

# PAAS: Fast Retrieval of Plant Amino Acid Substitutions and Their Effects on Comparative Phenomics

Junyang Yue · Wenjie Wang · Rongjun Ban · Yunye Zhu · Jia Liu · Yongsheng Liu

© Springer Science+Business Media New York 2014

**Abstract** PAAS (Plant Amino Acid Substitutions) is a web database that provides a comprehensive collection of single amino acid substitutions and plant phenotype effects. It gives direct access to the query for comparative phenotype among different species. From here, users can obtain detailed annotation of each variants, substitution combination, and corresponding protein effects. Currently, PAAS covers 4,749 manually curated single amino acid substitutions scattered throughout 1,313 proteins from 111 plant species. PAAS should be an efficient resource for researchers to facilitate

gene function understanding, produce rationally designed protein, as well as improve plant desirable traits. The database is implemented in PHP MySQL JavaScript and can be accessed at <http://paas.fruitech.org/> without registration.

**Keywords** Amino acid substitutions · Protein mutation · Plant phenotype · SNP · Database

## Abbreviations

PAAS Plant Amino Acid Substitutions  
SNP Single nucleotide polymorphism

Junyang Yue, Wenjie Wang, and Rongjun Ban contributed equally.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11105-014-0777-2) contains supplementary material, which is available to authorized users.

J. Yue · W. Wang · Y. Zhu · J. Liu · Y. Liu (✉)  
School of Biotechnology and Food Engineering, Hefei University of Technology, Hefei 230009, China  
e-mail: liuyongsheng1122@hfut.edu.cn

J. Yue  
e-mail: aaran.yue@gmail.com

W. Wang  
e-mail: wenjie\_wang@outlook.com

Y. Zhu  
e-mail: zhu-yun-ye@163.com

J. Liu  
e-mail: liujia@hfut.edu.cn

Y. Liu  
Ministry of Education Key Laboratory for Bio-resource and Eco-environment, College of Life Science, State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu 610064, China

R. Ban  
School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China  
e-mail: banrongjun1101@gmail.com

## Introduction

Amino acid substitutions are single residue changes leading to sequence variants of the corresponding encoded protein. Most variants in plants are from natural point mutation, TILLING and EMS mutagenesis, non-synonymous single nucleotide polymorphisms (nsSNPs), and in vitro mutation (Henikoff and Comai 2003). These variants could be neutral (e.g., with little or no appreciable effect), harmless, or valuable (Vert et al. 2005). Understanding the exact relationship between protein variants and phenotypic characteristics is important for both planning targeted mutation experiments in the laboratory and analyzing naturally occurring mutations found in plants. Systematic observation of mutational phenotype is an important technique for elucidating gene/protein functions. Unlike existed databases which are genecentric (O'Malley and Echer 2010) or algorithmic (Wainreb et al. 2011), PAAS is developed with more emphases on phenomics associating with protein properties and residues contribution.

## Implementation

The general process of variants collection, protein annotation, structure model, and data statistics was illustrated (Supplementary Fig. 1). We began by reviewing the UniProtKB/Swiss-Prot database (The UniProt Consortium 2014), obtained lists of candidate proteins with experimental mutations in plants, and performed extensive searches of the literature through exhaustive PubMed and Google Scholar to include all protein variants which occurred in nature or introduced experimentally. Only the single amino acid substitutions with effects on the biological properties were reserved. Following subsequent data incorporating, such as deleting multiple amino acids substitutions and removing the duplicates, we obtained 1,313 proteins with 4,749 variants from 111 organisms. Then, all the experimentally verified protein variants were annotated mainly in two parts: protein annotation and variants effects. After that, the Protein Data Bank (PDB) structures were built based on their original amino acid sequences through homology modeling using the fully automated server, SWISS-MODEL (Kiefer et al. 2009). Finally, the PDB structure of corresponding protein with amino acid substitutions was visualized using Jmol, a web browser applet integrated into our web pages (The OpenScience project 2010). The help information (USAGE) and frequently asked questions (FAQs) are available at <http://paas.fruitech.org/documentation.html/>.

## Database

### 1. Query options

PAAS is developed in an easy-to-use mode, providing five search categories: (i) with gene/protein names, (ii) with Uniprot AC/ID, (iii) with PMID or GenBank identifiers, (iv) with variants phenotype, and (v) with amino acid substitution combinations (Supplementary Fig. 2A).

Query with gene/protein names is automatically normalized with synonyms. The full name and abbreviation are both feasible.

The second and third search options are accession number centric. Users can find the proteins of interest with their Swiss-Prot AC/ID, PMID, as well as GenBank identifiers (e.g., DNA, mRNA, translation, and GI accession).

The fourth axis of query can be searched by a variant phenotype. Plant phenotype is classified according to visual inspection, physiological characterization, and biochemical analysis.

Finally, users can use two drop-down boxes to search amino acid substitution combinations of 20 different amino acid residues, with the left box indicating original residue and the right one showing mutational residue. This group could be

more useful when planning targeted mutation experiments in the laboratory.

### 2. Result pages

The search results are shown in a tabular format, containing PAAS accession, protein names, UniProt AC, original residue (Ori-Res), mutational residue (Mut-Res), molecular function, and biological phenotype (Supplementary Fig. 2B). From that table, users can browse the detailed annotation for variants (e.g., molecular function, biological phenotype, SNAP, WESA, residue size, polarity, and residue conservation) and access known database's interpretation of corresponding proteins [e.g., protein sequence, theoretical PI, molecular weight (MW), PIRSF, Pfam domain, SUPFAM, Prosite motif, and gene ontology annotation] (Supplementary Fig. 2C and 2D). In addition, biological function, tissue localization, cell localization, development stage, figures, and literature are also added through manual curation. It is hoped that these information will effectively aid the users exploring the relationship between plant genotype and phenotype.

## Discussion

Upon the evolutionary processes, spontaneous mutations were the main way to give rise to the great varieties in plants. A vital part is due to amino acid substitutions, which provides a rich resource in plant breeding and gene function exploration (Jia et al. 2013). But this pattern is difficult and passive. With the development of molecular techniques and transgenic breeding, introduced mutations have been frequently used to study the phenotypic effects of corresponding protein variants as well as to produce plant cultivars with desired traits (Zhang et al. 2013). Furthermore, the contribution of given amino acid to protein stability and interaction could be accessed. These increasingly detailed descriptions will help users to better utilization of functional protein in agriculture, bioenergy, and environment research.

However, current classification of plant phenotype in different databases is using slightly different terminology, which goes against comparative phenomics. In order to enhance the query power for plant phenotype across species, the uniform terminology created through the plant ontology was applied (Cooper et al. 2013). PAAS will be regularly updated as new plant species are developed in PO database. Finally, it is planned that multiple amino acid substitutions will be incorporated to investigate the cooperative functional effects of variants not entirely based on single amino acid substitutions. Comprehensive research with multiple amino acid substitutions should have more advantage than separate study on protein variants function, as it takes into consideration the interaction effects between amino acids as well as proteins.

**Acknowledgments** Supported by the National Science and Technology Key Project of China (2011CB100401), the National Natural Science Foundation of China (31171179), Advanced Program of Doctoral Fund of Ministry of Education of China (20110181130009), the National Science Fund for Distinguished Young Scholars (30825030), a Key Project from the Government of Sichuan Province (2013NZ0014), and a Project from the Government of Anhui Province (2012AKKG0739).

## References

- Cooper L, Walls RL, Elser J, Gandolfo MA, Stevenson DW et al (2013) The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol* 54:e1
- Henikoff S, Comai L (2003) Single-nucleotide mutations for plant functional genomics. *Annu Rev Plant Biol* 54:375–401
- Jia G, Huang X, Zhi H, Zhao Y, Zhao Q et al (2013) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat Genet* 45:957–961
- Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL repository and associated resources. *Nucleic Acids Res* 37: D387–D392
- O'Malley R, Echer J (2010) Linking genotype to phenotype using the Arabidopsis unimutant collection. *Plant J* 61:928–940
- The OpenScience Project (2010) Jmol: an open-source Java viewer for chemical structures in 3D. <http://www.jmol.org/>
- The UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 42:D191–D198
- Vert G, Nemhauser J, Geldner N, Hong F, Chory J (2005) Molecular mechanisms of steroid hormone signaling in plants. *Annu Rev Cell Dev Biol* 21:177–201
- Wainreb G, Wolf L, Ashkenazy H, Dehouck Y, Ben-Tal N (2011) Protein stability: a single recorded mutation aids in predicting the effects of other mutations in the same amino acid site. *Bioinformatics* 27: 3286–3292
- Zhang H, Xu C, He Y, Zong J, Yang X, Si H, Sun Z, Hu J, Liang W, Zhang D (2013) Mutation in CSA creates a new photoperiod-sensitive genic male sterile line applicable for hybrid rice seed production. *Proc Natl Acad Sci U S A* 110:76–81