ARTICLE ADDENDUM

New insights into the evolution and structure of *Colletotrichum* plant-like subtilisins (CPLSs)

Vinicio D Armijos Jaramillo,[1] Walter A Vargas,[1,†] Serenella A Sukno[1] and Michael R Thon[1,*]

[1]Centro Hispano-Luso de Investigaciones Agrarias; Departamento de Microbiología y Genética; Universidad de Salamanca; Villamayor, Spain

[†]Current affiliation: Centro de Estudios Fotosintéticos y Bioquímicos-CONICET; Facultad de Ciencias Bioquímicas y Farmacéuticas-UNR; Rosario, Argentina

The *Colletotrichum* plant-like subtilisins (CPLSs) are a family of proteins found only in species of the phytopathogenic fungus *Colletotrichum*. CPLSs have high similarity to plant subtilisins and our previous work has shown that they were acquired by an ancient horizontal gene transfer event from plants. The rapid growth of sequence data in public databases enabled us to reexamine the structure and evolution of the CPLSs. A new plant subtilisin structural model aided us in refining the tertiary structure of CPLSs. Also, new information about protein interactions of plant subtilisin has provided new insights into the putative function of CPLSs. The availability of new genome sequences of members of the genus *Colletotrichum* gave us the opportunity to further validate our hypothesis that the CPLSs are unique to the *Colletotrichum* lineage. Together, this information furthers our knowledge of the potential role of the CPLSs in pathogenicity and the role of HGT in the genome evolution of plant pathogenic fungi.

The increasing availability of genomic sequences that are available in public databases has enabled researches to detect rare evolutionary events such as Horizontal Gene Transfer (HGT).[1-10] Today it is widely accepted that HGT is not an unusual phenomenon. There are many reports of HGT between bacteria and from bacteria to eukaryotes. However, less common are the reports of HGT events from plants to fungi.[11-14] Our recent report of the horizontal transfer of a subtilisin (CPLS; *Colletotrichum* plant-like subtilisin) from plants to an ancestor of the fungal genus *Colletotrichum* is particularly interesting because it shows the horizontal transfer of a subtilisin S8 (family of proteases involved in many different pathways) from plants to a phytopathogen.[15]

A New Structural Model for CPLS

We previously reported (see ref. 15) a prediction of the tertiary protein structure of a *C. graminicola* CPLS (locus ID GLRG_05578) using crystallized protein structures of subtilisins with sequence similarity to CPLSs. The most similar one was the SBT3 of *Solanum lycopersicum* (PDB ID 3I6S). Recently, the crystal structure of the subtilisin Cucumisin of *Cucumis melo* was published.[16] With this new structure available, we recalculated the 3-D model of GLRG_05578 (the CPLS of *Colletotrichum graminicola*) using the Phyre 2 server.[17] The new structural model has only 57 residues predicted *ab initio* (highly unreliable sites) instead of 73 as reported in Armijos Jaramillo et al.[15] The new GLRG_05578 structural model was aligned to the structural models of SBT3 and Cucumisin to determine the similarities and the differences with the previous model. The structural alignment of the previous model of GLRG_05578 and SBT3 has a maximum RMSD (root–mean–square–deviation) of 11.68 and a minimum of 0.02. The new model has a highest RMSD value of 12.65 and the lowest value of 0.03. These results show that the new model is less similar to SBT3 than the one reported previously.[15]
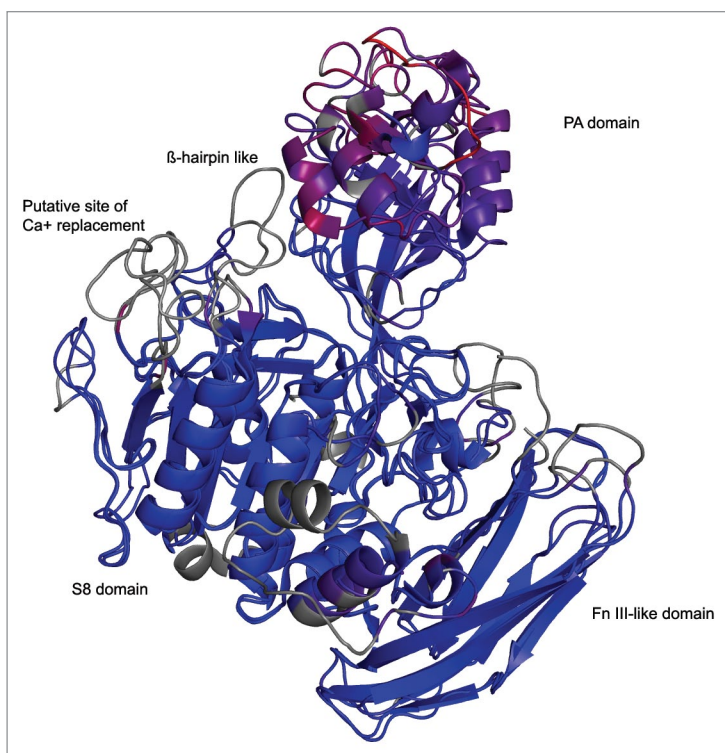
Figure 1. Structural alignment of the new model predicted for the tertiary conformation of CPLS GLRG_05578 of *Colletotrichum graminicola* and one monomer of the crystallized subtilisin Cucumisin (3VTA) of *Cucumis melo*. The color spectrum represents the values of pairwise RMSD, with blue specifying minimum values and red indicating maximum values. Gray sections are not aligned. Labels indicate the relevant domains and sites of the proteins conserved in subtilisin SBT3 of Tomato in Ottman et al.[18]

However, the new model is more similar to the SBT3 subtilisin than to the Cucumisin (**Fig. 1**) (maximum RMSD of 17.59 and minimum RMSD of 3.09 with the old model and maximum RMSD of 15.12 and minimum RMSD of 1.84 with the new model). The new model reveals differences in the tertiary protein structures of the PA domain, β-hairpin arrangement and one of the putative Ca^{+2} replacement sites between CPLS and tomato subtilisin. These sites are important in dimerization and stabilization of tomato SBT3[18] and the differences with CPLS could imply differences in the quaternary structure (conformation of dimers) and/or the specificity of interaction with other proteins.

Interactions of Plant Subtilisins could Reveal the Function of CPLSs

Although the protein targets of the CPLSs are unknown, analysis of the most similar plant proteins could reveal potential clues. A BLASTP search in the UniProt/SwissProt database shows that the most similar proteins to CPLSs are 3 subtilisins of *Arabidopsis thaliana* (O65351, O64495, Q9LLL8) and the Cucumisin of *Cucumis melo* (Q39547). Information about the interaction of the *Arabidopsis* subtilisins is available in the STRING database,[19] but it is not available for Cucumisin. This database shows that O64495 is co-expressed with several proteins involved in the development of stomata, including AT1G63700, AT5G62230, AT1G80080,[20] AT5G53210, AT3G06120,[21] and others. The protein O65351 (ARA12) is predicted to interact with AT1G22300 and AT4G09000, 2 general regulatory factors.[22] Gene Q9LLL8 is co-expressed with AT4G35350 and AT1G20850,[23] which have peptidase activity and cysteine-type peptidase activity, respectively. The interaction and co-expression of these proteins could help to explain the role of CPLSs in the host and help to plan future studies. However, as we commented in Armijos

Jaramillo et al.,[15] the subtilisins in plants have a variety of metabolic functions. Plant subtilisins with high percentage of sequence similarity to CPLSs are involved in varied, and sometimes unrelated biological process. The CPLS homologs in *Arabidopsis* are a good example of this phenomenon. Protein O65351 (ARA12) is essential for mucilage release from seed coats, whereas O64495 is involved in stomatal development and distribution, and Q9LLL8 is involved in negative regulation of catalytic activity. The functions of the CPLSs are not clear yet, but they could be involved in the mis-regulation or mimicry of any of those processes. Alternatively, novel functions might also be expected for CPLS.

CPLS are Unique to *Colletotrichum*

Our hypothesis is that the CPLS HGT event took place in an ancestor of the genus *Colletotrichum*. Therefore, CPLSs should be present in all of the extant species of *Colletotrichum* and absent in species outside the genus. We previously showed that the CPLSs are present in the genome sequences of all 3 species of *Colletotrichum* that were available at that time but absent in all other fungi. Recently, Gan et al.,[13] reported the presence of subtilisins, likely resulting from an HGT event, in the genomes of *C. orbiculare* (Cob 12233, Cob 06327) and *C. gloeosporioides* (CGGC5 2662), adding further support to our hypothesis. In addition, genome sequences for several more species of *Colletotrichum* have recently become available, enabling us to further validate our hypothesis. A BLAST search of the GenBank nr database[24] revealed the presence of CPLS homologs in the genome of *C. gloeosporioides* Nara gc5 (GenBank accession numbers: ELA37268 and ELA38265). We also performed BLAST searches against the proteomes of 2 additional species of *Colletotrichum* (*C. gloeosporioides* 23 and *C. fiorinae* MH 18, http://www.jgi.doe.gov) and found that both species have CPLS homologs. A phylogenetic analysis (data not shown) shows that all of the putative CPLS homologs share common ancestry with the CPLSs identified in our previous report. This idea is consistent

with our hypothesis of a plant protein that was transferred to an ancestor of the genus *Colletotrichum*. HGT events among eukaryotes are presumed to be rare, and we assume that those that we observe in populations, such as the CPLSs, have been fixed due to strong selective pressure. In the case of plant pathogenic fungi, such as *Colletotrichum*, horizontally transferred genes may have some role in pathogenicity, enabling the fungus to adapt to new hosts or enhancing its ability to colonize its host. Research is already underway which is aimed at determining whether CPLSs have a role in pathogenicity in *Colletotrichum*.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature 2000; 405:299-304; PMID:10830951; http://dx.doi.org/10.1038/35012500
2. Dutta C, Pan A. Horizontal gene transfer and bacterial diversity. J Biosci 2002; 27:27-33; PMID:11927775; http://dx.doi.org/10.1007/BF02703681
3. Brown JR. Ancient horizontal gene transfer. Nat Rev Genet 2003; 4:121-32; PMID:12560809; http://dx.doi.org/10.1038/nrg1000
4. Gogarten JP. Gene transfer: gene swapping craze reaches eukaryotes. Curr Biol 2003; 13:R53-4; PMID:12546803; http://dx.doi.org/10.1016/S0960-9822(02)01426-4
5. Rolland T, Neuvéglise C, Sacerdot C, Dujon B. Insertion of horizontally transferred genes within conserved syntenic regions of yeast genomes. PLos ONE 2009; 4:e6515; PMID:19654869; http://dx.doi.org/10.1371/journal.pone.0006515
6. Fitzpatrick DA. Horizontal gene transfer in fungi. FEMS Microbiol Lett 2012; 329:1-8; PMID:22112233; http://dx.doi.org/10.1111/j.1574-6968.2011.02465.x
7. Gardiner DM, McDonald MC, Covarelli L, Solomon PS, Rusu AG, Marshall M, et al. Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts. PLoS Pathog 2012; 8:e1002952; PMID:23028337; http://dx.doi.org/10.1371/journal.ppat.1002952
8. Kim SE, Moon JS, Choi WS, Lee SH, Kim SU. Monitoring of horizontal gene transfer from agricultural microorganisms to soil bacteria and analysis of microbial community in soils. J Microbiol Biotechnol 2012; 22:563-6; PMID:22534306; http://dx.doi.org/10.4014/jmb.1110.10066
9. Moran Y, Fredman D, Szczesny P, Grynberg M, Technau U. Recurrent horizontal transfer of bacterial toxin genes to eukaryotes. Mol Biol Evol 2012; 29:2223-30; PMID:22411854; http://dx.doi.org/10.1093/molbev/mss089
10. Techtmann SM, Lebedinsky AV, Colman AS, Sokolova TG, Woyke T, Goodwin L, et al. Evidence for horizontal gene transfer of anaerobic carbon monoxide dehydrogenases. Front Microbiol 2012; 3:132; PMID:22529840; http://dx.doi.org/10.3389/fmicb.2012.00132
11. Richards TA, Soanes DM, Foster PG, Leonard G, Thornton CR, Talbot NJ. Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. Plant Cell 2009; 21:1897-911; PMID:19584142; http://dx.doi.org/10.1105/tpc.109.065805
12. de Jonge R, van Esse HP, Maruthachalam K, Bolton MD, Santhanam P, Saber MK, et al. Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. Proc Natl Acad Sci USA 2012; 109:5110-5; PMID:22416119; http://dx.doi.org/10.1073/pnas.1119623109
13. Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y, et al. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. New Phytol 2013; 197:1236-49; PMID:23252678; http://dx.doi.org/10.1111/nph.12085
14. Sun BF, Xiao JH, He S, Liu L, Murphy RW, Huang DW. Multiple interkingdom horizontal gene transfers in *Pyrenophora* and closely related species and their contributions to phytopathogenic lifestyles. PLos ONE 2013; 8:e60029; PMID:23555871; http://dx.doi.org/10.1371/journal.pone.0060029
15. Armijos Jaramillo VD, Vargas WA, Sukno SA, Thon MR. Horizontal transfer of a subtilisin gene from plants into an ancestor of the plant pathogenic fungal genus *Colletotrichum*. PLos ONE 2013; 8:e59078; PMID:23554975; http://dx.doi.org/10.1371/journal.pone.0059078
16. Murayama K, Kato-Murayama M, Hosaka T, Sotokawauchi A, Yokoyama S, Arima K, et al. Crystal structure of cucumisin, a subtilisin-like endoprotease from *Cucumis melo* L. J Mol Biol 2012; 423:386-96; PMID:22841692; http://dx.doi.org/10.1016/j.jmb.2012.07.013
17. Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 2009; 4:363-71; PMID:19247286; http://dx.doi.org/10.1038/nprot.2009.2
18. Ottmann C, Rose R, Huttenlocher F, Cedzich A, Hauske P, Kaiser M, et al. Structural basis for Ca2+-independence and activation by homodimerization of tomato subtilase 3. Proc Natl Acad Sci USA 2009; 106:17223-8; PMID:19805099; http://dx.doi.org/10.1073/pnas.0907587106
19. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013; 41:D808-15; PMID:23203871; http://dx.doi.org/10.1093/nar/gks1094
20. Tricker PJ, Gibbings JG, Rodríguez López CM, Hadley P, Wilkinson MJ. Low relative humidity triggers RNA-directed de novo DNA methylation and suppression of genes controlling stomatal development. J Exp Bot 2012; 63:3799-813; PMID:22442411; http://dx.doi.org/10.1093/jxb/ers076
21. Pillitteri LJ, Peterson KM, Horst RJ, Torii KU. Molecular profiling of stomatal meristemoids reveals new component of asymmetric cell division and commonalities among stem cell populations in Arabidopsis. Plant Cell 2011; 23:3260-75; PMID:21963668; http://dx.doi.org/10.1105/tpc.111.088583
22. Swatek KN, Graham K, Agrawal GK, Thelen JJ. The 14-3-3 isoforms chi and epsilon differentially bind client proteins from developing Arabidopsis seed. J Proteome Res 2011; 10:4076-87; PMID:21766784; http://dx.doi.org/10.1021/pr200263m
23. Zhao C, Johnson BJ, Kositsup B, Beers EP. Exploiting secondary growth in Arabidopsis. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. Plant Physiol 2000; 123:1185-96; PMID:10889267; http://dx.doi.org/10.1104/pp.123.3.1185
24. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2011; 39:D32-7; PMID:21071399; http://dx.doi.org/10.1093/nar/gkq1079