In silico structural and functional characterization of six potential virulence genes identified in *Clarireedia spp.*

Harshita Saxena[1](harshita.saxena@uga.edu), Rajiv K. Parvathaneni[1,2], Willis T. Spratling[2], Paul L. Raymer[1,3], Alfredo Martinez-Espinoza[2], Bochra A. Bahri[1,2]

[1]Institute of Plant Breeding, Genetics, and Genomics, University of Georgia, Griffin, GA, 30223; [2]Department of Plant Pathology, University of Georgia, Griffin, GA, 30223, USA; [3]Department of Crop and Soil Science, University of Georgia, Griffin, GA, 30223, USA

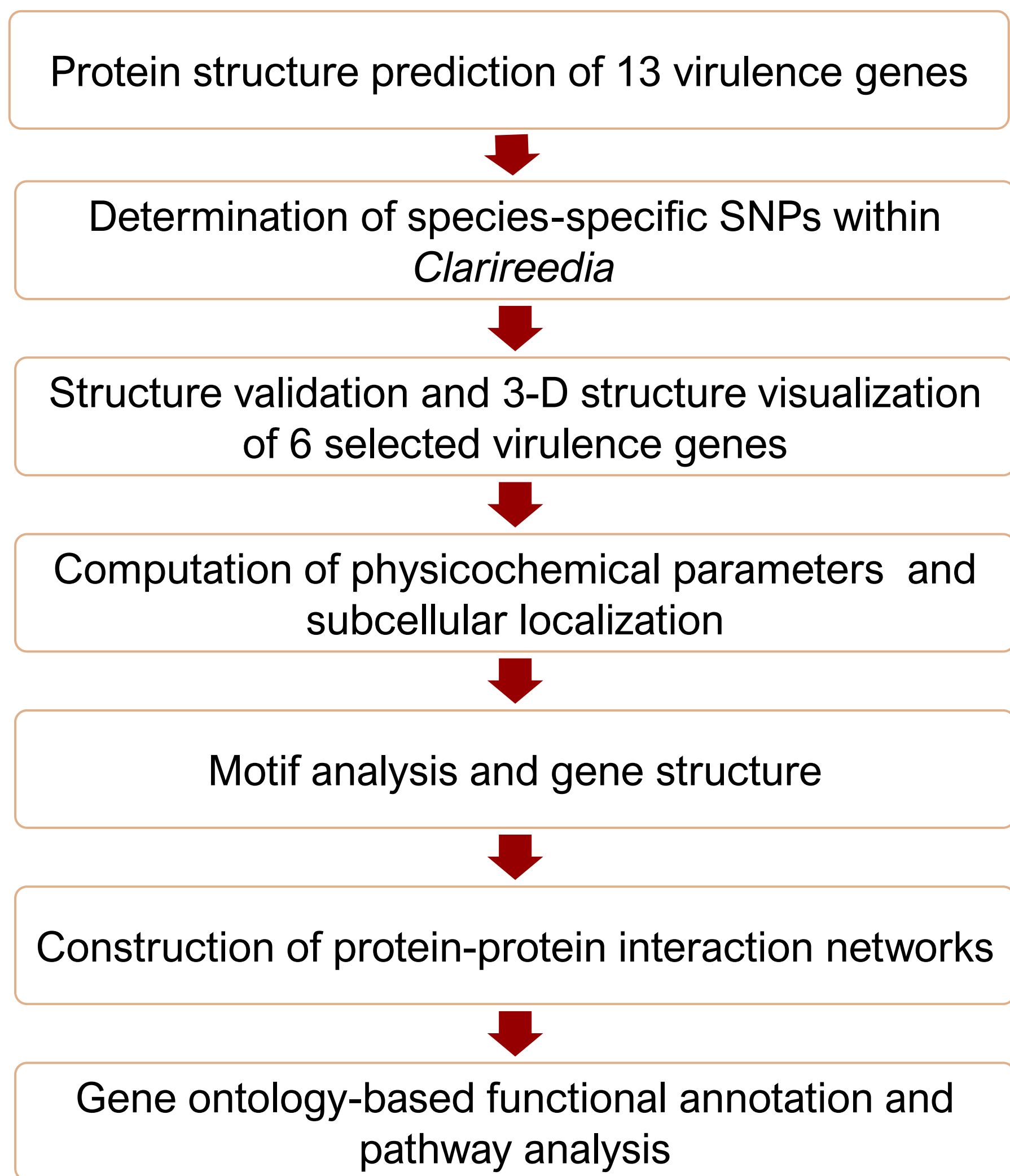Plant Center Retreat, December 15-16, 2021, Brasstown Valley Resort, Young Harris, GA, 30582

INTRODUCTION

Dollar spot is a foliar disease, having a wide host range, infecting both cool- and warm-season turfgrasses [1]. The causal agent for dollar spot was first identified in 1937, an ascomyete and a member of Sclerotiniaceae family, *Sclerotinia homoeocarpa*. However, in 2018, a new fungal genus *Clarireedia* was discerned comprising of four pathogenic species, *C. homoeocarpa, C. bennettii, C. jacksonii,* and *C. monteithiana*, that were responsible for dollar spot on turfgrass [2]. In US, two species, *C. jacksonii,* and *C. monteithiana* have been reported to cause dollar spot [3], and through next-generation sequencing, a novel species *Clarireedia aff. paspali* have been identified (Bahri, unpublished data) as another causal pathogen in the US. The pathogenicity of the three species was analysed by identifying orthologs in *Clarireedia* genomes of virulence genes previously identified in *S. sclerotiorum* [4]. Of the 13 potential virulence genes detected commonly in all three *Clarireedia* species, six were selected for this analysis.
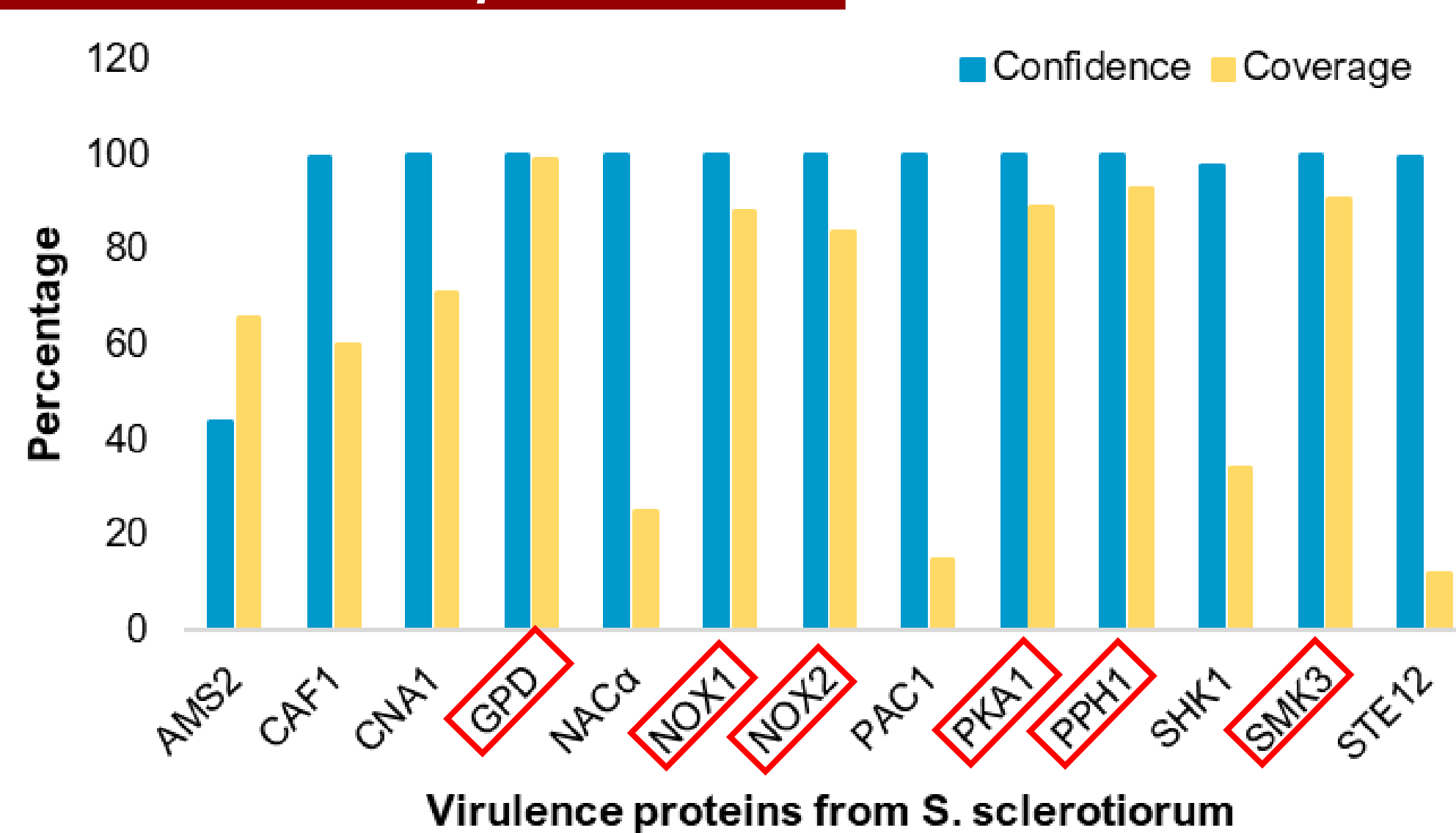
OBJECTIVES

1. To analyse the structural attributes of six potential virulence genes.
2. To functionally annotate the six virulence genes with its interacting protein partners.

METHODOLOGY

Protein structure prediction of 13 virulence genes
↓
Determination of species-specific SNPs within *Clarireedia*
↓
Structure validation and 3-D structure visualization of 6 selected virulence genes
↓
Computation of physicochemical parameters and subcellular localization
↓
Motif analysis and gene structure
↓
Construction of protein-protein interaction networks
↓
Gene ontology-based functional annotation and pathway analysis

RESULTS

Protein structure prediction

Figure 1: Percentage coverage and confidence of prediction of 13 putative virulence proteins from *S. sclerotiorum,* based on known templates to predict the 3-D protein structures using Phyre2 server [5]. Six virulence proteins with good prediction were selected. *Clarireedia aff. paspali, C. jacksonii* and *C. monteithiana* presented similar coverage (86 - 99%) and confidence of prediction (99.90 - 100%) for the 6 selected virulence proteins than *S. sclerotiorum*.

SNP variations
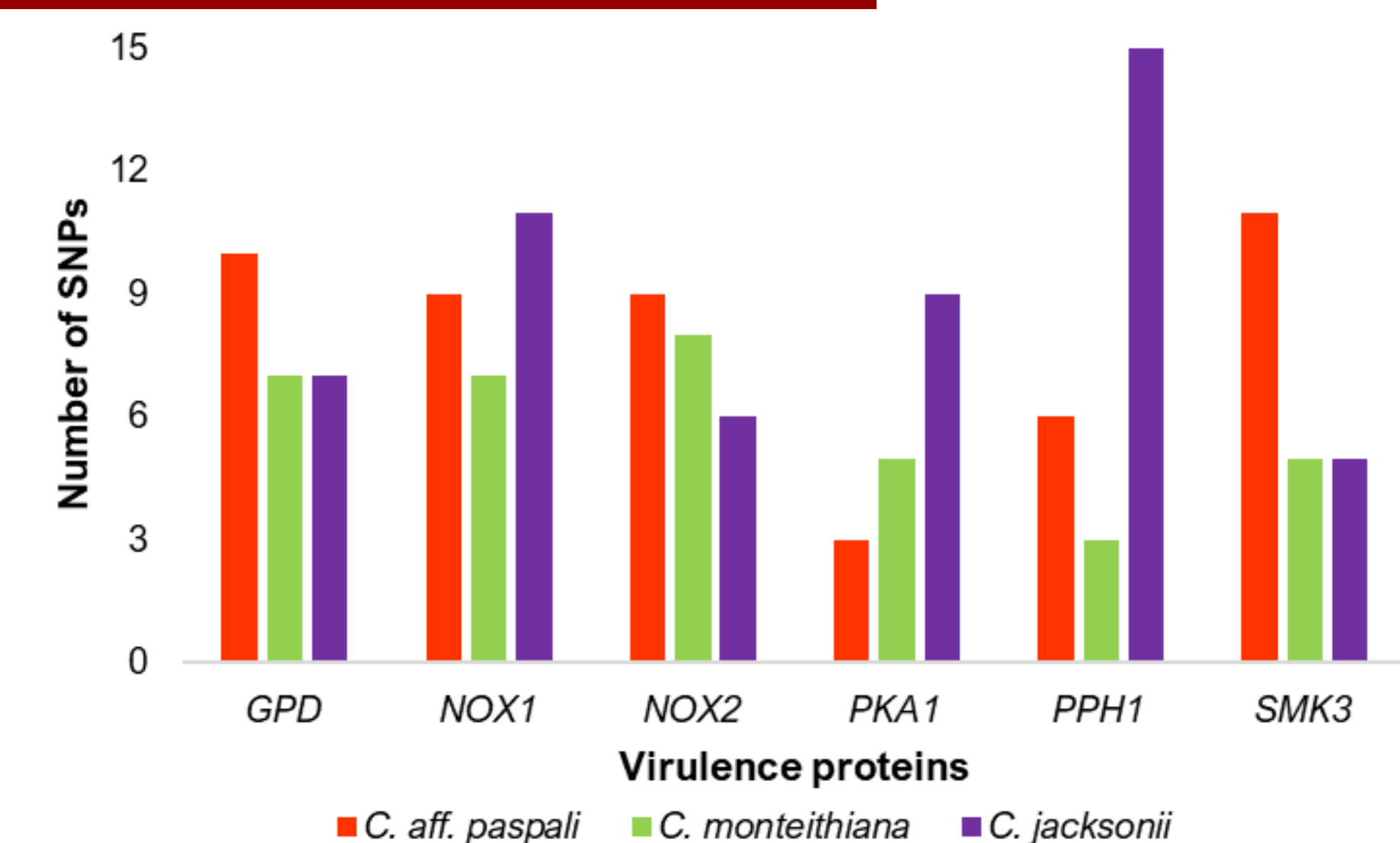
Figure 2: Number of species-specific SNPs detected between *Clarireedia* species in the six virulence proteins

Structure validation

Table 1: Structure validation (quality assessment and model validation) values of the six selected virulence proteins assessed using SAVES v6.0 [6–8] and Qualitative Model Energy ANalysis (QMEAN) [9] tools. *Clarireedia* is represented by *C. aff. paspali, C. jacksonii* and *C. monteithiana*

Organism	Ramachandran Plot Statistics (in %)				Verify 3D	ERRAT Quality Factor	QMEAN Z-score
	Most favored regions	Additional allowed regions	Generously allowed regions	Disallowed regions			
S. sclerotiorum	70.8 - 89	10.7 - 22.1	0.2 - 6.4	0 - 1.7	59.27 - 93.33	50.6667 - 82.0163	(-8.79) - (-1.4)
*Clarireedia**	73.7 - 89.3	10.4 - 21.5	0.2 - 3.6	0 - 1.4	65.86 - 98.02	54.3689 - 81.3699	(-8.5) - (-1.27)

* represents *C. paspali, C. jacksonii* and *C. monteithiana*

Visualization of 3-D structure
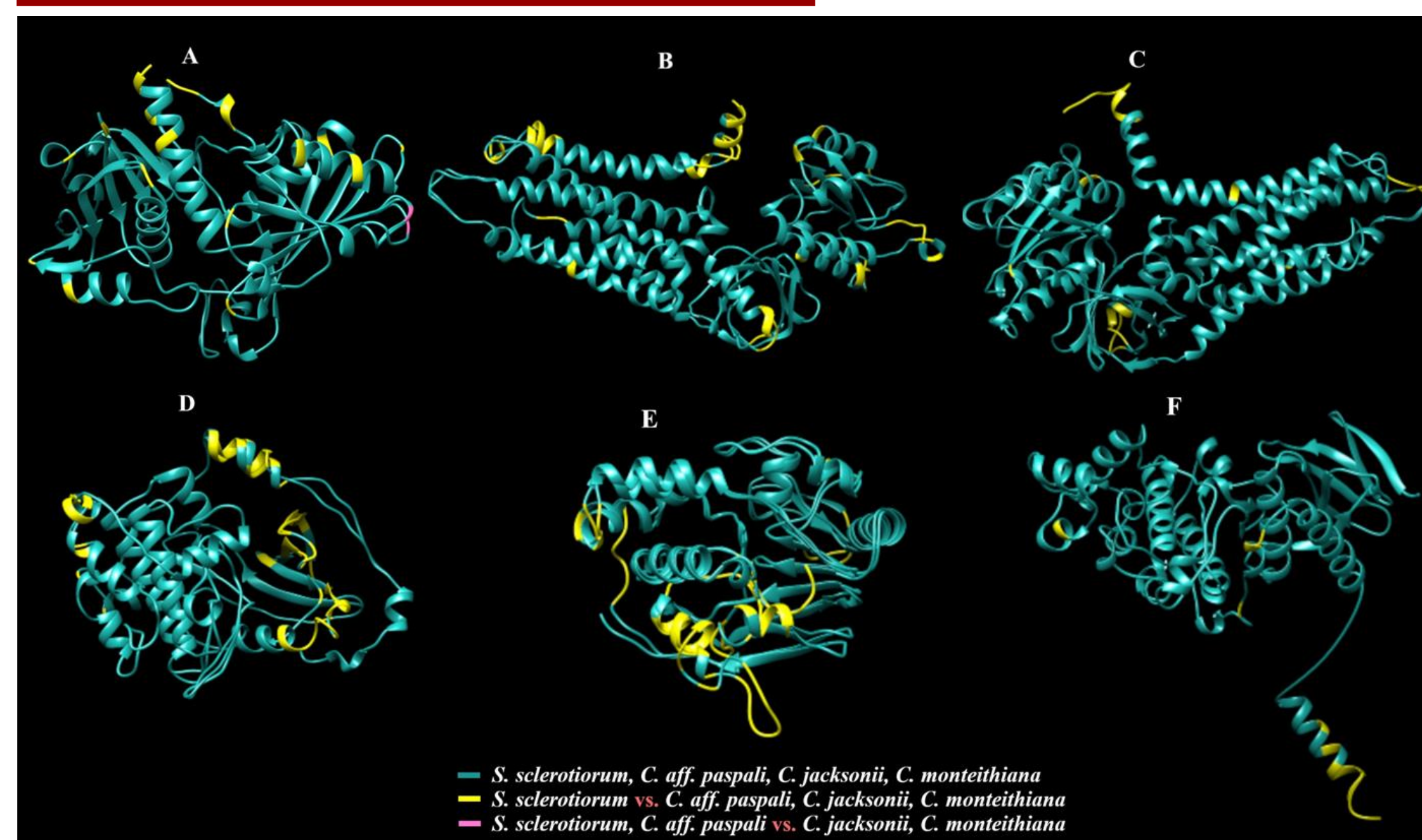
Figure 3: Three-dimensional protein models of six virulence proteins (A: *GPD*, B: *NOX1*, C: *NOX2*, D: *PKA1*, E: *PPH1*, and F: *SMK3*) from *S. sclerotiorum* aligned with the respective protein structures from *C. jacksonii, C. monteithiana*, and *C. aff. paspali*, performed with Chimera 1.15 [10].

Physiochemical Characteristics

Table 2: Physicochemical characterization of the six virulence proteins in *Clarireedia* using ExPASy ProtParam [11]. *Clarireedia* is represented by *C. aff. paspali, C. jacksonii* and *C. monteithiana*.

| Protein	pI	Mol. Wt.	Amino Acids	Instability Index	Aliphatic Index	GRAVY
GPD		36563.64 -36772.67	338			
NOX1		58498.43 - 58697.64	510-511			
NOX2	5.49 - 9.28	66458.86 - 66693.26	580-581	24.24 - 41.18	76.83 - 90.24	(-0.369) - (-0.003)
PKA1		38814.19 - 38890	340-341			
PPH1		29527.26 - 29576.52	260			
SMK3		47113.89 - 47424.31	410-412			

Subcellular Localization
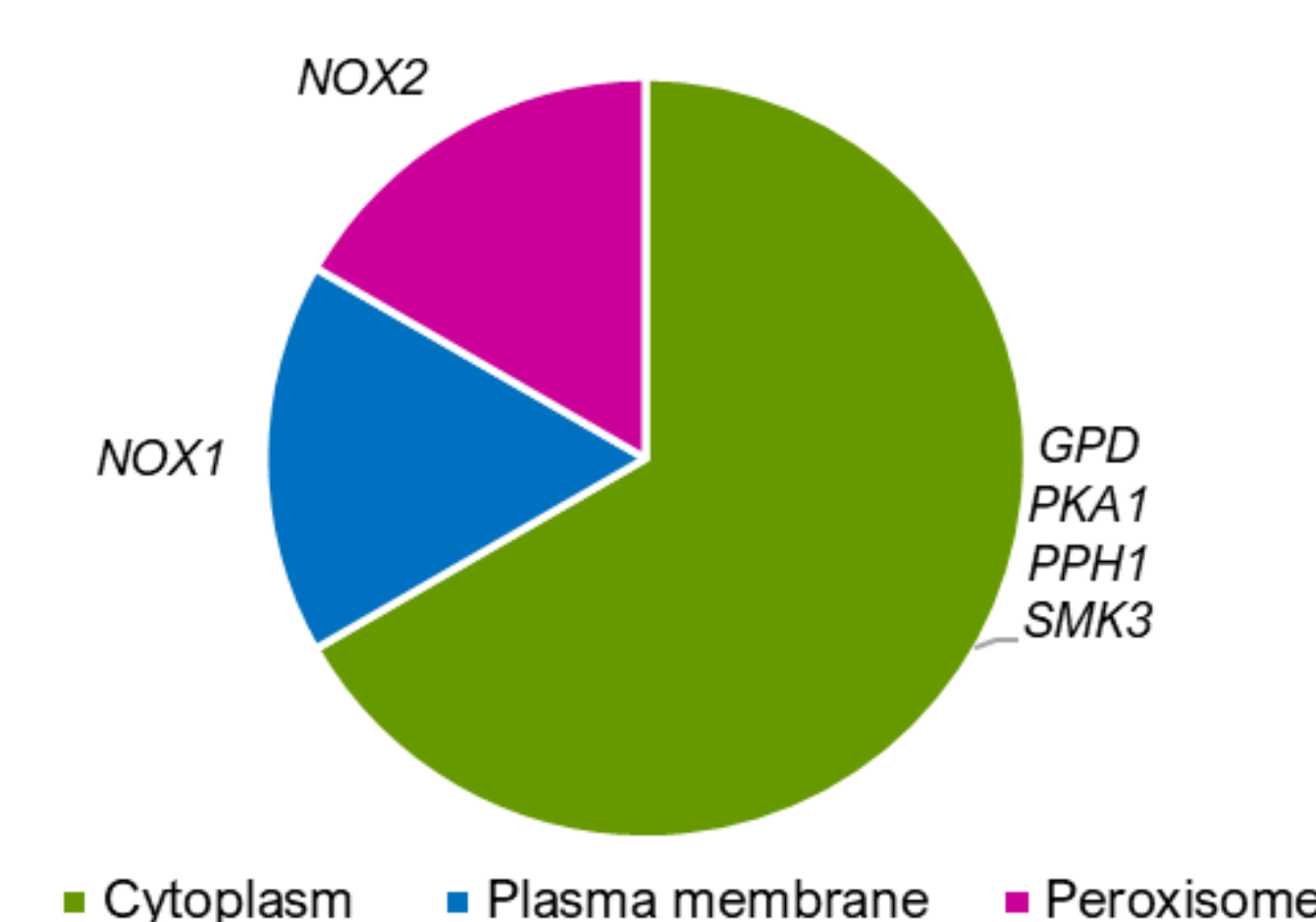
Figure 4: Subcellular localization of the six virulence proteins in *Clarireedia* into different sub-cellular compartment: cytoplasm, plasma membrane and peroxisome, predicted using an Interpretable Subcellular Localization Prediction tool, Yloc [12]. *Clarireedia* is represented by *C. aff. paspali, C. jacksonii* and *C. monteithiana*.

Motif Analysis
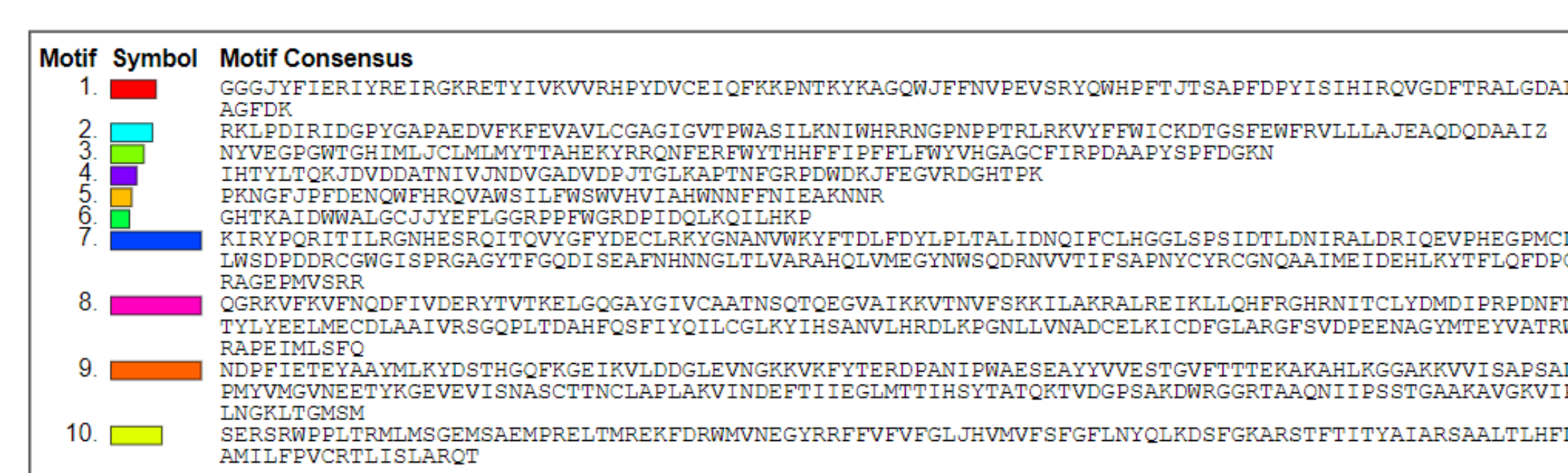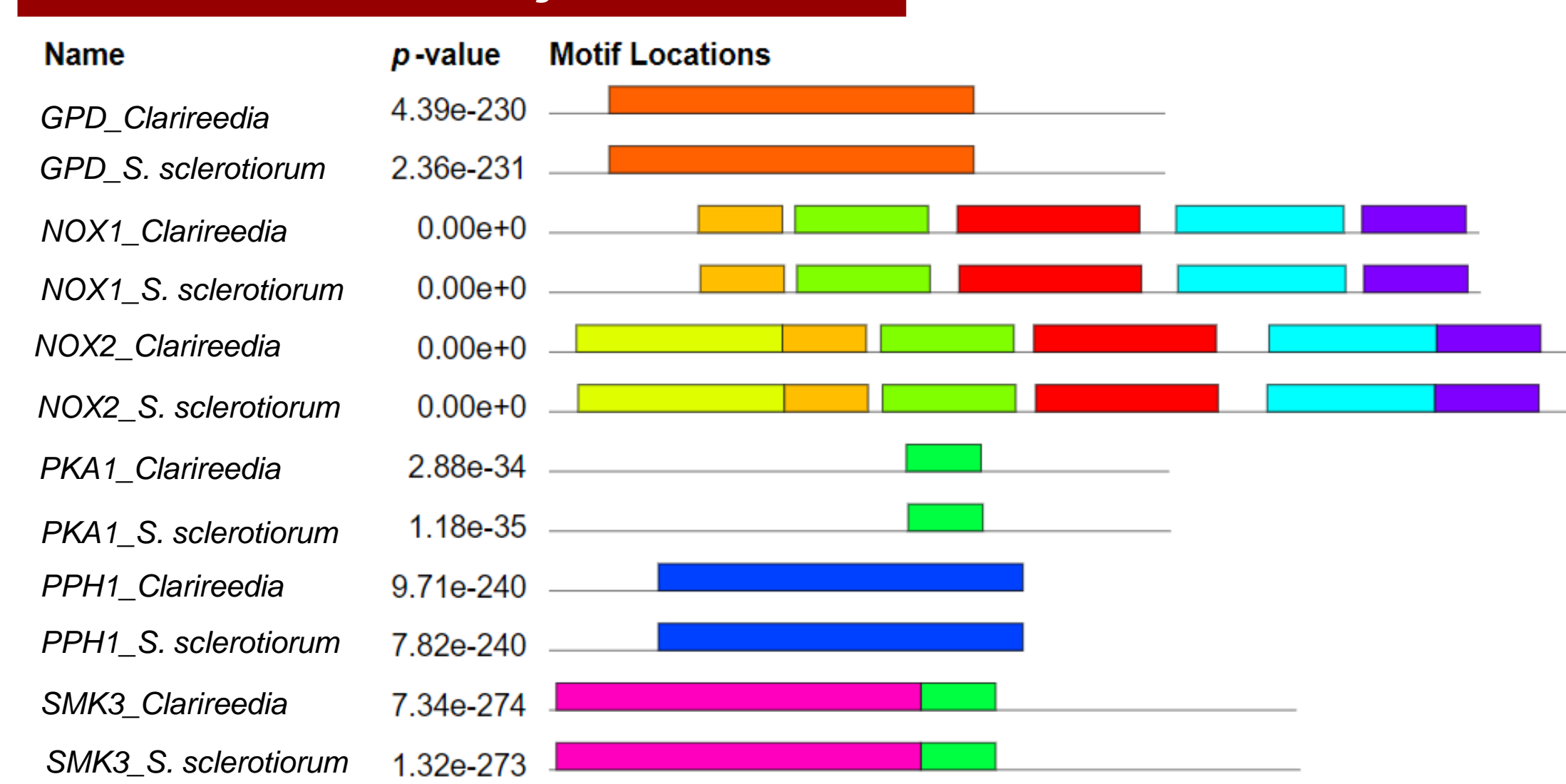
Figure 5: Presence of the ten novel conserved motifs in the six virulence proteins computed using MEME suite [13].

Gene Structure Analysis
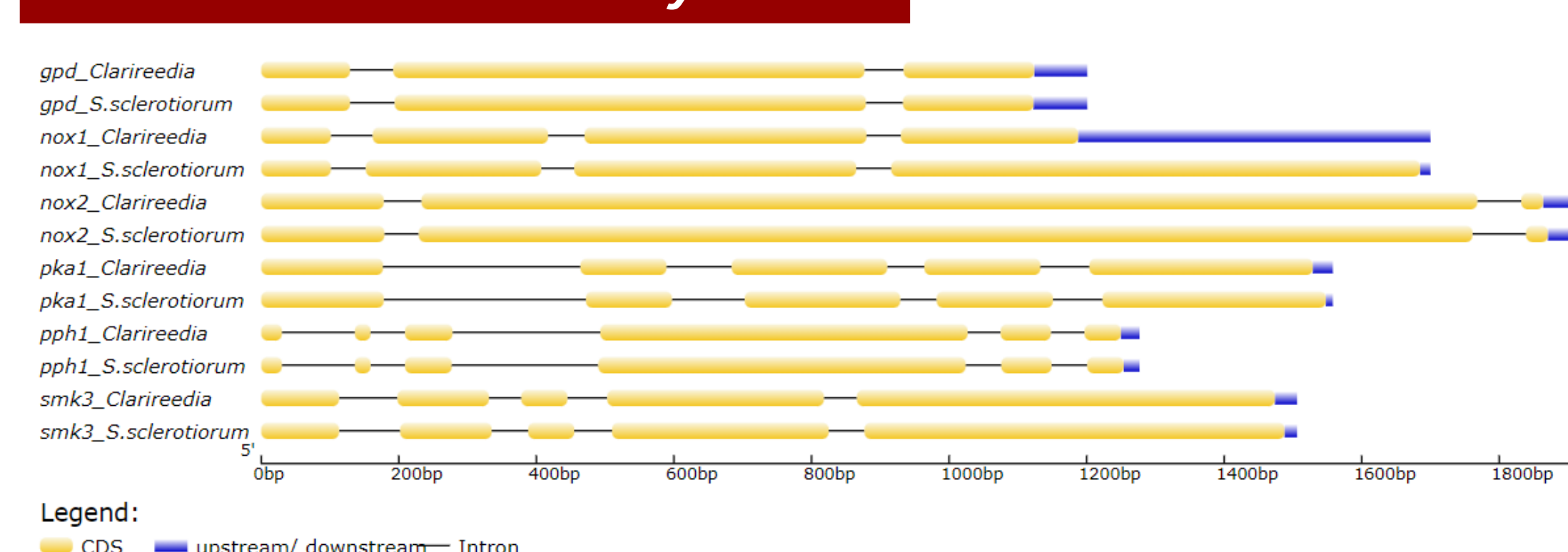
Figure 6: The distribution of introns, exons and upstream/downstream in six virulence genes computed using Gene Structure Display Server 2.0 [14]. *Clarireedia* is represented by *C. aff. paspali*.

Protein-protein interaction
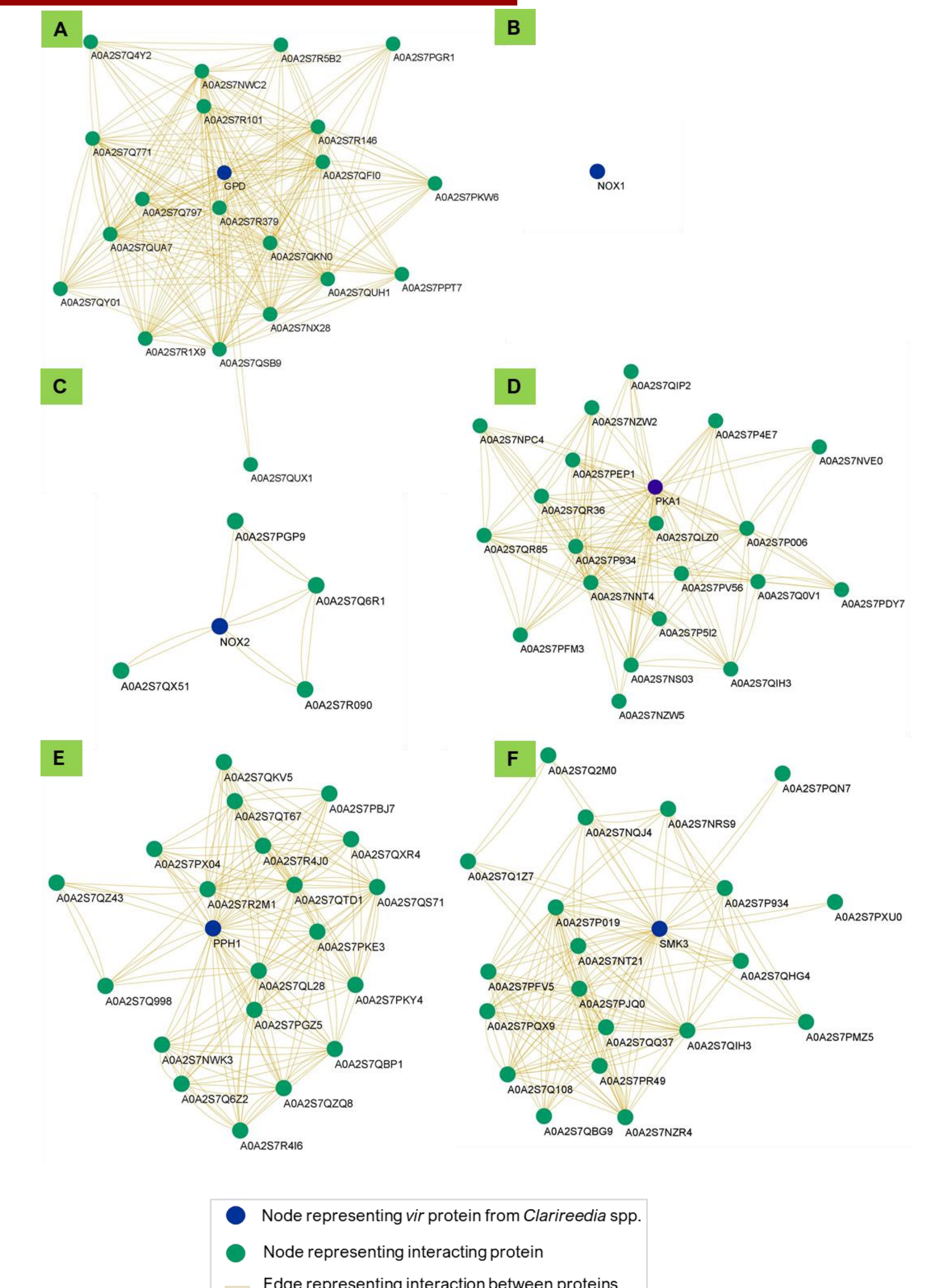
Figure 7: Protein-protein interaction networks of the six virulence proteins (A-F: *GPD, NOX1, NOX2, PKA1, PPH1,* and *SMK3*) from *C. aff. paspali*, predicted using STRING v11 [15] and visualized using Cytoscape 3.9.0 [16]. The interacting proteins with the virulence proteins were belongs to *Rutstroemia* genera, extracted from UniProtKB.

Functional Characterisation
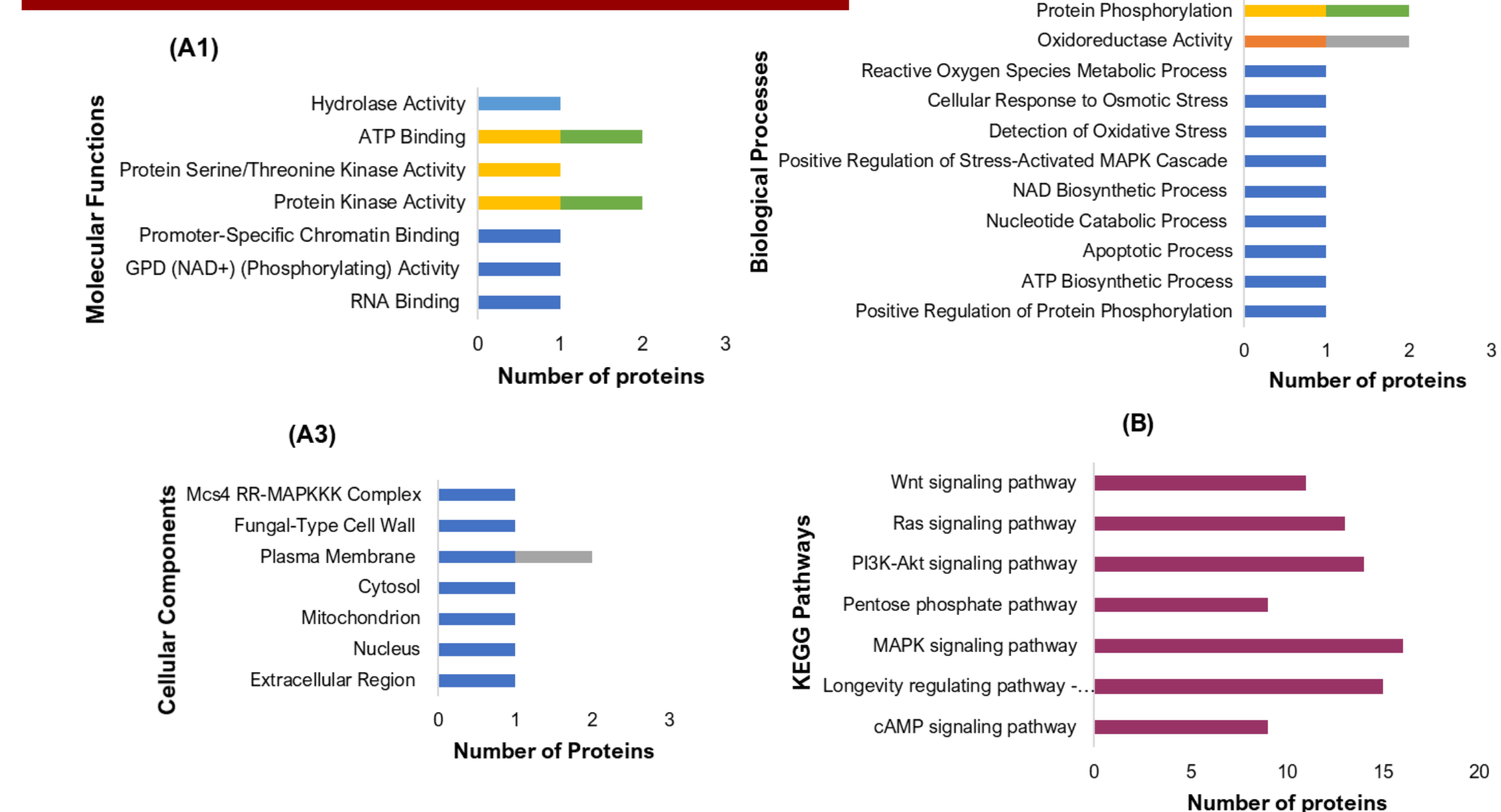
Figure 8: GO-based functional annotation (A1:Molecular Functions, A2:Biological Processes, A3:Cellular Components) of six virulence proteins, and pathway analysis of the proteins with their interacting partners (B) using Blast2GO[18].

SUMMARY/ CONCLUSION

✓ Six potential virulence genes in *Clarireedia*, *gpd, nox1, nox2, pka1, pph1,* and *smk3,* reproduced good three-dimensional protein structures and were selected for further analysis. The 3-D structure visualization showed shared and distinct regions between *S. sclerotiorum* and *Clarireedia* spp.

✓ Species-specific SNP variations were identified between the three *Clarireedia* species.

✓ Physicochemical characteristics gave insight about the stability of the proteins; sub-cellular localization suggested that they majorly localised in cytoplasm.

✓ Motif analysis and gene structure analysis identified similar conserved motifs, and intron-exon arrangement between *Clarireedia* and *S. sclerotiorum*.

✓ Gene ontology-based functional annotation revealed that *PKA1* and *SMK3; NOX1* and *NOX2* had similar functions.

✓ Pathway analysis highlighted the involvement of the virulent proteins and their interactors in major pathways like MAPK signalling pathway.

✓ Further studies are needed to understand the function of these virulence genes in *Clarireedia* and gain insights into the structural and functional disparities of these virulence genes within *Clarireedia*.

REFERENCES

1. Allen *et al.* (2005) doi:10.1094/PHI-1-2005-0217-02.
2. Salgado-Salazar *et al.* (2018) *Fungal Biology* 122:761–773.
3. Sapkota *et al.* (2020) *Plant Disease* 104: 3063.
4. Xia *et al.* (2019) *Pathogens (Basel, Switzerland),* 9(1): 27.
5. Kelley *et al.* (2015). *Nat Protoc* 10: 845–858.
6. Bowie *et al.* (1991). *Science* 253: 164–170.
7. Colovos *et al.* (1993). *Protein Sci* 2: 1511–1519.
8. Laskowski *et al.* (1993). *J Appl Cryst* 26: 283–291.
9. Benkert *et al.* (2011). *Bioinformatics* 27: 343–350.
10. Pettersen *et al.* (2004). *J Comput Chem* 25:1605–1612.
11. Gasteiger *et al.* (2005). *Humana Press*: 571–607.
12. Briesemeister *et al.* (2010). *Nucleic Acids Res* 38: W497–W502.
13. Bailey *et al.* (2015). *Nucleic Acids Research* 43: W39–W49.
14. Hu *et al.* (2015). *Bioinformatics* 31: 1296–1297.
15. Szklarczyk (2019). *Nucleic Acids Research* 47:D607–D613.
16. Shannon *et al.* (2003). *Genome Res* 13: 2498–2504.
17. The UniProt Consortium (2021). *Nucleic Acids Research* 49: D480–D489.
18. Conesa *et al.* (2008). *Int J Plant Genomics* 2008: 619832