

# *In silico* screening of genes coding for secondary metabolites in the phytopathogenic fungus *Stemphylium lycopersici*



Mario E. E. Franco<sup>1\*</sup>, Silvina López<sup>2</sup>, Mario C. N. Saparrat<sup>2,3</sup> & Pedro Balatti<sup>1</sup>.

Centro de Investigaciones de Fitopatología, Facultad de Ciencias Agrarias y Forestales, Universidad Nacional de La Plata. La Plata, Buenos Aires, Argentina<sup>1</sup>; Instituto de Fisiología Vegetal, UNLP-CONICET. La Plata, Buenos Aires, Argentina<sup>2</sup>; Instituto Carlos Spegazzini, Facultad de Ciencias Naturales y Museo. La Plata, Buenos Aires, Argentina<sup>3</sup>.

\*e-mail: ernesto.franco@agro.unlp.edu.ar

## INTRODUCTION

While the end products of the primary metabolism are energy and intermediates for synthesis of essential macromolecules such as lipids, proteins and nucleic acids, the secondary metabolism is the responsible for the production of a wide range of low molecular weight bioactive compounds that are dispensable when microorganisms are cultivated *in vitro* but provide adaptive advantages in nature (1).

The production of secondary metabolites is not universal among microbes, having a rather restricted distribution (1). These compounds are particularly abundant in soil-dwelling filamentous fungi, which exist as multicellular communities competing with each other for nutrients (2). Besides, in several pathogenic filamentous fungi of the ascomycete phylum, virulence is associated with the production of toxic secondary metabolites (3).

The genes responsible for the biosynthesis, export, and transcriptional regulation of secondary metabolites are often found in contiguous gene clusters that are characterized by the presence of “backbone” enzymes and several “decorating” enzymes with characteristic protein domain content. Actually, these hallmarks are the basis of several bioinformatic tools that has been developed for the genome mining of secondary metabolite gene clusters (2, 4).

## OBJECTIVE

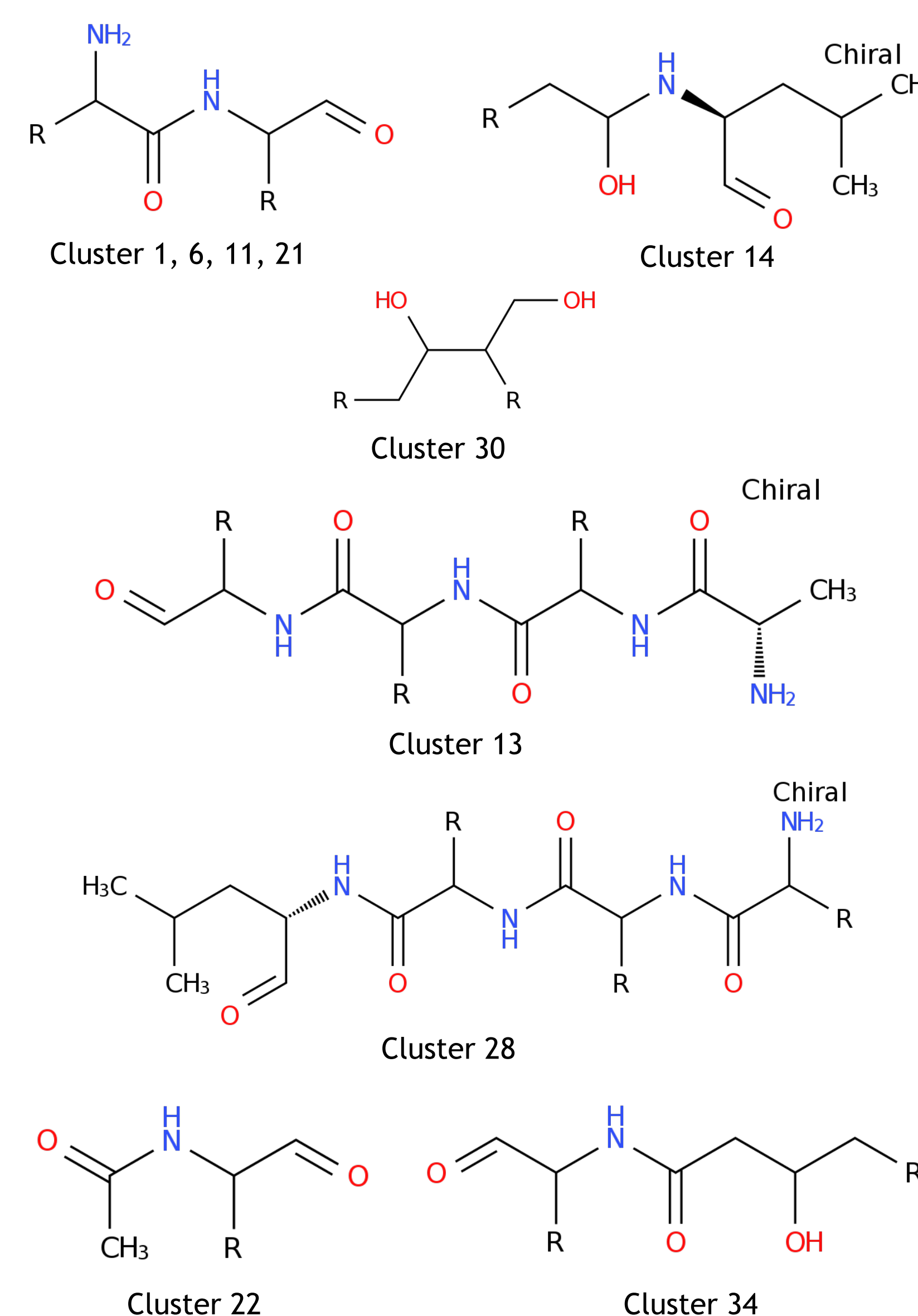
The purpose of this work was identify secondary metabolite gene clusters through the annotated draft genome sequence of the plant pathogenic fungus *Stemphylium lycopersici* CIDEFI-216, one of the etiological agents of tomato gray leaf spot disease.

## MATERIAL AND METHODS

A total of 8,997 protein-coding genes within the 414 annotated scaffolds of the *Stemphylium lycopersici* CIDEFI-216 (GenBank BioSample no. SAMN03332054) draft genome (GenBank Accession no. LGLR00000000) were queried with the bioinformatic tools Secondary Metabolite Unknown Region Finder (2) and Antibiotics and Secondary Metabolite Analysis Shell (4) enabling the ClusterFinder algorithm (5).

## RESULTS

CLUSTER	SCAFFOLD	BACKBONE GENE	ANTISMASH		SMURF		ANNOTATED GENE
			CLUSTER TYPE	CLUSTER POSITION	CLUSTER TYPE	CLUSTER POSITION	
1	LGLR01000164	KNG52744	NRPS-Terpene	1908 - 62553 nt	NRPS	16047-42553 nt	nonribosomal peptide synthase
		KNG52741			-	-	nonribosomal peptide synthase
		KNG52742			-	-	trichodiene synthase
2	LGLR01000165	KNG52671	t1PKS	301560 - 348140 nts	-	-	polyketide synthase
		KNG52673			-	-	1,3,8-naphthalenetriol reductase
3	LGLR01000170	KNG51973	t1PKS	1 - 37630 nt	PKS	5311-38298 nt	polyketide synthase
4	LGLR01000171	KNG51904	t3PKS	157263 - 201082 nt	-	-	chalcone synthase b
5	LGLR01000178	KNG51506	Other	942572 - 986420 nt	-	-	tyrocidine synthetase 1
		KNG51499			-	-	long-chain fatty acid transporter
6	LGLR01000181	KNG51071	NRPS	1 - 44001 nt	NRPS	16310-84650 nt	destruxin synthetase
		KNG51074			-	-	short-chain dehydrogenase reductase sdr
7	LGLR01000182	KNG51057	t1PKS	39016 - 86887 nt	PKS	59016-88712 nt	polyketide synthase
8	LGLR01000188	KNG50459	Other	19732 - 63494 nt	NRPS-Like	34735-43494 nt	hypothetical protein
		KNG50466			-	-	nrps-like enzyme
9	LGLR01000196	KNG50261	Other	931701 - 975289 nt	-	-	l-aminoacidipate-semialdehyde dehydrogenase large subunit
10	LGLR01000197	KNG49932	Other	75512 - 118709 nt	NRPS-Like	95512-100971 nt	NRPS-like enzyme
11	LGLR01000200	KNG49863	NRPS	432405 - 479637 nt	NRPS	424200-491794 nt	nonribosomal peptide synthetase 6
		KNG49866			-	-	long-chain-fatty-acid-CoA ligase
12	LGLR01000205	KNG49339	t1PKS	1 - 30070 nt	PKS	4439-22797 nt	polyketide synthase
13	LGLR01000205	KNG49613	NRPS	981167 - 1037425 nt	NRPS	995247-1017425 nt	nonribosomal peptide synthetase 2
14	LGLR01000205	KNG49728	NRPS-t1PKS	1508521 - 1546674 nt	PKS	1474846-1541190 nt	polyketide synthase
		KNG49726			-	-	hypothetical protein
15	LGLR01000212	KNG49242	Other	1022295 - 1065929 nt	NRPS-Like	1031872-1067043	ankyrin repeat-containing protein
		KNG49247			-	-	proteasome subunit beta type-1 precursor
16	LGLR01000224	KNG48389	t1PKS	210959 - 258797 nt	PKS	218359-259344 nt	polyketide synthase
		KNG48385			-	-	major facilitator superfamily general substrate transporter
17	LGLR01000226	KNG48283	-	-	PKS-like	904697-929482 nt	high affinity nickel transport protein nic1
18	LGLR01000227	KNG48017	Terpene	1 - 21613 nt	-	-	oxidosqualene:lanosterol cyclase
19	LGLR01000228	KNG48001	t1PKS	1 - 26873 nt	PKS	76-12557 nt	polyketide synthase
		KNG48002			-	-	hypothetical protein
20	LGLR01000241	KNG47890	t1PKS	79008 - 120787 nt	PKS	840650-120366 nt	phenolphthiocerol synthesis polyketide synthase ppsa



CLUSTER	SCAFFOLD	BACKBONE GENE	ANTISMASH		SMURF		ANNOTATED GENE
			CLUSTER TYPE	CLUSTER POSITION	CLUSTER TYPE	CLUSTER POSITION	
21	LGLR01000242	KNG47833	Terpene	148756 - 170136 nt	-	-	geranylgeranyl pyrophosphate synthetase
22	LGLR01000242	KNG47857	t1PKS-NRPS	232307 - 285126 nt	NRPS	250144-264441 nt	phenylloxazoline synthase mbtb
		KNG47856			-	-	polyketide synthase
23	LGLR01000245	KNG47736	Terpene	1 - 17198 nt	-	-	terpene synthase metal binding domain-containingprotein
24	LGLR01000249	KNG47342	Terpene	344315 - 365945 nt	-	-	squalene synthetase
25	LGLR01000251	KNG47232	Other	34559 - 61869 nt	NRPS-Like	38676-57833 nt	acetyl-CoA synthetase-like protein
26	LGLR01000258	KNG46901	-	-	NRPS-Like	12490-41637 nt	amp-dependent synthetase ligase
27	LGLR01000278	KNG46670	Terpene	3008 - 25070 nt	-	-	aristolochene synthase
		KNG46671			-	-	o-acetylhomoserine (thiol)-lyase
28	LGLR01000285	KNG46389	NRPS	433114 - 498048 nt	NRPS	61802-474901 nt	hypothetical protein
29	LGLR01000289	KNG46190	t1PKS	25936 - 76224 nt	PKS	36155-56224 nt	polyketide synthase
		KNG46188			-	-	zinc-binding dehydrogenase family oxidoreductase
30	LGLR01000318	KNG45813	t1PKS	377417 - 426290 nt	PKS	357121-421224 nt	lovastatin nonaketide synthase
		KNG45819			-	-	polyketide synthase
		KNG45818			-	-	polyketide synthase
31	LGLR01000318	KNG45775	t1PKS	240667 - 290893 nt	PKS	239819-315971 nt	polyketide synthase
		KNG45780			-	-	short-chain dehydrogenase reductase family 16c member 6
		KNG45776			-	-	short-chain dehydrogenase reductase family
32	LGLR01000381	KNG44599	t1PKS	123385 - 171344 nt	PKS	114632-168416 nt	polyketide synthase
		KNG44600			-	-	short-chain dehydrogenase
33	LGLR01000381	KNG44624	Lantipeptide	202410 - 240128 nt	-	-	pisatin demethylase
		KNG44619			-	-	lanthionine synthetase c-like protein 1
34	LGLR01000386	KNG44542	t1PKS-indole	110 - 73068 nt	PKS	4260-53068 nt	polyketide synthase
		KNG44541			-	-	hypothetical protein
		KNG44544			-	-	versicolorin reductase
		KNG44546			-	-	dimethylallyl tryptophan synthase 1
		KNG44548			-	-	polyketide synthase
35	LGLR01000392	KNG44310	t1PKS	1 - 9699 nt	-	-	6-methylsalicylic acid synthase
36	LGLR01000402	KNG44260	Terpene	188687 - 211923 nt	-	-	phytoene synthase lycopene cyclase
		KNG44257			-	-	phytoene dehydrogenase
		KNG44263			-	-	peroxiredoxin tsa1

Secondary metabolites gene clusters predicted by SMURF and antiSMASH are detailed in the table. Also antiSMASH predicted core structures are shown: rough prediction of core scaffold based on assumed PKS/NRPS colinearity; tailoring reactions not taken into account.

## CONCLUSION

*S. lycopersici* CIDEFI-216 is a source of a broad spectrum of secondary metabolites that can act as virulence factors.

The information generated in this work represents a valuable starting point for experimental validation and characterization of *S. lycopersici* secondary metabolism.

## BIBLIOGRAPHY

1. MALIK, V. S. Microbial secondary metabolism. Trends in Biochemical Sciences, 1980, vol. 5, no 3, p. 68-72; 2. KHALDI, Nora, et al. SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genetics and Biology, 2010, vol. 47, no 9, p. 736-741.; 3. THROCKMORTON, Kurt et al. Evolution of chemical diversity in a group of non-reduced polyketide gene clusters: using phylogenetics to inform the search for novel fungal natural products. Toxins, 2015, vol. 7, no 9, p. 3572-3607.; 4. WEBER, Tilmann, et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic acids research, 2015, vol. 43, no W1, p. W237-W243; 5. CIMERMANCIC, Peter, et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell, 2014, vol. 158, no 2, p. 412-421.