

The Curation and Analysis of Rice Stress-Resistance Genes Based on RiceWiki

Ye Zhang¹, Lingling Chen^{1*}, Zhang Zhang²

¹College of Informatics, Huazhong Agricultural University, Wuhan Hubei

²Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing

Email: lchen@mail.hzau.edu.cn, zhangzhang@big.ac.cn

Received: Aug. 18th, 2015; accepted: Sep. 5th, 2015; published: Sep. 8th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Curation is a frontier in China, the biocuration of all kinds of biological information is the development tendency. In this paper, 500 rice genes of eight kinds stress-resistance were collected from literature: salinity tolerance, drought tolerance, cold tolerance, heat tolerance, dwarf resistance, lodging resistance, bacterial blight resistance and blast resistance. Curating the rice stress-resistance genes based on the platform of RiceWiki and publishing on RiceWiki. The 500 rice stress-resistance genes were further analyzed. First, the protein-protein interactions between stress-resistance rice genes were analyzed which containing 78 protein interactions among 61 proteins. The results of co-expression analysis indicated that 21.61% protein pairs have a tendency to co-expression, which showed the protein encoded by the stress-resistance genes could interact with each other. Finally, the GO enrichment is analyzed by argiGO, the results indicated that the significant level of GO terms related to stress is higher.

Keywords

RiceWiki, Curation, Stress-Resistance

基于RiceWiki的水稻抗逆基因审编与分析

张 焯¹, 陈玲玲^{1*}, 章 张²

¹华中农业大学信息学院, 湖北 武汉

*通讯作者。

²中国科学院北京基因组研究所基因组科学与信息重点实验室, 北京
Email: luchen@mail.hzau.edu.cn, zhangzhang@big.ac.cn

收稿日期: 2015年8月18日; 录用日期: 2015年9月5日; 发布日期: 2015年9月8日

摘要

在中国, 审编是一个新领域, 对各种生物学信息进行审编是今后的发展趋势。本文收集了文献中耐盐、耐旱、耐热、耐寒、抗植株矮小、抗倒伏、抗白叶枯病、抗稻瘟病这8种水稻抗逆基因, 共计500个。以RiceWiki为平台, 对搜集的水稻抗逆基因进行审编, 发布到RiceWiki上, 并对搜集的500个基因进行了分析。首先分析了抗逆基因之间的蛋白互作, 得到了61个蛋白质间的78个蛋白互作。共表达分析表明有明显共表达趋势的蛋白对占21.61%, 进一步表明了这些抗逆基因编码的蛋白可以互作。最后, 利用argiGO对GO富集度进行了分析, 结果表明和压力相关的GO语义的显著性水平较高。

关键词

RiceWiki, 审编, 抗逆性

1. 引言

水稻是人类重要的粮食作物, 也是生物研究的关键模式物种之一, 在比较基因组学中有重要地位[1]。它的产量受到非生物胁迫和生物胁迫等压力胁迫的影响。

非生物胁迫, 如干旱、盐、极端温度、化学有毒物质等, 对农业环境产生了严重威胁[2], 导致水稻产量减少, 生长发育延迟甚至死亡[3]。如全球范围内, 超过 8000 万公顷的灌溉地受到盐害影响[4]; 700 万公顷的水稻受到寒冷的胁迫[5]; 而干旱在水稻萌芽、营养生长、开花等不同的生长阶段影响水稻的发育和粮食产量[6]; 在水稻的生长发育季节, 预计温度每升高 1°C, 水稻的粮食产量就会减少 10% [7]。

生物胁迫由病原体(细菌、真菌、病毒和卵菌等)、食草动物和线虫等生物体引起。白叶枯病、纹枯病和稻瘟病是水稻的三大病害, 致使水稻产量锐减, 植株死亡[8]。由于稻瘟病的侵害, 水稻产量每年损失 10%~30%, 损失的这一部分可满足 6000 万人一年的粮食需求。白叶枯病也称为白叶瘟, 是检疫性病害的一种, 可引起 10%左右的水稻减产, 严重时可达 50%~60%。虽然水稻纹枯病的发病面积大于稻瘟病和白叶枯病, 其发病频率也高于这两者, 但危害程度比较轻, 水稻纹枯病发生毁灭性灾害的情况较少。

水稻基因信息不断增多, 其处理速度远落后于数据增长速度, 两者间差距增大[9] [10], RiceWiki (<http://ricewiki.big.ac.cn>)为解决这个问题提供了平台。以中国科学院北京基因组研究所章张研究员课题组为主, 联合北京理工大学、中国林业科学研究院及华中农业大学合作开发完成了 RiceWiki。RiceWiki 是基于维基百科开发的水稻基因信息的平台, 能够编辑, 内容公开, 是一个公众的水稻基因注释系统。很多数据库需要专业的审编人员进行维护和更新, 这不仅耗时耗力而且速度较慢, RiceWiki 依靠的是群体审编而不仅是少数的专业审编[11], 因而 RiceWiki 旨在综合整个水稻科研工作者的智慧, 共同完善水稻基因的相关信息, 从而构建一部水稻基因的百科全书。

RiceWiki 是对 Bio-Wikis (Biological Wikis) [12]的应用, 也是 IC4R 项目(Information Commons for Rice, <http://www.ic4r.org/>)的一部分。RiceWiki 主要收集了 *indica* 93-11 和 *japonica Nipponbare* 两个水稻栽培亚种基因的数据。迄今为止, RiceWiki 网站上提供了 31,527 篇与水稻基因相关的文献, 列出了 86,216 个水稻基因, 其中 1001 个基因得到了较全面的审编。

目前大部分的数据库都是通过整合资料和审编得到的，审编是英文单词“Curation”的中文译名。审编通常是指通过对生物学数据进行转化和整合，形成数据库，从而保证这个数据库既是文献的集合又是生物数据的集合。标准化、质量化的审编能够增加生物学数据的相互操作性和一致性[13]。其过程主要包括三个步骤，首先对生物学文献和数据进行翻译，然后将生物学相关的信息进行整合，最后将整合后的信息综合成数据库中的信息。审编的主要目标有两个：一是对生物学的知识进行准确而综合的展示，二是方便科研工作者获取相关信息。审编也可以理解为一种转化，将生物学的数据转化成相应的某种格式的数据。生物审编人员处理的数据是多样化的，有核酸序列、相互作用途径和文献等[14]，本文在审编过程中处理的数据主要以文献为主。

2. 材料和方法

2.1. 水稻抗性的选取

水稻的抗性主要体现在对生物胁迫和非生物胁迫的抗性，及形态表型方面的抗性。本文选取了 8 类对水稻产量影响最大的压力胁迫因素：在水稻面临的生物胁迫中选取三大病害中较为严重的白叶枯病和稻瘟病，在非生物胁迫中选取了盐害、低温、干旱、高温这四类，在水稻的生理形态方面选择了倒伏和植株矮小。

2.2. 水稻抗性基因的查找和审编流程

如图 1，从搜集基因到审编的流程分为四部分：1) 水稻抗逆基因的查找；2) 水稻基因名称的转换；3) 相关文献的搜集；4) 对抗逆基因进行审编。

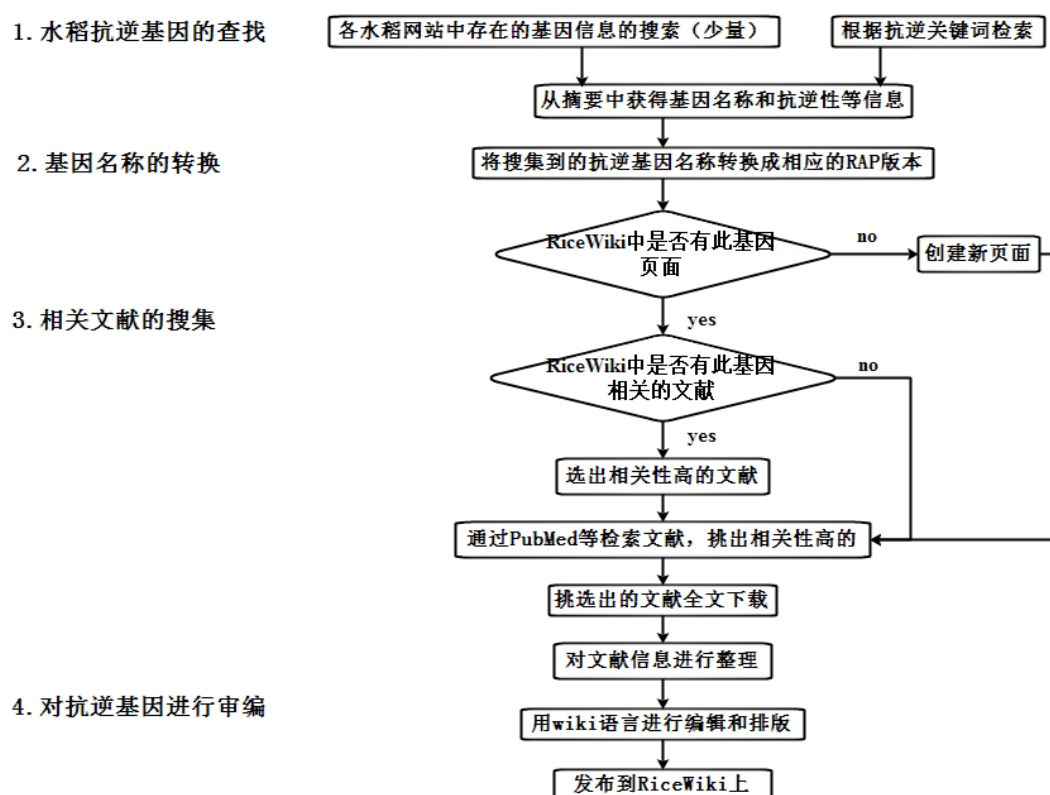


Figure 1. The flow chart of biocuration of rice gene

图 1. 对水稻基因进行生物审编的流程图

抗逆基因的查找主要有两部分：一，从国家水稻数据中心、PLANTSTRESS 等水稻相关网站上查找，通过这种方法得到的基因较少。二，在 PubMed 等网站上根据水稻抗逆性的关键词进行文献的检索。通过阅读摘要获得基因名称和其抗逆性。本研究对不同抗性的水稻基因分别从这两个方面进行收集和整理，并将其发布在 RiceWiki 页面上。

目前水稻的命名规则是以水稻遗传协作组织(CGSNL)的相关规则为基础制定的[15]。随着水稻全基因组测序的发展，出现了多个水稻测序和注释的版本，目前主要有 NCBI、水稻注释计划(RAP) [16]、北京基因组研究所(BGI) [17]、美国基因组研究所(TIGR) [18]等版本，这些版本之间的组装和注释是相互独立的，但彼此有些许差别。因此同一个基因在不同的组装注释版本中有不同的名称[19]。而水稻基因在文献中多以基因符号、基因名称、Genbank 登录号这三种水稻名称出现。RiceWiki 目前以 RAP 版本为主，需要将基因符号和 RAP 版本号相对应，将文献中的 TIGR 版本号或 Genbank 登录号转化成 RAP 版本。

从两方面进行文献搜集：一，RiceWiki 网站上此抗逆基因的相关文献；二，分别以 RAP 号、TIGR 号和基因符号为关键词在 PubMed 和 Textpresso 等网站上检索。

仔细阅读挑选出的高质量的文献，累计超过 1000 篇，筛选基因相关信息。整理文献中获得的抗逆基因的功能、突变等信息。审编后的基因信息经过 Wiki 语言的加工(排版、图片、链接等)，发布在 RiceWiki 上。

2.3. 审编质量评分办法

RiceWiki 中的 AuthorReward 会列出各版本的贡献值、编辑数量和编辑质量等[20]。将任意一个版本与最新版本进行比较，所有编辑过这个页面的审编人员，其相应的贡献值都会被详细的列出，随着审编内容的更新，编辑人员的贡献值也会随之变化[21]，编辑质量的取值范围区间为(-1, 1)。

AuthorReward 的原理为[20]：任意页面 P ，假设此页面有很多编辑版本 v_0, v_1, \dots, v_n ，版本 v_0 为空白页 ($n > 0$)。任意一个编辑成功的版本均视为一个新版本，任意相邻版本 v_{i-1} 和 v_i ($1 \leq i \leq n$) 由不同研究人员完成。版本 v_i 和 v_j 间的差距用 $d(v_i, v_j)$ 表示 ($i < j$)，通过 LD 算法对编辑的相关信息进行统计。版本 v_i 的贡献值为 $CS(v_i)$ 。

$$CS(v_i) = c[d(v_{i-1}, v_n) - d(v_i, v_n)]$$

其中 c 是比例系数， $d(v_{i-1}, v_n)$ 是版本 v_{i-1} 和 v_n 间的编辑差距， $d(v_i, v_n)$ 是版本 v_i 和 v_n 的差距。版本 v_i 的编辑数量(Edit quantity)为 $QTY(v_i)$ ，编辑质量(Edit quality)为 $QAL(v_i)$ 。

$$QTY(v_i) = d(v_{i-1}, v_i)$$

$$QAL(v_i) = \frac{d(v_{i-1}, v_n) - d(v_i, v_n)}{d(v_{i-1}, v_i)}$$

审编人员 r 在网页 p 上的贡献值为 $S(r, p)$ ，在整个 RiceWiki 中的贡献值为 $S(r)$ 。

$$S(r, p) = \sum_{v_i \in E(r, p)} CS(v_i)$$

$$S(r) = \sum_{p \in P} S(r, p)$$

RiceWiki 会给每个审编人员进行系统性的打分，将贡献值和审编质量等信息放在相应的网页上。

3. 结果与分析

3.1. 水稻抗逆基因的查找结果

对 8 类水稻抗性基因进行搜集和整理，如表 1 所示，共计 500 个。其中有一些有多重抗逆性，即有两种或

以上抗逆性。水稻抗逆性和抗性基因在 RiceWiki 网页上列出(<http://RiceWiki.big.ac.cn/index.php/RiceWiki:Stress>)。对这 500 个水稻抗性基因的基因符号、RAP 版本号、TIGR 版本号等基因名称进行了统一。

3.2. 对水稻抗逆基因的审编

在 RiceWiki 上, 每一个基因对应一个 Wiki 页面, 对各水稻抗逆基因的审编结果放在对应的 RiceWiki 的 Wiki 页面上。从已经审编的基因中挑选几个典型的基因作为例子(图 2~图 7), 对审编的内容和结果做以下说明。

以基因 *SNAC1*(<http://RiceWiki.big.ac.cn/index.php/Os03g0815100>)为例, 如图 2, 首先在页面上方用一句话概括, 之后分别对注释信息、研究此基因的实验室、参考文献、结构信息四部分进行审编, 关键词汇通常加粗显示。

3.2.1. 注释信息

这是审编过程中最重要的一部分。本文将其大致分为 8 部分, RiceWiki 用户可根据各基因的研究状

Table 1. The number of rice stress-resistance genes
表 1. 水稻抗性基因数量

水稻抗性	耐盐	耐寒	耐热	耐旱	抗矮小	抗倒伏	抗白叶枯病	抗稻瘟病	多重抗逆性	单一抗逆性	共计
基因数目	97	55	47	128	161	16	66	87	117	383	500

SNAC1 (STRESS-RESPONSIVE NAC 1) is a member of rice stress-responsive NAC family in rice.

Contents [hide]
1 Annotated Information
1.1 Function
1.2 Genes related to <i>SNAC1</i>
1.3 Mutation
1.4 Expression
1.5 Evolution
1.6 Knowledge Extension
2 Labs working on this gene
3 References
4 Structured Information

Figure 2. The illustration of the contents in curation page
(<http://RiceWiki.big.ac.cn/index.php/Os03g0815100>)

图 2. 审编页面的内容图示(<http://RiceWiki.big.ac.cn/index.php/Os03g0815100>)

Function

[edit]

- *OsNHX1*, functions as a Na⁺ and H⁺ exchanger and plays important roles in salt tolerance of rice^{[1][2][3][4][5][6]}. *OsNHX1* may play an important role in the salt tolerance of shoots rather than roots^[6].
- *OsNHX1* plays important roles in the compartment of Na⁺ from the cytoplasm of cells into the vacuoles in both of the tissues^{[2][5]}.
- Expression of *OsNHX1* under the constitutive promoter Actin1D can confer enhanced salinity tolerance in transgenic rice plants. It also function towards alleviation of toxic effects of salt stress at all stages of the life cycle from seedling to maturity^[3].
- *OsNHX1* has the ability to suppress Na⁺, Li⁺ and hygromycin sensitivity of yeast *nhx1* mutants and sensitivity to a high K⁺ concentration, a novel phenotype of the *nhx1* mutants^[6].

GO assignment(s): [GO:0006814](#), [GO:0006885](#), [GO:0015299](#), [GO:0015385](#), [GO:0016021](#)

Figure 3. The illustration of the function and GO
(<http://RiceWiki.big.ac.cn/index.php/Os07g0666900>)

图 3. 功能和 GO 号的图示(<http://RiceWiki.big.ac.cn/index.php/Os07g0666900>)

Mutation

[edit]

OsDEG10 RNAi transgenic plants^[1]:

- OsDEG10 RNAi plants showed **more remarkable decline of Fv/Fm ratio** than wild-type plants after **high light treatment**.
- OsDEG10 RNAi plants had **lower xanthophyll pigment contents** than wild-type plants, which suggest that OsDEG10 RNAi plants are **more sensitive to high light** than wild-type plants **probably due to lower xanthophyll pigment pool size**.
- OsDEG10 RNAi transgenic plants showed **more decline of Fv/Fm ratio** than wild type plants **after cold treatment**, which suggest that OsDEG10 RNAi transgenic plants are **more sensitive to cold stress** and that OsDEG10 might play an important role in the **response to cold stress** as well as **photooxidative stress**.
- Under optimal growing conditions, no phenotypic variations between the OsDEG10 RNAi plants and wild-type plants were observed.
- Overall, OsDEG10 RNAi transgenic plants were **more sensitive to high light and cold stresses** compared to wildtype plants.

Figure 4. The illustration of the mutation (<http://RiceWiki.big.ac.cn/index.php/Os08g0139000>)
图 4. 突变体的图示(<http://RiceWiki.big.ac.cn/index.php/Os08g0139000>)

Genes related to *SNAC1*

- **Target gene of *SNAC1***^[4]:
 - **OsSR01c**, was identified as a direct target gene of **SNAC1** involved in the regulation of stomatal aperture and oxidative response^[4].
 - **SNAC1** could bind to the promoter of *OsSR01c* and **activate** the expression of *OsSR01c*. *OsSR01c* has dual roles in drought and oxidative stress tolerance of rice by promoting stomatal closure and H₂O₂ accumulation through a novel pathway involving regulators *SNAC1* and *DS1*^[4].
- Gene regulated by *SNAC1*:
 - Some Stress Related Genes^[1]:
 - *SNAC1* encodes a NAC transcription factor that regulates many stress related genes, such as *DREB1A*, *CBF*, *SCOF*, *Tsi*, and *OSISAP1*^[1].
 - **OsPP18**
 - *OsPP18* is a novel PP2C gene which is regulated by *SNAC1* and confers drought and oxidative stress tolerance by regulating ROS homeostasis through ABA-independent pathways^[3].
 - *SNAC1* can bind to the promoter of the *OsPP18* gene and **activates** its expression^[3].

Figure 5. The illustration of the relationships between the gene with the others (<http://RiceWiki.big.ac.cn/index.php/Os03g0815100>)
图 5. 水稻基因与其它基因间的关系的图示 (<http://RiceWiki.big.ac.cn/index.php/Os03g0815100>)

Subcellular localization

[edit]

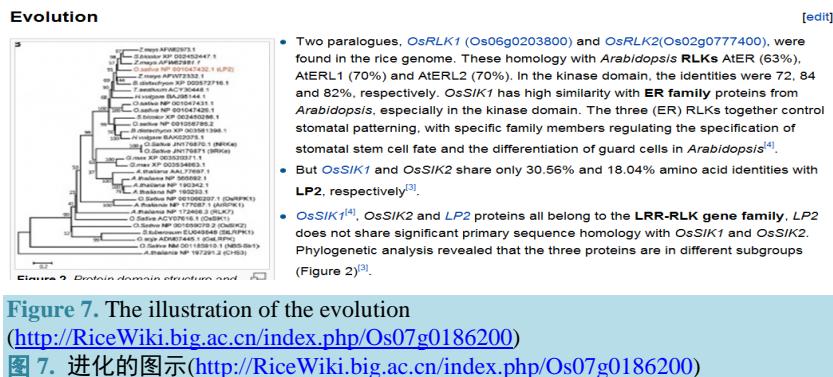
- Upon stress, OsTZF1-green fluorescent protein localization was observed in the **cytoplasm and cytoplasmic foci**^[1].
 - *OsTZF1* has **dynamic** subcellular localization patterns in the **cytoplasm** and probably the **nucleus**:
 - Under **normal growth conditions**, *OsTZF1* was predominantly localized in the cytoplasm of root meristem cells and was occasionally observed in **cytoplasmic foci**. *OsTZF1*-GFP was **rarely** observed in the **nucleus** of root cells at the young seedling stage.
 - **ABA and NaCl enhanced** the formation of **cytoplasmic foci** in **root cells**. The size, shape, and number of these cytoplasmic foci **varied** depending on the condition of plants and the type of roots. These cytoplasmic foci resembled processing bodies (PBs) and stress granules (SGs) where mRNA turnover and translational repression take place^{[3][4]}.
 - *OsTZF1* and *PABP8* colocalized in comparatively large cytoplasmic foci, considered to be SGs.
 - In onion (*Allium cepa*) epidermal cells, *OsTZF1* was localized in the cytoplasm and cell membran.
- The *OsTZF1*-GFP signal was detected in the **cytoplasm**, whereas GFP alone was distributed throughout the cell, the cytoplasmic localization of TZF proteins is likely a common feature in plants^[2].

Figure 6. The illustration of the subcellular localization (<http://RiceWiki.big.ac.cn/index.php/Os05g0195200>)

图 6. 亚细胞定位的图示(<http://RiceWiki.big.ac.cn/index.php/Os05g0195200>)

况自行添加或删除。如图 2 中，基因 *SNAC1* (<http://RiceWiki.big.ac.cn/index.php/Os03g0815100>) 注释信息的审编内容分成了 6 部分。由于各水稻抗逆基因的研究程度不相同，这 8 部分的信息并非全部都可获得，故而以不同的基因举例进行说明。

1) 从文献中获取抗逆基因的功能。如图 3，经过审编从 6 篇文献中总结了 4 条与基因 *OsNHX1* 相关的功能信息 (<http://RiceWiki.big.ac.cn/index.php/Os07g0666900>)。 *OsNHX1* 作为 Na⁺ 和 H⁺ 的交换器在水稻的耐盐性中扮演重要角色，与水稻根部相比， *OsNHX1* 更倾向于在水稻芽的耐盐性中起作用； *OsNHX1* 在



Na⁺从细胞质到液泡的划分中起到重要作用;*OsNHX1*在Actin1D启动子存在时,其转基因植株可增强耐盐性,在水稻的整个生命周期中,对盐胁迫产生的毒性效应具有缓冲和减轻作用;*OsNHX1*能够抑制Na⁺, Li⁺等。

2) GO(Gene Ontology)号[22]。从NCBI中可以得到水稻基因对应的GO号信息(并非所有基因都有对应的GO号),在AmiGO网站[23]上有GO号对应的详细信息。整理后将GO号和GO号在AmiGO网站上的链接放到RiceWiki网页上。如图3,以水稻基因*OsNHX1*(<http://RiceWiki.big.ac.cn/index.php/Os07g0666900>)为例,*OsNHX1*对应5个GO号,在AmiGO网站上可以找到GO的相关信息。这5个GO号中,有2个属于生物过程的GO号:GO:0006814和GO:0006885;有两个属于分子功能类的GO号:GO:0015299和GO:0015385;还有一个属于细胞组件的GO号:GO:0016021。这5个GO号代表的信息分别为:Na⁺运输、PH调节、溶质的逆向转运活性、钠的质子逆向转运活性和膜的组成部分。

3) 突变。如图4,主要列举突变体与野生型之间的区别。以基因*OsDEG10*(<http://RiceWiki.big.ac.cn/index.php/Os08g0139000>)为例,主要搜集了其RNAi转基因植株与野生型植株之间的差别:在高光处理下,RNAi植株显示出更明显的Fv/Fm率;比野生型的叶黄素色素含量低;冷处理后,*OsDEG10* RNAi的转基因植株表现出的Fv/Fm率比野生型要低,这表明*OsDEG10* RNAi的转基因植株对冷胁迫更敏感,*OsDEG10*可能在冷胁迫和氧化胁迫的应答反应中有着重要作用;在理想的生长条件下,*OsDEG10* RNAi的转基因植株和野生型植株没有任何形态变异。总的来说,与野生型相比,RNAi的转基因植株对强光和冷胁迫更敏感。

4) 此基因与其他基因的关系。在图5中,以基因*SNAC1*为例,审编后得到两类和*SNAC1*[24]相关的基因:*SNAC1*的靶基因和受*SNAC1*调控的基因(<http://RiceWiki.big.ac.cn/index.php/Os03g0815100>)。*OsSRO1c*[24][25]是*SNAC1*的靶基因,*SNAC1*能够连接到*OsSRO1c*的启动子并激活*OsSRO1c*的表达。*OsSRO1c*在干旱和氧化胁迫的抗性中扮演双重角色,促进气孔的关闭和H₂O₂的累积。受到*SNAC1*调控的基因又分为两类,一个是*OsPP18*[26],*OsPP18*是一种新型的PP2C基因,受到*SNAC1*的调控,通过调节ROS的稳态能够增强抗旱性和抗氧化性;另一个是和胁迫相关的基因,*SNAC1*编码一个NAC转录因子,能够调控许多和压力胁迫相关的基因,如*DREB1A*,*CBF*,*SCOF*,*Tsi*,*OSISAP1*。

5) 表达。主要收集基因表达的部位(如根,叶等)、受到的胁迫(生物和非生物胁迫等)、过度表达产生的生理形态和功能差别等。以*OsNHX1*(<http://RiceWiki.big.ac.cn/index.php/Os07g0666900>)为例,审编之后得到了7条表达信息,如前三条:在年幼的叶片中,三个*OsNHX1*的过度表达系(T4, T5, T6)表现出高表达;*OsNHX1*在根与芽的表达是不同的,在这两个组织中都会受到盐胁迫的刺激;通过半定量RT-PCR和Western印迹分析,对Actin1D启动子下的*OsNHX1*的过度表达进行了研究。在160 mM的盐胁迫下,转基因的籽苗生长良好,和对照组相比,显示出最小程度的根和芽长度的减少。预测分析表明在160 mM的盐胁迫下,野生型中出现更高的萎黄病。

6) 亚细胞定位。一般情况下, 经过实验验证的亚细胞定位信息是确定的。如基因 *OsSPX1* (<http://RiceWiki.big.ac.cn/index.php/Os06g0603600>), 实验结果表明 *OsSPX1* 是一个核定位蛋白。但有些特殊基因, 如图 6, 以 *OsTZF1* 为例(<http://RiceWiki.big.ac.cn/index.php/Os05g0195200>), 其亚细胞定位信息会受到 ABA 和 NaCl 的影响, 导致 *OsTZF1* 在细胞质和细胞核中可能有动态的亚细胞定位模式。

7) 进化。包括此基因所在家族、进化树、序列相似基因、同源基因等。图 7 中以 *OsSIK2* (<http://RiceWiki.big.ac.cn/index.php/Os07g0186200>) 为例进行说明。*OsSIK2* 的两个旁系同源, *OsRLK1*(Os06g0203800)和 *OsRLK2*(Os02g0777400); *OsSIK2* 和 *LP2* 氨基酸的同一性只有 18.04%, *OsSIK1*、*OsSIK2* 和 *LP2* 蛋白都属于 LRR-RLK 基因家族, 系统进化分析表明三个蛋白都属于不同的亚组。

8) 知识的延伸。这一环节主要涉及到基因所在家族以及基因家族中其它重要基因的信息概括、和此基因拥有类似功能的基因的相关信息等。如基因 *OsSIK2*, 属于 LRR-RLK 基因家族, 审编后用图和文字对 RLK 的相关信息和功能进行了说明(<http://RiceWiki.big.ac.cn/index.php/Os07g0186200>)。

3.2.2. 研究此基因的实验室

以基因 *OsNHX1* 为例, 审编后得到 14 个研究此基因的实验室 (<http://RiceWiki.big.ac.cn/index.php/Os07g0666900>)。将研究此基因的相关实验室放到 RiceWiki 的网页中一方面保护了各实验室的科研著作权, 另一方面可以使研究人员清楚的了解有哪些实验室参与了这个基因的研究, 方便研究人员之间的交流沟通, 此外可以吸引更多实验室的科研人员可以加入到 RiceWiki 中来, 将其实验室研究基因的相关信息放到 RiceWiki 页面上, 实现资源共享。

3.2.3. 参考文献

以 *OsNHX1* 为例, 审编过程中引用了 13 篇文献(<http://RiceWiki.big.ac.cn/index.php/Os07g0666900>)。RiceWiki 网页上关于基因的信息均来自于相关的文献, 对审编的基因信息的各个环节都表明出处。另外, 引用的文章在审编过程中出现的次数也会加以标明, 如此一来, 不仅保护了原文献作者的著作权, 而且对今后的修改和查阅提供了方便。如果用户对某个审阅的内容感兴趣或出现疑虑, 都可以通过后面标注的相关引用找到审阅内容的原出处。

3.2.4. 结构信息

以表格的形式出现, 使研究人员直观了解基因结构信息。包括基因版本、结构图示、编码序列等。

3.3. 审编质量的评分

根据评分细则, 贡献值 CS 为 782.82, 总编辑量为 782820, 平均编辑质量为 1。

3.4. 水稻抗逆基因的分析

本文主要从 PPI、共表达和 GO 分析对 500 个水稻抗逆基因进行了分析。

3.4.1. 水稻蛋白互作网络

水稻的各项生命活动不是单独靠某个蛋白完成的, 而是依靠蛋白互作完成的。基于对水稻抗逆基因之间是否会通过蛋白互作完成某些生物功能的猜想, 对水稻抗逆基因对应的 PPI 进行了分析。利用 PRIN(Predicted Rice Interactome Network)的方法[27], 首先对 500 个抗逆基因对应的蛋白互作进行统计, 如图 8(a), 得到 2334 个蛋白之间的 4068 个蛋白互作的网络图, 即 500 个抗逆基因对应 4068 个蛋白互作, 这些水稻抗逆基因对水稻各项生命活动至关重要。之后分析了 500 个水稻抗逆基因之间的蛋白互作, 得到了 135 个蛋白间的 213 个蛋白互作网络。在这 213 个蛋白互作中, 有 135 个是自身的蛋白互作。去除这些自身互作(可能形成了二聚体), 如图 8(b), 得到了 61 个蛋白质间的 78 个蛋白互作的网络图。

3.4.2. 共表达分析

共表达水平由皮尔森系数(PCC)决定[28], 由下面方程式计算 PCC [28]。若 r 大于零, 值越高, 表达正相关性越大; 若 r 小于零, 值越低, 表达负相关性越大。

$$r = \frac{1}{N} \sum_{i=1, \dots, N} \left(\frac{X_i - \bar{x}}{\sigma_x} \right) \left(\frac{Y_i - \bar{y}}{\sigma_y} \right)$$

鉴定 Ho 等人的数据[28], 整理 165 个不同处理下的芯片数据, 利用 R 软件包 AFFY 计算, 得到各基因表达, 随后计算 PCC。

对图 8(b)中的 78 个蛋白互作进行了 PCC 的计算, 将其 r 值从(-1, 1)这个区间每 0.1 为一个间隔做 r 值的百分比统计。结果显示在(-1.0, -0.6)区间的百分比统计为 0。如图 9 所示, 发现有 9.46%的蛋白对会互相抑制(PCC 值小于-0.3), 协同作用的可能性(PCC 值大于零)比抑制作用的可能性(PCC 值小于零)要大, 共表达趋势明显的占 21.61% (PCC 值大于 0.3)。从侧面说明这些抗逆基因编码的蛋白能够互作。

3.4.3. GO 分析

GO 用三种语义(分子功能、细胞组件、生物过程)描述基因及其产物。本文应用 argiGO [29] 的 SEA 工具和 Fisher 统计法。将显著性水平高的 GO 语义和对应各 GO 语义的基因数占总数的百分比进行统计。如图 10, 蓝色代表水稻抗逆基因, 绿色代表背景基因(*Oryza sativa* MSU7.0)。基因百分比比较高的有: 细胞过程、代谢过程、结合和对刺激的应答等。其中, 对刺激的应答(生物过程类)基因百分比为 50.1%, 而

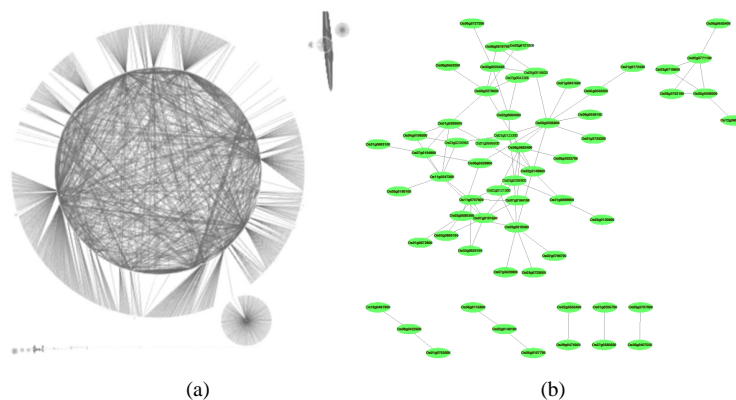


Figure 8. PPI network of rice stress-resistance genes (a) is PPI network of proteins which 500 rice stress-resistance genes participate in; (b) is PPI network between 500 rice stress-resistance genes

图 8. 水稻抗逆基因的蛋白互作网络 (a) 为 500 个水稻抗逆基因的蛋白参与的蛋白互作网络; (b) 为 500 个水稻抗逆基因之间的蛋白互作网络

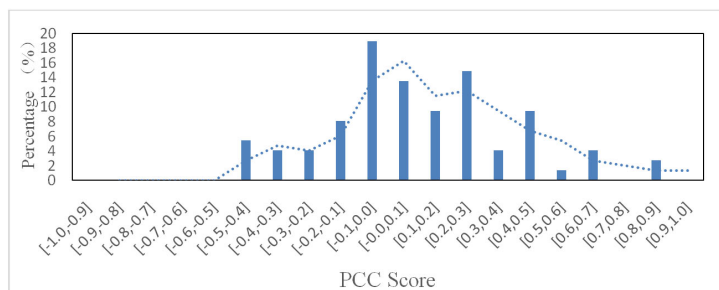


Figure 9. The PCC score of PPI between the rice stress-resistance genes

图 9. 水稻抗逆基因之间的 PPI 中的 PCC 值

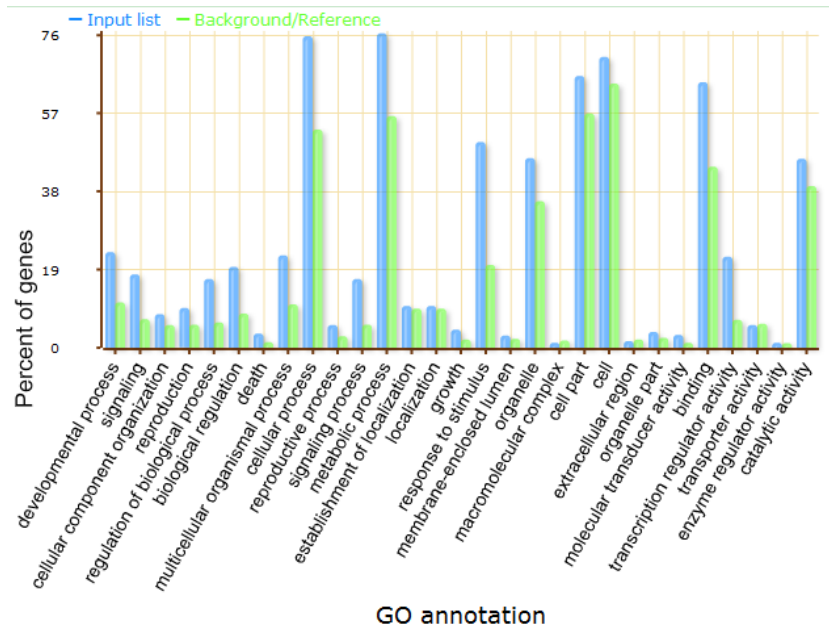


Figure 10. The percent of genes in GO annotation
图 10. GO 注释中的基因百分比

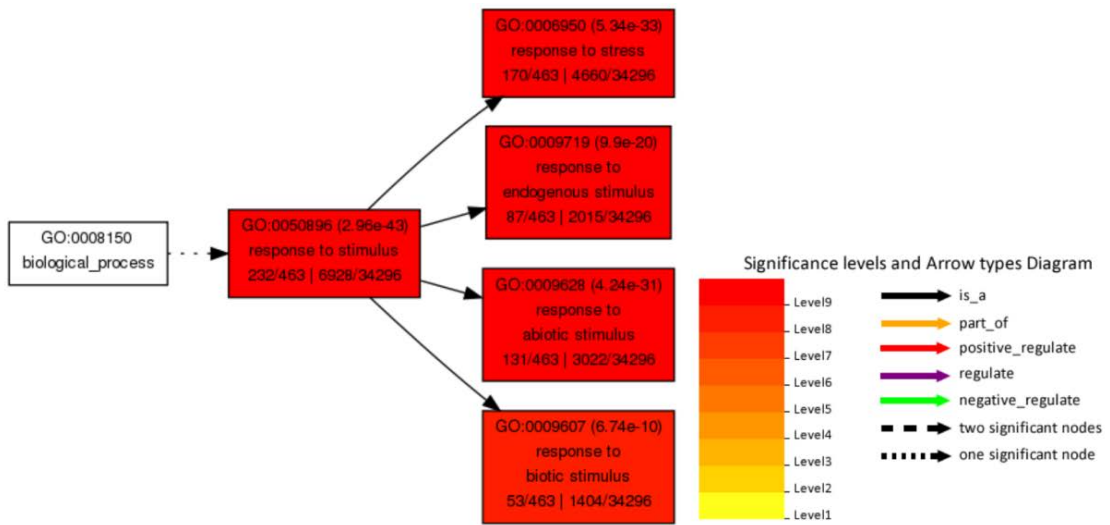


Figure 11. The significance levels of the GO terms related to stress
图 11. 和胁迫相关的 GO 语义的显著性水平

在背景组中为 20.2%。

在生物过程类的 GO 语义中，和压力胁迫相关的基因节点颜色深，表明其显著性高。选其作为结点，边表示关系。每个 GO 语义框中通常会有以下信息：GO 语义及其描述、对应此 GO 语义的抗逆基因的数目和对应到 GO 语义的背景(*Oryza sativa* MSU7.0)的基因的数目。如图 11，463 个抗逆基因有 GO 语义；背景(*Oryza sativa* MSU7.0)有 34296 个基因有 GO 语义。以“对刺激的应答”语义框为例，有 232 个抗逆基因有此 GO 语义，表明和生物胁迫相关的语义的显著性水平非常高。

在生物过程类的 GO 语义中，GO 的显著性水平比较高的有：对刺激的应答、生物调节、细胞过程、代谢过程等。图 11 中，我们选取了 5 个和胁迫相关的深颜色的结点：1) GO: 0050896 (对刺激的应答)；

2) GO: 0009719 (对生物体内部刺激的应答); 3) GO: 0009628 (对非生物刺激的应答); 4) GO: 0009607 (对生物刺激的应答); 5) GO: 0006950 (对压力的应答)。在 463 个有 GO 语义的水稻抗逆基因中, 分别对应上述 5 个 GO 语义的基因数目为: 232、87、131、53、170 个。根据箭头的颜色和形状可以得出这 5 个 GO 语义的关系: 对刺激的应答是对生物体内部刺激的应答、是对非生物刺激的应答、是对生物刺激的应答、是对压力的应答。

4. 结论

本文总结了盐害、干旱等 8 类对水稻产量影响较大的胁迫压力, 并对这 8 类水稻抗性基因进行了搜集, 获得了 500 个抗逆基因。通过生物审编, 对水稻抗逆基因的信息进行了整理汇总, 审编后的结果发布到 RiceWiki 网页上。虽然在审编的过程中, 将基因的注释信息分为了功能、表达等独立的 8 部分, 但每部分之间是相互关联的。如基因的 GO 号能体现生物学功能, 从突变体与野生型的差别及表达等信息中可推测其功能等。

搜集水稻的抗逆基因以及审编基因信息对水稻的抗逆研究有重要意义。PPI 和共表达等分析, 有助于系统地了解这些基因之间的联系和共同点。RiceWiki 是以群体审编为基础的开放性平台, 随着审编的普及和发展, 会有更多审编人员的加入, 相信 RiceWiki 会成为一部名副其实的“水稻百科全书”。

此外, 在审编过程中发现, 很多水稻的抗性基因, 如基因 *OsbZIP71*、*OsHsfB2b*、*OsCPK21*、*OsCPK13* 等, 和激素之间有重要的关系, 这些激素在水稻的抗逆性中有可能起到重要作用。水稻植株体内的某些物质, 如糖等, 有可能影响生物的抗逆性。日后对水稻抗逆机制的研究和推断可以借鉴从 RiceWiki 中获得的相关信息。

参考文献 (References)

- [1] Shimamoto, K. and Kyozuka, J. (2002) Rice as a model for comparative genomics of plants. *Annual Review of Plant Biology*, 53, 399-419. <http://dx.doi.org/10.1146/annurev.arplant.53.092401.134447>
- [2] Atkinson, N.J. and Urwin, P.E. (2012) The interaction of plant biotic and abiotic stresses: From genes to the field. *Journal of Experimental Botany*, 63, 3523-3543. <http://dx.doi.org/10.1093/jxb/ers100>
- [3] Zhu, J.-K. (2001) Plant salt tolerance. *Trends in Plant Science*, 6, 66-71. [http://dx.doi.org/10.1016/S1360-1385\(00\)01838-0](http://dx.doi.org/10.1016/S1360-1385(00)01838-0)
- [4] Xiong, L. and Zhu, J.K. (2001) Abiotic stress signal transduction in plants: Molecular and genetic perspectives. *Physiologia Plantarum*, 112, 152-166. <http://dx.doi.org/10.1034/j.1399-3054.2001.1120202.x>
- [5] Nishiyama, I. (1984) Climatic influence on pollen formation and fertilization. *Biology of Rice*, 153-171. <http://dx.doi.org/10.1016/B978-0-444-99615-2.50012-0>
- [6] Wade, L., McLaren, C., Quintana, L., Harnpichitvitaya, D., Rajatasereekul, S., Sarawgi, A., Kumar, A., Ahmed, H., Singh, A. and Rodriguez, R. (1999) Genotype by environment interactions across diverse rainfed lowland rice environments. *Field Crops Research*, 64, 35-50. [http://dx.doi.org/10.1016/S0378-4290\(99\)00049-0](http://dx.doi.org/10.1016/S0378-4290(99)00049-0)
- [7] Peng, S., Huang, J., Sheehy, J.E., Laza, R.C., Visperas, R.M., Zhong, X., Centeno, G.S., Khush, G.S. and Cassman, K.G. (2004) Rice yields decline with higher night temperature from global warming. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 9971-9975. <http://dx.doi.org/10.1073/pnas.0403720101>
- [8] Skamnioti, P. and Gurr, S.J. (2009) Against the grain: Safeguarding rice from rice blast disease. *Trends in Biotechnology*, 27, 141-150. <http://dx.doi.org/10.1016/j.tibtech.2008.12.002>
- [9] Schatz, M.C., Langmead, B. and Salzberg, S.L. (2010) Cloud computing and the DNA data race. *Nature Biotechnology*, 28, 691-693. <http://dx.doi.org/10.1038/nbt0710-691>
- [10] Eisenstein, M. (2012) Oxford nanopore announcement sets sequencing sector abuzz. *Nature Biotechnology*, 30, 295-296. <http://dx.doi.org/10.1038/nbt0412-295>
- [11] Zhang, Z., Sang, J., Ma, L.N., Wu, G., Wu, H., Huang, D.W., Zou, D., Liu, S.Q., Li, A., Hao, L.L., Tian, M., Xu, C., Wang, X.M., Wu, J.Y., Xiao, J.F., Dai, L., Chen, L.-L., Hu, S.N. and Yu, J. (2013) Ricewiki: A wiki-based database for community curation of rice genes. *Nucleic Acids Research*, 42, D1222-D1228.
- [12] Callaway, E. (2010) No rest for the bio-wikis, biologists' collaborative data repositories come of age. *Nature*, 468,

- 359-360. <http://dx.doi.org/10.1038/468359a>
- [13] Zhang, Z., Zhu, W.M. and Luo, J.C. (2014) Bringing biocuration to China. *Genomics Proteomics Bioinformatics*, **12**, 153-155. <http://dx.doi.org/10.1016/j.gpb.2014.07.001>
- [14] Burge, S., Attwood, T.K., Bateman, A., Berardini, T.Z., Cherry, M., O'Donovan, C., Xenarios, L. and Gaudet, P. (2012) Biocurators and biocuration, surveying the 21st century challenges. *Database (Oxford)*, **2012**, bar059. <http://dx.doi.org/10.1093/database/bar059>
- [15] Kinoshita, T. (1995) Report of committee on gene symbolization, nomenclature and linkage groups. *Rice Genetics Newsletter*, **12**, 9-153.
- [16] Sakai, H., Lee, S.S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., Wakimoto, H., Yang, C.-C., Iwamoto, M., Abe, T., et al. (2013) Rice annotation project database (RAP-DB): An integrative and interactive database for rice genomics. *Plant and Cell Physiology*, **54**, e6. <http://dx.doi.org/10.1093/pcp/pcs183>
- [17] Zhao, W.M., Wang, J., He, X.M., Huang, X.B., Jiao, Y.Z., Dai, M.T., Wei, S.L., Fu, J., Chen, Y., Ren, X.Y., Zhang, Y., Ni, P.X., Zhang, J.G., Li, S.G., Wang, J., Wong, G.K.-S., Zhao, H.Y., Yu, J., Yang, H.M. and Wang, J. (2004) BGI-RIS: An integrated information resource and comparative analysis workbench for rice genomics. *Nucleic Acids Research*, **32**, D377-D382. <http://dx.doi.org/10.1093/nar/gkh085>
- [18] Ouyang, S., Zhu, W., Hamilton, J., Lin, H.N., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y.D., Zheng, L., Orvis, J., Haas, B., Wortman, J. and Buell, C.R. (2007) The TIGR rice genome annotation resource, improvements and new features. *Nucleic Acids Research*, **35**, D883-D887. <http://dx.doi.org/10.1093/nar/gkl976>
- [19] McCouch, S.R. (2008) Gene nomenclature system for rice. *Rice*, **1**, 72-84. <http://dx.doi.org/10.1007/s12284-008-9004-9>
- [20] Dai, L., Tian, M., Wu, J.Y., Xiao, J.F., Wang, X.M., Townsend, J.P. and Zhang, Z. (2013) AuthorReward: Increasing community curation in biological knowledge wikis through automated authorship quantification. *Bioinformatics*, **29**, 1837-1839. <http://dx.doi.org/10.1093/bioinformatics/btt284>
- [21] Bateman, A. (2010) Curators of the world unite: The international society of biocuration. *Bioinformatics*, **26**, 991. <http://dx.doi.org/10.1093/bioinformatics/btq101>
- [22] Consortium, G.O. (2004) The gene ontology (GO) database and informatics resource. *Nucleic Acids Research*, **32**, D258-D261. <http://dx.doi.org/10.1093/nar/gkh036>
- [23] Carbon, S., Ireland, A., Mungall, C.J., Shu, S.Q., Marshall, B. and Lewis, S. (2009) AmiGO: Online access to ontology and annotation data. *Bioinformatics*, **25**, 288-289. <http://dx.doi.org/10.1093/bioinformatics/btn615>
- [24] You, J., Zong, W., Li, X.K., Ning, J., Hu, H.H., Li, X.H., Xiao, J.H. and Xiong, L.Z. (2013) The SNAC1-targeted gene OsSRO1c modulates stomatal closure and oxidative stress tolerance by regulating hydrogen peroxide in rice. *Journal of Experimental Botany*, **64**, 569-583. <http://dx.doi.org/10.1093/jxb/ers349>
- [25] Hu, H.H., Dai, M.Q., Yao, J.L., Xiao, B.Z., Li, X.H., Zhang, Q.F. and Xiong, L.Z. (2006) Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12987-12992. <http://dx.doi.org/10.1073/pnas.0604882103>
- [26] You, J., Zong, W., Hu, H.H., Li, X.H., Xiao, J.H. and Xiong, L.Z. (2014) A SNAC1-regulated protein phosphatase gene OsPP18 modulates drought and oxidative stress tolerance through ABA-independent reactive oxygen species scavenging in rice. *Plant Physiology*, **166**, 2100-2114.
- [27] Gu, H.B., Zhu, P.C., Jiao, Y.M., Meng, Y.J. and Chen, M. (2011) PRIN: A predicted rice interactome network. *BMC Bioinformatics*, **12**, 161. <http://dx.doi.org/10.1186/1471-2105-12-161>
- [28] Ho, C.L., Wu, Y.Z., Shen, H.B., Provart, N.J. and Geisler, M. (2012) A predicted protein interactome for rice. *Rice*, **5**, 15. <http://dx.doi.org/10.1186/1939-8433-5-15>
- [29] Du, Z., Zhou, X., Ling, Y., Zhang, Z.H. and Su, Z. (2010) AgriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Research*, **38**, W64-W70. <http://dx.doi.org/10.1093/nar/gkq310>