PROGRAMME

MONDAY, MAY 28, 2007: GENOMICS

Morning Session (Chair: Olivier Gascuel)

08.00–09.00 **Registration**

09.00–09.30 **Opening Welcome:**
- Christian Borgemeister, Alia Benkahla, Olivier Gascuel and Daniel Masiga

09.30–10.30 **Conference Keynote**
- Winston Hide (South Africa)
  Translation of genomics in endemic countries
- Group Photograph

10.30–11.00 COFFEE BREAK

11.00–12.00 **Keynote 2:**
- Appolinaire Djikeng (USA)
  Genomics and emerging viral infectious diseases: Influenza and coronaviruses

12.00–13.00 **Poster session 1 (Poster Nos 1–20)**

13.00–14.00 LUNCH

Afternoon Session (Chair: Raphael Isokpehi)

14.00–15.00 **Keynote 3:**
- Enrique Morett (Mexico)
  Taenia genomics

15.00–15.40 **Paper Presentations**
- Nicola Mulder (South Africa)
  Comparative analysis of pathogenic and non-pathogenic microbial genomes
- Karyn Mégy (UK)
  VectorBase, Resource Center for Invertebrate Vectors of Human Pathogens

15.40–16.10 COFFEE BREAK

16.10–17.10 **Paper Presentations**
- Adane Abraham (Ethiopia)
  Characterisation, genome organisation and phylogenetic analysis of two new viruses infecting legume crops in Ethiopia
- Daniel Masiga (Kenya)
  DNA barcoding and taxonomy of Glossina
- Ezekiel Adebiyi (NG)
  A convoluted problem and heuristic solution for predicting ensembles and transitions for pseudoknots polymers

17.30 Cocktail Party at Duduville International Guest House (DIGC) Terrace – *icipe*

TUESDAY, MAY 29, 2007: FUNCTIONAL GENOMICS AND DATABASES

Morning Session (Chair: Nicola Mulder)

09.00–10.00 **Keynote 4:**

- Matthew Berriman (UK)
  Data mining parasite genomes

10.00–10.30 COFFEE BREAK

10.30–11.30 **Keynote 5:**

- Tulio De Oliveira (South Africa)
  Resources and databases for HIV

11.30–11.50
- Gonedelé Bi Sery (Côte d'Ivoire)
  Clade diversity of HIV-1 subtype A pool gene in Côte d'Ivoire: Evidence of underestimated rate of recombinant form

11.50–12.50 **Poster Session 2 (Poster Nos 21–38)**

12.50–14.00 LUNCH

Afternoon Session (Chair: Ikram Guizani)

14.00–15.00 **Keynote 6:**

- Amos Bairoch (Switzerland)
  The annotation of proteins from pathogens in UniProtKB/Swiss-Prot: Current status and future plans

15.00–15.40 **Paper Presentations**

- Sheila Cecily Ommeh (Kenya)
  Analysis of signatures of polymorphisms and positive adaptive selection at candidate genes for genetic resistance to avian viral diseases
- Heikki Lehvaslaiho (South Africa)
  A molecular integration database system for all: Integration of HIV clinical and molecular data using open source genome management tools

15.40–16.10 COFFEE BREAK

16.10–17.10 **Paper Presentations**

- Laurent Bréhélin (France)
  GONNA: A Gene Ontology Nearest Neighbour Approach for the Functional Prediction of Plasmodium falciparum *Orphan Genes: The Database of the Predictions*
- Sondos Smandi (Tunisia)
  Global analysis of Leishmania genes expression using SAGE libraries
- Amel Ghouila (Tunisia)
  Multi-SOM: A novel clustering approach for gene expression analysis

17.10–17.20 **ISCB-SC, Segun Fatumo (Nigeria)**

17.30–18.30 **International *Glossina* Genome Initiative (IGGI) Meeting (Closed Session) (Chairman: Win Hide)**

WEDNESDAY, MAY 30, 2007: EVOLUTION (AM) AND TARGET DISCOVERY (PM)

06.15–13.00 **Visit to Nairobi National Park**

13.00–14.30 LUNCH

Afternoon Session (Chair: Alia Benkahla)

14.30–15.30 **Keynote 8:**

- Michel Tibayrenc (France)
  Trypanosomia, host/parasite/vector coevolution

15.30–16.00 COFFEE BREAK

16.00–17.00 **Paper Presentations**

- Niven Abdel Wahab Salih (Sudan)
  Dynamics of migrations of malaria protective polymorphisms across Africa's Sahel
- Etienne de Villiers (Kenya)
  Identification of Theileria parva *vaccine candidate genes using a bioinformatics approach*
- Lillian Wambua (Kenya)
  Discovery of novel drug targets against pathogenic protozoa: The promise of metabolic reconstruction
- Ikram Guizani (Tunisia)
  LBTx: A tool box for the selection and analysis of potential Leishmania targets for intervention and disease control"

17.10–18.10 **ASBCB Annual General Meeting (AGM) (Chairman: Raphael Isokpehi)**

THURSDAY, MAY 31, 2007: STRUCTURE AND DRUGS

Morning Session (Chair: Appolinaire Djikeng)

09.00–10.00 **Keynote 9:**

- Raphael Isokpehi (USA)
  Information superstructure for protozoan aquaporins

10.00–10.30 COFFEE BREAK

10.30–11.30 **Keynote 10:**

- Mohammad Afshar (France)
  In silico *drug design*

11.30–12.30 **Paper Presentations**

- Marion O. Owolabi (Nigeria)
  Computational discovery of drugs resistance mechanism(s) of the malaria parasite (Plasmodium falciparum (P.f))
- Leo Ghemtio (France)
  Large-scale distributed in silico *drug discovery using VSM-G (Virtual screening manager for computational grids)*
- Vincent Breton/Vinod Kasam (France)
  In silico *docking against malaria: The WISDOM initiative*

12.30–13.30 LUNCH

13.30–14.30 **Poster Session 3 (Special software/Database Session/Poster Nos 30, 33, 35–37)**

14.30–15.30 **Keynote 11:**

- Eric Maréchal (France)
  In silico *strategies for target discovery*

15.30 ***Capacity building Session*** **(Chair: Win Hide, SANBI)**

- Anne Corval, CNRS (Pretoria Office)
  Perspectives of the Africa – France cooperation in bioinformatics

Open session on Capacity Building (with Tea break)

18.00 Conference Dinner: "African Paradise—Safari Park Hotel"

SOFTWARE DEMONSTRATIONS (ILRI CAMPUS)

Transport leaves icipe (DIGC) at 08.00 am, returns from ILRI at 17.00 pm on 1st June and 16.00 pm on 2nd June

Friday, June 1, 2007

09.00–17.30 Amos Bairoch (Switzerland)

Saturday, June 2, 2007

09.00–12.30 Matthew Berriman (UK)
Artemis: A DNA sequence viewer and annotation tool
Artemis Comparison Tool (ACT)

14.00–16.00 Karyn Mégy (UK)
VectorBase

Bioinformatics of African Pathogens and Disease Vectors

ABSTRACTS

Session Theme: Genomics

Translation of Genomics in Endemic Countries

Winston Hide
SANBI: University Western Cape

Abstract: In 1994, TDR helped establish five international parasite genome networks and opened the door for scientists from disease endemic countries to participate and collaborate in genome and post-genome projects. A great deal of parasite, vector, and host genome data has been produced. The challenge now, is developing understanding to allow translation of genome sequence in order to deliver novel drug targets, structural predictions, characterisation of biodiversity, reconstruction of metabolic pathways, systems biology and identification of vaccine candidate molecules. In order to take these areas beyond the concept of buzzword, a great deal of knowledge transfer has to occur. This presentation will provide context to bioinformatics research in Africa, outline the challenges inherent in understanding of genome data, explore strategies, and provide examples of 'translation'.

Genomics and Emerging Viral Infectious Diseases: Influenza and Coronaviruses

Appolinaire Djikeng
Viral Genomics Group; The Institute for Genomic Research (TIGR), A Division of the J. Craig Venter Institute (JCVI); 9704 Medical Center Drive, Rockville, Maryland 20850, USA

Abstract: The majority of emerging diseases threats are of zoonotic origin and of these pathogens, the overwhelming majority are RNA viruses. Examples of important viral threats include agents such as Influenza, SARS, Ebola, Dengue, and Hantavirus. Genomic approaches provide a window on viral evolution and facilitate the study of viral ecology and the relationship between humanity and zoonoses. Due to the relative simplicity of RNA viral genomes the compilation of viral sequence data allows mass comparative genomic analysis and association with epidemiological data at a scale which is currently impossible for other organisms.

We are specifically interested in Influenza and Coronoviruses that have the ability to cause diseases that significantly affect both humans and animals. In addition we use novel and high throughput genome sequencing and analysis of viral samples in the context of emerging viral infectious diseases. Our technology platform and results will be presented and discussed.

The Genome Project of Taenia solium[1]

E. Morett, A. Garciarubioc, R. J. Bobesa, J. C. Carreroa, M. A. Cevallosb, G. Fragosoa, P. Gaytánc, V. M. Gonzálezb, M. V. Joséa, L. Jiménezd, A. Landaa, C. Larraldea, J. Morales-Montora, E. Sciuttoa, X. Soberónc, P. de la Torrea, V. Valdése, J. Yánezc and J. P. Laclettea*
a Instituto de Investigaciones Biomédicas, bCentro de Ciencias Genómicas, cInstituto de Biotecnología, dFacultad de Medicina and eFacultad de Ciencias, Universidad Nacional Autónoma de México
*Project Coordinator.

Abstract: We have constituted a consortium of key laboratories at the National Autonomous University of Mexico to carry out a full genomic project for Taenia solium. This project is providing invaluable resources for the study of taeniasis/cysticercosis, and, in conjunction with the *Echinococcus granulosus* and *E. multilocularis* genome projects of expressed sequence tags (EST's), is marking the advent of cestode parasites genomics. The *T. solium* genome size was estimated at about 270 Mb by cytofluorometry on isolated cyton nuclei. In contrast, several probabilistic calculations based on shotgun sequenced genomic clones, resulted in size estimates of 120–140 Mb. The sequencing is being carried out through a combined strategy of 454 pyrosequencing and traditional capillary

sequencing. We consider a 20X coverage by pyrosequencing and at least 3X by capillary sequencing from small insert shotgun and fosmid end libraries, to obtain a 95% completeness of the genome.

About 40,000 ESTs have been obtained; 14,113 adult and 9,157 larva distinct ESTs were made public through GenBank. Unique genes were identified by clustering all EST-fragments with an assembler (minimus). Out of the initial 23,290 ESTs, 19,067 were incorporated into 2,564 genes (contigs). We have ~6,800 "genes", including the contigs, plus 2,592 larva and 1611 adult ESTs that remained as solitons. We have found highly expressed genes in both adult and larvae ESTs. There are 349 "genes" with 10 or more sequences that account for 50% of all transcripts. Many are differentially expressed. Approximately 1/3 (2,038) of the 6,800 genes have a significant match in SwissProt. Gene Ontology (GO) and Gene Ontology broad categories (GoSlim) can be assigned to 36% (2448 genes) of our genes. About 27% of the genes have no match in SProt + TREMBL, and could constitute new genes. The highest expressed genes support a very active fat and carbohydrate metabolism.

As for the genomic sequences, we are analysing correlated polymorphism: sequences that share non-consensus nucleotides at different positions. These can hint to heterozygosis, gene duplications, and alternative splicing. Our results suggest that the genome of *T. solium* is not highly repetitive (< 7%). One small (53 bp) tandem-repeat represented 0.5% of the genome. Different tetranucleotide repeats of the form (TXXX) account for about 4.5% of the genome.

Key words: Parasite, genomics, *Taenia solium*, cysticercosis

[1]*This project is supported by a special grant IMPULSA-UNAM.*

Comparative Analysis of Microbial Genomes to Study Expanded Gene Families in *Mycobacterium tuberculosis*

N. Mulder, H. Rabiu, G. Jamieson, V. Vuppu

Abstract: *Mycobacterium tuberculosis*, the causative agent of tuberculosis, is the leading infectious disease agent, causing millions of deaths annually. The incidence of disease is increasing with the AIDS pandemic, and current vaccines and therapies are not 100% efficient, resulting in the emergence of drug resistance. The ability of the organism to evolve resistance to drugs with enhanced pathogenicity appears, at least in part, to be provided by the mechanism of gene duplication. This evolutionary mechanism results in expansion of gene families, thereby providing the organism with extra copies of the gene and thus the opportunity to evolve new functions. This project aimed to identify the expanded gene families in *M. tuberculosis* and investigate the potential contribution of gene duplication events to pathogenicity. Comparative genomics tools were used to compare the *M. tuberculosis* proteome to itself and the full protein complement from over 80 other microorganisms, including both pathogens and non-pathogens, to determine the extent of family expansion in each. We investigated the link between gene duplication, genome size and GC content, and selected expanded families from *M. tuberculosis* that were unique to either *M. tuberculosis* or pathogens for further analysis. Up to half of all *M. tuberculosis* proteins belong to expanded families or gene duplicate sets, some of which are unique to this organism or to *M. tuberculosis* complex organisms, suggesting that they have a role to play in evolution of these genomes. Although the evolution of *M. tuberculosis* is thought to be relatively recent, the maintenance of these duplicated families in the genome suggests they have a role to play in the pathogenic lifestyle of the organism.

Characterization, Genome Organization and phylogenetic Analysis of Two New Viruses Infecting Legume Crops in Ethiopia

Adane Abraham[1], M. Varrelmann[2] and H. J. Vetten[3]

[1]Ethiopian Institute of Agricultural Research, National Plant Protection Research Centre,
[2]University of Goettingen, Institute for Plant Pathology and Protection, Goettingen, Germany
[3]Federal Biological Research Centre for Agriculture and Forestry, Institute for Plant Virology, Microbiology and Biosafety, Messeweg, Germany

Abstract: While attempting to investigate the viruses associated with yellowing and stunting diseases of cool season food legumes in Ethiopia, two new viruses were discovered and characterised. A distinct luteovirus tentatively named as Chickpea chlorotic stunt virus (CpCSV) was found to

be associated with most legume crop samples namely faba bean, chickpea, lentil, grass pea and fenugreek collected from Ethiopia. Biological studies indicated that CpCSV is transmitted by the aphid *Aphis craccivora* but not by *A. fabae, Acyrthosiphon pisum* or *Myzus persicae* and has an experimental host range limited to species of cool season food legumes. Electron microscopy showed isometric particles of 28 nm in diameter typical of luteoviruses. A rabbit polyclonal antiserum and 10 mouse monoclonal antibodies were produced, characterised and evaluated for routine use in its diagnosis. Sequencing of the complete genome of an Ethiopian (Ambo) isolate revealed that its linear RNA genome has 5900 nucleotides arranged in six major open reading frames typical to members of genus *Polerovirus* in the family *Luteoviridae.* Phylogenetic analysis of the polymerase and coat protein (CP) sequence using Clustal-X program confirmed this classification. Recombination analysis of the genome using Sister scanning procedure suggested that although the virus genome is generally close to *Cucurbit aphid-borne yellows virus*, it has acquired a stretch of about 90 amino acids at the C-terminal part of its readthrough domain from ancestral Soybean dwarf virus-like ancestor. The CP protein gene sequence of 18 CpCSV isolates from five countries indicated that the virus occurs as two geographically differentiated strains significantly differing in symptom severity, serological properties and nucleotide sequence information.

Serological analysis of 73 nanovirus isolates from faba bean samples collected from Ethiopia indicated that the isolates could be categorised into at least three distinct serotypes designated A, B and C. The nucleotide sequence of at least two genes encoding for the CP and U1 protein were obtained for representative samples of each serotype. Whereas serotype A and B isolates were similar to those reported elsewhere previously, Serotype C represented a new group. To understand the phylogeny of the later in detail, the complete sequence of the eight distinct circular single stranded DNAs believed to form entire nanovirus genome was obtained from a representative sample (Eth-231). Each of these DNAs was about 1 kb in size and code for proteins of 12–35 kDa. Sequence comparison indicated that overall nucleotide similarity of only 70% with *Faba bean necrotic yellows virus*, its closest relative, suggesting that it is a distinct nanovirus belonging to family *Nanoviridae*. For this apparently new virus, a tentative name Faba bean yellow leaf virus (FBYLV) is proposed.

DNA Barcoding and Taxonomy of *Glossina* spp.

Daniel Masiga[1] and Johnson Ouma[2]
[1]Molecular Biology and Biotechnology Department, *icipe*, P. O. Box 30772-00100 Nairobi, Kenya
[2]Trypanosomiasis Research Centre, Kenya Agricultural Research Institute, P. O. Box 362, Kikuyu

Abstract: The study of the broad field of biology relies extensively on the ability to classify and identify individuals to the lowest taxonomic level. It is estimated that 250 years of dedicated taxonomic activity has resulted in an impressive 1.5–1.8 million descriptions, representing only about 2% of all species, estimated at 5–100 million. Unfortunately, expertise in taxonomy based entirely on morphology is becoming more difficult to find, and largely resident in national life science museums. Identifying species accurately is foundational to the ecology of insects that spread disease, and the pathogens they transmit. Although application of unique DNA sequences is not new to species identification, the application of a standardised locus for implementation on a broad scale is, and continues to elicit much debate. The Consortium for the Barcode of Life (CBOL, www.barcoding.si.edu) has identified mitochondrial Cytochrome Oxidase I (COI) as the barcode sequence for animals. Using universal primers described by Folmer et al (1994), the barcoding region (as defined by CBOL) we have obtained amplicons from 6 species of tsetse (*G. fuscipes fuscipes*, *G. pallidipes*, *G. swynnetoni*, *G. morsitans morsitans*, *G. m. centralis* and *G. m. submorsitans*). Under our PCR conditions, *G. brevipalpis* and *G. longipennis* failed to amplify, hence more optimisation is needed. However, all the 6 produced high quality sequences in excess of 650 bp from the PCR products. These data demonstrate the utility of COI in identifying tsetse flies to the level of sub-species.

A Convoluted Problem and Heuristic Solution for Predicting Ensembles and Transitions for Pseudoknots Polymers

Ezekiel F. Adebiyi
Department of Computer and Information Sciences, Covenant University, PMB 1023, Ota, Nigeria

Abstract: Note that tRNAs provides potential drug targets and are readily available. Presently, little is known about the structures of tRNA in *Plasmodium falciparum*, so development of computational techniques toward the prediction of tRNA structures will enhance the work on structural based drug design for malaria treatments. Chen and Dill (PNAS, 97, 646-51, 2000) developed a method capable of predicting structures, but its can not work on tRNAs. Lucas and Dill (*Journal of Chemical Physics*, 119(4), 2414-21, 2003) and Kopeikin and Chen (*Journal of Chemical Physics*, 122(9), 1–13, 2005) obtained one independently but their techniques worked only for simple structures. Infact, the prediction of Lucas and Dill overestimate in its prediction. In the work that follows, we took a broader look at a larger class of structures and developed a heuristic method for the resulting convoluted problem. Mainly, our technique targets extending the theory on polymer graphs that exist (PNAS, 97, 646–651, 2000) to include pseudoknots. Our experimentation produced enhanced results. Our future work focused on the improvement of our present framework.

Key words: Malaria, Drugs design, tRNA, Polymer graphs, Pseudo-knot, Conformation, Graph counting, RNA folding, Drug docking.

Session Theme: Functional Genomics and Databases

Data Mining Parasite Genomes

Matthew Berriman (UK)

Resources and Databases for HIV

Tulio De Oliveira (South Africa)

Clade Diversity of HIV-1 Subtype A Pol Gene in Côte d'Ivoire: Evidence of Underestimated Rate of Recombinant Form

Gonedelé Bi Sery[1,2], Sangaré Abdourahmane[1], Cissé Gueladio[2], Gnangbé Félix[1], N'Goran Eliezer[2,3]
[1]Département de génétique, Université d'Abidjan-Cocody, Côte d'Ivoire
[2]Centre Suisse de Recherches Scientifiques en Côte d'Ivoire
[3]Département de parasitologie, Université d'Abidjan-Cocody, Côte d'Ivoire

Abstract: Based on partial env and pol (protease and RT) subtyping, it was recently documented that the majority (>80%) of the HIV-1 strains that circulate in Côte d'Ivoire were CRF02_AG and about 11% could not be clearly assigned to a known subtype or CRF (Circulating Recombinant Form). Thus many subtypes A sequences from this country submitted to HIV genome databases are still not assigned to any sub-subtype. So the number of subtype A sequences not assigned to any sub-subtype continues to grow. Inter and intra clade variations within pol sequence are particularly relevant as this region encodes RT and protease proteins, against which many antiviral drugs are directed.

We attempted to identify and characterise HIV-1 subtype A using all pol gene sequences of this subtype isolated from Côte d'Ivoire and registered in the NCBI GenBank and Los Alamos database (n = 85). Rega was used for subtype identification and Simplot was used for Bootscan analysis to detect recombinant forms. A maximum-likelihood tree of pol gene sequences was produced by DNAML program in the PHYLIP package version 3.6 using a transition/transversion ratio of 1.5 and empirical base frequencies. Sequences were gap stripped. Pairwise genetic distances were calculated by using the DNADIST program in PHYLIP with an F84 model of evolution and a transition/transversion ratio previously defined. Calculations were done with the reference sequences A1, A2, A3, and AG.

All the clade previously defined as subtype A clustered with the recombinant forms CRF02_AG and AG. Distance calculations between these sequences and reference sub-subtypes A1, A2, A3 (7–11.5%) and CRF02_AG (1.7–3.8%) fell in the range of distances previously characterised between and within sub-subtype groups. Our results provide insights into the importance of recombinant

forms in Côte d'Ivoire which seem have been underestimated, and thus indicate that more than 90% of HIV-1 strains are recombinant forms.

Key words: HIV-1, subtype A, pol gene, phylogeny, recombinant form, Côte d'Ivoire.

The Annotation of Proteins from Pathogens in UniProtKB/Swiss-Prot: Current Status and Future Plans

A. Bairoch
Swiss Institute of Bioinformatics
amos.bairoch@isb-sib.ch

Abstract: The Swiss-Prot knowledgebase [1] was created in 1986. It is now the cornerstone of the UniProt consortium [2] efforts. We will briefly describe what Swiss-Prot has to offer to the life scientists. The major part of the presentation will be devoted to our current efforts in annotating proteins originating from microbial and viral pathogenic organisms. We will also describe our plan to initiate a concerted effort for a distributed annotation project concerning proteins originating from eukaryotic pathogens.

References: [1] Bairoch A., Boeckmann B., Ferro S., Gasteiger E. Swiss-Prot: Juggling between evolution and stability. Briefings Bioinform. 5, 39–55(2004). [2] Wu C. H., Apweiler R., Bairoch A., Natale D. A., Barker W. C., Boeckmann B., Ferro S., Gasteiger E., Huang H., Lopez R., Magrane M., Martin M. J., Mazumder R., O'Donovan C., Redaschi N., Suzek B. The Universal Protein Resource (UniProt): An expanding universe of protein information. Nucleic Acids Res. 34, D187–D191(2006).

VectorBase, Resource Center for Invertebrate Vectors of Human Pathogens

Karyn Mégy, P. Arensburger[1], P. Atkinson[1], N. Besansky[2], R. Bruggner[2], R. Butler[2], K. Campbell[3], G. Christophides[4], S. Christley[2], E. Dialynas[5], D. Emmert[3], M. Hammond, C. Hill[6], R. Kennedy[2], D. Lawson, N. Lobo[2], G. Madey[2], R. MacCallum[4], S. Redmond[4], S. Russo[3], D. Severson[2], E. O. Stinson[2], P. Topalis[5], E. Zdobnov[7], E. Birney, W. Gelbart[3], C. Louis[5,8], F. Kafatos[4] and F. Collins[2]
European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD UK [1]University of California, Riverside, 900 University Avenue, Riverside, CA. 92521 USA
[2]Center for Global Health and Infectious Diseases, University of Notre Dame, Notre Dame IN 46656-0369, USA
[3]The Biological Laboratories, 16 Divinity Avenue, Harvard University, Cambridge, MA 02138 USA
[4]Cell and Molecular Biology Department, Imperial College London, South Kensington Campus, London SW7 2AZ UK
[5]Institute of Molecular Biology and Biotechnology, FORTH, Vassilika Vouton, P. O. Box 1385, Heraklion, Crete Greece
[6]Department of Entomology, Purdue University, West Lafayette, IN 47907 US
[7]Department of Genetic Medicine and Development, University of Geneva Medical School, 1 rue Michel-Servet, 1211 Geneva Switzerland
[8]University of Crete, Heraklion, Crete Greece

Abstract: Until recently, the malaria vector *Anopheles gambiae* was the only sequenced and annotated mosquito. But in 2006, the genome of *Aedes aegypti,* the yellow fever vector mosquito, has been released and annotated. Complementing the gene sets, a large amount of data was available: expression and comparison data, domain content information (signal peptide, transmembrane and protein domains) and control vocabulary. With the upcoming sequencing of more disease vectors (*Culex pipiens, Ixodes scapularis, Glossina morsitans etc.*), it is becoming importance to organise the storing and access to this data. VectorBase is a Resource Centre for Invertebrate Vectors of Human Pathogens and regroup information about these organisms: gene sets and related information, sequences (genomic, ESTs, traces) and pictures. But more than storing, VectorBase is a place for mining data: walk along the genomes, visualise DNA and protein similarities between organisms, compare sequences *vs.* the hosted genomic sequences (blast) and build alignments (ClustalW). VectorBase aims to be a main resource centre for the invertebrates disease vector communities, involving the scientists, generating, updating and giving an easy access to the data [http://www.vectorbase.org/index.php].

A Molecular Integration Database System for All: Integration of HIV Clinical and Molecular Data Using Open Source Genome Management Tools

Heikki Lehvaslaiho*[1], Adam Dawe*[1], Allan Kamau[1], Ruby van Rooyen[1], Tulio de Oliveira[1], Anelda Boardman[1], Alan Powell[1], Salim Abdool Karim[2], Koleka Mlisana[2], Lynn Morris[3], Clive Gray[3], Carolyn Williamson[4] and Winston Hide[1]

[1]SANBI, UWC, Bellville, South Africa

[2]Centre for AIDS Programme of Research in South Africa (CAPRISA), Nelson R Mandela, School of Medicine, Durban, South Africa

[3]National Institute for Communicable Diseases, Johannesburg, South Africa

[4]Institute for Infectious Diseases and Molecular Medicine, University of Cape Town, South Africa

Abstract: HIV studies can combine long term clinical monitoring, sequencing of large number of whole or partial viral genomes to assess viral diversity, multi-institutional molecular laboratory assays including immune monitoring, and host genetics. Integrating this form of data across molecular and non-molecular domains is a challenge. In order to maximise the rate at which discoveries can be achieved in our study, we have developed a strategy of leveraging global open source developments to apply and refine in a developing world context. Capturing contextual information together with time series genome sequences from infected subjects is non-trivial. As part of an NIH acute infection study of the Centre for AIDS Programme for Research in South Africa, we have developed a Molecular Integration Database (MID). It securely and reliably integrates clinical, immunological and molecular data from infected subjects. The MID leverages open source software developed for model organism genomes such as BioMart for querying and the Generic Genomic Browser (GBrowse) for visualisation. Use of genome project software means that the resultant saving of time and resources in the database's development has been substantial and in return, the developing world, where infection is endemic, and its open source community will benefit from the release of the source code of the MID and its HIV sequence assembly pipeline.

GONNA: A Gene Ontology Nearest Neighbour Approach for the Functional Prediction of *Plasmodium falciparum* Orphan Genes: The Database of the Predictions

L. Brehelin, J-F. Dufayard and O. Gascuel

Abstract: *Plasmodium falciparum*, the pathogenic agent responsible for malaria, causes close to 3 millions human deaths each year in the world. Its genome, published in 2002, remains poorly understood. Among its ~5000 genes, ~3000 have a totally unknown biological function. Several reasons explain this situation, the first one being a totally atypical genome composition which renders ineffective the usual methods of functional annotation based on sequence comparison (alignment) with homologous genes in nearby organisms.

Non-homology methods are thus needed to obtain functional clues for the orphan genes. Notably, transcriptomic analysis using DNA microarrays have been proposed. These approaches are based on the hypothesis that uncharacterised genes may potentially share functional roles with annotated genes of similar transcriptomic profile, a principle known as "Guilt by Association" (GBA). Interestingly, this intra-species principle can also be apply to other type of available postgenomic data, such as proteomic data or protein interaction data.

The GO Consortium (http://www.geneontology.org) has developed a systematic and standardised nomenclature for annotating genes in various organisms, including *P. falciparum*. The GO annotation is a hierarchical structure describing generalisation relationships between hundreds of terms. So each gene can have several associated GO terms, due to the hierarchical structure of the ontology, and of the existence of multi-functional genes.

Here, we propose a low computing time method that allows the application of the GBA principle on the Gene Ontology in a very extensive way, and for several postgenomic data. Notably, it allows intensive used of a "cross-validation" procedure to provide high quality assessment of the predictions of the method for each GO term individually. The result is an assessed functional draft of the *P. falciparum* orphan genes. This draft has been integrated in a database available on the web and that can be accessed in different ways. A global view allows to rapidly inspect the branches of the ontology that are more suitable for the GBA principle using the considered type of data (transcriptome, proteome, and protein interaction data). Moreover, user can query the database to search for the potential GO terms of one particular gene, or to search for all the genes that potentially belong to a given GO term.

Global Analysis of *Leishmania* Genes Expression Using SAGE Libraries

S. Smandi[1], F. Guerfali[1], D. Mallouche[2], K. Ben-Aissa[1], O. Mghirbi[1], L. Guizani[1], D. Laouini[1], A. Benkahla[1] and K. Dellagi[1]
[1]LIVGM, Institut Pasteur, Tunis
[2]LEGI-EPT, Ecole Supérieure de la Statistique et de l'Analyse de l'Information de Tunis

Abstract: *Leishmania* is a major parasitic disease in the world for which no vaccine is available and drugs are toxic. As part of its research programme on host-parasite interaction, the LIVGM (*Laboratoire d'Immunologie Vaccinologie et Génétique Moléculaire*) has constructed four human macrophage SAGE libraries, either non-infected or infected by *Leishmania*, and one *Leishmania* promastigote library. The aim of this "SAGE project" is to dissect the impact of intracellular infection on the expression of macrophage genes. The aim of the present project is to identify and classify *Leishmania* derived tags, to evaluate the expression of *Leishmania* genes and to identify those up- or down-regulated in the amastigote stage. The challenge is due to the fact that: (i) very few full length *Leishmania* cDNAs are available in public databases; (ii) the only *Leishmania* gene evidences are the CDS predictions available in GeneDB (www.genedb.org); and (iii) tags are normally located in the 3'-UTR (outside the CDS).

We will present here the strategy followed in order to assign *Leishmania* tags to the corresponding transcripts/gene and to analyse their expression level: (i) *Leishmania* tags mapping was done using BLAST; (ii) the association tag-to-gene was done using Gaussian kernels; (iii) the analysis of differential expression was done using the binomial distribution.

Key words: comparative genomics, gene expression data, parasite, SAGE tags

Multi-SOM: A Novel Clustering Approach for Gene Expression Analysis

Amel Ghouila[1,3] , Haifa Jmel[3], Sadok Ben Yahia[1], Dhafer Malouche[2], Sonia Abdelhak[3]
[1]Computer Science Department, Faculty of Sciences of Tunis, Tunisia
[2]LEGI-EPT-ESSAI, 7 November University, Carthage, [3]Pasteur Institute of Tunis, Tunisia

Abstract: The production of increasingly reliable and accessible expression data has stimulated the development of computational tools to interpret such data and to organise them efficiently.The clustering techniques are largely recognised as useful exploratory tools for gene expression data analysis. Genes that show similar expression patterns over a wide range of experimental conditions can be clustered together. This relies on the hypothesis that genes that belong to the same cluster are co-regulated and involved in related functions. Clustering involves working with a data matrix in which the rows represent the genes, the columns represent the experiments, and each entry represents the expression level of a gene in a given experiment. There exist many clustering algorithms which take microarray data sets as input and produce clusters as output. In spite of the fact that a variety of different clustering algorithms is now available, a number of important questions remain to be addressed. Especially, the estimation of the number of clusters represented in an expression data set is a crucial and complex task, which may significantly influence the outputs of an analysis process.

We propose here a multi level SOM based clustering algorithm titled Multi-SOM. Through the use of clustering validity indices, Multi-SOM overcomes the problem of the estimation of clusters number. It tends to find the optimal clustering of data. To test the validity of the proposed clustering algorithm, we first tested it on supervised training data sets. We evaluate our results by computing the number of misclassified samples. We have then validated Multi-SOM on publicly available biological data. Finally, we discuss the use of Multi-SOM on microarray data of gene expression in the native skin compared to skin derivative cells and on macrophage data infected with different pathogens.

Gene ontology tools are then used to uncover the properties shared among, and specific to, clusters of genes produced by applying Multi-SOM algorithm.

To better visualise and analyse these clusters, we used the web based software, GOTM, FatiGo, ONTO-Express and GoSurfer. These tools are very useful to look for the common traits that are shared within gene clusters, since terms are organised in three general categories, "biological process", "molecular function", "cellular process". To show the specificity of each cluster, statistical tests are also performed.

Session theme: Evolution and Target discovery (p.m.)

Evolution of HIV

Allen Rodrigo
Bioinformatics Institute, University of Auckland, Private Bag 92019, Auckland, New Zealand

Abstract: The Human Immunodeficiency Virus (HIV) evolves very rapidly, a consequence of an error-prone viral reverse transcriptase. The rapidity with which HIV accumulates mutations challenges our efforts to control the virus and design suitable vaccines; it also challenges standard methods of evolutionary analyses that are not designed for rapidly evolving populations. Over the last 10 years, considerable effort has gone into the development of new computational and evolutionary methods of analyses that provide some quite remarkable tools to understand the biology of HIV. In this talk, I will discuss the factors that influence HIV genetic diversity, and the new methods that have been developed to study HIV evolution.

Reference: Drummond A. J., Pybus O. G., Rambaut A., Forsberg R. and Rodrigo A.G. (2003) Measurably evolving populations. *Trends in Ecology and Evolution* 18, 481–488.

Computing Integrated Genetic Epidemiology

Michel Tibayrenc
IRD Representative Office, French Embassy 29, Thanon Sathorn Tai, Bangkok 10120, Thailand

Abstract: Genetic epidemiology considers the role played by genetic factors in the transmission and severity of infectious diseases. However, specialists of this field usually consider only the host (generally humans). Now, pathogens' genetic diversity is generally considerable and has a strong impact on the epidemiology of transmissible diseases, including resistance to antibiotics and other drugs. Moreover, the genetic diversity of pathogens can be used for epidemiological tracking (strain typing). Lastly, in the case of vector-borne diseases, vectors' genetics needs also to be explored to characterise their populations of vectors, as well as their vectorial capacity and their resistance to insecticides.

The integrated approach proposed by us consists in exploring together the impact, on the transmission and severity of infectious diseases, of the host's, the pathogen's and the vector's genetic diversity. As a matter of fact, these three parameters actually correspond to a unique biological phenomenon (coevolution). Genetics is taken here in a broad sense and includes genomics, proteomics, population biology and evolution.

This approach is the scope of the MEEGID (Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases) congresses and of the journal Infection, Genetics and Evolution.

By taking the example of American trypanosomiasis (Chagas disease), I will illustrate the difficulties of this approach and will discuss the possibilities to apply it to other infectious diseases, including African trypanosomiasis.

Analysis of Polymorphisms and Signatures of Positive Adaptive Selection at Candidate Genes for Genetic Resistance to Avian Viral Diseases

Sheila Ommeh[1, 2], Daniel Masiga[1], David Lynn[3], Han Jianlin[2] & Olivier Hanotte[2]
[1]Department of Biochemistry and Biotechnology, Kenyatta University, P. O. Box 43844, Nairobi, Kenya
[2]International Livestock Research Institute (ILRI), P. O. Box 30709, Nairobi, Kenya
[3] Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland
E-mail: s.ommeh@cgiar.org

Abstract: The objective of this study is to assess polymorphisms at candidate genes associated with genetic resistance to viral diseases in chicken, such as Marek's disease and avian influenza. Two candidate genes were shortlisted from literature and their sequences obtained from Ensembl. BLB2 (class II) major gene is present in chicken within the MHC region on chromosome 16. It has been shown to be associated with resistance to avian viral infection, including Marek's disease. Mx gene on chromosome 1 is a putative candidate gene for genetic resistance to avian influenza at position 631 of the encoded amino acid. Species homologs of these two genes were obtained from the NCBI BLAST and analysed with the PAML package for detection of signature of positive adaptive selection (*dN/dS* ratio). Lineage analysis of species homologs of the Mx gene revealed a *dN/dS* ratio of 0.64

implying purifying selection, but site analysis revealed five amino acids under positive selection. These sites under positive selection lie in the central interaction domain CID of the Mx gene that mediates the activities of the GTPase domain and the leucine zipper domain which is important in the oligomerisation of the Mx protein. For the BLB2 gene lineage analysis of species homologs did not reveal positive selection but site analysis revealed seven amino acids under positive selection which lie in the peptide binding region PBR of the protein. This region is encoded by exon 2 of the BLB2 which and is important in antigen presentation. In summary site analysis reveals that some amino acids in the CID and PBR domains for the Mx and BLB2 genes respectively are under positive selection and this warrants further research on these two protein domains for their exact role towards genetic disease resistance.

Dynamics of Migrations of Malaria Protective Polymorphisms Across Africa's Sahel

N. A. Saleh, H. Y. Hassan, A. Hussein, D. Kwiatkowski and M. E. Ibrahim

Abstract: The Sahel which extends from the Atlantic Ocean to the Ethiopian highland is a historical reservoir of Africa cultures and grandest populations and a known arena of ancient and recent migrations. We study polymorphisms of mitochondrial and Y chromosome haplotype and SNPs of known protective effects against malaria in populations that lives across the Sahel, e.g. Hausa, Massalit, Bergu as well as other groups extending into the eastern Sudan. Populations were classified according to their linguistic affiliation which comprised the main linguistic families of Africa (Afro-Asiatic, Niger Kordofanian and Nilo-Saharan). We present evidence that despite the known historical endemicty of malaria in the eastern Sahel, some of the major protective polymorphisms in malaria have in fact found their way only recently to the gene pool of the populations in eastern Sahel. We give estimates of the age of the introduction of such alleles and discuss the possible dynamics of the process and its impact on the epidemiology of the diseases.

In Silico Strategies for Target Discovery

Eric Maréchal
Comparative chemogenomics team, UMR 5168, Institut de Recherche en Sciences et Technologies pour le Vivant, CEA Grenoble, France
emarechal@cea.fr

Abstract: In medical sciences, and in widest terms in biology, a target is a broad concept to qualify a biological entity and/or a biological phenomenon, on which one aims to act as part of a therapeutical intervention. Intervention on the target depends on the level of knowledge of the disease and the cultural background of the physician [traditional medicine will focus more on the symptoms in order to end the symptoms; allopathic (western, orthodox) medicine will focus on the causes (i.e. the etiology) in order to oppose the causes of the diseases; homeopathic medicine (in the sense of non-allopathic medicine) will focus on the etiology in order to treat likes with likes; in the widest sense of the word, vaccination is a category of homeopathic treatment]. Intervention also relies on the availability of techniques and technologies for the intervention, without exogenous agents (musculoskeletal manipulations, surgery, diets), with exogenous biochemicals (natural substances based on traditional pharmacopoeia, drugs derived from traditional pharmacopoeia or from synthetic chemistry, vaccines) or with exogenous genes (gene therapy).

Based on an example of a target/drug discovery project undergone in our laboratory to introduce new antimalarials, we show how current strategies for the discovery of target for new drugs actively combine *in silico* and "wet" experiments and analyses. *In silico* organisation of genomic and post-genomic information complying with knowledge of the disease in etiologic terms, appears as an efficient source of information to help a target discovery project, as long as accurate analysing and mining tools are available. Objects, attributes and data are consistent with biological entities as defined or understood by bench biologists in reductionist terms, i.e. genes, RNA, proteins, solutes, cellular structures, tissues, organs, etc., and their physiological measures and scales, i.e. mass, size, lengths, concentrations. Processes and functions represent biological operations and alterations over time, growth, transformations, movements, reactions. Accurate integration of data and assessment of the relations between objects and processes are critical. However, the most striking limitations lie in the analytic and mining tools to explore the stored biological information. In the case of malaria, in spite of an unprecedented effort to organise genomic and post-genomic information regarding its

causative agents (*Plasmodium* sp.) and make it accessible to the largest scientific audience, atypical features of the malarial genomes limit the use of conventional analytical tools.

Subsequent objective of a target discovery project, which is from target to lead, drug, vaccine or gene therapy, should not be ignored. The majority of the targets reported in the literature did not open to any realistic therapeutical treatment. An important drawback is the cost of drug development. In this concern, future prospects include the integration of biological and chemical information.

Reference: Birkholtz L.M., Bastien O., Wells G., Grando D., Joubert F., Kasam V., Zimmermann M., Ortet P., Jacq N., Saidani N., Roy S., Hofmann-Apitius M., Breton V., Louw A.I. & Maréchal E. (2006) Integration and mining of malaria molecular, functional and pharmacological data: how far are we from a chemogenomic knowledge space? *Malar. J.* 5(1):110.

Identification of *Theileria parva* Vaccine Candidate Genes Using a Bioinformatics Approach

Etienne de Villiers
International Livestock Research Institute, P. O. Box 30709-00100 Nairobi Kenya

Abstract: Immunity against the bovine protozoan parasite *Theileria parva* has previously been shown to be mediated through lysis of parasite-infected cells by MHC class I restricted CD8+ cytotoxic T lymphocytes. It is hypothesised that identification of CTL target schizont antigens will aid the development of a sub-unit vaccine. We exploited the availability of the complete genome sequence data and bioinformatics tools to identify genes encoding secreted or membrane anchored proteins that may be processed and presented by the MHC class I molecules of infected cells to CTL. A 525 base pair ORF encoding a 174 amino acid protein, designated Tp2, was identified by *T. parva*-specific CTL from four animals. These CTL recognised and lysed Tp2 transfected skin fibroblasts and recognised four distinct epitopes. Significantly, Tp2 specific CD8+ T cell responses were observed during the protective immune response against sporozoite challenge. This makes Tp2 an attractive candidate for evaluation of its vaccine potential.

Discovery of Novel Drug Targets Against Pathogenic Protozoa: The Promise of Metabolic Reconstruction

L. W. Wambua[1 2], G. McConkey[1], D. R. Westhead[1]
[1]Leeds University, School of Biological Sciences, Leeds, UK
[2]International Livestock Research Institute, Nairobi, Kenya.

Abstract: Protozoan organisms as *Plasmodium*, *Toxoplasma*, *Eimeria*, *Babesia*, *Theileria*, *Trypanosoma, Leishmania, Cryptosporidium* and *Entamoeba* species require intensive research for discovery of new drugs to circumvent drug resistance. The availability of complete genome sequences for many protozoans has set stage for mining of their genomes through systems biology approaches for novel drug targets. The metabolic architecture of the organisms in particular, provides a rich resource for potential targets.

We sought to identify presence of unique metabolic enzymes and pathways that would make possible drug targets in the Protozoans of interest by exploring their metabolism through online database searches. However, profound under-representation of these organisms on online public databases rendered this approach unfruitful. We therefore performed metabolic reconstruction of their metabolic pathways from their genome sequences using metaSHARK, a metabolic reconstruction tool developed by the Leeds University. This software combines a gene-detection package which identifies metabolic enzymes within raw un-annotated DNA sequences or ESTs. Gene detection is based on a two-stage protocol involving a PSI-TBLASTN search on the PRIAM database followed by application of Wise2 algorithm with Hidden Markov Models to assert the enzymes.

We studied five metabolic pathways: Glycolysis, Krebs's cycle, Shikimate, Pantothenate, Coenzyme A and Folate biosynthetic pathways. Our results showed potential targets in shikimate and folate pathways in *Plasmodium*, *Toxoplasma* and *Eimeria* species. *Babesia, Theileria and Plasmodium* had highly similar metabolic profiles, suggesting prospects of using anti-malarial drugs for treatment of Babesiosis and Theileriosis. *Toxoplasma* and *Eimeria* species had robust metabolic capacities compared to all other protozoans explored and therefore had several potential drug targets in the folate, pantothenate and shikimate metabolic pathways. *Cryptosporidium* and *Entamoeba*, have drastically reduced metabolic repertoire. However, hope lies in the mining of their genomes for unique bacterial genes acquired by possible lateral gene transfer. *Trypanosoma* and *Leishmania* species also

showed reduced metabolic capacities, apart from their elaborate carbohydrate metabolism pathways. The evolutionary compartmentalisation of their glycolysis pathway in specialised organelles referred to as glycosomes may have rendered these enzymes distinct from those of the human host, thereby making them potential drug targets. This is subject to further comparative genomic analysis.

We found the metabolic capacities of the organisms to be intricately associated with the endosymbiotic evolutionary origins of the organelles where these pathways were localised in the cell, most notably the mitochondria and the apicoplast. Loss of these organelles through secondary events was congruent with the absence of these pathways in the organism.

LBTx: A Tool Box for the Selection and Analysis of Potential Leishmania Targets for Intervention and Disease Control

A. S. Chakroun, L. Turki, H. Kebaier, M. Barhoumi, Y. Saadi, I. Smeti, A. Sahli, S. Guerbouj, R. Lahmadi, M. Ben Fadhel, M. Driss, B. Kaabi, K. Ajroud, J. Chemkhi, S. Ben Abderrazak, I. Guizani
Laboratoire d'Epidémiologie et d'Ecologie Parasitaire: Applied biotechnologies and genomics to parasitic diseases research programme, Institut Pasteur de Tunis, Tunisia

Abstract: Protozoan parasites of the genus *Leishmania* are responsible for leishmaniases, neglected diseases having a wide geographical distribution. These highly prevalent diseases cover a wide range of clinical presentations and consitute major public health problems. In spite of the importance and the number of studies achieved on these diseases, they remain poorly controlled. The identification of new targets/novel approaches for diagnostics, vaccine or treatment remains a research priority.

As a part of our work, we developed a bioinformatic's based support that we customised to the specific needs of our research approaches. This tool has been implemented to analyse polymorphic DNA markers from different Leishmania parasite species like *L. major, L. aethiopica, L. tropica, L. infantum, L. donovani* and *L. archibaldi*, identify polymorphic microsatellite loci, and design of specific PCR primers using a comparative genomic approach. It has also been instrumental for the comparative genome analysis and annotation of differentially expressed Leishmania transcripts, or Leishmania antigen genes and families.

This tool called LBTx for "Leishmania Bioinformatic's ToolboX" contains interfaced programmes and scripts among which BLAST where its searching database became only constituted by public releases of the different Leishmania genomes data. Other scripts were developed in Perl language (MutationCheckUp, pepHasher, MisaCluster). As a matter of example, MisaCluster is a microsatellite viewer and a primers/probes mapper that takes multiple sequence alignments as input data. This tool is used in association with other bioinformatic programs like Artemis, ACT, Clustal, MEME, primer 3, etc. This Toolbox keeps under development and can be expanded to the needs on request.

This work received financial support from MESRST–Tunisia (Contrat-programme 2004–2008), from the UNICEF/UNDP/World Bank/WHO special program for research and training TDR (A30380, A11032, A30134) and from RAB & GH programme of WHO/ EMRO-COMSTECH.

Session theme: Structure and Drugs

Information Superstructure for Protozoan Aquaporins

Raphael D. Isokpehi[1], Nyasha Chambwe[1], Jessica M. Murray[1], Hari H. P. Cohly[1], Subhangi Varadharajan[2] and Rajendram V. Rajnarayanan[3]
[1]Department of Biology and RCMI Centre for Environmental Health, Jackson State University, Jackson, Mississippi, USA
[2]Department of Chemistry, D. G. Vaishnav College, TN, India
[3]Department of Chemistry, Tougaloo College, Jackson Mississippi, USA

Abstract: Protozoan parasites of the genera *Plasmodium*, *Trypanosoma* and *Leishmania* contribute significantly to the burden of global infectious diseases. Genome sequencing projects on protozoan parasites combined with high-throughput gene and protein expression experiments are producing wealth of datasets for the discovery of novel drug targets. Water channel proteins collectively termed aquaporins are increasingly recognised as potential drug targets or transport routes for anti-protozoan drugs. We are developing an information superstructure on protozoan aquaporins by integrating heterogeneous datasets including genomic sequences, gene and protein expression profiles, protein structures and literature. Conventional aquaporins are six-transmembrane proteins with two

Asparagine-Proline-Alanine (NPA) motifs found in the inter-transmembrane loops B and E that form an aqueous pore. There are 13 known human aquaporins (AQP0 to AQP12) while in sequenced protozoan genomes the abundance of aquaporins range from 0 (*Cryptosporidium* and *Theileria*) to 7 (*Trypanosoma cruzi*). Protozoan aquaporins have been shown to also transport nutrients and metabolites between host and parasite that are crucial for initiating an infection and survival during parasite's life cycle. We report the generation of motif signatures from multiple sequence alignment of 26 aquaporins from 7 kinetoplastids (*L. major, L. infantum, L. braziliensis, T. brucei, T. congolense, T. cruzi,* and *T. vivax*) and the human aquaporins. These signatures are being examined as tools for aquaporin data integration and prioritization of binding sites for drug design/development.

In Silico Drug Design

Mahammad Afshar
Ariana Pharma–Pasteur Biotop, France

Computational Discovery of Drugs Resistance Mechanism(s) of the Malaria Parasite *Plasmodium falciparum*

Marion Owolabi[1], Segun Fatumo[1,2], Rainer Koenig[2], Gunnar Schramm[3], Anna-Lena Kranz[2], Roland Eils[2,3], Ezekiel Adebiyi[1]
[1]Department of Computer and Information Sciences, Covenant University, PMB 1023, Ota, Nigeria
[2]Department of Bioinformatics and Functional Genomics, Institute for Pharmacy and Molecular Biotechnology, University of Heidelberg, 69120 Heidelberg, Germany
Theoretical Bioinformatics, German Cancer Research Centre (DKFZ), 69120 Heidelberg, Germany

Abstract: At the genomics level, we use the microarray technology to produce gene expression data for various organisms under many conditions. A desirable advancement is the need to extract from this gene expression data, information that will be useful toward given answers to the questions targeted at the design of the microarray experiment. Analysis of gene expression data of *P. falciparum* when induced with two anti-malaria drugs (such as chloroquine and choline analogue T4) has shown that the parasite resistance mechanisms may not be elucidate-able at the genomics level. But on a proteomic level, biochemical research has elucidated an increasingly complete image of the metabolic architecture of organisms that included that of *P. falciparum*.

In this work, we sought to use the biochemical network of *P. falciparum* to deduce its drugs resistance mechanism(s) using the two gene expression data obtained when the parasite is treated with chloroquine and choline analogue T4. We do this by mapping the gene expression data onto the enzymatic reaction nodes of the metabolic network. A consecutive clustering method is used to derived important clusters. Further, a wavelet and a feature extraction method is applied to study these clusters which gives important insight into the mechanisms that *P. falciparum* deplore for resisting anti-malaria drugs.

Key words: Genomics, Proteomics, Microarray technology, *Plasmodium falciparum*, Biochemical network, Clustering, and Wavelet.

Large-scale Distributed *in silico* Drug Discovery Using VSM-G (Virtual Screening Manager for computational Grids)

Léo Ghemtio * ,Bernard Maigret, Alexendre beautrait, Vincent Leroux , Mathieu Chavent
Orpailleur and Algorille teams, LORIA, UMR CNRS 7503, Nancy University, France

Abstract: The Virtual Screening Manager platform, dedicated to automated virtual high-throughput screening using cluster grids, is presented here. The VSM-G package is constituted of pre-processing engines for both protein targets and small molecules putative ligands to be screened on, and of a funnel docking strategy. The latter is architectured as a sequence of different docking modules ranging from a fast but less accurate surface-matching procedure for crude rigid dockings to slower but more elaborated molecular dynamics calculations. At each step of the funnel and depending on tuning, a small proportion of molecules can be prioritised as more promising, discarding a large body of inappropriate ligands. VSM-G is able to handle several millions of compounds versus hundreds of targets (which can be different targets or several conformations of same one, e.g. extracted from molecular dynamics sampling or NMR data)), yielding to a small set of putative hit compounds to be proposed for real biological testing.

The high throughput virtual screening experiment which will be described aimed to identify, from a database consisting of about 600,000 drugable molecules, putative hits for the LXR receptors. In order to handle both compounds and receptor flexibilities, 400 conformers were generated for each molecule, leading to a total of about 91 millions of 3D objects to be processed. On the receptor side, 4 different crystal conformations were used for the docking calculations.

Because considerable computing power is needed for performing searches of this type within a reasonable time, VSM-G targets the use of computational grid technology, especially for the fast-matching procedure used in the first selection docking filter. This filter, which had to approximate all molecules and target surfaces using spherical harmonics polynomials, was used to eliminate most of the tested compounds in order to pass only about 2000 molecules to the next flexible docking filter. In order to obtain the molecular surfaces of the 91 M 3D objects, the French GRID'5000 facilities were used adding a total of 1360 processors.

At the end of this experience, we were able to propose a list of only 20,000 molecules to really be tested by our pharmaceutical company partner, reducing therefore the cost for an in vitro HTS campaign while increasing the chances of successes.

Several perspectives for expanding VSM-G functionality, some of which are currently developed, will also be discussed.

In silico docking against malaria: The WISDOM initiative

V. Breton, A. Da Costa and V. Kasam
LPC Clermont-Ferrand, Campus des Cézeaux 63177 Aubière Cedex, France on behalf of the WISDOM collaboration
Contact: V. Breton, breton@clermont.in2p3.fr

Abstract: The WISDOM initiative aims at developing new drugs for neglected and emerging diseases with a particular focus on malaria. Its specificity is to extensively rely on emerging information technologies to provide new tools and environments for drug discovery and development. Its main goal is to boost research and development on neglected diseases by fostering the use of open source information technology for drug discovery.

The WISDOM collaboration is presently using the grid paradigm to identify potential hits in virtual screening. Started in 2005, the initiative has been initially focusing its effort on the haemoglobin metabolism of the malaria vector which is one of its key metabolic processes. Plasmepsin, the aspartic protease of Plasmodium, is responsible for the initial cleavage of human haemoglobin. The goal was to identify which molecules could dock on the protein active sites in order to inhibit its action and therefore interfere with the molecular processes essential for the pathogen.

During the summer of 2005, a first large scale deployment allowed achieving in silico docking of 500.000 compounds in about 6 weeks against different structures of plasmepsines II and IV [1]. Results were analysed and the top 1000 compounds were further processed using a Molecular Dynamics procedure to select finally 25 compounds which are now being tested in vitro.

With the success achieved by this first initiative both on the computation and biological sides, several scientific groups around the world including groups in Africa proposed targets implicated in malaria which led to a second assault on malaria during the fall of 2006 [2]. WISDOM-II project dealt with several targets which are both X-ray crystal models and homology models. Targets from different classes of proteins were tested: reductases like DHFR as well as transferases such as GST. Analysis of this second initiative is under way.

The WISDOM production environment, which was initially designed to allow docking at a very large scale, is now evolving towards a fully grid-enabled virtual screening process where the compounds go through several selection steps using docking and molecular dynamics procedures.

In the coming months, we aim at expanding our collaboration with the research communities on neglected diseases to inform them of this service which allow cost-effective screening of their most promising targets. We also foresee to apply our virtual screening process to a subset of the targets proposed by the TDR drug target portfolio network (http://www.who.int/tdr/tropics/discovery_research/drug_target.htm).

In a longer term, our vision goes beyond virtual screening and we call for a distributed, internet-based collaboration [3] to address one of the worst plagues of our present world, malaria. The spirit is a non-proprietary peer-production of information-embedding goods. And we propose to use the grid technology to enable such a world wide open-source like collaboration.

References: (1) N. Jacq et al. Grid enabled virtual screening against malaria, accepted for publication in *Journal of Grid Computing* (2) J. Salzemann et al. Grid Enabled High Throughput Virtual Screening Against Four Different Targets Implicated in Malaria, Proceedings of Healthgrid conference 2007, Studies in Health Technology and Informatics, 126 (2007) (3) V. Breton et al Grid Added Value to Address Malaria, to be published in IEEE Transactions on NanoBiosciences.

POSTER ABSTRACTS

POSTER 1: Sequence Analysis of the Entire RNA Genome of Sweet Potato Chlorotic Fleck Virus Reveals that it Belongs to a Distinct Carlavirus Species

V. Aritua [1], E. Barg[2], E. Adipala[3] and H. J. Vetten[2]

[1]National Agricultural Biotechnology Centre, Kawanda Agricultural Research Institute, P. O. Box 7065, Kampala, Uganda

[2]Biologische Bundesanstalt fur Land und Forstwirtschaft, Institut fur Pflanzenvirologie, Mikrobiologie und biologische Sicherheit, Messeweg 11-12, D-38104, Braunschweig, Germany

[3]Department of Crop Science, Makerere University, P. O. Box 7062, Kampala, Uganda.

Abstract: The complete nucleotide sequence of the single-stranded RNA genome of a Ugandan isolate of Sweet potato chlorotic fleck virus (SPCFV) was determined. Excluding the 3′-terminal poly(A) tail the genome is 9104 nucleotides in length. It comprises six open reading frames (ORFs). ORF1 codes for a 238 kDa protein with characteristic replicase motifs. ORF2, -3 and -4 code for proteins of 27.5kDa, 11.5kDa and 7.3kDa, respectively, shows strong homologies with triple gene block proteins. ORF5 encodes the capsid protein of 33 kDa and ORF6 encodes a protein of 15kDa with a nucleic acid binding zinc finger motif. Based on the genomic organisation SPCFV appears to be a member of the genus *Carlavirus* with the family *Flexiviridae*. Comparison of the nucleotide and predicted amino acid sequences of the putative polymerase and coat protein with the members of *Flexiviridae* also showed highest similarities with the corresponding gene products of carlaviruses. However, SPCFV takes a position close, but clearly separate from the carlaviruses, as none of its putative gene products shared amino acid sequence similarities of >40% with the homologues of other carlaviruses. Based on CP sequences, its closest relative among the carlaviruses is *Melon yellowing-associated virus,* a proposed carlavirus from Brazil with which it shares 46% aa homology. The larger RNA (9104 nt) of SPCFV in comparison to other carlaviruses (7.4 to 8.5 kb) is largely due to the considerably larger replicase (238 vs. 200-223 kDa) and the considerably long un-translated region between ORF4 and the putative CP (213 vs. ca. 34 nt for PVM). Although SPCFV had earlier been regarded as a potyvirus, our data provided unambiguous evidence for its assignment as a distinct species to the genus *Carlavirus*.

POSTER 2: In Silico Identification of Candidate Pathogenicity Islands in Complete *Xanthomonas* Genomes

Mtakai Vald Ngara, Kenneth Babu Opap, Dr. Dong-Jin Kim

International Institute of Tropical Agriculture (IITA), P. O. Box 30709-00100, Nairobi, Kenya (dj.kim@cgiar.org)

Abstract: The need to integrate computational power and techniques in biological research in the African continent is crucial in tackling disease, food insecurity and managing genetic resources. The International Institute of Tropical Agriculture (IITA) is supporting efforts that will enhance the integration of computation in biological research and exploitation of the numerous bioinformatics resources in research. Currently, IITA is conducting basic in silico research on phytopathogens of the African major food crops.

The genus *Xanthomonas* is a diverse and economically important group of phytopathogens, belonging to the γ-subdivision of the Proteobacteria. The species and pathovars, cause distinct disease phenotypes in a wide range of plants.

Pathogenicity islands (PAIs), distinct genomic segments encoding virulence factors, represent a subgroup of genomic islands (GIs) that have been acquired by horizontal gene transfer event. Identification of potential PAI is useful in understanding disease pathogenicity, host specificity and genomic evolution in microbial genomes.

In this study, we are utilising an in silico approach to detect potential PAI(s) in six complete *Xanthomonas* genomes by combining sequence homology and abnormalities in genomic composition. We will first identify the GI(s) in the genomes, by calculating the G + C content anomalies and codon usage bias. All the PAI loci will be collected and their homologs identified in the genomes to come up with genomic strips. An algorithm will then be developed to merge overlapping or adjacent strips into large genomic regions. Among the defined genomic regions, PAI-like regions will be identified by the presence of homolog(s) of virulence genes. Candidate PAI (cPAI) will be considered only if the PAI-like region partly or entirely spans the GI. The identification and analysis of cPAI in *Xanthomonas* will broaden our knowledge on disease phenotype diversity, host specificity and genomic evolution in these phytopathogenic bacteria.

POSTER 3: Differential Expression of the CrV1 Haemocyte Inactivation-associated Polydnavirus Gene in the African Maize Stemborer Larvae Parasitized by Two Biotypes of the Endoparasitoid *Cotesia sesamiae* (Cameron)

Catherine Wanjiru Gitau[1]; Dawn Gundersen-Rindal[2]; Monica Pedroni[2]; Dupas Stephane[3]; J. P. Mbugi[1]

[1]Kenyatta University, Department of Zoology, School of Pure and Applied Sciences
[2]USDA-ARS. Insect Biocontrol Laboratory, Beltsville, MD, 20705, USA
[3]IRD c/o CNRS, Laboratoire Populations Génétique et Evolution, 1 av. de la Terrasse, 91198 Gif-sur-Yvette, France.

Abstract: The braconid *Cotesia sesamiae* is the most common endoparasitoid of the cereal stem borers *Busseola fusca* and *Sesamia calamistis* in sub-Saharan Africa. In Kenya, two biotypes of *C. sesamiae* exist. The *C. sesamiae* population from the western highlands completes development in *B. fusca* whereas eggs by *C. sesamiae* from the coastal population get encapsulated and ultimately, no parasitoids emerge from the *B. fusca* larvae. Both populations successfully complete development in *S. calamistis* larvae. Results from this study showed that *B. fusca* larvae parasitised by the avirulent *C. sesamiae* had frequently more encapsulated eggs at 24 hours post parasitism compared to earlier time points. Expression of the *C. sesamiae* polydnavirus (PDV) CrV1 gene associated with haemocyte inactivation in the *Cotesia rubecula*/*Pieris rapae* system, was detected in fat body and haemolymph samples from both *B. fusca* and *S. calamistis* larvae parasitised by the virulent *C. sesamiae* from Kitale and Meru regions. There were no or faint amplified CrV1 products on *B. fusca* fat body and haemolymph samples parasitised by the avirulent *C. sesamiae* from Mombasa and Kitui regions. The difference in expression of CrV1 gene in *B. fusca* and *S. calamistis* suggests the involvement of the CrV1 gene in immune suppression and sheds light on the refractoriness of *B. fusca* towards *C. sesamiae* biotypes.

POSTER 4: Genetic Structure and Analysis of *Striga gesnerioides* (Witchweed) from Senegal

D. C. Tonessia[1], M. Wadw[2], C. Ndiaga[3] and M. P. Timko[4]

[1]Dotorante au CERAAS bp 3320 Thiès- escale Senegal
[2]Malherbologiste au CNRA de Bambey (Senegal)
[3]Plant Breeder ISRA-CNRA BP 53Bambey Senegal
[4]Department of biology, university of Virginia, Charlottesville 22904

Abstract: The parasitic angiosperm, *Striga gesnerioides* is a root parasite of wild and cultivated legumes, among which cowpea (*Vigna unguiculata*) is suitable host. In West Africa, it causes serious yield losses up to 100% in some instances. Based upon the differential resistance response of various cultivars and breeding lines, five distinct races of *Striga gesnerioides* parasitic on cowpea have been proposed to exist in west and central Africa. But the race of the parasite present in Senegal is not known. Understanding the genetic structure and host-parasite interaction of *Striga gesnerioides* from Senegal is important because it allows identification of races or biotypes thus improving chances of breeding success.

Using Amplified fragment length polymorphism (AFLP) profile analysis and statistical clustering methods, we have examined the genetic variability and phenotpicic relationships of 215 individuals plants representing 12 different populations of *S. gesnerioides* collected from various areas throughout their suspected distribution range in Senegal.

Our results showed that genetic variability within and among populations are extremely low. Based on its genetic profile and host preference/resistance responses, we show that *S. gesnerioides* is parasitic on cowpea in Senegal and the race present in Mali are closely related.

POSTER 5: Analysis of *Plasmodium falciparum* var Gene Repertoire Expressed in Children with Severe Malaria from Tanzania

Joseph Paschal Mugasa[§], Matthias Rottman,[*] Weihong Qi[*], Hassan Mshinda[§] and Hans-Peter Beck[*]

[§]Ifakara Health and Research Development Center, Ifakara, Tanzania
*Swiss Tropical Institute, Basel, Switzerland

Abstract: The *var* gene family of *Plasmodium falciparum* encodes for the variant surface antigen PfEMP1 (*Plasmodium falciparum* erythrocyte membrane protein 1). PfEMP1 is considered an important pathogenicity factor in *P. falciparum* infection by mediating cytoadherence to host cell endothelial receptors. *var* genes can be grouped into the three major groups (A, B and C) and two intermediate groups (B/A and B/C) on the basis of their genomic location and upstream sequence similarities in coding and non-coding upstream regions. Severe malaria may be caused by parasites expressing restricted subset of PfEMP1 variants that confer optimal sequestration in immunologically naïve individuals. Studies on *var* gene expression are difficult due to the broad variability of *var* genes in field isolates. A case control study was set in Ifakara, Tanzania, children under age of five years with severe malaria were recruited using WHO 2000 guideline of case definition, and children with asymptomatic malaria were enrolled in a study as a control group. Expression patterns of different fragments of var genes sequences were compared. We found that *P. falciparum* isolates from children with severe malaria predominantly transcribe var genes with DBL1-like domains that are characteristic of Group A or B/A var genes with reduced cysteine residues. In contrast, isolates from children with asymptomatic malaria predominantly transcribe var genes with DBL1-like domains that are characteristic of the B and C-related var gene groups. Initial phylogenic analysis showed that dominant expressed var genes from severe malaria isolates cluster together, if compared with dominant expressed var genes from asymptomatic isolates. These preliminary results support the hypothesis that var genes with DBL1-like domains (Group A or B/A) may be implicated in the pathogenesis of severe malaria.

Poster 6: K-Means Clustering and the Analysis of Malaria Microarray Data (MMD)

Victor Osamor[1], Ezekiel Adebiyi[1] and Seydou Doumbia[2]

[1]Department of Computer and Information Sciences (Bioinformatics Unit) College of Science and Technology Covenant University, Ota, Ogun State, Nigeria
[2]Malaria Research Training Center (MRTC) Bamako, Mali

Abstract: On regular basis, enormous amounts of genomic data are generated by researchers. Forming a large part of these huge raw data resources is the malaria microarray data arising from the study of gene expression in both *Plasmodium* infected mosquito and human beings to understand their biology and evolve new drug remedies. Malaria, caused by *Plasmodium falciparum,* is lethal and responsible for major losses and death in sub-Saharan Africa. The purpose of microarray is to measure the concentration of mRNA for thousands of genes at given time points. Since data analysis have profound influence on interpretation of the final results, basic understanding of the underlying model surrounding computational tools is required for optimal experimental design and data analysis by biology researchers who are target users of such tools. K-means is one of the popular and simple partition computational models for clustering microarray data but has the problem of requesting for number of clusters from researchers who may not have idea about the structure of the data to be clustered.

In this work, we analyse the operational iteration mechanism of k-means by what we call *Step-by-step k-means walk* using synthetic gene expression data. However, we also examine the results of analysis of k-means and other clustering techniques on malaria microarray data (MMD) with the objective of revealing their strength and weakness on existing experimental works. We discover open problems that are both computational and experimental based, with k-means clustering having further problems of sensitivity to initialisation and also often find solutions that are local optimum and apparently far from global optimum. There is also the need for a new microarray experiment which ultimately raises some design issues. Our new simplified K-means algorithm for clustering (for example) gene expression data arising from malaria microarray experiment will be capable of fitting

adequate number of genes clusters required for any given microarray input data set automatically without guessing k (number of clusters) input.

Key words: Malaria, Microarray, Gene expression, *Plasmodium*, mRNA, Clustering

POSTER 7: Bioinformatics Analysis of Dynein Light Chains of *Plasmodium falciparum*

Elijah K. Githui[1], Etienne De Villiers[2] and Andrew G. McArthur[3]
[1]*Laboratory of Molecular Genetics, National Museums of Kenya*
[2]*Bioinformatics facility, ILRI*
[3]*Marine Biological Laboratories, Woodshole, MA.*

Abstract: *Plasmodium* spp. belong to the phylum apicomplexa. *Plasmodium*, *Toxoplasma* and *Cryptosporidium* are parasites of considerable medical importance while *Theireria* and *Eimeria* are animal pathogens with extensive impact on food production. *P. falciparum* is particularly important because it causes malaria that has 300–500 million clinical infections, mainly in sub-Saharan Africa and results in more than 1 million deaths each year. The malaria parasite actively invades the host cell in which they propagate and several proteins associated with the apical organelles have been implicated to be crucial in the invasion process. The biogenesis of the apical organelles is not well understood but several studies indicate that microtubule-based vesicular transport is involved. In *P. falciparum*, the microtubule destabilising drugs, colchicine, dinitoanilines and taxotere considerably reduce the number of new rings in culture or completely abolish parasitaemia. *Toxoplasma gondii* tachyzoites treated with oryzin or colchicine loses bulk of the subpellicular microtubules and the ability to re-invade host cells. At a higher drug concentration, the nascent spindle and the subpellicular microtubules are disrupted and the parasites are incapable of nuclear division. Immunofluorescence studies indicate the presence of microtubule based proteins at the apical organelle and have suggested that cytoplasmic dyneins maybe involved in the biogenesis of this organelle, and also in the replication of these parasites. Other studies have shown the presence homologues of vesicle mediated transport (COP, Sec23/31, NSF) inside the parasite and in the trans-cellular transport in *P. falciparum* infected erythrocytes but their precise role is not investigated. Dynein is a multi-subunit complex composed of heavy chains whose C-terminus form a globular head containing multiple ATP-hydrolysis sites and a small stalk that binds to microtubules while the N-terminus is a flexible stem that interacts with intermediate chains and light chains to form the cargo binding domain of the enzyme. In this study, we analysed the light chains of *P. falciparum* because they provide adaptor surface to the cargoes and, therefore, there is likelihood for variations. The cytoplasmic dynein light chains consist of three different families: TcTex1/2, LC8 and LC7/roadblock. TcTex1/2 are differentially expressed in various tissues and have been implicated in the attachment of cargoes to dynein motor. Vertebrates TcTex1 binds directly to rhodosin and disruption of the protein leads to retinitis pigmentosa. The LC8 protein family is highly conserved and is found in many multimeric enzymes including Nitric oxide synthase and myosin V. Complete loss of LC8 function is embryonic lethal. The LC7/roadblock family has been shown to bind directly to the intermediate chains of dynein complex. The data presented demonstrate that the *P. falciparun* dynein light chains sequence and functional domains show high sequence homology within the apicomplexa and that only the dynein LC 8 group2 (97 a.a.) has a high homology (e -40) to human while TcTex1 and dynein LC7 group1 (93 a.a.) have low homology (e -08, e -05, respectively) to the human homologues. This suggests that the sequence/domain differences in these proteins can be targeted for vaccine or drug design.

POSTER 8: Managing Energy Consumption and NAD Metabolism as a Basis to Enhance Stress Tolerance in Tropical Maize

Sylvester Anami[a,b], Hilde Nelissen[a], Jesse Machuka[b], Dirk Inze[a], and Mieke Van Lijsebettens[a]
[a]Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology, Ghent University, Technologiepark 927, B-9052 Gent, Belgium
[b]Plant Transformation Facility, Department of Biochemistry and Biotechnology, Kenyatta University, P. O. Box 43844-00100 Nairobi, Kenya

Abstract: In sub-Saharan Africa, tropical maize has traditionally been the main staple of the diet, 85% of the maize grown is used directly as human food. However its productivity has been hampered mainly by drought, salinity and extreme heat. In the present study, ZMPARP1 gene that breaks down

the cells energy homeostasis under drought stress conditions was targeted for RNAi and artificial miRNA (amiRNA) as a way of reducing the impacts of water deficit on tropical inbred maize lines. We have built an RNAi and amiRNA constructs for ZMPARP1 gene that will be transformed in local tropical maize variety via *Agrobacterium* mediated transformation and test whether the cells energy homeostasis will be maintained. We have also achieved successful regeneration and transformation of tropical inbred line using EHA101 Agrobacterium strain habouring pTF102 binary vector.

POSTER 9: Functional Studies of a Recombinant Anti-TcP2β Antibody in *Trypanosoma brucei*

B. Nyambega[1,4], C. Smulski[1], M. Juri Ayub[2], L. Simonetti[1], V. Owino[4], J. Hoebeke[3], D. Masiga[4] and M. J. Levin[1]

[1] Laboratorio de Biología Molecular de la Enfermedad de Chagas (LaBMECh), Instituto de Investigaciones en Ingeniería Genética y Biología Molecular (INGEBI), National Research Council (CONICET), Buenos Aires, Argentina

[2]Fundación Instituto Leloir, CONICET, Buenos Aires, Argentina

[3]Institut Cochin, Département Maladies Infectieuses, INSERM U567, Paris F-75014, France

[4] International Centre of Insect Physiology and Ecology (*icipe*), Kenya

Abstract: The trypanosome ribosomal P proteins are located on the stalk of the ribosomal large subunit they where play a critical role during the elongation step of protein synthesis. The single chain recombinant antibody (scFv) C5 recognises the conserved C-terminal end of the *Trypanosoma cruzi* ribosomal P proteins although it possesses very low affinity for the corresponding mammalian epitope. We show that scFvC5 is able to specifically block the protein synthesis *in vivo* in Try*panosoma brucei,* the causative agent for sleeping sickness in humans and nagana in cattle. In transgenic parasites, the mRNA corresponding to scFvC5 could be amplified by RT-PCR and the protein product was detectable by *Western blot*. Indirect immunofluorescence assay shows a cytoplasmic distribution similar to that obtained with purified scFvC5 on *wild type* parasites. Transfected parasites showed abnormal cell size, accompanied by growth arrest, consistent with a blockade of protein synthesis. Analogues to trypanosome scFv C5 antibody could be potential candidates as novel anti-parasitic agents.

Key words: *Trypanosoma cruzi*; *Trypanosoma brucei;* ribosomal P proteins; single-chain Fv fragments (scFv); intrabodies.

POSTER 10: In Silico Design of Diagnostic Primers for the Identification of Phytopathogen of the Genus *Xanthomonas*

Nelson Ndegwa, Gurdeep Lall, and Dong-Jin Kim

International Institute of Tropical Agriculture (IITA), P. O. Box 30709-00100, Nairobi, Kenya

Abstract: Genome-scale comparisons of plant pathogen sequences are being made possible by the increasing availability of completely sequenced bacterial genomes. There is need to design pathogen diagnostic kits for the rapid and specific identification of newly emerging bacterial strains. These comparisons offer the opportunity of designing polymerase chain reaction (PCR) primers based on the variability of the intergenic regions for the amplification of products from target taxa and not from non-target taxa for the identification of pathogens at various taxonomic levels. In this study, an in silico bioinformatics approach is used to design primers from the *Xanthomonas* genomes available at the Comprehensive Microbial Resource website of The Institute of Genomic Research (http://www.cmr.tigr.org). *Xanthomonas axonopodis* pv. citri 306 was chosen as the reference species and the intergenic regions that are 400–800bp in length were viewed to see genes that flank the 5′ and 3′ end of these regions, conserved and in synteny across the *Xanthomonas campestris, X. axonopodis* and *X. oryzae* species after performing a homology search. A multiple alignment of the intergenic regions flanked by these conserved genes was done and analysed using Amplicon software. Preliminary results support the logic that intergenic-targeted primers could be designed from the conserved region for specific identification of *Xanthomonas* bacteria at the genus, species and pathovar level.

POSTER 11: The Identification, Isolation and Biochemical Characterization of the Virulent Surface Protein (Pf-EMP1) from Placental Extracts in Pregnant Women and the Use of Synthetic Peptides in Malarial Vaccine Research

P. K. Onyango and O. C. Onyango
Egerton University, P. O. Box 536, Njoro, Kenya

Abstract: The process whereby erythrocytes containing mature forms of *P. falciparum* adhere to micro vascular endothelial cells of vital organs and thus disappear from circulation is known as sequestration (cytoadherence). This process is mediated by a family of strain specific high molecular weight parasite derived proteins known as P.*falciparum* Erythrocyte Membrane Protein 1. This protein is exported to the surface of infecting erythrocyte where it is anchored to a submembraneous accretion of parasite derived Histidine-rich protein which becomes the points of attachment to the vascular endothelium. This adhesive protein is the only parasite protein unequivocally present on the outside of the erythrocyte.

This project therefore, aims at identification, isolation and biochemical characterisation of this variant surface protein (pf–EMP 1) which mediates sequestration in placenta of infected women and is believed to be the main antigen determining the parasite population structure during chronic malarial infection of *falciparum*. The isolated protein will be characterised to determine the amino acid composition as well as the N-terminal amino acid sequence. Depending on the sequence obtained, chemically synthesised peptides will be used to study immunological reactions of the protein when induced as an immunogen.

Monoclonal antibodies will be produced against the purified protein and since they are site specific reagents, they will be applied in the investigations of the protein to identify the immunogenic regions in the conformation of the protein. The synthesised peptides will act as an antigen for specific immunological purpose for example, vaccine production to intractable or difficult pathogens like *Plasmodium*.

POSTER 12: Screening for Anti-*Plasmodium* Activity Using Extracts from Extremophiles Found in Lake Bogoria, a Kenyan Soda Lake

S. W. P. Okanya[1], S. Martin[2], F. J. Mulaa[1]
[1]Department of Biochemistry, University of Nairobi, P. O. Box 30197-00100 Nairobi, Kenya
[2]US Army – Walter Reed Project/ KEMRI, KEMRI HQS, Nairobi, Kenya

Abstract: Recent studies have shown that microorganisms living in extreme environments are a potential source of medicinal compounds from their secondary metabolites. The Kenyan Rift Valley has several soda Lake and hot spring environments that are inhabited by extremophiles, whose medicinal value has not been investigated. This study therefore aimed at isolating and extracting secondary metabolites from extremophiles with bioactivity against *P. falciparum* parasites and eventual characterisation and molecular identification of these isolates.

Water samples were collected from soda lakes and streaked on alkaline agar media. The bacterial colonies that grew were picked and replica plated on fresh agar plates. Forty-three pure colonies were isolated based on colony morphology and were designated as independent isolates. These colonies were incubated in liquid media for ten days at 45° C to produce secondary metabolites. The metabolites were extracted by methanol and screened for anti-malarial activity.

Five isolates with the highest activity against *P. falciparum* were identified and classified using the 16S rDNA analysis. PCR amplification of the 16S rDNA produced a 1.5kb fragment, which was sequenced. Phylogenetic analysis of the partial sequences showed that the isolates cluster together with *Bacillus licheniformis, Bacillus sonorensis, Bacillus subtilis,* and *Pseudobacillus carolinae* species.

Key words: Extremophiles, secondary metabolites, anti-malarial activity, 16S rDNA, *P. falciparum*

POSTER 13: Identification and Characterization of Deoxyribose Phosphate Aldolase (dpa) in *Toxoplasma gondii*

Biscah Munyaka[1], Gabriela Colman[2], Makoto Igarashi[3], George Dautu[3]
[1]Kenya Medical Research Institute (KEMRI), P. O. Box 54840-00200, Nairobi, Kenya
[2]Servicio Nacional de Salud y Calidad Animal (SENACSA), 595-21-584-496, San Lorenzo, Paraguay
[3]National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080-8555, Japan

Abstract: *Toxoplasma gondii* is an obligate intracellular parasite which has emerged as one of the most common opportunistic infections in HIV/AIDS patients. Parasites can escape immune attack, by differentiation into bradyzoites (cysts) which reside predominantly in the brain, and are able to persist for the life of the host. Toxoplasmosis in AIDS patients is considered to be as a result of reactivation of latent infection. In the expressed sequence tag data base, deoxyribose phosphate aldolase (Dpa) was found to be specifically expressed in the *Toxoplasma gondii* cyst (bradyzoites). We analysed the expression of *Toxoplasma gondii* Dpa (TgDPA) in toxoplasma bradyzoites and tachyzoites by RT-PCR and for the first time we here by report that TgDPA is highly expressed in the *Toxoplasma bradyzoites*. The hypothesis of the present study was that dpa is involved in cyst formation and therefore the involvement of TgDPA in cyst formation is very crucial as an entry point in drug development against Toxoplasmosis based on cyst formation. A gene encoding TgDPA was identified in the *Toxoplasma gondii* strain RH by comparative sequence analysis and gene cloning. The gene encoded a protein of 286 amino acids, having a predicted molecular weight of 31kDa. The gene was cloned into pGEX-5X1 vector and the resulting recombinant plasmid pGEX-5X1/dpa was transformed into *Escherichia coli* DH5α and the recombinant protein was expressed then used for antibody production in mice and in rabbit. The antibody titers against the protein were analysed by western blotting. The results indicated the successful production of anti-Dpa antibody, which react with a 34kDa protein. Immunofluorescent antibody test in the parasites showed that TgDPA was localised in bradyzoites cytoplasm. These findings suggest that dpa could be involved in cyst formation and has a potential to impact on public health initiatives

Key words: *Toxoplasma gondii*, Deoxyribose phosphate aldolase, Recombinant protein, latent infections, AIDS-Toxoplasmosis.

POSTER 14: Trends in *Plasmodium falciparum* Genotype Mutations Associated with Antimalarial Drug Resistance in Kenya

Pamela Liyala[1] Fredrick Eyase[1], Rachel Achilla[1], Hoseah Akala[1], Julia Wangui1, Josphat Mwangi[1], Finnley Osuna[1], Meshack Wadegu[1], Rosemary Nzunza[1] , Rodney L. Coldren[1], Norman C. Waters[1], ±, *, and Sheryl Bedno[1]
[1]United States Army Medical Research Unit-Kenya/Centre for Clinical Research, KEMRI, P. O. Box 29893-00202 Nairobi
[2]Division of Experimental Therapeutics, Walter Reed Army Institute of Research, Silver Spring, Maryland, 20910

Abstract: *Plasmodium falciparum* is responsible for the high mortality rate of malaria in sub-Sahara Africa. In an attempt to monitor this disease and provide valuable information to the anti-malarial drug programme, we have surveyed malaria parasites from several geographically distinct areas in Kenya that represent varying degrees of drug resistance, drug usage patterns and levels of malaria transmission. Malaria specimens collected from Busia, Malindi, Mumias, and Kisumu were evaluated for drug susceptibility against a panel of fifteen known anti-malarial drugs. DNA was isolated from these samples for the determination of specific mutations in the genes that have been correlated with drug resistance which include *Pfdhfr*, *Pfdhps*, *Pfmdr1*, and *Pfcrt*. We find a decrease in the number of mutations in *Pfmdr1* and *Pfcrt* from parasites isolated from areas where chloroquine is rarely used to treat malaria. In contrast, parasites collected from areas that are believed to still treat malaria with chloroquine have substantial mutations in *Pfmdr1* and *Pfcrt*. *In vitro* drug sensitivity assays demonstrate that a decrease in the number of mutations in *Pfmdr1* and *Pfcrt* correlates with an increase in chloroquine susceptibility. In those areas of Kenya where Fansidar (Sulphadoxine-Pyrimethamine combination) has replaced chloroquine as the first line treatment for malaria, we find an increase in mutations in *Pfdhps* that corresponds to an increase in sulfadoxine resistance. We also did not observe any mutations at codons 164 for *Pfdhfr* and 581 and 613 for *Pfdhps*. The

absence of these mutations is consistent with parasites isolated from East Africa during previous studies and reflects a lesser degree of drug resistance as that seen in Southeast Asia.

POSTER 15: Characterization of Antimalarial Drugs Resistant *P. falciparum* Isolates, Using Dot Blot Hybridization Molecular Technique [Poster was not presented]

F. T. Kimani[1, 2], S. A. Omar[1], J. N. Ngeranwa[2], J. Machuka[2]
[1]Centre for Biotechnology Research and Development- Kenya Medical Research Institute (KEMRI), [2]Department of Biochemistry and Biotechnology, Kenyatta University

Abstract: One of the major obstacles to the control of malaria is the emergence and spread of parasites resistant to anti-malarial drugs. With the current trend in malaria treatment policies, we aimed at assessing the level of resistance to one of the artemesinin combination therapy candidate sulfadoxine/ pyrimethamine in an epidemic area of Kenya. Standard clinical efficacy testing *in vivo,* as recommended is time consuming and expensive method to survey levels of drug resistance, especially for large scale epidemiological surveys. In this study we aimed at adopting and evaluating a radionuclide based technique for use in our settings and for use in drug monitoring at an epidemiological scale.

In this study the dot-blot/probe hybridisation technique was adopted alongside the conventional PCR-RFLP technique Pre treatment peripheral blood were collected by lancet prick from a fingertip for direct microscopic diagnosis and some amount of blood obtained by lancet prick was spotted directly onto Glass Fiber Membrane (GFM) or on 3MM Whatman filter paper. Parasitic DNA was extracted using GFM and Chelex-100® methods respectively. PCR amplifications were done using specific primers for RFLP and blotting. RFLP was by the use of specific restriction enzymes by overnight digests. Single stranded fragments were blotted on nitrocellulose membranes for probing using [Gama-^{32}P] dATP 5′ end labeled allele specific probes The nature and distribution of mutations in the *dhfr* gene (chromosome 4) in codons 51, 59 and 108 were profiled in Kisii, a *P. falciparum* malaria epidemic area in Kenya.

Mutation profiles after dot blotting *dhfr* 108 showed that there were mutations in 69 (58.9%) of the study sample with no wild type, *dhfr* 59 showed 14 (11.97%) wild and 67 (59.82%) being mutant type while *dhfr* 51 had 2 (1.7%) wilds and 70 (59.82%) being mutants. Fifty six (47.8%) of our study population carried mutations in the three codons that were studied and these represented 80% of the PCR positive samples. Sixty one (52.1%) of these samples had double mutations in *dhfr* codons 108 and 59 and this was equivalent to 87.1% of the PCR positive isolates. Fifteen (12.82%) isolates which had been characterised as wild type by RFLP turned out to be mixed infections by radio probing. Moreover 29 (24.79%) of the isolates which had been characterised as mutant type were actually mixed by radio probing.

There was high failure rate on the drugs and thus S/P drugs are not good combination candidates in this epidemic area. The drug being a failing option would lead to resistance due to drug pressure selection. The dot blot technique was found to have similar levels of specificity but more sensitive than PCR/RFLP. Moreover it was found to be suited for large-scale epidemiological surveys of genes associated with antimalarial drug resistance. From the study we recommend that this technique be adopted even in the monitoring of the efficacy of the new antimalarial and thus detect any possible resistance early enough for appropriate policy change.

POSTER 16: Rate of Substitution, Date of Emergence and Speed of Dispersal of Rice Yellow Mottle Virus in Africa

F. Sorho [1,6], A. Pinel-Galzi[2], O. Traore[3], M. S. Rakotomalala[4], E. Sangu[5], Z. Kanyeka[5], E. Hebrard[2], G. Konte[3], Y. Sere[1], D. Fargette[2], S. Ake[6]
[1]Africa Rice Center (WARDA), 01 BP 2031, Cotonou, Bénin
[2]IRD, BP 64501 F-34394 Montpellier cedex 5 France
[3] INERA, Laboratoire de virologie et de biotechnologie végétale, 01 BP 476, Ouagadougou, Burkina-Faso
[4]FOFIFA, BP289 Mahajanga 401 Madagascar
[5]University of Dar es Salam, P. O. Box 35060, Tanzania
[6]Laboratoire de Physiologie végétale, UFR Biosciences, Université Cocody-Abidjan, 22 BP 582 Abidjan 22, Côte d'Ivoire

Abstract: Rice yellow mottle virus sobemovirus (RYMV), causal agent of the Rice yellow mottle disease transmitted by beetles, is at the origin of an emergent disease of rice in Africa. Isolates of each of the countries where the disease was observed were sequenced. The spatial structure of the diversity, the relations between genetic and geographical distance, as well as the phylogeny suggest an initial diversification of RYMV in East Africa, followed by a temporal spread from east to west of the African continent. The objective of this study is to assess the rate of substitution of RYMV, then to infer the dates of emergence and the speed of dispersal throughout Africa. In this study, we use BEAST, a cross-platform programme for Bayesian Markov Chain Monte Carlo (MCMC) analysis of molecular sequences. The rate of nucleotide substitutions (per site and per year) of the RYMV dates of divergence from temporally spaced sequence data were estimated by Bayesian inference from the sequences of a large collection of isolates collected throughout the epidemic history of the virus (40 years). The temporally spaced data help to follow the accumulation of mutations over time and thus to estimate the mutation rate using coalescent-based population genetic inference. This rate of change, the first one for a plant virus, was compared with those of a set of animal and human viruses. It then served to calibrate the phylogenetic tree of RYMV. Dates of appearance of the virus in various regions of Africa were determined under the hypothesis of relaxed molecular clock using Bayesian MCMC method implemented in BEAST. Finally, the speed of dispersal of RYMV throughout Africa was estimated. These estimations were compared with the epidemiological data on rice yellow mottle disease, with the history of rice in Africa and with the dispersal modes of the vector beetles to identify factors responsible of the emergence of the RYMV.

POSTER 17: An *in silico* Immunological Approach for Prediction of T Cell Epitopes of *Leishmania major* Proteins in BALB/c Susceptible and C57BL/6 Resistant Murine Models of Infection

H. Ben-Abdallah, F. Z. Guerfali, R. M. Sghaier, G. Mkannez, K. Dellagi and D. Laouini
Laboratoire d'Immuno-pathologie, Vaccinologie et Génétique Moléculaire and Laboratoire International Associé CNRS. Institute Pasteur of Tunis, Tunisia.

Abstract: It is well established that MHC class II restricted CD4 T cells are dominant during the development of immunity against *Leishmania* (*L*) in C57BL/6 resistant mouse strain. However, and in agreement with a number of previous observations indicating that specific CD8 T cells are primed during natural infection or vaccination in humans, several evidences obtained recently with the BALB/c susceptible murine model of infection by *L. major* indicate that CD8 T cells participate in both pathogenesis and immunity to cutaneous lesihmaniasis.

Our goal herein was to identify *in silico*, using several public computational systems for the prediction of peptides binding to all MHC (histocompatibility complex-2) molecules in BALB/c and C57BL/6 mice (Syfpeithi, Rankpep, PRED$^{BALB/c}$ and BIMAS), all parasitic peptides present in the whole *L. major* predicted proteome (available at GeneDB-Sanger Institute). Peptides that were predicted to bind to different H2 molecules were then analysed for their homology with any of the murine proteins annotated so far, using the BLAST algorithm. Sets of selected peptides for each H2 molecule, defined by different prediction systems, were then compared to each other.

Surprisingly, the results showed that a higher number of *L. major* peptides was predicted to bind H2 BALB/c molecules and very few or none to bind H2 C57BL/6 ones. These results can explain experimental findings showing a stronger role, played by CD8 T cells in BALB/c susceptible strain than in C57BL/6 resistant strain. This shows how a hybrid immuno-computational approach may be useful for biologists to target an *in silico* set of selected proteins to define potential candidate antigens for experimental vaccination with more accuracy and yet reduced number of T cell antigens.

POSTER 18: Population Genetic Studies of Tsetse Flies (Diptera: Glossinidae): The Usefulness of Genomic Resources

J. O. Ouma[1], G. Marquez[2], E. S. Krafsur[2]
[1]Department of Biochemistry, Trypanosomiasis Research Centre, Kenya Agricultural Research Institute, P. O. Box 362-00902, Kikuyu, Kenya
[2]Department of Entomology, Iowa State University, Ames, Iowa 50011 USA

Abstract: Tsetse flies, *Glossina* spp. (Diptera: Glossinidae) are medically and economically important insects confined to sub-Saharan Africa. They are exclusively blood feeding and are the sole vectors

of African trypanosomiases. Proper understanding of rates of gene flow and genetic differentiation among tsetse populations is key to the success of area-wide tsetse control using genetic methods. Ecological studies indicate that tsetse flies in the *morsitans* group have great capacity for dispersal and colonisation of suitable habitats. However, genetic data from nuclear and mitochondrial DNA markers seem to indicate otherwise. The apparent inconsistency between ecological and genetic estimates begs further research. We investigated the breeding structure of the tsetse fly, *Glossina pallidipes* at micro- and macro-geographic scales by analysing spatial and temporal variation at eight microsatellite loci. Our data suggest that there is little gene flow between populations separated by as few as 50 km leading to strong measures of genetic differentiation. These results show that *G. pallidipes* tend to remain close to their neighbourhoods. However, the number of simple sequence repeat (SSR) markers used in this study were very few and could have led to substantial inter- and intra-locus variance. Ongoing *Glossina* genome sequencing project offers an opportunity to mine the data for more SSR markers. The potential benefits of this approach are herein discussed.

POSTER 19: Host-plant Diversity of *Sesamia calamistis* Hampson (Lepidoptera: Noctuidae): Cytochrome *b* Gene Sequence Reveals Genetic Differentiation in Host Utilization

G. O. Ong'amo[1,3*], B. P. Le Ru[1] P. Moyal[2] C. K. P. O. Ogol[3], E. D. Kokwaro[3] and J.-F. Silvain[2]

[1]Unité de Recherche IRD 072, International Centre of Insect Physiology and Ecology (*icipe*), P. O. Box 30772-00100 Nairobi, Kenya

[2] Unité de Recherche IRD 072, CNRS, Laboratoire Populations, Génétique et Evolution, BP1, 91198 Gif-sur-Yvette cedex, France

[3]Department of Zoological Sciences, Kenyatta University, P. O. Box 43844 Nairobi, Kenya.

Abstract: *Sesamia calamistis* is one of the indigenous stem borer pests associated with maize and sorghum in Africa. In a study conducted at Mtito Andei in Kenya to assess its diet breadth and genetic structure, *S. calamistis* larvae were found on eight different plant species from which 26 haplotypes were found. The haplotypes differentiated in two clades with respect to host plants. The first clade of 16 haplotypes was found mainly on maize while the other clade of 10 haplotypes was found mainly among wild hosts. These clades were separated by 10 hypothetical haplotypes which are probably extinct suggesting that differentiation may have taken place long time ago.

Key words: *Sesamia calamistis*, *Zea mays*, wild hosts, genetic differentiation, haplotypes

POSTER 20: Molecular Epidemiology of *Xanthomonas campestris* pv *musacearum*, The Causal Agent of *Xanthomonas* Wilt of Banana and Enset: 1968–2006

V. Aritua[1], N. Parkinson[2], R. Thwaites[2], D. R. Jones[2], W. Tushemereirwe[1] and J. Smith[2]

[1]National Agricultural Biotechnology Centre, Kawanda Agricultural Research Institute, P. O. Box 7065, Kampala, Uganda

[2]Central Science Laboratory, Sand Hutton, York YO41 1LZ, UK

Abstract: *Xanthomonas* wilt of enset and banana was first described in Ethiopia in 1968 and is now a serious bacterial disease that is spreading in East and Central Africa. Earlier studies identified *Xanthomonas campestris* pv. *musacearum* as the causal agent of the wilt, but analyses have not been undertaken using modern techniques to confirm this taxonomy. This paper describes the characterisation of strains of the bacteria that were isolated between 1968 and 2006. Phylogenetic analysis of partial nucleotide (nt) sequences of the gyrase B gene and Internal transcribed spacer (ITS) region, genomic amplicon fingerprints using Repetitive sequence PCR and fatty acid methyl esters showed that all strains of *X. campestris* pv. *musacearum* belonged to same genotype. Twenty strains originating from Ethiopia, Uganda, Democratic Republic of Congo and Rwanda had no sequence identities above 98%. When compared to other bacterial species, the pathogen was found to not be related to *X. campestris* but very similar to *X. vasicola*, which formerly comprised pathogens of sorghum (*X. vasicola* pv *holcicola*). Further, a group of sugarcane and maize pathogens, that were known to be atypical of *X. axonopodis* pv. *vasculorum* and had a proposed re-classification as *X. vasicola* pv *vasculorum*, were found to be the most similar to the banana and enset strains. Together, our data supports the reclassification of X. *campestris* pv. *musacearum* as *X. vasicola* pv. *musacearum*. In addition, this study revealed that the recent occurrence of the banana disease in Uganda and other East African countries is a consequence of the spread of the previously recognised *musacearum* pathovar. Possible hypothesis to explain evolutionary pathways of the three strains of *Xanthomonas* infecting banana, sorghum, sugarcane and maize are proposed.

POSTER 21: HIV/AIDS Susceptibility and Resistance. Immunoinformatics Approach

Muriira Geoffrey Karau
Department of Biochemistry and Molecular Biology, Egerton University, P. O. Box 536 Njoro, Kenya

Abstract: Genetic resistance and susceptibility to infectious diseases involves a complex array of immune-response and other genes with variants that impose subtle but significant consequences on gene expression or protein function. Numerous studies have identified a role for HLA genotype in AIDS outcomes, implicating many HLA alleles in various aspects of HIV disease. The complexity of the immune responses can be tackled by high thorough put computational methods. Here we would like to show that there are individual susceptibility and protections to HIV-1 due to diversity of HLA-A in the population.

We eluted envelop glycoprotein of 10 HIV-1 isolates against 17 HLA-A alleles with an external MHC binding prediction tool freely available on line. We then calculated all the top ten epitopes from each alleles and ranked them with the highest number of epitopes representing the resistant and lowest the susceptible group.

In the 17 alleles queried 8 were susceptible to HIV-1/AIDS with number of epitopes less than 38, while 9 were protective with epitopes number greater than 38. HLA-A *3002 allele had the highest number of epitopes suggesting that individuals with this HLA-A are likely to generate strong immune responses to HIV-1 envelop glycoprotein. Our data suggests that HIV-1 immune responses on the envelop glycoprotein depends on the HLA-A allele. This approach is significant in determining epitopes that can be used in the design of the epitope based vaccine and in epidemiological studies of resistance and susceptibility to HIV-1.

POSTER 22: Phylogenetics of HIV-1 in Heterosexual Cohort of Discordant Couples in Nairobi, Kenya

N. F. Nindo[1], M. Matu[2], C. Obuya[1]
[1]Clinical Trials Laboratory, Department of Obstetrics and Gynaecology, Kenyatta National Hospital, Nairobi Kenya
[2]Centre for Microbiology Research (CMR), Kenya Medical Research Institute

Abstract: Epidemiological and cohort approaches to HIV-1 transmission dynamics can now be complemented by phylogenetic data. HIV-1 transmissions in Africa occur between married adults who are discordant for their HIV-1 infection status. In Africa, all known HIV-1 genetic subtypes and groups including groups N and O are present. Whether the various groups, subtypes and recombinant forms of HIV-1 have biological differences with respect to transmissibility and the course of disease progression is not known and most studies have focused on human host immunological responses and not the evolutionary mechanisms of the HIV-1 virus. For this reason it is important to study the phylogenetic distribution of the HIV-1 genetic subtypes in Kenyan cohort of discordant couples with major focus on the HIV-1 seronegative partners who seroconvert during follow-up in prophylactic interventions and establish any epidemiological linkage. To date, 3000 couples attending voluntary counseling and testing centres (VCTs) in Nairobi, have been tested for HIV-1, of whom 8% were HIV-1 discordant, 9.2% were concordant HIV-1 positive and 82.8 % were concordant HIV-1 negative. The HIV-1 discordant group will be enrolled and incidence and predictors of heterosexual transmission monitored at 3-month intervals for seroconversion of the seronegative partner. Gp120 and P24 regions will be amplified by PCR analysis of blood samples from both partners and sequenced. In the final set of experiments, evolutionary linkage to incidence will be assessed by phylogenetic tree analysis. It is expected that phylogenetic data derived from this study will shade important insights into HIV-1 transmission dynamics in discordant couples in the developing world.

POSTER 23: Genetic Polymorphism and Positive Selection in a 'Concealed' Gut Potential Vaccine Antigen from *Rhipicephalus appendiculatus*

Kamau Lucy[1], R. Skilton[2],T. Shah, E. Kabiru[1], A. Orago[1], T. Musoke[2], R. Bishop[2]
[1]Kenyatta University, P. O. Box 43844 Nairobi, Kenya
[2]International Livestock Research Institute (ILRI), P. O. Box 30709 Nairobi, Kenya

Abstract: *Boophilus microplus* Bm86 based commercial vaccines control *Boophilus decoloratus* and *B. annulatus* from different parts of the world. They showed potential for control of *Hyalomma*

anatolicum and *H. dromedarii* but insufficient cattle protection against *Rhipicephalus appendiculatus*, an important livestock tick in sub-Saharan Africa. In this study, polymorphism in the Bm86 homolog in *R. appendiculatus* isolated from four Kenyan field populations and a laboratory stock was characterized in view of developing an effective vaccine against this tick.

An estimated 2 Kb cDNA encoding Bm86 homologue in *Rhipicephalus appendiculatus,* Ra86 Muguga was isolated from a cDNA library constructed using cDNA synthesised from gut RNA extracted from *R. appendiculatus* Muguga Laboratory strain. Nine full-length, including predicted signal peptide and membrane anchor and 10 truncated sequences were analysed. Two allele types were demonstrated; a full-length type, potentially encoding 693 amino acids (aa) and a shorter type (654 aa). Five of 19 (26%) sequences were of shorter type and the remaining 74% were of the longer type. A 40 amino acids deletion or insertion near the C terminal, and 129 (20%) amino acid substitutions from single nucleotide polymorphisms differentiated the two proteins. Combining Ra86 Muguga and field isolates, additional deletions (39-48 aa) led to more Ra86 size types, designated 1, 2, and 3 in addition to type 4 and 5 resulting in a total of 5 size types. Type 5 was isolated only from Ra86 Muguga while types 1, 2 and 3 were isolated from the field. Type 4 was isolated from the field and laboratory. It was shown that the polymorphism was due to positive selection, using selective neutrality and neutral evolution tests. Regions under-going positive were demonstrated along the gene using WINA window analysis. Positive selection in the gene was associated with antigenic and other functional role of the concealed antigen.

POSTER 24: Rational Vaccine Design: The Challenges of Adenoviral Vectors

C. O. Esimone[1], C. S. Nworu[2]*, M. U. Adikwu[1], P. A. Akah[2]

[1]Department of Pharmaceutics

[2]Department of Pharmacology and Toxicology Faculty of Pharmaceutical Sciences, University of Nigeria, Nsukka

Abstract: In our recent study, we developed HIV and adenoviral vectors as model probes for antiviral screening. This viral–vector-based assay were found to be rapid, specific, reproducible and sage. Our present research focuses on the use of computational and predictive tools to circumvent the major challenges limiting the optimal use of adenoviral vectors in vaccine deliveries. Low pathogenicity, high infectivity and ability to accommodate large size transgene have made adenoviruses useful and popular candidate in vaccine delivery. The occurrence of circulatory adenoviral antibodies in human, induced by passive infection limits the vector titre and hence the delivered vaccine titre. The second challenge is the use of computational methods to circumvent the possibly of disseminated adenoviral diseases in replication-competent viruses and the problem of low vector titre with replication-defective vectors. The third challenge is to predictively construct adenoviral vectors with high ability to infect replicating and non-replicating cells.

Key words: Vaccines, adenoviral vectors, antibody titre

POSTER 25: Bioinformatics Characterisation of Ligand-binding Hotspots in *P. falciparum*

Daudi Jjingo

Makerere University, Faculty of Computing and IT, P. O. Box 7062, Kampala

Abstract: In this poster I will use an example of a *P. falciparum* Surface protease (enzyme) to illustrate a putative methodology of using existing Bioinformatics software and tools, first to identify, and then characterise Binding sites/hotspots on the protein surfaces of putative drug targets. These characterised binding sites can then be used for subsequent anti-malarial drug design.

The WHO/TDR Malaria strategic research project, through its product research and development, has identified several molecular targets in *Plasmodium* that could be exploited for the development of new drugs against malaria. These include the protease responsible for secondary processing of the *P. falciparum* merozoite surface protein (MSP-1) (http://www.who.int/tdr/research/progress/mal_str/drugs.htm). Though the TDR Research and development has gone on to design and synthesise three possible inhibitors of this protease, there is still room for development of other inhibitors. Proper and accurate Drug (inhibitor) design is determined, among other things, by accurate knowledge of the target protein's surface. I use Sequence retrieval to obtain the protein sequence of the *P. falciparum* surface protease from the NCBI protein database (http://www.ncbi.nlm.nih.gov/entrz/query.fcgi?db=Protein). This is followed by sequence comparison (blast) with the entire Protein Data

Bank (PDB) (1, 2), to identify the closest protein (e-value must be >1* -10) with known 3D structure. The identified protein is then taken through QsiteFinder (3), a binding site prediction method which uses an energy based method to predict binding spots on the protein's surface. The predicted sites are then characterised by volume, coordinates and surrounding protein residues. This avails a reasonable catalogue of feasible binding hotspots on the *P. falciparum* surface protease, which could be used for drug (inhibitor) design.

POSTER 26: Exploring the Insect Acetylcholinesterase (AChE) Active Site Gorge: Toxicokinetic and AChE Sequences Analysis as Prospects to Molecular Design of Selective Insecticides

J. Mutunga, T. Anderson and J. Bloomquist
Neurotoxicology Laboratory, Entomology Department, Virginia Polytechnic Institute and State University, Blacksburg, VA-24061 USA

Abstract: Most insects express two acetylcholinesterase (AChE) genes, *ace*-1 and *ace*-2. The *ace*-1 codes for AChE-1, which has a functional role in hydrolysis of acetylcholine during signal transduction in insect nervous systems. Architecture of AChE active site gorge is characterised by functional motifs comprising a hydrophobic pocket, the H-bond network, the acyl pocket, and the peripheral site. Screening for differential interaction of bivalent ligands occupying both the catalytic and peripheral sites, with insect and vertebrate AChEs is important in the design of selective insecticides.

Dimeric tacrines with varying methylene linkers (C_2–C_{12}) were used to probe *B. germanica* (CSMA strain), *A. gambiae* and human AChE active site gorge and data for bovine AChE was adapted from a study by Munoz-Ruiz *et al* (2005). Dimeric tacrines were found to be more potent in vertebrate AChE, especially bovine, than in insects (Potency was ca. 16000 and 82500-fold higher for bovine AChE than for *Bg*AChE and *Ag*AChE respectively). The potency of dimeric tacrines was more tether length dependent in vertebrate AChE than in insect AChEs. Insect AChE had unique C_6 bump whereby the potency decreased 7-fold and 2-fold for *Bg*AChE and *Ag*AChE, respectively. Such a change occurs in a difference of a single carbon length indicating a significant interaction of the bivalents at C_5 compared to C_6 with the active site amino acid moieties. Further tests are on-going with other insects to confirm our hypothesis that this bump is unique in insects. We further performed a sequence analysis for the *ace* 1 sequence using the DeepView-Swiss-PdbViewer (http://expasy.org/spdbv/) to determine the 3D arrangement of catalytic signatures in the AChE active site gorge. We used ClustalW for alignment and homology analysis and the Treeview in the visualisation of the cladogram for the sequences. Possible applications of the 'C_6 bump' and differential interaction of catalytic motifs of AChEs to the molecular design of highly potent, insect-specific anticholinesterases that are less toxic humans will be discussed.

POSTER 27: Maesanin: A Benzoquinone from *Maesa lanceolata* that Completely Inhibits Respiration in Bloodstream *Trypanosoma brucei brucei*

L. W. Kariuki, E. K. Nguu, R. M. Njogu, P. W. Kinyanjui, J. O. Midiwo, W. D. Bulimo and J. K. Kiaira

Abstract: Maesanin a naturally occurring benzoquinone from plant *Maesa lanceolata* that inhibits growth of a wide range of microorganism. It has been shown to have antibacterial, antifungal antihelminth and more recently antitrypanosomal activity was established. This suggests that it could have some important pharmacological activity. To gain further insight into the mechanism of inhibition by maesanin, glucose catabolism in bloodstream and procyclic *T. b. brucei* was studied. Respiration by the bloodstream forms of *T. brucei brucei* was studied at increasing concentrations of maesanin with glucose as the substrate. This was done by incubating about 5×10^7 trypanosomes at different concentration of measanin for 0 to 30 minutes. At 10 µg/ml maesanin, all the parasites were immotile after 10 minutes. This was compared with motility when parasites were incubated with Salicyl Hydroxamic acid (SHAM), cyanide, antimycin and a combination of cyanide and SHAM. Maesanin unlike the other inhibitors showed complete inhibition. Respiration by the bloodstream forms of *T. b. brucei* incubated in phosphate saline glucose (5 mM, glucose) at pH 8.0 was studied at increasing concentrations of maesanin. It was observed that maesanin steadily inhibited the rates of oxygen consumption and the trypanosomes died rapidly within 5 minutes of incubation. At concentration of 1 µg/ml maesanin per 10^8 trypanosomes, the O_2 consumption was completely

inhibited. Inhibition by (SHAM) and cyanide was about 80% and 90% under the same conditions respectively. However the parasites were not completely immotile as in the case of measanin. The rate of pyruvate production was also studied at increasing concentrations of measanin. It was observed that increasing concentration of measanin decreased the rate of pyruvate production by *T. brucei brucei*. The above results suggested that maesanin could be inhibiting a vital energy production step in these parasites.

POSTER 28: Transfection of *Plasmodium berghei* to Express Green Fluorescence Protein (GFP) for Thiazole Kinase Drug Screening System

M. O. Gorreti[1], W. R. Mukabana[1], F. J. Mulaa[1] and H. S. Ozwara[2]
[1]School of Biological Sciences, University of Nairobi, P. O. Box 30197-00100 Nairobi, Kenya
[2]Institute of Primate Research, Nairobi, Kenya

Abstract: Malaria caused by the protozoa of the genus *Plasmodium* and transmitted by mosquitoes belonging to the genus *Anopheles gambiae* is the most serious tropical and sub tropical disease causing an estimated 300–500 million clinical malaria cases and between one and three million deaths (mostly children).

Genetic variation, a weak immune response against the parasites and a poor understanding of the malaria parasite biology has hindered vaccine and new drug development.

Genetic manipulation of malaria parasite would revolutionise the study of these parasites and have implications for vaccines and drug development. Transfection technology has laid down the ground work for performing targeted gene knockouts in *Plasmodium* and at the same time monitoring expression of transgenes as well as elucidating the functions of proteins by disrupting, modifying or replacing the genes encoding them.

In this study, Hydroxyl ethyl thiazole kinase gene that encodes an enzyme that is very essential in the synthesis of vitamin B1 in the plasmodial pathway will be engineered into a plasmid designed to express GFP gene and a selectable marker and this will be used to stably transfect the clinically relevant intracellular blood stages of non-human malaria parasite *Plasmodium berghei*. The Plasmodium parasite has been shown to synthesise vitamin B1 *de novo,* while the same pathway *is* absent in the human host rendering it as a potential drug target.

Successfully transformed parasites of the rodent parasite will be selected on the basis of resistance to pyrimethamine.

Once a drug-resistant population emerges post-transfection, the stable transfectants will be genetically monitored and finally the reproducibility of the GFP and thiazole kinase expressing clones will be tested in *In vitro* drug assays.

POSTER 29: Sustaining Capacity Building and Implementing Bioinformatics at *Institut Pasteur de Tunis*

Alia Benkahla, Ikram Guizani, Helmi Mardassi, Balkiss Bouhaouala, Souha Ben Abderrazak, Slimane Ben Miled, Lotfi Chargui, Mehdi Chenik, Sonia Abdelhak, Abdeladhim Ben Abdeladhim, and Koussay Dellagi
Institut Pasteur de Tunis

Abstract: *Institut Pasteur de Tunis* (IPT) is a Tunisian research institution founded in 1883. Its missions are to conduct Research and Training activities on infectious diseases, R/D on vaccines, and Public Health Laboratory activities (PHL). Research and training programmes are mainly oriented towards national health and/or economic problems such as rabies, leishmaniases, tuberculosis, viral hepatitis, bovine theileriosis, veterinary microbiology, genetic disorders and scorpion or snake envenomations. Molecular biology applications perfectly integrated in the research programmes of the Institute can be tracked back to the early 90's, providing a unique context for promoting bioinformatics and its applications at the national and regional level, in highly relevant research areas, particularly for Health. The diversity of the research activities, their complementarities, their unique combination within the health research system, the levels of expertise generated, and the North-South and South-South collaborations developed, contribute to this unique position.

Bioinformatics is crucial and takes advantage of the last decade explosion of molecular biology and genomics data, to support and to be supported by the "wet research" laboratories.

The implementation of bioinformatics at IPT was done in six steps, with the support and following the recommendations of the UNICEF/WHO/UNDP/World Bank Special Programme for Research

and Training in Tropical Diseases (TDR), the *Réseau International des Instituts Pasteurs* (RIIP), the Ministry of High Education, and the Ministry of Health: (i) PhD and Post-Doctoral training of a resource scientist having a mathematics background (Sep 1996–Oct 2004); (ii) Continuous short term training of staff and students (from 1998 to date) through some of the high quality bioinformatics courses organised by TDR (N = 8), *Institut Pasteur* in Paris (N = 4), and ICGEB-Trieste (N = 3); (iii) A dynamics of bioinformatics research projects development has been initiated in 2001 through a project supported by TDR promoting institutional networking on leishmaniases and tuberculosis followed by other projects successfully funded by TDR and by *Institut Français de la Coopération* or to be supported by the European Commission (SysCo project); (iv) Upgrading the institute's hardware (1 PC per scientist, servers, backup system), software and network infrastructure (Local network: Ethernet (10/100) and 1Go for the backbone; Internet connection: Special line of 2 Mb/s) and dedicated buildings; (v) Participating actively to the organisation of international introductory or advanced bioinformatics courses, and offers a national Master degree in bioinformatics; (vi) Recruiting the trained bioinformaticist (Dec 2004), starting and equipping a bioinformatics group (March 2005). Four core functions have been assigned to the bioinformatics group: user support, bioinformatics support and establishment of an interface with "wet laboratories", in-house training, and R/D. As an added value to the development of its activities, the bioinformatics group have started collaborations with different bioinformatics centers (e.g.: the SANBI-South Africa, the Max-Planck Institute for Molecular Genetics-Germany, BIOBASE-Germany, IBDML-France) and is currently contributing to the development of adequate human resources.

Key words: bioinformatics, capacity building, research, training, TDR, RIIP, North-South collaborations, South- South collaboration, Tunisia.

POSTER 30: *In silico* Prediction of Protein-protein Interactions in MTB Infected Macrophages

Ouissem Souiai[1,2], Haithem Essid[1,3], Ines Abdeljaoued[3], Mourad Masigh[4], Javier de las Rivas[5], Helmi Mardassi[1], Bernad Jacq[2], Alia Benkahla[1] and Christine Brun[2]

[1]Institut Pasteur, Tunis, Tunisia
[2]Institut de Biologie du Développement de Marseille Luminy, France
[3]Ecole Supérieure de la Statistique et Analyse de l'Information, Tunis, Tunisia
[4]Institut Supérieur de l'Informatique, Tunis, Tunisia
[5]Bioinformatics and Functional Genomics Research group, Salamanca, Spain

Abstract: Tuberculosis (TB) is responsible for about 3 million deaths each year. A third of the human population carries latent TB. The pathogen uses poorly understood complex strategies to subvert macrophage-killing activities. We will present here an *in silico* global integrative approach aiming at predicting protein-protein interactions in infected and non-infected macrophages (host-host, pathogen-pathogen and host-pathogen). To achieve this goal, the set of MTB protein-protein interactions available in the public domain and the large set of human protein-protein interactions available in the Agile Protein Interaction DataAnalyzer (APID: http://bioinfow.dep.usal.es/apid/index.htm) will be put into physiological context, using expression data from the Gene Expression Omnibus database (GEO), literature mining and Gene Ontology annotations (GO). Then, using PPI network analysis methods based on graph theory PRODISTIN (Brun *et al*., 2003, Genome Biology; Baudot *et al*., 2006, Bioinformatics; http://crfb.univ-mrs.fr/webdistin/), the modular composition of each network will be determined and compared to each other. Differences in module composition between infected and non-infected macrophages are expected to point towards the biological processes which are preferentially targeted by the pathogen, activated in response to the infection and are responsible for the disease phenotype.

Currently, a statistical test is being developed to rank the reliability of APID interactions in order to better predict the set of proteins that is effectively interacting in infected and non-infected macrophages.

Key words: MTB infected macrophages, protein-protein interaction network analysis, comparative genomics, gene expression data, literature mining

POSTER 31: Building Functional Genomics Research Capacity on Insects Vector of Human Disease Vector in Africa

S. Doumbia[1], M. B. Coulibaly[1], S. F. Traore[1], G. Dolo[1], E. Adebiyi[2]
[1]DMEVE/ MRTC/Faculty of Medicine, University of Bamako, Mali
[2]Department of Computer and Information Sciences (Bioinformatics Unit), College of Science and Technology
Covenant University, Ota, Ogun State, Nigeria

Abstract: The genome sequences for many insects vector of human diseases are now available and promise the development of a set of new, powerful tools that can be used to develop innovative approaches to control these diseases. The African continent, which is the most severely affected by vector borne diseases, lacks adequate infrastructures and personal resources required for rational uses of genomic information. To fill this gap, the African Centre for Training in Functional Genomics of Insect vectors of Human Disease (AFRO VECTGEN) was initiated by WHO/TDR and the Department of Medical Entomology and Vector Ecology (DMEVE) of the Malaria Research and Training Centre (MRTC) in Mali. The aim of the AFRO VECTGEN programme is to train young scientists in functional genomics who will ultimately use genome sequence data for research on insect vectors of human diseases. Over the past 3 years, the African centre for training in insect disease vector functional genomics has contributed to the training of more than 40 junior scientists across the continent through a series of two-week workshops sponsored by TDR/WHO. The training is focused on 3 main areas: Molecular Biology with emphasise on molecular entomology, applied bio-informatics and microarray technology and data analysis. Research and training opportunities for carrier development in genomics research on insects vector of human diseases at the Malaria Research and Training Center will be presented.

Key words: Disease vectors, Bioinformatics, Functional genomics, Training, Africa

POSTER 32: The International Centre of Reference "Chantal Biya" (CIRCB): From Basic Research to Patient Care

Odile Ouwe Missi Oukem-Boyer
International Centre of Reference "Chantal Biya" (CIRCB), P. O. Box 3077, Yaoundé, Cameroon

Abstract: The International Centre of reference "Chantal Biya" (CIRCB) is a new International Research Centre established in February 2006. It aims at supporting HIV/AIDS and STI control programmes through various activities namely: basic and applied research, disease prevention and management and training. The CIRCB seeks to become an open environment for researchers, either national or international, in view of collaborative research within a global perspective of the reduction of disease burden, with an emphasis on research on vaccines against HIV/AIDS, Tuberculosis and Malaria.

The CIRCB laboratories provide a high-level technical expertise for diagnosis, follow-up, clinical and vaccine research against AIDS and other tropical infectious diseases. The technical platforms comprise the immunology, virology, biochemistry and haematology laboratories. For 2007, our planned activities are, among others, to establish the epidemiology and bioinformatics unit in order to analyse all our data from HIV sequences, particularly in terms of resistance of some subtypes to antiretroviral drugs. Also, an Influenza genome sequencing project is expected to be carried out at CIRCB very soon, in which bioinformatics will be an important part of the work.

Through the general presentation of the CIRCB, the importance of the Bioinformatics and Statistic area in our various research projects will be emphasised, in order to establish fruitful specialised contacts in this area, hence making the Centre move ahead.

POSTER 33: PhyML: Fast and Accurate Phylogeny Reconstruction by Maximum Likelihood

S. Guindon, J. F. Dufayard, W. Hordijk, V. Lefort, O. Gascuel
LIRMM, CNRS, 161 rue Ada, 34392, Montpellier Cedex 5, France
E-mail: gascuel@lirmm.fr, URL: http://www.lirmm.fr/mab

Abstract: We will present and make a demo of PHYML, a software that implements a fast and accurate heuristic for estimating maximum likelihood phylogenies from DNA and protein sequences. This tool provides the user with a number of options (a large range of evolutionary models, estimation of various evolutionary parameters, bootstrapping, fast LRT of branch supports, several tree space

search strategies, etc.), in order to perform comprehensive phylogenetic analyses on large data sets in reasonable computing time. This software is widely used and highly cited, and a new version (PHYML 3.0) will be available shortly. Users can download the binaries (Windows, Mac, Linux, and several other OS) or analyse their data via our web server. Papers, documentation, binaries, web server are available from http;//atgc.lirmm.fr/phyml.

References: (1) PhyML - A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Guindon S., Gascuel O. Systematic Biology. 2003 52(5), 696–704. (2) PhyML Online: A web server for fast maximum likelihood-based phylogenetic inference. Guindon S., Le Thiec F., Duroux P., Gascuel O. Nucleic Acid Research 2005 33, 557–559. (3) Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. Hordijk W., Gascuel O. Bioinformatics 2005 21(24), 4338–4347. (4) Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. (5) Anisimova M., Gascuel O. Systematic Biology 2006 55(4), 539–552.

POSTER 34: Functional Genomic and Structural Bioinformatics for Rational Drug Design Against Malaria in a Developing Country, Mali

M. Coulibaly[1], B. Maigret[2], S. Doumbia[1], M. M. Desvigne [2] and O. K. Doumbo[1]

[1]University of Bamako Faculty of Medicine, Pharmacy and Odontostomatology Malaria Research and Training Center, Mali

[2]University of Henry Poincare Nancy, Faculty of Sciences and Techniques, Loria, France

Abstract: Despite major attempts over the past century to control malaria, this infection is still a life-threatening disease affecting half a billion humans in underdeveloped and developing countries. Its global heartland is Africa, with an appalling death toll of 1 to 2 million people every year. Drug and vaccine research in malaria has a high priority. However, identification of suitable enzyme or antigens as candidates against malaria parasites development in human or in its vector mosquito has been on molecular biology mechanism of malaria life cycle. Genomic and post-genomic studies has made the search of new drugs and vaccines easier. Urgent collaboration between endemic countries, and academic research institutions in developed country, i.e. required to address new and inexpensive drug discovery methods. We describe here a partnership development and implementation between the University of Bamako, Mali and the University of Henry Poincare of Nancy I, France.

This collaboration aims to strengthen Malian capacity on computational biology and to create a platform and network between North and South for post genomic data integration and interpretation. Share data and allow research institutions to focus their clinical trials on the most promising potential targets. We aim to perform large scale functional genomic study to understand molecular mechanism that govern the *Plasmodium* life cycle. Genomic and proteomic data provided by functional genomic study can be integrated and interpreted by using computational power full tool and could allow us to identify suitable metabolic pathway targets. Structural bioinformatic tools can then be used to determine and predict protein structures and their molecular docking with drug candidate structures for the specific inhibitor research.

We hope this collaboration will first strengthen our institute by allowing transfert of functional genomic and structural bioinformatics technology. In the long term we hope to develop new innovative drugs or vaccines that can inhibit parasite life cycle development in the host and vector.

POSTER 35: In-silico Screening for Inhibitors of CLPQY Protease of *Plasmodium falciparum*

C. Uzochukwu Ikemefuna[1] and Gupta Dinesh[2]

[1]Department of Pharmaceutical and Medicinal Chemistry,Faculty of Pharmaceutical Sciences, Nnamdi Azikwe University, Awka, Anambra State, Nigeria

[2]International Centre for Genetic Engineering and Biotechnology, New Delhi India

Abstract: The 3D molecular structure of ClpQy will be obtained from www.PDB.org. The physicochemical parameters of ClpQy (such as molecular weight, isoelectric point etc) will be determined using the SwissProt online machine. Ligand structures shall be obtained from the Cambridge Structural Database®, and/or Chemical Abstracts Online®, and/or LOPAC-1280®, and/or TRANSFAC 3.5® etc. Rational molecular modification of known proteases inhibitors shall be done to afford other compounds for computational screening. Phenyl and heterocyclic structures

will be selected using QUEST®. The generated structures will be screened against the ClpQy protein structure with DOCK3.5® in the contact score mode and structures with scores above acceptable value will be selected for further scoring based on steric fit within the active site of the ClpQy protease. The selectivity for the parasite enzyme over the human enzyme will be determined by substracting the contact score for parasite enzyme from the contact score for the human enzyme. Selected compounds will be visually screened for protein-ligand, electrostatic, H-bonding and hydrophobic interactions using MACROMODEL®. Chemical interactions will be assessed within DOCK 3.5® using the force field scoring. Successful compounds will be subjected to biological screening and *in vitro* assays. Lead protease inhibitors specific against *Plasmodium falciparum* ClpQy protease at a reasonable nanomolar range will be identified.

POSTER 36: *In silico* Knock-out Screening of the Metabolic Network of *Plasmodium falciparum* to Yield Potential Drug Targets

Segun Fatumo[1,2,3], Ezekiel Adebiyi[1], and Rainer König[3]
[1]Computer and Information Science Department, Covenant University, Ota, Nigeria
[2]Cologne University bioinformatics Center, University of Cologne, Germany
[3]Bioinformatics and Functional Genome Analysis, Institute of Pharmacy and Molecular Biotechnology, University of Heidelberg, Germany

Abstract: *Plasmodium falciparum* is responsible for about 90% of malaria deaths. Therefore there is a need to discover new drug targets since the parasite has become resistant to existing malaria drugs. We have established a simple and efficient tool which analyses the metabolic pathways and as well identifies essential reactions/enzymes as possible drug targets in the metabolic network of *Plasmodium falciparum*. We investigated the essentiality of a reaction in the metabolic network of *Plasmodium falciparum* by deleting such a reaction *in silico*, 30% of the neighbouring compounds of the investigated reaction were chosen as products. We identified 203 essential reactions, our algorithm yield about 40% consensus choke-points as presented elsewhere. Consequently, we did not take essential reactions/enzymes as drug targets into account which have similarity to any human enzyme. Furthermore, our enzymatic reaction data were grouped into six clusters of co-expression of their corresponding genes during the red blood cell life cycle. We effectively identified 19 possible drug targets to be validated experimentally. Our aim is to combine inhibitors in the same cluster group of these possible drug targets additionally improving malaria therapy.

Key words: Knock-out, Essential, *Plasmodium*, Flux balance analysis

POSTER 37: Integrating Malaria Data from Public Databases Using DGB

Sylvester Lyantagaye, Junaid Gamieldien, Darren Otgaar, Anita Schwegmann, Benjamin Nortier, Shaheem Price and Daniel Jacobson
National Bioinformatics Network, 2nd Floor, Bytes Technology Building, Raapenberg Road, Golf Park, Mowbray, Cape Town, South Africa. sylvester@nbn.ac.za

Abstract: The rapid growth of high-throughput technologies has stimulated interest in methods to model, discover, and validate biological relationships found in genomics and post-genomics generated data. A lot of the malarial molecular data have been collected and organised in the relational database PlasmoDB and other Malaria public databases, which could allows for in-silico analyses. However, a lot still needs to be done about the data. The functions of about two-third of *P. falciparum* genes for instance are not known yet, limiting any in-silico exploration of the malaria biology and of the new drug targets. More importantly, what would now be useful for Malaria research would be methods with which to integrate semantically rich data in ways that "wetlab" experimentation and in-silico discovery are supported and enhanced. To this end we have developed DigraBase (DGB), a distributed graph database, which enables the storing and querying of heterogeneous data in the form of a semantic network. We want to use the DGB to integrate malaria parasite, host, and vector related –omics data and ontologies from public sources to enable large-scale analysis of the data. The integrative model will give more insight to scientists working on anti-malaria vaccine and/ or drugs a better understand the relationships between the genotypes and phenotypes related to the disease.

POSTER 38: DNA Array Based Differentiation and Identification of Leishmania Species, Encountered in Tunisia and the Mediterranean Region, Using Targets Identified in silico

Y. Saadi, A.S Chakroun, M. Ben Fadhel, R. Lahmadi, A. Fathallah-Mili, S. Ben Abderrazak and I. Guizani

Laboratoire d'Epidémiologie et d'Ecologie Parasitaire: Applied biotechnologies and genomics to parasitic diseases research programme, Institut Pasteur de Tunis, Tunisia

Abstract: Leishmaniases constitute major public health problems in many countries over 4 continents, and are particularly endemic in Africa and the Mediterranean region. In Tunisia, the epidemiological situation is complex with proven existence of at least three Leishmania species: *L. infantum, L. major* and *L. tropica*, which are responsible for cutaneous leishmaniasis. Visceral leishmaniasis is caused by *L. infantum*. Simple, reliable, sensitive and rapid diagnostic tools still remain a research priority. The aim of this study is to develop diagnostic PCR based assays for the Leishmania species identification and discrimination using targets identified upon the comparative analysis of *L. major* and *L. infantum* genomes. We have exploited the genome project databases, available in the public domains: the *L. major* genome is completed while the *L. infantum* is still ongoing. We developed perl scripts to Blast sequences from *L. infantum* against *L. major* and inversely. The sequences that were identified this way were then blasted on other databases like Genebank and EMBL. Following this selection process, 2 sequences were identified for *L. major* and 4 for *L. infantum*. The sequences were analysed in silico, they were assigned to 6 different Leishmania chromosomes (Chr: 8, 21, 23, 30, 32 and 33), and they were annotated using the "Leishmania bioinformatic's toolbox'.

Twenty-four primers were designed using the software "primer 3". Different primer pairs were considered for each sequence. Their specificity was checked in silico and was assessed experimentally. A total of 18 primer pairs were tested on 3 DNAs representative of the species *L. major, L. infantum* and *L. tropica*. Interestingly, specificity of 4 sequences was confirmed towards *L. infantum* (2) or *L. major* (2). For the remaining 2, results are concordant with the latest genome releases. *L. tropica* DNA was amplified with all primer pairs tested with the exception of two primer pairs corresponding to one *L. infantum* marker. Three physically independent pairs were selected for the development of a multiplex PCR assay that allows the generation of 3 consistent and well differentiating amplification profiles, specific for the 3 Leishmania species encountered in Tunisia, *L. major, L. tropica* and *L. infantum*. A DNA array was further developed which allowed a drastic increase in sensitivity of detection and a significant reduction in results delivery time. Work is in progress to evaluate the potential of these tools for molecular diagnosis of Leishmaniasis in Tunisia.

This work received financial support from MRSTDC-Tunisia (contrat programme 2004–2008), from the RAB & GH programme of WHO/EMRO-COMSTECH and from AIEA Technical cooperation programme (TUN/06/12).

LIST OF PARTICIPANTS

Abraham Adane
Ethiopian Institute of Agricultural Research,
National Plant Protection Research Center
adaneabraham@yahoo.com

Adebiyi Ezekiel
Department of Computer and Information Sciences,
Covenant University, PMB 1023, Ota, Nigeria
eadebiyi@sdsc.edu

Afshar Mohammad
Ariana Pharma, Institut Pasteur-Biotop
m.afshar@arianapharma.com

Aritua Valentine
National Agricultural Biotechnology Center,
Kawanda Agricultural Research Institute,
P. O. Box 7065, Kampala, Uganda
arituavalentine@yahoo.com or
valentine@kari.go.ug

Bairoch Amos
Swiss Institute of Bioinformatics
ab@isb-sib.ch

Benkahla Alia
Group of Bioinformatics and Modelling,
Laboratory of Immunology, Vaccinology, and Molecular Genetics,
Institute Pasteur of Tunis
Ad: 13, place Pasteur BP 74 1002, Tunis
Alia.Benkahla@pasteur.rns.tn

Berriman Matthew
Sanger, UK
mb4@sanger.ac.uk

Biscah Munyaka
Kenya Medical Research Institute
P. O. Box 20752-00202 Nairobi, Kenya
bsyombua@yahoo.com

Bréhélin Laurent
CNRS-France
brehelin@lirmm.fr

Breton Vincent
CNRS-France
breton@clermont.in2p3.fr

Cho-Ngwa Fidelis
University of Buea, Cameroon
chongwa_ub@yahoo.co.uk

Coulibaly Modibo
Malaria Research and Training Centre,
University of Bamako,
B.P.1805, Bamako, Mali
cmodibo@usa.com

Daudi Jjingo
Makerere University, Uganda
jjingois@yahoo.co.uk

de Oliveira Tulio
SANBI, South Africa
tulio@sanbi.ac.za

de Villiers Etienne
International Livestock Reasearch Institute (ILRI)
P. O. Box 30709-00100, Nairobi, Kenya.
evilliers@cgiar.org

Djikeng Appolinaire
TIGR, US
adjikeng@tigr.org

Domelevo-Entfellner Jean-Baka
CNRS, France
domelevo@gmail.com

Doumbia Seydou
Malaria Research and Training Centre,
University of Bamako,
B.P.1805, Bamako, Mali
sdoumbi@MRTCBKO.org

Elikana Anami Sylvester
Gent University and Kenyatta University
syanakuti@gmail.com

Fatogoma Sorho
Africa Rice Center/ Université de Cocody Abidjan
fsorho@gmail.com

Fatumo Segun
Covenant University, Ota, Nigeria
f.segun@dkfz-heidelberg.de

Gascuel Olivier
LIRMM, CNRS, 161 rue Ada,
34392, Montpellier Cedex 5, France
gascuel@lirmm.fr

Gaunt Michael
SANBI, South Africa
mgaunt@sanbi.ac.za

Ghemtio Leo
INRIA-Lorraine Engineer, Algorille team,
INRIA LORRAINE,
615 rue du Jardin Botanique,
54600 Villiers les Nancy
leo.ghemtio@loria.fr

Ghouila Amel
Faculté des Sciences, Tunis
amel.ghouila@gmail.com

Gitau Catherine W.
icipe, P. O. Box 30772-00100
Nairobi, Kenya
cgitau@icipe.org

Githui Elijah Kem
National Museums of Kenya
kegithui@yahoo.com

Gonedelé Bi Sery
Université d'Abidjan
gonedelebi@yahoo.fr

Guizani Ikram
Institut Pasteur, Tunis
ikram.guizani@pasteur.rns.tn

Hide Winston
SANBI, South Africa
winhide@sanbi.ac.za

Isokpehi Raphael
Jackson State University, Jackson Mississippi USA
raphael.isokpehi@jsums.edu

Jaufeerally-Fakim Yasmina
Biotechnology Unit, University of Mauritius
yasmina@uom.ac.mu

Johnson O. Ouma
TRC-KARI,
P. O. Box 362, Kikuyu, Kenya
joouma@gmail.com

Joseph Paschal Mugasa
Ifakara Health Research and Dev Centre, Tanzania
jpmugasa@ihrdc.or.tz

Karau Muriira Geoffrey
Mother Offspring Malaria Study (MOMS)
Morogoro Regional Hospital
P. O. Box 2120, Morogoro, Tanzania
karaugm2@gmail.com

V. Kasam
LPC Clermont-Ferrand, Campus des Cézeaux
63177 Aubière Cedex, France
breton@clermont.in2p3.fr

Laouini Dhafer
Institut Pasteur, Tunis
dhafer_l@yahoo.ca

Lefort Vincent
CNRS, France
vincent.lefort@lirmm.fr

Lehvaslaiho Heikki
SANBI, South Africa
heikki@sanbi.ac.za

Lyantagaye Sylvester
National Bioinformatics Network, South Africa
sylvester@nbn.ac.za

Marechal Eric
CEA Grenoble
eric.marechal@cea.fr

Masiga Dan
icipe, P. O. Box 30772-00100
Nairobi, Kenya
dmasiga@icipe.org

Mathurin Koffi
CIRDES Bobo-Dioulasso/Burkina fasso
math_kof@yahoo.fr

Mégy Karyn
European Bioionformatic Institute (EMBL outstation, Cambridge, UK)
kmegy@ebi.ac.uk

Morett Enrique
Universidad Nacional Autónoma de México
emorett@ibt.unam.mx

Mtakai Vald Ngara
International Institute of Tropical Agriculture (IITA), Nairobi
ngaravald@yahoo.com

Mulder Nicky
University of Cape Town, South Africa
nmulder@science.uct.ac.za

Nelson Ndegwa
IITA, P. O. Box 30709-00100,
Nairobi, Kenya
dj.kim@cgiar.org

Nindo Nzabanyi Fredrick
Kenyatta Hospital P.I.P.S PROJECT, Kenya
fnindo@yahoo.com

Niven Abdel Wahab Salih
Institute of endemic diseases,
University of Khartoum, Sudan
neveensalih@yahoo.com

Nworu Chukwuemeka Sylvester
University of Nigeria, Nsukka, Enugu State, Nigeria
csnworu@yahoo.com

Nyakeriga Emmanuel
Egerton University, Kenya
kenyanbiomedic@yahoo.com

Nyambega Benson
icipe, P. O. Box 30772-00100
Nairobi, Kenya
bnyambega@icipe.org

Nzunza Rosemary
US Army Medical research Unit, Kenya/KEMRI
rnzunza@wrp-nbo.org

O. Owolabi Marion
Covenant University, Nigeria
olumarion@hotmail.com

Okanya Patrick W. S.
Department of Biochemistry,
University of Nairobi,
P. O Box 30197- 00100
Nairobi, Kenya
pokanya@yahoo.co.uk

Ommeh Sheila Cecily
Kenyatta University/ILRI, Kenya
s.ommeh@cgiar.org

Onyango Kasuku Philbert
Egerton University, Kenya
onphilbertus@yahoo.com

Onyango Maria Gorreti
Department of Biochemistry
University of Nairobi, Kenya
sharon_golder2004@yahoo.com

Onyango Odhiambo Christopher
Egerton University, Kenya
nyakwaracholla2006@yahoo.com

Osamor Victor
CIS/CST, Covenant University, OTA,
Ogun State, Nigeria
vcosamor@yahoo.com

Otieno Ong'amo George
IRD/*icipe*
P. O. Box 30772-00100
Nairobi, Kenya
gongamo@icipe.org

Oumou Maïga
Malaria Research and Training Center, Bamako, Mali
osame8@gmail.com

Ouwe Missi Oukem Odile
Centre International de Référence Chantal Biya
ooukem@yahoo.fr

Owino Vincent
icipe, P. O. Box 30772-00100
Nairobi, Kenya
vowino@icipe.org

Shah Trushar
BecA, Kenya
tm.shah@africabiosciences.org

Smandi Sondos
Institut Pasteur, Tunis
smandisondos@yahoo.fr

Souiai Ouissem
Institut Pasteur, Tunis & IBDML, France
souiai@gmail.com, souiai@ibdm.univ-mrs.fr

Tibayrenc Michel
IRD, France
Michel.Tibayrenc@ird.fr

Tonessia Dolou Charlotte
CERAAS Senegal
dolouc@yahoo.fr

Uwayo Jean Claude
National University of Rwanda
nellyhotboy2000@yahoo.fr

Uzochukwu Ikemefuna Chijioke
Nnamdi Azikiwe University, Awka,
icuzochukwu@yahoo.com

Wambua Lillian
International Livestock Research Institute (ILRI)
P. O. Box 30709-00100, Nairobi, Kenya
l.wambua@cgiar.org

Wanjiru Kariuki Lucy
University of Nairobi, Kenya
lkariuki@uonbi.ac.ke or
madamkariuki@yahoo.com