive compounds in foods.
We are collaborating with Professor Richard Mithen's group at the Institute of Food Research (IFR), to understand the physiological differences between transition and peripheral zones, in the prostate and to study the protective effects of broccoli consumption in prostate cancer. This study will follow the changes in metabolism and gene expression in prostate tissue from men at risk of developing prostate cancer. We will determine how these changes are affected by a diet enriched with sulforaphane. We will develop an integrative strategy to identify associations between genes and metabolites and to study the potential health benefits of broccoli in prostate cancer patients that answers questions about how specific diet intervention.

The cardiovascular and cognitive benefits of plant-derived bioactives such as flavonoids are increasingly recognised. We are collaborating with Professor Anne-Marie Minihane from UEA to determine the mechanism of flavonoid absorption and metabolism in humans. The response to flavonoid uptake is highly heterogeneous and the variability of dose-response relationships is poorly understood. This is likely to be explained, in large part, by individual genetics and microbiome differences underlying absorption, distribution, metabolism and elimination of flavonoid metabolites.

Studies of EPS in Bifidobacteria

Previous studies have suggested that the presence of the gut microbiota genus Bifidobacterium is associated with a range of health benefits, including pathogen protection. In collaboration with Dr Lindsay Hall's group at IFR, we are focusing on understanding the biology of Bifidobacterium and the role of the exopolysaccharide capsule by measuring and integrating gene expression, metabolite and phenotype data. The main aim of this project is to improve techniques for the integration of omics data in non-model species organisms. Secondly, we plan to develop an analytical framework for combining multiple genetic variants on combined bacterial and host molecular networks that will allow us to study the genetic background underpinning both bacterial adaptation and host susceptibility. Our ambition is to understand cross-talk between genotypes, metabolites and the transcriptome in microbial communities and their host organisms.

Publications

Schoenfelder S., Sugar R., Dimond A., Javierre B. M., Armstrong H., Mifsud B., Dimitrova E., Matheson L., Tavares-Cadete F., Furlan-Magaril M., Segonds-Pichon A., Jurkowski W., Wingett S. W., Tabbada K., Andrews S., Herman B., LeProust E., Osborne C. S., Koseki H., Fraser P., Luscombe N. M., Elderkin S. **Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome**. *Nature Genetics,* 47 (2015)

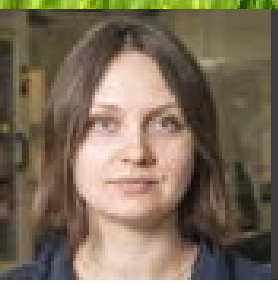
Jurkowski W. J. **ONION: Functional approach for integration of lipidomics and transcriptomics data**. *PLoS ONE*, 10 (2015)

Jurkowski W. J. **DNA Microarray integromics analysis platform**. *BioData mining,* 8 (2015)

Jurkowski W. J. (2015) E**volutionary Algorithms for the Inverse Protein Folding Problem**. *Handbook of Heuristics Springer*

Future Aims

- ESCAPE Effect of diet intervention on control of cancer progression
- Development of methods for analysis of exosomes to exploit potential for cancer biomarkers
- Enable omics data integration and network analysis to Norwich Research Park



KRASILEVA GROUP

Ksenia V Krasileva
BSc University of California Berkeley
PhD University of California Berkeley
Started at TGAC in 2014

O&E

The Krasileva Group studies fundamental principles of plant innate immunity and the use of genomics in basic and translational research. Wheat, a grass highly amenable to laboratory research and of immense importance in food security and human health, is the primary focus of this group. Recent advances in sequencing technologies are opening new frontiers in this area of research. Draft sequence of wheat genome, availability of TILLING populations and CRISPR technologies allow for pathway dissection and engineering of favourable traits in a short time span. Our interests include advancing functional genomics in wheat and applying them to generate increased resistance to wheat pathogens. This group is formed by partnership of The Genome Analysis Centre and The Sainsbury Laboratory.

Project Highlights

Sequencing of wheat TILLING populations

As a part of a large international effort and in collaboration with Platforms & Pipelines and the Ayling group at TGAC as well as partners at UC Davis, JIC and Rothamsted, we have sequenced and analysed over 2,000 mutagenized wheat lines from two TILLING populations of the elite wheat varieties, *Triticum aestivum* cv 'Cadenza' and *Triticum turgidum* cv 'Kronos'. We predicted mutation effects in the annotated wheat genes and set up online databases for searching induced alleles in genes of interested. This mutant database provides the first large-scale reverse genetics sequenced resource in wheat. We are preparing a publication describing the resource.

Rapid identification of wheat genes that increase resistance to yellow rust

Yellow rust is a devastating disease of wheat that threatens food production in the UK and globally. New sources of genetic resistance to yellow rust are urgently needed to boost wheat defenses and lessen our dependence on fungicides. We have previously phenotyped our Kronos TILLING population in the USA and identified mutants with gain-of-resistance. We have confirmed phenotypes in the field tests in the UK and in the controlled laboratory conditions. We plan to leverage recently developed wheat genomic resources to rapidly identify sources of resistance in these lines using exome capture and mapping-by-sequencing approaches.

Evolution in the architectures of plant immune receptors

Plants have powerful defence mechanisms, which rely on the diverse arsenal of plant immune receptors, particularly the NB-LRR (NLR) class. Understanding the architecture of plant immune receptor proteins is key to unravelling their pathogen recognition potential. We developed pipelines to identify NLR complements and their architectures in plant genomes and applied them to 41 publicly available datasets spanning major plant families and major crop species. Our current analyses show exciting new trends that flowering plants use to generate new sources of diversity in plant NLRs. Specifically, novel gene fusions between receptors and additional plant proteins provide a rapid way of evolving pathogen recognition specificities. We plan to investigate the mechanisms how these NLR-fusions are generated. The long-term goal of this project is genetic engineering of plant immune receptors with novel pathogen recognition specificities.

Selected Publications

Schwessinger B, Bart R, Krasileva KV and Coaker G. **Focus issue on plant immunity: from model systems to crop species**. *Frontiers Plant Science, 6* (2015)

Wu CH, Krasileva KV, Banfield MJ, Terauchi R, Kamoun S. **The "sensor domains" of plant NLR proteins: more than decoys?** *Frontiers Plant Science, 6* (2015)

Future Aims

- Robotic surveillance of wheat diseases
- Improvement of wheat genome reference (TGAC wheat LoLa)
- Grassroots Genomics (collaboration with the Davey group)
- Genomics of brusone and Fusarium head blight (collaboration with JIC/EMBRAPA Brasil – pump priming funded to start 1April 2015)

SWARBRECK GROUP

David Swarbreck
PhD Rothamsted Research/University of Bristol
Started at TGAC in 2010

O&E

The Swarbreck Group uses computational approaches to improve genome annotation and understand the complexity of eukaryotic transcriptomes and gene regulation. Our research involves the development of sequencing, assembly, and annotation strategies to characterise the complexity of eukaryotic transciptome landscapes, including coding and non-coding transcripts, alternative splice variants and small peptides. We work mainly on higher eukaryotes such as plants (ash, rubber, willow, barley, wheat). We are interested in also understanding evolution of expression, alternative splicing, and function of genes duplicated by polyploidy events and characterising the changes in gene expression and regulation in polyploidy species and interspecies hybrids.

Project Highlights

Functional genomics of aphid adaptation to plant defenses
In collaboration with the Hogenhout (JIC) and Van Oosterhout (UEA) labs, we are investigating the mechanisms that have given the green peach aphid (GPA) *Myzus persicae* its impressive phenotypic plasticity. *Myzus persicae* is an agronomically important pest worldwide able to colonizes over 400 different plant species from more than 50 plant families and has developed resistance to all insecticides that are currently in use. We are currently exploring the role of transcriptional and epigenetic regulation on the ability of a single GPA clone (consisting of genetically identical individuals) to colonise diverse plant species and on exposure to insecticides with different chemistries. We have sequenced and annotated the ~400 Mbp Myzus genome identifying ~18K genes and ~30K transcripts.

Gene models and sequences are available via aphidbase http://www.aphidbase.com/node_94263/Myzus-DB.

Effect of Chromatin modification on meiosis: wheat, a model for polyploid crops
Most related chromosomes of wild relatives of wheat exhibit extensive gene synteny along their chromosome length. Moreover, the genes on related chromosomes have more than 95 per cent homology at the sequence level. Despite this level of similarity, there is little recombination between wild relative and wheat chromosomes at meiosis due to the presence of the Ph1 locus. Deletion of the Ph1 locus allows the chromosomes to behave like homologous chromosomes and recombine. Swarbreck's Group lead TGAC's efforts to annotate wheat and are collaborating with Graham Moore (JIC) to perform studies in wheat-rye hybrids generated from crossing the wheat variety Chinese Spring with the rye variety Petkus in either the presence or absence of Ph1 in order to assess changes in transcription, methylation and chromatin structure.

Understanding small RNA-Mediated gene regulation
Simon Moxon (Project Leader, TGAC) is leading the group's efforts in collaboration with the Dalmay (UEA) and Moulton (UEA) groups to understand small RNA-mediated gene regulation. We are interested in applying next generation sequencing to address microRNA (miRNA) discovery, biogenesis and expression, miRNA target prediction in plants and animals and to discover sRNA regulatory pathways. We are currently developing tools that utilise data obtained from a Parallel Analysis of RNA Ends (PARE) experiments, sometimes called the degradome, to discover and characterise small RNAs as well as providing visualisation of small RNA/target interaction networks.

Selected Publications

RAMPART: a workflow management system for *de novo* genome assembly. Mapleson D, Drou N, Swarbreck D. *Bioinformatics*, 31 (2015)

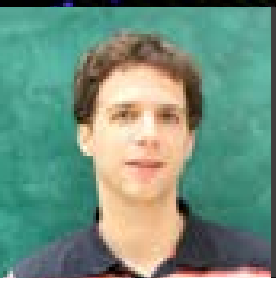
Oakleaf: an S locus-linked mutation of Primula vulgaris that affects leaf and flower development. Cocker JM, Webster MA, Li J, Wright J, Kaithakottil G, Swarbreck D, Gilmartin PM. New Phytologist, 208 (2015)

The RNAi machinery controls distinct responses to environmental signals in the basal fungus Mucor circinelloides. Nicolás FE, Vila A, Moxon S, Cascales MD, Torres-Martínez S, Ruiz-Vázquez RM, Garre V. *BMC Genomics, 16 (2015)*

A non-canonical RNA silencing pathway promotes mRNA degradation in basal Fungi. Trieu TA, Calo S, Nicolás FE, Vila A, Moxon S, Dalmay T, Torres-Martínez S, Garre V, Ruiz-Vázquez RM. *PLoS Genetics, 11* (2015)

Future Aims

- Exploiting natural and synthetic polyploids to investigate the immediate and long-term effects of polyploidy on alternative splicing
- Evolution of expression, regulation, and function of genes duplicated by polyploidy events
- Epigenetics and phenotypic plasticity in Aphids
- Develop tools and pipelines for the annotation of complex eukaryotc genomes, with the focus on the investigation of transcriptome complexity at individual exon and transcript levels
- Explore new technologies and protocols relevant to gene annotation, expression and regulation



TAMÁS KORCSMÁROS

PhD, University of Szeged, Hungary
Started at TGAC in 2014

O&E

TGAC-IFR Fellow

The Korcsmaros Group's main focus is to apply network biology approaches to uncover how the microbiota regulates autophagy in the gut, developing and utilising bioinformatics methods for computational analysis, alongside collaborators at IFR and the UEA.

Project Highlights

The group developed and host three gap-filling resources: SignaLink; Autophagy Regulatory Network; NRF2ome. SignaLink (http://SignaLink.org) is among the top ten signalling network resources according to independent evaluations, and received thousands of visitors in the past year. Data from SignaLink was used in many studies, including papers in *PLoS Computational Biology* and *Nature*.

In collaboration with the Baranyi and Kingsley groups, the Korcsmaros Group developed SalmoNet, the first integrated molecular network of Salmonella, containing metabolic, regulatory and protein-protein interactions (publication in preparation). SalmoNet allows the group to predict and validate potential connections between Salmonella and the host autophagy process. These connections could serve as mechanisms to how Salmonella is modulating the host upon infection.

As a translational project, the Korcsmaros Group have launched the development of TGAC's OmiX Navigator, an integrated software solution and workflow management tool to allow other research groups to perform systems biology projects.

Selected Publications

Türei D, Földvári-Nagy L, Fazekas D, Módos D, Kubisch J, Kadlecsik T, Demeter A, Lenti K, Csermely P, Vellai T, Korcsmáros T. **Autophagy Regulatory Network – a systems-level bioinformatics resource for studying autophagy components and their regulation**. *Autophagy*, 11 (2015)

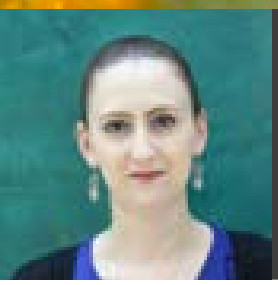
Veres DV, Gyurkó DM, Thaler B, Szalay KZ, Fazekas D, Korcsmáros T, Csermely P. **ComPPI: a cellular compartment-specific database for protein-protein interaction network analysis.** *Nucleic Acids Res.* 43 (2015)

Földvari-Nagy L, Ari E, Csermely P, Korcsmáros T*, Vellai T (2014) **Starvation- response may not involve Atg1-dependent autophagy induction in non-unikont parasites.** *Scientific Reports*, 4 (2014)

Fazekas D, Koltai M, Türei D, Módos D, Pálfy M, Szalay-Bekő M, Lenti K, Farkas IJ, Vellai T, Csermely P, Korcsmáros T **SignaLink 2 – A signaling pathway resource with multi-layered regulatory networks.** *BMC Systems Biology, 7* (2013)

Future Aims

- Develop intestinal cell (e.g., Paneth cell) specific networks from integrated transcriptomics and proteomics datasets.
- Investigate host-probiotics interactions with combined *in silico* and *in vivo* methods
- Analyse the systemic effect of bacterial mediated autophagy down-regulation in the gut
- Further develop the SignaLink resource to serve the model and non model organism community



DIANE SAUNDERS

PhD, University of Exeter
Started at TGAC in 2014

O&E

TGAC-JIC Fellow

The Saunders Group focuses on applying a multidisciplinary approach to the study of plant pathogen interactions, integrating molecular genetics, microbiology, plant pathology, population genetics, genomics and data mining to improve our understanding of the molecular mechanisms at the plant pathogen interface.

Project Highlights

From pathogenomics to mechanistic insights into adaptation

We are especially interested in how pathogens adapt to changing environments such as new host genotypes. This can be studied in depth by characterising and analysing the genes encoding pathogen effector repertoires. Plant pathogens deliver effector proteins to their hosts to reprogramme plant defense circuitry and enable parasitic colonisation. On certain plant genotypes, effectors that act within the host cell can be recognised by immune receptors, encoded by resistance genes, which initiate defence responses.

Apply sequencing technologies to the surveillance of re-emerging plant pathogens

Increasing our understanding of the population dynamics and the evolution rate of (re-) emerging plant pathogens may enhance the deployment of effective resistant genotypes that fully embody the pathogens that pose a significant threat to UK agriculture. For instance, we have numerous projects studying the wheat yellow rust pathogen, *Puccinia striiformis f. sp tritici* that is a substantial threat to wheat production worldwide and more recently re-emerged as a major constraint on UK agriculture.

Selected Publications

Hubbard A., Lewis CM., Yoshida K., Ramirez-Gonzalez RH., De Vallavieille-Pope C., Thomas J., Kamoun S., Bayles R., Uauy C., Saunders, D.G.O. **Field pathogenomics reveals the emergence of a diverse wheat yellow rust population**. *Genome Biology,* 16 (2015)

Saunders D.G.O. **Hitchhiker's guide to multi-dimensional plant pathology**. *New Phytologist*, 10 (2014) (Invited review)

Cantu D, Segovia V, MacLean D, Bayles R, Chen X, Kamoun S, Dubcovsky J, Saunders D.G.O.* and Uauy C*. **Genome analyses of the wheat yellow (stripe) rust pathogen Puccinia striiformis f. sp. tritici reveal polymorphic and haustorial expressed secreted proteins as candidate effectors.** *BMC Genomics,* 14 (2013) *Co-corresponding

Saunders, D.G.O, Breen, S, Win, J, Schornack, S, Hein, I, Bozkurt, T.O, Champouret, N, Birch, P.R.J, Gilroy, E.M and Kamoun, S. **Host Protein BSL1 Associates with Phytophthora infestans RXLR Effector AVR2 and the Solanum demissum Immune Receptor R2 to Mediate Disease Resistance**. *The Plant Cell,* 24 (2012)

Future Aims

- Establish platforms for rust effector function analysis
- Identify host targets of rust effector proteins
- Develop models for how rust pathogens manipulate host defence mechanisms
- Develop methods to enhance the speed of pathogen surveillance
- Establish a low cost surveillance system for (re-)emerging pathogens
- Develop predictive models for pathogen transmission and dispersal

OPERATIONS

The Operations Division includes the Platforms and Pipelines Group, Scientific Computing Group, Business Development & Communications Group and Business Support.

DANIEL SWAN

Head of Platforms & Pipelines
PhD in Developmental Genetics, Imperial College
Started at TGAC in 2015

EB

Platforms & Pipelines

The Platforms and Pipelines (P&P) Group, led by Dr Daniel Swan, are responsible for high-throughput genomics at TGAC as part of our National Capability in Genomics. Maintaining communications throughout, our dedicated project management team advise on experimental design to sample submission procedures as well as carrying out the QC of incoming samples.

The P&P lab team deliver high quality library construction and sequencing across our range of Illumina, Ion Torrent and PacBio platforms - from *de novo* genome sequencing, genome resequencing and transcriptomics to long mate-pair libraries and custom targeted capture. They can also apply our Opgen and BioNano Genomics optical mapping platforms to aid genome assembly projects. Data is analysed by our versatile bioinformatics team who provide bespoke data analysis on our high performance computing systems, maintained by our Scientific Computing Group.

National Capability

TGAC's computing platforms and laboratories provide the BBSRC National Capability in Genomics (NCG).

Four critical objectives define the NCG:

- Provide state-of-the-art technology platforms for genomics research
- Deliver a high quality service to enable researchers to work with TGAC's technologies
- Enable scientific impact by developing and deploying novel techniques and platforms
- Deploy a national resource for surveillance or emergency response requiring genomics expertise

The NCG is delivered jointly by the Scientific Computing Group and the Platforms and Pipelines Group at TGAC, highlighting the importance of Integrated delivery of laboratory and bioinformatics solutions to tackle bottlenecks in modern genome science. With **4,000** processing cores, **7 petabytes** of storage and laboratories that cover the full suite of sequencing platforms, the NCG is a modern, high-throughput facility capable of tackling everything from small prokaryotic genome projects to the sequencing and assembly of complex eukaryotic genomes.

In the last year, the NCG has dealt with nearly 300 diverse genomics projects resulting from responding to over 600 customer enquires.

Our largest projects this year were to sequence thousands of crop exomes, and means that the NCG has a huge amount of expertise in the automation of in-solution target capture. Our high-throughput BAC sequencing pipeline sequenced **115,000** BACs in the last quarter of the year alone.

We are continually reviewing technologies and bringing new protocols online to service the ever changing face of sequence-based science.

For more information contact tgac.projects@tgac.ac.uk

SCIENTIFIC COMPUTING

"Intel is delighted to welcome TGAC to the Scientific Computing Conference 2014 where will we continue our close collaboration to help accelerate the pace of discovery of TGAC's research agenda.

Robert Maskell, High Performance Computing at Intel

TIM STITT

Head of Scientific Computing

PhD Computational Science, Queen's University Belfast

Started at TGAC in 2014

DB

The Scientific Computing Group (SCG), led by Tim Stitt, is responsible for ensuring that TGAC researchers (and external collaborators) have ready access to all the computing resources they need to carry out their computational research effectively. Associated with this mission, the SCG Group provides dedicated user support on bioinformatics tools and libraries, trains users to use the computing resources productively while applying their expertise in code tuning, optimisation and parallelisation to provide fast, efficient application codes on TGAC hardware. The SCG Group also works to evaluate new and novel hardware and software technologies with the goal of keeping TGAC researchers at the vanguard of computational bioscience.

Key Achievements

Energy-efficient bioinformatics

While energy-efficient computing has garnered much interest in traditional HPC computing domains in recent years its ingress to the bioinformatics domain has been more muted. In this project, we are investigating novel processing hardware for performing energy-efficient bioinformatics calculations. Current investigations focus around the latest ARMv8 64-bit processors and GPU accelerators.

Elastic high-performance computing for external users

In this project, we are investigating mechanisms for provisioning elastic cloud computing resources for external TGAC users. This work is in association with the NCG with the goal of opening up TGAC HPC resources to the wider bioinformatics community.

Tuning of the DISCOVAR assembler on the SGI UV2000 architecture

Assembly of large genomes is a critical computational activity at TGAC. Our recent wheat assemblies can take weeks to complete so improvements in code performance can have an important impact on the productivity of wheat researchers. In this project, the SCG has teamed up with SGI and Intel developers to profile and improve the runtime performance of the DISCOVAR assembly software on the SGI UV2000 supercomputer.

Project GENESYS - energy efficient comparison of DNA gene sequences using a revolutionary optical processor

In this project, the SCG is collaborating with an industrial partner (Optalysys) to design, implement and test a revolutionary new optical processor for sequence alignment calculations.

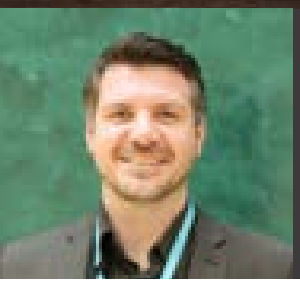
This project is funded through an Innovate UK award with the goal of performing BLAST-like sequence searches using optical processing techniques at 95 per cent energy efficiency of traditional computing approaches. Awarded a **£600K** grant via Innovate UK to develop and test a revolutionary optical processor for DNA sequence alignments.

- **"Introduction to High-Performance Computing"** University of East Anglia (November 2014)
- **"The World's Next Top Model – The Beauty of Mathematics"** BBSRC Scientists in the School Program, INTO UEA (January 2015)
- **"Beyond Human - Sequencing the Complex Wheat Genome to Advance Global Food Security"** The Annual SGI User Group Meeting, SC14, New Orleans (2014) and the 12th EMEA HPC Intel Roundtable, St. John's College, Oxford University (2015)
- Invited Speaker - **BioData World Congress 2015**, Cambridge, UK

KNOWLEDGE EXCHANGE & COMMERCIALISATION

"We are thrilled to be working with the leading research Institute in genomics, TGAC on this exciting project. The GENESYS system has the potential to fundamentally change the field of DNA analysis.

Optalysys CEO, Dr Nick New



STUART CATCHPOLE

Head of Business Development & Communications

BSc Business Management, University of East Anglia

Started at TGAC in 2013

The Business Development & Communications Group covers all Knowledge Exchange & Commercialisation (KEC) and Marketing & Communications activity at TGAC. TGAC's KEC vision is to deliver long-term social and economic impact from our research, world class science and capability.

The KEC team raises awareness of funding opportunities with our staff, offering customised advice on best suited grants. They also assist in market research to inform grant applications. Relationships between TGAC, other research organisations and industry are carefully nurtured, facilitating introductions and then structuring engagement. Projects are guided to success via continuous monitoring of intellectual property, industry feedback and detailed asset exploitation plans. The team also helps to develop important networking skills necessary for partnership working, removing obstacles to high-quality research.

Key Achievements

- 6 out of 6 KEC projects funded spanning 6 sectors: Agri-tech, Energy Efficiency, Human Health, Big Data, Surveillance & Diagnostics, Synthetic biology
- More than £4M won in collaborative and translational grants
- 15 new collaborative projects, 10 of those with industry partners
- Individualised market reports in Agri-tech, DNA synthesis and Optical Processing to inform strategy
- 2 CASE studentships

"Our joint Innovate UK bid was successful with much of the credit due to TGAC's KEC team. They helped us by providing a market analysis and insights as to how we could understand the product potential and exploitation opportunities. We are very much looking forward to working with them further on this when the project kicks off next month." Emma Blaylock, Head of Business Development at Optalysys.

"KEC has been extremely helpful in identifying opportunities for funding and making it possible for me to succeed in getting my NRP Translational Fund Award. I have had a lot of support at every stage of the grant writing process and they have helped me understand how to give a business spin to the scientific work I do. I thoroughly commend their work." Dr Manuel Corpas, Bioinformatics Project Leader at TGAC.

"Our recent Innovate UK award win would not have been possible without the dedicated support and efforts of the TGAC KEC team. Their guidance and expertise is an invaluable resource when developing and managing these types of industrial collaborations and I look forward to leveraging their skills on upcoming projects." Dr Tim Stitt, Head of Scientific Computing at TGAC.

TGAC Facilities

This year we have delivered a capital programme to develop our facilities to enable us to deliver world leading research and embed TGAC on the Norwich Research Park.

Refurbishment

We have refurbished and extended our offices and laboratories, and developed a dedicated training suite for computational bioinformatics.

We have also opened a TGAC reception and upgraded the internal and external spaces and facilities for staff.

361º DIVISION

"It's great to see that TGAC4Kids on the road has been a success. The events have gone from strength to strength and the teachers were delighted with the activities we provided. The pupils clearly enjoyed themselves, judging by the positive feedback we received, and they learned a lot in just two hours. We look forward to taking TGAC4Kids on the road to as many schools as possible.

Dr Pete Bickerton
Public Engagement & Society Officer at TGAC

VICKY SCHNEIDER

Head of 361° Division

PhD University of Leiden and University of Lyon

Started at TGAC in 2013

DB

The Division comprises of four expert teams: Scientific Training and Education, Public Engagement & Society, Best practice & e-Science, and Research & Learning. We offer a variety of cutting-edge programmes and activities aimed to foster best practice in Open Science, data driven analytical skills, post graduate training expertise, including how to create scalability and long-term impact of such efforts. The Research & Learning team focuses on best practice in Open Science as a vehicle for enhancing data integration and interoperability of data analysis workflows, with a keen research interest in how adults learn new skills in bioinformatics.

The Division also aims to make science more accessible to the general public with an understanding of how science impacts our daily lives, by enhancing interdisciplinary collaborations and fostering knowledge exchange around the world.

Key Achievements

- Established TGAC as a Centre of Excellence in Training (Bioinformatics)
- Organised and hosted 25 Hands on Courses reaching more than 400 attendees with a widespread diversity (from 26 countries (52 per cent in UK) 36 nationalities (35 per cent UK))
- Organised and hosted 15 interactive best practice & e-Science workshops reaching more than 200 attendees
- Set up the BBSRC Bioinformatics and Biomathematics Training Hub (BBHub)
- Public Engagement & Society team events and reach: 6 onsite public events (402 participants), 3 off-site Public events (842 participants), 5 school visits off-site (327 pupils), school visits on-site (81 pupils). Total 1664 members of the public reached.
- 'I'm a Scientists Get me out of here!' zone: winner of the bioinformatics round.

Publications

D'Auria G., Schneider M. V., Moya A. **Live Genomics for Pathogen Monitoring in Public Health**. *Pathogens,* 3 (2014)

Welch L., LeWitter F., Schwartz R., Brooksbank C., Radivojac P., Gaeta B., Schneider M. V. **Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies**. *PLOS: Computational Biology,* 3 (2014)

Brazas M., Lewitter F., Schneider M. V., Van Gelder C., Palagi P. **A Quick Guide to Genomics and Bioinformatics Training for Clinical and Public Audiences**. *PLoS Computational Biology,* 10 (2014)

Aleksic J., Alexa A., Attwood T. K., Hong N. C., Dahlo M., Davey R., Dinkel H., Forstner K. U., Grigorov I., Heriche J. K., Lahti L., MacLean D., Markie M. L., Molloy J., Schneider M. V., Scott C., Smith-Unna R., Vieira B. M. **The Open Science Peer Review Oath**. *F1000Research,* 1 (2014)

Future Aims

The coming year promises exciting developments and new projects for 361° Division. We will be rolling out the Research and Learning Group activities, including analysing and gaining deeper understanding of the mechanism for effective data integration in the life sciences. The Best Practice and e-Science group will continue the 'Train the Trainer' collaboration with BioPlatforms Australia Ltd, applying the same methodology to roll-out the programme across NRP and the UK, as well as pursuing a training vehicle for best practice in Open Science. On the horizon for our Public Engagement & Society team is the launch of the TGAC4Kids online learning platform and consolidating the 'TGAC4Kids on the Road' national engagement programme. Our Scientific Training and Education team will deliver innovative training courses, which also incorporate peer review of our training, merged with collation of ongoing feedback to derive measurable impact of our training.

Review of the Reporting Year

April 2014 - March 2015

DNA Synthesis at the Norwich Research Park

APRIL 2014

TGAC was selected as one of only five centres across the UK to develop a state-of-the-art DNA synthesis centre in Norwich. The facility aims to support the design, generation and exploitation of high-value compounds and bioactives obtained from plants and microbes to contribute to areas of strength within the NRP.

Recent advances in DNA technologies have stimulated the development of innovative research in synthetic biology. The investment will equip the NRP with the cutting-edge technology in DNA synthesis that will build on the existing National Capability in Genomics at TGAC and help to propel the UK to the forefront of synthetic biology research.

We take part in the MinION Access Programme

TGAC was one of the first research Institutes to be part of the MinION's early access programme (MAP), where our Sequencing Informatics and Plant and Microbial Genomics groups trialled the miniaturised sensing system.

Introduced by Oxford Nanopore Technologies, the MinION compresses Nanopore-sensing technology into a portable low-powered device for electronic single-molecule sensing experiments. The MAP allows scientists to develop sensing applications such as DNA sequencing and provides immediate data for analysis.

As a self-contained disposable device to deliver real-time experimental data, the MinION can be plugged directly into a USB port in a laptop or desktop computer. The new device is aimed at enabling scientists to fit the technology to their applications, including using and developing open-source software.

We take the lead towards a sustainable future for wheat genomics

TGAC hosted the workshop "Towards a sustainable future for wheat genomics" that introduced the latest, freely available wheat genomic resources and the new opportunities these bring for sophisticated analysis.

Experts in genomics and bioinformatics from four world-leading, UK-based institutes: TGAC, John Innes Centre (JIC), European Bioinformatics Institute (EBI) and Rothamsted Research, came together to present the state-of-the-art public wheat genomic data sets which have the potential to accelerate traditional crop improvement methods. The workshop gave participants the opportunity to both learn about the new resources and to influence the direction of these developments.

"NGS Data after the Gold Rush" with European partners

MAY 2014

TGAC led the training workshop with European partner SeqAhead, Cooperation in Science and Technology (COST) Action, which aims to comprehend Next Generation Sequence (NGS) data and galvanise efficient workflows for NGS data storage, retrieval and analysis.

With 70 stakeholders from 28 countries, the 3-day meeting explored state-of-the art NGS data analysis, its current challenges and applications. Dr. Ana Conesa from the Prince Felipe Research Centre (Spain) and Aleksandra Pawlik from the Software Sustainability Institute (UK) chaired presentations from leading international scientists, covering data analysis and management, bioinformatics training, functional and pathogen genomics. The SeqAhead Management Committee meeting concluded the event, led by Chairman Erik Bongcam-Rudloff, and Vice-Chair Terri Attwood.

JUNE 2014

TGAC at the forefront of Next Generation Sequencing capability

Expanding our sequencing capabilities, we added two Illumina HiSeq 2500 to our cutting-edge platform suite. The new HiSeqs generate up to 1 Terabase (Tb) of sequence data per run, enhancing TGAC's role as a BBSRC National Capability in Genomics.

The new laboratory equipment provided the National Capability Institute with the most up-to-date level of next generation sequencing for genome analysis projects. Part of the latest innovations from Illumina's growing collection of sequencing technology, the HiSeq 2500 supports the broadest range of genome sequence applications and study sizes, with the greatest yield of sequence data per run.

New genetic data released to combat ash dieback epidemic

TGAC released new genetic data that will help understand the spread of the ash dieback epidemic, across Europe and the UK.

As part of the NORNEX consortium, we sequenced 20 genomes of the fungus (*Hymenoscyphus fraxineus*) responsible for the spread of the ash dieback epidemic that threatens our third most common broadleaf tree (after oak and birch). The data is available for analysis on the crowdsourcing site OpenAshDieBack.

Genetic blueprint of bread wheat genome unveiled to improve world's most popular cereal crop

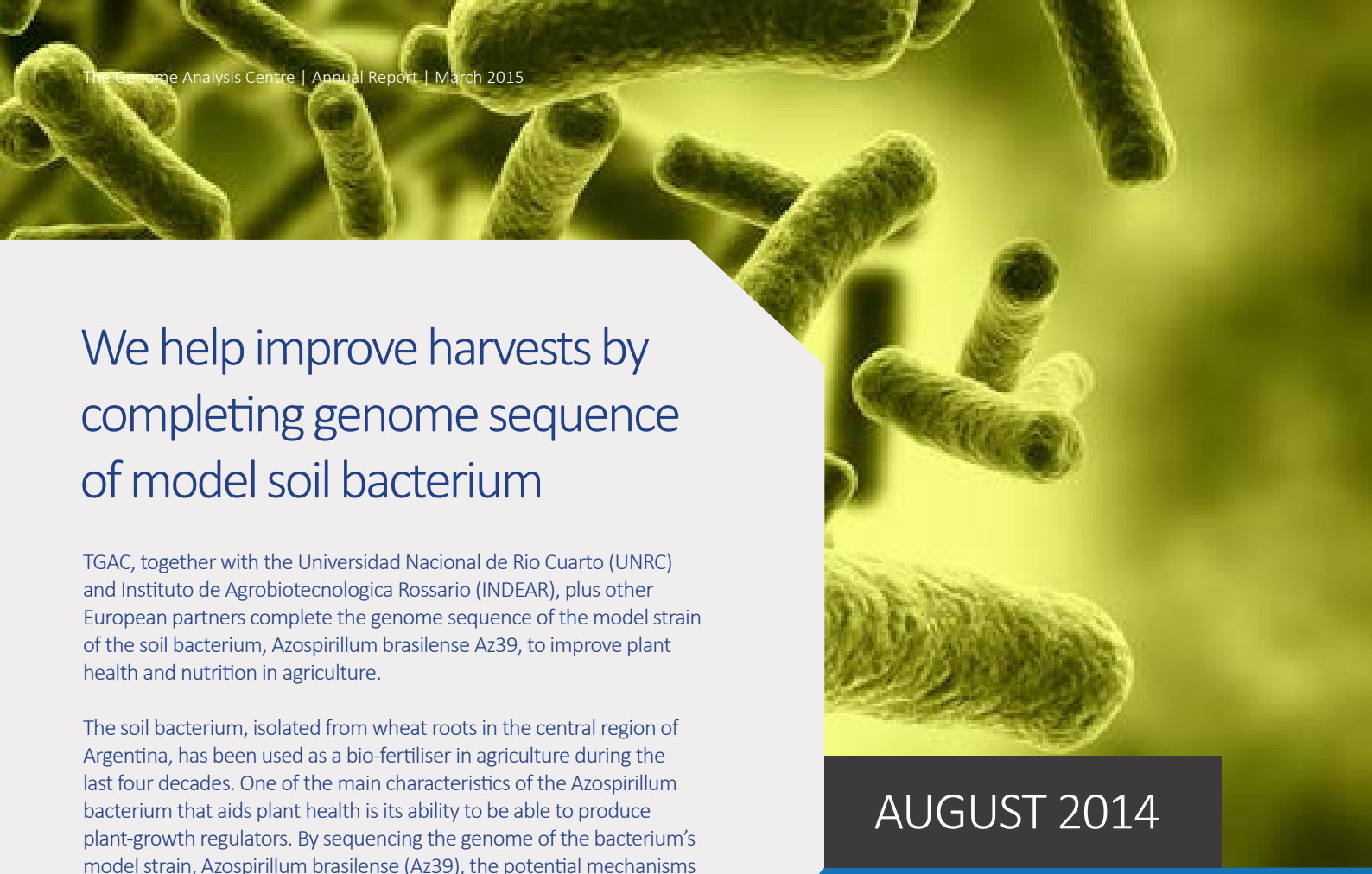
TGAC played a key part in the generation and analysis of the draft sequence of the bread wheat genome, published in the international journal *Science*. The work led by the International Wheat Genome Sequencing Consortium (IWGSC) provides new insight into the structure, organisation, and evolution of the large, complex genome of the world's most widely-grown cereal crop.

JULY 2014

UK Centre of Excellence for ground-breaking whole genome mapping technology

TGAC is the first research Institute in the UK to install the pioneering genome mapping Irys System technology by BioNano Genomics, adding to our existing state-of-the-art Next Generation Sequencing (NGS) suite.

The Irys System was initially used to improve the genome assembly of the British Ash tree sample from the Earth Trust, as part of our collaboration with Queen Mary University of London (UK) for research into the species' disease. Also used for TGAC's bread wheat genome sequencing project to improve the DNA make-up of this important crop. The work on the wheat genome will help to accelerate breeding, with a direct impact on increasing the crop and its yields, contributing to global food security.



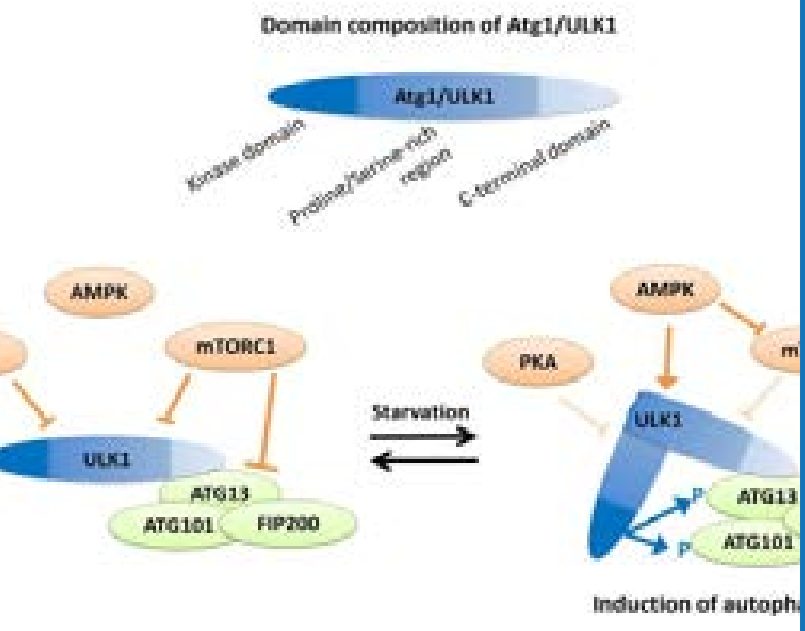
We help improve harvests by completing genome sequence of model soil bacterium

TGAC, together with the Universidad Nacional de Rio Cuarto (UNRC) and Instituto de Agrobiotecnologica Rossario (INDEAR), plus other European partners complete the genome sequence of the model strain of the soil bacterium, Azospirillum brasilense Az39, to improve plant health and nutrition in agriculture.

The soil bacterium, isolated from wheat roots in the central region of Argentina, has been used as a bio-fertiliser in agriculture during the last four decades. One of the main characteristics of the Azospirillum bacterium that aids plant health is its ability to be able to produce plant-growth regulators. By sequencing the genome of the bacterium's model strain, Azospirillum brasilense (Az39), the potential mechanisms responsible for growth improvement have been subsequently unravelled.

AUGUST 2014

Novel insights revealed into the induction of autophagy



Tamás Korcsmáros, TGAC-IFR Computational Biology Fellow, in collaboration with the NetBiol group, released a new study on the bioinformatics analysis of the possibilities of autophagy induction in 40 unicellular parasitic species that could help to identify novel therapeutic targets, especially for patients with colon cancer or inflammatory bowel disease (IBD), published in *Scientific Reports*.

Autophagy is a highly conserved self-degradation process of eukaryotic cells, important in various cellular processes including stress-response, protein metabolism, differentiation and ageing. In the gut, autophagy provides a powerful means of removing intracellular pathogens, and the malfunction of autophagy is related to IBD and cancer progression. A better understanding of the effect of particular bacterial species on the regulation of human intestinal autophagy could help to identify prognosis markers to the common diseases.

SEPTEMBER 2014

Landmark study reveals diverse molecular mechanisms underlying evolution

Reservoir of mutations enabled cichlid fish to adapt to varied environments; results shed light on mechanisms of vertebrate evolution.

The new study, led by Director of Science at TGAC Federica Di Palma, for the first time, examines the molecular basis of adaption and evolution in vertebrates by sequencing the genomes of African cichlid fish species, published in major science publication *Nature*.

In an effort to understand the molecular basis of adaptation in vertebrates, researchers sequenced the genomes and transcriptomes of five species of the African cichlid fish. The researchers uncovered a variety of features in the cichlid genome that enabled the fishes to thrive in new habitats and ecological niches within the Great Lakes of East Africa. In addition to helping explain the complex genomic mechanisms that give rise to incredible diversity among cichlid fishes, the findings from these 'natural mutants' shed new light on the molecular process of evolution in all vertebrate species.

TGAC's new training suite opens its doors to open-access science

In collaboration with the bioinformatics Coordination Action 'AllBio', we host three-day workshop on open science and reproducibility in our brand new training suite.

Held in our new training suite part-funded by BBSRC, the interactive workshop brought together both researchers and non-researchers from different life science fields with one common interest: open science and reproducibility.

Openness not only exists in terms of publications and data, but can be applied to the research process itself. More than ever, given the magnitude of data in bioinformatics, the ability to adopt best practice in workflows, standards, tools for data storage and sharing as well as its analysis, is imperative to making progress in science discovery.

We lead research to help identify animal-to-human transmitted diseases

TGAC leads research into the development of bioinformatics to support the identification and characterisation of viruses through metagenomics.

The BBSRC-funded research, led by TGAC's Dr Richard Leggett, aims to develop computational algorithms that can accurately assemble viral genomes contained within metagenomic samples. These microbial samples pose a challenge to researchers as, not only do they contain numerous different viral species; it is also difficult to locate precisely which species are present.

OCTOBER 2014

Norwich Research Park receives £12.5M investment for next generation of scientists to lead the revolution in bioscience

TGAC is part of the consortium of research institutions on NRP, led by the John Innes Centre that received an additional £12.5M from the BBSRC for its doctoral training programme, confirming the Park's status as a world-class centre of excellence for bioscience.

The endorsement will fund 125 PhD students through the NRP Doctoral Training Partnership over the next five years. This investment further cements Norwich Research Park's growing reputation as a leading, internationally recognised centre of excellence in bioscience. The new funding significantly builds on a £4M BBSRC investment made in 2012, which supported 39 studentships over three years.

Ferret genome study provides clues to how the respiratory system responds to pandemic flu and cystic fibrosis

Genetic analysis reveals airway and lung responses to pandemic flu and cystic fibrosis.

In what is likely to be a major step forward in the study of influenza, cystic fibrosis and other human disease, an international research effort has sequenced the ferret genome. The sequence was then used to analyse how influenza and cystic fibrosis affects respiratory tissues at the cellular level.

The National Institute of Allergy and infectious Diseases, of the National Institutes of Health, funded the project that was coordinated by Michael Katze and Xinxia Peng at the University of Washington in Seattle and TGAC's Federica Di Palma, and Jessica Alfoldi at the Broad Institute of MIT and Harvard.

TGAC and JIC identify new genetic markers to combat yellow rust disease in wheat

New study released identifying genetic markers that signal resistance to the wheat yellow rust pathogen.

To counter the threat of wheat yellow rust, breeders have developed wheat variants that incorporate resistance genes (R-genes). Yr15 is an R-gene taken from wild emmer (Triticum dicoccoides) that provides resistance to this disease. Yet it remains difficult, due to the complex hexaploid structure of wheat DNA, to accurately identify which plants should be used in breeding schemes to ensure that the next generation of wheat will have the resistance gene.

Funded by the BBSRC, the study was led by Ricardo Ramirez-Gonzalez and Dr Mario Caccamo at TGAC and Dr Cristobal Uauy at the John Innes Centre (JIC), as part of an international team of scientists from agri-tech Institutes.

NOVEMBER 2014

DECEMBER 2014

Understanding Life on Earth

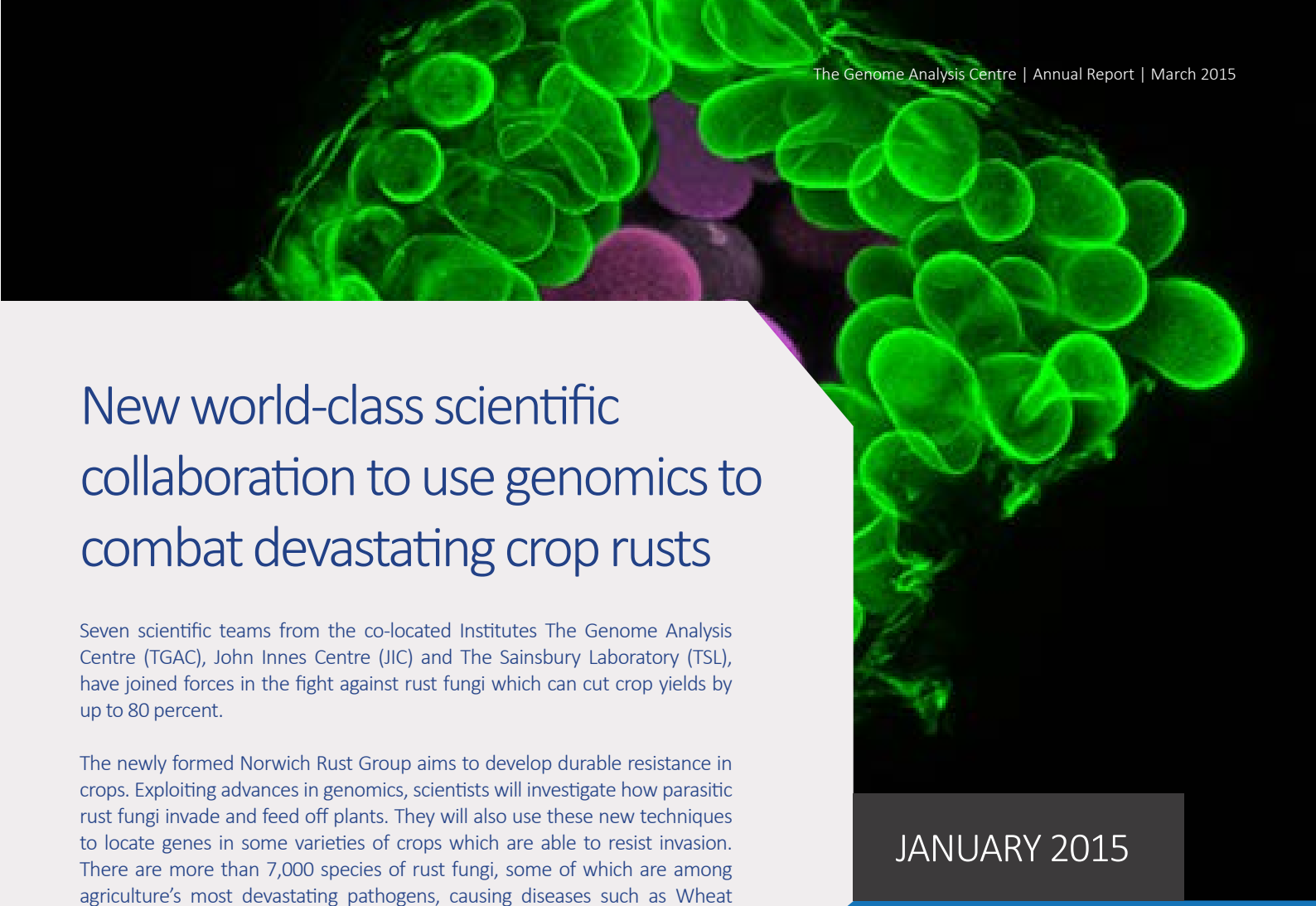
TGAC uses Intel-powered super computers to collect and crunch complex data that is shaping the future of science.

Tackling problems like population growth and securing food supply are what TGAC aims to contribute towards. Having a supercomputer that keeps up with the vast level of data delivered is crucial.

The Institute's research output requires a significant amount of computing power to handle the hundreds of samples and millions of sequences analysed every day. In the human genome alone, there are three billion letters of DNA. At TGAC, these super computers are used to categorise, process, and analyse the genome sequences of a diverse range of plants, animals, and microbes.

TGAC post-grad student awarded IWGSC Early Career accolade

Ricardo Ramirez-Gonzalez was awarded an IWGSC (International Wheat Genome Sequencing Consortium) Early Career Award for developing bioinformatics pipelines that can be used for wheat improvement using next-generation sequencing (NGS), at the International Plant & Animal Genome XXIII (PAG) conference in San Diego.



New world-class scientific collaboration to use genomics to combat devastating crop rusts

Seven scientific teams from the co-located Institutes The Genome Analysis Centre (TGAC), John Innes Centre (JIC) and The Sainsbury Laboratory (TSL), have joined forces in the fight against rust fungi which can cut crop yields by up to 80 percent.

The newly formed Norwich Rust Group aims to develop durable resistance in crops. Exploiting advances in genomics, scientists will investigate how parasitic rust fungi invade and feed off plants. They will also use these new techniques to locate genes in some varieties of crops which are able to resist invasion. There are more than 7,000 species of rust fungi, some of which are among agriculture's most devastating pathogens, causing diseases such as Wheat Stem Rust, Wheat Yellow (Stripe) Rust, Asian Soybean Rust and Coffee Rust.

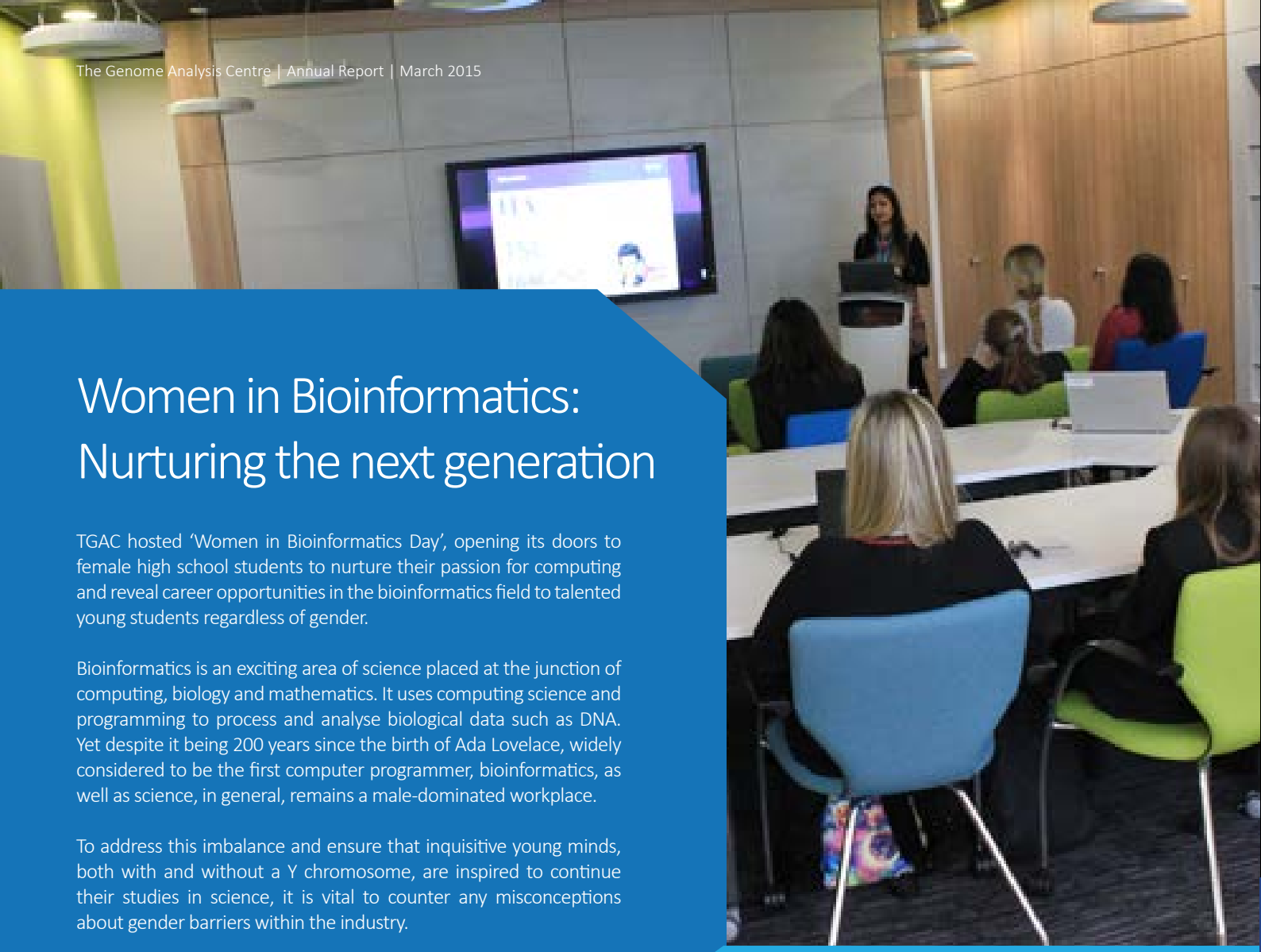
JANUARY 2015

Black-footed Ferret: Reviving a declining population

The Vertebrate & Health Genomics group at TGAC is preparing to start work on a project, in collaboration with the US Fish and Wildlife Service, the San Diego Frozen Zoo and the Longnow Revive and Restore Program, to help conservation efforts for the Black-footed Ferret (Mustela nigripes).

The Black-footed Ferret is an endangered nocturnal predator of the North American prairies, feeding almost solely on various prairie dog species, with which they share their burrows.

Researchers hope to identify genetic diversity responsible for disease resistance in previous Black-footed Ferret populations, which are not present in the sole surviving group, to boost its chance of a successful revival.



Women in Bioinformatics: Nurturing the next generation

TGAC hosted 'Women in Bioinformatics Day', opening its doors to female high school students to nurture their passion for computing and reveal career opportunities in the bioinformatics field to talented young students regardless of gender.

Bioinformatics is an exciting area of science placed at the junction of computing, biology and mathematics. It uses computing science and programming to process and analyse biological data such as DNA. Yet despite it being 200 years since the birth of Ada Lovelace, widely considered to be the first computer programmer, bioinformatics, as well as science, in general, remains a male-dominated workplace.

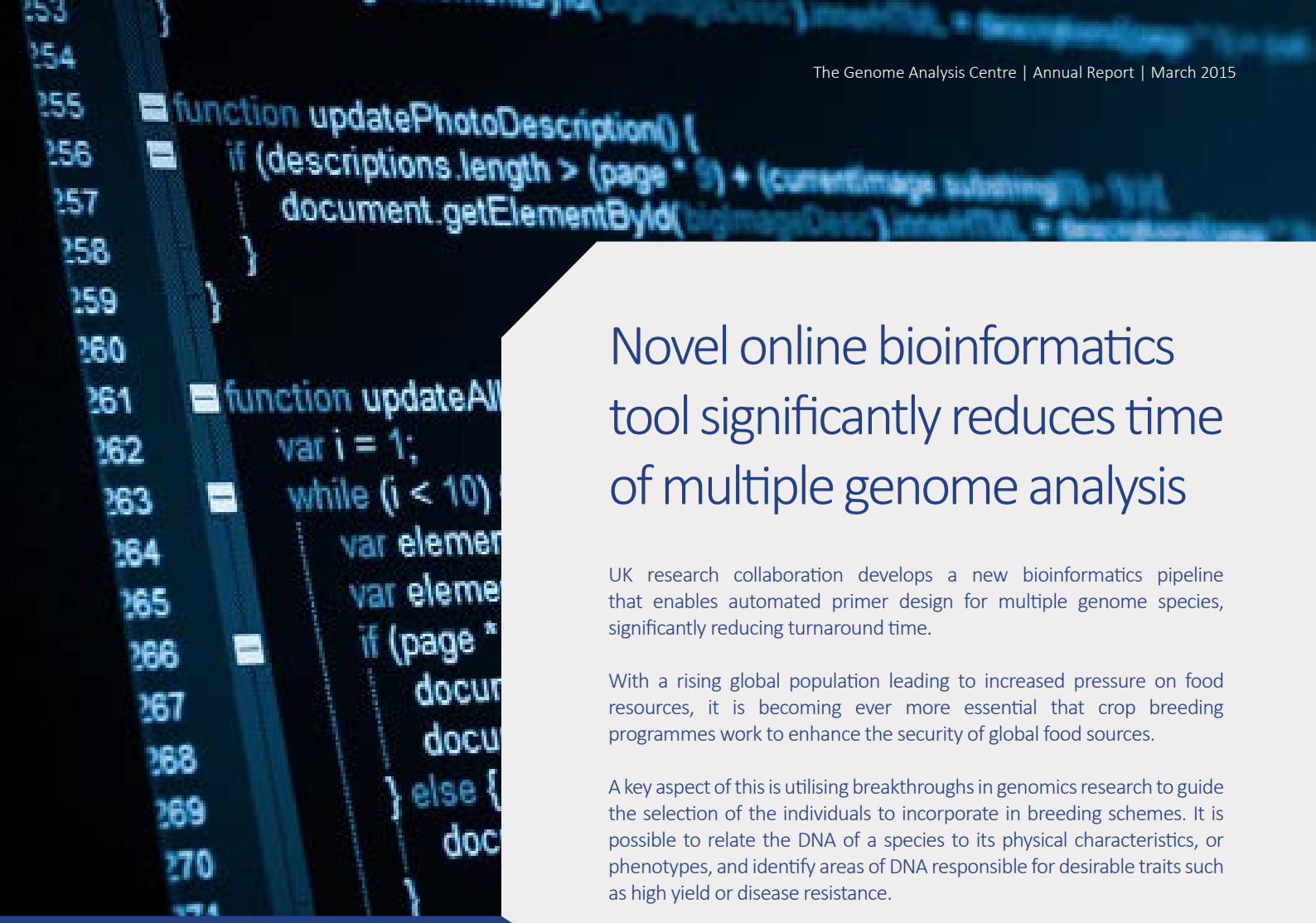
To address this imbalance and ensure that inquisitive young minds, both with and without a Y chromosome, are inspired to continue their studies in science, it is vital to counter any misconceptions about gender barriers within the industry.

FEBRUARY 2015

TGAC and scientific partners awarded £6m to tackle big data challenges in bioscience

As part of the UK's Biotechnology and Biological Sciences Research Council (BBSRC) big data infrastructure announcement at the AAAS 2015 Annual Meeting, TGAC, with partner Institutes were awarded £6m for three joint projects: Big Data Infrastructure for Crop Genomics; iPlant UK – creating a cyber-infrastructure for plant sciences; and Establishing the infrastructure for functional annotation of farmed animal genomes.

BBSRC invested £7.5M in new infrastructure to tackle bioscience big data challenges. The new funding will improve the storage and curation of enormous datasets that will unlock untold discoveries in important areas like health, agriculture and sustainable fuels.



Novel online bioinformatics tool significantly reduces time of multiple genome analysis

UK research collaboration develops a new bioinformatics pipeline that enables automated primer design for multiple genome species, significantly reducing turnaround time.

With a rising global population leading to increased pressure on food resources, it is becoming ever more essential that crop breeding programmes work to enhance the security of global food sources.

A key aspect of this is utilising breakthroughs in genomics research to guide the selection of the individuals to incorporate in breeding schemes. It is possible to relate the DNA of a species to its physical characteristics, or phenotypes, and identify areas of DNA responsible for desirable traits such as high yield or disease resistance.

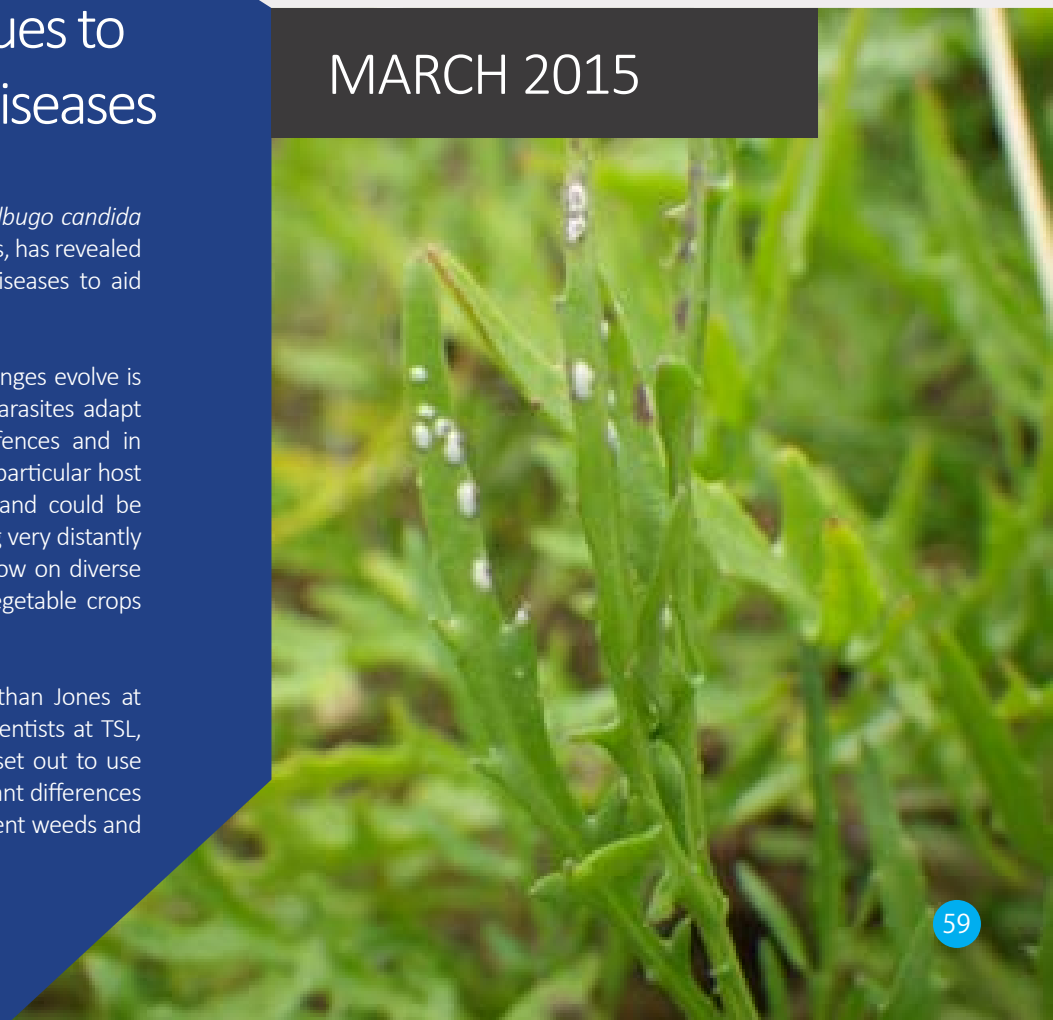
Parasite provides clues to evolution of plant diseases

A new study into the generalist parasite *Albugo candida* (A. candida), cause of white rust of brassicas, has revealed key insights into the evolution of plant diseases to aid agriculture and global food security.

How generalist parasites with wide host ranges evolve is a central question in parasite evolution. Parasites adapt in response to their host organisms' defences and in many cases this adaptation is specific to a particular host species. *A. candida* is a plant pathogen and could be easily confused with a fungus despite being very distantly related to fungi. The plant parasite can grow on diverse plants of the cabbage family, including vegetable crops and common weeds.

The project ("Albugon") led by Prof Jonathan Jones at The Sainsbury Lab (TSL) and a team of scientists at TSL, including Mark McMullan, now at TGAC, set out to use genome sequencing to identify the important differences between *A. candida* races that infect different weeds and crops.

MARCH 2015



Company Registration: 6855533
Charity Number: 1136213

Contact Us
Norwich Research Park
Norwich
NR4 7UH UK

T: +44 1603 450861
E: tgac.enquiries@tgac.ac.uk