# Why do we need Bioinformatics Degree programmes in Africa?

**Sumir Panji**

# Introduction

- Depending on who you ask, the definition of "bioinformatics" varies e.g:

- Wikipedia: "an interdisciplinary scientific field that develops methods for storing, retrieving, organizing and analyzing biological data. A major activity in bioinformatics is to develop software tools to generate useful biological knowledge." : http://en.wikipedia.org/wiki/Bioinformatics

- Russ Altman: "Generally speaking, we define it as the creation and development of advanced information and computational technologies for problems in biology, most commonly molecular biology (but increasingly in other areas of biology). As such, it deals with methods for storing, retrieving and analyzing biological data, such as nucleic acid (DNA/RNA) and protein sequences, structures, functions, pathways and genetic interactions." : http://www-helix.stanford.edu/people/altman/bioinformatics.html

- From T.K. Attwood and D.J. Parry-Smith's *Introduction to Bioinformatics*, Prentice-Hall 1999 (Longman Higher Education; ISBN 0582327881):

"The term bioinformatics is used to encompass almost all computer applications in biological sciences, but was originally coined in the mid-1980s for the analysis of biological sequence data." : http://www.bioinformatics.org/wiki/Origins_of_bioinformatics

# Introduction

- From Mark S. Boguski's article in the "Trends Guide to Bioinformatics" Elsevier, Trends Supplement 1998 p1:

"The term "bioinformatics" is a relatively recent invention, not appearing in the literature until 1991 and then only in the context of the emergence of electronic publishing... "...However, some of my role models when I was a graduate student (Margaret O. Dayhoff, Russell F. Doolittle, Walter M. Fitch and Andrew D. McLachlan) had been building databases, developing algorithms and making biological discoveries by sequence analysis since the 1960s---long before anyone thought to label this activity with a special term (if anything it was called `molecular evolution'). Even a relatively new kid on the block, the National Center for Biotechnology Information (NCBI), is celebrating its 10th anniversary this year, having been written into existence by US Congressman Claude Pepper and President Ronald Reagan in 1988. So bioinformatics has, in fact, been in existence for more than 30 years and is now middle-aged." :http://www.bioinformatics.org/wiki/Origins_of_bioinformatics

- A paper was even published on the definition of bioinformatics: "**What is bioinformatics? A proposed definition and overview of the field.**" - PMID: 11552348 (2001)

- One idea for a definition: *(Molecular)* **Bio** - **informatics =** is conceptualizing **biology in terms of molecules** (in the sense of physical-chemistry) and then applying **"informatics" techniques** (derived from disciplines such as applied math, CS, and statistics) to understand and **organize the information associated** with these molecules, **on a large-scale. (http://bioinfo.mbb.yale.edu/what-is-it/)**

# Why have a Bioinformatics Programme?

- So Bioinformatics appears to be an amalgamation of traditional fields of computer science, molecular biology, genetics, statistics and mathematics

- Why is it important to have programme that encompasses all these well established fields?

**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# Why have a Bioinformatics Programme?

- So Bioinformatics appears to be an amalgamation of traditional fields of computer science, molecular biology, genetics, statistics and mathematics

- Why is it important to have programme that encompasses all these well established fields?

- **Technology:** All scientific research is driven by technological advances in its own field, which in turn formulates the skills to handle and interpret the new paradigms brought about by these advances

# Why have a Bioinformatics Programme?

- So Bioinformatics appears to be an amalgamation of traditional fields of computer science, molecular biology, genetics, statistics and mathematics

- Why is it important to have programme that encompasses all these well established fields?

- **Technology:** All scientific research is driven by technological advances in its own field, which in turn formulates the skills to handle and interpret the new paradigms brought about by these advances

- A major technological advance in molecular biology is the sophistication, diversity, scale and decreasing cost of the data being generated i.e. high throughput platforms

**H3ABioNet**
Pan African Bioinformatics Network for H3Africa

# Technology and Biological "Big data"

- Human genome sequencing project started in 1990, published in 2003 (13 years)

- 13 years 3 billion bases sequenced at x 10 coverage

- Cost = ~ $3,000,000,000 (including expertise and other scientific endeavours and not sequencing alone (~ $1,000,000,000)

- Old sequencing machines (AB3730xl) used for the human genome project can output 2.8 million bp per 24 hours (in an optimal run)

- 3,000,000,000 / 2,800,000 = 1,071 days for x1 coverage (2.96 years to get x1 sequence coverage of the human genome

http://www.454.com/downloads/news-events/history-of-genome-sequencing_FINAL.pdf
http://www.oxbridgebiotech.com/review/research-and-policy/whats-so-special-about-next-generation-sequencing/

**H3ABioNet**
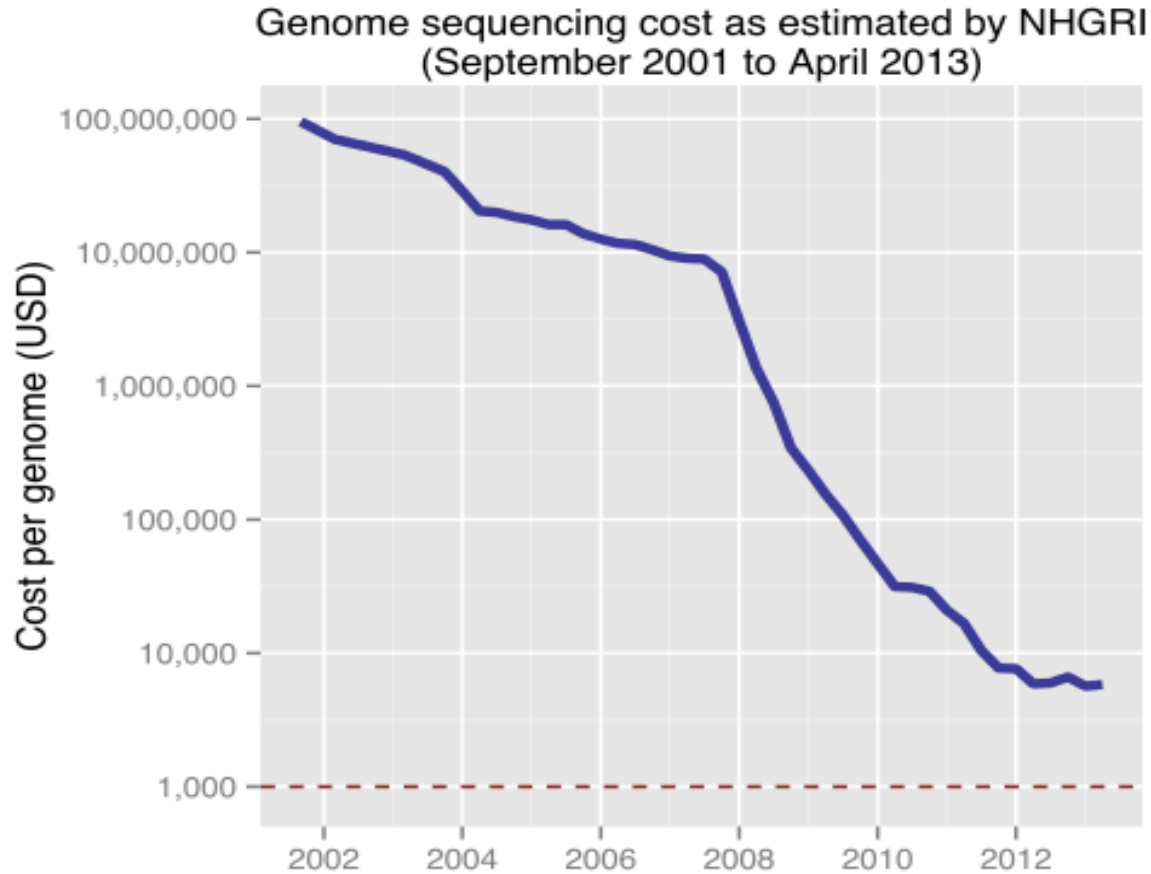Pan African Bioinformatics Network for H3Africa

# Technology and Biological "Big data"

- Next (now!!!) generation technologies can sequence 1 million nucleotides per a second and give x1 coverage of the human genome in 1 hour (Illumina / Solexa sequencing)

- x30 sequence coverage of the human genome achievable in 1 day

- Estimated cost for a human genome sequence is below $10,000

http://www.454.com/downloads/news-events/history-of-genome-sequencing_FINAL.pdf

http://www.oxbridgebiotech.com/review/research-and-policy/whats-so-special-about-next-generation-sequencing/

**H3ABioNet**
Pan African Bioinformatics Network for H3Africa

# Technology and Biological "Big data"



Genome sequencing cost as estimated by NHGRI
(September 2001 to April 2013)

http://en.wikipedia.org/wiki/File:Historic_cost_of_sequencing_a_human_genome.svg
Author = Ben Moore

# Why have a Bioinformatics Programme in General?

- Multi-disciplinary nature of bioinformatics makes it difficult to train individuals by a single department / faculty

- Global shortage of bioinformaticians – demand outstrips supply

- Biggest current bottleneck is not the data, it's the people to help make sense of it

- Personalised medicine is gaining momentum with "translational" or "clinical" bioinformatics (PMID: 23046162)

- Increasing digitisation of genomic, medical and clinical data will make the requirement for bioinformatics more pervasive

**H3ABioNet**
Pan African Bioinformatics Network for H3Africa

# Why have a Bioinformatics Programme in Africa?

- Technology advancement and decreasing costs of platforms makes data generation more "democratic" and accessible to African research groups

- To help define the genomic identity of various populations and the diverse resources within Africa, African scientists must participate in the process

- Scale and quantitative nature of date generated ensures bioinformatics is a pre-requisite for working within the modern day, data driven life sciences field

- Modern day research grants from funding agencies **require** a data management and bioinformatics analysis plan, so can not get away from it

- Imperative to have a programme to train individuals in a systematic manner with comparable measures / outcomes

H3ABioNet
Pan African Bioinformatics Network for H3Africa

# Benefits of a Bioinformatics MSc Programme

- Bioinformatics develops computational, analytical and quantitative data science skills

- Various Faculties working together are able to leverage a diverse range of skills for new collaborations, funding opportunities

- Allows flexibility in research, strengthens the capacity and output of an Institution

- Increasing compute power and decreasing hardware costs plus the quantity of data available makes bioinformatics easily accessible, do not require large overhead costs, consumables i.e. a sequencing factory

- Skills obtained from a bioinformatics degree programme are highly transferable to different industries

**H3ABioNet**
Pan African Bioinformatics Network for H3Africa

# Benefits of a Bioinformatics MSc Programme

- Graduates from bioinformatics programmes can work in:

- Academic research (numerous fields from cancer genomics to statistical methodology development)

- Biotech and pharma

- Health care

- Consultancy

- Forensics

- Software

- High performance computing

- Data management

- Finance / commerce

- Data analysts

- Scientific systems administration

**H3ABioNet**
Pan African Bioinformatics Network for H3Africa

# Why a Bioinformatics Curriculum in Africa?

- Not all bioinformatics programmes are the same, skills and expertise vary widely based on background and training provided

- Is a biologist who uses a piece of software a bioinformatician?

- Is an individual whose whole research thesis comprises of developing an analytical piece of software a biologist (mainly because that is the degree programme that fits "best" at their Institution)?

- Standard outcomes from various programmes that are comparable across Institutions is important

- To have standard outcomes from a bioinformatics degree programme, the standards must first be defined

- Bioinformatics curricula enables these standards to be defined, assessed, compared and accepted as the expected output from a graduate with a bioinformatics degree

## H3ABioNet
Pan African Bioinformatics Network for H3Africa

# Challenges in implementing a Bioinformatics curriculum

- Traditional University departmental / faculty structures hinder the implementation of a multi-disciplinary program like bioinformatics e.g. under Faculty of Computer Science, Statistics, Health Sciences?

- Getting different people from different fields to communicate with each other

- Infrastructure, need a good network, computing, data storage hardware, unix / linux support etc.

- Usually University IT departments are ill equipped to provide support for this (although this would benefit other research fields in the University)

- Experienced faculty to teach (poached by other industries) and Africa's "brain drain" so what is the best model to use?

- What topics are required for bioinformatics?

- How will the degree be accredited?

**H3ABioNet**
Pan African Bioinformatics Network for H3Africa

# Challenges in implementing a Bioinformatics curriculum

- What topics are required for bioinformatics curriculum?

- Who and how will the degree be accredited?

- What Institutional stakeholders are needed to implement a bioinformatics curricula?

- What will be the structure / model of the programme?

- What can we learn from previous experiences and improve on?

- Would it make more sense to build a bioinformatics programme around existing expertise / specialization?

**H3ABioNet**
Pan African Bioinformatics Network for H3Africa

# Challenges in implementing a Bioinformatics curriculum

- Hopefully these are the questions we can address during this workshop

- What would you do with ~ 10 Gigs of data?

**H3ABioNet**

Pan African Bioinformatics Network for H3Africa

# Challenges in implementing a Bioinformatics curriculum

- Hopefully these are the questions we can address during this workshop

- What would you do with ~ 10 Gigs of data?

- **Analyse it or get a bioinformatician!**

# Thank you

**H3ABioNet**

Pan African Bioinformatics Network for H3Africa