Original article

SuCComBase: a manually curated repository of plant sulfur-containing compounds

Sarahani Harun[1], Muhammad-Redha Abdullah-Zawawi[1], Mohd Rusman Arief A-Rahman[1], Nor Azlan Nor Muhammad[1] and Zeti-Azura Mohamed-Hussein [ID][1,2,*]

[1]Centre for Bioinformatics Research, Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia and [2]Centre for Frontier Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

*Corresponding author: Tel.: +603 8921 4547; Fax: 603 8921 3398; Email: zeti.hussein@ukm.edu.my

Abstract

Plants produce a wide range of secondary metabolites that play important roles in plant defense and immunity, their interaction with the environment and symbiotic associations. Sulfur-containing compounds (SCCs) are a group of important secondary metabolites produced in members of the Brassicales order. SCCs constitute various groups of phytochemicals, but not much is known about them. Findings from previous studies on SCCs were scattered in published literatures, hence SuCComBase was developed to store all molecular information related to the biosynthesis of SCCs. Information that includes genes, proteins and compounds that are involved in the SCC biosynthetic pathway was manually identified from databases and published scientific literatures. Sets of co-expression data was analyzed to search for other possible (previously unknown) genes that might be involved in the biosynthesis of SCC. These genes were named as potential SCC-related encoding genes. A total of 147 known and 92 putative *Arabidopsis thaliana* SCC-related genes from literatures were used to identify other potential SCC-related encoding genes. We identified 778 potential SCC-related encoding genes, 4026 homologs to the SCC-related encoding genes and 116 SCCs as shown on SuCComBase homepage. Data entries are searchable from the Main page, Search, Browse and Datasets tabs. Users can easily download all data stored in SuCComBase. All publications related to SCCs are also indexed in SuCComBase, which is currently the first and only database dedicated to plant SCCs. SuCComBase aims to become a manually curated and *au fait* knowledge-based repository for plant SCCs.

Database URL: http://plant-scc.org

Introduction

Secondary metabolites are metabolites synthesized by living cells that are not directly involved in cellular metabolism of organisms such as growth, development and reproduction (1, 2). In plants, secondary metabolites act as chemical defense against plant pests and pathogens. Many experimental findings revealed various important functions of secondary metabolites indicating their role and importance in defense response against pathogens (1, 3–6). Apart from being toxic or repellant to herbivores or pests and microbes, they also regulate beneficial interactions such as attracting pollinators or seed dispersal and modulation of abiotic stress responses (7–9).

Different characteristic plant species–specific mix of these chemicals can be used as taxonomic identifier in plant classification (10, 11). Secondary metabolites are extremely diverse and usually belong to one of the three major classes, namely terpenes, phenolics and alkaloids (5, 12). However, there is one unusual plant constituent that was found to be highly involved in plant defense system known as sulfur-containing compounds (SCCs) (11). SCCs are very important in plant–pest interaction in various types of plant families comprising species-specific defense chemicals such as camalexin in Brassicaceae, glucosinolates (GSLs) in Brassicales, alline in Alliaceae, thiopene in Asteraceae and defensins in other plant families (13).

Currently, there are almost 200 SCCs found in Brassicaceae (11, 14). Camalexin for example, is a major phytoalexin SCC found in *Arabidopsis thaliana* that plays an important role in deterring pathogens such as *Botrytis cinera* (15) and *Alternaria brassicicola* (16). Research on SCCs has contributed to new knowledge on their important biological induction: for example, antioxidant activity, chemoprevention effects and apoptosis (17). However, among known SCCs, only GSLs and isothiocyanates have become popular research subjects due to their anticancer activities (18–22). Isothiocyanates in cruciferous vegetables were shown to induce glutathione S-transferase and NAD(P)H:quinone oxidoreductase 1 that act as cell protectants by detoxifying against potential carcinogens and oxidants (23), which explain their effect in reducing the risk of bladder cancer in individuals who consume loads of these vegetables (24, 25). Besides, they also have broad antibiotics properties such as antimicrobial, nematocidal, antifungal and antiprotozoal (26, 27). Furthermore, sulforaphane is one of the most studied isothiocyanates in numerous animal studies [e.g. samples taken from lung (28), colon (29), breast (30), skin (31), stomach (32), small intestine (33) prostate (34), pancreas (35) and oral cells (36)]. The findings reported the ability of sulforaphane to inhibit the carcinogenic cells at either in the early or late stages of malignant tissues. Meanwhile, GSLs are potential putative repellents and attractants of diamondback moth (DBM), *Plutella xylostella* L., which is a widespread destructive pest on Brassicales crops (37, 38). A number of studies revealed the contribution of different GSL profiles toward the behavior of DBM feeding that serve as novel findings in plant–herbivore interactions (39–41).

Information regarding SCCs genes, proteins and compounds that are involved in SCC biosynthetic pathway is abundant due to their potential contribution in pharmaceutical and agricultural industries. The availability of co-expression studies has significantly contributed to the search of new or potential SCC-related encoding genes. However, this information was distributed in various literature and biological databases, causing difficulties in finding all genes and molecular information of SCCs. This limitation has prompted the need to develop a digital repository that provides a platform for easy access of comprehensive information on SCCs, hence the development of a manually curated database called SuCComBase. The collective data in this database aims to provide valuable resources for genomic studies on potential SCC-related genes that might be involved in the SCC biosynthetic pathway. Herein, the development and current status of SuCComBase were described with the web interface systematically elaborated. SuCComBase is accessible at http://plant-scc.org (SuCComBase ver1.2, last updated on 4 January 2019).

Materials and methods

Data collection

Extensive bibliomic (all related publications published between early 2001 and 2017) and database searching from publication and biological databases were carried out to identify all genes responsible in encoding the SCCs related proteins in *A. thaliana*. Known SCC-related genes and compounds were identified using various keywords such as 'sulfur containing compounds', 'sulfur containing secondary metabolites', 'sulphur containing compounds', 'sulphur containing secondary metabolites', 'glucosinolate', 'phytoalexin', 'camalexin' and 'Brassicaceae', '*Arabidopsis thaliana*', '*Brassica rapa*', 'cabbage', '*Brassica oleracea*', 'broccoli', '*Carica papaya*' and 'papaya'. These keywords were searched in publication databases such as PubMed, ScienceDirect and Scopus and biological databases, e.g. AraCyc v8.0 (42), KEGG v88.2 (43), KNApSAcK (last updated on 29 June 2018) (44) and PubChem v1.6.2 beta (45). Putative SCC-related genes were also identified using the keyword search in AraCyc and KEGG; however, no experimental evidence can be found to support their contribution in SCC biosynthesis.
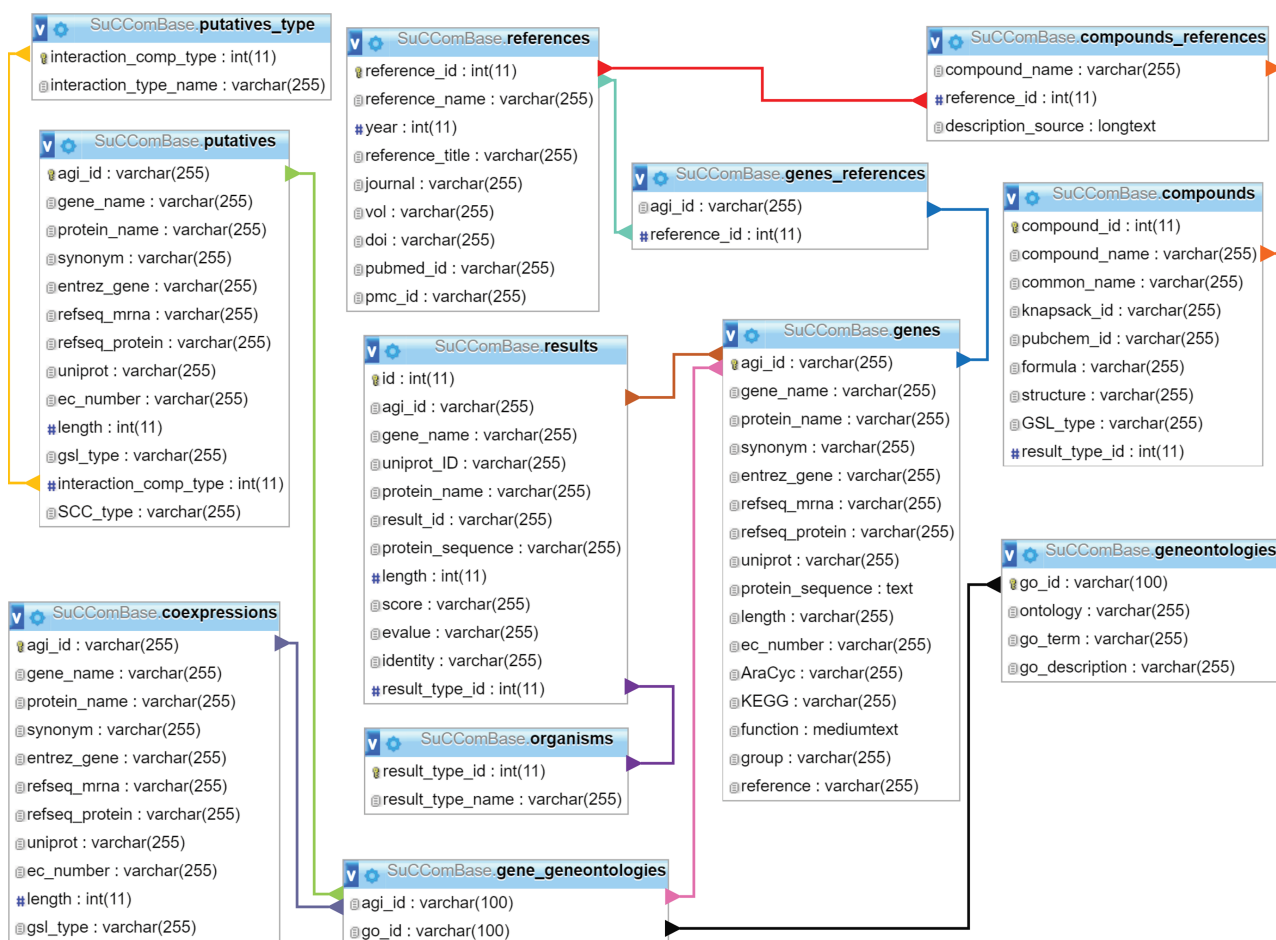
Figure 1. SuCComBase schema contains 12 tables with the connections from table to table.

Functional annotation

Genes identified from database searching were manually curated using all relevant information obtained from various databases such as NCBI Gene (last updated 4 December 2018) (46), UniProt (last updated 5 December 2018) (47), AraCyc (42), KEGG (43), Ensembl Plants v41 (48), KNApSAcK (44) and PubChem (45). Three co-expression databases, i.e. ATTED v9 (49), AraNet v2 (50) and Gene-MANIA (last updated on 14 March 2017) (51), were used to identify potential SCC-related encoding genes. The protein sequence of known SCC encoding genes of *A. thaliana* were retrieved from Phytozome v12.1.6, which is a Plant Comparative Genomics Portal containing 93 plant genomes (52). SCC homologs were identified via BLAST (53) against *B. rapa*, *B. oleracea* and *C. papaya* in Phytozome database using the known SCC protein sequences as queries. Gene Ontology Consortium (last updated on 26 October 2018) (54) was used to identify gene ontology (GO) in known SCC-related genes, potential SCC-related genes and putative SCC-related genes to provide a clearer understanding of SCC biosynthesis in *A. thaliana*.

Database organization and architecture

SuCComBase consisted of 12 linked tables (Figure 1) with information on the SCC encoding genes and compounds, their functional information and references. MySQL Server 5.0.11 was used to host SuCComBase relational database. The SuCComBase web interfaces were developed using Laravel 5.3.31 (PHP web framework), HTML and JavaScript.

Results and Discussion

Database summary

Figure 2 shows the organization of data types whereby SCC-related genes in *A. thaliana* are assigned as the main data in SuCComBase. A total of 147 known SCC-related genes were manually curated and supported with added information obtained from KEGG and AraCyc. We have identified 92 computationally predicted genes that might be involved in the production of SCCs and classified them as putative SCC-related genes. The known SCC-related genes were used as queries in identifying potential
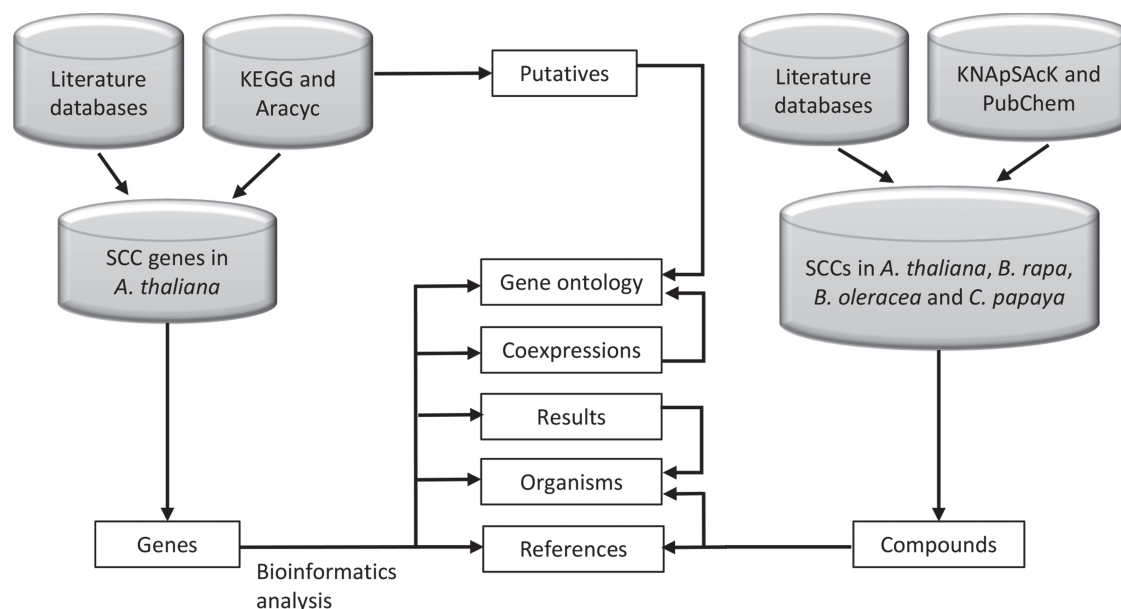
Figure 2. SuCComBase data types structure organization. These data types are tables that can be found in 'Browse' and 'Datasets' menu.
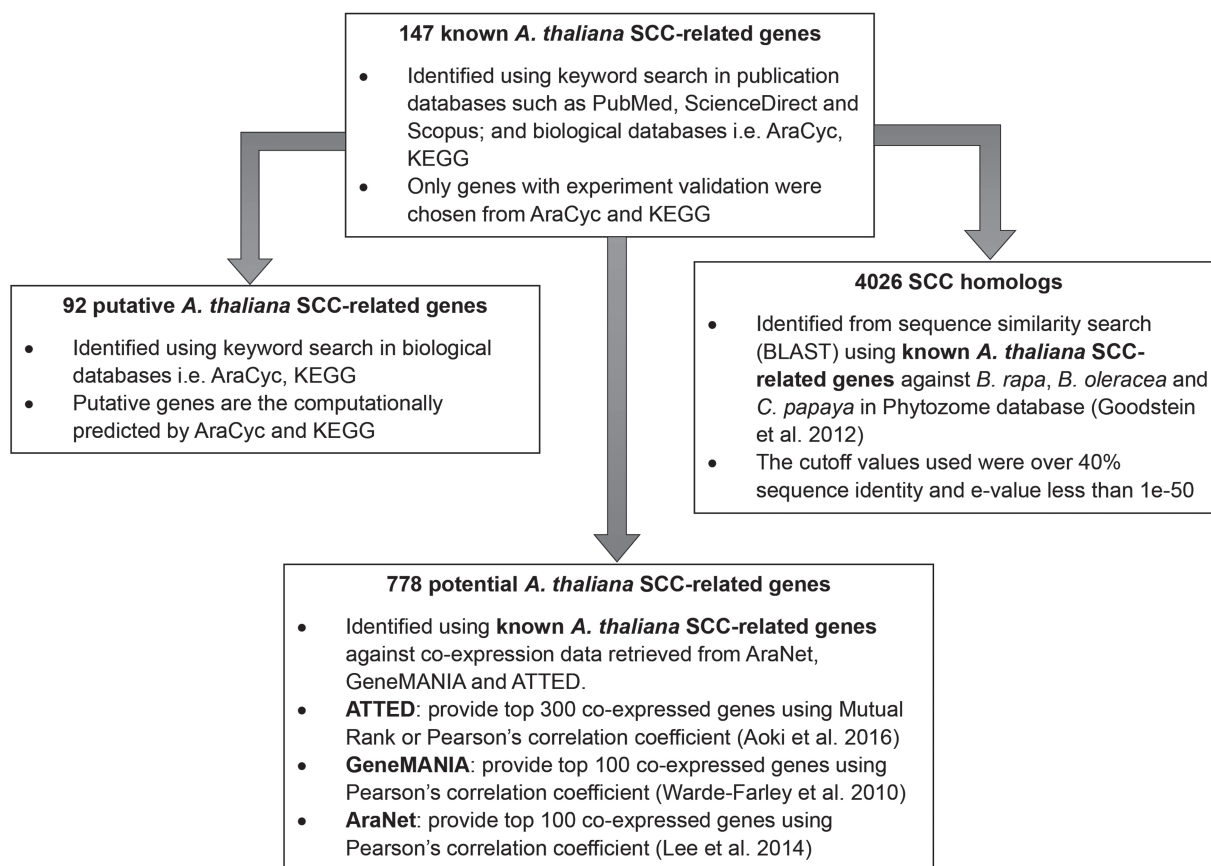
Figure 3. Identification of 92 putative *A. thaliana* SCC-related genes, 778 potential *A. thaliana* SCC-related genes and 4026 SCC homologs using 147 known *A. thaliana* SCC-related genes as queries.

SCC-related genes from the co-expression data, and 778 potential SCC-related genes were successfully identified from three co-expression databases: e.g. ATTED (49), AraNet (50) and GeneMANIA (51). BLAST analysis against Phytozome database has identified 4026 SCC homologs from three Brassicales plant genomes, i.e. 1970

Locus ID
eg. AT1G01030

| 147 known *A. thaliana* SCC-related genes | 778 potential *A. thaliana* SCC-related genes | 92 putative *A. thaliana* SCC-related genes |

The locus ID from each classes of dataset (known, putative and potential) were used to search for the GO annotation information. The GO annotation information search was performed separately according to the respective datasets (known, putative and potential genes).

Date: July 5th 2018

GO annotation search

TAIR – Bulk Data Retrieval

https://www.arabidopsis.org/tools/bulk/go/index.jsp

Dataset version: July 1st 2018

The GO annotation search was performed on July 5th 2018 using the Bulk Data Retrieval from TAIR database. The GO dataset version was July 1st 2018 and the GO annotation data in TAIR is updated on a monthly basis.

Total number of GO

Known *A. thaliana* SCC-related genes	Potential *A. thaliana* SCC-related genes	Putative *A. thaliana* SCC-related genes
CC: 38	CC: 112	CC: 14
MF: 176	MF: 602	MF: 44
BP: 269	BP: 742	BP: 43

Merge   Merge   Merge

CC: 119; MF: 656; BP: 800

CC: cellular component; MF: molecular function; BP: biological process

GO annotation search result composed of:
1. **GO ID**: the unique identifier for a GO term.
2. **GO term**: the actual string of letters corresponding to the GO ID.
3. **aspect**: MF=molecular function, CC=cellular component, BP=biological process.

A total of **119 cellular component** (CC), **656 molecular function** (MF) and **800 biological process** (BP) were identified from 3 datasets (known, putative and potential). The GO IDs were merged to remove the duplicate ID present in respective datasets.

SQL Inventory

agi_id	ontology	go_id	go_term	go_description
AT1G01120	BP	GO:0000038	very long-chain	The chemical reactions and pathw
AT1G01120	CC	GO:0005783	endoplasmic ret	The irregular network of unit men
AT1G01120	CC	GO:0005789	endoplasmic ret	The lipid bilayer surrounding the e
AT1G01120	BP	GO:0006633	fatty acid biosyn	The chemical reactions and pathw
AT1G01120	BP	GO:0009409	response to cold	Any process that results in a chan

All the GO annotation information were stored into the SQL. It consists of:
1. **AGI ID**: Locus gene identifier
2. **GO ID**: Identifier corresponds to the GO term found.
3. **GO term**: the actual string of letters corresponding to the GO ID.
4. **Ontology/aspect**: MF=molecular function, CC=cellular component, BP=biological process.
5. **GO descriptions**: Explanation on three aspects; BP, MF and CC to describe gene function.

The relationship between genes and GO annotation information were presented into the pivot table. The many-to-many relationship were linked between the AGI ID with the matched GO ID.

Figure 4. Detailed overview on the approaches used to identify and annotate GO terms to each SCC-related genes entry. Known, potential and putative *A. thaliana* SCC-related genes were used to search for the GO annotation information.

Table 1. Number of entries in SuCComBase

Data set	Entries
Known *A. thaliana* SCC-related genes	147
Putative *A. thaliana* SCC-related genes	92
KEGG putative *A. thaliana* SCC-related genes	3
AraCyc putative *A. thaliana* SCC-related genes	89
Potential *A. thaliana* SCC-related genes	778
SCC homologs	4026
B. rapa SCC homologs	1970
B. oleracea SCC homologs	1319
C. papaya SCC homologs	737
Compounds	116
A. thaliana SCCs	47
B. rapa SCCs	28
B. oleracea SCCs	40
C. papaya SCCs	1
Publications	206

SCC homologs in *B. rapa*, 1319 SCC homologs in *B. oleracea* and 737 SCC homologs in *C. papaya*. Figure 3 summarizes the steps involved in identifying putative *A.*

thaliana SCC-related genes, potential *A. thaliana* SCC-related genes and the SCC homologs. Furthermore, we have included GO terms of the SCC-related genes where a total of 800 biological process, 656 molecular functions and 119 cellular components were included in this database. Figure 4 shows the steps performed in identifying the GO terms and the results obtained from the analysis. We have also identified 116 SCCs in *A. thaliana*, *B. rapa*, *B. oleracea* and *C. papaya* and included them in SuCComBase. Table 1 shows the summary of each data set in SuCComBase.

Database interface and access

The interface of SuCComBase contains a homepage and seven main menus, i.e. About, Browse, Search, Datasets, Download, Help and Contact. These menus are used to facilitate the users in navigating the pages effortlessly.

(i) SuCComBase Homepage displays data statistics in each table and provides a brief overview of this database.

(ii) The background of SuCComBase can be viewed in the 'About' page.
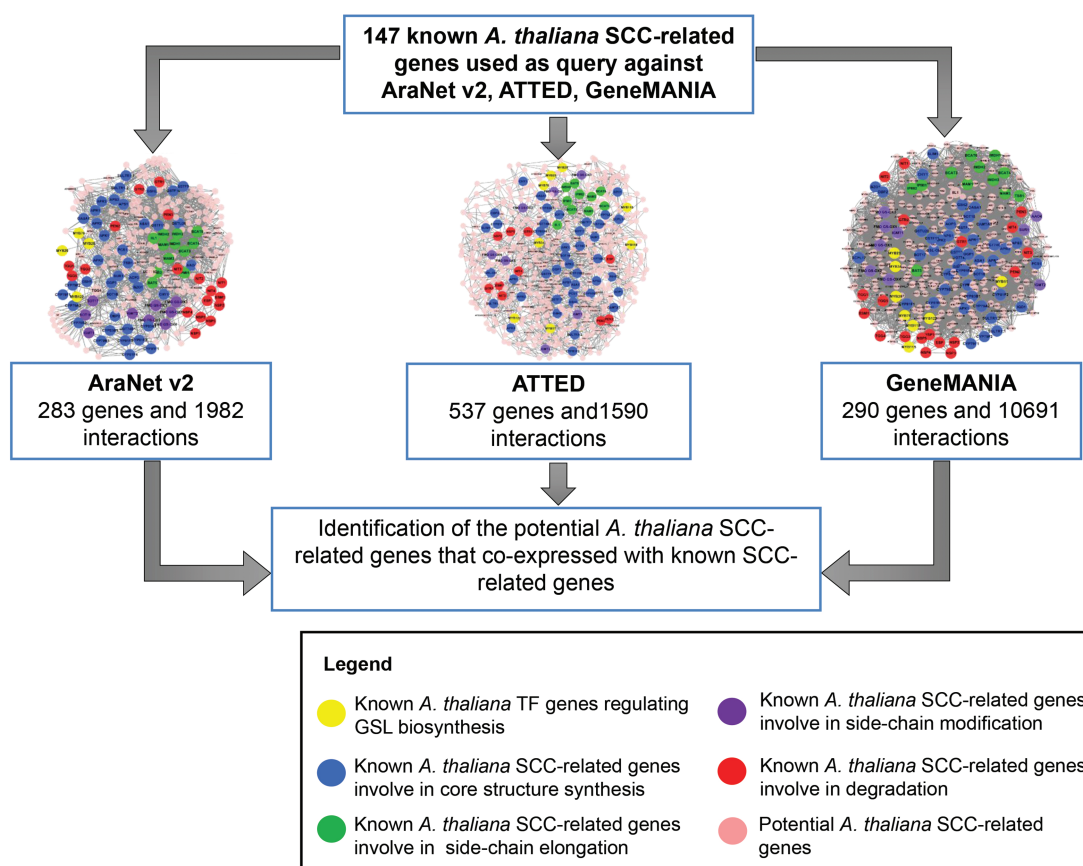
Figure 5. The integration of three co-expression gene networks reveals potential SCC-related genes. Different colors refer to the function of known SCC-related genes in GSL biosynthesis: yellow (transcription factor), blue (core structure synthesis), green (side-chain elongation), purple (side-chain modification) and red (GSL degradation). Known SCC-related genes were used as query to identify the co-expressed genes to be classified as potential SCC-related genes.

(iii) The 'Browse' page allows users to assess all six data sets in SuCComBase. Each data set varied based on their biological information as described below:

    (a) Known *A. thaliana* SCC-related gene data set: contains SCC-related encoding genes in *A. thaliana* including protein-encoding genes that involve in the GSL and camalexin biosynthetic pathways based on experiments reported in various publications and pathway databases (KEGG and AraCyc).

    (b) Putative *A. thaliana* SCC-related gene data set: contains computationally predicted GSL and camalexin genes from KEGG and AraCyc databases.

    (c) Potential *A. thaliana* SCC-related gene data set: contains 778 *A. thaliana* genes that might be involved in the SCC biosynthetic pathway based on the bioinformatic analysis using co-expression data retrieved from AraNet, GeneMANIA and ATTED. Figure 5 shows an example of the network constructed between the potential SCC-related encoding genes.

    (d) SCC homolog data set: contains SCC homologs in *C. papaya*, *B. rapa* and *B. oleracea*.

    (e) Compound data set: contains SCCs produced in *A. thaliana*, *C. papaya*, *B. rapa* and *B. oleracea*. All information was obtained from extensive literature search and from KNApSAcK as well as PubChem databases.

    (f) Publication data set: contains 206 published articles that were used in identifying SCCs and all SCC-related encoding genes.

(iv) The 'Search' page is used to search for genes, compounds or any biological term that match to a particular keyword. For example, if 'sulfur' keyword is searched, all entries (Genes, Putative genes, Potential genes, SCC homologs, Compounds, Publications) in SuCComBase that contain 'sulfur' will appear.

(v) The 'Datasets' dropdown menu provides links to all data sets in SuCComBase. These Dataset tabs are located at the header of SuCComBase to help users navigate the database.

(vi) The 'Download' page provides access to current and archived data sources in SuCComBase.

(vii) The 'Help' menu contains manual of SuCComBase, database schema, data sources and all references used to retrieve the information in SuCComBase. All scientific terms, definition and FAQs that are related to SuCComBase were also provided in the 'Help' page.

(viii) The 'Contact' menu provides information on the SuCComBase developer contacts and email address.

Conclusion and future work

SuCComBase is publicly available online at http://plant-scc.org and will be periodically updated. Currently, SuCComBase is the first and only database that provides the information on SCCs that are related to plant defense system in Brassicales. All information provided in this database is important to plant scientists, synthetic biologists, systems biologists, chemists or anyone who are interested working on the secondary metabolites or potential compounds, as well as those who study plant–host interactions specifically in Brassicales crops, hence continuous comprehensive cataloguing and curation is a priority.

Acknowledgements

Funding

References

1. Scharf,D.H., Heinekamp,T. and Brakhage,A.A. (2014) Human and plant fungal pathogens: the role of secondary metabolites. *PLoS Pathog.*, **10**, 10–12.

2. Hartmann,T. (2008) The lost origin of chemical ecology in the late 19th century. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 4541–4546.

3. Miedes,E., Vanholme,R., Boerjan,W. *et al.* (2014) The role of the secondary cell wall in plant resistance to pathogens. *Front. Plant Sci.*, **5**, 1–14.

4. Bednarek,P. (2012) Chemical warfare or modulators of defence responses—the function of secondary metabolites in plant immunity. *Curr. Opin. Plant Biol.*, **15**, 407–414.

5. Piasecka,A., Jedrzejczak-Rey,N. and Bednarek,P. (2015) Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. *New Phytol.*, **206**, 948–964.

6. Voigt,C.A. (2014) Callose-mediated resistance to pathogenic intruders in plant defense-related papillae. *Front. Plant Sci.*, **5**, 1–6.

7. Moghe,G. and Last,R.L. (2015) Something old, something new: conserved enzymes and the evolution of novelty in plant specialized metabolism. *Plant Physiol.*, **169**, 1512–1523.

8. Guerriero,G., Berni,R., Muñoz-Sanchez,J.A. *et al.* (2018) Production of plant secondary metabolites: examples, tips and suggestions for biotechnologists. *Genes (Basel)*, **9**, 1–22.

9. Wink,M. (2018) Plant secondary metabolites modulate insect behavior-steps toward addiction? *Front. Physiol.*, **9**, 1–9.

10. Noel,J.P., Austin,M.B. and Bomati,E.K. (2005) Structure-function relationships in plant phenylpropanoid biosynthesis. *Curr. Opin. Plant Biol.*, **8**, 249–253.

11. Bednarek,P. (2012) Sulfur-containing secondary metabolites from Arabidopsis thaliana and other Brassicaceae with function in plant immunity. *ChemBioChem*, **13**, 1846–1859.

12. Patra,B., Schluttenhofer,C., Wu,Y. *et al.* (2013) Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim. Biophys. Acta*, **1829**, 1236–1247.

13. Burow, M., Wittstock, U. and Gershenzon, J. (2008) In: Hell, R.; Dahl, C.; Knaff, D.; Leustek, T. (eds). *Sulfur Metabolism in Phototrophic Organisms*, Vol. **27**. Switzerland: SpringerLink, pp. 201–222.

14. Pedras,M.S.C., Yaya,E.E. and Glawischnig,E. The phytoalexins from cultivated and wild crucifers: chemistry and biology. *Nat. Prod. Rep.*, **28**, 1381–1405.

15. Cargnel,M.D., Demkura,P.V. and Ballaré,C.L. (2014) Linking phytochrome to plant immunity: low red: far-red ratios increase Arabidopsis susceptibility to Botrytis cinerea by reducing the biosynthesis of indolic glucosinolates and camalexin. *New Phytol.*, **204**, 342–354.

16. Pedras,M.S.C., Minic,Z. and Abdoli,A. (2014) The phytoalexin camalexin induces fundamental changes in the proteome of Alternaria brassicicola different from those caused by brassinin. *Fungal Biol.*, **118**, 83–93.

17. Wu,X., Kassie,F. and Mersch-Sundermann,V. (2005) Induction of apoptosis in tumor cells by naturally occurring sulfur-containing compounds. *Mutat. Res.*, **589**, 81–102.

18. Warin,R., Xiao,D., Arlotti,J.A. *et al.* (2010) Inhibition of human breast cancer xenograft growth by cruciferous vegetable constituent benzyl isothiocyanate. *Mol. Carcinog.*, **49**, 500–507.

19. Warin,R., Chambers,W.H., Potter,D.M. *et al.* (2009) Prevention of mammary carcinogenesis in MMTV-neu mice by cruciferous vegetable constituent benzyl isothiocyanate. *Cancer Res.*, **69**, 9473–9480.

20. Pintão,A.M., Pais,M.S., Coley,H. *et al.* (1995) In vitro and in vivo antitumor activity of benzyl isothiocyanate: a natural product from *Tropaeolum majus*. *Planta Med.*, **61**, 233–236.

21. Wattenberg,L.W. (1981) Inhibition of carcinogen-induced neoplasia by sodium cyanate, tert-butyl isocyanate, and benzyl isothiocyanate administered subsequent to carcinogen exposure. *Cancer Res.*, **41**, 2991–2994.

22. Rao,C.V. (2013) Benzyl isothiocyanate: double trouble for breast cancer cells. *Cancer Prev. Res.*, **6**, 760–763.

23. Zhang,Y., Munday,R., Jobson,H.E. *et al.* (2006) Induction of GST and NQO1 in cultured bladder cells and in the urinary bladders of rats by an extract of broccoli (Brassica oleracea italica) sprouts. *J. Agric. Food Chem.*, **54**, 9370–9376.

24. Tang,L., Zirpoli,G.R., Guru,K. *et al.* (2010) Intake of cruciferous vegetables modifies bladder cancer survival. *Cancer Epidemiol. Biomarkers Prev.*, **19**, 1806–1811.

25. Tang,L., Zirpoli,G.R., Guru,K. *et al.* (2008) Consumption of raw cruciferous vegetables is inversely associated with bladder cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **17**, 938–944.

26. Assayed,M.E. and Abd El-Aty,A.M. (2009) Cruciferous plants: phytochemical toxicity versus cancer chemoprotection. *Mini Rev. Med. Chem.*, **9**, 1470–1478.

27. Nowicki,D., Rodzik,O., Herman-Antosiewicz,A. *et al.* (2016) Isothiocyanates as effective agents against enterohemorrhagic Escherichia coli: insight to the mode of action. *Sci. Rep.*, **6**, 1–12.

28. Conaway,C.C., Wang,C.-X., Pittman,B. *et al.* (2005) Phenethyl isothiocyanate and sulforaphane and their N-acetylcysteine conjugates inhibit malignant progression of lung adenomas induced by tobacco carcinogens in A/J mice. *Cancer Res.*, **65**, 8548–8557.

29. Chung,F.-L., Conaway,C.C., Rao,C.V. *et al.* (2000) Chemoprevention of colonic aberrant crypt foci in Fischer rats by sulforaphane and phenethyl isothiocyanate. *Carcinogenesis*, **21**, 2287–2291.

30. Zhang,Y., Kensler,T.W., Cho,C.-G. *et al.* (1994) Anticarcinogenic activities of sulforaphane and structurally related synthetic norbornyl isothiocyanates. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 3147–3150.

31. Gills,J.J., Jeffery,E.H., Matusheski,N.V. *et al.* (2006) Sulforaphane prevents mouse skin tumorigenesis during the stage of promotion. *Cancer Lett.*, **236**, 72–79.

32. Fahey,J.W., Haristoy,X., Dolan,P.M. *et al.* (2002) Sulforaphane inhibits extracellular, intracellular, and antibiotic-resistant strains of Helicobacter pylori and prevents benzo[a]pyrene-induced stomach tumors. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 7610–7615.

33. Myzak,M.C., Dashwood,W.M., Orner,G.A. *et al.* (2006) Sulforaphane inhibits histone deacetylase in vivo and suppresses tumorigenesis in Apc-minus mice. *FASEB J.*, **20**, 506–508.

34. Singh,A.V., Xiao,D., Lew,K.L. *et al.* (2004) Sulforaphane induces caspase-mediated apoptosis in cultured PC-3 human prostate cancer cells and retards growth of PC-3 xenografts in vivo. *Carcinogenesis*, **25**, 83–90.

35. Kallifatidis,G., Rausch,V., Baumann,B. *et al.* (2009) Sulforaphane targets pancreatic tumour-initiating cells by NF- B-induced antiapoptotic signalling. *Gut*, **58**, 949–963.

36. Liu,C.-M., Peng,C.-Y., Liao,Y.-W. *et al.* (2017) Sulforaphane targets cancer stemness and tumor initiating properties in oral squamous cell carcinomas via miR-200c induction. *J. Formos. Med. Assoc.*, **116**, 41–48.

37. Henniges-Janssen,K., Heckel,D.G. and Groot,A.T. (2014) Preference of diamondback moth larvae for novel and original host plant after host range expansion. *Insects*, **5**, 793–804.

38. Moreira,L.F., Teixeira,N.C., Santos,N.A. *et al.* (2016) Diamondback moth performance and preference for leaves of Brassica oleracea of different ages and strata. *J. Appl. Entomol.*, **140**, 627–635.

39. Robin,A.H.K., Hossain,M.R., Park,J.-I. *et al.* (2017) Glucosinolate profiles in cabbage genotypes influence the preferential feeding of diamondback moth (*Plutella xylostella*). *Front. Plant Sci.*, **8**, 1–13.

40. Santolamazza-Carbone,S., Sotelo,T., Velasco,P. *et al.* (2016) Antibiotic properties of the glucosinolates of *Brassica oleracea* var. *acephala* similarly affect generalist and specialist larvae of two lepidopteran pests. *J. Pest Sci.*, **89**, 195–206.

41. Santolamazza-Carbone,S., Velasco,P., Soengas,P. *et al.* (2014) Bottom-up and top-down herbivore regulation mediated by glucosinolates in *Brassica oleracea* var. *acephala*. *Oecologia*, **174**, 893–907.

42. Mueller,L.A., Zhang,P. and Rhee,S.Y. (2003) AraCyc: a biochemical pathway database for Arabidopsis. *Plant Physiol.*, **132**, 453–460.

43. Kanehisa,M., Sato,Y., Kawashima,M. *et al.* (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.

44. Nakamura,Y., Mochamad Afendi,F., Kawsar Parvin,A. *et al.* (2014) KNApSAcK metabolite activity database for retrieving the relationships between metabolites and biological activities. *Plant Cell Physiol.*, **55**, 1–9.

45. Kim,S., Thiessen,P.A., Bolton,E.E. *et al.* (2016) PubChem substance and compound databases. *Nucleic Acids Res.*, **44**, D1202–D1213.

46. Brown,G.R., Hem,V., Katz,K.S. *et al.* (2015) Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.*, **43**, D36–D42.

47. Bateman,A., Martin,M.J., O'Donovan,C. *et al.* (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, D158–D169.

48. Kersey,P.J., Allen,J.E., Armean,I. *et al.* (2016) Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res.*, **44**, D574–D580.

49. Aoki,Y., Okamura,Y., Tadaka,S. *et al.* (2016) ATTED-II in 2016: a plant coexpression database towards special online collection. *Plant Cell Physiol.*, **57**, 1–9.

50. Lee,T., Yang,S., Kim,E. *et al.* (2014) AraNet v2: an improved database of co-functional gene networks for the study of *Arabidopsis thaliana* and 27 other nonmodel plant species. *Nucleic Acids Res.*, **43**, D996–D1002.

51. Warde-Farley,D., Donaldson,S.L., Comes,O. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, **38**, W214–W220.

52. Goodstein,D.M., Shu,S., Howson,R. *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.*, **40**, 1178–1186.

53. Altschul,S.F., Gish,W., Miller,W. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

54. Blake,J.A., Christie,K.R., Dolan,M.E. *et al.* (2015) Gene ontology consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.