Logical Operation Based Literature Association with Genes and its application, PosMed.

Yuko Makita[1]◯, Rinki Bhatia[1], Mrinalini Deshpande[1], Akihiro Matsushima[1] , Manabu Ishii[1] ,
Yoshiki Mochizuki[1] , Yuko Yoshida[1], Norio Kobayashi[1], Tetsuro Toyoda[1]
1. Bioinformatics And Systems Engineering (BASE) division, RIKEN

PosMed prioritizes candidate genes for positional cloning by employing our original database search engine GRASE, which uses an inferential process similar to an artificial neural network comprising documental neurons (or 'documentrons') that represent each document contained in databases such as MEDLINE and OMIM (Yoshida, et al. 2009). PosMed immediately ranks the candidate genes by connecting phenotypic keywords to the genes through connections representing gene–gene interactions other biological relationships, such as metabolite–gene, mutant mouse–gene, drug–gene, disease–gene, and protein–protein interactions, ortholog data, and gene–literature connections.

To make proper relationships between genes and literature, we manually curate queries, which are defined by logical operation rules, against MEDLINE. For example, to detect a set of MEDLINE documents for the AT1G03880 gene in *A. thaliana*, we applied the following logical query: ('AT1G03880' OR 'CRU2' OR 'CRB' OR 'CRUCIFERIN 2' OR 'CRUCIFERIN B') AND ('Arabidopsis') NOT ('chloroplast RNA binding'). Curators refined these queries in mouse, rice and *A. thaliana*. For human and rat genes, we use mouse curation results via ortholog genes in PosMed.

Fig. 1 PosMed-plus accelerates forward-genetics gene discoveries (left half of the chart) by integrating the omics knowledge collected from reverse genetics (right half of the chart).

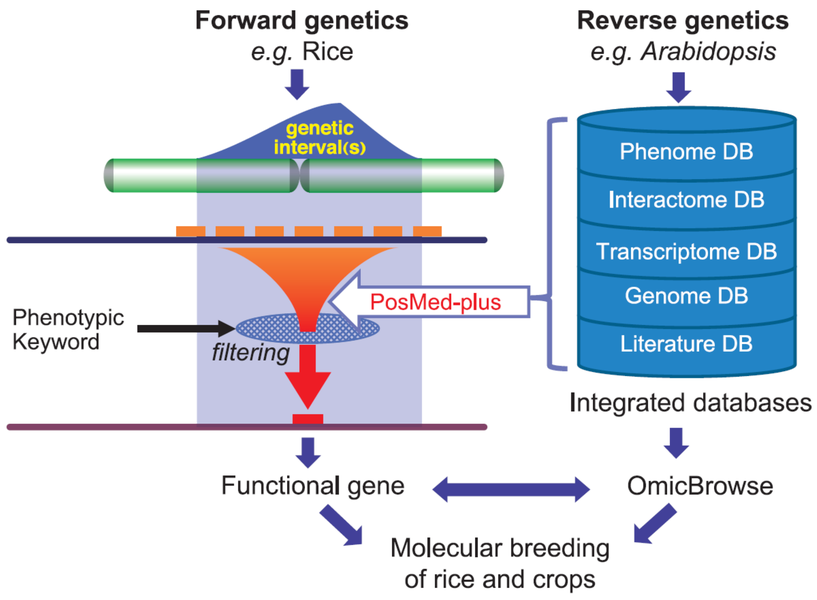


Fig. 2 Model of the intelligent search engine for PosMed-plus. PosMed-plus handles highly connected network and reply candidate genes depending on the correlation with the phenotypic keyword.

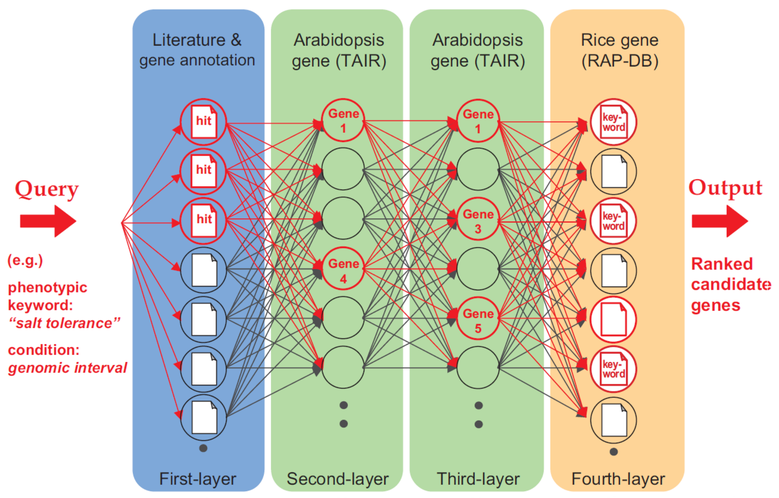


Fig. 3 Data flow of PosMed-plus search and comparison with the direct searches and the inference searches.

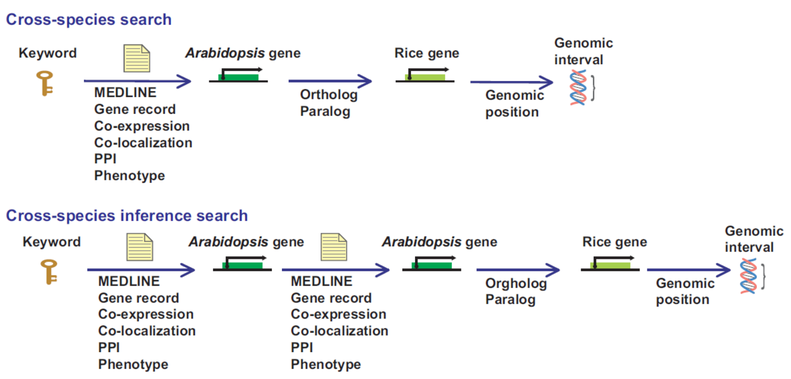
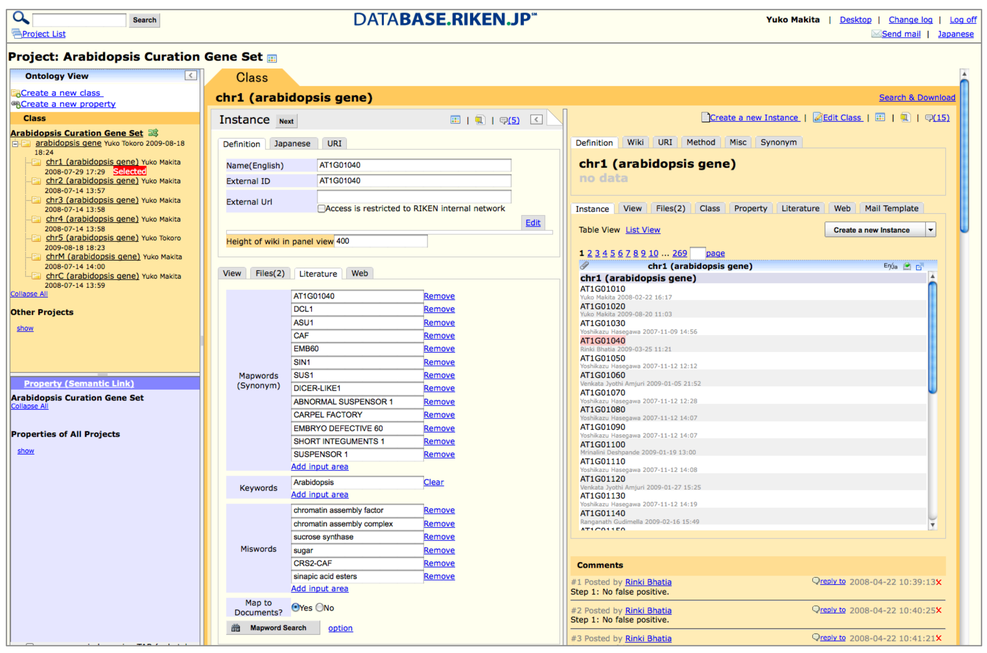
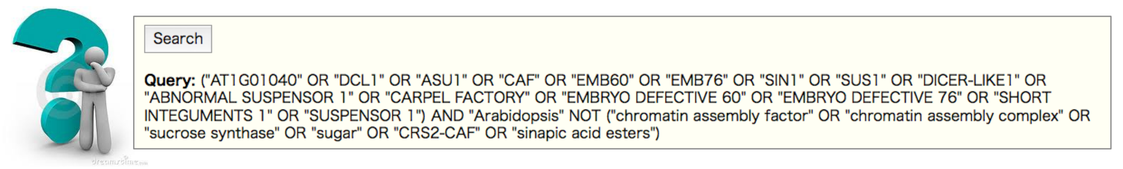


Fig. 5 Curation system on SciNetS, RIKEN BioCuration cloud system



Curators consider the best queries to retrieve related publications.



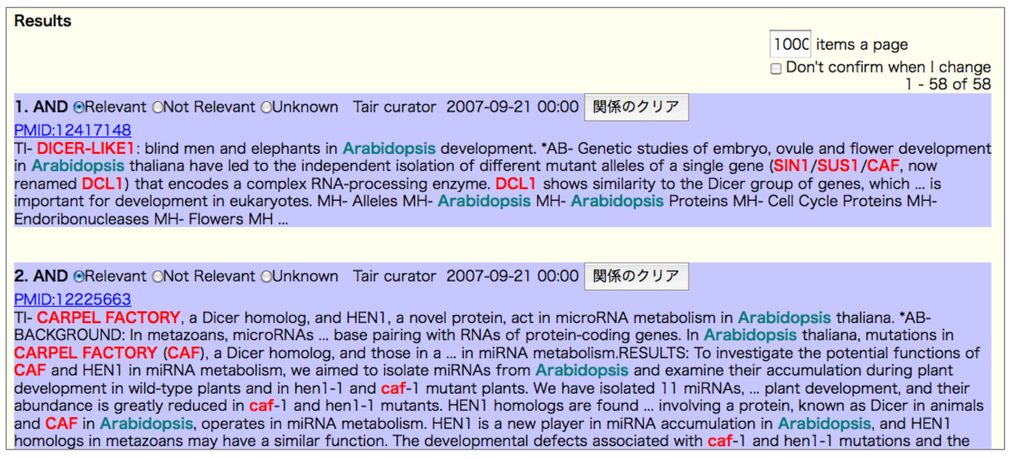
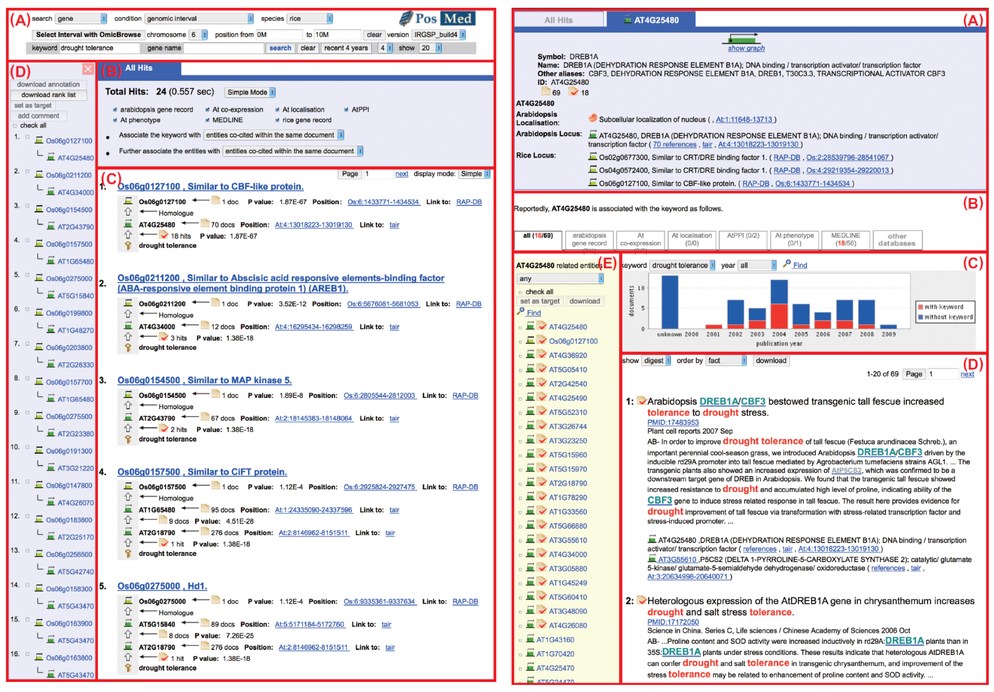


Table 1 Comparison of curation results between PosMed–plus and TAIR.

	# of MEDLINE data	# of gene	# of gene–reference pairs
PosMed	15165	4773	30155
TAIR	11894	11503	38905

Fig. 4 Example search result for *Oriza sativa* genes against the query keyword "drought tolerance.



Reference

Makita Y, Kobayashi N, Mochizuki Y, Yoshida Y, Asano S, Heida N, Deshpande M, Bhatia R, Matsushima A, Ishii M, Kawaguchi S, Iida K, Hanada K, Kuromori T, Seki M, Shinozaki K, Toyoda T. PosMed-plus: an intelligent search engine that inferentially integrates cross-species information resources for molecular breeding of plants. ***Plant Cell Physiol.*** 2009 50(7):1249-59. PMID: 19528193

Yoshida Y, Makita Y, Heida N, Asano S, Matsushima A, Ishii M, Mochizuki Y, Masuya H, Wakana S, Kobayashi N, Toyoda T. PosMed (Positional Medline): prioritizing genes with an artificial neural network comprising medical documents to accelerate positional cloning. ***Nucleic Acids Res.*** 2009 37(Web Server issue):W147-52. PMID: 19468046

PosMed http://omicspace.riken.jp/PosMed