



OPEN

Integration of genome-wide association studies and gene coexpression networks unveils promising soybean resistance genes against five common fungal pathogens

Fabricio Almeida-Silva[✉] & Thiago M. Venancio[✉]

Soybean is one of the most important legume crops worldwide. However, soybean yield is dramatically affected by fungal diseases, leading to economic losses of billions of dollars yearly. Here, we integrated publicly available genome-wide association studies and transcriptomic data to prioritize candidate genes associated with resistance to *Cadophora gregata*, *Fusarium graminearum*, *Fusarium virguliforme*, *Macrophomina phaseolina*, and *Phakopsora pachyrhizi*. We identified 188, 56, 11, 8, and 3 high-confidence candidates for resistance to *F. virguliforme*, *F. graminearum*, *C. gregata*, *M. phaseolina* and *P. pachyrhizi*, respectively. The prioritized candidate genes are highly conserved in the pangenome of cultivated soybeans and are heavily biased towards fungal species-specific defense responses. The vast majority of the prioritized candidate resistance genes are related to plant immunity processes, such as recognition, signaling, oxidative stress, systemic acquired resistance, and physical defense. Based on the number of resistance alleles, we selected the five most resistant accessions against each fungal species in the soybean USDA germplasm. Interestingly, the most resistant accessions do not reach the maximum theoretical resistance potential. Hence, they can be further improved to increase resistance in breeding programs or through genetic engineering. Finally, the coexpression network generated here is available in a user-friendly web application (<https://soyfungigcn.venanciogroup.uenf.br/>) and an R/Shiny package (<https://github.com/almeidasilvaf/SoyFungiGCN>) that serve as a public resource to explore soybean-pathogenic fungi interactions at the transcriptional level.

Soybean (*Glycine max* (L.) Merr.) is a major legume crop worldwide, contributing to global food security and economy. However, soybean yield is significantly affected by diseases, with an estimated economic loss of 95.8 billion dollars from 1996 to 2006 in the US¹. Most of the yield loss has been linked to foliar and stem/root diseases, which are mostly caused by phytopathogenic fungi¹. Fungal diseases, such as sudden death syndrome, Fusarium wilt, brown stem rot and asian rust, can impact soybean crops through leaf damage, necrosis, chlorosis, and death¹⁻³.

Over the past decade, several genome-wide association studies (GWAS) have uncovered multiple single-nucleotide polymorphisms (SNPs) associated with resistance to pathogenic fungi in soybean populations³⁻⁹. Nevertheless, GWAS often fail to accurately pinpoint the causative genes¹⁰. GWAS limitations are particularly challenging for self-pollinating plants (e.g., soybean) because of limited recombination and strong linkage disequilibrium between causative and non-causative variants¹¹. Such limitations ultimately lead to large genetic intervals with several genes, hindering causative gene identification. Because of the exponential accumulation of genomic and transcriptomic data in public databases¹²⁻¹⁶, integrative analyses to prioritize candidate genes have become a promising approach. This strategy consists in investigating the transcriptional patterns of all the genes near a significant SNP. Hence, the combination of multiple sources of evidence can result in richer and

Laboratório de Química e Função de Proteínas e Peptídeos, Centro de Biociências e Biotecnologia, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Av. Alberto Lamego 2000, P5, sala 217, Campos dos Goytacazes, RJ, Brazil. ✉email: fabricio_almeidasilva@hotmail.com; thiago.venancio@gmail.com

References	Pathogen	Resistance SNPs
8	<i>F. graminearum</i>	12
32	<i>F. virguliforme</i>	8
9	<i>C. gregata</i> / <i>F. virguliforme</i> / <i>P. pachyrhizi</i>	2/1/2
7	<i>F. virguliforme</i>	32
33	<i>F. virguliforme</i>	27
34	<i>M. phaseolina</i>	4
35	<i>M. phaseolina</i>	12
3	<i>C. gregata</i>	7

Table 1. GWAS included in this work.

narrower sets of high-confidence candidate genes for downstream experimental validation towards biotechnological applications.

Here, we integrated multiple publicly available RNA-seq and GWAS datasets to identify high-confidence candidate genes for resistance to five phytopathogenic fungi. The prioritized resistance genes are species-specific and highly conserved in the pangenome of cultivated soybeans. The candidate resistance genes against each species are involved in various immunity-related processes, such as recognition, signaling, oxidative stress, and apoptosis. Finally, we highlighted the five most resistant accessions against each fungal species in the USDA germplasm, uncovering important information for breeding programs and genetic engineering initiatives. Finally, the co-expression network resulting from this work was also made available as a publicly available web application (<https://soyfungigcn.venanciogroup.uenf.br/>) and R/Shiny package (<https://github.com/almeidasilva/SoyFungiGCN>).

Materials and methods

Curator of resistance-associated SNPs and pan-genome data. SNPs that contribute to resistance against phytopathogenic fungi were manually curated from the scientific literature (Table 1; Supplementary Table S1). SNPs that were identified using the Gmax_a1.v1 genome were converted to their corresponding sites in the Gmax_a2.v1 assembly using the .vcf files for both assemblies available at Soybase¹⁷. A matrix of gene presence/absence variation (PAV) in the pan-genome of cultivated soybeans ($n = 204$ genomes from 24 countries and 5 continents) was obtained from the Supplementary Data in¹⁸.

Transcriptome data. Gene expression estimates in transcripts per million mapped reads (TPM, Kallisto estimation) were retrieved from the Soybean Expression Atlas¹⁹. Additional RNA-seq samples comprising soybean tissues infected with fungal pathogens were retrieved from a recent publication from our group²⁰. We filtered the SNP and transcriptome datasets to keep only fungal species that were represented by both data sources. A total of 150 RNA-seq samples from soybean tissues infected with fungal pathogens were selected (Supplementary Table S2). Finally, genes with median expression values lower than 5 were excluded to attenuate noise, resulting in an $18,748 \times 150$ gene expression matrix for downstream analyses.

Selection of guide genes. MapMan annotations for soybean genes were retrieved from the PLAZA 3.0 Dicots database²¹. Genes assigned to defense-related pathways (e.g., pathogenesis-related proteins, lignin biosynthesis, oxidative stress, and phytohormone regulation) were used as guides (Supplementary Table S3).

Candidate gene mining and functional analyses. Gene expression data were adjusted for confounding artifacts and quantile normalized with the R package BioNERO²². An unsigned coexpression network was inferred with BioNERO using Pearson's r as correlation. All genes located in a 2 Mb sliding window relative to each SNP were selected as putative candidates, as previously proposed²³. Candidate genes were prioritized using the algorithm implemented in the R package cageminer²⁴, with an r_{pb} threshold of 0.2 for gene significance (gene-trait correlation). Enrichment analyses were also performed with BioNERO, using functional annotations from the PLAZA 4.0 database²⁵. To rank the prioritized candidates, they were given scores using the formula:

$$S = r_{pb}\kappa$$

where

r_{pb} = point-biserial correlation coefficient (cageminer algorithm)

$\kappa = 2$ if the gene is a transcription factor

$\kappa = 2$ if the gene is a hub

$\kappa = 3$ if the gene is a hub and a transcription factor

$\kappa = 1$ if the gene is neither a hub nor a transcription factor.

Selection of most resistant accessions from the USDA germplasm. The VCF file with genotypic information for all accessions in the USDA germplasm was downloaded from Soybase¹⁷. For each locus, scores 0, 1, or 2 were attributed if accessions had 0, 1, or 2 beneficial SNPs (effect size > 0), respectively, whereas scores

2, 1, or 0 were attributed if accessions had 0, 1, or 2 deleterious SNPs (effect size < 0). Total resistance scores for each accession were calculated as the sum of scores S_i for all n loci as follows:

$$S_{total} = \sum_{i=1}^n S_i, \text{ where } S_i = \{0, 1, 2\}$$

Total resistance scores were ranked from highest to lowest, and ranks were used to select the most resistant accessions. The resistance potential of the best accessions was calculated as a ratio of the attributed scores to the theoretical maximum score (all beneficial SNPs and no deleterious SNPs).

Results and discussion

Data summary and genomic distribution of SNPs. After filtering the datasets to keep only fungal species represented by both SNP and transcriptome information, we kept five common phytopathogenic fungi: *Cadophora gregata*, *Fusarium graminearum*, *Fusarium virguliforme*, *Macrophomina phaseolina*, and *Phakopsora pachyrhizi* (Fig. 1A). Overall, SNPs were located in gene-rich regions of the genome (Fig. 1B). SNPs were unevenly distributed across chromosomes, except for *F. virguliforme* (Fig. 1C). Further, we found that most SNPs were located in intergenic regions (Fig. 1D). Hence, predicting SNP effect on genes would not be suitable for this trait.

Candidate gene mining reveals a highly species-specific immune response. Using defense-related genes as guides, the cageminer algorithm identified 188, 56, 11, 8, and 3 high-confidence genes for *F. virguliforme*, *F. graminearum*, *C. gregata*, *M. phaseolina*, and *P. pachyrhizi*, respectively (Fig. 2). Only three genes were shared between species, revealing a high specificity in plant-pathogen interactions for these species. The three genes are shared by *F. virguliforme* and *F. graminearum*, suggesting that some conservation can occur at the genus level, but not at other broader taxonomic levels.

The specificity of resistance genes to particular species has been widely reported^{26–29}. This phenomenon imposes a challenge for biotechnological applications, as it requires pyramiding many different genes to render elite cultivars resistant to different pathogens. However, we cannot rule out that the species-specific trend we observed results from low diversity in the association panels in the GWAS we analyzed. Additionally, as SNP and transcriptome data are not available for multiple pathogen strains, we might overlook broad-spectrum resistance genes that confer resistance to multiple strains of the same species²⁷.

Further, we manually curated the high-confidence candidate resistance genes to predict the putative role of their products in plant immunity (Supplementary Table S4). Most of the prioritized candidates (28%) encode proteins involved in immune signaling, although this does not apply to all fungi species (Fig. 3). The main discrepancy in the functional classification of candidates was observed for candidate resistance genes against *P. pachyrhizi*. However, this is likely due to sampling bias, as the number of SNPs associated with resistance to *P. pachyrhizi* is limited as compared to other species. Candidates also encode proteins that play a role in recognition, phytohormone metabolism, systemic acquired resistance, transport, transcriptional regulation, oxidative stress, apoptosis, physical defense, and direct function against fungi (Fig. 3).

Interestingly, 21 candidate genes lack functional description and, hence, we could not infer their roles in plant immunity ($n = 2, 4, 14$, and 1 for *C. gregata*, *F. virguliforme*, and *P. pachyrhizi*, respectively). Nevertheless, as they were identified as high-confidence candidate genes, we hypothesize that they encode defense-related proteins. This finding reveals that besides the identification of high-confidence candidate genes, our algorithm can serve as a network-based approach to predict functions of unannotated genes, similar to previous approaches^{30,31}.

We also developed a scheme that was used to rank high-confidence candidate genes (Table 2). Ranking candidates is particularly useful to prioritize genes when there are several candidates, such as for *F. virguliforme* and *F. graminearum*. Here, we suggest using the top 10 candidate resistance genes against each pathogen for experimental validation in future studies. Experimental tests with transgenic or edited soybeans using our set of target genes will likely reveal which genes are more suitable to develop soybean lines with increased resistance to each fungal disease.

Pangenome presence/absence variation analysis demonstrates that most prioritized genes are core genes.

We analyzed PAV patterns for our prioritized candidate genes in the recently published pangenome of cultivated soybeans to unveil which soybean genotypes contain prioritized candidate genes and explore gene presence/absence variation patterns across genomes¹⁸. We found that most candidates are present in all 204 accessions (Supplementary Fig. 1A). This trend is not surprising, as the gene content in this pangenome is highly conserved, with ~91% of the genes being shared by >99% of the genomes. Although the variable genome is enriched in genes associated with defense, signaling, and plant development, this trend was not found in our gene set.

Further, we investigated if gene PAV patterns could be explained by the geographical origins of the accessions (Supplementary Fig. 1B). We observed no clustering by geographical origin, suggesting that gene PAV is not affected by population structure. As this pangenome is comprised of improved soybean accessions¹⁸, the lack of population structure effect can be due to breeding programs targeting optimal adaptation to different environmental conditions (e.g., latitude and climate), even if they are in the same country.

Screening of the USDA germplasm reveals a room for genetic improvement. We inspected the USDA germplasm to find the top 5 most resistant genotypes against each fungal pathogen (see Materials and Methods for details). Strikingly, the most resistant genotypes do not contain all resistance alleles, revealing that,

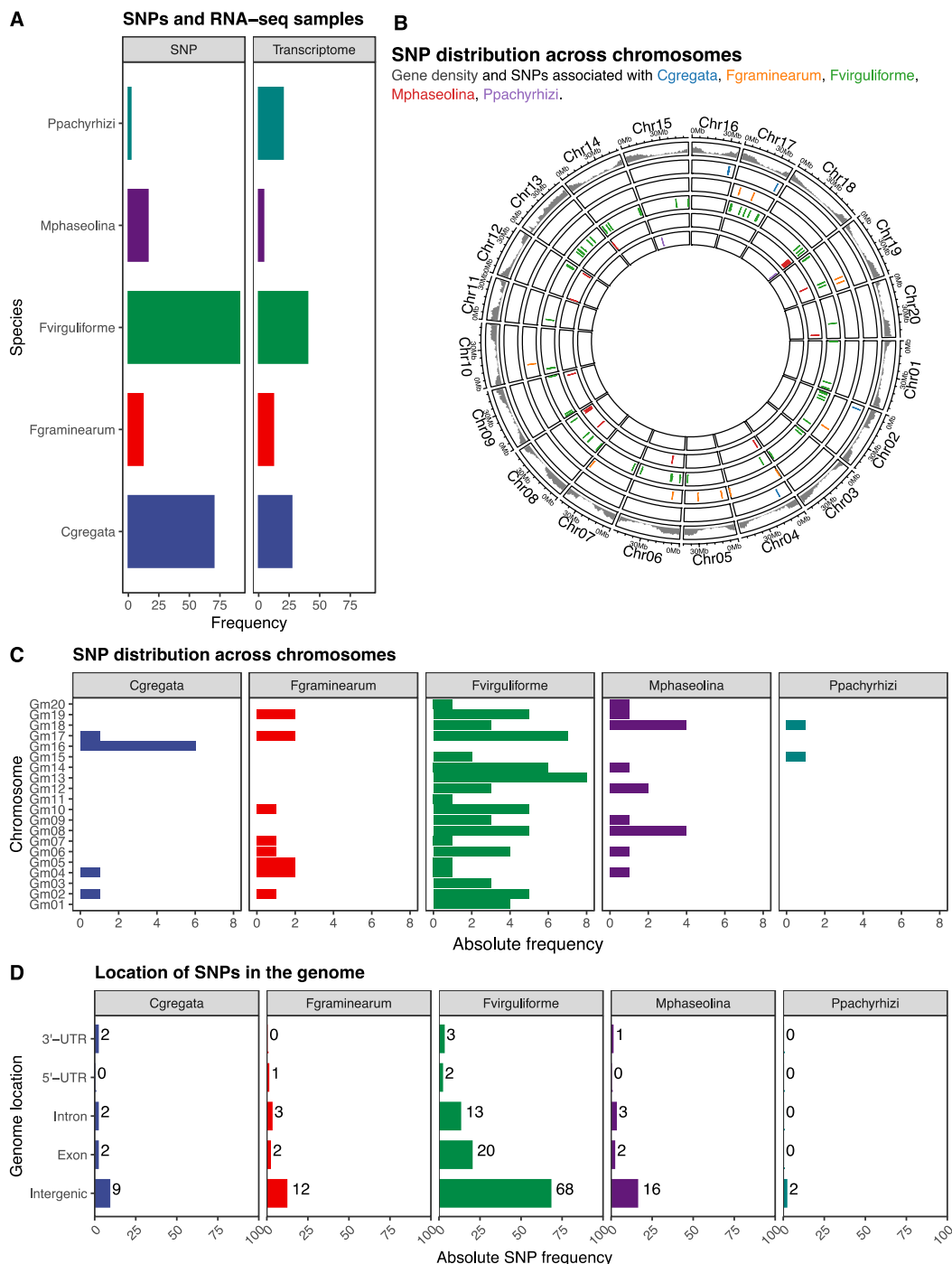


Figure 1. Data summary and genomic distribution of SNPs. (A) Frequency of SNPs and RNA-seq samples included in this study. (B) Genomic coordinates of resistance SNPs against each fungal pathogen. The outer track represents gene density, whereas inner tracks represent the SNP positions for each species. (C) SNP distribution across chromosomes. Overall, there is an uneven distribution of SNPs across chromosomes. (D) Genomic location of SNPs. Most SNPs are located in intergenic regions.

theoretically, they could be further improved to increase resistance (Table 3). All resistance-associated SNPs against *P. pachyrhizi* are present in some accessions, but this is because only two SNPs have been reported for this species. Additionally, none of the reported SNPs for *F. graminearum* have been identified in the SoySNP50k collection. Hence, we could not predict the most resistant accessions to this fungal pathogen in the USDA germplasm.

Although some individual genes can confer full race-specific resistance to some pathogens, their durability in the field is often short because of pathogen evolution²⁷. Thus, pyramiding quantitative trait loci (QTL) that confer partial resistance has been proposed as a strategy to confer long-term resistance²⁸. To accomplish this, the

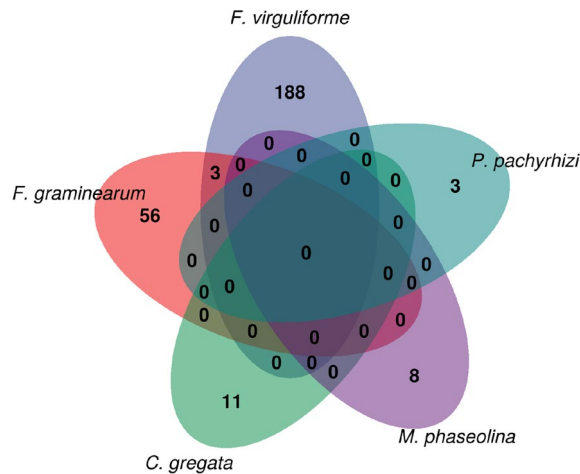


Figure 2. Venn diagram of prioritized candidate resistance genes against each species. The diagram demonstrates a high species-specific response to each pathogen, as genes are mostly not shared. Only three genes are shared between *F. graminearum* and *F. virguliforme*, suggesting some conservation at the genus level.

most resistant genotypes identified here can be targets of allele pyramiding in breeding programs using marker-assisted selection. Alternatively, these genotypes might have their genomes edited with CRISPR/Cas systems to introduce beneficial alleles or remove deleterious alleles, ultimately boosting resistance.

Development of a user-friendly web application for network exploration. To facilitate network exploration and data reuse, we developed a user-friendly web application named SoyFungiGCN (<https://soyfungigcn.venanciogroup.uenf.br/>). Users can input a soybean gene of interest (Wm82.a2.v1 assembly) and visualize the gene's module, scaled intramodular degree, and hub status (Fig. 4A). Additionally, users can explore enriched GO terms, Mapman bins and/or Interpro domains associated with the input gene's module (Fig. 4A). Users can also visualize a network plot with the input gene and its coexpression neighbors (Fig. 4B). This resource can be particularly useful for researchers studying soybean response to other fungal species, as they can check if their genes of interest are located in defense-related coexpression modules. Also, researchers studying other species can verify if the soybean ortholog of their genes of interest is located in a defense-related module. The application is also available as an R package named SoyFungiGCN (<https://github.com/almeidasilvaf/SoyFungiGCN>). This package lets users run the application locally as a Shiny app, ensuring the application will always be available, even in case of server downtime.

Conclusions

By integrating publicly available GWAS and RNA-seq data, we found promising candidate genes in soybean associated with resistance to five common phytopathogenic fungi, namely *C. gregata*, *F. graminearum*, *F. virguliforme*, *M. phaseolina*, and *P. pachyrhizi*. The prioritized candidates encode proteins that play a role immunity-related processes such as in recognition, signaling, transcriptional regulation, oxidative stress, and physical defense. We have also found the top 5 most resistant soybean accessions against each fungal species and hypothesize that they can be further genetically improved in breeding programs with marker-assisted selection or through genome editing. The coexpression network generated here was also made available in a web resource and R package to help in future studies on soybean-pathogenic fungi interactions.

Data availability

All data and code used in this study are available in our GitHub repository (https://github.com/almeidasilvaf/SoyFungi_GWAS_GCIN) to ensure full reproducibility.

Received: 7 October 2021; Accepted: 3 December 2021

Published online: 27 December 2021

References

- Bandara, A. Y., Weerasooriya, D. K., Bradley, C. A., Allen, T. W. & Esker, P. D. Dissecting the economic impact of soybean diseases in the United States over two decades. *PLoS ONE* **15**(4), 1–28. <https://doi.org/10.1371/journal.pone.0231141> (2020).
- Pandey, A. K. *et al.* Functional analysis of the asian soybean rust resistance pathway mediated by Rpp2. *Mol. Plant-Microbe Interact.* **24**(2), 194–206. <https://doi.org/10.1094/MPMI-08-10-0187> (2011).
- Rincker, K., Lipka, A. E. & Diers, B. W. Genome-wide association study of brown stem rot resistance in soybean across multiple populations. *Plant Genome* <https://doi.org/10.3835/plantgenome2015.08.0064> (2016).
- Iquira, E., Humira, S. & François, B. Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. *BMC Plant Biol.* **15**(1), 1–12. <https://doi.org/10.1186/s12870-014-0408-y> (2015).

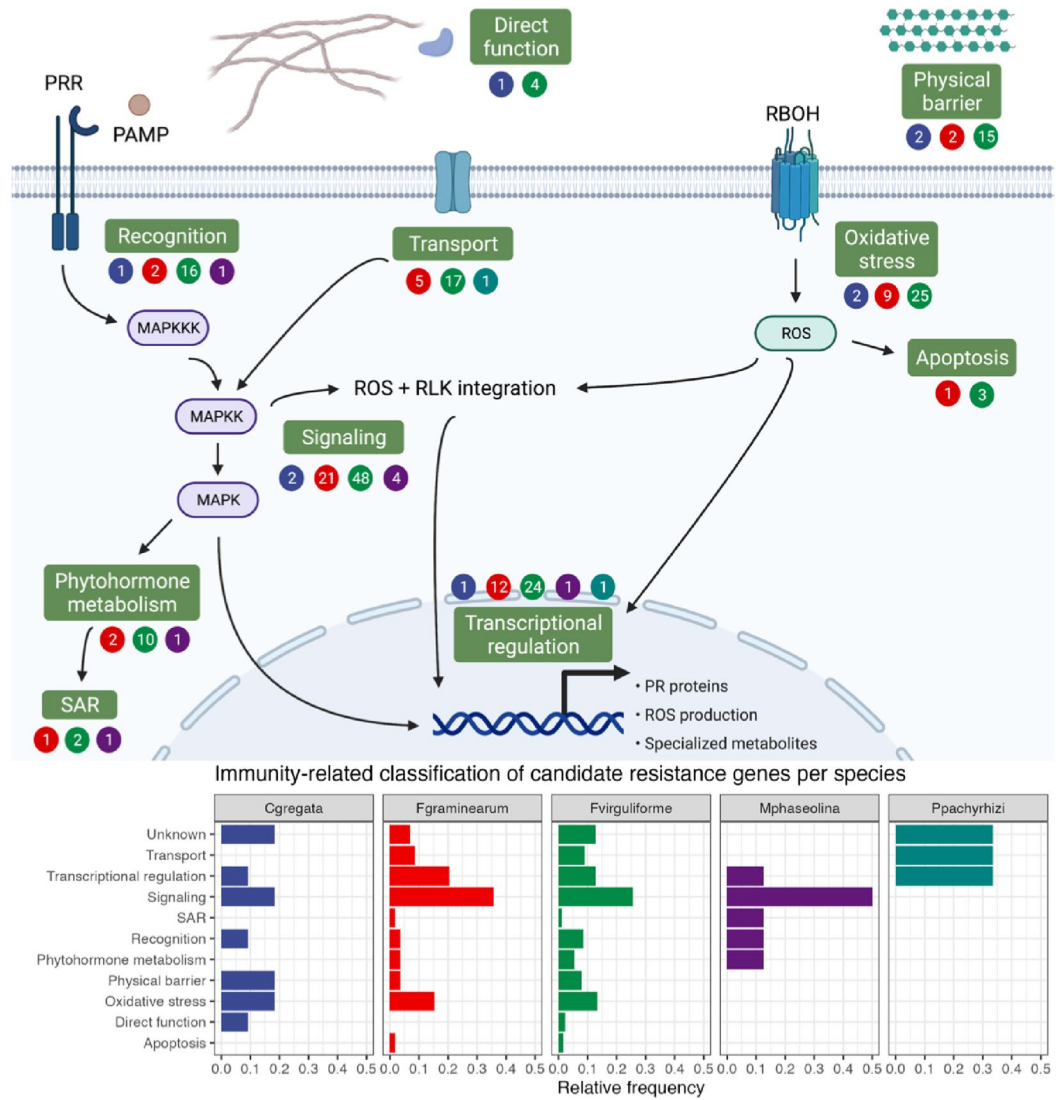


Figure 3. Prioritized candidate resistance genes and their putative role in plant immunity. Numbers in circles represent absolute frequencies of resistance genes against *C. gregata* (blue), *F. graminearum* (red), *F. virguliforme* (green), *M. phaseolina* (purple), and *P. pachyrhizi* (turquoise). PRR, pattern recognition receptor. PAMP, pathogen-associated molecular pattern. MAPKKK, mitogen-activated protein kinase kinase. MAPKK, mitogen-activated protein kinase kinase. MAPK, mitogen-activated protein kinase. SAR, systemic acquired resistance. RBOH, respiratory burst oxidase homolog. ROS, reactive oxygen species. RLK, receptor-like kinase. PR, pathogenesis-related. Figure designed with Biorender (biorender.com).

- Sun, M. *et al.* Genome-wide association study of partial resistance to sclerotinia stem rot of cultivated soybean based on the detached leaf method. *PLoS ONE* **15**(5), 1–15. <https://doi.org/10.1371/journal.pone.0233366> (2020).
- Kandel, R. *et al.* Soybean resistance to white mold: Evaluation of soybean germplasm under different conditions and validation of QTL. *Front. Plant Sci.* **9**(April), 1–12. <https://doi.org/10.3389/fpls.2018.00505> (2018).
- Zhang, J., Singh, A., Mueller, D. S. & Singh, A. K. Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. *Plant J.* **84**(6), 1124–1136. <https://doi.org/10.1111/tpj.13069> (2015).
- Zhang, C. *et al.* Loci and candidate genes in soybean that confer resistance to Fusarium graminearum. *Theor. Appl. Genet.* **132**(2), 431–441. <https://doi.org/10.1007/s00122-018-3230-3> (2019).
- Chang, H. X., Lipka, A. E., Domier, L. L. & Hartman, G. L. Characterization of disease resistance loci in the USDA soybean germplasm collection using genome-wide association studies. *Phytopathology* **106**(10), 1139–1151. <https://doi.org/10.1094/PHYTO-01-16-0042-FI> (2016).
- Baxter, I. We aren't good at picking candidate genes, and it's slowing us down. *Curr. Opin. Plant Biol.* **54**, 57–60. <https://doi.org/10.1016/j.pbi.2020.01.006> (2020).
- Michno, J. M., Liu, J., Jeffers, J. R., Stupar, R. M. & Myers, C. L. Identification of nodulation-related genes in *Medicago truncatula* using genome-wide association studies and co-expression networks. *Plant Direct* **4**(5), 1–10. <https://doi.org/10.1002/pld3.220> (2020).
- Schwartz, T. S. The promises and the challenges of integrating multi-omics and systems biology in comparative stress biology. *Integr. Comp. Biol.* **53**(9), 1689–1699. <https://doi.org/10.1017/CBO9781107415324.004> (2020).

Gene	Predicted function	Resistance to	Role
Glyma.16G170100	Cell wall biogenesis-related extensin 3	<i>C. gregata</i>	Physical barrier
Glyma.02G026700	Transcriptional repressor SIN3	<i>C. gregata</i>	Transcriptional regulation
Glyma.02G026900	Galacturonosyltransferase	<i>C. gregata</i>	Physical barrier
Glyma.02G029300	SAM domain-containing	<i>C. gregata</i>	Unknown
Glyma.16G155100	Aquaporin	<i>C. gregata</i>	Oxidative stress
Glyma.17G217000	Class V chitinase	<i>C. gregata</i>	Direct function
Glyma.17G213600	Calcium-binding EF hand	<i>C. gregata</i>	Signaling
Glyma.17G231800	Clathrin adaptor EPSIN1	<i>C. gregata</i>	Recognition
Glyma.02G047000	Thiosulfate sulfurtransferase/rhodanese	<i>C. gregata</i>	Oxidative stress
Glyma.16G150500	Unknown	<i>C. gregata</i>	Unknown
Glyma.17G087500	SOUL heme-binding protein	<i>F. graminearum</i>	Oxidative stress
Glyma.06G121300	GRAS transcription factor	<i>F. graminearum</i>	Transcriptional regulation
Glyma.05G070300	Tobamovirus multiplication 2A	<i>F. graminearum</i>	Recognition
Glyma.04G013500	BURP domain-containing protein	<i>F. graminearum</i>	Physical barrier
Glyma.06G105000	ERF/AP2 transcription factor	<i>F. graminearum</i>	Transcriptional regulation
Glyma.05G062400	2OG-Fe(II) oxygenase	<i>F. graminearum</i>	Oxidative stress
Glyma.05G063600	ERF/AP2 transcription factor	<i>F. graminearum</i>	Transcriptional regulation
Glyma.05G115700	RING domain ubiquitin E3 ligase	<i>F. graminearum</i>	Signaling
Glyma.17G116100	MAPK signaling-related protein	<i>F. graminearum</i>	Signaling
Glyma.05G103600	Peroxidase	<i>F. graminearum</i>	Oxidative stress
Glyma.13G081000	Nodulin-like amino acid transporter	<i>F. virguliforme</i>	Transport
Glyma.01G225600	Unknown	<i>F. virguliforme</i>	Unknown
Glyma.02G210500	bHLH transcription factor	<i>F. virguliforme</i>	Transcriptional regulation
Glyma.01G162500	BIG1 protein	<i>F. virguliforme</i>	Apoptosis
Glyma.17G061400	Peroxidase	<i>F. virguliforme</i>	Oxidative stress
Glyma.19G010100	HD-Zip transcription factor	<i>F. virguliforme</i>	Transcriptional regulation
Glyma.18G276800	Amino acid transporter	<i>F. virguliforme</i>	Oxidative stress
Glyma.05G209900	PLAC8 family protein	<i>F. virguliforme</i>	Apoptosis
Glyma.14G025100	Inositol-1,4,5-trisphosphate 5-phosphatase	<i>F. virguliforme</i>	Signaling
Glyma.19G117800	Unknown	<i>F. virguliforme</i>	Unknown
Glyma.20G203900	Type I serine/threonine protein phosphatase	<i>M. phaseolina</i>	Signaling
Glyma.08G316500	Calmodulin-dependent protein kinase	<i>M. phaseolina</i>	Signaling
Glyma.06G187200	R-gene-mediated resistance, lipase	<i>M. phaseolina</i>	SAR
Glyma.09G218600	Cytochrome P450, family 707, subfamily A	<i>M. phaseolina</i>	Phytohormone metabolism
Glyma.09G216800	Pectin acetyltransferase	<i>M. phaseolina</i>	Signaling
Glyma.20G216600	Dof-type transcription factor	<i>M. phaseolina</i>	Transcriptional regulation
Glyma.08G332800	Calcineurin B-like calcium sensor	<i>M. phaseolina</i>	Signaling
Glyma.18G301700	Leucine-rich repeat receptor kinase (LRR-RK)	<i>M. phaseolina</i>	Recognition
Glyma.15G125900	Magnesium transporter CorA-like	<i>P. pachyrhizi</i>	Transport
Glyma.18G286900	Unknown	<i>P. pachyrhizi</i>	Unknown
Glyma.15G123900	CBF1 interacting co-repressor CIR	<i>P. pachyrhizi</i>	Transcriptional regulation

Table 2. Top 10 candidate resistance genes against each fungal species and their putative roles in plant immunity. The predicted function for each gene was manually curated from the description of the best ortholog in *Arabidopsis thaliana*, using functional annotations from Soybase and TAIR.

- Deshmukh, R. *et al.* Integrating omic approaches for abiotic stress tolerance in soybean. *Front. Plant Sci.* **5**, 1–12. <https://doi.org/10.3389/fpls.2014.00244> (2014).
- Schaefer, R. J. *et al.* Integrating coexpression networks with GWAS to prioritize causal genes in maize. *Plant Cell* **30**(December), 2922–2942. <https://doi.org/10.1105/tpc.18.00299> (2018).
- Baker, R. L. *et al.* Integrating transcriptomic network reconstruction and eQTL analyses reveals mechanistic connections between genomic architecture and Brassica rapa development. *PLOS Genet.* **15**(9), e1008367. <https://doi.org/10.1371/journal.pgen.1008367> (2019).
- Wen, Z. *et al.* Integrating GWAS and gene expression data for functional characterization of resistance to white mould in soya bean. *Plant Biotechnol. J.* **16**(11), 1825–1835. <https://doi.org/10.1111/pbi.12918> (2018).
- Brown, A. V. *et al.* A new decade and new data at SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* **13**(3), 1–6. <https://doi.org/10.1093/nar/gkaa1107> (2020).
- Torkamaneh, D., Lemay, M.-A. & Belzile, F. The pan-genome of the cultivated soybean (pansoy) reveals an extraordinarily conserved gene content. *Plant Biotechnol. J.* **19**, 1852–1862. <https://doi.org/10.1111/pbi.13600> (2021).

Accession	Score	Potential	Species
PI594466	102	0.73	<i>C. gregata</i>
PI578477A	100	0.71	<i>C. gregata</i>
PI437571	100	0.71	<i>C. gregata</i>
PI567520A	100	0.71	<i>C. gregata</i>
PI274507	100	0.71	<i>C. gregata</i>
PI339871C	82	0.60	<i>F. virguliforme</i>
PI378694	80	0.59	<i>F. virguliforme</i>
PI407145	80	0.59	<i>F. virguliforme</i>
PI424107A	80	0.59	<i>F. virguliforme</i>
PI479753A	80	0.59	<i>F. virguliforme</i>
PI594760B	24	0.75	<i>M. phaseolina</i>
PI479752	24	0.75	<i>M. phaseolina</i>
PI603706A	24	0.75	<i>M. phaseolina</i>
PI603531A	24	0.75	<i>M. phaseolina</i>
PI603412A	24	0.75	<i>M. phaseolina</i>
PI603547	4	1	<i>P. pachyrhizi</i>
PI639559A	4	1	<i>P. pachyrhizi</i>
PI639559B	4	1	<i>P. pachyrhizi</i>
PI326582A	4	1	<i>P. pachyrhizi</i>
PI407057	4	1	<i>P. pachyrhizi</i>

Table 3. Top 5 most resistant soybean accessions against each fungal pathogen. Overall, the best genotypes do not reach the maximum potential. An exception is observed for *P. pachyrhizi*-resistant genotypes, but this is likely due to the small number of resistance SNPs. None of the resistance SNPs for *F. graminearum* have been identified in the USDA SoySNP50k compendium and, hence, we could not predict resistance potential against this species.

19. Machado, F. B. *et al.* Systematic analysis of 1,298 RNA-Seq samples and construction of a comprehensive soybean (*Glycine max*) expression atlas. *Plant J.* **103**, 1894–2190. <https://doi.org/10.1111/tpj.14850> (2020).
20. Almeida-Silva, F. & Venancio, T. M. Pathogenesis-related protein 1 (PR-1) genes in soybean: Genome-wide identification, structural analysis and expression profiling under multiple biotic and abiotic stresses. *Gene* **809**, 146013. <https://doi.org/10.1016/j.gene.2021.146013> (2022).
21. Proost, S. *et al.* PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res.* **43**(D1), D974–D981. <https://doi.org/10.1093/nar/gku986> (2015).
22. Almeida-Silva, F. & Venancio, T. M. BioNERO: an all-in-one R/Bioconductor package for comprehensive and easy biological network reconstruction. *Funct. Integr. Genom.* <https://doi.org/10.1007/s10142-021-00821-9> (2021).
23. Brodie, A., Azaria, J. R. & Ofran, Y. How far from the SNP may the causative genes be?. *Nucleic Acids Res.* **44**(13), 6046–6054. <https://doi.org/10.1093/nar/gkw500> (2016).
24. Almeida-Silva, F. & Venancio, T. M. cageminer: an R/Bioconductor package to prioritize candidate genes by integrating GWAS and gene coexpression networks. *bioRxiv* **54**, 57. <https://doi.org/10.1101/2021.08.04.455037> (2021).
25. Van Bel, M. *et al.* PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* **46**(D1), D1190–D1196. <https://doi.org/10.1093/nar/gkx1002> (2018).
26. Kourelis, J. & Van Der Hoorn, R. A. L. Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *Plant Cell* <https://doi.org/10.1105/tpc.17.00579> (2018).
27. Ning, Y. & Wang, G. L. Breeding plant broad-spectrum resistance without yield penalties. *Proc. Natl. Acad. Sci. USA* **115**(12), 2859–2861. <https://doi.org/10.1073/pnas.1801235115> (2018).
28. Li, W., Deng, Y., Ning, Y., He, Z. & Wang, G. L. Exploiting broad-spectrum disease resistance in crops: From molecular dissection to breeding. *Annu. Rev. Plant Biol.* **71**, 575–603. <https://doi.org/10.1146/annurev-arplant-010720-022215> (2020).
29. Durrant, W. E. & Dong, X. Systemic acquired resistance. *Annu. Rev. Phytopathol.* **42**, 185–209. <https://doi.org/10.1146/annurev.phyto.42.040803.140421> (2004).
30. Almeida-Silva, F., Moharana, K. C., Machado, F. B. & Venancio, T. M. Exploring the complexity of soybean (*Glycine max*) transcriptional regulation using global gene co-expression networks. *Planta* **252**, 1–12. <https://doi.org/10.1007/s00425-020-03499-8> (2020).
31. Depuydt, T. & Vandepoele, K. Multi-omics network-based functional annotation of unknown Arabidopsis genes. *Plant J.* **108**, 1198–1212. <https://doi.org/10.1111/tpj.15507> (2021).
32. Bao, Y., Kurle, J. E., Anderson, G. & Young, N. D. Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. *Mol. Breed.* **35**(6), 1–14. <https://doi.org/10.1007/s11032-015-0324-3> (2015).
33. Swaminathan, S. *et al.* Genome wide association study identifies novel single nucleotide polymorphic loci and candidate genes involved in soybean sudden death syndrome resistance. *PLoS ONE* **14**(2), 1–21. <https://doi.org/10.1371/journal.pone.0212071> (2019).
34. Vinholes, P., Rosado, R., Roberts, P., Borém, A. & Schuster, I. Single nucleotide polymorphism-based haplotypes associated with charcoal rot resistance in Brazilian soybean germplasm. *Agron. J.* **111**(1), 182–192. <https://doi.org/10.2134/agronj2018.07.0429> (2019).
35. Coser, S. M. *et al.* Genetic architecture of charcoal rot (*Macrophomina phaseolina*) resistance in soybean revealed using a diverse panel. *Front. Plant Sci.* **8**(September), 1–12. <https://doi.org/10.3389/fpls.2017.01626> (2017).

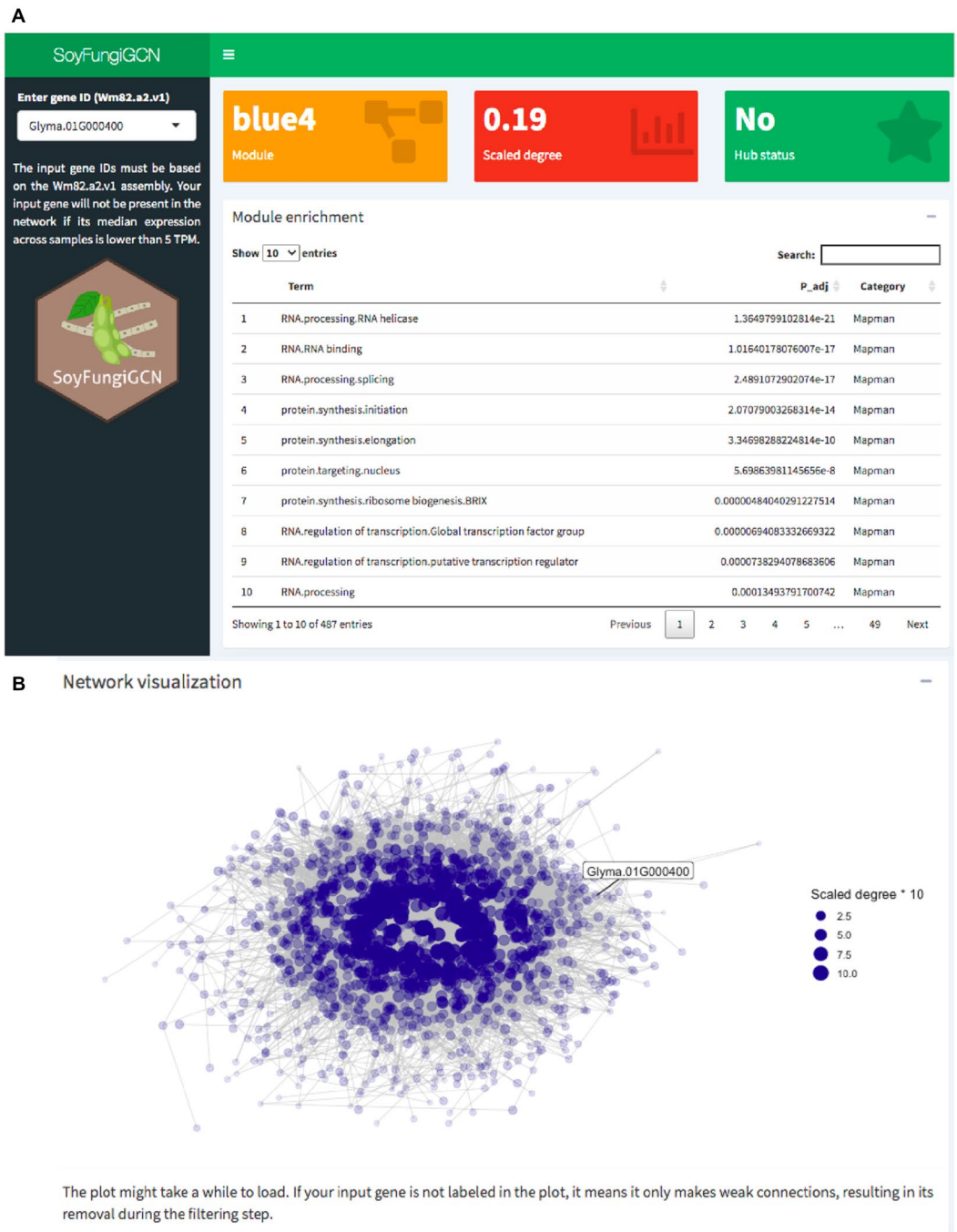


Figure 4. Functionalities in the SoyFungiGCN web application. **A.** Screenshot of the page users see when they access the application. In the sidebar, users can specify the ID of a gene of interest (Wm82.a2.v1 assembly). For each gene, users can see the gene's module (orange box), scaled degree (red box), hub gene status (green box), and an interactive table with enrichment results for MapMan bins, Interpro domains and Gene Ontology terms associated the gene's module. *P* values from enrichment results are adjusted for multiple testing with Benjamini–Hochberg correction. **B.** Network visualization plot. Users can optionally visualize the input gene and its position in the module by clicking the plus (+) icon in the “Network visualization” tab below the enrichment table. As the plot can take a few seconds to render (~2–5 s), it is hidden by default.

Acknowledgements

This work was supported by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ; grants E-26/203.309/2016 and E-26/203.014/2018), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES; Finance Code 001), and Conselho Nacional de Desenvolvimento Científico e

Tecnológico. The funding agencies had no role in the design of the study and collection, analysis, and interpretation of data and in writing.

Author contributions

Conceived the study: F.A.-S. and T.M.V. Data analysis: F.A.-S. Funding, project coordination and infrastructure: T.M.V. Manuscript writing: F.A.-S and T.M.V.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-03864-x>.

Correspondence and requests for materials should be addressed to F.A.-S. or T.M.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021