

Development of Laos Khao Kai Noi rice landrace (*Oryza sativa* L.) core collection as a model for rice genetic resources management in the Laos National Genebank

K Vilayheuang¹, E Borrayo², M Kawase² and K N Watanabe^{2,3*}

¹ Rice Research Center, National Agriculture and Forestry Research Institute (NAFRI), Vientiane, PO Box 811, Laos P.D.R.

² Tsukuba Plan Innovation Research Center, University of Tsukuba, 1-1-1 Tennodai 305-8572 Tsukuba, Ibaraki, Japan

³ Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8572, Japan

*E-mail: watanabe.kazuo.fa@u.tsukuba.ac.jp

Abstract. Khao Kai Noi rice is considered as an elite quality landrace in Laos, which has led to its germplasm conservation in the Laos National Genebank. As happens with other germplasm collections, a manageable yet representative sub collection has become an essential element for researchers and breeders to simplify many activities, including those related to crop improvement, phenotype-genotype correlation and determination of diversity hotspots. In this study, 109 accessions were used as a test collection for core collection development to determine the feasibility of collection reduction in a closely related rice group. Three core collections were developed by two established methodologies and evaluated by diversity indexes, allele retention, phylogenetic distribution and geographical location. Based on SSR molecular markers and PowerCore, a reduction to 24 accessions was achieved with the conservation of complete genetic diversity. A K-means based on reduction to 24 accessions rendered slightly lesser results while based on 12 accessions resulted in a 17% diversity loss. These core collections may be useful for genebank management, research and breeding activities in the future. Also, they may as well serve to estimate core collection development behavior in other landraces and cultivars, which is fundamental in genetic resources management and utilization.

Keywords: rice, core collection, Khao Kai Noi, PowerCore, K-means.

1. Introduction

Genebanks play a key role in the conservation, availability and use of a wide range of plant genetic resources for crop improvement in order to provide food and nutrition security. They help to ensure the continued availability of genetic resources that may have become disused in order to aid future research and breeding programs [1]. As collections tend to increase in numbers as time goes by, maintenance costs and management effort increase as well. Also, diversity analysis, phenotype-genotype correlation and low redundancy becomes challenging, which in turn affect the genebank ability to supply users and breeders with the most adequate materials to perform their activities. A suitable solution to this issue is the development of a “core collection” derived from the existing



germplasm collection, which consists of a limited set of accessions, chosen to represent the genetic spectrum in the whole collection in order to facilitate mentioned management, research and supply [2].

This process is generally known as core collection development (CCD) and was first proposed by Frankel and Brown [3], and as stated, it is of high importance as it plays a significant role in the management and use of genetic resources [2]. CCD has been implemented by many genebanks in diverse crops such as rice [4–6], common beans [7], chilis [8], barley [9,10], soybeans [11] and sesame [12]. Although there exist several approaches and different particular objectives, it has become clear that many factors need to be considered for an adequate CCD, where selection criteria and genetic structure play an important role [13], as well as the evaluation procedures that best fit the particular objectives, as some core collections are being created to represent specific sections of germplasm collections [14]. Strategies for CCD of these specific sections need to be further explored and implemented.

The Laos National Genebank (LNG) is located in the Rice Research Center, National Agriculture and Forestry Research Institute (NAFRI). Among its different functions, LNG provides both active and base germplasm collections, which include more than 14,000 rice accessions. This collection includes a group of Khao Kai Noi (KKN), a high quality regional Laos landrace with high importance in local consumption and export value. Currently, about 200 accessions of KKN germplasm are conserved in the LNG, collected in different efforts since 1995 [15–17].

KKN was chosen as the model group for CCD in LNG for two principal reasons. The first one is related to the direct application of the KKN core collection in research of genetic diversity, genetic structure and diversity hotspots determination; which in turn would lead to adequate *in situ/ex situ* conservation recommendations, breeding improvement and supply. The second one has to do with CCD in a closely related rice landrace, in order to evaluate the properties of a core collection of such nature, and to be able to extrapolate the CCD methodologies to the entire Laos rice collection, as well as to other important landraces and cultivars in the near future.

PowerCore [18] is a reliable and open access software that has been successfully applied in the development of several subsets from rice collections [19,20], as well as from other crops like chili [8] and Turkish melon [21]. PowerCore implementation provides 100% allelic coverage, yet the target core element number depends on the original collection selecting parameters distribution. A K-means based algorithm [22] may complement our implementation of PowerCore, as it allows determination of a target core element number *a priori*.

In this work, we propose the establishment of core collections from KKN collection in LNB, which could be useful for breeding purposes, identifying diversity hotspots, phenotype-genotype correlation, genebank germplasm management and recommendation for on-farm conservation sites for this important rice landrace, as well as serves as a model for CCD implementation in other landraces and cultivars.

2. Materials and methods

One hundred and nine non-redundant accessions were selected from the available KKN in LNG as a whole collection model (Supplemental Table 1). In order to construct our whole collection model genetic data set, a random individual for each non redundant accession was selected and associated with 24 highly informative genome spread SSR data. Molecular characterization procedures have been published elsewhere [23].

Data was analyzed with PowerCore V1.0 in its default parameters, which provided a core collection number that preserves the totality of allele diversity. This number was used as fixed target in the K-means algorithm for CCD to construct a second core collection with the same number of elements. A third core collection was built with K-means, using half previous fixed target to determine impact on such reduction.

Evaluation of the selected core collections was performed by comparing original collection and selected core collections in terms of: Nei and Shannon-Weaver diversity indexes [18], allelic representation, phylogenetic clustering representation, and, when possible, geographical distribution.

Diversity indexes and allelic representation were calculated with PowerCore. Phylogenetic clustering was determined by an UPGMA dendrogram, which was first constructed by PowerMarker V3.25 software [24] and then visualized by MEGA6 [25], where the representative accessions were tracked. Geographic distribution of those accessions with available geographical reference data were plotted in a Xiengkhouang and Houaphan Laos Provinces Map by ArcGIS (<http://doc.arcgis.com/en/arcgis-online/>).

3. Results and discussion

To establish representative accessions of KKN as a tool for breeding purposes, diversity hotspots identification, genetic resources management and on-farm conservation sites recommendations for this rice landrace, 109 accessions were analyzed for 24 SSR genome spread molecular markers. The minimum number for complete allelic representation was determined as the 24 accessions selected by PowerCore. Therefore, 24 and 12 accessions were the target for K-means CCD algorithm.

As established above, the PowerCore core collection was able to represent 100% of allele diversity, with a Nei index of 0.48 and a Shannon-Weaver of 0.94. K-means core collections represented 92% of allele diversity, with a Nei index of 0.51 and a Shannon-Weaver of 0.94 when 24 accessions were selected; 83% allelic diversity representation with a Nei index of 0.51 and a Shannon-Weaver of 1.0 when 12 accessions were selected (Figure 1).

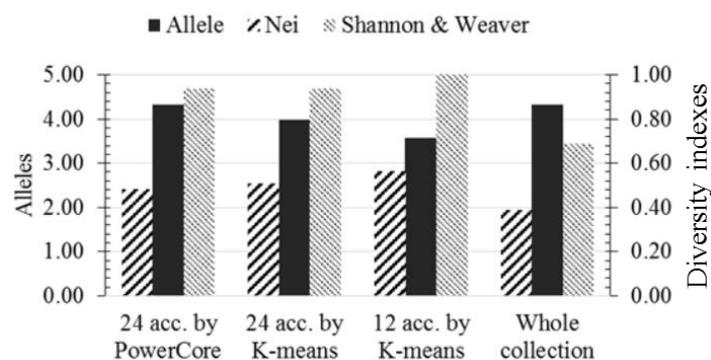


Figure 1. Diversity indexes and allele representation of core accessions of Khao Kai Noi and its whole collection.

Cluster representation among the constructed phylogenetic dendrogram is presented in Figure 2, where constructed core collections with 24 accessions included an accession from each of the 12 major clusters. Constructed core collection with 12 accessions included an accession from 9 of mentioned clusters.

KKN can be classified by its phenotype in 6 groups: “Deng” (red), “Leuang” (yellow), “Hay” (upland), “Khao” (white), “Lai” (striped) and “Lai Dam” (striped and black). Twenty-four accessions included at least a member for each group, while 12 accession core collection failed to represent “Khao” (Supplementary Table 1). The geographical distribution of all core collection accessions were distributed along and covered all provinces where KKN is primarily produced (Figure 3).

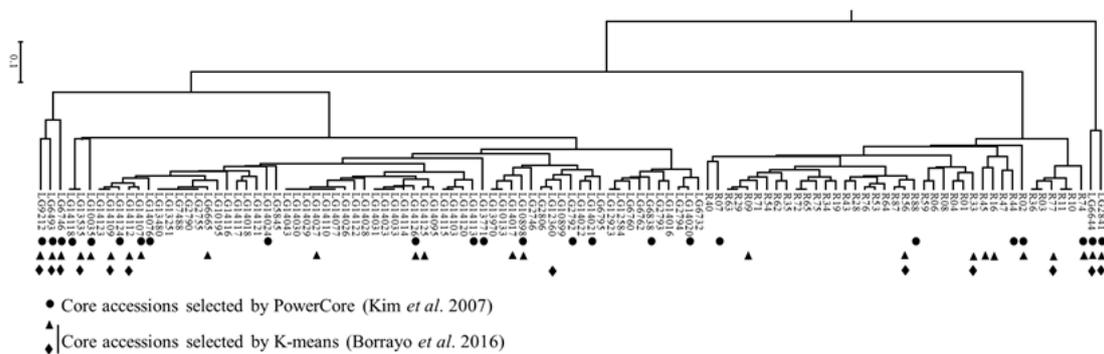


Figure 2. UPGMA dendrogram of 109 accessions of Khao Kai Noi. Marks below each accession indicate its selection as a core collection element.

The core collection consists of a limited set of accessions derived from an existing germplasm collection, chosen to represent the genetic spectrum in the whole collection and should include as much as possible of its genetic diversity [2]. In this study, we aimed to determine the feasibility to create a useful core collection from a closely related rice landrace. A CCD within KKN would prove handfull in more than one way, as it can provide priority accessions for *in situ/ex situ* conservation, as well as to optimize resources on breeding and research, and provide an important insight of how these CCD methodologies may respond to other landraces or cultivars.

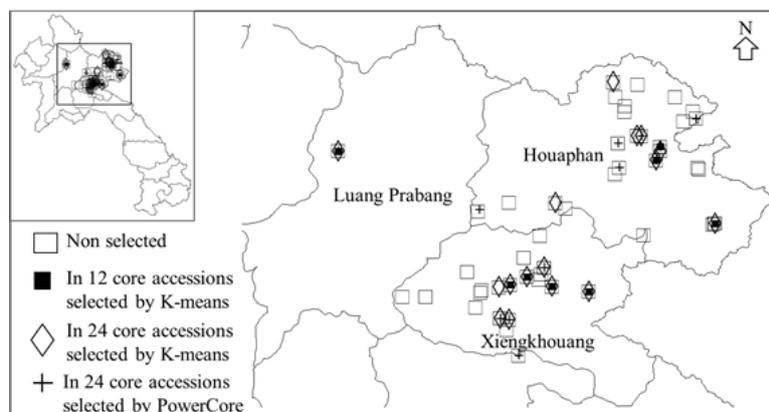


Figure 3. Map of Laos showing the distribution of 62 of 109 accessions of Khao Kai Noi used for collection development. The map was created by using ArcGIS.

The first approach was to determine a core collection that would maintain all possible diversity, which is possible by PowerCore implementation. However, a 22% core collection appear as a high value, distant from the recommended consensus percentage mentioned by Frankel and Brown [3], who indicated that core collections should be reduced to 5% from large and 10% from small collections. This led to the implementation of another methodology that could establish a target value regardless of allele retention percentage. By K-means 24-accession implementation, we wanted to compare methodologies under similar conditions, while 12-accession implementation allow us to determine the behavior under mentioned percentages that have been used in similar CCD [8].

Both PowerCore and K-means 24-accession behaved very similar in terms of diversity indexes, as both clearly reduced redundancy from the original collection. K-means is not able to represent the full allele diversity. This is expected as the selection method for this algorithm is to select a sample closest to centromere of each of the target-number generated groups, which explain why Nei index is higher

in K-means than in PowerCore's selected accessions. In terms of hierarchical clustering and geographic representation, these two core collections had again similar results, proving to be effective methodologies for CCD in the evaluated landrace.

K-means 12-accessions did not comply as effectively as its evaluated counterparts, and a phenotype was not included among the selected elements. However, considering the fact that this core collection had a 50% accession deficit compared to the other ones, it is interesting to note that there was only a 17% allele reduction representation; 2 of the 3 missing cluster representation in the hierarchical clustering were single accession groups; diversity indexes were maximized. It is also important to consider that the main strength of K-means is the inclusion of both agromorphological traits and genetic information, although in this preliminary study only the second one was included. There may be some circumstances where this core collection may prove useful, particularly when priority for conservation or distribution must be provided for budget limitations.

4. Conclusions

In summary, we present three KKN core collections that have their valuable for the landscape's genetic resources management, research and breeding. We believe that the methodologies implemented in this work may be successfully extrapolated to other rice landraces and cultivars, with similar results when selecting a core collection from a highly related original collection.

5. Acknowledgement

This research was supported by Grants-in-Aid #24405049 and #25257416 from the Japan Society for the Promotion of Science.

6. References

- [1] FAO 1994 *Genebank Standards* (Rome, Italy: International Plant Genetic Resources Institute)
- [2] Brown A H D 1995 The core collection at the crossroads *Core Collections of Plant Genetic Resources* ed T Hodgkin, A H D Brown, T J van Hintum and E A Morales (Chichester, United Kingdom: John Wiley and Sons) pp 3–19
- [3] Frankel O H and Brown A H D 1984 Plant genetic resources today: a critical appraisal *Crop Genetic Resources: Conservation & Evaluation* ed J Holden and J Williams (London: Allen & Unwin) pp 249–57
- [4] Chung H, Kim K, Chung J, Lee J, Lee S, Dixit A, Kang H, Zhao W, McNally K L and Hamilton R S 2009 Development of a core set from a large rice collection using a modified heuristic algorithm to retain maximum diversity *J. Integr. Plant Biol.* **51** 1116–25
- [5] Ebana K, Kojima Y, Fukuoka S, Nagamine T and Kawase M 2008 Development of mini core collection of Japanese rice landrace *Breed. Sci.* **58** 281–91
- [6] Tiwari K K, Singh A, Pattnaik S, Sandhu M, Kaur S, Jain S, Tiwari S, Mehrotra S, Anumalla M and Samal R 2015 Identification of a diverse mini-core panel of Indian rice germplasm based on genotyping using microsatellite markers *Plant Breed.* **134** 164–71
- [7] Zeven A C, Waninge J, Van Hintum T and Singh S P 1999 Phenotypic variation in a core collection of common bean (*Phaseolus vulgaris* L.) in the Netherlands *Euphytica* **109** 93–106
- [8] Mongkolporn O, Hanyong S, Chunwongse J and Wasee S 2015 Establishment of a core collection of chilli germplasm using microsatellite analysis *Plant Genet. Resour.* **13** 104–10
- [9] Igartua E, Gracia M P, Lasa J M, Medina B, Molina-Cano J L, Montoya J L and Romagosa I 1998 The Spanish barley core collection *Genet. Resour. Crop Evol.* **45** 475–81
- [10] Muñoz-Amatriáin M, Cuesta-Marcos A, Endelman J B, Comadran J, Bonman J M, Bockelman H E, Chao S, Russell J, Waugh R and Hayes P M 2014 The USDA barley core collection: genetic diversity, population structure, and potential for genome-wide association studies *PLoS One* **9** 1–13
- [11] Kaga A, Shimizu T, Watanabe S, Tsubokura Y, Katayose Y, Harada K, Vaughan D A and

- Tomooka N 2012 Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections *Breed. Sci.* **61** 566–92
- [12] Xiurong Z, Yingzhong Z, Yong C, Xiangyun F, Qingyuan G, Mingde Z and Hodgkin T 2000 Establishment of sesame germplasm core collection in China *Genet. Resour. Crop Evol.* **47** 273–9
- [13] Odong T L, Van Heerwaarden J, Jansen J, van Hintum T J L and Van Eeuwijk F A 2011 Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data *Theor. Appl. Genet.* **123** 195–205
- [14] Odong T L, Jansen J, Van Eeuwijk F A and van Hintum T J L 2013 Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation *Theor. Appl. Genet.* **126** 289–305
- [15] Rao S A, Schiller, J M, Bounphanousay C and Jackson M 2006 Diversity within the traditional rice varieties of Laos *Rice in Laos* ed J M Schiller, M B Chanphengxay, B Linquist and S Rao (Los Baños, Philippines: IRRI) pp 123–40
- [16] Bounphanousay C, Bounphanousay V, Kanyavong K and Inthapanya P 2009 Khao Kay Noi (“Small Chicken Rice”) is high grain quality confined to higher altitude *Lao J. Agric.* **5** 1–22
- [17] Vilayheuang K, Okuizumi H, Yamamoto S and Watanabe K N 2014 Rice landrace (*Oryza sativa* L.) field survey and collection in Xiengkhouang and Houaphan Provinces of Laos in 2014 *Annu. Rep. an Explor. Introd. Plant Genet. Resour.* **31** 314–7
- [18] Kim K W, Chung H K, Cho G T, Ma K H, Chandrabalan D, Gwag J G, Kim T S, Cho E G and Park Y J 2007 PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets *Bioinformatics* **23** 2155–62
- [19] Agrama H A, Yan W, Lee F, Fjellstrom R, Chen M H, Jia M and McClung A 2009 Genetic assessment of a mini-core subset developed from the USDA rice genebank *Crop Sci.* **49** 1336–46
- [20] Zhang B, Tian F, Tan L, Xie D and Sun C 2011 Characterization of a novel high-tillering dwarf 3 mutant in rice *J. Genet. Genomics* **38** 411–8
- [21] Frary A, Şığva H Ö, Tan A, Taşkın T, İnal A, Mutlu S, Haytaoğlu M and Doğanlar S 2013 Molecular genetic diversity in the Turkish national melon collection and selection of a preliminary core set *J. Am. Soc. Hortic. Sci.* **138** 50–6
- [22] Borrayo E, Machida-Hirano R, Takeya M, Kawase M and Watanabe K 2016 Principal components analysis-K-means transposon element based foxtail millet core collection selection method *BMC Genet.* **17** 42
- [23] Vilayheuang K, Machida-Hirano R, Bounphanousay C and Watanabe K N 2016 Genetic diversity and population structure of “Khao Kai Noi”, a Lao rice (*Oryza sativa* L.) landrace, revealed by microsatellite DNA markers *Breed. Sci.* **66** 204–12
- [24] Liu K and Muse S V 2005 PowerMarker: an integrated analysis environment for genetic marker analysis *Bioinformatics* **21** 2128–9
- [25] Tamura K, Stecher G, Peterson D, Filipski A and Kumar S 2013 MEGA6: molecular evolutionary genetics analysis version 6.0 *Mol. Biol. Evol.* **30** 2725–9