Bioinformatics investigations in **Africa**

The **South African National Bioinformatics Institute** plays a vital role in computational biology research in Africa, as well as training future leaders in the field. Its importance is set to grow over the coming years

SOUTHERN AFRICA, AND particularly the country of South Africa, is currently an exciting location for those in the field of bioinformatics. 2011 saw the National Institutes of Health (NIH) and Wellcome Trust make a strong commitment to support research on the continent with the Human, Heredity and Health (H3Africa) programme, as well as a Pan African Bioinformatics Network to support genetics projects funded by H3Africa and biobanks across the African continent. 2012, similarly, was the inaugural year of the South African Society for Bioinformatics and Computational Biology and saw the establishment of the Southern Africa Human Genome Programme, designed to increase South Africa's capacity for genomics research. Overall, the region has poised itself to become a major contributor to genomic research in the near future.

The continent of Africa is plagued by a number of major health problems that could benefit hugely from genomic, and subsequently bioinformatics, investigations. This approach may, for example, help to shed light on the parasite-host interactions that result in malaria and sleeping sickness, or to unravel the question of why some people do not respond to currently available drugs for the treatment of HIV and tuberculosis. Even with these specific questions aside, the countries that make up Southern Africa have been found to be among the most genetically diverse in the world; therefore, the region simultaneously reflects a great need for genetic study and a great opportunity for it.

FILLING THE GAP

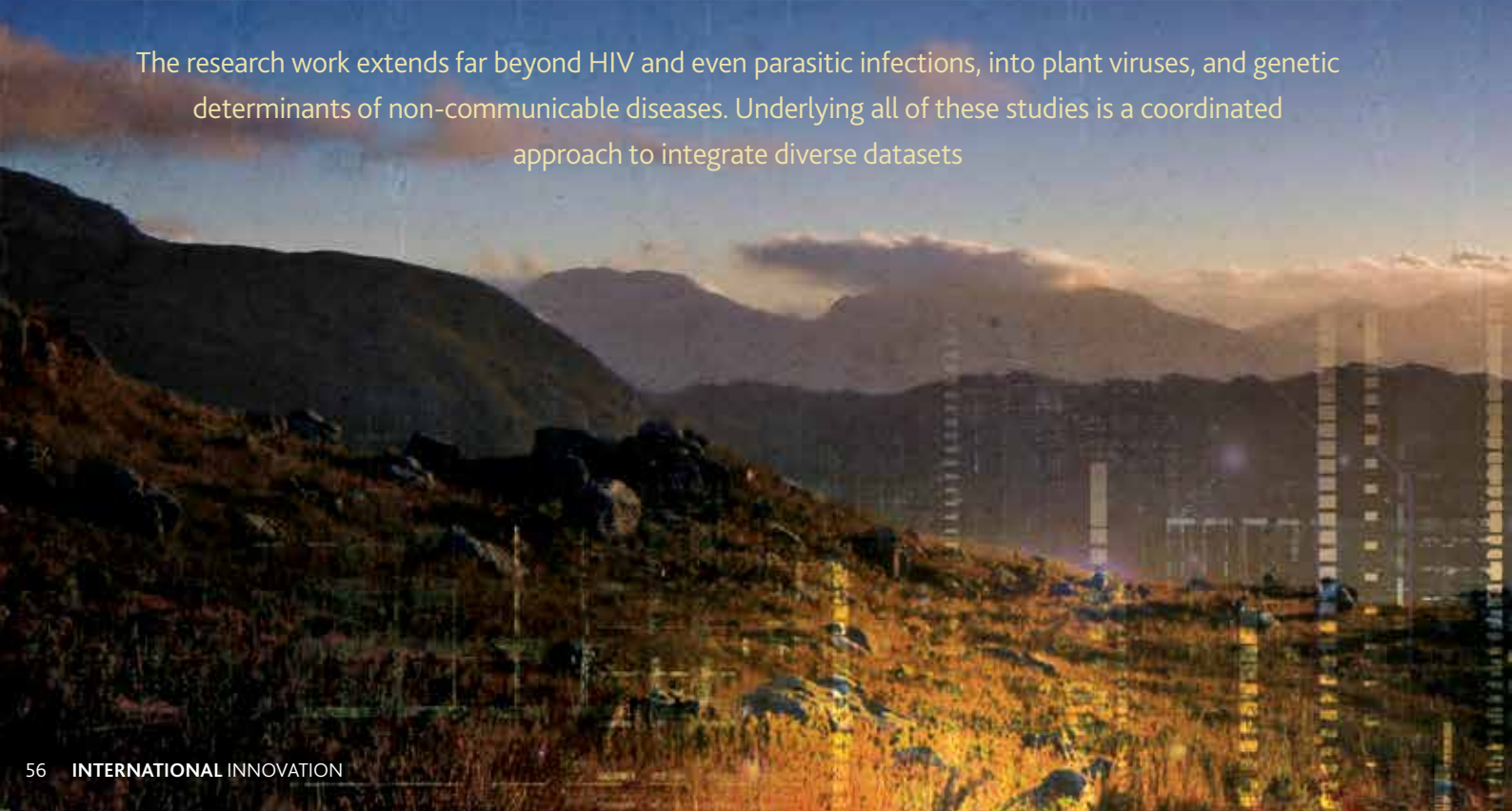
According to the World Health Organization (WHO), no more than 10 per cent of the total staff required to provide adequate genetic services were available in South Africa in 2008 – and human health is only one of the areas that has suffered because of this deficiency. Greater powers of genomic investigation could help Africa in its production of food crops and animals, for example, and there are also unique opportunities to conduct more fundamental research into plant, animal and human populations. Southern Africa's promising initiatives are the first step on the path towards this bright future, but the next will be to ensure that researchers and scientists in the field of computational biology and bioinformatics are being sufficiently trained.

A major player in this research landscape, and a vital provider of new scientists and researchers, has been the South African National Bioinformatics Institute (SANBI), located at the University of the Western Cape in South Africa. SANBI was founded in 1996, and in 2000 it was responsible for hosting South Africa's first workshop on genomics. Comprising five research teams headed by Drs Junaid Gamieldien, Simon Travers, Nicki Tiffin and Gordon Harkins, under the directorship of Professor Alan Christoffels, the institute has continued to achieve its important goals: to conduct nationally and internationally relevant, cutting-edge research, and to educate and mentor biological and computational scientists. The institution produces around six graduates per year, half of these being Master's students and the other half PhDs. Many of these students are drawn from historically disadvantaged backgrounds, and all of them go on to further study or to take positions amongst the faculty of other national or international institutions.

PROFILING PARASITES

Concerning human health, one of SANBI's most important projects has been investigating the genetic adaptations that hosts and parasites have developed towards one another, with particular regard to three pairings which are responsible for tuberculosis, malaria and sleeping sickness. In the case of sleeping sickness, a disease caused by the protozoan *Trypanosome brucei*, its work focuses on analysing the genome of the tsetse fly. This fly is a common vector for parasitic infection, but is itself immune to the

The research work extends far beyond HIV and even parasitic infections, into plant viruses, and genetic determinants of non-communicable diseases. Underlying all of these studies is a coordinated approach to integrate diverse datasets



protozoan's attacks; studying its genes may reveal why that is. At the same time, the team is using comparative phylogenetic analyses to identity SNARE proteins within the parasites.

In terms of tuberculosis, the abundance of mycobacterial samples in South Africa has led to the rapid sequencing of hundreds of mycobacterial genomes. Christoffels' team is in the process of building workflows to automate the assembly of sequencing data from Pac Bio and illumina. Ultimately, the goal is to map the genetic variation of the bacteria against their phenotypic traits to inform their drug discovery platform.

EVOLUTION AND IMMUNITY

The SANBI also devotes much of its research to the problems presented by HIV, looking at the genomics of both the virus and the host from a number of angles. In collaboration with groups in South Africa, the US and Ireland they are studying the role of sugars bound to the surface of a HIV vision in host cell recognition and entry, as well as a target for anti-HIV therapeutics. They are involved in a number of projects with the Centre for the AIDS programme of research (CAPRISA) in South Africa including a collaborative project studying variation in the human immunoglobulin heavy chain variable region genes with the aim of identifying individuals capable of producing broadly cross-clade neutralising antibodies. A third project aims to use their next-generation sequencing analysis pipeline to facilitate high-throughput, low cost HIV drug resistance testing.

But all of these studies represent only a fraction of the diverse, ongoing research projects at SANBI, and their medical work extends far beyond HIV and even parasitic infections, covering studies into plant viruses, genetic determinants of non-communicable diseases and personalised medicine. Underlying all of these studies is a coordinated approach to integrate diverse datasets. The Institute's committed approach to collaboration and training also ensures that the information, practices and resources cultivated as part of these projects are widely shared – SANBI participates in the International Glossina Genome Initiative, which is concerned with assembling the tsetse fly genome. As part of this, the SANBI decided to implement the Glossina Functional Genomics Network, to provide African researchers with bioinformatics training and, more recently, internship opportunities in American and European laboratories.

AGRICULTURAL APPLICATIONS

The institute's research even extends to the genomic investigation of fungal pathogens affecting plants. *Venturia inaequalis*, or apple scab, is a fungal disease that can ruin fruit crops, and so SANBI has been working in collaboration with the Agricultural Research Council for the last three years to assemble and annotate the *V. inaequalis* genome. The team has identified a few classes of genes that are expanded specifically in this organism and not in other fungi, and this discovery provides examples of genes that could be key to the species' pathogenicity. The findings provide a starting point for wet-lab experiments. Moreover, the collaborators have developed a web portal to share the genomic data with other users in the agricultural research community.

THE AGE OF BIOINFORMATICS

SANBI has served the South African community well over the last 18 years, but it seems that its best years are still ahead; as bioinformatics becomes increasingly important in the African research environment, and worldwide, so too SANBI will continue its ascension, providing new scientists, cutting edge research, and practical solutions for Africans along the way.

INTELLIGENCE

SOUTH AFRICAN NATIONAL BIOINFORMATICS INSTITUTE

OBJECTIVES

The South African National Bioinformatics Institute delivers biomedical discovery appropriate to an international and African context. The institute's mission is:

- To conduct cutting-edge bioinformatics and computational biology research relevant to South African, African and global populations
- To develop human resources in bioinformatics and computational biology by educating and mentoring scientists
- To increase awareness of and access to bioinformatics and computational biology resources

KEY COLLABORATORS

National Institute for Communicable Diseases • University of Cape Town • University of Kwa-Zulu Natal • University of Witwatersrand • Centre for the AIDS Programme of Research in South Africa (CAPRISA) • University of Stellenbosch, South Africa • **International Centre for Insect Physiology and Ecology**, Kenya • **Cambridge University**, UK

FUNDING PARTNERS

Atlantic Philanthropies • South African Medical Research Council • South African National Research Foundation • Wellcome Trust • World Health Organization (WHO) • EU Sixth Framework Programme • International Business Machines Inc. South Africa • South African Department of Science and Technology

CONTACT

Dr Alan Christoffels
Director

South African National Bioinformatics Institute
University of the Western Cape
Private Bag X17, Bellville 7535
South Africa

T +27 21 959 2969
E alan@sanbi.ac.za

www.sanbi.ac.za

DR ALAN CHRISTOFFELS is the current Director & DST/NRF Research Chair in Bioinformatics and Public Health Genomics at the South African National Bioinformatics Institute. His bioinformatics lab focuses on host-pathogen interactions and is developing high throughput genomics methods, including next generation sequencing data analysis approaches to study communicable diseases such as tuberculosis, malaria and sleeping sickness.