Table S1. Overview of existing pan-genome studies of large eukaryotes.

Organism	# genomes	Method	Sequencing depth	Gene PAV	Pan-genome	Reference & Year
Homo sapiens	3	assembly only	>45x	no	yes	Li, et al. 2010
Glycine soja	7	traditional	>104x	yes	yes	Li, et al. 2014
Zea Mays [a]	503	pan-transcriptome	NA	no	yes	Hirsch, et al. 2014
Oryza sativa L.	3	traditional	110x	yes	yes	Schatz, et al. 2014
Oryza sativa L.	1,483	metagenome-like	1-3x	no	yes	Yao, et al. 2015
Oryza sativa L [b]	453(3,010)	EUPAN	>20x(~14x)	yes	yes	Wang, et al. 2016 Sun, et al. 2016

[a] This is a pan-transcriptome study, in which mRNA-seq was carried out for each sample.

[b] In the 3,000 Rice Genome Project, the pan-genome sequences were derived from assemblies of 3,010 accessions with an average sequencing depth of 14x and gene PAVs were detected only for 453 accessions with sequencing depths >20x.

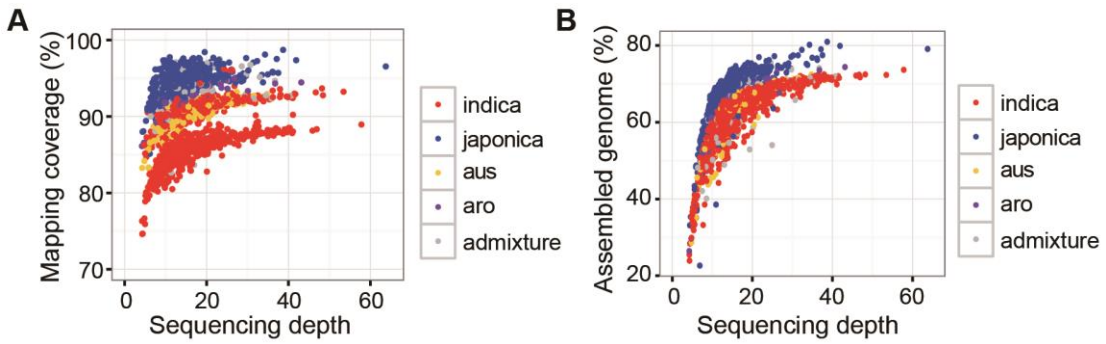

Figure S1. Genomic read mapping and *de novo* genome assembly of 3,024 rice accessions against the japonica reference genome (Nipponbare accession). The colors indicated 5 subpopulations of rice accessions. **A:** mapping coverage vs. sequencing depth. The trend of mapping coverage was stable at sequencing depth >20X, with the differences among subpopulations reflecting subpopulation differentiation. The sequenced Nipponbare rice accession, the same as the reference genome possessed a sequencing depth of ~20X and a mapping coverage of 98.43%. Considering Ns in the reference genome, we concluded that whole genome could be covered at sequencing depth >20X. This depth might be somehow different for other species. **B:** Proportion of *de novo* assembly vs. sequencing depth. The y-axis showed the how much of the genome can be assembled. This is evaluated by aligning contigs to the reference genome and calculating what proportion of the genome is covered. In general, no more than 75% of the genome could be assembled at sequencing depth of ~20X and only about 80% of the genome could be assembled at sequencing depth >40X.

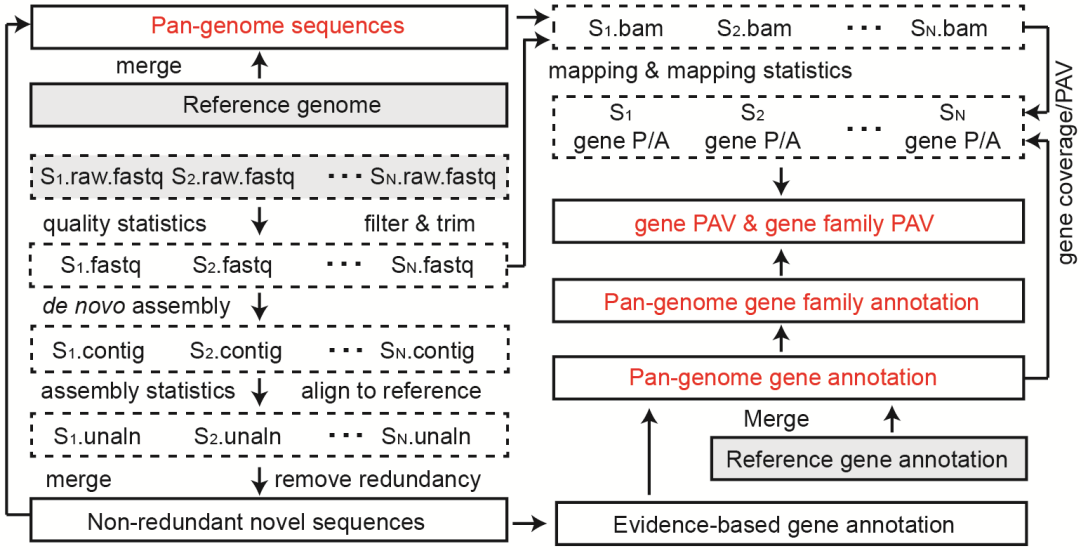

Figure S2. The flow diagram and functions of EUPAN tools: (1) EUPAN checks and plots the overall sequencing qualities and then extracts high-quality reads by filtering or trimming; (2) EUPAN carried out de novo assemblies and automatically selects the best Kmer parameter for each sample; (3) EUPAN aligns the assemblies to a reference genome and evaluates the assemblies; (4) EUPAN builds the pan-genome by combining the reference genome and a set of non-redundant novel sequences; (5) EUPAN maps high-quality reads to the pan-genome/reference genome and checks the mapping statistics; (6) EUPAN calculates gene coverage and gene PAVs based on pan-genome gene annotation; (7) EUPAN calculates gene family PAVs from gene PAVs and gene family annotation. Raw data for EUPAN are shown in grey boxes and key outputs of EUPAN are shown in red. Please note that gene annotation and gene family annotations are not integrated into EUPAN, since its high complexity and flexibility.