

---

# BioKit

stand alone bioinformatics  
platform

---

# New research strategies



# Problem

---

Bad or non existing Internet connection

Limit the progress for African researchers

# Solution

Bioinformatics made easy x

www.ebiokit.eu

**Home**

- Ensembl
- Ensembl bacteria
- Ensembl fungi
- Ensembl metazoa
- Ensembl plants
- Ensembl protists
- BioMart
- Galaxy
- Jalview
- NCBI Blast
- MRS
- RSAT
- wEMBOSS
- WebLogo
- Yasara
- Other tools

eBioKit PROJECT

# What is this?

A powerful but small  
computer

A 10 TB harddisk



# ENSEMBL

www.ebiokit.eu

Home

**Ensembl**

Ensembl bacteria

Ensembl fungi

Ensembl metazoa

Ensembl plants

Ensembl protists

BioMart

Galaxy

Jalview

NCBI Blast

MRS

RSAT

wEMBOSS



Launch Ensembl

Ensembl provides genes and other annotation such as regulatory regions, conserved base pairs across species, and sequence variations. The Ensembl gene set is based on protein and mRNA evidence in UniProtKB and NCBI+ RefSeq+ databases, along with manual annotation from the VEGA/Havana group.

A screenshot of the Ensembl website interface. At the top, there is a navigation bar with links for BLAST/BLAT, BioMart, Tools, Downloads, Help &amp; Documentation, Blog, and Mirrors. A search bar is located on the right side of the navigation bar. Below the navigation bar, there is a search box with a dropdown menu set to 'All species' and a 'Go' button. Below the search box, there is a section titled 'Browse a Genome' with a description: 'The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.' Underneath, there are 'Popular genomes' listed: Human (GRCv37), Mouse (GRCm38), and Zebrafish (Zv9). There is also a link to 'Log in to customize this list' and a section for 'All genomes' with a dropdown menu to 'Select a species'. On the right side of the main content area, there is a 'New to Ensembl?' section with several links: 'Learn how to use Ensembl', 'Add custom tracks', 'Upload and analyse your data', 'Search for a DNA or protein sequence', 'Fetch only the data you want', 'Download our databases via FTP', and 'Mine Ensembl with BioMart'. There is also a 'Did you know...?' section with a link to 'New to Ensembl? Try our online course'. At the bottom, there is a 'What's New in Release 70 (January 2013)' section.

# Genome localisation

7.235.253.122/index.php?location=ensembl

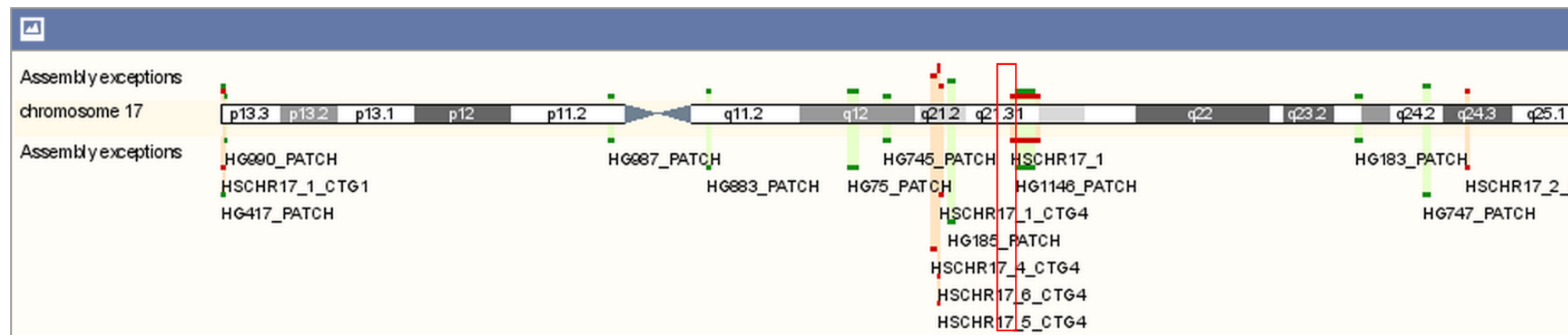
**Ensembl** BLAST/BLAT | Tools | Downloads | Help & Documentation

Search Human...

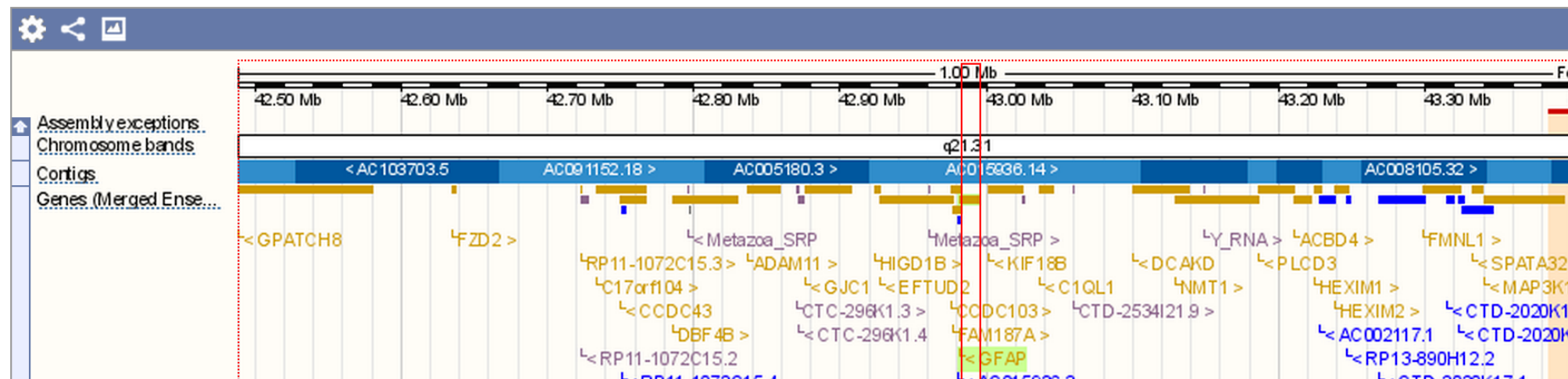
Location: 17:42,982,376-42,994,305

Gene: GFAP

## Chromosome 17: 42,982,376-42,994,305



## Region in detail



Cortigs  
Genes (Merged Ense...

protein coding

AC015336.14 >

< GFAP-025  
processed transcript

< GFAP-002  
protein coding

< GFAP-001  
protein coding

< GFAP-017  
retained intron

< GFAP-024  
protein coding

< GFAP-016  
processed transcript

< GFAP-005  
protein coding

< GFAP-008  
protein coding

< GFAP-014  
retained intron

< GFAP-010  
retained intron

< GFAP-006  
protein coding

< GFAP-015  
protein coding

< GFAP-011  
protein coding

< GFAP-023  
protein coding

< GFAP-019  
protein coding

< GFAP-004  
nonsense mediated decay

< GFAP-012  
processed transcript

< GFAP-020  
nonsense mediated decay

< GFAP-013  
processed transcript

< GFAP-009  
protein coding

< GFAP-021  
protein coding

< GFAP-018  
protein coding

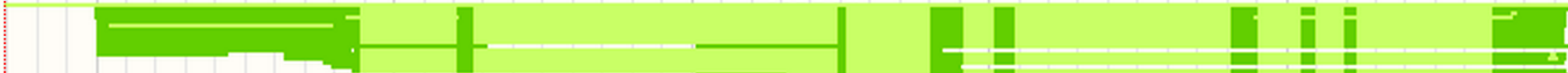
< GFAP-007  
processed transcript

CCDS set

< CCDS11491.1  
CCDS set

< CCDS45708.1  
CCDS set

Human cDNAs (RefS...





# Protein details

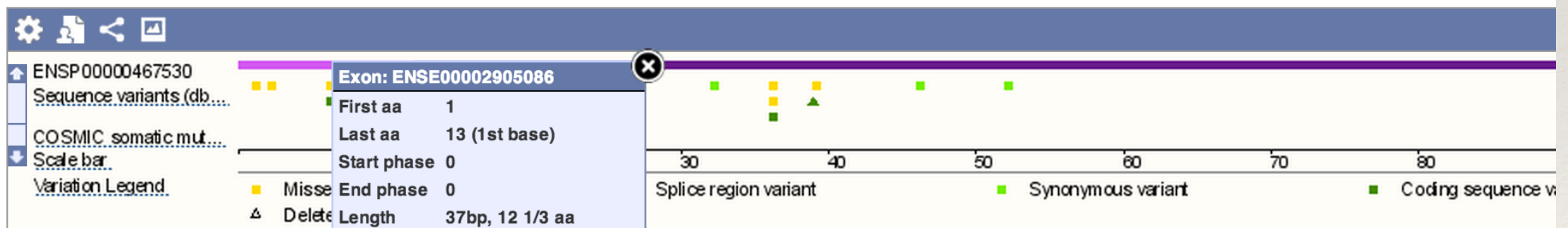
GFAP-012	<a href="#">ENST00000588640</a>	566	No protein product	-	<a href="#">Processed transcript</a>	-	-
GFAP-013	<a href="#">ENST00000591719</a>	585	No protein product	-	<a href="#">Processed transcript</a>	-	-
GFAP-016	<a href="#">ENST00000592706</a>	579	No protein product	-	<a href="#">Processed transcript</a>	-	-
GFAP-025	<a href="#">ENST00000592065</a>	193	No protein product	-	<a href="#">Processed transcript</a>	-	-
GFAP-010	<a href="#">ENST00000590922</a>	556	No protein product	-	<a href="#">Retained intron</a>	-	-
GFAP-014	<a href="#">ENST00000585543</a>	571	No protein product	-	<a href="#">Retained intron</a>	-	-
GFAP-017	<a href="#">ENST00000589701</a>	663	No protein product	-	<a href="#">Retained intron</a>	-	-

## Transcript and Gene level displays

Views in Ensembl are separated into gene based views and transcript based views according to which level the information is more appropriately associated. This view is a transcript level view. To flip between the two sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page.

## Protein summary

Ensembl version ENSP00000467530.1



## Statistics

Charge: 3.0

Isoelectric point: 9.4146

Molecular weight: 11,266.03 g/mol

# Plants

77.235.253.122/index.php?location=plants

EnsemblPlants

BLAST | BioMart | Tools | Downloads | Help & Documentation

 Search Ensembl Plants...

Search:  for

e.g. **Carboxy\*** or **chx28**

Genomes

**Arabidopsis thaliana**

Col-0

**Zea mays**

B72

**Nicotiana glauca**

W3

**Solanum tuberosum**

Species --

[View all Ensembl Plants species](#)



**Oryza sativa**

MSU6



**Brachypodium distachyon**

v1.0



**Physcomitrella patens**

ASM242v1

## Featured content

- This release of Ensembl Plants includes an extensive set of variations for [Hordeum vulgare \(barley\)](#) [1]. Barley serves as a model for adaptation, coping with a range of biotic and abiotic stresses. Here we show variations between the reference sequence and cultivars Barke, Bowman, Igri, Haruna Nijo and a single accession of wild barley (*H. spontaneum*). [Read more...](#)
- A new sequence search service specific to the [Triticum aestivum \(bread wheat\)](#) species has been created. In the absence of a reference genome for wheat, this search service aligns a DNA input sequence against wheat ESTs, publically available in ENA, and has been adapted to link the alignments to the *brachypodium* genomic locations where the wheat sequences map. This mapping between the wheat EST sequences and the *Brachypodium* genome relies on the computed alignments performed using Exonerate. These alignments can be visualized independently on *Brachypodium* location view. [Click here for example.](#)

## References

1. The International Barley Genome Sequencing Consortium (IBSC). [A physical, genome-wide, high-resolution, and gene-based map of the barley genome](#). *Nature*. 2012.
2. Brenchley R, *et al.* [Analysis of the bread wheat genome using whole-genome sequencing](#). *Nature*. 2012.

Ensembl Plants is developed in coordination with other plant genomics and bioinformatics projects.

# Biological databases

Home

Ensembl

Ensembl bacteria

Ensembl fungi

Ensembl metazoa

Ensembl plants

Ensembl protists

BioMart

Galaxy

Jalview

NCBI Blast

## MRS

Launch MRS

This is MRS, a search engine for biological and medical databanks. Use it to search well over a terabyte of indexed text. MRS was designed and implemented by Maarten Hekkelman at the CMBI with the help and contributions from many others. The biological data explosion of the 'omics' era requires fast access to many data types in rapidly growing data banks. The MRS software provides the tools to rapidly and reliably download, store, index, and query flat-file databanks. Data stored and indexed by MRS takes considerably less space on disk than the raw data, despite that these raw data are included. The MRS index information is part of the stored data. Therefore, public and private data can be combined by simple concatenation and thus without computational overheads.

<a href="#">Home</a>	<a href="#">Blast</a>	<a href="#">Align</a>	<a href="#">Status</a>	<a href="#">Databank: sprot</a>	<a href="#">Admin</a>
Search	SwissProt	▼	for	aaa	Search
Records 1-15 of 11,485					1 2 3 4 5 next

# Similarity search

Fast Align Status

All Databanks for

protein sequence in *FastA* format

```

>prot|GFAP_CARAU
GLNDRFASYIEKVRFLFLEQQNKMLVAELNQLRGKEPSRLGDIYQEELRELRRQVDGLNAGKARLEI
DLATLKQRLQEENALRQEAENNLNTRQDVDEAALNRVQLERKIDALQDEISFLRKVHEEEMRQL
QVHVDLDVSKPDLTTALKEIRAQFEAMATSNMQETEEWYRSKFADLTDAAGRNAEALRQAKQEAN
GLTCDLESLRGSNESLERQLREMEERFAIETAGYQDTVARLEDEIQMLKEEMARHLQEYQDLLNV
    
```

to search

complexity


cutoff

options


## Blast results

Nr	ID	Databank	Status
1	gnllsprot GFAP_CARAU	sprot	250 hits found
2	gnlltrembl Q8SQ32_SHEEP	trembl	250 hits found
3	null	sprot	250 hits found
4	gnllsprot GFAP_CARAU	sprot	250 hits found
5	anllsprot GFAP_CARAU	sprot	250 hits

## Hits for gnllsprot|GFAP\_CARAU

ID	Coverage	Description	Hsps	Bits
1 GFAP_CARAU			1	701

Hsp nr	Alignment	Score	Bitscore	E-value	Length	Id
1		1808	701	2.97e-201	365	

Q: 1 REVDRVMGLNDRFASYIEKVRFLFLEQQNKMLVAELNQLRGKEPSRLGDIYQEELRELRRQV 60  
 REVDRVMGLNDRFASYIEKVRFLFLEQQNKMLVAELNQLRGKEPSRLGDIYQEELRELRRQV

# Non coding discovery

77.235.253.122/rsat/

AT

Regulatory  
Sequence  
Analysis  
Tools

Programs

Tools

ance

BL seq

words)

quick)

genes

ols

sets

ry

hing

eq

ilities



BiGRe - ULB

## Regulatory Sequence Analysis Tools

Lab

[Tool Map](#) [Introduction](#) [Forum](#) [Tutorials](#) [Publications](#) [Credits](#) [People](#) [Data](#) [Download](#) [\[Forum\]](#)

Welcome to **Regulatory Sequence Analysis Tools (RSAT)**. This web site provides a series of modular computer programs specifically designed for the detection of regulatory signals in non-coding sequences.

- **New !** Learn how to use **Peak-motifs** with a **Nature Protocol** [\[view article\]](#)
- **Peak-motifs** is now published in **NAR** [\[view article\]](#)
- Latest features of RSAT presented in the **2011 NAR Web server issue** [\[PubMed 21715389\]](#)

- Try our **new programs**
- Check the **latest news** in [our \[Forum\]](#)
- Stay Tuned !! **RSS feed** to all RSAT news
- [How to cite RSAT ?](#)

This website is free and open to all users.

---

**Regulatory Sequence Analysis Tools - Web servers**

Marseille TAGC - France

# Many extra tools

## MEGAN

[Get MEGAN!](#)

MEGAN - *MEtaGenome ANalyzer* provides tools for optimized analysis of large metagenomic datasets.

## plink

[Launch PLINK!](#)

**PLINK** is a free, open-source whole genome association analysis toolset, designed to perform a range of basic, large-scale analyses in a computationally efficient manner. The focus of PLINK is purely on analysis of genotype/phenotype data, so there is no support for steps prior to this (e.g. study design and planning, generating genotype or CNV calls from raw data). Through integration with gPLINK and Haploview, there is some support for the subsequent visualization, annotation and storage of results. PLINK (one syllable) is being developed by Shaun Purcell at the Center for Human Genetic Research (CHGR), Massachusetts General Hospital (MGH), and the Broad Institute of Harvard & MIT, with the support of others.

Home

Ensembl

l bacteria

embl fungi

metazoa

embl plants

ol protists

BioMart

Galaxy

Jalview

CBI Blast

MRS

RSAT

EMBOSS

Uniprot

# GALAXY

77.235.253.122/index.php?location=galaxy

galaxy

Analyze Data

Workflow

Shared Data ▾

Visualization ▾

Help ▾

User ▾

Search tools

ta

ata

Tools

ver

Manipulation

and Sort

Subtract and Group

Text Formats

Features

Sequences

Alignments

Genomic Scores

Work on Genomic Intervals

cs

Text Analysis

Display Data

Genetic Variation

Linear regression

Statistical Analysis

## Welcome to Galaxy

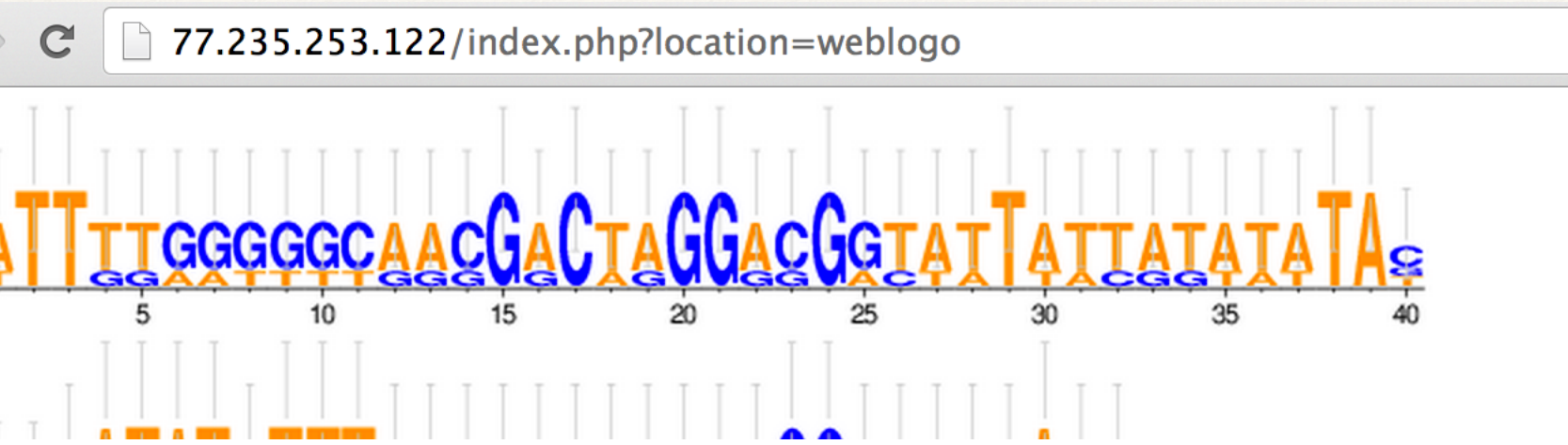
### Introducing Galactic Quickies

Galactic quickies are *super-short* screencasts that are *always* under 5 minutes. We thought it may be a good way to spread the word about Galaxy's functionality while keeping the "annoyance factor" to the minimum. The quickies will be updated weekly.

Sequences as Tab  
delimited data

Galactic quickie # 1

# Also used in Europe



ACCESS Freely available online

 PLOS ONE

## Conserved Motifs in the Clr2 Protein from Fission Yeast, *Schizosaccharomyces pombe*

Christoph Steinhauf<sup>1</sup>, Alejandro Rodriguez<sup>1</sup>, Dimitrios Vlachakis<sup>2</sup>, Gordon Virgo<sup>1</sup>, Vladimir Maksimov<sup>1</sup>,  
Ida Kristell<sup>1</sup>, Ida Olsson<sup>1</sup>, Tomas Linder<sup>3</sup>, Sophia Kossida<sup>2</sup>, Erik Bongcam-Rudloff<sup>4</sup>,  
Björn Bjerling<sup>1\*</sup>

<sup>1</sup>Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Biomedical Research Foundation of the Academy of Health Sciences, Stockholm, Sweden, <sup>2</sup>Department of Biotechnology, University of Athens, Athens, Greece, <sup>3</sup>Department of Biotechnology, University of Copenhagen, Copenhagen, Denmark, <sup>4</sup>Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

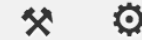
**B**





# Education pages

eBioKit information site



[Welcome to eBioKit information site](#)

[Introduction to Bioinformatics](#)

[How to use the eBioKit](#)

[UNIX command line](#)

[Introduction to Biological Databases](#)

[Basics of Blast search](#)

[Ensembl](#)

[What is EMBOSS](#)

[Multiple Alignments on eBioKit](#)

[Phylogeny](#)

[Galaxy screencasts](#)

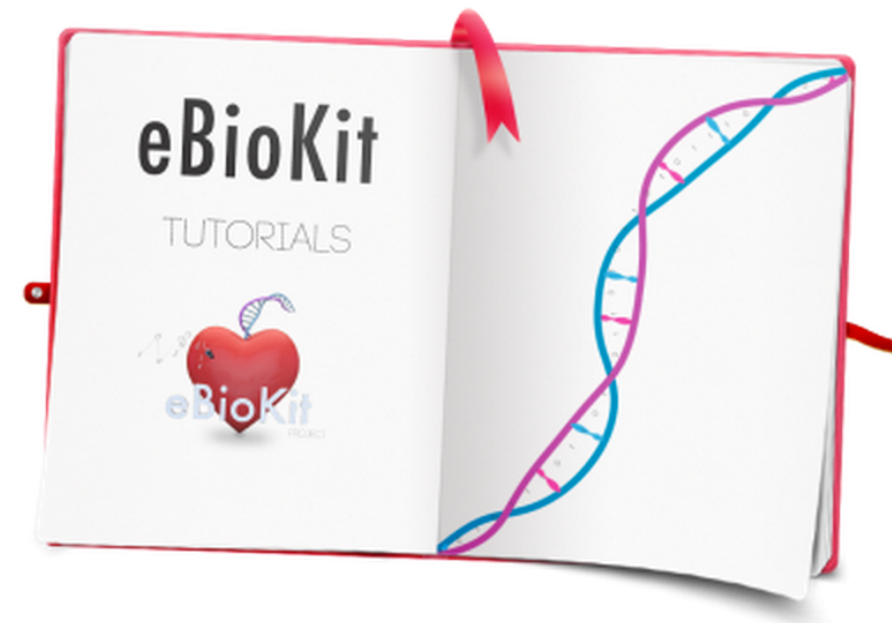
[NGS data analysis with Galaxy](#)

[Personalised mugs are great](#)

## Welcome to eBioKit information site

eBioKit is a compilation of online open source databases and software hosted on the same server. Included in the kit are some tutorials in order to get you started in no time.

- [Introductions to Bioinformatics](#)
- [How to use the eBioKit](#)
- [UNIX command line](#)
- [Introduction to Biological Databases](#)
- [Basics of Blast search](#)
- [Ensembl](#)
- [Introduction to EMBOSS](#)
- [Multiple alignments](#)
- [Introduction to Phylogenetics](#)
- [Galaxy screencasts](#)
- [NGS data analysis with Galaxy](#)



# Financing

## Funding Partners



**SWEDEN**

Swedish Ministry for  
Foreign Affairs



SWEDISH INTERNATIONAL DEVELOPMENT  
COOPERATION AGENCY



Swedish University of  
Agricultural Sciences



**H3ABioNet**

Pan African Bioinformatics Network for H3Africa



# H3ABioNet

Pan African Bioinformatics Network for H3Africa

[About](#)

[Training & Education](#)

[User Support](#)

[Research & Tools](#)

[Events](#)

[Log in](#)

**H3ABioNet Introduction to Bioinformatics using the eBioKit**  
**from 29th of July to 2nd of August 2013 International**  
**Center for Insect Physiology and Ecology (ICIPE), Nairobi,**  
**Kenya**

H3Africa initiative is aimed at studying the genomic and environmental determinants of disease in Africa. The goals of H3Africa are to enhance the necessary genomic expertise among African scientists and to encourage collaborations between African investigators by supporting infrastructure development and research projects. H3ABioNet aims to create a sustainable African Bioinformatics Network to support H3Africa researchers through the development of bioinformatics capacity on the continent.

It is a system for teaching bioinformatics, which is running multiple open source web services on an Apple Mac-mini where all databases are stored locally. This reduces the need for a fast internet connection

[Search](#)

[Latest News](#)

[H3ABioNet Joins the Global Alliance for Genomics and Health](#)

[Helpdesk](#)

[iAnn](#)

# Articles

## ISCB Africa ASBCB Conference on Bioinformatics and BioKit Workshop



**Anne de Villiers<sup>1</sup>, Judit Kumuthini<sup>2</sup>, Erik Bongcam-Rudloff<sup>3</sup>**

<sup>1</sup> I Bioinformatics group, Kenya

<sup>2</sup> Centre for Proteomic and Genomic Research (CPGR), South Africa

<sup>3</sup> Swedish University of Agricultural Sciences, Uppsala, Sweden

The International Society for Computational Biology ([ISCB\[1\]](#)) and the African Society for Bioinformatics and Computational Biology ([ASBCB\[2\]](#)) held the ISCB Africa ASBCB Conference on Bioinformatics in Cape Town, South Africa, in March 2011. The meeting constituted the second joint meeting of ISCB and ASBCB, and the third conference of the ASBCB on the Bioinformatics of African Pathogens, Hosts and Vectors. The conference was preceded by a two-day workshop at the [University of the Western Cape\[3\]](#). ASBCB is a society dedicated to the advancement of bioinformatics and computational biology in Africa. The society works with the [ISCB African Regional Student Chapters\[4\]](#) to provide training courses and a mentorship programme, to help train the current and next generation of African bioinformatics students.



# Collaboration

## GOBLET

Global Organisation for Bioinformatics  
Learning, Education & Training

[Training portal](#)[About us](#)[Join us](#)[Login](#)

[Home](#) / [eBioKit](#)

## eBioKit

Submitted by [bongcam](#) on Thu, 2013-10-03 14:00

Extensive teaching experience gained by conducting bioinformatics training courses in Kenya, Uganda, Mauritius (UoM), SriLanka, Sweden, Chile and Zimbabwe showed that it was difficult to successfully teach and demonstrate several bioinformatics resources, due to and often limited by slow Internet access. For that reason a bioinformatics platform, eBioKit was engineered to ease the administrative burden of regularly updating large databases and installing software. This platform contains more than 300 bioinformatics applications (EMBOSS, Galaxy, Blast, RSAT etc), and most relevant databases (ENSEMBL, UniProt, OMIM, PDB, etc) locally, solving the network speed related problems and problems associated with the installation of software. Version 2 of this system has been successfully tested in real world situations both for capacity building and research in Kenya at ILRI, the Biosciences eastern and central Africa (BecA), EMRI Wellcome Trust Research Programme (KWTRP) and The International Centre of Insect Physiology and Ecology (ICIPE). The system has further been deployed in the SANBio bioinformatics network in



## Training portal

[Trainers and organisers](#)[Training materials](#)[Course pages](#)[FAQ](#)

# Workshops: 25 since 2009

ISCB Africa ASBCB  
5th 2013 Morocco  
12th March

All Official Conference of the International Society for Computational Biology and the African Society for Bioinformatics and Computational Biology



REGISTRATION KEY DATES

INFORMATION

ISCB Africa ASBCB Conference on Bioinformatics  
Agenda (suggested topics subject to change)

ABOUT US

- Brief History
- Message from the Director
- Vision, Mission & Core Values
- Newsletters
- Service Charter
- Annual Report
- Contacts
- Alumni
- Useful Links


RESEARCH / PROJECTS

- Research Projects
- Publications
- Journals
- Mainstream Consultancy
- Teaching & Research Facilities

Home

BIOINFORMATICS

Head of Thematic Area



Prof. James Ochanda

Thematic Area Title and Description

This program is designed to provide sound knowledge of various aspects of

EMBO Global Exchange Lecture Course  
European Molecular Biology Organization

EMBO Courses, Workshops & Conference Series

- HOME
- ORGANISERS
- SPEAKERS
- PROGRAMME
- VENUE
- TRANSPORT
- REGISTRATION

Next Generation Sequencing for Africa

Lecture Course

31 August - 9 September, 2011 | Nairobi | Kenya

Programme

WEDNESDAY, 31-AUG

16:00 Arrival at AICAD

THURSDAY, 01-SEP

9:00-10:00 *Welcome/Introduction* *Etienne de Villiers/Vish Nene*

NEWS AND EVENTS

- Events
- News
- Outreach activities

# Participants: 300-400



# To do

---

Education: enhance and improve the tutorials

Contents:

Add new Genomes: Cassava

Add new more specific tools, e.g. plant specific

Create an organisation for support, education and upgrades



# How To ?

---

Webpage (ip)

Command-line (ssh)

Students accounts were created and are required for  
some tools (e.g. wEMBOSS)