

Advancing Research and Training in Genomics to address Biosecurity Threats caused by Emerging Infectious Diseases in Africa (ART2BioSec)

Project Framework

Milestone 1: Developing the next generation of leaders in research and technical support				
Objective 1.1: Develop curriculum for system administrators				
Purpose: Training material is required so that we can system administrators annually at co-applicant sites and in general for sustainability via co-applicant institutions				
Responsibility: Training co-ordinator with system administrators at each of the collaborating centers				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
An online training resource for system administrators	Develop training material for the system administrators	Online material that will be used for subsequent training	Online teaching and learning platform Method of evaluation Sufficient scope of training materials	Sufficient internet bandwidth Practical evaluation protocol Standardised and broad-based training curriculum
6 system administrators trained	Train system administrators through curriculum online and site visits for 1-week to SANBI-UWC	Engineers able to configure HPC environment and install modules on their servers	Recruitment Retention	Capacity of interested individuals Competitive work environment for

		and maintain a core set of bioinformatics tools	Training material longevity	retention of trained administrators Training material updated as technology changes
Competencies from systems administration training aligned with HPC-Certification Forum (https://www.hpc-certification.org/) skills tree	Extract competencies from training resources' learning objectives and ensure they are aligned with HPC-CF or are represented as additions to the HPC-CF skills tree (admin section)	Competencies represented as learning objectives in training material align with HPC-CF skills tree.		
Accreditation of system administrators 12 months after training	System administration lead at SANBI-UWC to conduct virtual assessment of trainees.	Trainee passes accreditation examination.	Standard method of evaluation developed Availability to travel	According to accredited standard or framework Can accreditation be adapted should restrictions on physical travel occur?
Objective 1.2: Develop pathogen bioinformatics curriculum				
Purpose: Training material is required so that we can train bioinformatics staff and and in general for sustainability via co-applicant institutions adopting these modules more formally				
Responsibility: Training co-ordinator with system administrators at each of the collaborating centers				

Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
An online training resource for bioinformatics researchers	Develop training material for the bioinformatics researchers	Online material that will be used for subsequent training	Needs assessment Access to training resources over and above those developed Continuous development in line with need	Targeted module development and needs assessment Infrastructure and capacity. Communication of additional resource calls. Interactive engagement as new material emerges in line with specific need
Formalize bioinformatics training at co-applicant institutions	Co-applicant lead to engage with institutional management (academic planning)	Adoption of the bioinformatics modules as part of a MSc programme	Institutional and National Higher Education approval	Stakeholder engagement and buy-in Course accreditation in line with institutional policy

Objective 1.3: Develop a network for system administrators across the consortium

Purpose: The system administrators will support a computing environment for all collaborating centers to enable genomic surveillance

Responsibility: Training co-ordinator, and personnel at collaborating centers, coordinates the community of practice network (rsse-africa.sanbi.ac.za) activities (website, webinars).

Deliverable	Activity	Indicator	Dependencies	Constraints/
--------------------	-----------------	------------------	---------------------	---------------------

				Assumptions
A network of system administrators	coordinates the community of practice network activities (webinars)	Establish a community of practice to drive skills development	Community driven management Communication, workshops, webinars Increase interest and leverage external networks	Proactive management Infrastructure availability and coordinated scheduling Ensure interest is increased by creative approaches Means to access global networks
Objective 1.4: Integrate bioinformatics trainees into existing bioinformatics networks				
Purpose: The current African-based bioinformatics networks (African Society for Bioinformatics and Computational Biology, East and West Africa bioinformatics hubs, South African Society for Bioinformatics, H3ABionet) provide networking opportunities that will benefit our trainees career development.				
Responsibility: Training co-ordinator and personnel at collaborating centers will communicate activities of the above-mentioned networks				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
Consortium bioinformatics trainees are networked	Promote the work of the range of bioinformatics networks	Consortium trainees register with any of the bioinformatics networks	Annual subscription fees are not a limitation	Funding for subscriptions are not for personal costs Ability and availability to travel. If events are

		Trainees develop and drive interest groups	Workshop and conference opportunities Bioinformatics trainees actively train others	hosted online, sufficient internet access to attend Bioinformatics trainees are confident and skilled to become teaching assistants
Objective 1.5: Provide LIMS training to laboratory staff				
Purpose: Train laboratory staff to manage biospecimens and good laboratory practice				
Responsibility: LIMS lead and her team				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
12 Staff trained on Baobab LIMS	Training provided virtually and face-to-face	Ability to use the LIMS	Requires LIMS installed in the laboratory Previous knowledge of laboratory workflows Train the trainer	Suitable infrastructure and administrative staff for maintenance Well-defined laboratory workflows and SOPs in place Ability to adapt software to workflows and vice versa Trained staff return to train other staff and

				create an assessment protocol Helpdesk management for support
Milestone 2: Genomic Surveillance in Africa				
Objective 2.1: Install a fit-for-purpose Computing infrastructure				
Purpose: To establish a fit-for-purpose computing environment for collaborating partners.				
Responsibility: System engineer based at SANBI-UWC will oversee and contribute to the hardware installation and configuration. Require additional system administrators at each of the African collaborating sites (SANBI-UWC, AAU, USL, TUK)				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
Minimal computer hardware	Purchase of a minimal computing environment	Procurement of hardware	Procurement of hardware	1. Suitable vendors present and able to supply server within budget. Suitable physical infrastructure for server present (space, cooling, power, wide area network (WAN)).
Working sever at each center	Installation of computing hardware and software	Working server	Infrastructure such as a power, electricity, backup	

Updated software and general systems management	Maintenance of hardware and software systems	Maintain software updates and general systems management	Staff skills	
Objective 2.2: Edge processor development				
Purpose: There is a requirement for users to access bioinformatics tools using easy-to-use interfaces. Our work on integrating GALAXY as part of our COMBAT-TB Platform demonstrates that a user-friendly graphical interface that allows a biomedical researcher to point-and-click bioinformatics pipelines is effective. This system can be deployed across the continent with limited resources.				
Responsibility: system administrators at collaborating centers (SANBI-UWC, AAU, USL, TUK, LSB) main a toolkit of analytics modules that can be shared.				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
Functioning Galaxy server	Host a user-friendly userGalaxy.africa server located at at each site (AAU, USL, TUK, SANBI-UWC)	Locally maintained useGalaxy.africa at each site	Objective2.1 Supporting institutional infrastructure including electricity, backup power	
A minimal galaxy server (edge processor) with relevant pathogen analysis tools.	Installation of a workstation with appropriate software at Lagos State biobank	A workstation that Lagos state biobank can connect to the AAU server	Successful installation of software	
Objective 2.3: Analysis of genetic diversity of SARS-COV-2 isolates				
Purpose: phylogenetic profiling of viral isolates to transmission chains and understand disease spread				

Responsibility: Researchers and students and each research collaborating center				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
500 SARS-COV-2 genetic material	Collection of 500 nasopharyngeal samples. Isolation of virus material	Non-contaminated viral RNA	Available reagents and standardized RNA extraction protocols	
500 sequenced viral genomes for each of 3 sites (Sierra Leone, Nigeria, Kenya)	Sequencing using an illumina machine	Depth of coverage		
Assembled viral genomes submitted to the AGA	QC, base calling and assembly	Assembled genomes		
Transmission chains	Phylogenetic analysis			
Objective 2.4: Predict the future distribution of viral hemorrhagic fevers using genetic and ecological data to				
Purpose: Understand seasonal and climatic cycles that influence the dynamics and abundance of reservoir hosts (ecology)				
Responsibility: Researchers and students and each research collaborating center				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
10000 ticks at each of three sites	capture and sampling of wild and domesticated animals with traps. Use non-invasive technique (insect-based animal surveillance)	Captured animals	Use of an innovative non-invasive technique for surveillance in the field.	

10,000 ticks at each of three sites	Collection of ticks	The number of collected ticks per month per location	Access to various sites for sample collection	Assumption: In country travel to various remote locations for ticks collection will proceed as COVID-19 restrictions are lifted.
Tick-borne viruses identified	RT-PCR assay of material found in the ticks and sanger sequencing	The number of samples analysed by RT-PCR and sequenced	The number of identified virus will depend on the prevalence of each virus Reproducible extraction techniques for genetic material. Enrichment of viral material.	Constraints: A prevalence as low as 0.1% would allow the identification of 10 viral sequence Assumption: International travel restriction will be resumed allowing in-person training at African sites. It's possible that our insect-based surveillance is unable identify novel animal species susceptible to Ebola, or Lassa. This is mitigated by sampling from immune privilege sites where virus persists.
phylogenetic and environmental correlates of reservoir	1*Phylogenetic comparison of obtained	Generation of a phylogenetic tree for each pathogen	Identification and sequencing of	Assumption: The sampling magnitude is sufficient to capture

host seroprevalance across Africa	and existing pathogen sequences 2*Collection of environmental parameters at the tick collection sits		circulating emerging pathogens	pathogens of interest even those with low frequencies
Understand seasonal and climatic cycles that influence the dynamics and abundance of reservoir hosts (ecology)	Combining environmental factors with viral prevalence data	Obtaining from existing data climatic and environmental data for each tick collection sites such as temperature, rainfall, elevations,...	Various international database recording climatic variables are freely available. Additional information should be collected during tick sampling	Assumption: The existing database are accurate
Predicted future global changes will influence the distribution and probability of spillover and transmission of VHF	Modeling the movement of the pathogens of interest based on environmental parameters	Development of one mathematic model for the spillover of each studied pathogen	Sufficient amount of segregated data will be collected to improve the accuracy of the model	Constraints: The more segregated the obtained data, the more accurate the final model will be.

Objective 2.5: Analysis of genetic diversity of Biobanked Ebola biospecimens

Purpose: Demonstrate the utility of the a newly established nanopore sequencing platform for analysing emerging pathogens

Responsibility: Researchers and students and USL

Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions

50 Sequenced Ebola samples from Sierra Leone Biobank	Isolation of virus material from biobanked samples. Sequencing of samples using nanopore platform	Non-contaminated viral RNA Assembled genomes	Available reagents and standardized RNA extraction protocols	Assumption: installation of nanopore platform
--	--	---	--	---

Milestone 3: Development of an African Genome Archive (AGA)

Objective 3.1: Develop a datastore

Purpose: We need a data store that facilitates data mining that is scalable to account for the range of genomics and epidemiological data that will be collected

Responsibility: Research engineer, bioinformatics scientist, application specialist, computer science student

Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
Meta data schema for pathogen data	implement an internationally acceptable meta data standard	Ability to import meta data for pathogens Compatibility with other archive submission protocols	Requires an assessment of the current global standards used for pathogen data. Community input and evaluation	Assumption that this will feed into the genomic surveillance platform Community adoption Bioinformatics Knowledge for support purposes
Openstack Infrastructure as a service	Installation of openstack to manage virtual instances and network connectivity	Successful installation of OpenStack as shown by server log reports.	Requires systems engineers working on the surveillance platform to provide computing infrastructure	Skilled individuals to maintain service

				Bioinformatics Knowledge for support purposes
Meta-data aware storage	Installation of iRODS and NoSQL database	Ability to import pathogen data and epi data using APIs	Infrastructure International Archive Standards	Skilled individuals to maintain service
Server-side architecture	Installation and configuration of FLASK.	A working web front end for the AGA	Infrastructure	Skilled individuals to maintain service
Front-end web functions	Configure/customize FLASK templates for the AGA	Utility of the front end to upload data	Infrastructure Software Maintenance	Skilled individuals to maintain service
A catalogue for the archive	Develop a mechanism to track data holdings so that we know where to find different data	A web-based interface that shows our data holdings	A function website (FLASK)	Skilled individuals to maintain service Bioinformatics Knowledge for support purposes
A centralized pathogen dashboard	Develop a graphical interface that summarizes the info in the datastore	Visualisation tool that reduces dimensionality of the data and summarizes what we have in real-time	Network Access	Skilled individuals to maintain service Bioinformatics Knowledge for support purposes

Authentication for user access to the federated datastore	Build authorization and authentication protocols for the datastore	A user can access the federated datastore using sign-in credentials	Infrastructure Network Access	Skilled individuals to maintain service Network Access
---	--	---	----------------------------------	---

Objective 3.2: Grow a network of data curators for QC and analysis

Purpose: To screen data for QC that is to be submitted to the AGA

Responsibility: 6 bioinformatics graduates at each of five research collaborating sites

Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
One science graduate as a curator at each of 5 research centers	Recruitment of one science graduate as a curator at each of 5 research centers	Recruitment		
a network of curators in Africa	Create a network of curators in Africa that can connect with the Biocuration community	Active curators who participate in international biocuration activities		

Objective 3.3: Establish a helpdesk

Purpose: Provide an electronic tracker for all queries to, and about, the AGA

Responsibility: Sysadmin at SANBI-UWC to install a help desk. Someone to monitor helpdesk and assign tasks to individuals

Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
Email submissions to the help desk	Submit queries	Record of all queries		

Monthly helpdesk reports	Generate monthly helpdesk reports	reports submitted to the lims team lead		
Milestone 4: Effective Project Management				
Objective 4.1: Develop a project management framework for the Consortium				
Purpose: Ensure accountability and good governance within the consortium				
Responsibility: Consortium management team and project leads each of five research collaborating sites				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
Regular documented meetings, oversight and timely reporting	Minutes for all meetings	Records of all meetings		Assume meetings take place or are rescheduled
	Collaborating centers generating quarterly reports	Narrative and Financial reports on time for SAB meeting	Lead institution require reports from partners Availability of parallel plans such as data and risk management plans.	Assume that reports are submitted on time by all collaborating centers
Objective 4.2: Develop a governance framework for the AGA				
Purpose: Ensure accountability and good governance at the AGA				
Responsibility: Consortium management team and project leads each of five research collaborating sites				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
A framework document that outlines the scope of the AGA	Develop a ToR for the AGA that outlines the scope of AGA	ToR document		
A code of conduct policy that outlines expectations for staff	A code of conduct policy for AGA	Code of conduct policy document	Policy is linked to the consortium policy	

and an SOP for conflict resolution				
A plan for how AGA will manage data including but not limited to documents, sequencing data	Describe a data management plan (DMP) for all data at the AGA	Data management plan policy	Linked to DMP plans across the consortium	
Objective 4.3: Reliable Monitoring and Evaluation				
Purpose: Monitor progress of all work packages				
Responsibility: M&E team at GET. design templates in collaboration with administrators at partner sites.				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
quarterly updates from each project	Produce quarterly reports for each partner site	Submit quarterly reports	Require partners to submit report templates	Assume each partner site is on time
Annual financial reporting to the lead institution	Track financial reporting from partner institutions to the lead institution	Successful financial reports delivered on time	Invoicing and payments	Institutional Procurement processes are not be efficient
Skill transfer to students	Track student participation and completion	Number of students trained		
Skill transfer to support staff	Track staff participation and completion	Number of staff trained (engineers, sysadmin, scientists)		
Milestone 5: Biospecimen Management				

Objective 5.1: Customized Baobab LIMS for an infectious disease laboratory

Purpose: Pathogen Biospecimen associated metadata are an essential part of an African genomics archive and laboratories need to be provided with the tools for pathogen samples management and storage of biospecimens

Responsibility: Dr Dominique Anderson: oversee the product development process. Software developer will be based at each of the five research collaborating centers to customize BaobabLIMS.

Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
Customized Baobab LIMS for managing pathogen biospecimens	Customization of BaobabLIMS – instrument integration, lab workflows, temperature monitoring, virtual catalogue development etc.	<p>A dockerized installation module</p> <p>Metadata needs assessment</p> <p>Validation of function, mechanism to report errors/bugs by end-users</p> <p>Development of documentation, manuals and videos.</p>	<p>Infrastructure on site</p> <p>Adequate training and communication with laboratory staff</p> <p>User uptake</p> <p>Software installation and management</p> <p>Dedicated testing and validation of functionality</p> <p>Collaboration between laboratory and developers</p> <p>Metadata needs assessment</p>	<p>Evaluation of computational capacity based on need and the availability of capacity</p> <p>Developer and lab personnel communication platforms</p> <p>Scheduled feedback and hackathons</p> <p>Dedicated repository, code validation protocols, on site testing by developers and end users, bug fixes.</p> <p>Assessment of metadata standards applicable to general and specific need</p>

			Developer skills Access to third-party resources Dedicated documentation with version control. Visual manuals for end-users	Recruitment of skilled personnel, retention and training if required Resource availability
Milestone 6: Communications strategy				
Objective 6.1: Develop a user-friendly communication and dissemination platform				
Purpose: sharing and distributing information at multiple stakeholder levels. Ensure communication is up-to-date and report milestones				
Responsibility: Communications team based across all partner centers – community engagement researcher at each of the 5 sites, videographer at 3 sites, 1 web site developer, web content developers at each site.				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
A user-friendly website	Develop a website	Working website	Responses from content contributors, sysadmin support	
Social media presence	Post announcements on twitter, Facebook and website	weekly announcements on twitter, Facebook and website	Regular inputs Consortium-wide involvement	
YouTube videos of every event organized	Generate video material of consortium events	Recording highlights of events/symposia/training sessions	Budget for videographers to attend consortium meetings	

			Skilled staff at each co-applicant institution	
Objective 6.2: Community Engagement to assess perceptions				
Purpose: Gain a deeper understanding of lay-persons' perceptions of data use and establish a dialogue between researchers and lay-communities to enable researchers to address community concerns (possibly through policy changes) and build trust in communities				
Responsibility: A researcher based across all partner centers – community engagement researcher at each of the 4 sites,				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
Public perceptions of data use	Meeting with communities and relevant stakeholders to assess public opinion on data use during a pandemic	Survey of participants	Community participation	Language barrier and language equivalence / communities want to be involved
Objective 6.3: Develop an audio book to communicate biosecurity concepts				
Purpose: Create awareness and understanding of biosecurity, thereby building community trust through transparency and increase research participation. Offering a fun and interactive educational tool that can increase people's interest in science.				
Responsibility: A researcher based across all partner centers – community engagement researcher at each of the 4 sites; illustrator, audio				
Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
Speaking book on biosecurity	Develop a speaking book- illustrations, narrative	Working speaking book	Community participation	Communities want to be involved

Speaking book available in 2 local languages in West and East Africa	Translate the narrative into 4 languages;	Audio translations of the speaking book	Knowledgeable, expert translators	Language equivalence difficulties due to scientific terms that do not exist in local languages
--	---	---	-----------------------------------	--

Milestone 7: Biosecurity Framework in Africa

Objective 7.1: Improve the health-security-legal interface to ensure adoption of a genomics and biosecurity framework

Purpose: To ensure adoption of a genomics and biosecurity framework that promotes peaceful custodianship and research on dangerous category “A” infectious pathogens in Africa

Responsibility: Team led by the GET consortium with all partner centers

Deliverable	Activity	Indicator	Dependencies	Constraints/ Assumptions
Genomics and biosecurity policy and laws established in sub Saharan Africa countries	Conduct an analysis of existing laws, regulations and policy for genomics, biosecurity and biosafety	Analysis report	government priorities and interest in participating	Lack of interest by government
	Stakeholders meeting			Government bureaucratic processes Security challenges in some African countries Assume government shows interest in the project