THE FUTURE OF THE WESTERN CAPE AGRICULTURAL SECTOR IN THE CONTEXT OF THE 4TH INDUSTRIAL REVOLUTION

Review: Bioinformatics

October 2017

University of Stellenbosch Business School

Table of Contents

1. Technology Overview and Detailed Description

Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data and combines computer science, statistics, mathematics, as wells as engineering to analyse and interpret biological data. Bioinformatics has been used for *in silico* analyses of biological queries using mathematical and statistical techniques.

Bioinformatics is both an umbrella term for the body of biological studies that use computer programming as part of their methodology, as well as a reference to specific analysis "pipelines" that are repeatedly used, particularly in the field of genomics. Common uses of bioinformatics include the identification of candidate genes and nucleotides (SNPs). Often, such identification is made with the aim of better understanding the genetic basis of disease, unique adaptations, desirable properties (especially in the agricultural species), or differences between populations. In a less formal way, bioinformatics also tries to understand the organisational principles within nucleic acid and protein sequences, called proteomics[1].

The field of bioinformatics has evolved such that it now primarily involves the analysis and interpretation of various types of data. This includes nucleotide and amino acid sequences, protein domains, and protein structures[2].

Bioinformatics has become an important part of many areas of biology. In experimental molecular biology, bioinformatics techniques such as image and signal processing allow extraction of useful results from large amounts of raw data. In the field of genetics and genomics, it aids in sequencing and annotating genomes and their observed mutations. It plays a role in the text mining of biological literature and the development of biological and genetic data to organize and query biological data. It also plays a role in the analysis of gene and protein expression and regulation. Bioinformatics tools aid in the comparison of genetic and genomic data and more generally in the understanding of evolutionary aspects of molecular biology. At a more integrative level, it helps analyse and catalogue the biological pathways and networks that are an important part of systems biology. In structural biology, it aids in the simulation and modelling of DNA, RNA and protein[3,4,5], including biomolecular interactions[6,7,8].

The field of bioinformatics experienced significant growth starting in the mid-1990s, driven largely by the Human Genome Project and by rapid advances in DNA sequencing technology. The actual process of analysing and interpreting data is referred to as computational biology. Important sub-disciplines within bioinformatics and computational biology include:

- Development and implementation of computer programs that enable efficient access to, use and management of, various types of information
- Development of new algorithms (mathematical formulas) and statistical measures that assess relationships among members of large data sets. For example, there are methods to locate a gene within a sequence, to predict protein structure and/or function, and to cluster protein sequences into families of related sequences.

What is the goal of bioinformatics?

The primary goal of bioinformatics is to increase the understanding of biological processes. What sets it apart from other approaches, however, is its focus on developing and applying computationally intensive techniques to achieve this goal. Examples include: pattern recognition, data mining, machine learning algorithms, and visualization. Major research efforts in the field include sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction, prediction of gene expression and protein–protein interactions, genome-wide association studies, the modelling of evolution and cell division/mitosis.

Analysing biological data to produce meaningful information involves writing and running software programs that use algorithms. The algorithms in turn depend on theoretical foundations such as discrete mathematics, control theory, system theory, information theory, and statistics. Bioinformatics now entails the creation and advancement of databases, algorithms, computational and statistical techniques, and theory to solve formal and practical problems arising from the management and analysis of biological data.

Bioinformatics also should be differentiated from related scientific fields such as biological computation and computational biology. Biological computation aims to develop biological computers using advances of bioengineering, cybernetics, robotics, and molecular cell biology. In contrast, bioinformatics develops and utilizes computational algorithms to understand and interpret biological processes based on genome-derived molecular sequences and their interactions. Therefore, in many aspects, bioinformatics seems similar to computational biology objectives. A computational biology is concentrated on building and/or developing theoretical models for biological analyses, whereas bioinformatics focuses on providing practical tools to organize and analyse basic genomic, proteomic and other "omics" data, including sequence analysis and its visualization[9].

2. Application Examples and Case Studies

The first part of this section of this report will describe the **tools used in bioinformatics**, and is not intended to overwhelm the reader with technical detail. However, for the technically inclined, the literature in the public domain may be consulted, including several reviews.

The second part of this section will deal specifically with **applications of bioinformatics in agriculture.**

Tools used in bioinformatics

The scientific literature on the applications of bioinformatics tools is overwhelming in extent, however, in this section of this report, a selection of the various tools of bioinformatics will be described to illustrate how bioinformatics can be used to extract, and analyse useful information from biological data that has been generated. The areas that will be discussed to illustrate the application of bioinformatics will include analyses of DNA/protein sequences, phylogenetic studies, predicting 3D structures of protein molecules, molecular interactions and simulations[10].

Gene Identification and Sequence Analyses

Sequence analyses refer to the understanding of different features of a biomolecule like nucleic acid or protein, which give to it its unique function(s).

These analyses are popular due to their huge applications in biological sciences, the simplicity, and the capacity to generate a wealth of knowledge about the gene/protein in question. sequence analyses are one of the frequently performed analyses of bioinformatics. Table 1 below shows a list of tools used in primary sequence analyses.

The tools used for primary sequence analyses are described extensively in the literature[11]

Phylogenetic Analyses

Phylogenetic analyses are procedures used to reconstruct the evolutionary relationship among a group of related molecules or organisms, to predict certain features of a molecule with unknown functions, to track gene flow, and to determine genetic relatedness[12]. A phylogenetic comparative analysis is widely used to control for the lack of statistical independence among species[13]. The methods to construct a phylogenetic tree are divided into three major groups: distance methods, parsimony methods, and likelihood methods. None of the methods is perfect; each one has its own strengths and weaknesses, which is beyond the scope of this report, but is referenced in detail elsewhere[14].

Sequence Databases

Biological sequence database refers to a vast collection of information about biological molecules such as nucleic acids, proteins and polymers, each molecule to be identified by a unique key. The stored information is not only important for future use but also serves as a tool for primary sequence analyses. With the advancement of high throughput sequencing techniques, the sequencing has reached to a whole-genome scale, which is generating a massive amount of data every day, at affordable costs. The submission and storage of this information to become freely available to the scientific community has led to the development of various databases worldwide.

Genome Sequence Databases

The GenBank, built by the NCBI (National Centre for Biotechnology Information), is a vast collection of genome sequences of over 250,000 species. The data from GenBank can be accessed through the NCBI's integrated retrieval system, Entrez, while the literature is accessible via PubMed (a publication database). Each sequence carries information about the literature, bibliography, organism, and a set of various other features, which include coding regions, promoters, untranslated regions, terminators, exons, introns, repeat regions, and translations.

Protein Sequence Databases

The most significant protein sequence databases include SWISSPROT (Swiss Protein) Databank, TrEMBL (translation of DNA sequences in EMBL), UniProt (Universal Protein Resource), PIR (Protein Information Resource) and wwPDB (worldwide Protein DataBank).
A selection of the most popular Miscellaneous databases as well as Signalling & Metabolic Pathway Databases, collectively known as genome databases, are listed in the literature[15].

Predicting Protein Structure and Function

Protein molecules begin their life as shapeless amino acid strings, which ultimately fold up into a three-dimensional (3D) structure to become biologically active. The information of 3D structure of a protein is necessary to gain an insight into the function of a specific protein. Some of the tools used to predict the secondary structure of proteins are described in the literature[16].

Molecular Interactions

Proteins seldom perform their functions in isolation, and therefore often interact with other molecules all the time to execute a certain process. Understanding how biomolecules interact with other molecules holds numerous implications, and the understanding of molecular interactions is also essential to elucidate the biological functions of a molecule. Selected tools to study protein-protein interactions is in the literature[17].

An extensive review of **plants databases and economically important crops** was conducted[18], and is recommended as a reference for scientists.

The applications of bioinformatics in agriculture

The genome sequencing of the plants and animals has also provided benefits to agriculture. Tools of bioinformatics are playing significant role in providing the information about the genes present in the genome of these species. These tools have also made it possible to predict the function of different genes and factors affecting these genes. The information provided about the genes by the tools makes the scientists to produce enhanced species of plants which have drought, herbicide, and pesticide resistance in them. Similarly, specific genes can be modified to improve the production of meat and milk. Certain changes can be made in their genome to make them disease resistant.

The sequencing of the genomes of plants and animals will provide enormous benefits for the agricultural community. Bioinformatics tools can be used to search for the genes within those genomes that are useful for the agricultural community and to elucidate their functions. This specific genetic knowledge could then be used to produce stronger, more drought, disease and insect resistant crops and improve the quality of livestock making them healthier, more disease resistant and more productive.

Crop breeding

The goal of plant genomics is to understand the genetic and molecular basis of all biological processes in plants that are relevant to the species. This understanding is fundamental to allow efficient exploitation of plants as biological resources in the development of new cultivars with improved quality and reduced economic and environmental costs. This knowledge is also vital for the development of new plant diagnostic tools. Traits considered of primary interest are, pathogen and abiotic stress resistance, quality traits for plant, and reproductive traits determining yield[19,20]. A genome program is now an important tool for plant improvement. Such an approach to identify key genes and understand their function will result in a "quantum leap" in plant improvement. Additionally, the ability to examine gene expression will allow researchers to understand how plants respond to and interact with the physical environment and management practices. This information, in conjunction with appropriate technology, may provide predictive measures of plant health[21], and to assist in decision-making at the level of commercial planting[22].

Insect Resistance

Many plants have been made insect resistant by incorporating the desired genes. *Bacillus thuringiensis* is a bacterial species which increases the soil fertility and protects plants against pests. When researchers mapped its genome, they used its genes to incorporate into the plant to make it resistant against insects. For example, corn, cotton and potatoes have been made insect resistant. This new ability of the plants to resist insect outbreak may reduce the amount of insecticides being used and hence, the nutritional quality of the crops will be increased[23].

Improved Nutritional Quality

Gene-Diet-Disease interaction of Nutritional genomics aims to study the susceptible genes and provide dietary interventions for individuals at risk of such diseases. Scientists have recently succeeded in transferring genes into rice to increase levels of Vitamin A, iron and other micronutrients. This work could have a profound impact in reducing occurrences of blindness and anaemia caused by deficiencies in Vitamin A and iron respectively[24]. In another example, scientists have inserted a gene from yeast into the tomato, and the result is a plant whose fruit stays longer on the vine and has an extended shelf life[25].

When the changes are made in the genome of the plants, the nutritional value of plants also increases. Golden rice is an important achievement in this endeavour. Here the genes are inserted in the rice genome to increase the Vitamin A level in the crop.
Information obtained from the model crop systems has been used to suggest improvements to other food crops. Arabidopsis thaliana (water cress) *Oryza sativa* (rice), *Triticum aestivum* (wheat) and *Zea mays* (Maize) are examples of available complete land plant genomes.

Plant- pathogen interactions

Organisms that cause infectious disease include fungi, oomycetes, bacteria, viruses, viroids, virus-like organisms, phytoplasmas, protozoa, nematodes and parasitic plants. Plant pathology involves the study of pathogen identification, disease etiology, disease cycles, economic impact, plant disease epidemiology, plant disease resistance, how plant diseases affect humans and animals, pathosystem genetics, and management of plant diseases. Genome sequenced from fungi, oomycetes, bacteria, viruses, viroids, virus-like organisms, phytoplasmas, protozoa, nematodes and parasitic plants gives opportunities to understand plant-pathogen interaction which helps to management, diagnose the disease and make disease resistance transgenic plant. In its most simple form, the gene-for-gene hypothesis states that plants contain single dominant resistance R genes that specifically recognize pathogens that contain complementary avirulence genes. Avirulence genes can be defined as genes in the pathogen that encode a protein product that is conditionally recognized directly or indirectly only by those plants that contain the complementary R gene.

To survive, plants must defend themselves from numerous pathogens. Some defences are constitutive, such as various pre-formed anti-microbial compounds, whereas others are activated by pathogen recognition. The recognition process includes the product of a dominant or semi-dominant resistance R gene present in the plant and the corresponding dominant avirulence (*Avr*) factor encoded by or derived from the pathogen. The recognition of the *Avr* factor by the host plant starts one or more signal transduction pathways that activate several of the plant's defences, thus compromising the ability of the pathogen to colonize the plant. The interactions between plants and pathogens are specific, complex and dynamic[26]. The identification of resistant genes in the germplasm of wild species of field crops and their subsequent integration into commercial cultivars has been the main approach of many plant breeders. Several strategies for the identification, characterization and functional analysis of plant genes involved in the triggering, signaling and response to biotic and abiotic factors have been recently envisaged. In-silico biology therefore plays an important role to understand the plant pathogen interaction at gene and genome of plants and pathogens[27].

Agriculturally Important Microorganisms

With the help of bioinformatics, researchers can under understand the genetic architecture of a microorganism and other pathogens, and to elucidate how these microbes affect the host plant using a metagenomics and transcriptomics approach. The result is the ability to develop pathogen resistant crops, and identify those microbes which are beneficial for the host[28, 29, 30].

Growth in poorer soils

Bioinformatics play an important role to detect the metal from Metagenomic sequencing obtains from contaminated soil[31]. Soil arguably houses the most complex microbial communities because of its ancient history, complex sets of interrelating gradients, and protective, isolating and relatively resource poor and stable physical structure. This results in an incredibly diverse set of gene sequences; at least at the scale soils are normally sampled. The challenge is no longer sequence yield, but the analysis of those sequences, and especially so due to the short sequence products of current sequencing technologies.

Progress has been made in developing cereal varieties that have a greater tolerance for soil alkalinity, free aluminium and iron toxicities. These varieties will allow agriculture to succeed in poorer soil areas, thus adding more land to the global production base. Research is also in progress to produce crop varieties capable of tolerating reduced water conditions[32,33].

Improvement for plant resistance against biotic and abiotic stresses

Genes from *Bacillus thuringiensis* that can control several serious pests have been successfully transferred to cotton, maize and potato. This new ability of the plants to resist insect attack means that the amount of insecticides being used can be reduced. A plant's first line of defence against abiotic stress is in its roots. If the soil holding the plant is healthy and biologically diverse, the plant will have a higher chance of surviving stressful conditions.

Renewable Energy

Plant based biomass is one of the resource for obtaining energy by converting into biofuels such as ethanol which could be used to drive the vehicles and fly the planes. Biomass based crop species such as maize (corn), switch grass and lignocellulosic species like bagasse, and straw are widely used for biofuel production. Sequence variants in biomass-based crop species were detected to maximize biomass production and recalcitrance. Recently, genome of *Eucalyptus grandis* has been released which is also one of major resource of biomass components and all the genes take part in conversion of sugars into biomass components have already been deciphered, therefore provides great insight into mechanisms and pathways responsible for this conversion so in future it will be possible to enhance production of biomass components in eucalyptus and other relevant plants[34]. Hence, the use of genomics and bioinformatics in combination with breeding would likely increase the capability of breeding crop species to be being used as biofuel feedstock and consequently keep increasing the use of renewable energy in modern society[35,36].

Comparative genomics is based on the fact that gene order and content among related plant species are largely conserved over millions of years of evolution[37]. Genome sequencing of several important plants species has enabled researchers to identify 'chromosome' and 'difference' factor in sequences. This in turn has been used to identify value traits for crop improvement. For instance, the barley stem rust resistance gene has been identified from rice-barley comparisons and the sugarcane rust resistance gene based on maize-sorghum comparisons[38]. Comparative genomics could help in achieving improvement of yields in rice, maize, and other related grass crops such as barley, rye, sugarcane and wheat. The ability to represent high resolution physical and genetic maps of plants has been one of the great applications of informatics tools. With several display formats available it has been possible to look for specific positions on the chromosome for large plant genomes.

This kind of cross-genome referencing will lead to a convergence of economically relevant breeding information with basic molecular genetic information. The specific phenotypes of commercial interest that are expected to be dramatically improved by these advances include both the improvement of factors that traditionally limit agronomic performance (input traits) and the alteration of the amount and kinds of materials that crops produce (output traits). Examples include: abiotic stress tolerance (cold, drought, submergence, salt), biotic stress tolerance (fungal, bacterial, viral, chewing and sucking insect attack (feeding)), nutrient use efficiency (manipulation of plant architecture and development (size, organ shape, number, and position, timing of development, senescence)), and metabolite partitioning (redirecting of carbon flow among existing pathways, or shunting into new pathways)[39].

Animal production and animal health

Genomics, proteomics and bioinformatics offer the potential to identify genes or loci regulating traits of economic importance in animals. Production traits that are proposed as eligible for transgenic modification include increased growth rate and improved carcass composition, improved feed utilisation, modified milk composition, improved mohair production improved reproductive performance and increased disease resistance[40]. In veterinary research, bioinformatics tools were used in the detection of new castle diseases and to generate novel solutions for the continued improvement and development of molecular diagnostics[41].

3. Technology or Application Life Cycle: Current Status and Expected Development in 2020 and 2025

Table 1: Heading

Technology Area	Current application in agriculture	Expected applications in agriculture by 2020	Expected applications in agriculture by 2025
Bioinformatics	Crop breeding. Insect Resistance. Improved Nutritional quality. Plant- pathogen interactions. Better understanding of and interaction with agriculturally important microorganisms. Animal production and animal health. Control of infectious diseases in animals.	Improvement for plant resistance against biotic and abiotic stresses. Weather prediction Increased cultivation of crops in poorer soils. Renewable energy applications	Faster detection of disease outbreaks at an early stage globally (not limited by geography). Faster elucidation of causes of disease outbreaks. Risk analysis or a prediction of the future.

4. Business Eco-System View

Bioinformatics has, with time, become integrated with other disciplines, in a multidisciplinary manner, to address the need for solutions that deliver a comprehensive approach to a challenge.

Bioinformatics overlaps with:

- Smart farming
- Genetics
- Sensor technology
- IT and IT Infrastructure

5. Benefits and Risks

The benefits and risks of bioinformatics, from development to application are listed below in point form for ease of reference.

Benefits

- Increase crop performance and yield, and therefore increased productivity (disease resistance, drought tolerance)
- Pest resistant crops – cost effective solutions to problems of economically relevant crops such as cotton, corn and potato through the introduction of beneficial genes

- Improved nutritional value, flavour and texture, through the incorporation of essential elements such as Vitamin A, increased starch content, or beans with more amino acids
- Reduced dependence on pesticides, and therefore increased environmental benefits (such as reduced leaching of chemicals into water reserves), less exposure by farmers to toxic pesticides
- Increased product shelf-life for better market prices and inclusion of smallholder farmers into the mainstream of the agricultural sector
- Manipulated fruit ripening, to exploit better market prices, and reduce competition by ensuring the availability of fruit when no other sources are available; an immediate result is the ability to demand premium prices for fruit
- Climate change studies - there are many organisms which use carbon dioxide as their sole carbon source and increasing levels of carbon dioxide emission is one of the major causes of the global climate change. The study of genomes of these microbial organisms, which is possible through bioinformatics, helps in proposing ways to decrease the carbon dioxide content[42]

Risks

- Unknown health effects (allergens and toxins) due to the scarcity of data on possible (or not) long-term effects (food safety concerns)
- Ecological risks and the requirement for long-term observation
- Lack of consumer acceptance of the products, and the requirement for onerous labelling legislation due to heightened consumer awareness
- Food safety concerns.
- Access to and use of proprietary information (intellectual property)

6. Potential Economic, Social, Ecological (Environmental) and Political Developments and Impacts

The following sections highlight the challenges faced by African farmers in a modern-day context, however, as well as the potential developments and impacts of Bioinformatics.

Economic Developments and Impacts

The market for bioinformatics

According to Reportlinker[43], the global bioinformatics market was valued at $2.9 billion in 2012 and is poised to reach $7.5 billion by 2017 at a CAGR of 20.9%. The growth of the bioinformatics market is driven by decrease in cost of DNA sequencing, increasing government initiatives and funding, and growing use of bioinformatics in biomarkers development processes, amongst others. It is expected that the market will offer opportunities for bioinformatics solutions manufacturers with the introduction and adoption of technologies such as nanopore sequencing and cloud computing. However, factors such as scarcity of skilled personnel to ensure proper use of bioinformatics tools and lack of integration of a wide variety of data generated through various bioinformatics platforms are hindering the growth of the market, and if allowed to prevail, will continue to hinder the growth of the market. Manufacturers of bioinformatics solutions will face further challenges regarding industry consolidation and management of high volume data, if the appropriate infrastructure investments are not made. The major players in the bioinformatics market are Accelrys, Inc. (U.S.), Affymetrix, Inc. (U.S.), Life Technologies Corporation (U.S.), Illumina, Inc. (U.S.), and CLC bio. (Denmark).

The OECD and FAO have predicted that sub-Saharan Africa's consumption and production of agricultural commodities over the period 2016–2025, will place a demand on imported foodstuffs, and that these imports will increasingly meet region's growing demand for food products of higher value commodities. The importation of processed foods may still stimulate job growth in food retailing, but jobs will be lost upstream stages of the food system, namely, on-farm production, agricultural input supply and agri-business services, financial services for the farming sector, storage, and local trading

The bulk of countries in Africa do not cultivate GM maize, but South Africa is the exception, since GM maize can be imported if it is registered domestically. Because of this condition, South Africa is currently unable to acquire white maize from the United States.

Another risk to domestic producers is high price volatility in the region, particularly in key food staple markets, and governments have responded to stabilise prices and supply through continued intervention.

The potential for urbanization and growth in income to stimulate job creation in downstream segments of the food value chain is dependent upon the source of the primary agricultural products. If domestic farm production (due primarily to farm commercialization) can meet rising urban demand, then job opportunities will be presented food assembly, wholesaling, the convenience foods market, processing and retail. However, if domestic production is unable to meet food demand, then imported food (both raw and processed) will take an increasing share of consumers' expenditure. A reduction in barriers to regional trade (i.e. the promotion of intraregional trade) offers an affordable means of reducing domestic price volatility[44]. The suggestion is that an enabling environment which allows intraregional trade to occur more efficiently has immense potential to ensure food security in the region. This is evident in regions where cross border trade is prevalent (Mozambique-Malawi, Malawi-Zambia, Uganda-Kenya), with neighbouring countries, essentially pooling production to stabilise markets[45].

Social Developments and Impacts

Agricultural transformation

In a recent BFAP report [46] changes in the macro and sector level policies that support the observed agricultural transformation, and emerging megatrends that may impact agriculture's contribution to future economic transformation in the Southern African region over the ten years were studied. Marked improvements (due to increased domestic and foreign private investment, increased tax revenues, and increased contribution to GDP) in macroeconomic management will prevail, and continuous and sustainable growth will rely on governments' ability to anticipate and absorb shocks to the system, whilst continuing to implement relevant policies that preserve macroeconomic stability.

African Union (AU) heads of states launched the Comprehensive Africa Agriculture Development Programme (CAADP), a programme of the New Partnership for Africa Development (NEPAD), which was interpreted as Africa's policy framework for agricultural transformation, wealth creation, food security and nutrition, economic growth and prosperity[47]. This was in response to, amongst others, the realization of Africa's population growth, labour demands, and global developments.

A surge in food prices was observed in 2007, and has continued to rise. Consequently, the demand for agricultural land in Africa has risen dramatically. This demand has been accelerated by recent agricultural subsidies and land policies in many countries. The effect of this land acquisition has been a rise in the numbers of medium scale farmers, and large-scale land acquisitions.

Africa has also experienced rising crop diversification. This is due to changing production systems in Africa, driven by rapid economic growth, extensive urbanization, changing consumer preferences when women take control over their own incomes (and make decisions on nutritional needs), as well as globalized markets and their drivers. The implications of crop diversification have been linked to enhanced food security, rural income generation, nutritional diversity needs, and the desire to reduce risks against climate change effects. Evidence for this is seen clearly in the choices made by smallholder famers, particularly those farming on rain-fed land, who have diversified their crop to counter climate change, meet consumer demand (the demand by China and India for cash crops such as soya beans, tobacco, ground nuts for example, and the commercialization opportunities on offer), adapt to the demands of changing from smallholder staple food production to commercial high-value products farming, or simply to avert crop failure[48,49].

Megatrends

According to the BFAP 2016 report referred to elsewhere, the following megatrends are envisaged for the future. These megatrends are interpreted as some of the drivers for investment in bioinformatics, with the aim of addressing these megatrends.

Youth numbers and employment: With over 60% of Africa's population currently under 25 years old, 17 million people will enter the labour force each year over the next decade[50]. Using current employment growth rates, less than 50% of those entering the labour force each year will be absorbed into meaningful off-farm wage-paying jobs. In a more hard-hitting report (McKinsey Global Institute[51]) it was stated that even under the most favourable conditions, the supply of wage jobs in manufacturing, services, and government is not growing fast enough to absorb more than two-thirds of the region's rapidly rising labour force. Therefore, family farming will become important to offer the excluded youth with employment in the agricultural sector and the informal sector.

Consistent with employment trends observed by the Groningen Global Development Centre[52], a recent World Bank report[53] projects that family farming will remain the single largest source of employment for the coming decades. Employment in some non-farm sectors will rise rapidly, but agriculture will dominate in most African countries. New and better crops from bioinformatics should attract youth into agriculture.

Land availability and soil degradation: Major bodies of evidence in Africa point to soil degradation arising from unsustainable cultivation practices in high-density areas of the continent[54,55,56]. Loss of micronutrients and soil organic matter presents with peculiar challenges because it cannot be overcome through the application of fertilisers, and crop output is depressed[57]. Because of continuous cultivation and lack of crop rotations, soil organic carbon levels are at their minimum in high population-density areas. Smallholder farmers in these areas do not benefit from the current yield gains offered by plant genetic improvement due to their farming on depleted soils that do not respond to fertilizer application[58]. Bioinformatics should deliver crops that are able to interact in useful ways with soils, even marginal soils.

Ecological (Environmental) Developments ad Impacts

The BFAP 2016 report also describes climate change as another megatrend envisaged for the future. There is consensus in the literature on two general predictions of climate change, namely, greater variability in agricultural production and a decline in crop productivity[59]. In sub-Saharan Africa, the predominance of rain-fed cropping systems and extensive, pasture-based livestock production means that this region is prone to climatic variability, relative to other regions in the world. Historical climate studies suggest that the region is to recurrent drought. This somewhat explains Africa's low yield, and although this may be the case, the opportunity is in realizing that there is potential to facilitate continued growth in food production, before the region's biophysical limits are reached.

Africa and Latin America are experiencing the world's fastest growth in the farmland under cultivation[60]. However, feeding the global population through expansion of agricultural land will surely promote the degradation of natural ecosystems. The alternative, which is ecological intensification of agriculture would require deliberate access to appropriate technology and its adoption, including sustainable water use through irrigation systems and the implementation of best farming practices.

Human capacity development

Statisticians and informatics experts have become advanced users of bioinformatics software and develop a capability to solve problems locally. To facilitate the process of skills development, biomathematics/bio-computing could be introduced at university level, and the establishment of small software groups and companies should be encouraged through government entrepreneurship programmes.

A challenge remains in the understanding of the results of the data generation and processing activities. At present, advanced bioinformatics is concentrated in a few research centres and private companies around the world that have the capacity to employ personnel with highly specialized training. Although bioinformatics methods are relatively freely accessible through public databases described above, and that here is a gap between the level of skills and technology application in developing countries and the industrialized world, bioinformatics is the enabling technology for several fields of biomedical and agricultural research.

Political Developments and Impacts

National Policy Mandates of South Africa (taken from the DAFF2015/16 to 2019/20 Strategic Plan)

The strategic goals and associated objectives of the Department of Agriculture, Forestry and Fisheries (DAFF), namely,

- effective and efficient strategic leadership, governance and administration;
- enabling environment for food security and sustainable agrarian transformation;
- enhance production, employment and economic growth in the sector,

are a response to achieve the National Development Plan's (NDP) objectives and targets. To align with these priorities, objectives and targets, the department aims to continue providing comprehensive support to all categories of producers. The support will increase the number of people participating in different agricultural sector activities.

In rural areas, the focus is on support to subsistence and smallholder producers in line with the expectations of the NDP, namely that a third of the food surplus should be produced from small-scale farmers or households. The department aims to support targeted land reform beneficiaries as the NDP also supports the land reform objective. DAFF will implement sustainable development programmes that ensure protection of biomes and endangered species, rehabilitation of degraded land and climate change mitigation and adaptation strategies. The department also aims to improve the production efficiencies for smallholder producers. These strategies include organising smallholder producers into commodity-based organisations, increasing their collective bargaining power in negotiations, as well as providing support and training to SMMEs. Implementation of transformation initiatives such as the AgriBEE Charter, Forestry Charter and allocation of commercial fishing rights will promote participation in the economy of the country.

Three key programmes are aligned with the priorities, namely Fetsa Tlala, aimed at massive production of staple foods on fallow land that has the potential for agricultural production; Ilima/Letsema, aimed at supporting sustainable agriculture and promoting rural development for smallholder producers; and LandCare to address land degradation problems and encourage sustainable use of natural resources.

The NDP also states that agriculture has the potential to create close to 1 million new jobs by 2030 through:

- Expanding irrigated agriculture—the 1,5 million ha under irrigation could be expanded by at least another 500 000 ha to 2 million ha • Cultivating underutilised land in communal areas and land-reform projects for commercial production
- Supporting commercial agricultural industries and regions with the highest growth and employment potential
- Supporting upstream and downstream job creation
- Finding creative opportunities for collaboration between commercial farmers, communal farmers and complementary industries
- Developing strategies that give new entrants access to value chains and support.

The New Growth Path (NGP)

The NGP is a national policy which broadly aims to unblock private investment and job creation to address systematic blockages to employment-creating growth (infrastructure, skills, regulatory framework, for example). It focuses on productive sectors and proactively intends to support industries, activities and projects that will generate employment. The NGP has identified job drivers for growth, namely, infrastructure, agricultural value chains, mining value chain, manufacturing, tourism and high-level services, green economy, knowledge economy, social economy, public sector, rural development and African regional development. The NGP manages the job drivers for growth such as in mining, commercial agriculture and smallholders, higher industries, and others.

Industrial Policy Action Plan

The IPAP takes place within the framework of continuous improvements and upscaling of concrete industrial development interventions, as set out in the National Industrial Policy Framework (NIPF). IPAP aims to upscale key interventions over a rolling three-year period, with a ten-year outlook on desired economic outcomes. The NIPF has the following core objectives, namely to:

- Facilitate diversification beyond the economy's current reliance on traditional commodities and non-tradable services that require the promotion of value addition, characterised particularly by the movement into non-traditional tradable goods and services that compete in export markets and against imports.
- Ensure the long-term intensification of South Africa's industrialisation process and movement towards a knowledge economy.
- Promote a labour-absorbing industrialisation path, with the emphasis on tradable labour-absorbing goods and services and economic linkages that create employment.
- Promote industrialisation, characterised by the increased participation of historically disadvantaged people and marginalised regions in the industrial economy.
- Contribute towards industrial development in Africa with a strong emphasis on building the continent's productive capacity and securing regional economic integration.

The Agricultural Policy Action Plan (APAP)

Based on this analysis of the various challenges within the Agriculture, Forestry and Fisheries (AFF) sectors the AFF Strategic Framework was developed to outline appropriate responses to these challenges. The APAP aims to translate these high-level responses offered in the Agricultural, Forestry and Fisheries Strategic Framework, into tangible, concrete steps. The IGDP identifies four broad sector goals:

- Equitable growth and competitiveness;
- Equity and transformation;
- Environmental sustainability; and
- Governance,

which translate into a comprehensive, abiding intervention framework, which will be supported through iterations of APAP via short and medium-term interventions targeting specific value chains (sectoral interventions) or transversal challenges (transversal interventions).

The APAP therefore aligns itself with the New Growth Path (NGP), the National Development Plan (NDP), as well as the Medium Term Strategic Framework in respect of Outcomes 4, 7 and 10.

Annexure 1 illustrates the alignment of bioinformatics with the key policy mandates of DAFF, articulated in the NDP, and APAP, and illustrates where bioinformatics and possibly technologies of the future may be used to support the delivery of the South African government's proposed interventions as articulated in the APAP.

7. Conclusions

Biology is a big-data discipline, and so is the study of Ecology. Major drivers of growth in this field include:

- Next Generation Sequencing
- Imaging
- Remote
- Demand in computer storage, and networking infrastructure
- The development of standards
- Data integration & visualisation

Bioinformatics is also an essential interdisciplinary scientific field to the life science helping to understand the "omics" field and technologies, as well as handling and analysing "omes" data produced. The accumulation of high-throughput biological data due to the technological advances in "omics" fields required and prioritized the use of bioinformatics resources and IT applications for the analysis of complex "Big Data", which would be impractical and useless without bioinformatics. Therefore, as highlighted herein, there is a critical need for the preparation of well-qualified, new generation scientists with integrated knowledge, multilingual ability, and cross-field experience who are capable of using sophisticated operating systems, software and algorithms, and database/networking technologies to handle, analyze, and interpret high-throughput and increasing volume of complex biological data[61], in order to ensure the continued feasibility, and scaling potential of this discipline.

There is a need for research communities to collaborate and share controlled vocabularies and data, as well as the results of analysis. The efforts of the plant ontology(PO) and gene ontology (GO) consortia would assist, and efforts are being carried out internationally to link existing related databases on the globe[62]. These efforts will enable the transfer of knowledge and information on agriculturally related matters and practices, for the benefit of both commercial and smallholder farmers seeking to meet the scalable deliverables of governments. The linking of agricultural information resources would be helpful for whole scientific community around the world in the pursuit of cultivating best variety of crops and plants[63] for inclusive development.

8. Synthesis and key trends from the literature

The growth of the bioinformatics market is driven by decrease in cost of DNA sequencing, increasing government initiatives and funding (both private and public funding), and the growing use of bioinformatics in agriculture. It is expected that the market will offer opportunities for bioinformatics solutions manufacturers with the introduction and adoption of upcoming technologies such as nanopore sequencing and cloud computing.

Process of life science research is changing, and so in time, data mining the public data resources will save time and money. Because of the scale of research, biologists will spend less time in the laboratory, more on the computer. This will create the demand for complex skills, namely Computational Biologists). The advent of bioinformatics will bring about outsourced data generation (i.e. research service companies, in particular NGS services)

Bioinformatics has created new infrastructure requirements for life sciences. These include, but are not limited to data centres with relevant expertise (European Bioinformatics Institute (EU), and the National Centre for Biotechnology Information (USA)). Such infrastructure has also necessitated the requirement for international collaborations, facilitating remote collaborations with data, data integration, analysis and visualization

Cross-disciplinary working, has resulted in new disciplines being created, namely, Computer Science and Electronic Engineering, Mathematics and Computer science, and the development of new informatics-led sub-disciplines such as Cheminformatics, as well as Data managers and Data scientists

End Notes

---

[1] Encyclopædia Britannica. *Bioinformatics*. [Online] Available: https://www.britannica.com/science/bioinformatics [Accessed: 23 October 2017].

[2] Attwood, T.K., Gisel, A., Eriksson, N.E. & Bongcam-Rudloff, E. 2011. Concepts, historical milestones and the central place of bioinformatics in modern biology: A European perspective. [Online] Available: http://cdn.intechopen.com/pdfs/22487/InTech-Concepts_historical_milestones_and_the_central_place_of_bioinformatics_in_modern_biology_a_european_perspective.pdf [Accessed: 8 January 2012].

[3] Sim, A.Y., Minary, P. & Levitt, M. 2012. Modeling nucleic acids: Current Opinion in Structural Biology. *Nucleic Acids/Sequences and Topology*, **22**(3), 273-278.

[4] Dawson, W.K., Maciejczyk, M., Jankowska, E.J. & Bujnicki, J.M. 2016. Coarse-grained modeling of RNA 3D structure: Methods. *Advances in RNA Structure Determination*, **103**, 138-156.

[5] Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A.E. & Kolinski, A. 2016. Coarse-grained protein models and their applications. *Chemical Reviews*, **116**, 7898-936.

[6] Wong, K.C. 2016. *Computational biology and bioinformatics: Gene regulation*. United States: CRC Press (Taylor & Francis Group).

[7] Joyce, A.P., Zhang, C., Bradley, P. & Havranek, J.J. 2015. Structure-based modeling of protein: DNA specificity. *Briefings in Functional Genomics*, **14**(1), 39-49.

[8] Spiga, E., Degiacomi, M.T. & Dal Peraro, M. 2014. Chapter Three - New strategies for integrative dynamic modeling of macromolecular assembly. *Advances in protein chemistry and structural biology*, **96**, 77–111.

[9] Mount, W.D. 2004. *Bioinformatics: sequence and genome analysis*. 2nd ed. New York: Cold Spring Harbor Laboratory Press.

[10] Mehmood, M.A., Sehar, U. & Niaz Ahmad, N. 2014. Use of bioinformatics tools in different spheres of life sciences. *Data Mining Genomics Proteomics*, **5**, 158.

[11] Various authors and websites in Mehmood, M.A., Sehar, U. & Ahmad, N. 2014. Use of bioinformatics tools in different spheres of life sciences. *Data Mining Genomics Proteomics*, **5**, 158.

[12] Khan, F.A., Phillips, C.D. & Baker, R.J. 2014. Timeframes of speciation, reticulation, and hybridization in the bulldog bat explained through phylogenetic analyses of all genetic transmission elements. *Syst. Biol*., **63**, 96-110.

[13] Freckleton, R.P., Harvey, P.H. & Pagel, M. 2002. Phylogenetic analysis and comparative data: A test and review of evidence. *Am. Nat*., **160**, 712-726.

[14] Mehmood, M.A., Sehar, U. & Niaz Ahmad, N. 2014. Use of bioinformatics tools in different spheres of life sciences. *Data Mining Genomics Proteomics*, 5, 158.

[15] Various authors and websites in Mehmood, M.A., Sehar, U. & Ahmad, N. 2014. Use of Bioinformatics Tools in Different Spheres of Life Sciences. *Data Mining Genomics Proteomics*, **5**, 158.

[16] Mehmood, M.A., Sehar, U. & Niaz Ahmad, N. 2014. Use of Bioinformatics Tools in Different Spheres of Life Sciences. *Data Mining Genomics Proteomics*, **5**, 158.

[17] Mehmood, M.A., Sehar, U. & Niaz Ahmad, N. 2014. Use of Bioinformatics Tools in Different Spheres of Life Sciences. *Data Mining Genomics Proteomics*, **5**, 158.

[18] Agarwal, R. & Narayan, J. 2015. Unravelling the impact of bioinformatics and omics in agriculture. *Int J Plant Biol Res*, **3**(2), 1039.

[19] Blum, A. 1988. *Plant breeding for stress environments*. Boca Raton, Fla: CRC Press, Inc.

[20] Xu, Y., 2010. *Molecular plant breeding*. Wallingford, UK: CABI.

[21] Rajamanickam, E. 2012. Application of bioinformatics in agriculture. In Singh, K.M. & Meena, M.S. (eds.), *ICT for Agricultural Development in Changing Climate*, 163-179.

[22] Langridge, P. & Fleury, D. 2011. Making the most of 'omics' for crop breeding. *Trends Biotechnol*, **29**, 33-40.

[23] Betz, F.S., Hammond, B.G. & Fuchs, R.L. 2000. Safety and advantages *of Bacillus thuringiensis*-protected plants to control insect pests. *Regul Toxicol Pharmacol*, **32**, 156-173.

[24] Paine, J.A., Shipton, C.A., Chaggar, S., Howells, R.M., Kennedy, M.J. *et al*. 2005. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat Biotechnol*., **23**, 482-487.

[25] Fraser, P.D., Enfissi, E. & Bramley, P.M. 2009. Genetic engineering of carotenoid formation in tomato fruit and the potential application of systems and synthetic biology approaches. *Archives of Biochemistry and Biophysics*, **483**, 196-204.

[26] Matsumura, H., Reich S., Ito A., Saitoh H., Kamoun S. *et al*. 2003. *PNAS*, **100**, 15718-15723.

[27] Wan, J., Dunning, F.M. & Bent, A.F. 2002. Probing plant-pathogen interactions and downstream defense signalling using DNA microarrays. *Functional and Integrative Genomics*, **2**(6), 259- 273.

[28] Berg, G. 2009. Plant-microbe interactions promoting plant growth and health: perspectives for controlled use of microorganisms in agriculture. *Appl Microbiol Biotechnol*, **84**, 11-18.

[29] McCarthy, F.M., Wang, N., Magee, G.B., Nanduri, B., Lawrence, M.L. *et al.* 2006. AgBase: A functional genomics resource for agriculture. *BMC Genomics*, **7**, 229.

[30] Schenk, P.M., Carvalhais, L.C. & Kazan, K. 2012. Unraveling plant-microbe interactions: can multi-species transcriptomics help? *Trends Biotechnol*., **30**, 177-184.

[31] Handelsman, J. 2004. Metagenomics application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*, **68**, 669-685.

[32] Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K. & Shinozaki, K. 1999. Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat Biotechnol*., **17**, 287-291.

[33] Wang, S., Wan, C., Wang, Y., Chen, H., Zhou, Z., Fu, H. & Sosebee, R.E. 2004. The characteristics of Na^+, K^+ and free proline distribution in several drought-resistant plants of the Alxa Desert, China. J. Arid Env., **56**, 525-539.

[34] . Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., *et al*. 2014. The genome of Eucalyptus grandis. Nature, **510**, 356-362.

[35] Boyle, G. 2004. *Renewable energy*. Oxford: Oxford University Press.

[36] Turner, J.A. 1999. A realizable renewable energy future. *Science*, **285**, 687-689.

[37] Tang, H., Bowers, .J.E., Wang, X., Ming, R., Alam, M., Paterson, A.H. 2008. Synteny and collinearity in plant genomes. *Science*, **320**, 486-488.

[38] Kilian, A., Kudrna, D.A., Kleinhofs, A., Yano, M., Kurata, N. *et al*. 1995. Rice-barley synteny and its application to saturation mapping of the barley Rpg1 region. *Nucleic Acids Res*., **23**, 2729-2733.

[39] Rajamanickam, E. 2012. Application of bioinformatics in agriculture. In K.M. Singh, M.S. Meena (eds.), *ICT for Agricultural Development in Changing Climate*, 163-179. New Delhi: Narendra Publishing House.

[40] Baldassarre, H., Wang, B., Keefer, C.L., Lazaris, A., Karatzas, C.N. 2004, State of the art in the production of transgenic goats. *Reprod Fertil Dev*, **16**(4), 465-470.

[41] Soetan, K.O., Abatan, M.O. 2008. Biotechnology a key tool to breakthrough in medical and veterinary research. *Bio & Mol. Biol. Rev*., **3**(4), 88-94.

[42] Jayaram, B., Dhingra, P. 2010. *Bioinformatics for a better tomorrow*. New Delhi: Indian Institute of Technology.

[43] Reportlinker. 2013. Genomics Industry: Bioinformatics Market by Sector (Molecular Medicine, Agriculture, Research & Forensic), Segment (Sequencing Platforms, Knowledge Management Tools & Data Analysis Services) & Application (Genomics, Proteomics & Drug Design). *Global Forecasts to 2017*. [Online] Available: http://www.reportlinker.com [Accessed: 15 November 2013].

[44] World Bank. 2012. *Africa can help feed Africa – Removing barriers to regional trade in food staples*. [Online] Available: http://siteresources.worldbank.org/INTAFRICA/Resources/Africa-Can-Feed-Africa-Report.pdf [Accessed: 21 October 2017].

[45] Chapoto, A. & Sitko, N. 2014. *Understanding the effects of trade restrictions on maize prices*. Presentation to COMESA/ACTESA Parliamentary Policy Seminar on Import and Export Bans, Lusaka, 11-12 August. [Online] Available: http://fsg.afre.msu.edu/zambia/COMESA_PRESENTATION.pdf [Accessed: 1 November 2017]

[46] BFAP Baseline. 2016. *Agricultural outlook – Putting plans into action: agriculture and economic growth in South Africa 2016-2025*. [Online] Available: http://www.bfap.co.za/documents/baselines/BFAP_Baseline_2016.pdf [Accessed 23 October 2017].

[47] NEPAD. 2016. *The CAADP Country Process*. NEPAD.

[48] Minot, N. 2003. *Income diversification and poverty reduction in the northern uplands of Vietnam*. Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Montreal, Canada.

[49] Delgado, C. & Siamwalla, A. 1997. *Rural economy and farm income diversification in developing countries*. Discussion Paper No. 20. Markets and Structural Studies Division. International Food Policy Research Institute. Washington, D.C.

[50] International Monetary Fund (IMF). 2015. *International Jobs Report*. Economist Intelligence Unit.

[51] Fine, D., Van Wamelen, A., Lund, S., Cabral, A., Taoufiki, M. *et al*. 2012. *Africa at work: Job creation and inclusive growth*. Boston: McKinsey Global Institute.

[52] Groningen Growth and Development Centre (GGDC). 2013. *Africa Sector Database (ASD).* [Online] Available: www.rug.nl/research/ggdc/data/ africa-sector-database [Accessed: 2 November 2017].

[53] Filmer, D. & Fox, L. 2014. *Youth employment in sub-Saharan Africa*. World Bank Publications.

[54] Stoorvogel, J. J. & Smaling, E. 1990. *Assessment of soil nutrient depletion in Sub-Saharan Africa: 1983–2000*. [Online] Available: http://library.wur.nl/isric/fulltext/isricu_i25331_001.pdf [Accessed: 2 November 2017].

[55] Drechsel, P., Gyiele, L., D. Kunze, D. & Cofie, O. 2001. Population density, soil nutrient depletion, and economic growth in sub-Saharan Africa. *Ecological Economics*, **38**, 251-258

[56] Tittonell, P. & Giller. K. 2012. When yield gaps are poverty traps: The paradigm of ecological intensification in African smallholder agriculture. *Field Crops Research*, **143**(1), 76-90.

[57] Vanlauwe, B., Kihara, J., Chivenge, P., Pypers, P., Coe, R. & Six, J. 2011. Agronomic Use Efficiency of N Fertilizer in Maize-Based Systems in Sub-Saharan Africa within the Context of Integrated Soil Fertility Management. *Plant and Soil*, **339**, 35-50.

[58] Tittonell, P. Zingore, S., van Wijk, M.T., Corbeels, M. & Giller, K.E. 2007. Nutrient Use Efficiencies and Crop Responses to N, P, and Manure Applications in Zimbabwean Soils: Exploring Management Strategies across Soil Fertility Gradients. *Field Crops Research*, **100**, 348-368

[59] Schlenker, W. & Lobell, D.B. 2010. Robust negative impacts of climate change on African agriculture. *Environmental Research Letters*, **5**(1), 014010.

[60] Headey, D. 2015. *The evolution of global farming land: Facts and interpretations*. Plenary paper presented at the 29th International Conference of Agricultural Economists, Milan, Italy, August 13.

[61] Abdurakhmonov, I.Y. 2016. *Bioinformatics: Basics, development, and future, Bioinformatics – updated features and applications*. [Online] Available: https://www.intechopen.com/books/bioinforamtics-updated-features-and-applications/bioinformatics-basics-development-and-future [Accessed: 2 November 2017].

[62] Plant Ontology Consortium. 2002. The plant ontology consortium and plant ontologies. *Comp. Funct. Genom.*, **3**, 137-142.

[63] Agarwal, R. & Narayan, J. 2015. Unraveling the impact of bioinformatics and omics in agriculture. *Int J Plant Biol Res*, **3**(2), 1039.