PROJECT PROPOSAL COVER SHEET

Project No. _____ *(For Treaty use. Do not write anything here)*

Project Title: Biomarker Development Toolkit for Improving Rice Production and Resistance to Climate Change in Indonesia

Project duration: 12 months

Target crops: Rice

Targeted developing country/ies   Indonesia

Other Contracting Party/ies involved   None

Project geographic extension (km²)   N/A

Total requested funding   US$ 150,000

Total co-funding   US$ 80,000

Please select the type of project you are applying for:

☐ Single-country Immediate Action Project (Window 2)

☐ Multi-country Immediate Action Programme (Window 2)

☒ Single-country Co-development and Transfer of Technology project (Window 3)

☐ Multi-country Co-development and Transfer of Technology project (Window 3)

Applicant

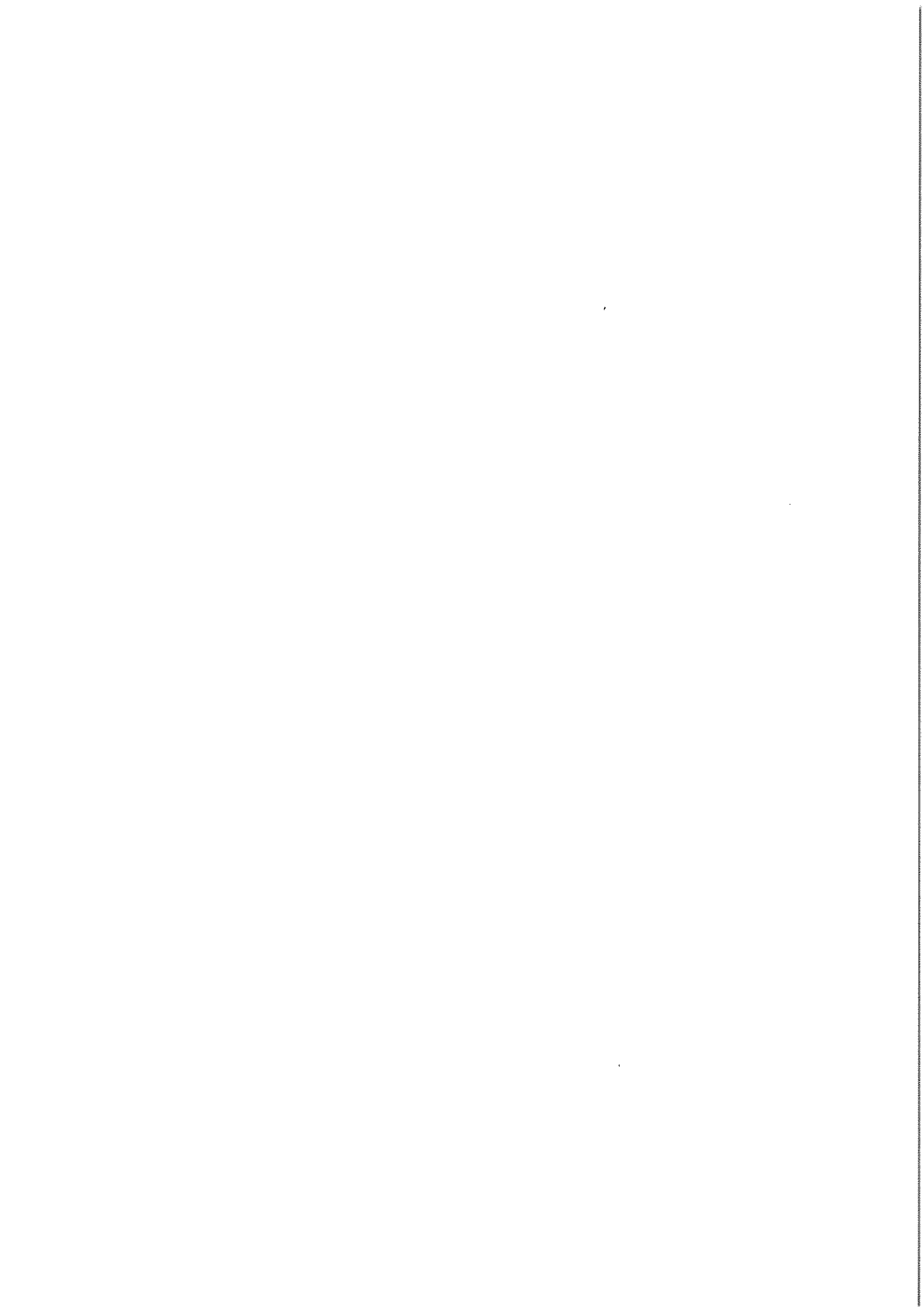Name of Organization: Bina Nusantara University

Type of organization: Higher Education Institution

Project Contact: Dr. Bens Pardamean
                   Director of Bioinformatics & Data Science Research Center

E-mail address: bpardamean@binus.edu

Telephone: +62 21 534 5830 ext. 1700

Fax: +62 21 530 0244

SECTION A: EXECUTIVE SUMMARY

1. Executive summary

Climate change is negatively impacting rice plants through its effect on environmental factors that determine crop yield. Due to the widespread consumption of rice by the world's populace, a threat to rice crops translates into food insecurity. Therefore, there is a dire need to revolutionize the way existing rice plant varieties are improved and new varieties are developed. Competent rice breeds should be able to survive in and adapt to climate change challenges such as prolonged droughts, extreme wet conditions, and water source depletion. Rice farmers' practical field experience can provide agricultural researchers and other relevant parties with valuable information on the current conditions of rice farming. Conversely, researchers are the main providers of efficient ways by which farmers obtain new and/or improved rice varieties. In an effort to face the impending environmental problems, Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development (ICABIOGRAD) has developed a genome-wide association study with the purpose of identifying genetic variants that are linked to yield, yield components, and maturity traits of rice, and the impact of climatic factors (Gene-Environment). With the state-of-the-art statistical methodology developed by Bina Nusantara University (Jakarta, Indonesia), ICABIOGRAD's database can be efficiently and accurately analyzed to facilitate novel productions and enhancements on rice varieties. The sophisticated statistical analyses will consider many factors together and provide predictive models for downstream breeding experiments. The biomarker development toolkit developed in this project, consisting of a database, web-based application, and statistical framework will allow for continued dissemination of knowledge. The fluidity of the toolkit allows for updated models as more data becomes available, enabling new biomarkers to be created based on specific and changing needs. The biomarkers uncovered in this project may help produce rice breeds of which farmers can take advantage to achieve higher rice yield. This can have a ripple effect of lower cost in rice production, leading to poverty alleviation through the decrease in the price of rice. The sustainability of the methodology will rely upon the collaborative efforts among agricultural researchers, basic scientists, rice breeders, and farmers through their contributions towards the established database. With the widespread applicability of the methodology and toolkit, the maintenance of the rice line database can become a joint global effort, leading to common food security.

SECTION B: PROJECT DESCRIPTION AND CONTENTS

2.1. Problem definition

Rice and Climate Change

Oryza sativa or Asian rice is a staple food in Asian countries and worldwide, and its continual production is essential to food security. As the fourth most populous country in the world, Indonesia is also one of the biggest producers and consumers of rice. With the current annual population growth rate at 1.21% (World Bank, 2013), the population in Indonesia is predicted to reach 337 million in 2050 (People Facts, 2012). With the current rate of rice consumption at 139 kg/capita/year (Shean, 2012), Indonesia must reach an annual rice production of 47 million tons by 2050.

Climate change will affect rice production through a rise in average temperatures and an increasingly frequent and/or prolonged episodes of extreme area floods and dry spells in Southeast Asia (Singh 2014). Drought is the largest constraint to rice production in the

rainfed agricultural systems of Asia (Pandey 2007). It is estimated that for every 1 degree Celsius increase in temperature, rice yields will decline by 7% (Matthews 1997). To address this challenging outlook, new varieties of rice must be developed to have resilience to adverse climate conditions and have improved production qualities.

Marker-assisted selection (MAS) is a molecular breeding strategy that uses DNA markers to develop varieties of plants with desirable traits, such as disease resistance and yield improvements (especially under conditions like drought stress). The process involves linking variations in the genome to important characteristics and then using those genetic variants to screen plants for breeding. MAS improves the efficiency of the plant breeding process by genetically predicting the phenotype of a plant (often more accurately than direct observation) as early as the seedling stage. MAS reduces the time and resources (such as human labor, water, and plot resource to grow plants) required for breeding because a smaller yet more relevant sample is used in the experiment. There are several breeding strategies based on MAS, such as marker-assisted backcrossing, pyramiding, and early generation screening.

Drought and flooding stress are among the major abiotic factors that deeply impact rice yield. Understanding the relationship through MAS is difficult. Many factors influence complex traits and each genetic variant usually has relatively small effects; it is the cumulative effect that leads to the complex trait. Additionally, the timing, duration, and intensity of the stress and overall climatic conditions are relevant (e.g. stress around time of flowering, lack of oxygen during germination). Traditional candidate gene association studies and single-variant statistical modelling have not made much progress in identifying marker(s) predictive of yield under stress. However, the application of MAS requires accurate and reliable biomarkers linked to the target gene(s) responsible for the phenotype of interest. As the environment influences the relationship between genes and traits, gene-environment interactions must be considered in biomarker development.

Gene-Environment-Wide (GxE) Association Study

The Indonesian Center for Agricultural Biotechnology and Genetic Resources Research and Development (ICABIOGRAD) have developed a genome-wide association study with the purpose of identifying genetic variants that are linked to yield, yield components, and maturity traits of rice and the impact of climatic factors (gene-environment).

A diverse Indonesian rice germplasm collection of 467 accessions including land rices, released varieties, improved lines, and wild species were used in this study. The land rice accessions were selected to represent the range of Indonesian geographical areas. The other accessions were chosen to build upon several previous studies and related breeding programs.

The rice collection was extensively phenotyped for complex traits (described in Table 1), including the flowering and harvesting time as well as the yield component characteristics. Overall phenotyping was conducted in the greenhouse and fields for two planting seasons. Greenhouse phenotyping was completed at the ICABIOGRAD greenhouse during 2011-2012. Field testing was conducted in three locations with different agro–ecosystem, Kuningan (rainy season 2010/2011), Subang (rainy season 2011/2012), and Citayam (rainy season 2012/2013).

The rice genome sequence, SNP database, and diverse linkage map of several candidate genes were used to design the 1536 SNP-chip. The markers were selected from several sources of rice SNP databases: Rice SNP Consortium (www.ricesnp.org); Illumina 1536 SNP chip on 395 *O. sativa* accessions (Zhao 2010); IRFGC-Perlegen SNP (TIGR Pseudomolecules v5); and SNPs of the Rice diversity projects developed by Susan McCouch Lab: www.ricediversity.org).

Table 1: Rice complex traits measured on 467 varieties in 4 locations

Trait	Units	Description
Days to flowering		days after planting when 50% of the plants have flowers
Days to harvest		days after planting until physiological maturity
Total tiller		number of tillers per hill
Productive tiller		number of tillers that produce panicles
Plant height	cm	measured from the ground to the base of the panicle, at the time of flowering
Total panicle		panicles in a square meter
Panicle length	cm	main stem panicle length, measured from the base to the tip of the panicle, 7 days after anthesis.
Filled grain		average number of filled grain clumps per panicle
Unfilled grain		average number of empty grain clumps per panicle
Grain per panicle		total number of grain per panicle
1000 grain weight	gr	weight of 1000 full grain
Yield	t/ha	tons of rice per hectare

The need for a biomarker development toolkit for molecular breeding

Given the complexity and volume of the data collected and the pressing need for breeders to develop robust varieties of rice, a molecular breeding toolkit is needed to accelerate the process of linking genetic variation in rice to important traits (such as morpho-agronomic traits), understanding the influence of climate change on these traits, and building the predictive gene-environment models needed for MAS breeding programs.

The proposed project will focus on building a toolkit for rice biomarkers development using state-of-the-art bioinformatics and statistical methodologies. The toolkit will include a database and web-based application for data management; a statistical framework for single-, multi-variant, and variant-environment interaction modeling and replication and validation; reporting of promising biomarkers for follow-on research from current data; and website for dissemination of results and resources to multiple beneficiaries.

References

Matthews, R. B., Kropff, M. J., Horie, T., & Bachelet, D. *Simulating the impact of climate change on rice production in Asia and evaluating options for adaptation.* Agricultural Systems, 1997. **54**(3): p. 399–425. doi:10.1016/S0308-521X(95)00060-I

Pandey, S. & Bhandari, H. *Drought, coping mechanisms and poverty.* IFAD Occasional Papers, 2009.

People Facts. (2012). Population Growth. Retrieved from http://os-connect.com/pop/p2ai.htm

Shean, M. (2012). INDONESIA: Stagnating Rice Production Ensures Continued Need for Imports. Retrieved from http://www.pecad.fas.usda.gov/highlights/2012/03/Indonesia_rice_Mar2012/

Singh, D., Tsiang, M., Rajaratnam, B., Diffenbaugh, N. S. (n.d.). *Observed changes in extreme wet and dry spells during the South Asian summer monsoon season,* **4**(6), 456–461. doi:10.1038/nclimate2208

World Bank. (2013). Indonesia Population Growth Rate. Retrieved from http://data.world bank.org/country/indonesia

Zhao, K., Wright, M., Kimball, J., Eizenga, G., McClung, A., Kovach, M., et al. *Genomic diversity and introgression in O. sativa reveal the impact of domestication and breeding on the rice genome.* PLoS One, 2010. **5**(5): e10780.

2.2. Project objectives: Overall and specific objectives

The overall objective of this project is building a molecular breeding toolkit for identifying, validating, and updating biomarkers for traits in Indonesian rice that could lead to increased production and adaptability to climate induced-stress. This project uses data from a unique gene-environment-wide (GxE) association study designed by ICABIOGRAD and a state-of-the-art statistical methodology developed by Bina Nusantara University (Jakarta, Indonesia). These biomarkers have strong potential to develop much more resilient rice varieties through marker-assisted selection (MAS) strategies, leading to improved food security due to climate change impact, not just for Indonesia but for many nations with rice as its primary carbohydrate source. The tools developed in this project will be a valuable asset for accelerating rice breeding programs in Indonesia and other member countries, and facilitating the coordination of future trials among rice breeders and farmers in determining the most promising rice varieties. Figure 1 depicts a schematic of the biomarker development toolkit and its relationship with the proposed database, web application, and statistical framework, and the beneficiaries who will utilize the toolkit.

The specifics of the project objectives include:

1. The integration of genotypes, trait, heredity, stress, and climate data for a diverse set of rice (n=467) planted in various locations with distinct climate conditions across multiple seasons in a relational database and the linking of these data to other national and international resources (e.g. data from International Rice Research Institute (IRRI)). The implementation of integration would require the project to develop protocols and a web-based application for beneficiaries to interface to these data, to build powerful queries, and to expand data collection for multiple trials and new rice varieties.

2. The performance of genome-wide and genome-environment-wide (GxE) analyses of complex traits in the diversity panel of 467 rice varieties using appropriate generalized linear modeling.

3. The development and utilization of statistical methodologies to assess systematically the relationships among numerous factors (genetic, ancestry, climatic, soil, and agronomic factors) within relevant abiotic biological pathways. For example, the enzyme RuBisCO connects the Calvin cycle and photorespiration cycle, which are two biochemical pathways crucial for rice plant's carbon source and storage management. During drought, RuBisCO is inactivated and its production lowered due to stomata closure, leading to low carbon dioxide availability for carbon fixation. Numerous factors will be comprehensively modeled using conventional and advanced multivariate techniques.

4. The combination of evidences from independent datasets and consortia (i.e. IRRI) and the evaluation of the replication and predictive properties of promising models from aims (2) and (3).

5. The development of a final report that provides prioritized lists of marginal and interaction associations, and well-fitting models for breeders to evaluate for marker-assisted selection programs; evaluation and advising on high-throughput/screening technologies and the design of follow-on trials; and the integration of findings into the database and web-based application, subsequently providing access to the toolkit to qualified rice breeding and agro-genomic researchers.
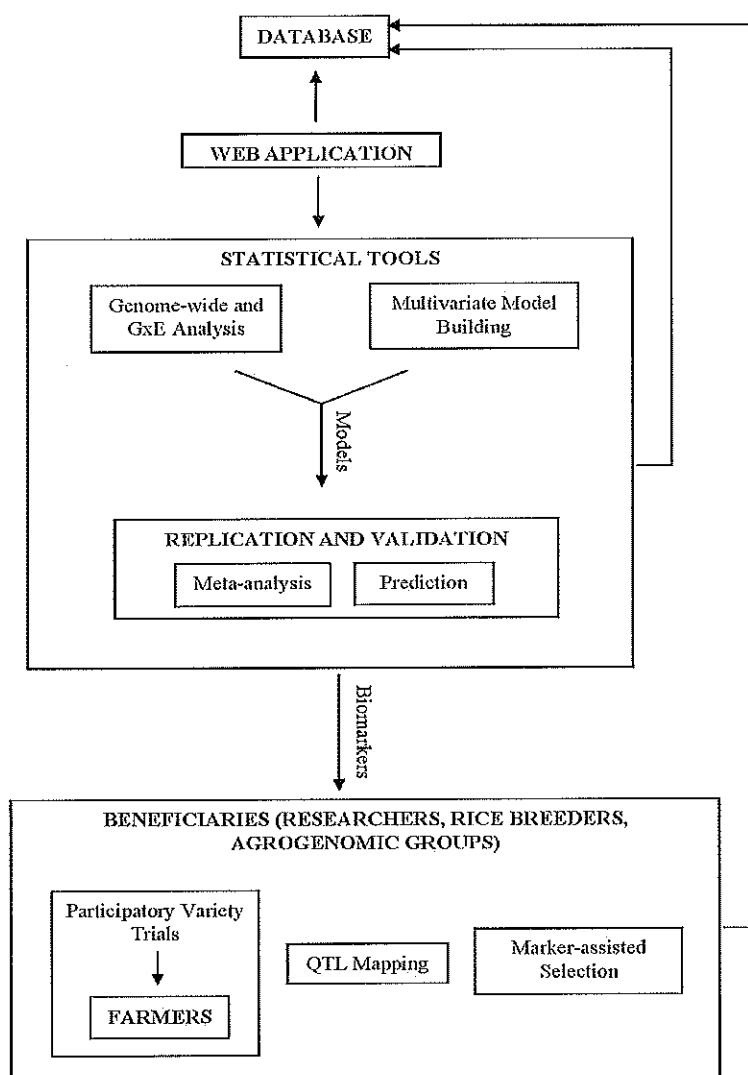
Figure 1. Biomarker Development Toolkit Schematic

2.3. Targeted outputs, activities and related methodology of implementation

Output 1: Integrated rice database and web-based application

Dr. Pardamean's team will develop a database (Activity 1.1) and web application (Activity 1.2) specifically tailored for rice biomarker development. The system will consist of a relational database implemented in PostgreSQL and a software application implemented in PHP. The purpose of the system is to manage and analyze data as well as create linkages to other rice databases for annotation and replication/validation studies (Activity 1.3).

Development will occur at the Bioinformatics and Data Science Research Center (BDSRC) at Bina Nusantara University (Jakarta, Indonesia) using standard software development and relational modelling methodologies. The production system will be housed on the servers at ICABIOGRAD in Bogor, Indonesia.

The BDSRC will integrate the following data into the database:

(1) Genetic data on 467 rice varieties found in Indonesia genotyped with two panels of genome-wide markers. This genotyping data is available from ICABIOGRAD.

(2) The rice diversity panel was extensively phenotyped for complex traits, including the flowering and harvesting time and the yield component characteristics. Overall phenotyping

was conducted in greenhouses and fields over the course of two planting seasons. Greenhouse phenotyping was completed at the ICABIOGRAD during 2011-2012. Field testing was conducted in three locations which have different agro–ecosystems: Kuningan (rainy season 2010/2011), Subang (rainy season 2011/2012) and Citayam (rainy season 2012/2013).

(3) Accession information and subpopulation ancestry data is available from the Indonesian Agency for Agriculture Research and Development (IAARD). In addition to this information, ancestry estimation will be performed in aim 2 by ICABIOGRAD and imported into the database for analysis.

(4) Climatic, soil, and other environmental data for the three field locations (available from IAARD).

(5) Annotation on the content of the custom SNP arrays (database source, references, chromosome, position, alleles, allele frequencies, functional consequence, gene and pathway involvement). This information is available from multiple external sources (e.g., OryzaSNP (http://oryzasnp.plantbiology.msu.edu/).

(6) Links to data from other national and international resources (e.g. IAARD, IRRI)

The web-based application developed by BDSRC will provide tools for searching and writing powerful queries. The application will provide a secure portal for sharing data, results, and statistical tools to qualified beneficiaries.

Output 2: *Genome-wide and genome-environment-wide (GxE) analysis results*

The purpose of data analysis is to identify rice characteristics (both genetic and non-genetic) that influence important rice production traits, and evaluate how the environment and climate impact these relationships. Dr. Baurley will supervise the data analysis effort, coordinating efforts closely with rice scientists at ICABIOGRAD and database developers at BDSRC.

A scientific workflow for data quality control and analysis of genotypes, phenotypes (traits), and environmental data will be developed by BDSRC (Activity 2.1). The workflow will be implemented using a combination of software tools and custom programming including the whole genome association analysis toolset PLINK, ancestry estimation using the software STRUCTURE, the statistical language [R], and packages designed for agriculture research (e.g. STAR and PBTools).

BDSRC will generate descriptive statistics for each variable in [R] and stratified according to location and season/year (Activity 2.2). Descriptive statistics for continuous variables are expressed as mean, median, standard deviation, and ranges. Analyses of continuous variables are performed using t-test or an analysis of variance (ANOVA), as appropriate. Discrete variables are expressed as frequencies and percentages. The analyses of discrete variables are performed using the appropriate chi-square test. Fisher's exact test is used for small cell sizes (< 5). For each trait, all tests of significance are two-tailed, with statistical significance set at $p < 0.05$.

Generalized linear models will be used as a flexible statistical framework for model building and validation,

$$g(\mu_Y) = \alpha_0 + \sum_P \theta_p X_p + \sum_V \beta_v G_v + \dots$$

Here, Y is the outcome of interest, X is a set of P predictor variables and corresponding effect estimates, θ, and G is a set of V genetic predictor variables and corresponding effect estimates, β. Additional variables that may need to be included in the model, such as potential interactions (including gene-environment and gene-gene), are represented by ". . .".

Genome-wide analyses will be performed by BDSRC for each of the complex traits, for greenhouse and the three fields across the two seasons (Activity 2.3). Genome-environment-wide (GxE) association analysis will be performed to identify genetic associations that vary by environmental factors (Activity 2.4). The top marginal and interaction effects will be reported for each scan, both with a naïve correction for ancestry and a mixed model that accounts for relatedness in the diversity panel. Family-wise error rate will be controlled based on a Bonferroni correction for the number of effective number of tests (accounting for linkage disequilibrium).

Output 3: Multivariate models

Multivariate modeling will be used to evaluate the relationships between genetic and environmental factors influencing complex rice traits. Multivariate models have better predictive properties than univariate models for complex traits, such as screening for rice varieties with better response to climatic stress. The multivariate analysis effort will be led by Dr. Baurley with domain expertise on climate modeling, hierarchical modeling, and rice breeding (Dr. Utami - ICABIOGRAD, Bogor, Indonesia).

Initially, conventional techniques will be used to determine the form of the model, i.e. what variables to include. Classical model building and search algorithms (backwards, forwards, stepwise, or all sub-set regression) with standard model selection criteria will be used (Activity 3.1). The resultant single "best" model will then be used to generate prediction for future observations.

Advanced modeling strategies, such as Bayesian hierarchical approaches, can reveal complex statistical associations that were not captured by the use of standard regression modeling. We will complement conventional analysis, with novel methods aimed at interrogating complex relationships among genetic variants and environmental exposures influencing rice traits (Activity 3.2). Bayesian pathway modeling approaches have important advantages over alternative analytic strategies, including the ability to leverage data from external biological knowledge-bases, and account for uncertainty in model variables and parameters (Thomas 2009).

We propose using the Algorithm for Learning Pathway Structure (ALPS) framework, developed by Dr. Baurley and used in many complex trait applications, in this project (Activity 3.2). The ALPS framework uses biological knowledge in the form of a "prior topology", a structure derived from databases or domain experts (Baurley 2010), which gives preference to existing multivariate models. ALPS provides an estimate of the net effect of the entire pathway which is easier to interpret than many interaction coefficients. For example, ALPS can comprehensively model variants in genes involved in signaling network responses to climatic stress. Bayes factors, the ratio of posterior to prior odds, will be used to evaluate the extent the data supports a particular variable (Raftery 1995). Multivariate associations with strong evidence (Bayes factors > 20) will be considered as potential predictive biomarkers and evaluated.

Output 4: Replication and prediction models

The previous steps produced (1) marginal associations between genetic factors and important traits in rice and interactions between genotypes and the environment; (2) the "best" models among many candidate variables using traditional model building strategies; and (3) a set of multivariate models with high Bayes factors.

BDSRC will harmonize data from available sources (e.g. IRRI, literature search) and fit similar models to combine p-values in meta-analyses using standard approaches (Activity 4.1). Consistent evidence across studies will be evaluated and reported.

Using the multivariate models, BDSRC will generate prediction values from additional observations (available from IAARD) and will evaluate them using standard evaluation techniques (Activity 4.2).

The advantage of a Bayesian approach for model selection is that the posterior distribution of the models can be used to account for the uncertainty in the model determination when performing subsequent prediction. Thus, for prediction of a future observation Yi, the posterior predictive distribution can be used. This distribution is a mixture of the model for Y given the specified model form M and observed predictor variables, $\mathbf{D}$; and the posterior distribution of models given the data used for model selection, $\mathbf{Y}$, $\mathbf{X}$, and $\mathbf{G}$. That is,

$$p(Yi|\mathbf{D}) = \int p(Yi|\mathbf{D}, M)p(M|\mathbf{Y}, \mathbf{X}, \mathbf{G})dM$$

The advantage of this approach for prediction is rather than determining a single "best" model, prediction is averaged or integrated over a set of models. This yields models that may have better predictive properties for marker assisted breeding programs because they incorporate uncertainty, leverages prediction performance from multiple models, and avoids issues of over-fitting of any single model to the training dataset.

A Bayesian analysis framework allows Bayesian updating; that is, the posterior distribution results from one study can be used as priors for the next subsequent study. Thus, information can be naturally incorporated into a single predictive modeling framework that uses information from additional trials, including participatory variety trials between breeders and farmers. The statistical framework for Bayesian updating of promising biomarkers will be integrated into the database and web application toolset (Activity 4.3).

Output 5: Final Report, Toolkit, and Workshop

The methods, results, interpretations, and recommendations will be presented in a final report (Activity 5.1). The final report will integrate (1) documentation for the final system (2) summaries of association results from marginal main effect, GxE interaction, and multivariate analyses; and (3) replication and predictive properties of select models. Within the report, recommendations will be made on additional replication and validation steps, experimental designs, and simpler or less expensive genotyping platforms for breeding.

The toolkit will be compiled (Activity 5.2) by Dr. Pardamean's team. The final toolkit will consist of the database, web-based application, statistical framework, and public website. Secure access to the toolkit will be available to research staff and collaborators.

Two international workshops will be organized by Bina Nusantara University with invited speakers to present on bioinformatics and statistical genetics, marker-assisted breeding in rice, and the toolkits developed during this project (Activity 5.3). In addition, this work will be presented at the annual International Rice Congress International meeting.

References

Thomas, D.C., D.V. Conti, J. Baurley, F. Nijhout, M. Reed, and C.M. Ulrich, *Use of pathway information in molecular epidemiology.* Hum Genomics, 2009. **4**(1): p. 21-42.

Baurley, J.W., D.V. Conti, W.J. Gauderman, and D.C. Thomas, *Discovery of complex pathways from observational data.* Stat Med, 2010. **29**(19): p. 1998-2011.

Raftery, A., *Bayesian model selection in social research.* Sociol Methodol, 1995. **25**: p. 111-163.

2.4. Targeted PGRFA

This project will study a diverse Indonesian rice germplasm collection of 467 accessions including land rice, released varieties held by the agricultural agencies, improved lines, and wild species. The land rice lines were selected to represent the multitude of geographical diversity in Indonesia. Other rice lines were chosen based on the need to perform follow-up studies and assist in related rice breeding programs.

It is expected that the toolkit will enable identification of genetic variants in rice related to production and resistance to climate changes such as temperature, drought, and flood. The widespread applicability of the toolkit will enable the analysis and improvement of other Indonesian rice varieties as well as those of other nations. The novel information, including the database, association analyses results, and toolkit will be made available per treaty policy within one year of study conclusion via the web-application. Eligible farmers, researchers, and other relevant parties will be able to access available information, contribute new data, and utilize the toolkit to answer their own queries. This contributes to the non-monetary benefits of the Annex I International Treaty since information and technology from this study will be shared to various groups interested in rice genetics and breeding programs.

Ultimately, adoption of the toolkit developed in this project will accelerate the research needed to create an inventory of rice varieties that can withstand and better adapt to extreme conditions due to climate change. Since the toolkit is capable of expansion in terms of studying additional varieties, traits, and genetic and environmental data, there will be minimal constraints in using the platform in a global effort with multiple nations contributing to and benefiting from adoption of the toolkit. Information resulting from the projects will be made available publicly through the global information system (GIS) infrastructure (to be created in the biennium 2016-2017).

2.5. Target groups and beneficiaries

The proposed project would have several groups of beneficiaries. Basic scientists, rice breeders, and agrogenomics groups will benefit directly from the project. Basic scientists will benefit from the toolkit enabling studies to improve the understanding of how genetic factors in rice found in Indonesia influence various desirable phenotypes. This is a particularly important query due to climate change, rendering the need for investigating the genetic determinants of durability and adaptability towards frequent extreme climatic conditions. More than 2,500 scientists will be impacted from this research through IRRI presentations and high-impact publications. The direct ramification of the toolkit for rice breeders will be the ability to conduct trials and marker assisted selection driven by biomarkers uncovered using the tools. There are approximately 30 rice breeders that may benefit from this project in Indonesia alone. The statistical framework that can account for both genetic and environmental factors will have general applicability to numerous rice breeders in other treaty countries within Southeast Asia. The toolkit will also serve as a model platform for agrogenomics research on other plant crops facing similar climate change challenges as rice crops.

Between 60-70% of Indonesian farmers produce their own seedlings every year instead of purchasing seeds from government agencies (Rachman et al., 2002) due to dissatisfaction with the available seeds. For example, one of the varieties approved by the agricultural department, labeled as the 'ES' variety, has low popularity with farmers due to its inability in withstanding droughts. However, it is distributed widely by the department due to its water immersion resistance. Molecular breeding trials should involve farmers, particularly in poverty areas, since their feedback on existing and new rice lines could help increase adoption of superior varieties, eventually reducing rice production costs and increasing yields significantly. The toolkit developed in this project could help gather and analyze the data

collected through these participatory variety trials to improve and update biomarkers over time.

References

Rachman, B., Rusastra, I.W., Kariyasa, K. *Seed & fertilizer distribution system within the context of government subsidy for rice crops*. Center for Socioeconomic Research and Development for Agriculture, **2002**. p. 204-226.

2.6. Impact and impact pathways

2.6.1. Food security and poverty alleviation

Rice price stability is vital in Indonesia since Indonesians spend a significant portion of their disposable income on rice; for poor households, this could be more than half their disposable income. (Source: http://www.indonesia-investments.com/doing-business/commodities/rice/item183)

Identifying genetic markers related to production traits will help breeders develop varieties of rice with improved yield, quality, and sustainability using molecular breeding methods. This is especially important due to increases in population size and crop stresses due to climate change, such as prolonged drought and extreme wet conditions. Breeding programs based on biomarkers will ultimately increase access and availability of seeds with superior qualities to farmers and ensure food security for the general population. The toolkit designed by this project will accelerate the process of identifying accurate and reliable genetic markers for selecting traits for improved durability.

Improving yield and decreasing the likelihood of failed crops through molecular breeding programs will lead to more affordable rice, thus helping poorer communities. The biomarkers uncovered in this project may help produce rice breeds of which farmers can take advantage, providing government agricultural agencies with improved rice varieties to sell. Consequently, farmers will have more incentive to take advantage of the subsidy that comes with purchasing rice seeds from the government. Due to the relatively low quality rice seeds sold by these agencies, about 60-70% of Indonesian rice farmers opt to spend significant amount of time, money, and resources into maintaining their own rice seed instead on an individual basis (Rachman et al., 2002). The combination of government subsidy, improved rice seed quality, and a more efficient allocation of resources would enable farmers to produce rice at a lower cost, thereby lowering the price of rice sold to the general public, especially in impoverished areas.

2.6.2. Adaptation to climate change and environmental sustainability

Rising global temperatures will lead to unpredictable weather patterns and prolonged periods of drought and wet conditions. Irrigation and flood control are not practical for most rain-fed areas. Genetic variation influences yield but the level of influence varies, depending on stresses from climate and other factors. Rice is a plant that is extremely sensitive to changes in abiotic factors; even slight changes in temperature, soil oxygen level, and water availability can lead to reduced yield. An approach for responding to such challenges would be developing biomarkers designed to screen rice varieties based on anticipated/predicted future changes caused by climate change (e.g. extreme drought). Thus, trials can be designed and performed to identify seeds that can withstand these sub-optimal environments. The toolkit is designed to allow the collection of new data from such trials and use the entire dataset in generating or revising biomarkers for MAS.

Climate change may lead to water depletion in certain regions but water excess in other areas. The timing, duration, and intensity of these conditions are important factors on yield, especially during the last two months of the growing season and around the time of

flowering. The toolkit will enable the capturing of this information and help in identifying the genetic variants related to yield and yield under stress. Since this project will take into account both environmental and genetic factors in interaction models, biomarkers developed using the toolkit may be able to identify subgroups of rice suited to particular environmental conditions. Developing new rice lines that are less sensitive to changes in these factors will prepare countries for the challenges posed by climate change.

2.6.3. Scientific impact

This project will identify genetic markers related to complex traits in rice and how those relationships are influenced by climate change. ICABIOGRAD previously has established a unique genome wide association study. The data produced by this study will be analyzed with the statistical methodology designed by Bina Nusantara University. The result of the analysis will confirm the reliability of existing markers and discover new biomarkers that can be used in rice breeding programs to screen for rice with superior production qualities and can withstand stresses brought about by the environment (e.g., prolonged drought and extreme wet conditions).

The sophisticated statistical analyses will consider many factors simultaneously and provide predictive models for downstream breeding experiments. The toolkit developed in this project, consisting of a database, web-based application, and statistical framework will allow for continued dissemination of knowledge. The fluidity of the toolkit allows for updated or new modeling as more data becomes available (increasing statistical power), enabling new biomarkers to be created based on specific and changing needs. The general and widespread applicability of the toolkit to investigating gene-environment interactions facilitates the ability to screen rice varieties for suitability to certain climatic conditions.

The results from this project, including the toolkit and the precipitating information from the use of the toolkit, will be disseminated to basic scientists, breeders, and agrogeneticists through a public website. The knowledge that the toolkit generates would be further distributed through publications in scientific journals and presentations at international conferences in various agronomic sub-fields such as plant genetics, agro-genetics, and sustainable agriculture. Results of impact to rice breeders in Indonesia will be published in regional journals such as the Indonesian Journal of Agricultural Science, Agro Biogen Journal, and Germplasm Bulletin.

2.7. Relevance to national or regional priorities in its plans and programmes for PGRFA

Indonesia's national rice breeding program is overseen by the Indonesian Agency for Agricultural Research and Development (IAARD), which is an institution established by the Ministry of Agriculture. Its goal is to develop and strengthen a global information system to facilitate the exchange of information, based on existing information systems, on scientific, technical, and environmental aspects relevant to plant genetic resources for agriculture. IAARD oversees eleven research centers with the main function of managing studies on food crops, horticulture, estate crops, livestock, veterinary, soil and agro-climate, agro-socio economics, machinery development, post-harvest, biotechnology and agricultural technology assessment. The toolkit developed in this project has general and widespread applicability, supporting the global sharing and knowledge exchange goals of IAARD in efforts of managing rice breeding programs.

To disseminate new knowledge and information within the country on a national level, IAARD established thirty one Assessment Institutes for Agricultural Technology (AIAT) across Indonesia's provinces. Each AIAT is located near specific farming areas they serve with several agricultural posts, enabling for fast adoption of new information and technology.

IAARD partakes in different international engagements through its collaborations with international contracting parties. IAARD's organization of international treaties, such as the third High-level Roundtable on the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRA). The treaty was established within the context of the International Conference on Plant Genetic Resources, Food Security, and Climate Change. The High-level Roundtable focuses on policy and technological advancement contributions to cope with global challenges related to biodiversity, climate change, and food security. IAARD also has regional collaborations with institutions in other nations that are major rice producers such as Vietnam, Thailand, India, and Korea. With such an extensive collaborative network across multiple nations, new knowledge resulted from the development of the toolkit can be distributed quickly to many groups across regions and nations.

SECTION C: OPERATIONS

3.1. Methodology of project implementation

The Bioinformatics and Data Science Research Center (BDSRC) of Bina Nusantara University (Lead: Dr. Bens Pardamean) will coordinate the project implementation and be responsible for the project management, database development, web-based application and public web site, workshop, and dissemination of the toolkit. The Indonesian Center for Agricultural Biotechnology and Genetic Resources (Lead: Dr. Dwinita Wikan Utami) will be responsible for data collection, interpretation of results, dissemination, and design of follow-up experiments. BDSRC's research partners (Lead: James Baurley, Ph.D.) will advise on the database design, develop novel statistical methodologies, build predictive models, and report results.

Appendix 3 presents the project workplan. Development of the integrated rice database and web-based application will be led by BDSRC and completed in the first three months of the project. Upon completion of the database in Month 2, BDSRC team will coordinate workflow development and univariate analysis efforts in Month 3 and Month 4. The results will be reviewed by ICABIOGRAD. Upon completion of quality control and descriptive analyses steps in Month 4, BDSRC team will proceed with genome-wide and GxE analyses in Months 5 and 6. Working with experts in rice breeding at ICABIOGRAD, climate modeling, and statistical methodology, BDSRC will develop multivariate models in Months 5-9. In Month 7, BDSRC will harmonize data from external sources and incorporate these data into the project database. In Month 8, BDSRC will perform meta-analyses; and in Months 9 and 10 will evaluate models for prediction. Models that perform well will be incorporated into the toolkit by BDSRC and tested. BDSRC will coordinate preparation of the final report, final toolkit, and workshops with ICABIOGRAD in Months 11 and 12.

Database, software, and statistical platforms will be developed using industry standard best practices and guidelines for data modeling, software engineering, and statistical reporting. Software tools for project management (https://basecamp.com/), software versioning (http://git-scm.com/), and bug tracking (https://www.atlassian.com/software/jira) will be utilized.

3.2. Partnerships and collaboration arrangements

This project will be a national and international collaboration among researchers at five institutions. The Bioinformatics and Data Science Research Center (BDSRC) of Bina Nusantara University (Jakarta, Indonesia) will lead the project, provide information technology support, and be responsible for project management, dissemination and reporting. The Indonesian Center for Agricultural Biotechnology and Genetic Resources (ICABIOGRAD), Indonesian Agency for Agriculture Research and Development, Ministry

of Agriculture (Bogor, Indonesia) will provide genomic and phenotypic data, review results, perform follow-up experiments, and assist in disseminating results to rice breeders, farmers, and international consortia. ICABIOGRAD has a long standing memorandum of understanding (MOU) with Bina Nusantara University. The Bina University has a good track record of multiyear research collaboration with multiple partners from countries such as Australia and the USA. The Bina University will seek support and expertise from relevant international partners during the implementation of this proposal.

3.3. Project management team

Name	Institution	Role/expertise	Responsibility
Dr. Bens Pardamean	Bina Nusantara University	Data scientist	Project Coordinator
James Baurley, PhD	Bina Nusantara University	Statistical geneticist and data scientist	Co-Project Coordinator
Fitra Kacamarga	Bina Nusantara University	Programmer and analyst	Database programming and analysis
Anzaludin Samsinga	Bina Nusantara University	Web progammmer	Website programming
Carissa Pardamean	Bina Nusantara University	Data analyst and geneticists	Statistical modeling of project data
Hery Muljo	Bina Nusantara University	Administration and finance	Project administration
Dr. Dwinita Wikan Utami	Indonesian Center for Agricultural Biotechnology and Genetic Resources	Bio-molecular and plant breeder	Providing genomic and phenotypic data
Dr. Muhammad Sabran	Indonesian Center for Agricultural Biotechnology and Genetic Resources	Statistician	Advising on statistical modeling

3.4. Sustainability

The toolkit developed in this project will have widespread applicability. The database is scalable, meaning additional variables and data may enter from other rice research programs. The toolkit represents a common platform for sharing individual level and summary data, enabled increased statistical power for producing and validating biomarkers. This will serve as the basis for continued use and adoption of the project database, website, application, and statistical tools beyond the completion of this project.

IAARD's active efforts in improving rice breeding programs at national, regional, and international levels mean that the toolkit will receive continued support, both in monetary and non-monetary aspects, such as funding to maintain the toolkit, database, and genebanks. The toolkit can also serve as a model for other market-assisted molecular studies and genome-wide association projects in terms of improving crop quality and endurance, be it rice or other major food crops. The propagation of knowledge and methodology from the study will be further amplified by the establishment of the global information system (GIS) infrastructure

between the years 2016-2017. Through GIS, information gathered by this toolkit will be readily available to the public. Therefore, sustaining the toolkit and database will not be an action performed only during the study period by the specific groups involved. It will be a maintenance achieved through the active engagement and collaboration among any formal bodies across the globe that strive to improve food security in the future.

SECTION D: APPENDIXES

By signing this submission form for full proposal, the applicant confirms that all the above statements, including the attached Appendixes, are true to the best of his/her knowledge. Any deliberately untruthful response will lead to the automatic exclusion from the further screening and appraisal process, and may lead to the denial of awarded grants from the Benefit-sharing Fund.

Signature of contact person: **Date and location:**

5 December 2014

Jakarta, Indonesia