PhosSNP Manual

Genetic polymorphisms that influence protein phosphorylation

Version 1.0
20/08/2009

Author: Yu Xue & Jian Ren

Contact: Dr. Yu Xue, xueyu@ustc.edu.cn; Dr. Jian Ren, renjian@ustc.edu.cn
The database is only free for academic research.
The latest version of PhosSNP database is available from http://phossnp.biocuckoo.org

Index

Statement

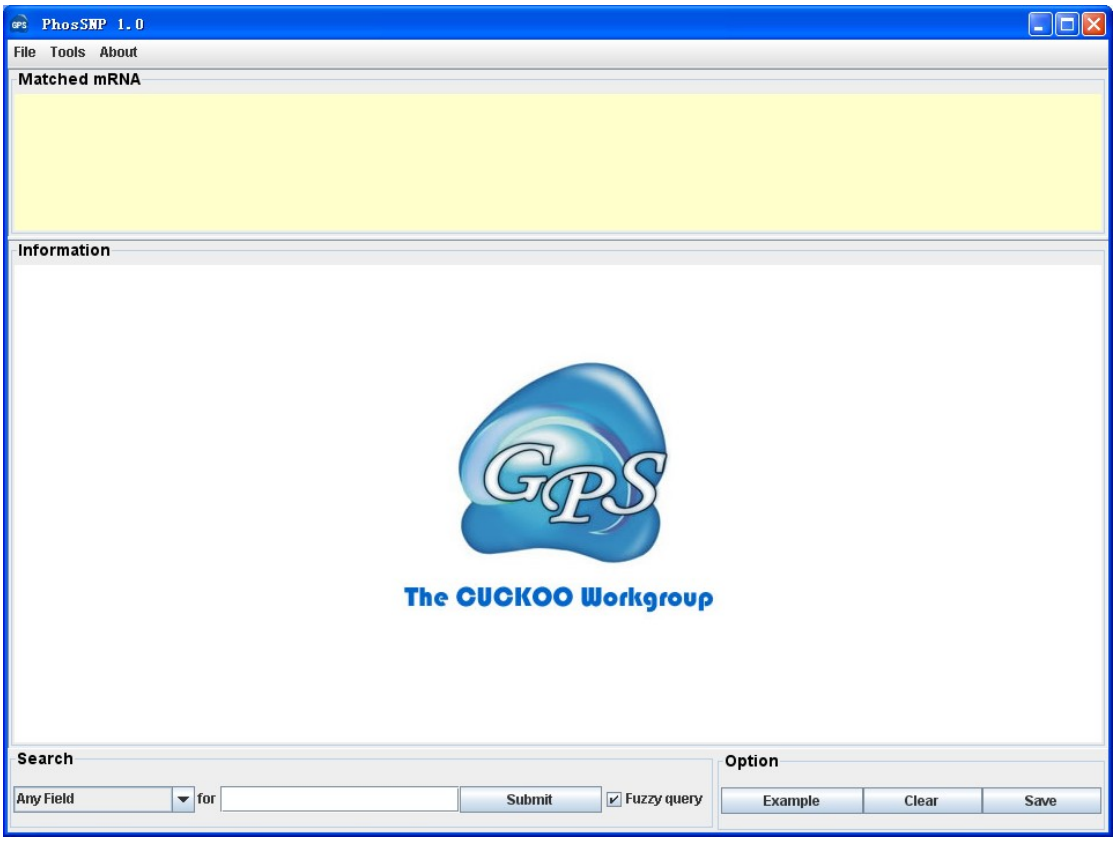
1. **Implementation**. The softwares/databases of the CUCKOO Workgroup are implemented in JAVA (J2SE). Usually, both of online service and local stand-alone packages will be provided.
2. **Availability**. Our softwares/databases are freely available for academic researches. For non-profit users, you can copy, distribute and use the softwares for your scientific studies. Our softwares are not free for commercial usage.
3. **GPS**. Previously, we used the GPS to denote our Group-based Phosphorylation Scoring algorithm. Currently, we are developing an integrated computational platform for post-translational modifications (PTMs) of proteins. We re-denote the GPS as Group-based Prediction Systems. This software/database is an indispensable part of GPS.
4. **Usage**. Our softwares/databases are designed in an easy-to-use manner. Also, we invite you to read the manual before using the softwares.
5. **Updation**. Our softwares/databases will be updated routinely based on users' suggestions and advices. Thus, your feedback is greatly important for our future updation. Please do not hesitate to contact with us if you have any concerns.
6. **Citation**. Usually, the latest published articles will be shown on the software/database websites. We wish you could cite the article if the software has been helpful for your work.
7. **Acknowledgements**. The work of CUCKOO Workgroup is supported by grants from the National Basic Research Program (973 project) (2006CB933300, 2007CB947401), National Natural Science Foundation of China (90919001, 30700138, 30830036, 30721002), Chinese Academy of Sciences (KSCX2-YW-R-139), the Cultivation Fund of the Ministry of Education of China (NO706035), and National Science Foundation for Post-doctoral Scientists (20080430100).

Introduction

As we are entering the age of “Personal Genomics” or “Personalized Medicine”, it has been expected that the knowledge of human genetic polymorphisms and variations could provide a foundation for understanding differences in susceptibility to diseases and designing individualized therapeutic treatments[1,2]. Recent progresses of the International HapMap Project and similar projects[3,4] have provided a wealth of information detailing tens of millions human genetic variations between individuals, including copy number variations (CNVs)[5] and single nucleotide polymorphisms (SNPs)[6]. It was estimated that ~90% of human genetic variations are due to SNPs[2]. In particular, by changing amino acids in proteins, non-synonymous SNPs (nsSNPs) in the gene coding regions could account for nearly half of the known genetic variations linked to human inherited diseases[7]. In this regard, numerous efforts have been contributed to elucidate how nsSNPs generate deleterious effects on the stability and function of proteins. Obviously, an nsSNP might change the physicochemical property of a wild-type amino acid to affect the protein stability and dynamics, or disrupt the interacting interface that prohibits the protein to form a complex with its partners[8-11]. Alternatively, nsSNPs could also influence post-translational modifications (PTMs) of proteins (eg., phosphorylation), by changing the residue types of the target sites or key flanking amino acids[12-14].

In this work, we performed a genome-wide analysis of genetic polymorphisms that influence protein phosphorylation in *H. Sapiens*. We collected 91,797 nsSNPs from NCBI dbSNP build 130[15]. The human mRNA/protein sequences were taken from RefSeq build 31[16]. We used our GPS 2.0 software[17] to predict kinase-specific phosphorylation sites for human proteins and nsSNP data. Here, we defined a phosSNP (Phosphorylation-related SNP) as an nsSNP that might influence protein phosphorylation status. We classified all phosSNPs into five groups. The first three types (I, II, and III) were similarly defined as previously described[14], including change of an amino acid with S/T/Y residue or *vice versa* to create a new [Type I (+)] or remove an original phosphorylation site [Type I (-)], variations to add [Type II (+)] or remove adjacent phosphorylation sites [Type II (-)], and mutations to change PK types of adjacent phosphorylation sites (Type III)[14]. Also, we observed that an amino acid substitution among S, T or Y could also change the PK types in the phosphorylated position (Type IV), say, the target site could still be phosphorylated but by a different type of kinase. Moreover, we defined the type V phosSNP as a variation that results in a stop codon, which might remove its following phosphorylation sites in the protein C-terminus. Unexpectedly, we computationally detected 69.76% of nsSNPs as potential phosSNPs (64, 035) in 17, 614 proteins. In this regard, we proposed that most of nsSNPs might affect protein phosphorylation and play ubiquitous roles in rewiring the biological pathways. More interestingly, we observed 74.58% of phosSNPs as type III phosSNPs (47, 760), which might suggest that nsSNPs prefer to alter PK types of flanking phosphorylation sites rather than

creating or removing phosphorylation sites. Taken together, we proposed that our results could be a useful resource for future disease diagnostics and provide basis for better and individualized. Finally, all phosSNPs data were integrated into PhosSNP 1.0 database, which was implemented in JAVA 1.5 (J2SE 5.0). The PhosSNP 1.0 supports Windows, Unix/Linux and Mac and is freely available for academic researches at: http://phossnp.biocuckoo.org/.

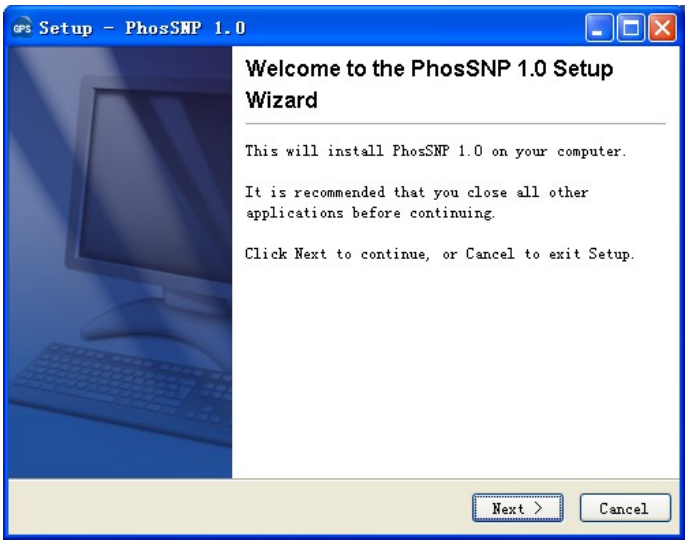


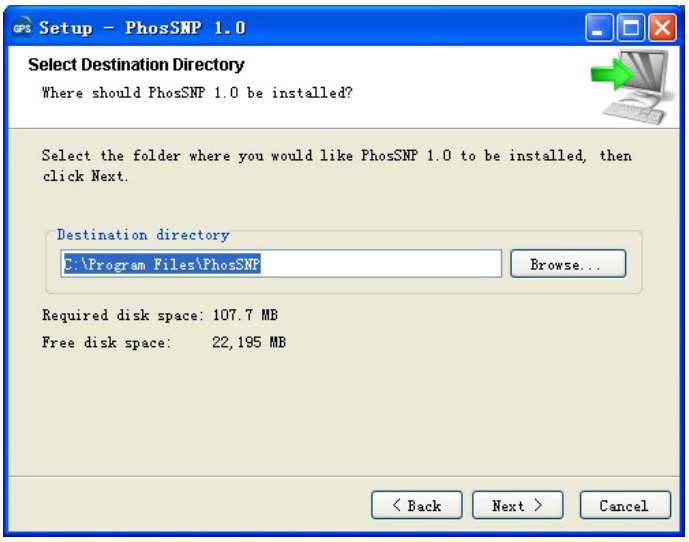
PhosSNP v1.0 User Interface

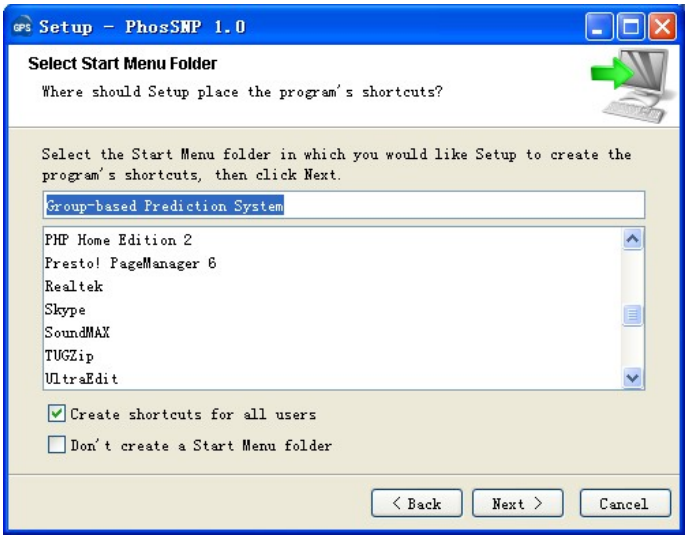
Download & Installation

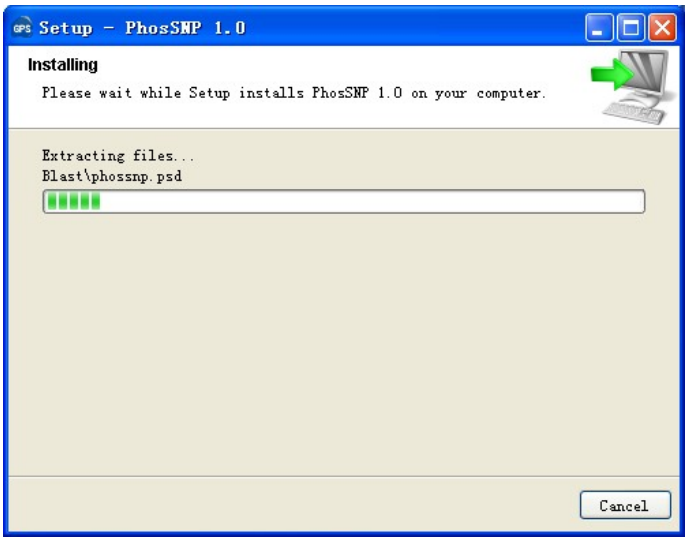
The local packages of PhosSNP 1.0 database were implemented in JAVA, and could be installed on Windows, Mac OS X or Linux systems. The latest distributions of PhosSNP database could be found at http://phossnp.biocuckoo.org/down.php. We recommend that users could download the latest release.

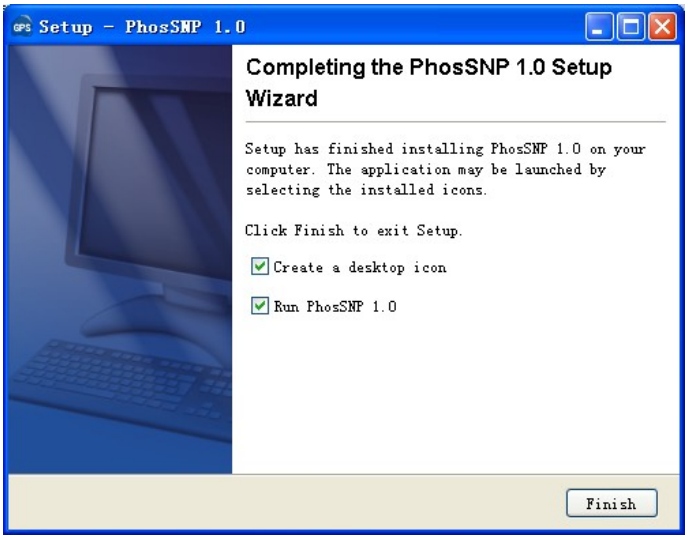
After downloading, please double-click on the install package to begin installation. Follow the user prompts through the installation. And snapshots of the setup program are shown below:









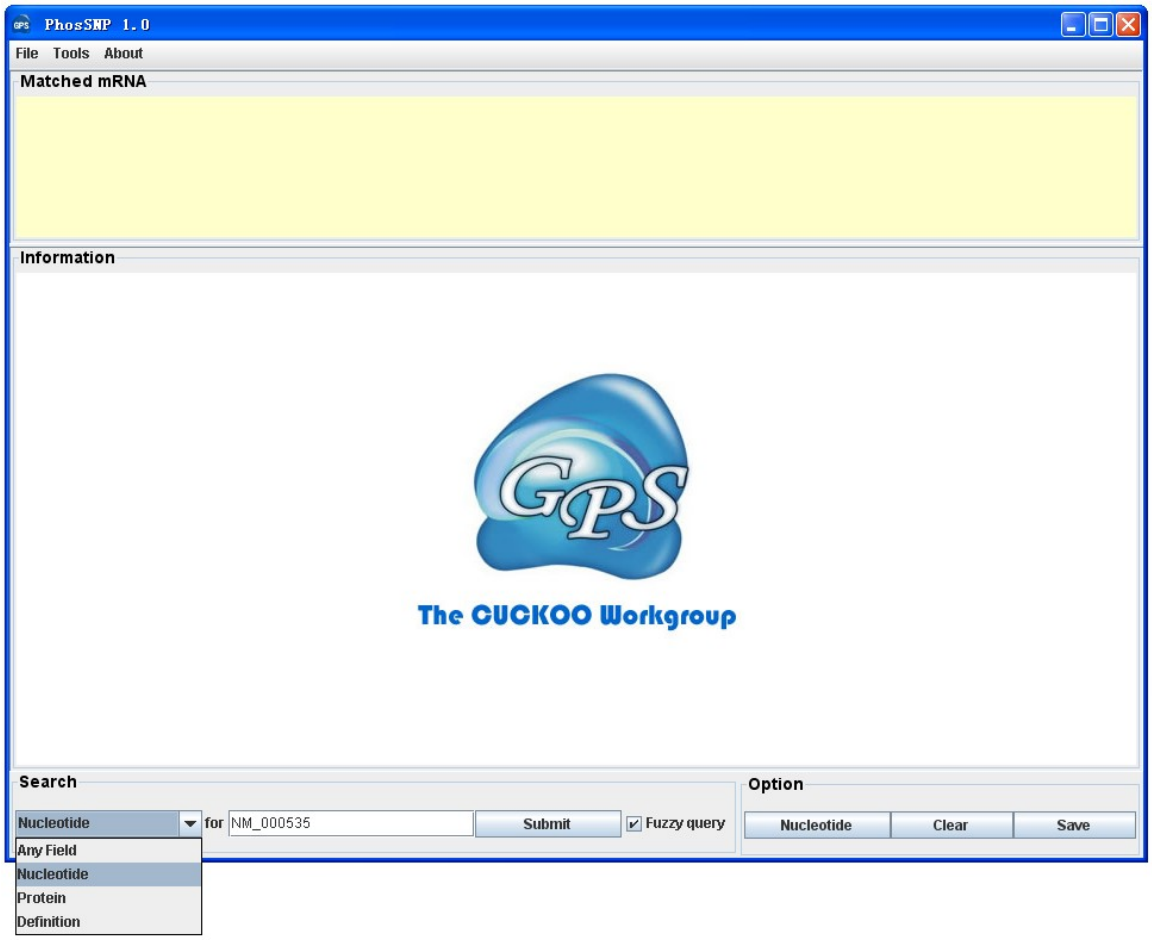


Click on the **Finish** button to complete the setup program.

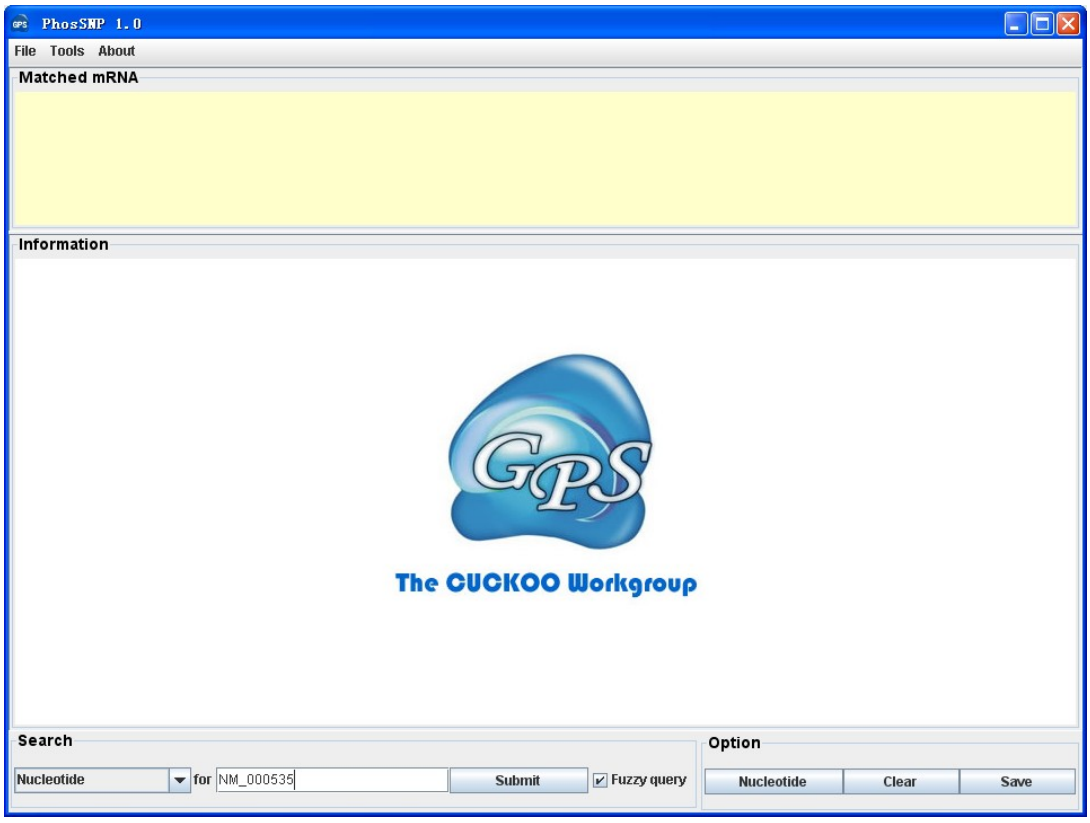
The usage of PhosSNP database

Simple search

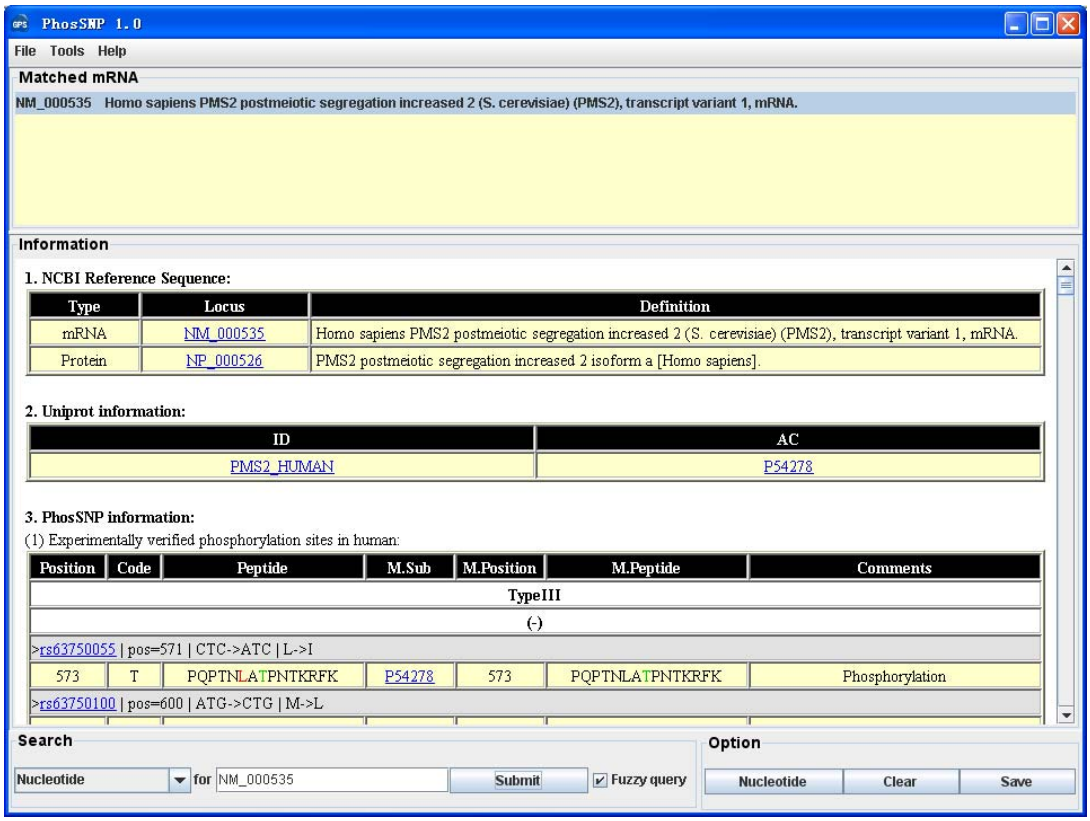
The PhosSNP database was designed in an easy-to-use manner. For simple search, users could input a RefSeq ID with NM_XX (mRNA ID) or NP_XX (Protein ID), and/or definition of the gene (gene or protein name).



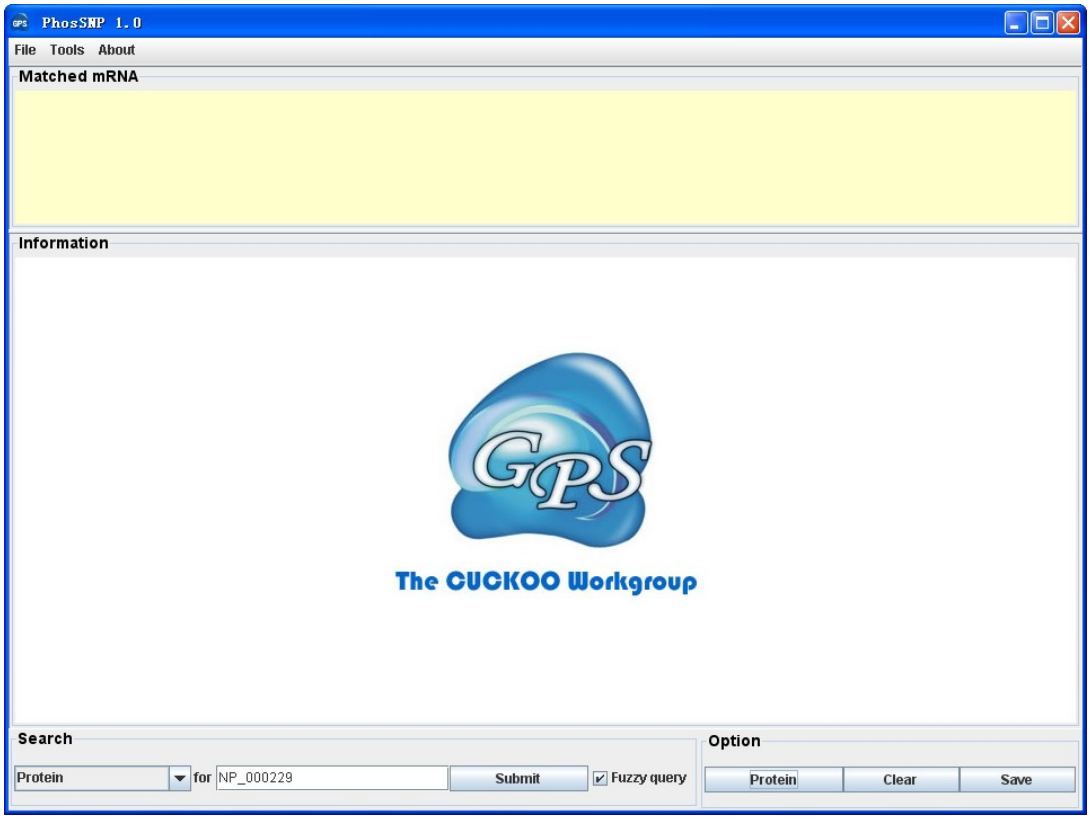
For example, users could input an mRNA ID, eg. NM_000535, specify the “Nucleotide”, then click on the “Submit” button to search the phosSNPs information for this entry.



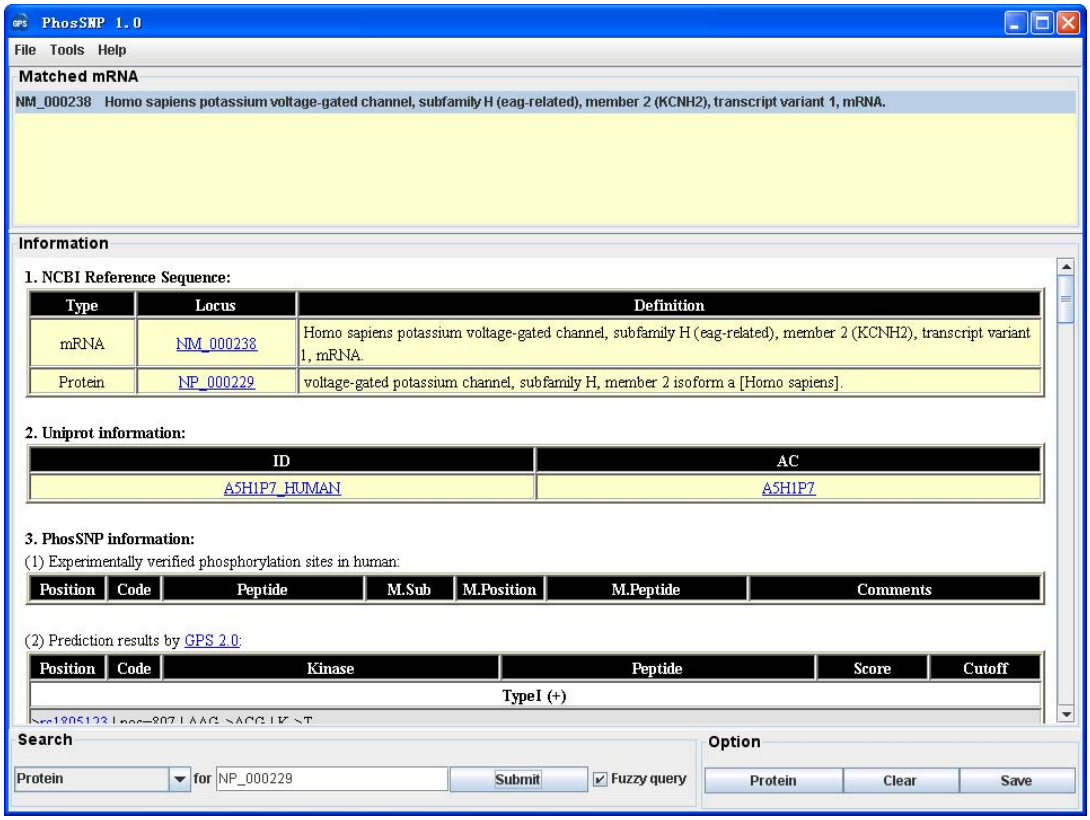
Then the phosSNPs information for NM_000535 will be shown in the “Information” form.



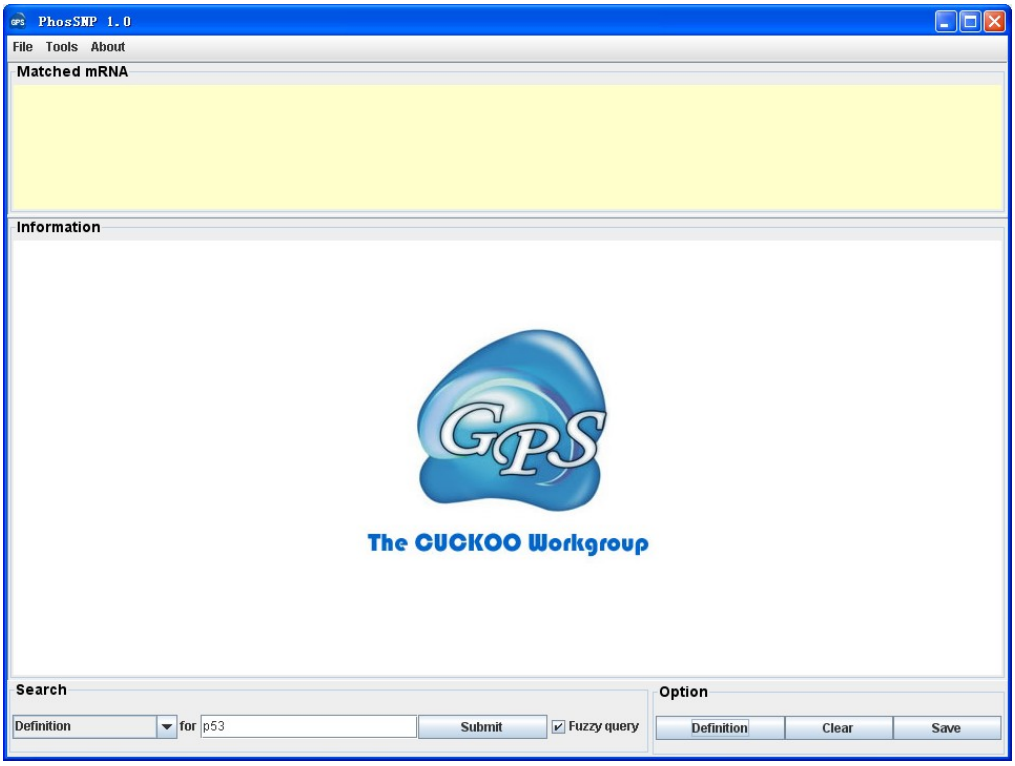
Also, users could input a protein ID, eg. NP_000229, specify the “Protein”, then click on the “Submit” button to search the phosSNPs information for this entry.



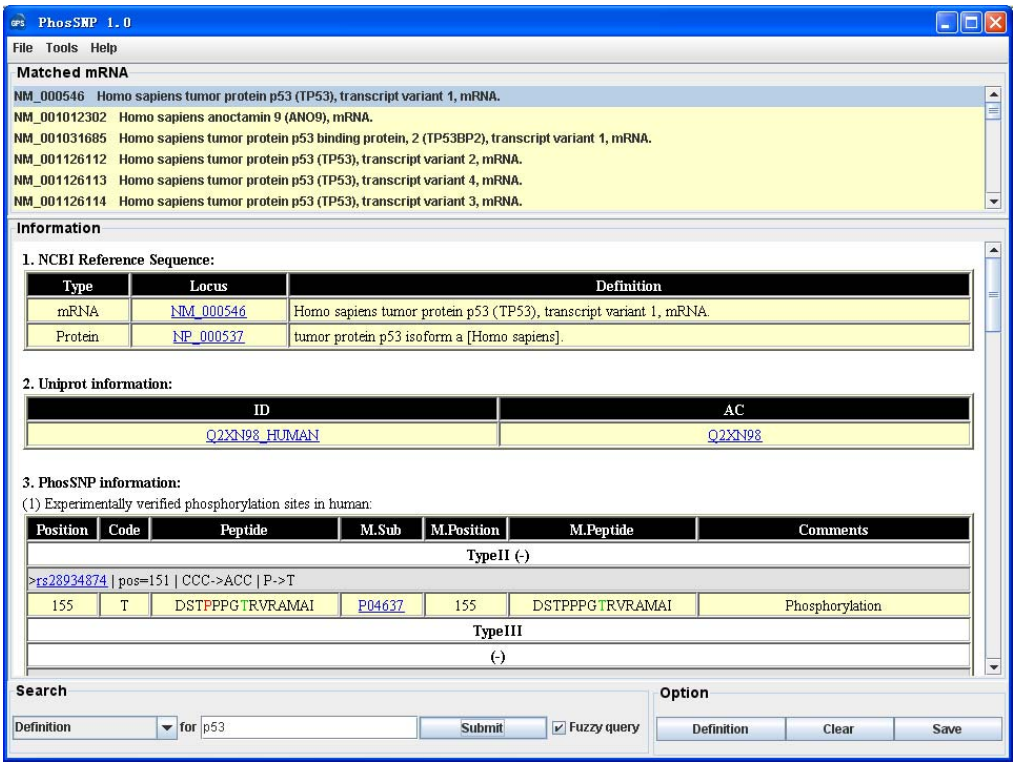
Then the phosSNPs information for NP_000229 will be shown.



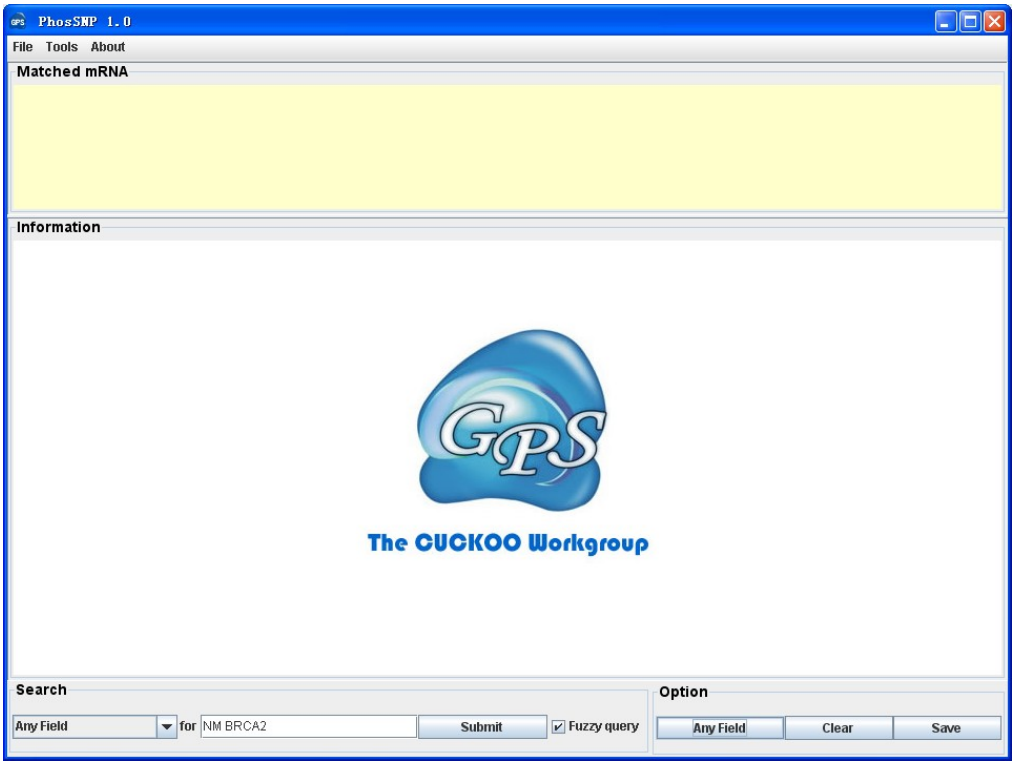
Moreover, users could input a definition of the gene or gene/protein name, eg. p53, specify the “Definition”, then click on the “Submit” button to search the phosSNPs information for p53.



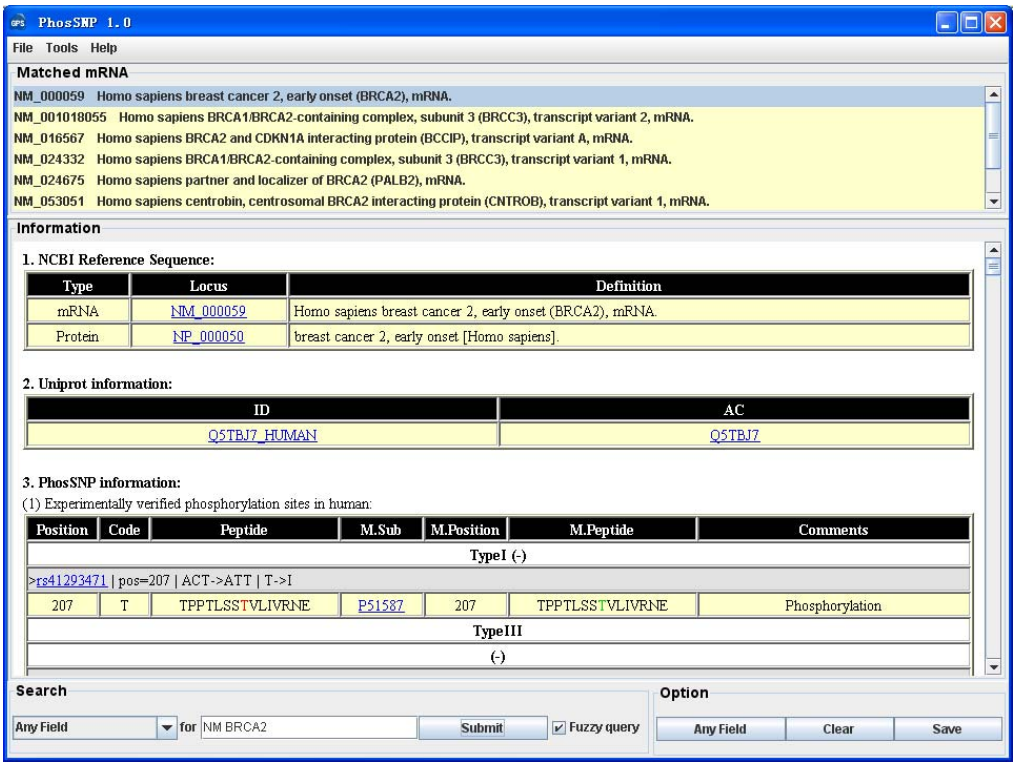
Then the phosSNPs information for p53 or related genes/proteins will be shown. Users could visualize p53 or other related genes/protein by click on the entries listed in the “Matched mRNA” form.



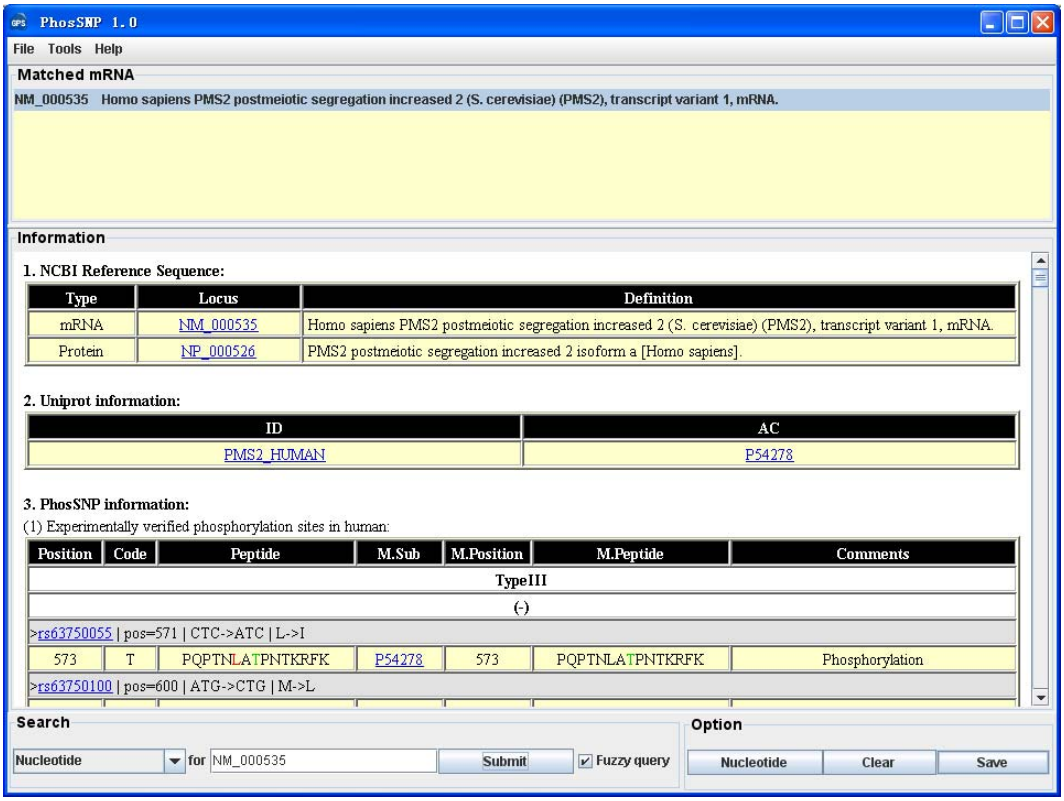
In addition, users could input several keywords together to search phosSNPs information, eg., NM BRCA2 (delimited by space), specify the “Any Field”, then click on the “Submit” button to search the phosSNPs information for BRCA2.



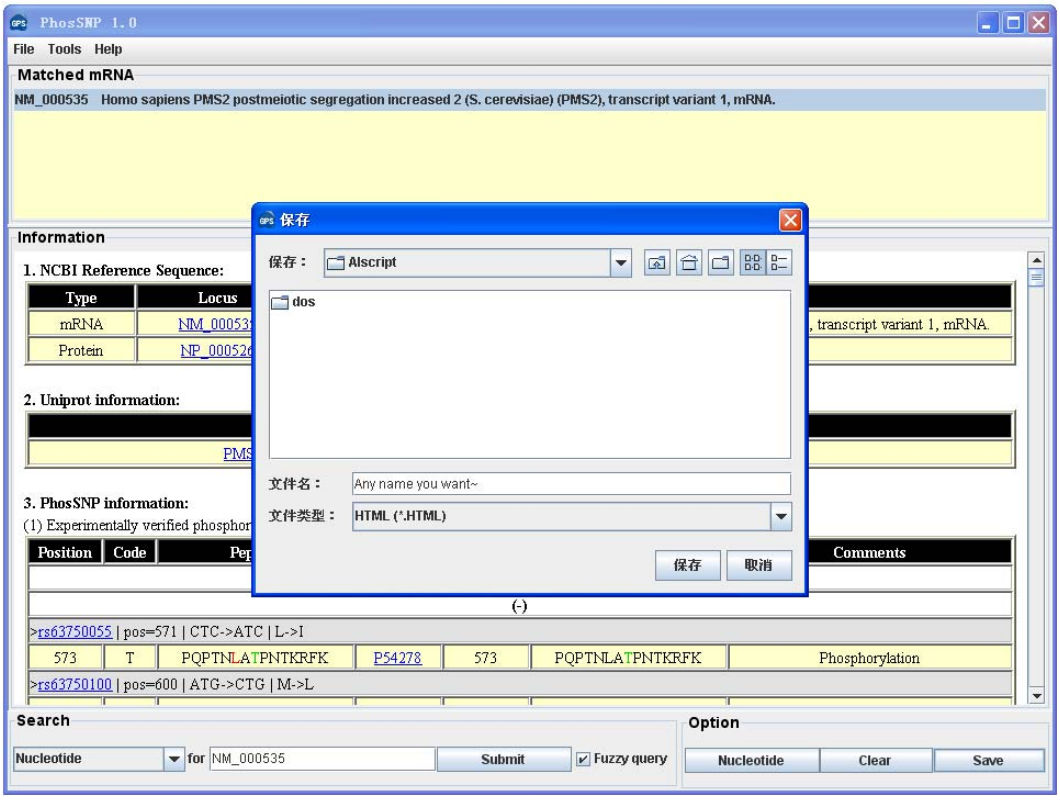
Then the phosSNPs information for BRCA2 or related genes/proteins will be shown. Users could visualize BRCA2 or other related genes/protein by click on the entries listed in the “Matched mRNA” form.



Finally, the searched result could be save in an HTML file by click on the “Save” button in the “Option” form.

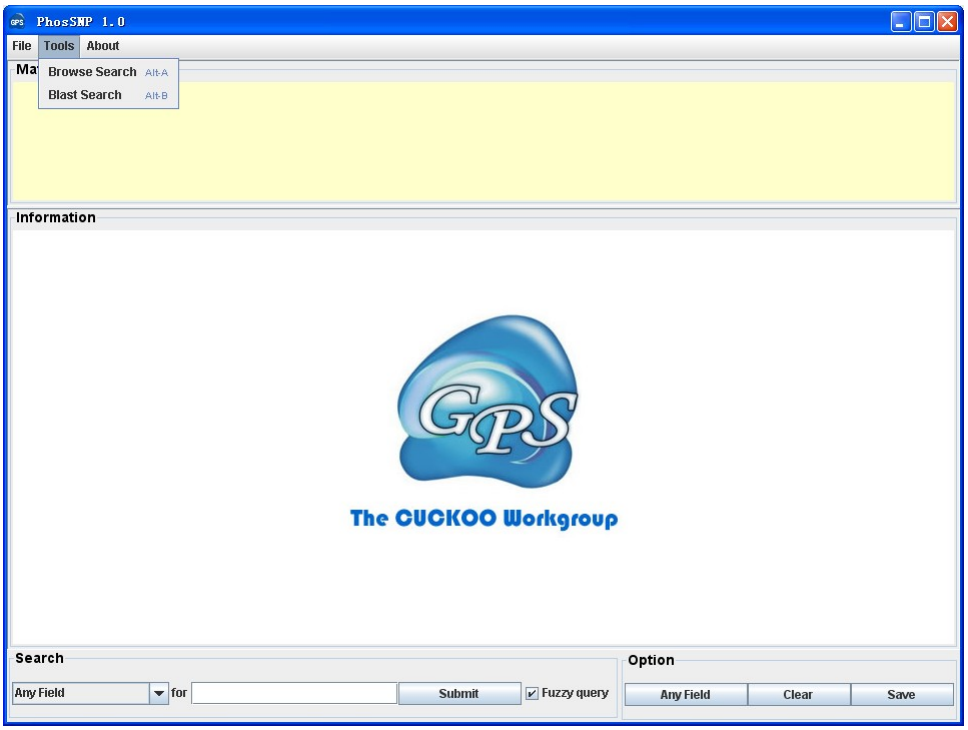


Users could save the results with any name.

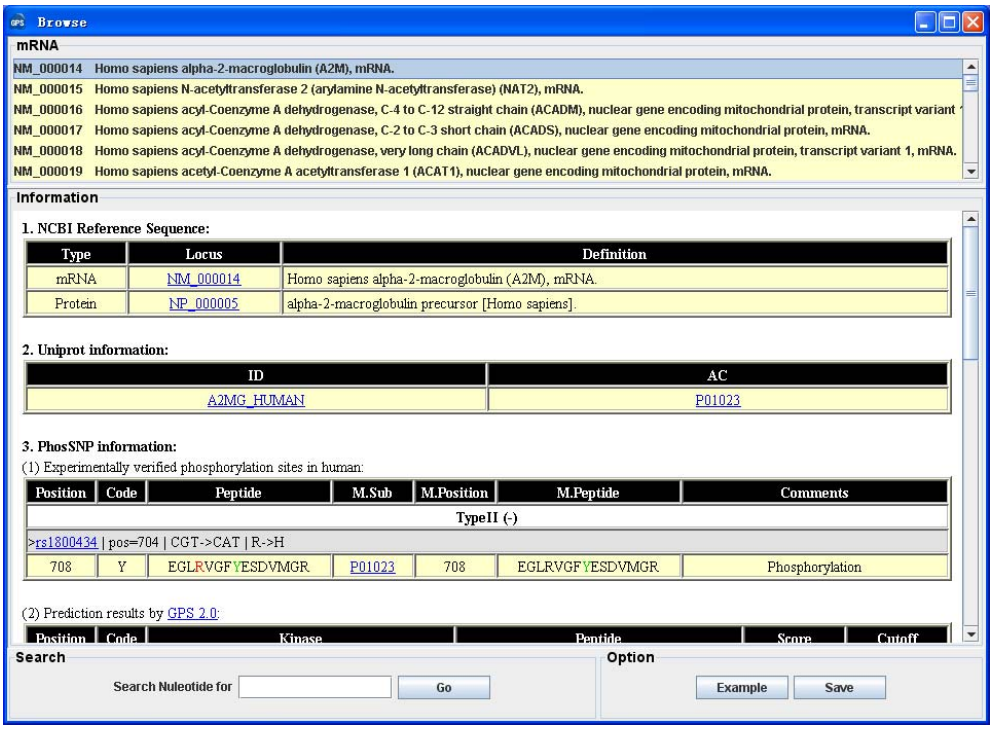


Browse all phosSNPs

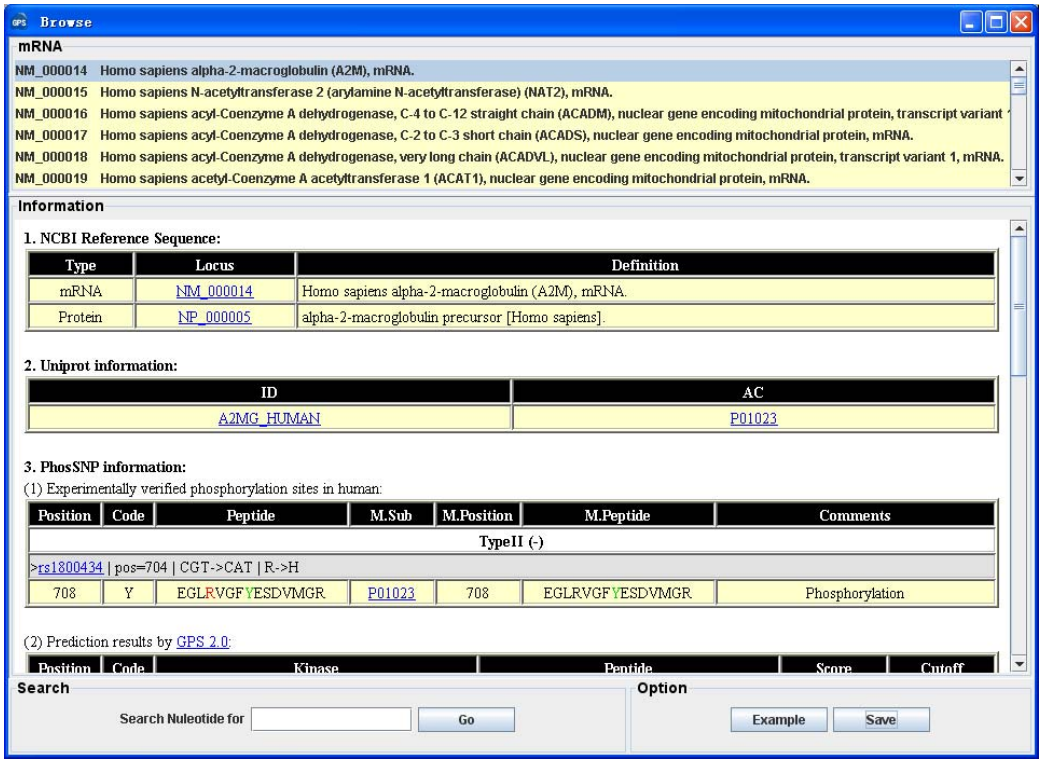
The PhosSNP database supports the browse function. The Browse search allows users to view all entries in PhosSNP database.



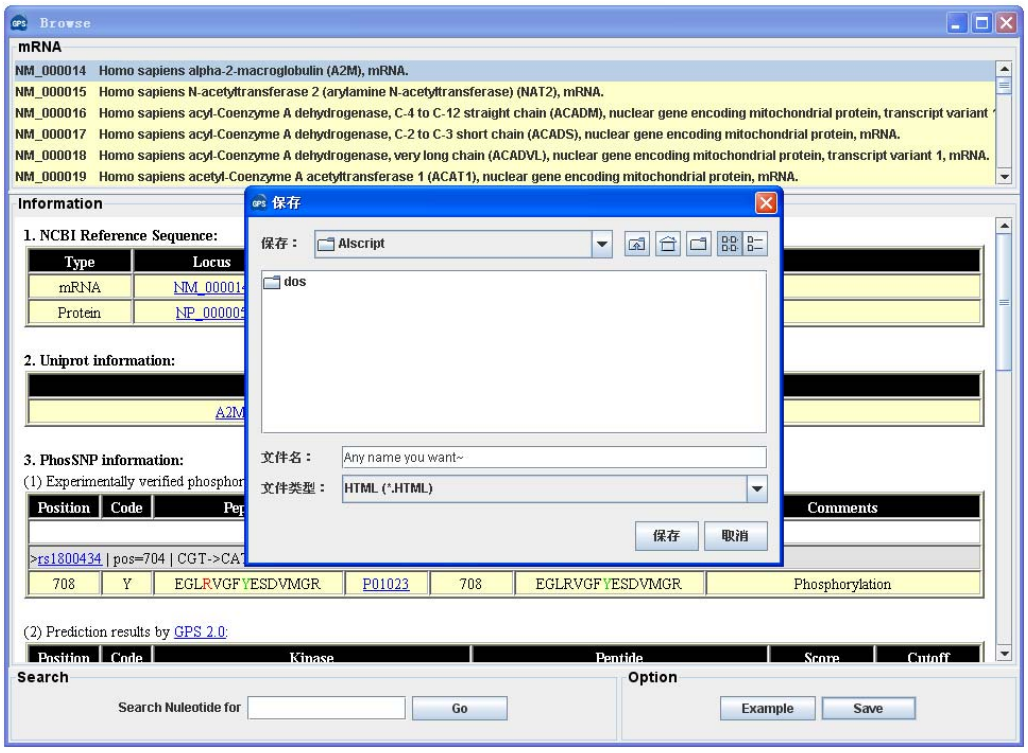
First, users could click on the “Tools” button then click on the “Browse Search” button to visualize all phosSNPs contained genes. Users could visualize any genes/protein by click on the entries listed in the “Matched mRNA” form.



Again, any browsed results could be save in an HTML file by click on the “Save” button in the “Option” form.

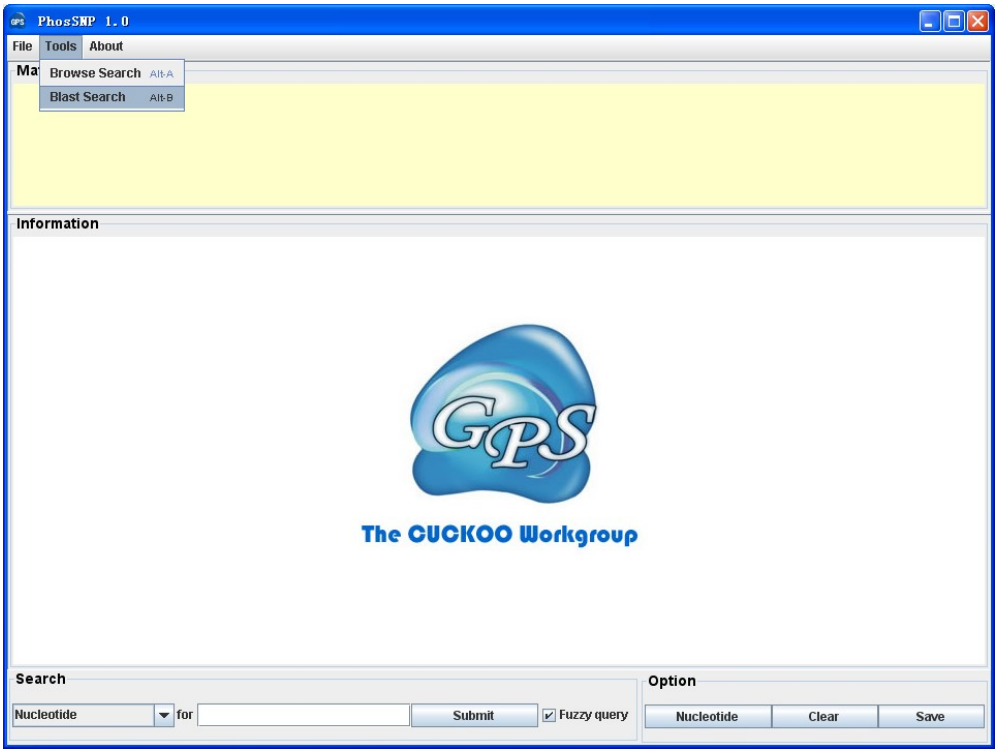


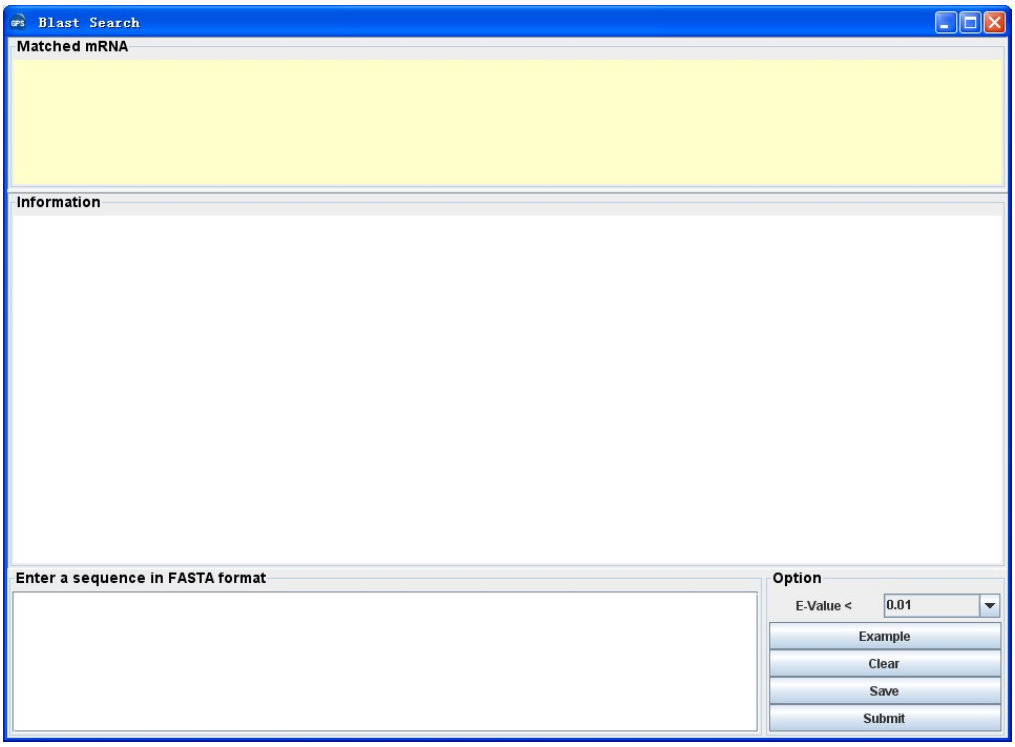
Users could save the results with any name.



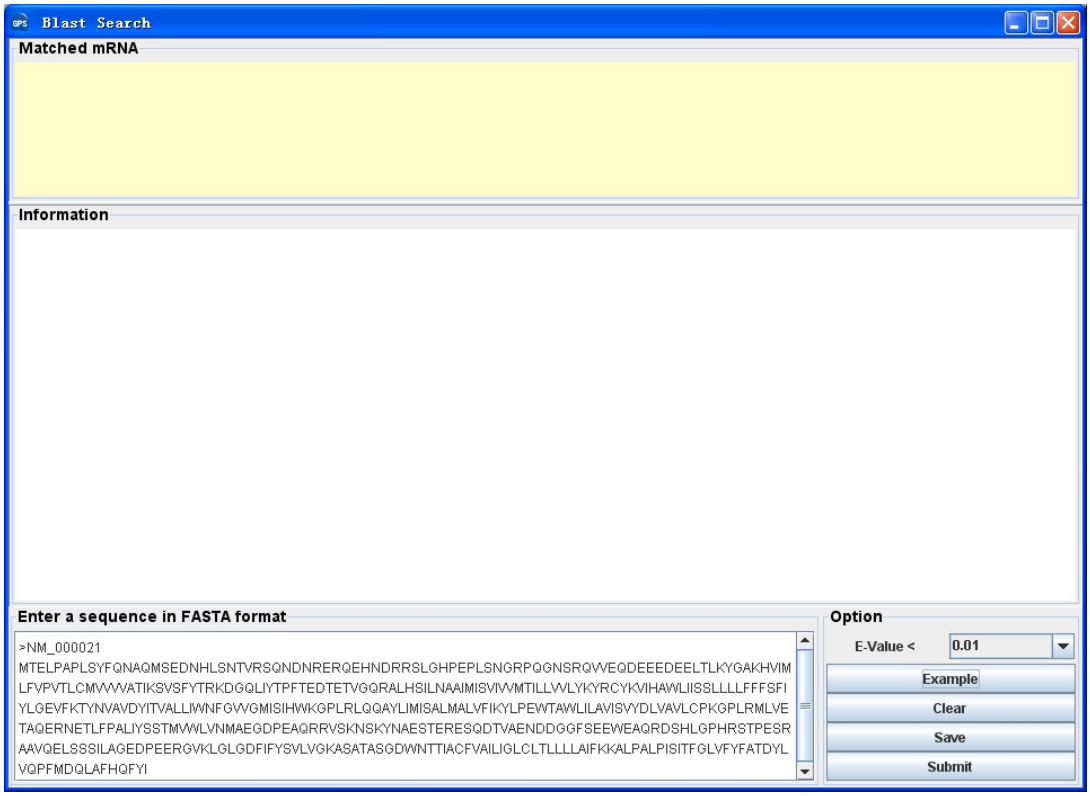
Blast search by sequence alignment

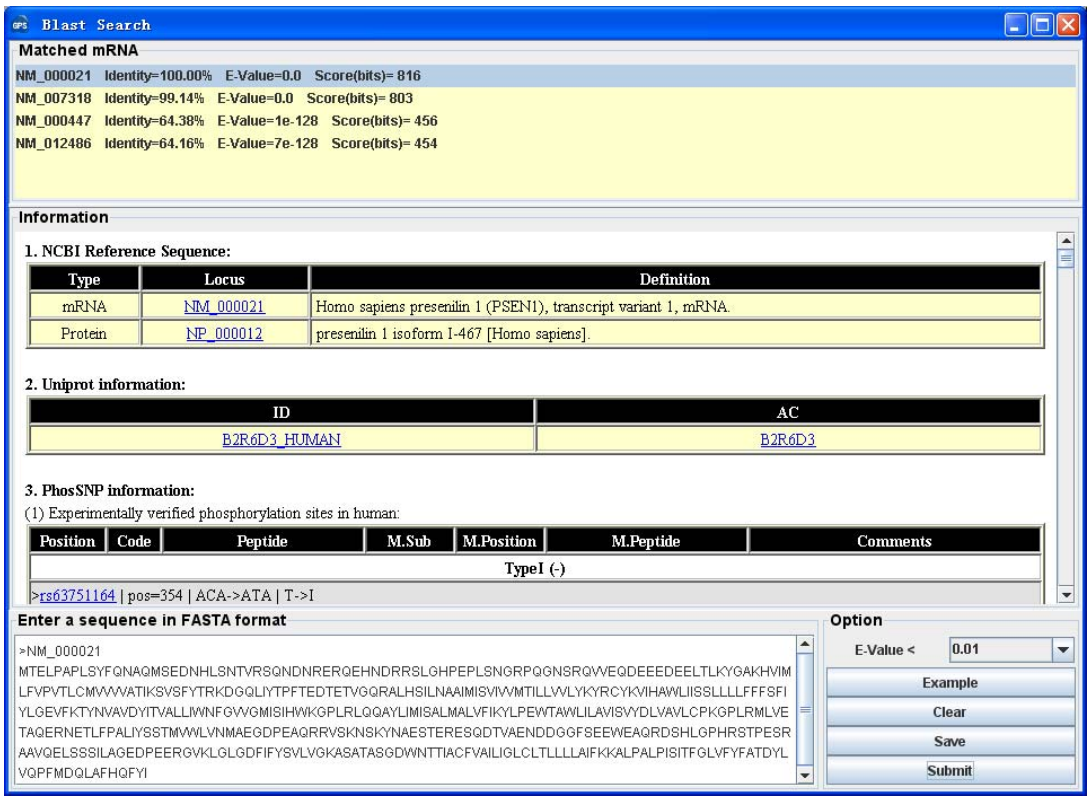
The PhosSNP database supports the searching function by sequence alignment. The blastall program from NCBI BLAST packages was included in PhosSNP database. Users could input one protein (not mRNA sequence) in FASTA or RAW format a time to search identical or homologous entries. First, users could click on the “Tools” button then click on the “Blast Search” button to open a Blast search window.



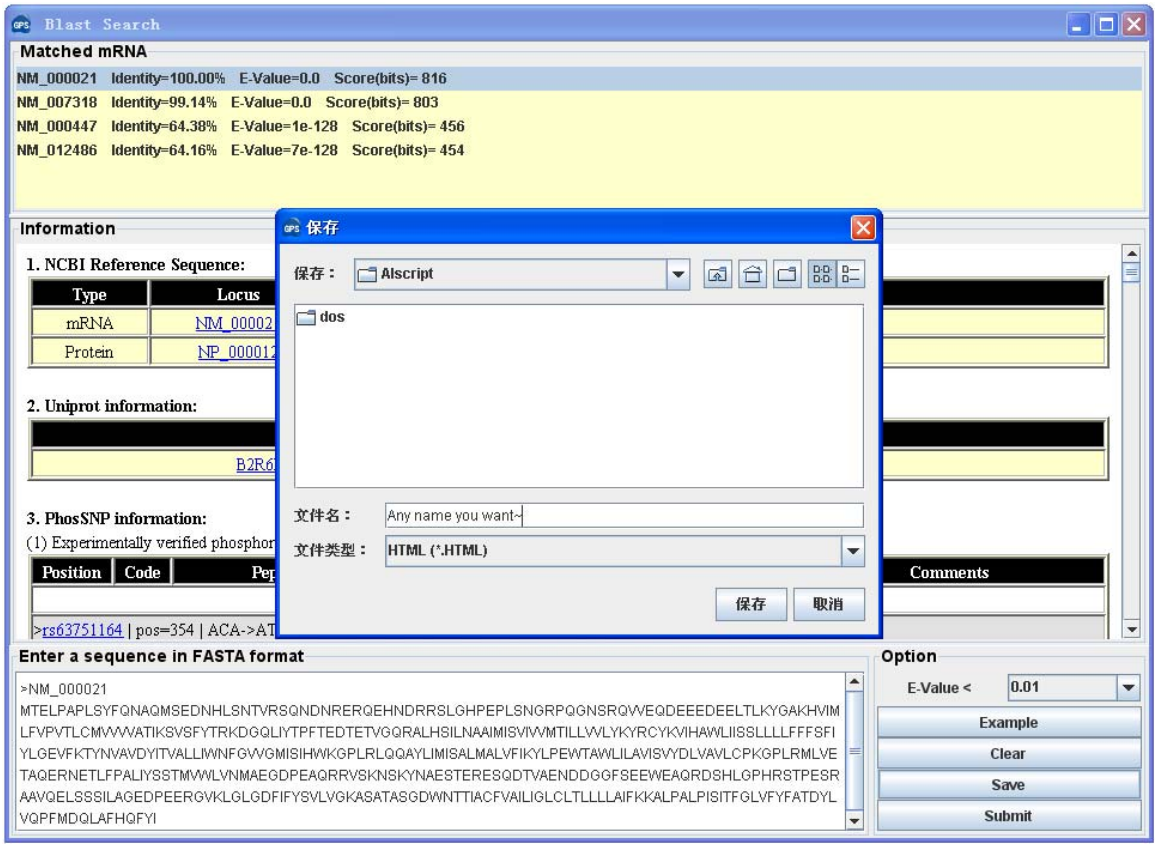


Then users could either click on the “Example” button in the Option form or directly input a protein sequence in FASTA or RAW format. Please note that only one protein is permitted a time. Then please click on the “Submit” button to search identical or homologous entries. The E-value cut-off could be user-defined in the Option form.





Again, users could visualize any genes/protein in the “matched mRNA” form by click on the entries listed in the “Matched mRNA” form. And the results could be saved by clicking on the “Save” button in the Option form.



References

1. M. Cargill, D. Altshuler, J. Ireland et al., *Nat Genet* **22** (3), 231 (1999).
2. F. S. Collins, L. D. Brooks, and A. Chakravarti, *Genome Res* **8** (12), 1229 (1998).
3. K. A. Frazer, D. G. Ballinger, D. R. Cox et al., *Nature* **449** (7164), 851 (2007).
4. *Nature* **437** (7063), 1299 (2005).
5. R. Redon, S. Ishikawa, K. R. Fitch et al., *Nature* **444** (7118), 444 (2006).
6. D. Hinds, L. Stuve, G. Nilsen et al., *Science (New York, NY)* **307** (5712), 1072 (2005).
7. P. D. Stenson, E. V. Ball, M. Mort et al., *Hum Mutat* **21** (6), 577 (2003).
8. P. Yue and J. Moult, *J Mol Biol* **356** (5), 1263 (2006).
9. N. O. Stitziel, T. A. Binkowski, Y. Y. Tseng et al., *Nucleic Acids Res* **32** (Database issue), D520 (2004).
10. A. Uzun, C. M. Leslin, A. Abyzov et al., *Nucleic Acids Res* **35** (Web Server issue), W384 (2007).
11. H. Kono, T. Yuasa, S. Nishiue et al., *Nucleic Acids Res* **36** (Database issue), D409 (2008).
12. S. Savas and H. Ozcelik, *BMC Cancer* **5**, 107 (2005).
13. C. Y. Yang, C. H. Chang, Y. L. Yu et al., *Bioinformatics* **24** (16), i14 (2008).
14. G. M. Ryu, P. Song, K. W. Kim et al., *Nucleic Acids Res* **37** (4), 1297 (2009).
15. S. T. Sherry, M. H. Ward, M. Kholodov et al., *Nucleic Acids Res* **29** (1), 308 (2001).
16. K. D. Pruitt, T. Tatusova, and D. R. Maglott, *Nucleic Acids Res* **35** (Database issue), D61 (2007).
17. Y. Xue, J. Ren, X. Gao et al., *Mol Cell Proteomics* **7** (9), 1598 (2008).
18. Y. Xue, F. Zhou, M. Zhu et al., *Nucleic Acids Res* **33** (Web Server issue), W184 (2005).
19. F. F. Zhou, Y. Xue, G. L. Chen et al., *Biochem Biophys Res Commun* **325** (4), 1443 (2004).
20. Y. Xue, A. Li, L. Wang et al., *BMC Bioinformatics* **7**, 163 (2006).
21. N. Blom, T. Sicheritz-Ponten, R. Gupta et al., *Proteomics* **4** (6), 1633 (2004).
22. J. C. Obenauer, L. C. Cantley, and M. B. Yaffe, *Nucleic Acids Res* **31** (13), 3635 (2003).
23. H. D. Huang, T. Y. Lee, S. W. Tzeng et al., *Nucleic Acids Res* **33** (Web Server issue), W226 (2005).
24. J. H. Kim, J. Lee, B. Oh et al., *Bioinformatics* **20** (17), 3179 (2004).
25. R. I. Brinkworth, R. A. Breinl, and B. Kobe, *Proc Natl Acad Sci U S A* **100** (1), 74 (2003).
26. T. Li, F. Li, and X. Zhang, *Proteins* (2007).
27. G. Neuberger, G. Schneider, and F. Eisenhaber, *Biology direct* **2**, 1 (2007).
28. Y. H. Wong, T. Y. Lee, H. K. Liang et al., *Nucleic Acids Res* **35** (Web Server issue), W588 (2007).

Release Note

1. Apr. 16th, 2009, the beta version of PhosSNP database was developed.
2. May 5th, 2009, the website of PhosSNP 1.0 database was constructed.
3. Jun. 12, 2009, the beta versions of PhosSNP 1.0 local packages were released.
4. Aug. 20, 2009, the final version of PhosSNP 1.0 was released.