Genomic Insights into Disease Resilience of East African Dairy Cows

Background

East Africa's dairy sector is a cornerstone of regional food security and economic stability. However, it faces significant challenges due to endemic diseases that can devastate livestock populations. Enhancing disease resilience in dairy cows is crucial for sustainable agricultural production and the well-being of local communities.

Objective

The project's primary goal is to investigate the genetic underpinnings of disease resilience in indigenous and locally adapted dairy cows in East Africa. By analyzing DNA sequences, genotypes, and phenotypes, we aim to uncover the genetic variations that confer resistance to local pathogens and stressors.

Methods

- DNA Sequencing: We will collect and sequence the genomes of various indigenous and locally adapted dairy cattle populations to identify genetic variants.
- Genotype Analysis: Utilizing advanced bioinformatics tools, we will analyze the genotypic data to pinpoint alleles associated with disease resistance among these cattle populations.
- Phenotype Correlation: Through field studies, we will correlate the identified genotypes with observed phenotypic traits, particularly those related to immune response and disease resilience.

Expected Outcomes

- A comprehensive genomic database of East African dairy cows.
- Identification of key genetic markers for disease resilience.
- Development of a breeding program to enhance these traits in local cattle populations.

Impact

The findings from this project have the potential to revolutionize dairy farming in East Africa by enabling farmers to breed cows with enhanced disease resilience. This will lead to increased productivity, reduced production costs such veterinary services and medicines, and improved livelihoods for farmers.

Conclusion

By leveraging cutting-edge genomic technologies, this project represents a significant step towards a more resilient, profitable and sustainable dairy industry in East Africa. The integration of local knowledge and scientific innovation will empower communities and contribute to global food security.

Data Management Plan (DMP)

1. Data Description

- **Data Types**: The project will generate raw and processed DNA sequence data, genotype arrays, and phenotypic measurements.
- **Data Formats**: Sequencing data will be in FASTQ and BAM formats, genotypes in PLINK binary format, and phenotypes in structured CSV files.

2. Data Collection and Documentation

- **Collection Methods**: DNA sequencing will be performed using next-generation sequencing. Genotyping will be done using SNP chips, and phenotypic data will be collected through veterinary assessments and laboratory tests.
- **Documentation**: All data will be accompanied by metadata following the MIAME and MINSEQE standards to ensure clarity and reproducibility.

3. Ethics and Legal Compliance

- **Consent**: Informed consent will be obtained from cattle owners for the collection of samples and data.
- **Privacy**: Personal data will be anonymized to protect the identity of participants involved in the study.

4. Data Storage and Security

- **Storage Solutions**: Data will be stored on secure, encrypted servers with backup systems in place.
- **Access Control**: Access to sensitive data will be restricted to authorized personnel only.

5. Data Sharing and Public Access

- **Sharing Policy**: Data will be shared with the broader research community through established databases like NCBI's dbSNP and dbGaP, following a data embargo period.
- **Public Access**: Summarized data, code, and findings will be made available to the public through open-access publications, public repositories like GitHub and presentations.

6. Roles and Responsibilities

- **Data Manager**: A designated data manager will oversee the implementation of the DMP.
- **Research Team**: Researchers will be responsible for data collection, quality control, and preliminary analysis.

7. DMP Budget

- **Costs and Funding**: The budget includes costs for data storage, backup services, and personnel for data management tasks. Funding sources include grants from research councils and contributions from collaborating institutions.
- **Data Storage:** Estimated at $2,000 per terabyte per year, with an expected need of 10 terabytes over the course of the project: $20,000.
- **Data Management Personnel:** A data manager's salary is estimated at $60,000 per year.
- **Backup Services:** Off-site backup services are estimated at $1,000 per year.
- **Data Analysis Software:** Licensing for bioinformatics software is estimated at $5,000 per year.
- **Genotyping and Sequencing:** The cost of genotyping per sample is approximately $50, and whole-genome sequencing is around $1,000 per sample. With an estimated 500 samples, the total cost for genotyping and sequencing would be $525,000.
- **Publication costs:** $5,500 is requested annually for publication fees of papers in an open access journal.
- **Ethical and Legal Compliance:** Costs for legal counsel and ethical review are estimated at $3,000.
- **Training:** Training for staff in data management practices is estimated at $2,000.
- **Total Estimated Budget:** $621,500 over the project duration.

8. Data Preservation and Archiving

- **Long-term Preservation**: Data will be archived in a sustainable format in institutional repositories with a minimum retention period of 10 years post-project.
- **Data Conversion**: Procedures will be in place to convert data to new formats if current ones become obsolete.

9. Data Quality Assurance

- Regular audits and quality checks will be conducted to ensure data integrity and quality. Data validation protocols will be established to maintain the accuracy of the data sets.

10. Data Ownership and Intellectual Property

- All data generated from this project shall remain as an intellectual property of the contributing researchers and institutions. Clear guidelines and agreements shall be established to protect the data's integrity and ensure equitable sharing of information and benefits arising from its use.