UNIVERSIDADE FEDERAL DE LAVRAS

FABÍOLA DE JESUS SILVA

CHALCONE ANALOGUES ON THE CONTROL OF *Meloidogyne incognita* AND GENOMIC STUDIES OF THE BIOCONTROL AGENT *Bacillus velezensis*

LAVRAS - MG
2020

FABÍOLA DE JESUS SILVA

CHALCONE ANALOGUES ON THE CONTROL OF *Meloidogyne incognita* AND GENOMIC STUDIES OF THE BIOCONTROL AGENT *Bacillus velezensis*

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Fitopatologia, área de concentração Fitopatologia, para obtenção do título de Doutora.

Dr. Vicente Paulo Campos
Orientador

Dra. Aline Ferreira Barros
Coorientadora

LAVRAS – MG
2020

FABÍOLA DE JESUS SILVA

CHALCONE ANALOGUES ON THE CONTROL OF *Meloidogyne incognita* AND
GENOMIC STUDIES OF THE BIOCONTROL AGENT *Bacillus velezensis*

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Fitopatologia, área de concentração Fitopatologia, para a obtenção do título de Doutora.

APROVADA em 05 de Março de 2020.

Dra. Aline Ferreira Barros	AGROTESTE
Dr. Valter Cruz Magalhães	UFLA
Dr. Sonia Maria de Lima Salgado	EPAMIG
Dr. Denilson Ferreira de Oliveira	UFLA

Dr. Vicente Paulo Campos
Orientador

Dra. Aline Ferreira Barros
Cooorientadora

LAVRAS – MG
2020

Aos meus pais, José Miguel e Maria Luzete, que muito me apoiaram e me

incentivaram a realiza-lo.

Dedico

AGRADECIMENTOS

Primeiramente, manifesto aqui minha gratidão a Deus, que me deu força e energia para que tudo isso acontecesse. Em todos os momentos, Ele foi o meu maior mestre.

Agradeço a Universidade Federal de Lavras, e ao programa de Pós-Graduação em Fitopatologia pela oportunidade concedida para a realização do Doutorado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela concessão da bolsa de estudos.

Meu muito obrigada ao Dr. Vicente Paulo Campos, pela orientação, apoio e confiança durante todos esses anos.

Meu agradecimento especial a minha coorientadora e amiga, Dr. Aline Ferreira Barros, que nunca negou uma ajuda durante esses anos de trabalho.

Agradeço também aos colegas e amigos do Laboratório de Nematologia, pelas colaborações na condução dos experimentos, momentos de descontração e amizade.

A Vanessa, Acleide e Gizeli, companheiras de trabalho e irmãs de amizade que fizeram parte da minha formação e que vão continuar presentes em minha vida com certeza.

Agradeço a Larissa Carvalho, pelo seu grande desprendimento em me ajudar. Você foi "peça" fundamental desse trabalho.

Aos meus pais, pelo amor, incentivo e apoio incondicional.

À minha irmã, Neilly Paula, e aos meus afilhados Carlos Alberto e Maria Izabel por todo carinho e alegria.

A toda minha família, apesar de estar distante, estão sempre presentes em meu coração.

Ao meu namorado, Gabriel, que sempre esteve ao meu lado. Obrigada por todo amor, carinho e paciência.

À Cleicy e ao Francisco José, pela confiança, carinho e amizade.

E a todos que direta ou indiretamente fizeram parte da minha formação, meus sinceros agradecimentos.

"Sem a curiosidade que me move, que me inquieta, que me insere na busca,

não aprendo nem ensino."

Paulo Freire

RESUMO GERAL

Os nematoides das galhas, *Meloidogyne* spp., são patógenos vegetais economicamente importantes e distribuídos em todo o mundo. Dentre as estratégias utilizadas para o controle desses patógenos estão os nematicidas químicos. No entanto, o uso excessivo dessas moléculas tóxicas na agricultura causam sérios problemas ambientais. Daí a necessidade de estratégias sustentáveis, como por exemplo, o uso de compostos naturais e agentes biológicos. As chalconas são produtos intermediários e finais na rota de biossintese dos flavonoides. Elas possuem um amplo espectro de atividades biológicas importantes. Neste trabalho, estudou-se a atividade de 12 análogos de chalconas contra o nematoide das galhas *Meloidogyne incognita*. Três deles mostraram forte ação nematicida e nematostática contra juvenis de segundo estádio de *M. incognita.* A chalcona (1E,4E)-1,5-di(4-nitrofenil)-2-butilpenta-1,4-dien-3-ona (composto 6) apresentou maior atividade nematicida do que o nematicida comercial Carbofuran, e quando aplicada em tomateiros infestados, reduziu o número de galhas e ovos do nematoides em 51% e 68%, respectivamente. Em estudo *in silico*, esta chalcona atua, presumivelmente, inibindo a enzima citocromo P450, que é importante na oxidação de diversas substâncias na fisiologia do nematoide. Ainda neste trabalho, estudou-se o genoma completo da estirpe *Bacillus velezensis* UFLA258, que é um agente de controle biológico de patógenos vegetais, desde a sua obtenção, seguida da montagem e anotação. Adicionalmente, usando uma abordagem genômica comparativa, realizaram-se avaliações *in silico* com todos os genomas completos de *B. velezensis* disponíveis no banco de dados, mais os genomas completos das espécies próximas *Bacillus amyloliquefaciens* e *Bacillus siamensis*. Assim, o genoma de *B. velezensis* UFLA258 consiste em um único cromossomo de 3,95 Mbp de comprimento, com um teor médio de GC de 46,69%. Contém 3.949 genes codificadores de proteína e 27 genes de RNA. Análises baseadas na identidade média dos nucleotídieos (ANI), hibridização DNA-DNA (dDDH) e filogenia com sequências completas do gene *rpoB* confirmaram que 19 cepas depositadas no banco de dados como *B. amyloliquefaciens* eram de fato *B. velezensis*. Embora essas espécies sejam filogeneticamente próximas, as análises combinadas de várias características genômicas, como presença de genes biossintéticos codificadores de metabólitos secundários, arranjos CRISPr/Cas, ANI, dDDH, e outras informações sobre as cepas, incluindo fonte de isolamento, permitiram sua classificação inequívoca como estas três espécies. Esta análise genômica amplia o conhecimento sobre as espécies aparentadas, *B. velezensis*, *B. amyloliquefaciens* e *B. siamensis*, com ênfase no status taxonômico.

Palavras-chave: Nematoide das galhas, chalconas, controle biológico, sequenciamento do genoma, *Bacillus* sp., taxonomia.

GENERAL ABSTRACT

Root-knot nematodes, *Meloidogyne* spp., are economically important plant pathogens and distributed worldwide. Among the strategies used to control these pathogens are chemical nematicides. However, the overuse of these toxic molecules in agriculture causes serious environmental problems. Hence the need for sustainable strategies, such as the use of natural compounds and biological agents. Chalcones are intermediate and final products in the flavonoid biosynthesis route. They have a wide spectrum of important biological activities. In this work, the activity of 12 chalcone analogues against the root-knot nematode *Meloidogyne incognita* was studied. Three of them showed strong nematicidal and nematostatic action against second stage juveniles of *M. incognita*. Chalcone (1E,4E)-1,5-di(4-nitrophenyl)-2-butylpenta-1,4-dien-3-one (compound 6) showed higher nematicidal activity than the commercial nematicide Carbofuran, and when applied in infested tomato plants, it reduced the number of nematode galls and eggs by 51% and 68%, respectively. In *in silico* study, this chalcone presumably acts by inhibiting the cytochrome P450 enzyme, which is important in the oxidation of various substances in nematode physiology. Also in this work, we studied the complete genomic sequence of the strain *Bacillus velezensis* UFLA258, which is a biological control agent of plant pathogens, since its obtaining, followed by assembly and annotation. In addition, using a comparative genomic approach, *in silico* evaluations were performed with all complete *B. velezensis* genomes available in the database, plus genomes of nearby species *Bacillus amyloliquefaciens* and *Bacillus siamensis*. Thus, the *B. velezensis* UFLA258 genome consists of a single chromosome of 3.95 Mbp in length, with an average GC content of 46.69%. It contains 3,949 protein coding genes and 27 RNA genes. Analysis based on mean nucleotide identity (ANI), DNA-DNA hybridization (dDDH), and complete sequence phylogeny of the *rpoB* gene confirmed that 19 strains deposited in the database as *B. amyloliquefaciens* were indeed *B. velezensis*. Although these species are phylogenetically close, combined analyzes of various genomic traits, such as the presence of biosynthetic genes encoding secondary metabolites, CRISPr/Cas arrays, ANI, dDDH, and other strain information, including source isolation, have allowed their unambiguous classification like these three species. This genomic analysis extends knowledge about related species, *B. velezensis*, *B. amyloliquefaciens* and *B. siamensis*, with an emphasis on taxonomic status.

Keywords: Root-knot nematodes, chalcones, biological control, genome sequencing, *Bacillus* sp., taxonomy.

LISTA DE FIGURAS

CAPÍTULO 1

LISTA DE FIGURAS

CAPÍTULO 2

LISTA DE TABELAS

CAPÍTULO 2

SUMÁRIO

CAPÍTULO 2

INTRODUÇÃO GERAL

As doenças de plantas são apontadas como um dos principais fatores de ordem bioeconômica a impor limitações ao desempenho das atividades agrícolas (BEBBER et al., 2014). Dentre os agentes causadores de doenças em plantas, os nematoides fitopatogênicos são responsáveis pela perda de 10% de toda a produção agrícola global (ZACHEO, 1993). Esta enorme perda é refletida pela extensa gama de hospedeiros destes fitopatógenos que engloba mais de 3.000 espécies de plantas, incluindo culturas economicamente importantes (ABAD et al., 2003; BEBBER et al., 2014). O impacto gerado por nematoides estão na produtividade, no aumento dos custos de produção e no incremento das interações biológicas danosas decorrentes da ação de outros microrganismos (KARSSEN e MOENS, 2006).

Durante as últimas décadas, grandes quantidades de pesticidas têm sido utilizadas principalmente na agricultura intensiva para proteção de cultivos (SIVASAKTHI et al., 2014). O uso abusivo desses produtos no campo levou ao surgimento de patógenos resistentes, resultando em fortes perdas econômicas (GEORGHIOU, 2012; MUÑOZ-LEOZ et al., 2013). Além disso, muitas moléculas foram retiradas do mercado devido aos seus efeitos residuais encontrados nos alimentos, a toxicidade ao homem e a contaminação ambiental (RUIZ-SUÁREZ et al., 2013; NORMAN et al., 2008; SANKHLA et al., 2018). Como consequência, novos métodos, mais eficientes no controle de fitonematoides tornaram-se desejados. Na busca por sustentabilidade na agricultura, as pesquisas foram direcionadas, em grande parte, para utilização de métodos ecológicos para o manejo de nematoides.

Neste contexto enfatizam-se a investigação do papel dos compostos naturais no controle de pragas e doenças (DIAZ-TIELAS et al., 2016). Explorar a química de moléculas já selecionadas pela natureza para desempenhar um papel na defesa das plantas surge como um ponto de partida essencial para procurar novos produtos ativos (DIAZ-TIELAS et al., 2016). Nos últimos anos, têm se intensificado os estudos empregando plantas que apresentam efeitos antagônicos a nematoides para serem utilizadas em rotação de culturas, aplicação direta no solo como tortas ou extratos vegetais e, também, para o isolamento das substâncias ativas (NTALLI e CABONI, 2012). Vários fitoquímicos, derivados do metabolismo secundário de plantas possuem atividades nematicidas comprovadas (BARROS et al., 2014; JARDIM *et al*., 2017, NTALLI et al., 2010). O objetivo dessas pesquisas é buscar compostos

com novos mecanismos de ação, diferentes daqueles utilizados pelos produtos químicos existentes no mercado, a fim de obter compostos eficientes e de baixo impacto ambiental.

Metabólitos secundários são continuamente encontrados em bactérias, fungos e plantas, e constituem manancial para descoberta de novas drogas. Asssim esses compostos são empregados como matéria-prima para a síntese de novos compostos que retenham ou aumentem suas propriedades de controle (CECHINEL-FILHO e YUNES, 2001). É importante mencionar que a utilização de produtos naturais ativos como molécula-protótipo para a síntese de análogos mais potentes e seletivos, tem contribuído significativamente para a obtenção de inúmeros fármacos utilizados na medicina, que podem, muitas vezes, ser obtidos mais facilmente e a custos menores (VIEGAS et al., 2006; SCHENKEL et al., 2001). Dessa forma, a síntese de análogos também representa um alvo promissor para o desenvolvimento de novos produtos nematicidas.

Os flavonoides são compostos biossintetizados a partir da via dos fenilpropanoides nas plantas, constituindo assim uma importante classe de polifenóis, presentes em relativa abundância entre os metabólitos vegetais (KUMAR e PANDEY, 2013). Nas plantas, as chalconas são precursores dos flavonoides. Quimicamente as chalconas são definidas como cetonas aromáticas α,β-insaturadas (PIÑERO et al., 2006). As chalconas são de grande interesse químico e farmacológico por apresentarem diversas atividades biológicas, as quais variam conforme os diferentes substituintes destas moléculas (SOMANI et al., 2017). São referenciados efeitos antibacterianos, antifúngicos, antitumoral, antioxidante, entre outros (DIAZ-TIELAS et al., 2016; ROZMER e PERJESI, 2016). Devido às numerosas atividades biológicas destes metabólitos secundários, existem muitas possibilidades para seu uso na agricultura.

Considerando o potencial das chalconas naturais e sintéticas como novos nematicidas para o controle de nematoides parasitas de plantas, este estudo objetivou sintetizar análogos de chalconas para verificar suas atividades contra o nematoide das galhas *Meloidogyne incognita*. Além disso, a estrutura química do análogo mais ativo foi empregada em um estudo *in silico* para descobrir seu alvo enzimático no nematoide.

REFERÊNCIAS

ABAD, P.; FAVERY, B.; ROSSO, M. N.; CASTAGNONE-SERENO, P. Root-knot nematode parasitism and host response: Molecular basis of a sophisticated interaction. **Molecular Plant Pathology**, v. 4, p. 217-224, 2003.

BARROS, A. F.; CAMPOS, V. P.; SILVA, J. C. P.; POZZOBON, M.; MEDEIROS, F. H. V.; POZZA, E. A.; REALE, A. L. Nematicidal activity of volatile organic compounds emitted by *Brassica juncea*, *Azadirachta indica*, *Canavalia ensiformis*, *Mucuna pruriens* and *Cajanus cajan* against *Meloidogyne incognita*. Applied Soil Ecology, v. 80, p. 34-43, 2014.

BEBBER, D. P.; HOLMES, T.; GURR, S. J. The global spread of crop pests and pathogens. **Global Ecology Biogeography**, v. 23, p. 1398–1407, 2014.

CECHINEL-FILHO, V; YUNES, R. A. Estudo Químico de Plantas Medicinais Orientado para a Análise Biológica. Obtenção, Determinação e Modificação Estrutural de Compostos Bioativos. In: YUNES, R. A.; CALIXTO, J. B. (ed.). **Plantas Medicinais: sob a ótica da Química Medicinal Moderna**. Chapecó: Argos, 2001. Cap. 2, p. 59-69.

DIAZ-TIELAS, C.; GRAÑA, E.; REIGOSA, M. J.; SÁNCHEZ-MOREIRAS, A. M. Biological activities and novel applications of chalcones. **Planta Daninha**, v. 34, n. 3, p. 607-616, 2016.

GEORGHIOU, G. P. **Pest resistance to pesticides**. Springer Science & Business Media, Boston, MA, 2012.

JARDIM, I. N.; OLIVEIRA, D. F.; SILVA, G. H.; CAMPOS, V. P.; SOUZA, P. E. (E)-cinnamaldehyde from the esssential oil of *Cinnamomum cassia* controls *Meloidogyne incognita* in soybean plants. **Journal of Pest Science**, v. 91, n. 1, p. 479-487, 2017.

KARSSEN, G.; MOENS, M. Root-knot nematodes. In: PERRY, R. N.; MOENS, M. **Plant Nematology**. Wallingford, UK: CABI Publishing; 2006. pp. 59–90.

KUMAR, S.; PANDEY, A. K. Chemistry and Biological Activities of Flavonoids: An Overview. **The Scientific World Journal**, v. 2013, 16 p., 2013.

MUÑOZ-LEOZ, B.; GARBISU, C.; CHARCOSSET, J. Y.; SÁNCHEZ-PÉREZ, J. M.; ANTIGUEDAD, I.; RUIZ-ROMERA, E. Non-target effects of three formulated pesticides on microbially-mediated processes in a clay-loam soil. **Science of the Total Environment**, v. 449, p. 345-354, 2013.

NORMAN, C. S., DE CANIO, S. J., FAN, L. The Montreal Protocol at 20: Ongoing opportunities for integration with climate protection. **Global Environment Change**, v. 18, p. 330-340, 2008.

NTALLI, N. G.; CABONI, P. Botanical nematicides: A review. **Journal of Agricultural And Food Chemistry**, v. 60, p. 9929-9940, 2012.

NTALLI, N. G.; VARGIU, S.; MENKISSOGLU-SPIROUDI, U.; CABONI, P. Nematicidal carboxylic acids and aldehydes from *Melia azedarach* fruits. **Journal Agricultural and Food Chemistry**, v. 58, p. 11390-11394, 2010.

PIÑERO, J.; TEMPORAL, R. M.; SILVA-GONÇALVES, A. J.; JIMÉNEZ, I. A.; BAZZOCCHI, I. L.; OLIVA, A.; PERERA, A.; LEON, L. L.; VALLADARES, B. New administration model of trans-chalcone biodegradable polymers for the treatment of experimental leishmaniasis. **Acta Tropica**, v. 98, p. 59-65, 2006.

ROZMER, Z.; PERJÉSI, P. Naturally occurring chalcones and their biological activities. **Phytochemistry Reviews**, v. 15, n. 1, p. 87-120, 2016.

RUIZ-SUÁREZ, N., BOADA, L. D., HENRÍQUEZ-HERNÁNDEZ, L. A., GONZÁLEZ-MOREO, F., SUÁREZ-PÉREZ, A., CAMACHO, M., ZUMBADO, M., ALMEIDA-GONZÁLEZ, M., TRAVIESO-AJA, M. D. M., LUZARDO, O. P. Continued implication of the banned pesticides carbofuran and aldicarb in the poisoning of domestic and wild animals of the Canary Islands (Spain). **Science of the Total Environment**, v. 505, p. 1093-1099, 2015.

SANKHLA, M. S.; KUMARI, M.; SHARMA, K.; KUSHWAH, R. S.; KUMAR, R. Water contamination through pesticides and their toxic effect on human health. **International Journal for Research in Applied Science and Engineering Technology**, v. 6, p. 967-970, 2018.

SCHENKEL, E. P.; GOSMANN, G.; PETROVICK, P. R. Produtos de origem vegetal e o desenvolvimento de medicamentos. In: SIMÕES, C. M. O.; *et al.* (orgs.). **Farmacognosia: da planta ao medicamento**. 3 ed. Florianópolis: Editora da UFSC, Porto Alegre: Editora da UFRGS, 2001. Cap. 15. p. 309-315.

SIVASAKTHI, S.; USHARANI, G.; SARANRAJ, P. Biocontrol potentiality of plant growth promoting bacteria (PGPR) – *Pseudomonas fluorescens* and *Bacillus subtilis*: A review. **African Journal of Agricultural Research**, v. 9, p. 1265-1277, 2014.

SOMANI, R. R.; DEVENDRA, M.; PRATIK, B.; REWA, B. Optimization and study of variables in microwave assisted organic synthesis of some biologically active chalcones. **Indian Journal of Pharmacy and Pharmacology**, v. 4, n. 3, p. 153-160, 2017.

VIEGAS Jr., C.; BOLZANI, V. S.; BARREIRO, E. J. Os produtos naturais e a química medicinal moderna. **Química Nova**, v. 29, n. 2, p. 326-337, 2006.

ZACHEO, G. Introduction. In: KHAN, W. W. **Nematode Interactions**. London: Chapman and Hall; 1993. pp. 1-25.

ARTIGO 1 - Chalcone analogues: synthesis, activity against *Meloidogyne incognita*, and *in silico* interaction with cytochrome P450

Authors

Fabíola J. Silva[1], Vicente P. Campos[1], Denilson F. Oliveira[2], Vanessa A. Gomes[1], Aline F. Barros[1], Zia Ud Din[3], Edson Rodrigues-Filho[3]

Institutional Affiliation

[1] Department of Plant Pathology, Federal University of Lavras, 37200-000, Lavras-MG, Brazil.

[2] Department of Chemistry, Federal University of Lavras, 37200-000, Lavras-MG, Brazil.

[3] Department of Chemistry, Federal University of São Carlos, 13565-905, São Carlos-SP, Brazil.

Abstract

To contribute to the development of new products to control plant-parasitic nematodes, twelve chalcone analogs were synthesized and screened for activity against *Meloidogyne incognita*. Three caused mortality greater than negative controls in second-stage juvenile *M. incognita*, with values varying from 19.9% to 100%. The most active chalcone analog was (1*E*,4*E*)-1,5-di(4-nitrophenyl)-2-butylpenta-1,4-dien-3-one (compound **6**), which had an LC_{50} value of 41 µg/mL. Under the same conditions, the commercial nematicide Carbofuran® (2,2-dimethyl-2,3-dihydro-1-benzofuran-7-yl methylcarbamate) presented an LC_{50} equal to 101 µg/mL. When this chalcone analog was applied to tomato plants infested with *M. incognita*,

reductions in the numbers of galls and eggs of 51% and 68% were observed, respectively. According to *in silico* studies, the enzyme target of compound **6** in *M. incongita* is cytochrome P450, which is important for the oxidation of several substances in the nematode. Therefore, compound **6** is potentially useful for the development of new products to control *M. incongita*.

1 Introduction

Nematode species of the genus *Meloidogyne*, also known as root-knot nematodes, are economically important phytopathogens (Abad et al., 2008). They have a universal distribution, parasitizing thousands of plants, including several crops of great economic importance, such as coffee, tomato and cotton (Jones et al., 2013). They are obligate parasites, feeding and reproducing within roots. They induce gall formation, which physiologically disorganizes the infected plant, and makes water and nutrients absorption through roots difficult, which reduce the quantity and quality of agricultural production (Bebber et al., 2014).

Despite the great damage that plant-parasitic nematodes can cause, currently available nematicidal products are not as efficient as desired, in part because some products have been withdrawn from the market due to their high toxicity to man and the environment. An example is the nematicide Aldicarbe [2-methyl-2-(methylthio)propanal *O*-(*N*-methylcarbamoyl)oxime], whose use has been forbidden in several countries (Ruiz-Suárez et al., 2013). Another example is methyl bromide, which has been banned because of its destructive effect on the ozone layer (Norman et al., 2008). Due to this gradual elimination of

nematicides from the market, there is an increasing global demand for new molecules and methods to control plant-parasitic nematodes, preferably at lower costs, better efficiencies and lower environmental and human toxicities hazard.

One possibility to overcome this problem is using phytochemicals as lead compounds in the development of new commercial products (Zhang et al., 2017; Kumari et al., 2014; Ntalli & Caboni, 2012; Attar et al., 2011). There are several metabolites produced by plants with a surprising diversity of chemical structures; some of them have proven activity against plant-parasitic nematodes (Nunes et al., 2013; Edens et al., 1995). This is the case, for example, of chalcones, which can be produced by plants in response to nematode attack (Edens et al., 1995). Gonzalez and Estevez-Braun (1997, 1998) demonstrated the high nematicidal and nematostatic activity of (*E*)-chalcones against the potato cyst nematodes *Globodera pallida* (Stone) Behrens and *G. rostochiensis* (Wollenweber) Behrens. In addition, *in vitro* studies have demonstrated that chalcones cause mortality in second-stage juvenile (J2) *Meloidogyne incognita* (Kofoid & White) Chitwood (Caboni et al., 2016), and inhibition of *Radopholus similis* (Cobb) Throne hatching (Wuyts et al., 2006). Motivated by the potential of chalcones to control plant-parasitic nematodes, Nunes et al. (2006) sought better analogs of the most efficient natural chalcones for controlling *Meloidogyne exigua* Goeldi. They synthesized the chalcone (*E*)-3-(2,4,5-trimethoxyphenyl)-1-(4-nitrophenyl)prop-2-en-1-one, which was active against *M. exigua* Goeldi, possibly through the inhibition of the enzyme caffeic acid 3-*O*-methyltransferase.

Considering the potential of natural and synthetic chalcones as new nematicides for the controlling plant-parasitic nematodes, this study initially aimed at synthesizing, through aldol condensations, chalcone analogs with carbon-carbon double bonds between both aromatic rings and the carbonyl group. In other words, instead of working with derivatives of (*E*)-1,3-diphenylprop-2-en-1-one, we studied derivatives of (*E*,*E*)-1,5-diphenylpenta-1,4-dien-

3-one. Then, the synthesized compounds were submitted to *in vitro* tests with *M. incognita*. The most active chalcone analog was then tested *in vivo* with nematode-infected tomato plants. Furthermore, the chemical structure of the most active chalcone analog was employed in an *in silico* study to find out its enzymatic target in the nematode.

2 Materials and Methods

2.1 Synthesis of chalcone analogues 1-6

All chemicals were purchased from Organics, Sigma-Aldrich, Acros Chemicals and Fisher Scientific Ltd, to be used without further purification. Deuterated solvents from Appolo were used for the nuclear magnetic resonance (NMR) analysis. Thin layer chromatography (TLC) analysis was performed with precoated silica gel G-25-UV254 plates, and detection was carried out at 254 nm under UV, and by vanillin in H_2SO_4 solution. Hydrogen (^1H) and carbon thirteen (^{13}C) NMR analyses were performed on a Brüker AVANCE 400 operating at 400.15 MHz and 100.62 MHz, respectively, $CDCl_3$ was used as solvent and tetramethylsilane (TMS) as an internal reference. Each compound was dissolved in $CDCl_3$ (5-10 mg/mL) and poured into a NMR tube (5 mm). Chemical shifts (δ in ppm) were measured with accuracies of 0.01 and 0.1 ppm for ^1H and ^{13}C, respectively.

The reactions of aldehydes AL1 (*p*-nitrobenzaldehyde, *p*-trifluoromethylbenzladehyde and *p*-methoxybenzaldehyde, 80-95 mmole) with asymmetric ketones K (butan-2-one, pentan-2-one and heptan-2-one, 18.9 mmole), using gaseous HCl as a catalyst, was performed at room temperature by stirring the reaction mixtures and passing dry gaseous HCl through until the medium turned color. In this step, the formation of intermediates AL1K, which underwent condensation to aldehydes AL2 under basic conditions to create chalcone analogs

1-6 (Figure 1a), occurred. A Solution of intermediates AL1K (3.125 mmole), and differently substituted benzaldehydes (3.75 mmole) in ethanol (5 mL) was stirred for 5 min at room temperature. Then, a solution of sodium hydroxide in ethanol (4 mL, 50 mmole) was added and stirring was continued for 7 h. Ethanol was removed on a rotary evaporator and the resulting residue was dissolved in ethyl acetate to result in a solution that underwent partition with an aqueous saturated $NaHSO_3$ solution. The ethyl acetate phase was dried over anhydrous Na_2SO_4 and concentrated to dryness on a rotary evaporator. Each crude product was collected as a yellow precipitate that was further purified by column chromatography (ethyl acetate/hexane) and recrystallization from ethanol.

2.2 Synthesis of chalcone analogues 7-12

The purchases of chemical reagents and the analyzes were done as described above, in item 2.1. The aldehydes AL1 (5.0-7.3 mmole) were stirred with dimethylammonium dimethylcarbamate (DIMCARB, 0.2 equivalent) in a water-ethanol (10 mL) solution for 5 min to generate the corresponding reactive imine species. The added cyclopentanone (C, 10.95 mmole) was converted to its enol form, which attacked the imines to give the products AL1C. These were condensed to aldehydes, AL2 (0.943 mmole), under basic conditions, as described for compounds **1-6**, to afford the chalcone analogs **7-12** (Figure 1b).

2.3 Obtaining eggs and second-stage juveniles (J2) of *M. incognita*

In accordance with the method described by Hussey and Barker (1973), *M. incognita* eggs were extracted from the roots of tomato plants (*Solanum lycopersicon* L. 'Santa Clara') grown under greenhouse conditions, which had been artificially infested with the nematode.

Eggs retained in a 500-mesh sieve (American Society for Testing and Materials, ASTM) were used in the experiments or transferred to hatching chambers and incubated at 28 $^{\circ}$C. Second-stage juveniles (J2) of the nematode that hatched during the first 24 h were discarded. Only J2 with a maximum age of 48 h after hatching were used in the experiments.

2.4 Selection of chalcone analogues active against *M. incognita in vitro*

Compounds **1–12**, obtained as described above, were employed in this experiment that was carried out acording to the method described by Chen and Dickson (2000) and adapted by Amaral et al. (2003). An aqueous suspension (20 μL) containing approximately 25 *M. incognita* J2 individuals and 100 μL of one of the test compounds dissolved in an aqueous Tween 80$^{\circledR}$ (0.01 g/mL) solution were pipetted into a well of a 96-well polypropylene plate. The final concentration of each compound in the well was 500 μg/mL. The commercial nematicide Carbofuran$^{\circledR}$ (2,2-dimethyl-2,3-dihydro-1-benzofuran-7-yl methylcarbamate, 98%, Sigma-Aldrich), at a final concentration (in the well) of 170 μg/mL, was used as a positive control. Negative controls were an aqueous solution of Tween 80$^{\circledR}$ (0.01 g/mL) and water. After exposing the J2 to compounds and controls for 48 hours at 28 $^{\circ}$C, mobile and immobile J2 were counted using an optical microscope. Mortality was assessed after the addition of one drop of an aqueous 1.0 mol/L NaOH solution to the contents of each well. Nematodes that changed their body shape within 2 minutes after addition were considered alive, whereas the others, which remained still, were considered dead. This experiment was performed once, with five replicates per treatment. Only chalcone analogs that caused mortalities above 80% were selected for subsequent trials.

2.5 Determination of the 50% lethal concentration (LC$_{50}$) to *M. incognita* J2

Carbofuran$^®$ solutions (final concentrations in the wells: 500, 250, 210, 170, 130 and 90 μg/mL) and chalcone analogue **6** (final concentrations in the wells: 500, 250, 125, 100, 75, 50 and 25 μg/mL) were used in the *in vitro* assay with *M. incognita* J2, following the same methodology described previously. Water and an aqueous solution of Tween 80$^®$ (0.01 g/mL) were employed as negative controls. This experiment was repeated twice, with five replicates per treatment.

2.6 Hatching of *M. incognita* J2 from eggs exposed to chalcone analogue 6

An aqueous suspension (500 μL) containing approximately 1,000 *M. incognita* eggs and 500 μL of chalcone analog **6** solution dissolved in an aqueous Tween 80$^®$ (0.01 g/mL) solution, were poured into microtubes (1.5 mL). The final concentrations of compound **6** were 500, 250, 125, 100, 75, 50 and 25 μg/mL. Carbofuran$^®$ (final concentrations in the tubes: 500, 250, 210, 170, 130 e 90 μg/mL) was used as the positive control. Negative controls were water and an aqueous Tween 80$^®$ (0.01 g/mL) solution. After seven days at a temperature of 28 ^{o}C, intact eggs and hatched J2 (alive or dead) were counted. This experiment was repeated twice, with five replicates per treatment.

2.7 *In Vivo* assay with tomato plants

Tomato seeds (*S. lycopersicum* L. 'Santa Clara') known to be susceptible to *M. incognita* were sown on a commercial substrate (Tropstrato, Vida Verde Indústria e Comércio de Insumos Orgânicos Ltda., Mogi Mirim, São Paulo, Brazil), contained in 72-well styrofoam

trays (125 cm^3), in a greenhouse. Chalcone analog and Carbofuran$^®$ was dissolved in an aqueous solution of Tween 80$^®$ (0.01 g/mL) to concentration of 800 μg/mL. At the time of inoculation (seedling with three pairs of leaves), each solution (4.0 mL) was combined with an aqueous suspension (4.0 mL) containing 500 J2, which reduced the concentrations by half (final concentrations 400 μg/mL). Water (4.0 mL) and an aqueous Tween 80$^®$ (0.01 g/mL) solution (4.0 mL) were used as negative controls, to which the J2 suspension (4.0 mL) was also added. Immediately after preparation, each suspension was applied on the substrate around a plant through three equidistant holes (1.0 cm wide and 4.0 cm deep) around the stem. The seedlings were kept in a room for 48 h to avoid sunshine dryness and then transferred to a greenhouse. The temperature range within the greenhouse was 25-29°C, with photoperiod of 12 hours. After 40 days, the roots were removed, washed thoroughly, dried on a paper towel and weighed. After counting galls, the roots were subjected to egg extraction according to the method described by Hussey and Barker (1973). Those eggs trapped in a 500 mesh sieve (ASTM) were suspended in 50 mL of water and counted in a Peters chamber using an optical microscope. The obtained data were expressed as the number of galls and eggs per root system. This experiment was repeated twice, with six replicates per treatment.

2.8 Experimental design and statistical analysis

For all experiments, the design was completely randomized. The replicates of the experiments were combined when there were no significant interactions between treatments and experiments ($P > 0.05$) and analyzed separately when interactions occurred ($P < 0.05$). The results were previously submitted to a normality test (Shapiro-Wilk test) and homogeneity of error variance (Bartlett´s test). As no transformation was required, the F-test was performed through analysis of variance (ANOVA). When the F-test was significant ($P <$

0.05), the mean values of the treatments were compared according to the Student´s T-test or Scott-Knott´s test ($P < 0.05$). Sisvar software v. 5.6 was used to carry out statistical analyses (Ferreira, 2014). To calculate LC_{50}, mortality values were converted into a percentage and submitted to Logit analysis using the drc package of the R® computer program (R Development Core Team) (Ritz, 2018). SigmaPlot® software v. 12 was used to prepare the graphs.

2.9 *In silico* conformational search

Using ChemSketch 12.01 (https://www.acdlabs.com/) software, the chemical structures of chalcone analogue **6**, (2*E*,6*E*)-2,6-bis(4-hydroxybenzylidene)cyclohexanone (A0), and (2*E*,6*E*)-2,6-dibenzylidenecyclohexanone (A2) (Appiah-Opong et al., 2008) were drawn to undergo conformational searches employing Open3Dalign 2.3 software (Tosco et al., 2011). Each search was done through 1,000 dynamic molecular simulations at 300 ^{0}C, with 1,000 1 fs steps in each. These calculations were carried out with the Merck Molecular Force Field (MMFF94), considering the solvent (water) implicitly through generalized Born and surface area continuum solvation (GBSA). The most stable conformation, as well as those up to 10 kcal/mol less stable, underwent optimization with Mopac2016 17.270L software (Stewart, 2012). Water (solvent) was implicitly considered through the conductor-like screening model (COSMO). Those conformations with the lowest energies were the most stable.

2.10 Pharmacophoric search

The three-dimensional structures of protein ligands in the Ligand Expo database (http://ligand-expo.rcsb.org/) (Feng et al., 2004) were submitted to hydrogen addition by

OpenBabel 2.3.2 software (O'Boyle et al., 2011), to be compared to the two most stable conformations of chalcone analogue **6** by Lisica 1.0.1 software (Lesnik et al., 2015), which used standard values for all parameters. The same procedure was applied to a search in the ZINC database (Irwin et al., 2012). Only those substances with a Tanimoto score equal to or greater than 0.5 were selected (Peón et al., 2017). Among the substances selected from the ZINC database, only those with known enzymatic activity were further selected.

2.11 Cytochrome P450

The amino acids sequences of four groups of cytochrome P450 were downloaded from the RSCB Protein Data Bank (http://www.pdb.org) (Berman et al., 2000): 1A2 (CYP1A2), 2C9 (CYP2C9), 2D6 (CYP2D6), and 3A4 (CYP3A4). The sequences in each group were aligned using the Clustal Omega 1.2.1 algorithm (Sievers et al., 2011) in Ugene 1.30.0 software (Okonechnikov et al., 2012). Except for increasing the number of iterations to 10, all parameters were set to default values. Hamming dissimilarities (%) were then calculated using the same software, which considered all gaps in the alignments. For each pair of sequences with greater than or equal to 90% similarity (100 − Hamming dissimilarity), one sequence was discarded. The National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) was searched for amino acid sequences in nematode genomes similar to the remaining amino acids, using Blastp 2.8.0+ software (Altschul et al., 1997; Schäffer et al., 2001) with Delta-Blast (Domain Enhanced Lookup Time Accelerated BLAST) (Boratyn et al., 2012) set to the default parameters. This search was carried out in all non-redundant (nr) databases. Only those sequences of cytochrome P450 downloaded from the RCSB Protein Data Bank similar (score > 200) to amino acids sequences of cytochrome P450 in nematode genomes of were selected for the next step of the *in silico* study.

2.12 Blind docking

The following calculations were performed only with non-mutant three-dimensional structures of the cytochrome P450 selected through the NCBI search and of those with similarities (100 – Hamming dissimilarity) equal to or greater than 90% to the selected cytochrome P450. Initially, they were submitted to the MakeMultimer.py Python script (http://watcut.uwaterloo.ca/tools/makemultimer/) and the resulting three-dimensional structures underwent missing amino acid reconstruction using Jackal 1.5 software (http://honig.c2b2.columbia.edu/jackal). The resulting three-dimensional structures were then optimized using Chimera 1.10.1 software (Pettersen et al., 2004) through 10 conjugate gradient steps, followed by 200 steepest descent steps. All molecules (water, ligands, metal etc.) initially present in the structures downloaded from the RCSB Protein Data Bank were conserved during this process. The three-dimensional structures were aligned to the optimized structure of the CYP1A2 2hi4 (Sansen et al., 2007) through Lovoalign 1.1.0 software (Martínez et al., 2007), which also calculated the root-mean-square deviations (RMSD) between the remaining enzymes. Only the optimized three-dimensional structures of one CYP1A2 (2hi4) (Sansen et al., 2007) and three CYP2C9 (1r9o, 5k7k, 5w0c) (Wester et al., 2004; Swain et al., 2017; Liu et al., 2017) were converted to the pdbqt format through AutodockTools 1.5.7rc1 software (Morris et al., 2009). In a similar way, the most stable conformations of A0 and A2, according to optimizations carried out with Mopac2016 as described above, were converted to the pdbqt format by AutodockTools 1.5.7rc1 (Morris et al., 2009). Using a of 30x30x30 Å (x, y, z) grid box, which was continually moved 10 Å in the x, y and z directions to cover the three-dimensional structures of the enzymes, A2 and A0 were respectively docked to enzymes in the CYP1A2 (2hi4) and CYP2C9 (1r9o, 5k7k and 5w0c) groups, using Autodock Vina 1.1.2 software (Trott & Olson, 2010). As the enzymes were contained in a 50x81x59 Å (x, y, z) box, a total of 72 dockings were done for each

ligand in one enzyme. Except for the exhaustiveness parameter that was set to 32, all parameters were used with the default values. The binding sites of A0 and A2 were those regions in the enzymes for which these substances presented more affinity according to the calculations carried out with Autodock Vina.

2.12 Docking to Selected Regions

The most stable conformations of chalcone analog **6**, A2 and A0, after conversion to the pdbqt format with Autodock Tools 1.5.7rc1 software (Morris et al., 2009), were docked to two regions selected during the blind docking, in the selected cytochrome P450. The first region (selected in the CYP1A2 group) had its center at 0.54, -2.70, and -25.01 Å (x, y, z), and a size of 21.00x29.00x20.00 Å (x, y, z). The second region (selected in the CYP2C9 group) had its center at 4.60, 3.76 and -7.10 Å (x, y, z), and a size of 12.23x13.87x17.66 Å (x, y, z). Chalcone analog **6** was docked to all enzymes in the CYP1A2 and CYP2C9 groups, while A2 and A0 were docked to CYP1A2 and CYP2C9, respectively. These calculations were done by Autodock Vina 1.1.2 (Trott & Olson, 2010) with the exhaustiveness set to 256. All the other parameters were set to default values. The most stable conformations of chalcone analog **6**, A2 and A0, in the pdbqt format, were also used to calculate atomic affinities with Autogrid 4.2.6 software (Morris et al., 2009) in the two regions selected during blind docking. Then, Autodock 4.2.6 (Morris et al., 2009) used such affinities to dock A2 and A0 to CYP1A2 and CYP2C9, respectively. Chalcone analog **6** was docked to enzymes in both groups using Autodock 4.2.6. Except for the parameter ga_num_evals_parameter, which was raised from 2500000 to 5000000, all other parameters were used with the default values.

3 Results

3.1 Synthesis of chalcone analogues

The use of aldol condensations allowed the synthesis of the chalcone analogs **1–12** in 50–83% overall yields (Figure 1). The reaction of AL1 aldehydes with asymmetrical ketones K, using gaseous HCl as a catalyst, resulted in almost 100% conversion to the intermediate AL1K, that underwent condensations with AL2 aldehydes to generate chalcone analogs **1–6** 60–75% yields (Figure 1a). When dimethylammonium dimethylcarbamate (DIMCARB) amine was employed as a catalyst in the reaction of AL1 aldehydes with cyclopentanone (C), the AL1C intermediates were obtained in 80% yields, to be further reacted with the AL2 aldehyde to form chalcone analogues **7–12** in 50–83 % yields (Figure 1b).

3.2 Selection of chalcone analogues active against *M. incognita* J2

Of the twelve chalcone analogs studied (Figure 1), three of them (**3**, **6**, **8**) increased immobility and mortality of *M. incognita* J2 compared to the negative controls (Figure 2). Among them, compound **6** showed the highest nematicidal activity, causing 100% immobility and mortality to J2 exposed to this compound at 500 µg/mL for 48 h. The mortality value obtained for chalcone analog **6** was statistically higher than that observed for the positive control (71% mortality) at a concentration of 170 µg/mL. Regarding chalcone analogs **3** and **8**, the mortality caused to J2 were 19.9% and 44.0%, respectively.
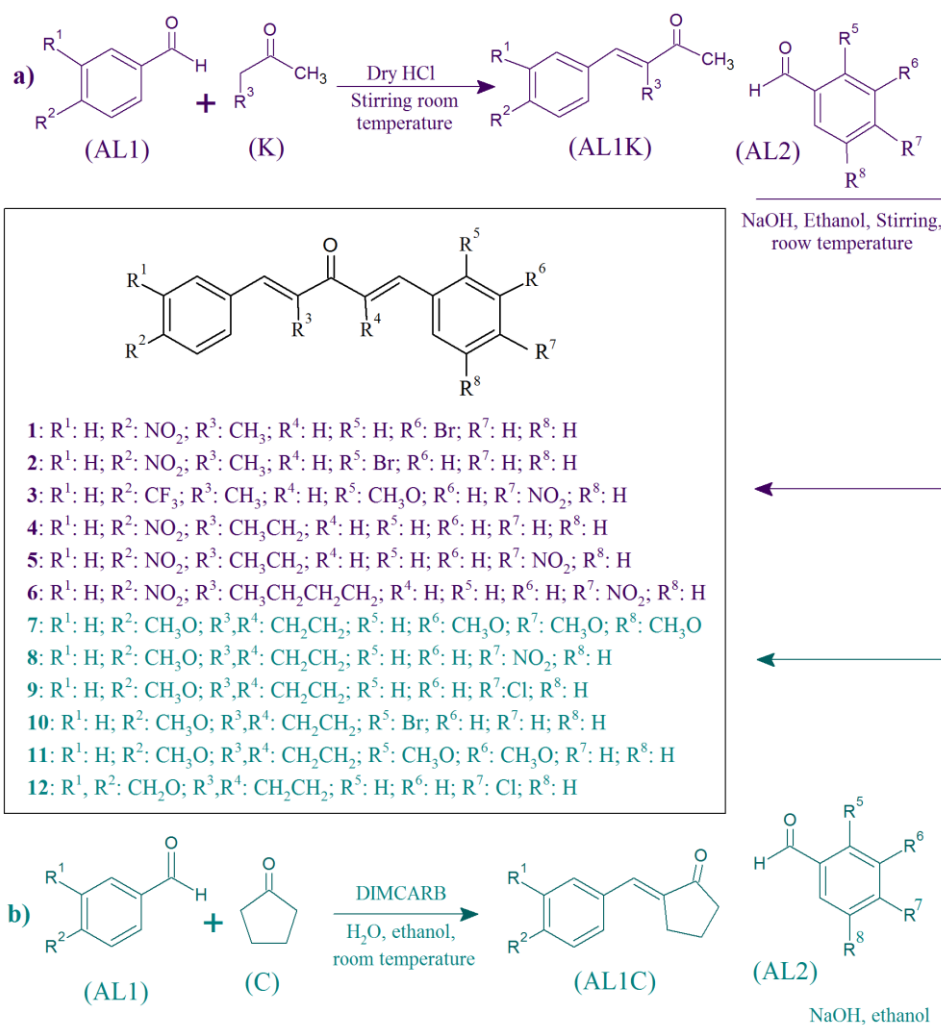
Figure 1. A) Acid-catalyzed aldol condensation of aldehydes (AL1) with ketones (K) to form intermediates (AL1K) that underwent basic catalyzed aldol condensations with aldehydes (AL2) for form chalcone analogs **1–6**. B) DIMCARB (dimethylammonium dimethylcarbamate) catalyzed aldol condensation of aldehydes (AL1) with cyclopentanone (C) to form the intermediates (AL1C) that are submitted to basic catalyzed aldol condensations with an aldehyde (AL2) to form chalcone analogs **7–12**.
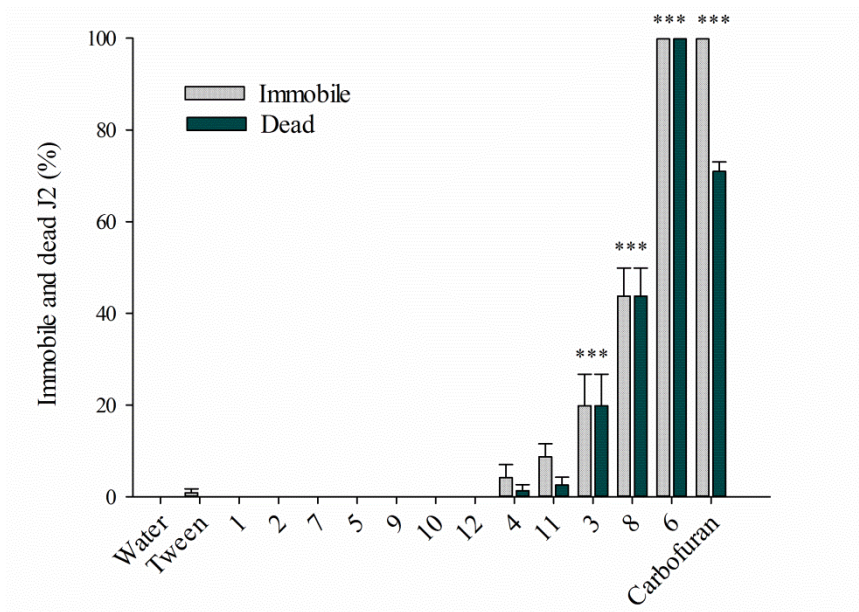
Figure 2. Immobile and dead second-stage juveniles (J2) of *Meloidogyne incognita*, after 48 h of exposure to chalcone analogs **1–12**, at the concentration of 500 µg/mL. Carbofuran[®] (170 µg/mL), water and Tween 80[®] were used as controls. Vertical bars represent the standard error of the mean. *** Significance at the 0.01 probability level according to the Student's T-test when compared to the negative control (Tween 80[®]).

3.3 Determination of the 50% lethal concentration (LC_{50}) to *M. incognita* J2

Increasing concentrations of chalcone analog **6** from 50 µg/mL to 250 µg/mL promoted progressive quadratic increases in the percentage of *M. incognita* J2 mortality, while increases in Carbofuran[®] concentration provided a linear behavior in J2 mortality (Figure 3). LC_{50} values of 41 µg/mL and 101 µg/mL were obtained for compound **6** and Carbofuran[®], respectively.
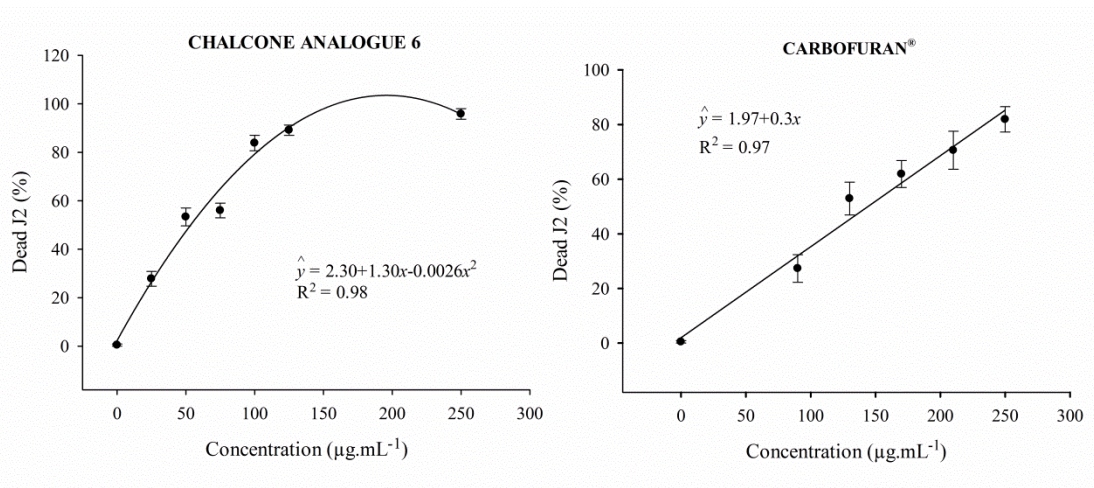
Figure 3. Mortality of *Meloidogyne incognita* second-estage juveniles (J2), after exposure to solutions of chalcone analog **6** at concentrations in the 25-250 µg/mL range, and to solutions of the commercial nematicide Carbofuran[®] at concentrations in the 90-250 µg/mL range. The graphs comprise averages of two experiments, with five replicates for each concentration.

3.4 Hatching of *M. incognita* J2 from eggs exposed to chalcone analogue 6

Exposure of *M. incognita* eggs to solutions of chalcone analog **6** for seven days significantly reduced the number of J2 hatched when compared to the negative controls (Figure 4). The highest hatching decrease occurred at concentrations of 500 and 250 µg/mL, reaching, respectively, 56% and 69% reductions compared to those observed in water (control). The J2 hatching at a concentration of 75 µg/mL was similar to the negative controls in both experiments and significantly equal to that at a concentration of 100 µg/mL in experiment 2 ($P < 0.05$). Carbofuran[®] did not reduce hatching at any of the concentrations studied (P = 0.3928).
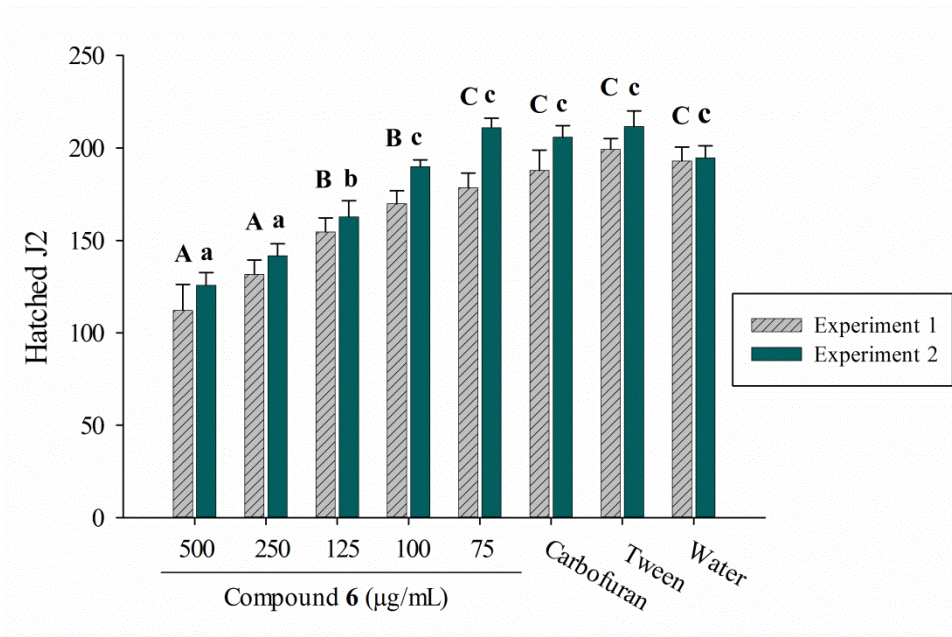
Figure 4. Hatching of *Meloidogyne incognita* second-stage juveniles (J2) from eggs exposed for seven days to chalcone analog **6** at different concentrations and Carbofuran® at a concentration of 500 µg/mL. Water and aqueous Tween 80® solution were used as controls. Vertical bars represent standard error of the mean. The graphs comprise averages of two experiments, with five replicates for each concentration. Bars with the same capital letter (experiment 1) or small letters (experiment 2) do not differ from each other at the 0.05 probability level according to Scott-Knott's test.

3.5 *In vivo* assay with tomato plants

The application of either chalcone analog **6** or Carbofuran®, both at a concentration of 400 µg/mL, on tomato roots infested with *M. incognita*, resulted in the reduction ($P < 0.05$) of galls and eggs per root system compared to water and Tween 80® (controls) (Figure 5). When compared to the negative control (water), the number of galls was reduced by 65% and 75% by chalcone analog **6** and Carbofuran®, respectively, while the number of eggs was reduced,

respectively, by 63% and 79%. Root masses were similar in all treatments (P = 0.4703), ranging from 4.7 to 5.2 grams.
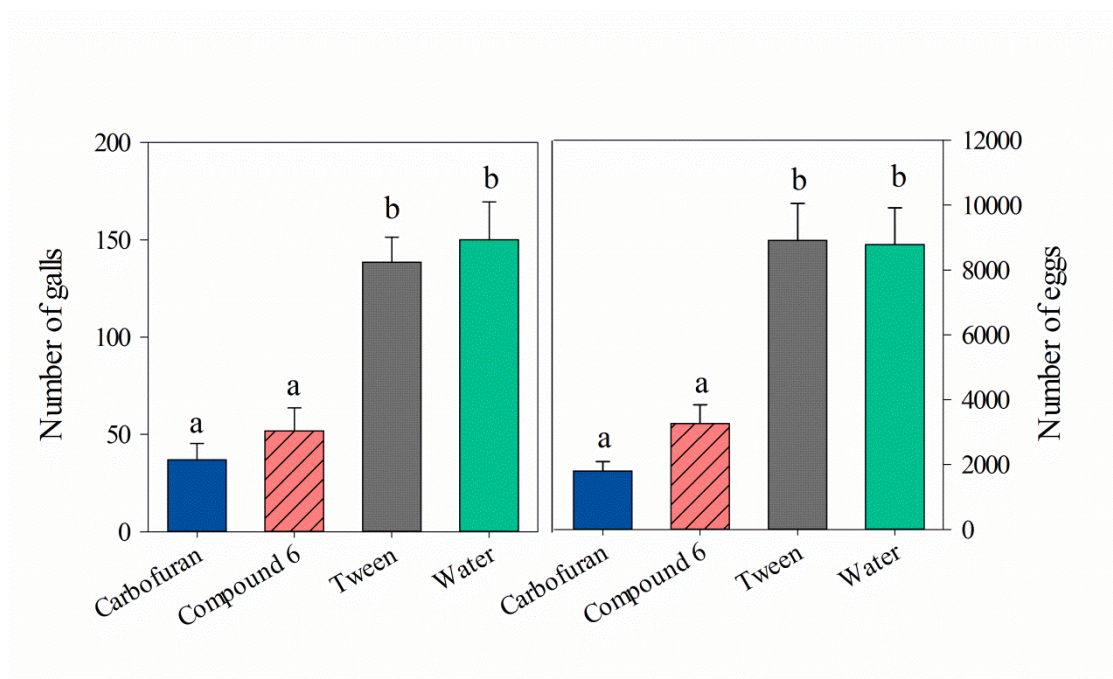
Figure 5. Numbers of galls and eggs of *Meloidogyne incognita* per root system of tomatoes inoculated with second-stage juveniles and treated with chalcone analog **6** and Carbofuran[®] solutions, both at a concentration of 400 µg/mL. Water and Tween 80[®] were used as controls. Vertical bars represent standard error of the mean. Columns with the same letter in each graph do not differ from each other at the 0.05 probability level according to Scott-Knott's test. The graphs comprise averages of two experiments, with six replicates for each treatment.

3.6 *In silico* study

During the conformational search, several conformations with close values of energies were observed for chalcone analog **6**. When two representative conformations were employed in a pharmacophoric search in the Ligand Expo database, no result was obtained with a Tanimoto score above 0.4. The same search performed in the ZINC database resulted in the

selection of several substances with Tanimoto scores above 0.5, among which the only substance with reported enzyme inhibitory activity was (2*E*,6*E*)-2,6-dibenzylidenecyclohexanone (A2, code ZINC38528035, http://zinc.docking.org/substance/38528035), with a Tanimoto score of 0.51. According to Appiah-Opong et al. (2008) A2 is a strong inhibitor of cytochrome P450 1A2 (CYP1A2).

In the search for amino acid sequences similar to those of the cytochrome P450 1A2 (CYP1A2), 2C9 (CYP2C9), 2D6 (CYP2D6) and 3A4 (CYP3A4), which was carried out in the nematode genomes, the following numbers of sequences with scores above 200 were respectively found: 35, 370, 235 and 110. After the elimination of mutant enzymes, followed by reconstruction of missing amino acid residues and optimization of the three-dimensional structure of the enzymes, only one CYP1A2 enzyme remained: 2hi4 (Sansen et al., 2007). After the same process, three CYP2C9 enzymes remained: 1r9o, 5k7k and 5w0c (Wester et al., 2004; Swain et al., 2017; Liu et al., 2017). For CYP2D6 and CYP3A4 no enzymes remained.

The blind docking of A2 to 2hi4 resulted in the selection of a cavity (CAV-1) on the enzyme surface, formed by the amino acid residues ASP119, GLN121, GLN265, VAL268, GLN269, GLU305, LYS306, VAL308 and ASN309 (Figure 6). In the five poses with more affinities (-8.3 kcal/mol) of A2 for 2hi4, the A2 positions in that cavity were identical. However, in the blind docking of A0 to 1r9o, 5k7k and 5w0c, a cavity (CAV-2) in the center of these enzymes, close to the protoporphyrin IX containing an iron (Fe) ion (HEME), was selected. In the case of the 5w0c enzyme, CAV-2 was formed by the amino acid residues PHE83, ARG91, VAL96, PHE97, ASN187, ILE188, LEU191, VAL220, MET223, ASN272, VAL275, ASP276, GLY279, ALA280, LEU345, LEU349, PRO350 and PHE459 (Figure 6).
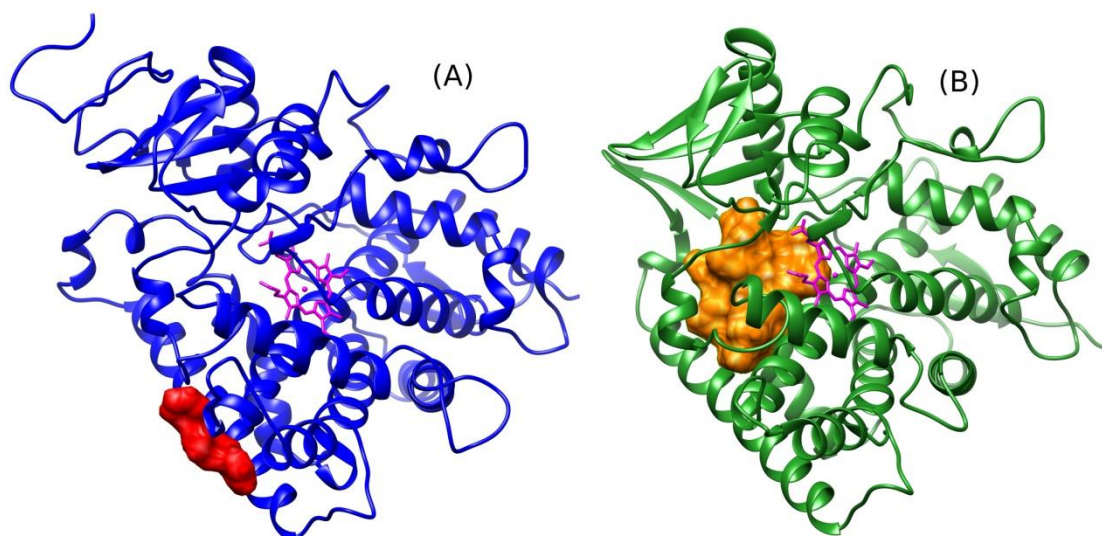
Figure 6. A) Three-dimensional structure of the cytochrome P450 1A2 2hi4 (dark blue), protoporphyrin IX containing an iron ion (HEME – magenta) and the cavity (CAV-1, red) on the surface of the enzyme that was selected through the blind docking of (2*E*,6*E*)-2,6-dibenzylidenecyclohexanone (A2) to 2hi4. B) three-dimensional structure of the cytochrome P450 2C9 5w0c (green), protoporphyrin IX containing an iron ion (HEME – magenta) and the cavity (CAV-2, orange) inside the enzyme, close to HEME, which was selected through the blind docking of (2*E*,6*E*)-2,6-bis(4-hydroxybenzylidene)cyclohexanone (A0) to 1r9o, 5k7k and 5w0c. Images were generated with the UCSF Chimera 1.11.1 program (Pettersen et al., 2004).
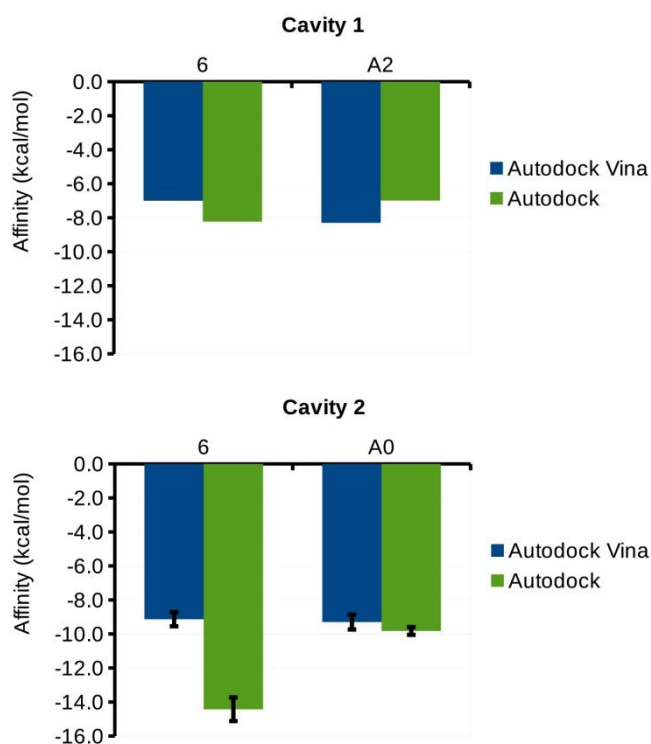
Figure 7. Affinities of (1*E*,4*E*)-1,5-di(4-nitrophenyl)-2-butylpenta-1,4-dien-3-one (**6**) and (2*E*,6*E*)-2,6-dibenzylidenecyclohexanone (A2) for the cavity CAV-1, localized on the surface of the cytochrome P450 1A2 2hi4 (Sansen et al., 2007), and affinities of **6** and (2*E*,6*E*)-2,6-bis(4-hydroxybenzylidene)cyclohexanone (A0) for the cavity CAV-2 that is inside the cytochrome P450 2C9 1r9o, 5k7k and 5w0c (Wester et al., 2004; Swain et al., 2017; Liu et al., 2017). Calculations were done with Autodock Vina 1.1.2 and Autodock 4.2.6 (Morris et al., 2009; Trott & Olson, 2010). Error bars correspond to the standard deviation. It is possible to observe that poses of compound A2 and **6** are reasonably overlapped in CAV-1. Since chalcone analog **6** is larger, it has more interactions with the amino acid residues of the enzyme 2hi4 (Figure 8).
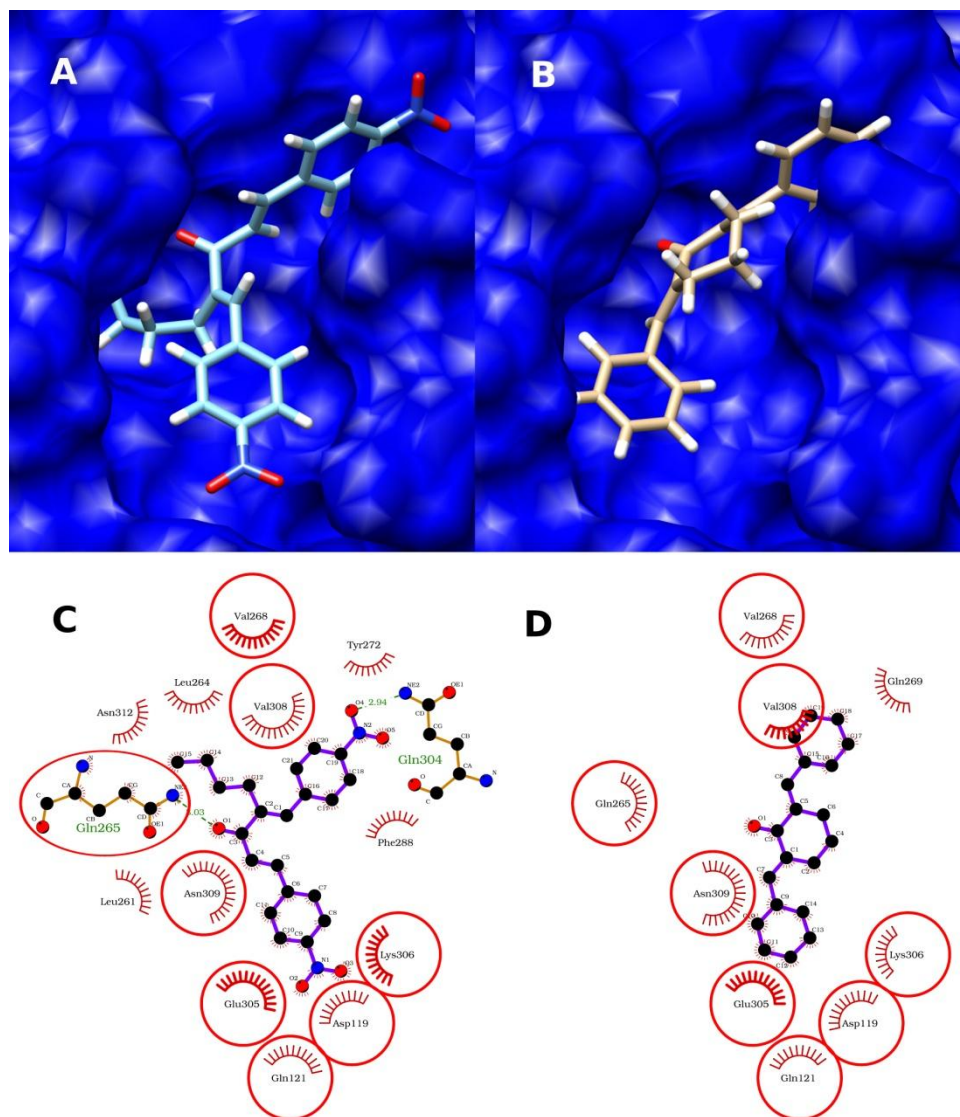
Figure 8. A) Three-dimensional structure of chalcone analog **6** docked to the surface cavity CAV-1 of cytochrome P450 1A2 2hi4 by the computer program Autodock 4.2.6 (Morris et al., 2009), B) Three-dimensional structure of the compound A2 docked to CAV-1 of cytochrome P450 1A2 2hi4 by Autodock 4.2.6, C) Two-dimensional representation of chalcone **6** docked to CAV-1, D) Two-dimensional representation of A2 docked to CAV-1. The three-dimensional images were generated with the UCSF Chimera 1.11.1 program (Pettersen et al., 2004), while the two-dimensional images were generated with LigPlot+ 1.4.5 software (Wallace et al., 1995).
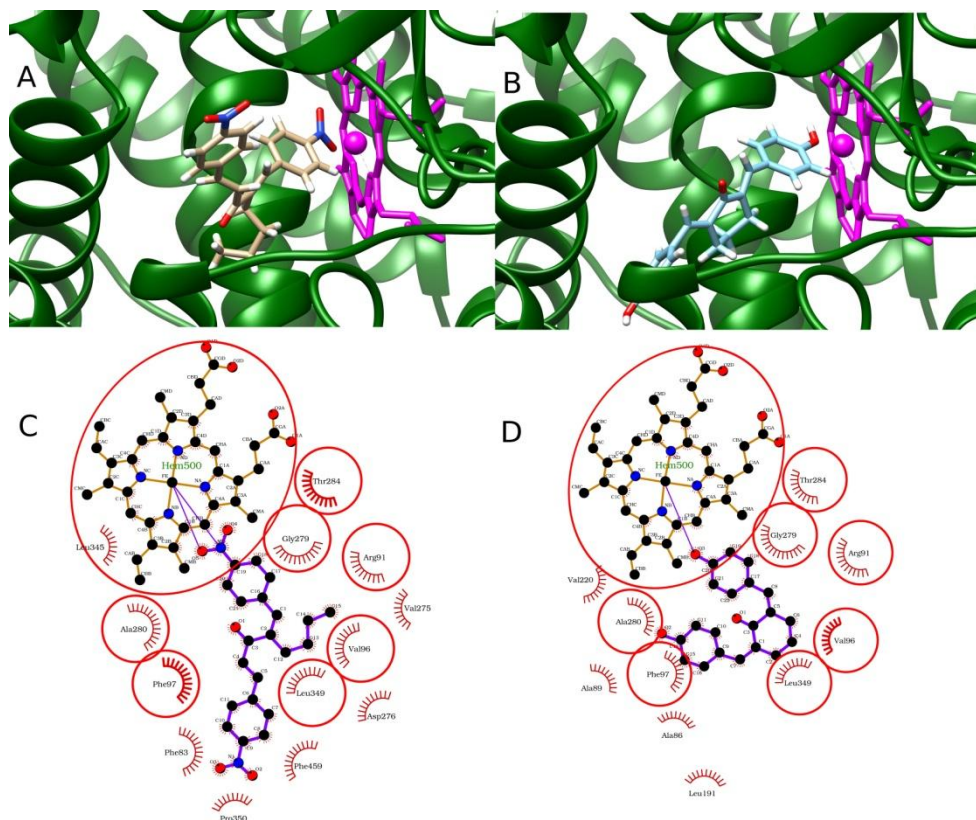
Figure 9. A) Three-dimensional structure of chalcone analog **6** docked to cavity CAV-2 of cytochrome P450 2C9 5w0c by the Autodock 4.2.6 (Morris et al., 2009), B) Three-dimensional structure of the compound A0 docked to CAV-2 of the cytochrome P450 2C9 5w0c by Autodock 4.2.6, C) Two-dimensional representation of chalcone 6 docked to CAV-2, D) Two-dimensional representation of A0 docked to CAV-2. The three-dimensional images were generated with the UCSF Chimera 1.11.1 program (Pettersen et al., 2004), while the two-dimensional images were generated with LigPlot+ 1.4.5 software (Wallace et al., 1995).

Apparently, compound **6** binds more efficiently to CAV-2 than to CAV-1. According to the Autodock Vina 1.1.2, compound A2, which is an inhibitor of CYP1A2, binds more efficiently than chalcone analog **6** to CAV-1. However, the values calculated with the Autodock 4.2.6 suggests the opposite. For CAV-2, Autodock Vina calculated similar values

for compound **6** and A0, which is an inhibitor of CYP2C9, However, according to Autodock, compound **6** binds more efficiently to the enzymes than compound A0 (Figure 7).

One of the aromatic rings of compound A0 and **6** are overlapped, but the other aromatic rings are pointing in different directions. In the overlapping rings, both the OH group from compound A0 and the nitro group from compound **6** interact with the iron íon of HEME (Figure 9).

4 Discussion

In the reactions of AL1 aldehydes with asymmetrical ketones K, using gaseous HCl as a catalyst, only the intermediates AL1K were detected in the products. No aldol condensation via kinetic enolates was observed. This means that these reactions occur via thermodynamic enolates, giving good regiochemistry control (Figure 1a). Furthermore, the enolization of the AL1K intermediates in an acidic medium must be minimal, which allows its use for the formation of chalcone analogs **1-6** through aldol condensations in a basic medium.

The aldol condensations of aldehydes with symmetrical ketones in conventional acidic or basic conditions yield symmetrical diaryl alkadienones (Shetty et al., 2015). However, unsymmetrical diarylidene alkanones are promising building blocks for pronounced pharmacological or agricultural chemistry (Zia et al., 2014). Thus, it was necessary to control the reaction of symmetrical cycloketones to obtain a monoarylidene product, to be further used as an intermediate for the design of unsymmetrical diarylidene cycloalkanones. To circumvent the enolization at both sides of cycloalkanones in acidic or basic conditions, the use of an amine was proposed to form Schiff-bases, which are less reactive than the parent aldehydes, as intermediates, allowing mono-aldol formation by reacting with enolates from

ketones. As a few studies show the synthesis of monoarylidene cicloalkanones using the ionic liquid amine-based catalyst dimethylammonium dimethylcarbamate (DIMCARB), this amine was employed as a catalyst in the reaction of AL1 aldehydes with cyclopentanone (C). This allowed the formation of AL1C intermediates, to be further reacted with the AL2 aldehyde to form chalcone analogs **7-12** (Figure 1b).

Since compound **6** was the only one of the 12 chalcone analogs studied that showed potential for the development of new nematicides, the activity against nematodes is apparently very specific and depends on the position and nature of the substituents in the molecule structure. For example, the exchanging the butyl group of compound **6** for the ethyl group of compound **2** (Figure 3) caused the biological activity to change completely. Other chalcones and analogs have already demonstrated nematicidal activities (Caboni et al., 2016; Nunes et al., 2013; Wuyts et al., 2006; Gonzalez et al., 1998, 1997). Nunes et al. (2013) suggested that the presence of three methoxy groups at positions 2, 4 and 5 of the chalcone structure was required for activity against *M. exigua*.

It is also worth mentioning that compound **6** has reduced the number of J2 hatched, is of great importance for nematode control since most of the nematode population at field conditions is in the egg-stage (Karssen & Moens, 2006). This may be an advantage over nematicides such as Carbofuran[®], which does not affect the hatching of *M. incognita*. This result agrees with Terra et al. (2018) who also found no reduction in hatching when *M. incognita* eggs were exposed to Carbofuran[®] solutions.

Since chalcone analog **6** presented results statistically equal to those observed for the commercial nematicide Carbofuran[®] in the *in vivo* assay and apparently did not have any phytotoxic effect, it seems that this compound has great potential to be used in the development of new nematicides. As a result, we sought to identify the possible enzymatic

target of this substance in the nematode through *in silico* studies. As compound **6** was pharmacophoricaly similar to A2, which is an inhibitor of cytochrome P450 1A2 (Appiah-Opong et al., 2008), it seemed reasonable to consider that compound **6** could also be able to inhibit this enzyme.

Cytochrome P450 enzymes (CYPs) belong to the superfamily containing HEME (porphyrin coordinated to iron ion) as a cofactor (Gonzalez & Gelboin, 1992). They have a large variety of substrates and usually are terminal oxidases in the electron transport chain. They have been identified in all kingdoms in nature, even in viruses (Lamb et al., 2009). CYPs can be divided into several groups (Hanukoglu, 1996), among which 1A2 (CYP1A2) that is strongly inhibited by compound A2 (Appiah-Opong et al., 2008). The 2C9 groups (CYP2C9), 2D6 (CYP2D6), and 3A4 (CYP3A4), are also worth mentioning because they are inhibited by compound A0, (1*E*,5*E*)-2,6-di(4-hydroxy-3-methoxyphenyl) penta-1,5-dien-3-one (C1), and (2*E*,5*E*)-2,5-bis(4-hydroxybenzylidene)cyclopentanone (B0), respectively (Appiah-Opong et al., 2008). As the chemical structures of A0, A2, B0 and C1 were similar to the chemical structure of compound **6**, it seemed reasonable to investigate whether compound **6** could be an inhibitor of cytochrome P450 belonging to such groups. Unfortunately, there was no available three-dimensional structure of this enzyme as produced by nematodes. However, as the amino acid sequences of CYPs in the RCSB Protein Data Bank were similar to the amino acid sequences produced by nematodes, it was assumed that the three-dimensional structures downloaded from this databank could be employed in a preliminary study to investigate the possibility of cytochrome P450 as the enzymatic target of chalcone analog **6** in *M. incognita.*

Several three-dimensional structures of cytochrome P450 downloaded from the RCSB Protein Data Bank were discarded because they were mutants or because the software packages used were unable to cope with some flaws in their structures. Therefore, this study

focused only on the CYP1A2 and CYP2C9 groups. In the case of the CYP1A2 group, blind docking indicated that compound A2 binds to 2hi4 through a surface cavity (CAV-1, Figure 6), suggesting that the inhibition caused by this substance is allosteric. Therefore, chalcone analog **6** is believed to be able to inhibit the 2hi4 enzyme in the same way, since its interaction with the same cavity, according to calculations carried out in this study, was as energetically favorable as that calculated for compound A2 (Figure 7). To a large extent this is due to compound **6** being able to establish two hydrogen bonds with amino acid residues of the enzyme: one with GLN265 and other with GLN304 (Figure 8). In addition, the butyl group is inserted into a cavity that favors nonpolar interactions and at the same time keeps it away from the polar solvent to which the enzyme surface is exposed.

Regarding the CYP2C9 group, blind docking suggests that the enzyme's active site (CAV-2, Figure 6) is the binding site for compound A0, which clearly interacts with an iron ion through the OH group. Similarly, the interaction of compound **6**, through its nitro group, with an iron ion seems to be very important to binding to CYP2C9 (Figure 9). It is worth mentioning that, although calculations carried out with Autodock Vina suggest that the interactions of compound **6** with the enzymes are as favorable as those of compound A0, calculations carried out with Autodock suggests that compound **6** should be a stronger inhibitor of these enzymes than compound A0 (Figure 7).

In conclusion, compound **6** was synthesized with good yield using inexpensive reagents and simple methods. The compound presented no phytotoxic activity and was more active against *M. incognita in vitro* than the commercial nematicide Carbofuran[®]. Furthermore, it controlled this nematode in tomato plants to the same extent as Carbofuran. According to calculations carried out in this study, compound **6** act against *M. incognita* by inhibiting its cytochrome P450, differing from Carbofuran[®] that acts on the

acetylcholinesterase produced by the nematode. Therefore, chalcone analog **6** has potential as a building block for the development of new products to control *M. incognita*.

SUPPORTING INFORMATION

Percent yields and spectroscopic data of the 12 compounds synthesized (Apendix A).

ACKNOWLEDGEMENTS

REFERENCES

Abad, P., Gouzy, J., Aury, J. M., Castagnone-Sereno, P., Danchin, E. G. J., Deleury, E., Perfus-Barbeoch, L., ..., Wincker, P. (2008). Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology*, *26*(8), 909-915.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389-3402.

Amaral, D. R., Oliveira, F. E. R., Oliveira, D. F., Campos, V. P. (2003). Purification of two substances from bulbs of onion (*Allium cepa* L.) with nematicidal activity against

Meloidogyne exigua Goeldi. *Nematology*, *5*, 859-864.

    https://doi.org/10.1163/156854103773040763

Appiah-Opong, R., Esch, I., Commandeur, J. N. M., Andarini, M., Vermeulen, N. P. E.

    (2008). Structure-activity relationships for the inhibition of recombinant human

    cytochromes P450 by curcumin analogues. *European Journal of Medicinal Chemistry*,

    43(8), 1621-1631. https://doi.org/10.1016/j.ejmech.2007.10.034

Attar, S., O´Brien, Z., Alhaddad, H., Golden, M. L., Calderón-Urrea, A. (2011). Ferrocenyl

    chalcones versus organic chalcones: a comparative study of their nematocidal activity.

    Bioorganic Medicinal Chemistry, *19*(6), 2055–2073.

    https://doi.org/10.1016/j.bmc.2011.01.048

Bebber, D. P., Holmes, T., Gurr, S. J. (2014). The global spread of crop pests and pathogens.

    Global Ecology Biogeography, *23*(12), 1398–1407. https://doi.org/10.1111/geb.12214

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I.

    N., Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research*, *28*(1), 235-242.

Boratyn, G. M., Schäffer, A. A. Agarwala, R., Altschul, S. F., Lipman, D. J., Madden T. L.

    (2012). Domain enhanced lookup time accelerated BLAST. *Biology Direct*, *7*, 1-12.

    https://doi.org/10.1186/1745-6150-7-12

Caboni, P., Aissani, N., Demurtas, M., Nikoletta, N., Onnis, V. (2016). Nematicidal activity

    of acetophenones and chalcones against *Meloidogyne incognita* and structure–activity

    considerations. *Pest Management Science*, *72*(1), 125–130.

    https://doi.org/10.1002/ps.3978

Chen, S. Y., & Dickson, D. W. (2000). A technique for determining live second-stage

    juveniles of *Heterodera glycines*. *The Journal of Nematology*, *32*, 117-121.

Edens, R. M., Anand, S. C., Bolla, R. I. (1995). Enzymes of the phenylpropanoid pathway in

soybean infected with *Meloidogyne incognita* or *Heterodera glycines*. *Journal of*

Nematology, *27*(3), 292–303.

Feng, Z., Chen, L., Maddula, H., Akcan, O., Oughtred, R., Berman, H. M., Westbrook, J.

(2004). Ligand Depot: a data warehouse for ligands bound to macromolecules.

Bioinformatics, *20*, 2153-2155. https://doi.org/10.1093/bioinformatics/bth214

Ferreira, D. F. (2014). Sisvar: a computer statistical analysis system. *Ciência e*

Agrotecnologia, *35*, 1039-1042.

Gonzalez, F. J., Gelboin, H. V. (1992). Human cytochromes P450: evolution and cDNA-

directed expression. *Environmental Health Perspectives*, *98*, 81-85.

https://doi.org/10.1289/ehp.929881

Gonzalez, J. A., & Estevez-Braun, A. (1997). Phytonematicidal activity of aromatic

compounds related to shikimate pathway. *Pesticide Biochemistry Physiology*, *58*(3),

193–197. https://doi.org/10.1006/pest.1997.2294

González, J. A., & Estévez-Braun, A. (1998). Effects of (E)-chalcone on potato-cyst

nematodes (*Globodera pallida* and *G. rostochiensis*). *Journal of Agriculture and Food*

Chemistry, *46*(3), 1163–1165. https://doi.org/10.1021/jf9706686

Hanukoglu, I. (1996). Electron transfer proteins of cytochrome P450 systems. *Advances in*

Molecular and Cell Biology, *14*, 29-56. https://doi.org/10.1016/S1569-2558(08)60339-2

Hussey, R. S., & Barker, K. R. (1973). A comparison of methods for collecting inocula of

Meloidogyne spp., including a new technique. *Plant Disease Report*, *57*, 1025–1028.

Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., Coleman, R. G. (2012). ZINC: A free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling*, *52*(7), 1757-1768. https://doi.org/10.1021/ci3001277

Jones, J. T., Haegeman, A., Danchin, E. G. J., Gaur, H. S., Helder, J., Jones, M. G. K., ..., Perry, R. N. (2013). Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology*, *14*(9), 946–961. https://doi.org/10.1111/mpp.12057

Karssen, G., & Moens, M. (2006) Root-knot nematodes. In: Perry, R. N., Moens, M. (eds) Plant Nematology.Wallingford, UK: CABI Publishing, pp 59–90.

Kumari, S., Singh, R., Walia, R. K. (2014). Synthesis and bioevaluation of 3-(substitutedphenyl)-1-(4-hydroxyphenyl)-2-propen-1-ones and their carbamate derivatives against root-knot nematode (*Meloidogyne javanica*). *Oriental Journal of Chemistry*, *30*(3), 1293-1302. http://dx.doi.org/10.13005/ojc/300346

Lamb, D. C., Lei, L., Warrilow, A. G., Lepesheva, G. I., Mullins, J. G., Waterman, M. R., Kelly, S. L. (2009) The first virally encoded cytochrome p450. *Journal of Virology*, *83*(16), 8266-8269. https://doi.org/10.1128/JVI.00289-09

Lesnik, S., Stular, T., Brus, B., Knez, D., Gobec, S., Janezic, D., Konc, J. (2015). LiSiCA: A software for ligand-based virtual screening and its application for the discovery of butyrylcholinesterase inhibitors. *Journal of Chemical Information and Modeling*, *55*(8), 1521-1528. https://doi.org/10.1021/acs.jcim.5b00136

Liu, R., Lyu, X., Batt, S. M., Hsu, M. H., Harbut, M. B., Vilcheze, C., …, Wang, F. (2017). Determinants of the inhibition of DprE1 and CYP2C9 by antitubercular thiophenes. *Angewandte Chemie International Ed. English*, *56*, 13011-13015. https://doi.org/10.1002/anie.201707324

Martínez, L., Andreani, R., Martínez, J. M. (2007). Convergent algorithms for protein structural alignment. *BMC Bioinformatics*, *8*:306. https://doi.org/10.1186/1471-2105-8-306

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., Olson, A. J. (2009). Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *Journal of Comput Chemistry*, *30*(16), 2785-2791. https://doi.org/10.1002/jcc.21256

Norman, C. S., De Canio, S. J., Fan, L. (2008). The Montreal Protocol at 20: Ongoing opportunities for integration with climate protection. *Global Environment Change*, *18*, 330-340. https://doi.org/10.1016/j.gloenvcha.2008.03.003

Ntalli, N. G., Caboni, P. (2012). Botanical nematicides: a review. *Journal of Agricultural and Food Chemistry*, *60*(40), 9929-9940. http://dx.doi.org/10.1021/jf303107j

Nunes, A. S., Campos, V. P., Mascarello, A., Stumpf, T. R., Chiaradia-Delatorre, L. D., Machado, A. R. T., Santos, H. M. Jr., Yunes, R. A., Nunes, R. J., Oliveira, D. F. (2013). Activity of chalcones derived from 2,4,5-trimethoxybenzaldehyde against *Meloidogyne exigua* and *in silico* interaction of one chalcone with a putative caffeic acid 3-O-methyltransferase from *Meloidogyne incognita*. *Experimental Parasitology*, *135*(4), 661–668. https://doi.org/10.1016/j.exppara.2013.10.003

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., Hutchison, G. R. (2011). Open Babel: na open chemical toolbox. *Journal of Cheminformatics*, 3, 1-33. https://doi.org/10.1186/1758-2946-3-33

Okonechnikov, K., Golosova, O., Fursov, M., UGENE team. (2012). Unipro UGENE: a

unified bioinformatics toolkit. *Bioinformatics*, *28*(8), 1166-1167.

https://doi.org/10.1093/bioinformatics/bts091

Peón, A., Naulaerts, S., Ballester, P. J. (2017). Predicting the reliability of drug-target

interaction predictions with maximum coverage of target space. *Scientific Reports*,

7:3820. https://doi.org/10.1038/s41598-017-04264-w

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C.,

Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research

and analysis. *Journal of Computacional Chemistry*, *25*(13), 1605-1612.

https://doi.org/10.1002/jcc.20084

Ritz, C. (2018) Analysis of dose-response curve data. Package 'drc'. Internet Resource:

http://cran.r-project.org/web/packages/drc/drc.pdf (verified June 14, 2018).

Ruiz-Suárez, N., Boada, L. D., Henríquez-Hernández, L. A., González-Moreo, F., Suárez-

Pérez, A., Camacho, M., Zumbado, M., Almeida-González, M., Travieso-Aja, M. D. M.,

Luzardo, O. P. (2015). Continued implication of the banned pesticides carbofuran and

aldicarb in the poisoning of domestic and wild animals of the Canary Islands (Spain).

Science of the Total Environment, *505*, 1093-1099.

https://doi.org/10.1016/j.scitotenv.2014.10.093

Sansen, S., Yano, J. K., Reynald, R. L., Schoch, G. A., Griffin, K. J., Stout, C. D., Johnson, E.

F. (2007). Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited

by the structure of human P450 1A2. *Journal of Biological Chemistry*, *282*(19), 14348-

14355. https://doi.org/10.1074/jbc.M611692200

Schäffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V., Altschul, S. F. (2001). Improving the accuracy of PSI_BLAST proteim database searches with composition-based statistics and other refinements. *Nucleic Acids Research*, *29*(14), 2994-3005.

Shetty, D., Kim, Y. J., Shim, H., Snyder, J. P. (2015). Eliminating the heart from the curcumin molecule: monocarbonyl curcumin mimics (MACs). *Molecules*, *20*(1), 249–92. https://doi.org/10.3390/molecules20010249

Sievers, F., Wilm, A., Dineen, D. G., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*:539. https://doi.org/10.1038/10.1038/msb.2011.75

Stewart, J. J. P. (2012) MOPAC2012, Stewart computational chemistry, colorado springs, CO, USA. Internet Resource: http://OpenMOPAC.net (verified June 14, 2018).

Swain, N. A., Swain, N.A., Batchelor, D., Beaudoin, S., Bechle, B.M., Bradley, P.A., …, West, C. W. (2017). Discovery of clinical candidate 4-[2-(5-Amino-1H-pyrazol-4-yl)-4-chlorophenoxy]-5-chloro-2-fluoro-N-1,3-thiazol-4-ylbenzenesulfonamide (PF-05089771): design and optimization of diaryl ether aryl sulfonamides as selective inhibitors of naV1.7. *Journal of Medicinal Chemistry*, *60*(16), 7029-7042. https://doi.org/10.1021/acs.jmedchem.7b00598

Terra, W. C., Campos, V. P., Martins, S. J., Costa, L. S. A. S., Silva, J. C. P., Barros, A. L., Lopez, L. E., Santos, T. C. N., Smant, G., Oliveira, D. F. (2018). Volatile organic molecules from *Fusarium oxysporum* strain 21 with nematicidal activity against *Meloidogyne incognita*. *Crop Protection*, *106*, 125–131. https://doi.org/10.1016/j.cropro.2017.12.022

Tosco, P., Balle, T., Shiri, F. (2011). Open3DALIGN: na open-source software aimed at unsupervised ligand alignment. *Journal of Computer-Aided Molecular Design*, *25* (8), 777-783.

Trott, O., Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Comput Chemistry*, *31*(2), 455-461. https://doi.org/10.1002/jcc.21334

Wallace, A. C., Laskowski, R. A., Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering*, *8* (2), 127-134.

Wester, M. R., Yano, J. K., Schoch, G. A., Yang, C., Griffin, K. J., Stout, C. D. Johnson, E. F. (2004). The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0 a resolution. *Journal of Biological Chemistry*, *279*, 35630-35637. https://doi.org/10.1074/jbc.M405427200

Wuyts, N., Swennen, R., Waele, D. D. (2006). Effects of plant phenylpropanoid pathway products and selected terpenoids and alkaloids on the behaviour of the plant-parasitic nematodes *Radopholus similis*, *Pratylenchus penetrans* and *Meloidogyne incognita*. *Nematology*, *8*(1), 89-101. https://doi.org/10.1163/156854106776179953

Zhang, Y. N., Feng, Y. A., Li, Z., Shao, X. S. (2017). Synthesis and insecticidal evaluation of phytoalexin phenalenones derivatives. *Chinese Chemical Letters*, *28*, 1228–1231. https://doi.org/10.1016/j.cclet.2017.04.003

Zia, U. D., Fill, T. P., Assis, F. F. de, Lazarin, D. B., Kaplum, V., Garcia, F. P., Nakamura, C. V., Oliveira, K. T., Rodrigues-Filho, E. (2014). Unsymmetrical 1,5-diaryl-3-oxo-1,4-pentadienyls and their evaluation as antiparasitic agents. *Bioorganic and Medicinal Chemistry*, *22*(3), 1121–1127. https://doi.org/10.1016/j.bmc.2013.1

Percent yields and spectroscopic data of the 12 compounds synthesized.

(1*E*,4*E*)-5-(3-bromophenyl)-2-methyl-1-(4-nitrophenyl)penta-1,4-dien-3-one (1). Percent yield: 51.4 %; ^{1}HH NMR (400 MHz, CDCl$_3$) δ 8.31 (d, *J* = 8.8 Hz, 2H), 7.80 (t, *J* = 1.8 Hz, 1H), 7.67 (d, *J* = 15.6 Hz, 1H), 7.64 − 7.60 (m, 2H), 7.58 (s, 1H), 7.57 − 7.51 (m, 2H), 7.39 (d, *J* = 15.6 Hz, 1H), 7.35 − 7.29 (m, 1H), 2.21 (d, *J* = 1.4 Hz, 3H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 191.64 (1C), 147.33 (1C), 142.73 (1C), 142.40 (1C), 141.40 (1C), 136.92 (1C), 135.75 (1C), 133.29 (2C), 130.69 (1C), 130.40 (1C), 130.29 (1C),127.28 (1C), 123.75 (2C), 123.12 (1C), 122.55 (1C), 14.03 (1C).

(1*E*,4*E*)-5-(2-bromophenyl)-2-methyl-1-(4-nitrophenyl)penta-1,4-dien-3-one (2). Percent yield: 57 %; ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.28 (d, *J* = 7.5 Hz, 2H), 7.80 − 7.54 (m, 8H), 7.52 − 7.37 (m, 1H), 2.19 (s, 3H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 191.62, 147.35, 142.41, 141.35, 138.20, 135.98, 130.30, 128.46, 125.94, 123.66, 14.00 (s).

(1*E*,4*E*)-5-(2-methoxy-4-nitrophenyl)-2-methyl-1-[4-(trifluoromethyl)phenyl]penta-1,4-dien-3-one (3). Percent yield: 68 %; ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.95 (d, *J* = 15.9 Hz, 1H), 7.83 (d, *J* = 8.4 Hz, 1H), 7.75 (s, 1H), 7.72 (d, *J* = 8.5 Hz, 1H), 7.67 (d, *J* = 8.0 Hz, 2H), 7.61 − 7.49 (m, 4H), 4.00 (s, 3H), 2.16 (s, 3H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 192.14 (1C), 158.55 (1C), 149.35 (1C), 140.22 (1C), 139.40 (1C), 137.28 (1C), 136.72 (1C), 130.39 (4C), 129.82 (1C), 128.96 (1C), 125.82 (1C), 125.81 (1C), 125.01 (1C), 115.83 (1C), 106.22 (1C), 56.24 (1C), 13.78 (1C).

(1*E*,4*E*)-2-ethyl-5-(4-nitrophenyl)-1-phenylpenta-1,4-dien-3-one (4). Percent yield: 70%;
^{1}H NMR (400 MHz, CDCl$_3$) δ 8.25 (d, *J* = 8.8 Hz, 2H), 7.73 (d, *J* = 15.7 Hz, 1H), 7.62 (dd, *J* = 6.6, 2.9 Hz, 2H), 7.55 (d, *J* = 8.7 Hz, 2H), 7.47 − 7.34 (m, 5H), 2.65 (q, *J* = 7.5 Hz, 2H), 1.16 (t, *J* = 7.5 Hz, 3H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 192.32 (1C), 147.65 (1C), 147.23 (1C), 144.52 (1C), 142.57 (1C), 134.60 (1C), 131.03 (1C),130.59 (1C), 129.85 (2C), 128.99 (2C), 128.42 (2C), 123.77 (2C), 122.22 (1C), 20.76 (1C), 13.52 (1C).

(1*E*,4)-1,5-di(4-nitrophenyl)-2-ethylpenta-1,4-dien-3-one (5). Percent yield: 75 %; ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.33 − 8.24 (m, 4H), 7.75 (t, *J* = 12.1 Hz, 3H), 7.57 (d, *J* = 8.6 Hz, 2H), 7.53 − 7.41 (m, 2H), 2.65 (q, *J* = 7.5 Hz, 2H), 1.16 (t, *J* = 7.5 Hz, 3H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 191.33 (1C), 148.58 (1C), 147.45 (1C), 142.09 (1C), 141.27 (1C), 140.94 (1C), 135.73 (1C), 129.84 (2C), 129.31 (1C), 128.89 (2C), 125.71 (1C), 124.24 (2C), 123.90 (2C), 20.64 (1C), 13.52 (1C).

(1*E*,4*E*)-1,5-di(4-nitrophenyl)-2-butylpenta-1,4-dien-3-one (6). Percent yield: 68 %; ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.33 − 8.23 (m, 4H), 7.81 − 7.70 (m, 3H), 7.56 (d, *J* = 8.7 Hz, 2H), 7.51 − 7.39 (m, 2H), 2.68 − 2.54 (m, 2H), 1.55 − 1.42 (m, 2H), 1.43 − 1.31 (m, 2H), 0.91 (t, *J* = 7.2 Hz, 3H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 191.60 (1C), 148.59 (1C), 147.44 (1C), 146.51 (1C), 142.19 (1C), 141.30 (1C), 140.93 (1C), 135.75 (1C), 129.83 (2C), 128.89 (2C), 125.78 (1C), 124.24 (2C), 123.88 (2C), 31.08 (1C), 27.10 (1C), 22.89 (1C), 13.80 (1C).

(2E,5E)-2-(4-methoxybenzylidene)-5-(3,4,5-trimethoxybenzylidene)cyclopentanone (7). Percent yield: 80 %, ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.56 (d, *J* = 8.7 Hz, 14H), 7.49 (s, 4H), 6.96 (d, *J* = 8.9 Hz, 9H), 6.84 (s, 8H), 3.90 (d, *J* = 2.5 Hz, 37H), 3.85 (s, 14H), 3.08 (d, *J* =

10.8 Hz, 17H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 196.18, 160.80, 153.36, 139.55, 136.81, 135.01, 133.94, 133.67, 132.72, 131.63, 128.75, 114.47, 108.20, 61.11, 56.29, 55.51, 26.56.

(2E,5E)-2-(4-methoxybenzylidene)-5-(4-nitrobenzylidene)cyclopentanone **(8)**. Percent yield: 71 %, ^{1}H NMR (400 MHz, CDCl$_3$) δ 8.30 (d, J = 8.6 Hz, 2H), 7.74 (d, J = 8.6 Hz, 2H), 7.67 − 7.53 (m, 5H), 6.99 (t, J = 7.9 Hz, 3H), 3.88 (d, J = 4.5 Hz, 3H), 3.13 (d, J = 21.6 Hz, 4H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 195.63 (s), 141.65 (s), 135.27 (s), 133.29 (s), 132.87 (s), 132.51 (s), 130.88 (s), 130.08 (s), 128.30 (s), 123.93 (s), 114.38 (d, J = 16.7 Hz), 55.42 (s), 26.82 − 26.10 (m).

(2E,5E)-2-(4-chlorobenzylidene)-5-(4-methoxybenzylidene)cyclopentanone **(9)**. Percent yield: 70 %, ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.57 (d, J = 8.8 Hz, 3H), 7.52 (d, J = 8.4 Hz, 3H), 7.40 (d, J = 8.6 Hz, 2H), 6.97 (d, J = 8.9 Hz, 2H), 3.86 (s, 3H), 3.08 (s, 4H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 196.07, 160.76, 138.13, 135.14, 134.57, 134.18, 132.67, 131.79, 129.02, 128.56, 114.37, 55.40, 26.46.

(2E,5E)-2-(2-bromobenzylidene)-5-(4-methoxybenzylidene)cyclopentanone (10). Percent yield: 75 %, ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.82 (t, J = 2.5 Hz, 2H), 7.65 (dd, J = 8.0, 1.1 Hz, 2H), 7.61 − 7.51 (m, 8H), 7.40 − 7.32 (m, 2H), 7.20 (td, J = 7.8, 1.6 Hz, 2H), 6.96 (d, J = 8.8 Hz, 4H), 3.85 (s, 5H), 3.08 − 2.93 (m, 8H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 195.73, 160.75, 139.93, 135.73, 135.29, 134.68, 134.27, 133.29, 132.58, 131.51, 130.11, 128.61, 127.23, 126.38, 114.35, 55.39, 26.59, 26.31.

(2E,5E)-2-(2,3-dimethoxybenzylidene)-5-(4-methoxybenzylidene)cyclopentanone **(11)**. Percent yield: 82 %, ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.91 (s, 1H), 7.56 (d, J = 8.7 Hz, 3H),

7.17 (dd, J = 7.9, 1.3 Hz, 1H), 7.10 (t, J = 8.0 Hz, 1H), 7.01 − 6.90 (m, 3H), 3.88 (s, 3H), 3.86 (s, 3H), 3.85 (s, 3H), 3.04 (s, 4H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 196.24, 160.60, 153.02, 149.26, 138.77, 135.17, 133.69, 132.55, 130.31, 128.70, 127.72, 123.75, 121.59, 114.31, 113.39, 61.52, 55.89, 55.38, 26.59.

(2E,5E)-2-(benzo[d][1,3]dioxol-5-ylmethylene)-5-(4-chlorobenzylidene)cyclopentanone (12). Percent yield: 50 %, ^{1}H NMR (400 MHz, CDCl$_3$) δ 7.59 − 7.50 (m, 24H), 7.43 (dt, J = 10.1, 3.3 Hz, 14H), 7.28 (s, 7H), 7.19 − 7.12 (m, 12H), 6.91 (dd, J = 8.0, 2.5 Hz, 5H), 6.08 − 6.02 (m, 11H), 3.11 (d, J = 5.4 Hz, 22H). ^{13}C NMR (101 MHz, CDCl$_3$) δ 137.96, 135.15, 134.33, 132.70 , 132.31, 131.67, 129.07 , 126.93, 109.78, 108.81, 101.59, 26.45.

CAPÍTULO 2

INTRODUÇÃO GERAL

O uso de bactérias como agentes de controle de pragas e doenças de plantas cresceu rapidamente nas últimas décadas. Entre as suas diferentes espécies, membros do gênero *Bacillus* são candidatos potenciais (SHAFI et al., 2017). *Bacillus* spp. constituem um grupo de bactérias encontradas de forma onipresente no meio ambiente. São capazes de formar endósporos e exibem uma grande versatilidade na proteção de plantas contra infecções por patógenos (FIRA et al., 2018). E possibilitam, dessa forma, longa manutenção e sobrevivência em nichos ecológicos específicos (MUKHERJEE e DAS, 2005). Em decorrência, *Bacillus* se tornou o gênero bacteriano mais extensivamente estudado sob vários aspectos e com os maiores conjuntos de genomas sequenciados (RAVEL e FRASER, 2005).

Os avanços recentes nas tecnologias de sequenciamento de DNA de última geração permitiram estudos aprofundados sobre os genomas microbianos (KIM et al., 2017). Os sistemas de sequenciamento possibilitam, por exemplo, investigar quais determinantes genéticos bacterianos são responsáveis pelo desenvolvimento de atividades de biocontrole (VERNIKOS et al., 2015; KIM et al., 2017). A aquisição de sequências genômicas de múltiplas espécies bacterianas gerou uma enorme quantidade de informações que, em combinação com algumas ferramentas de bioinformática, aumentaram nosso conhecimento do potencial genético que pode contribuir para a adaptação ao meio ambiente, interação com plantas, entre outras características (CAI et al., 2014; EARL et al., 2008). O sequenciamento do genoma completo de numerosas cepas de *Bacillus* possibilitou a descoberta da base molecular de seu desempenho versátil sob diferentes ambientes (KIM et al., 2017).

Surgiram, por conseguinte, várias maneiras possíveis de analisar um conjunto de genomas, com a possibilidade de estabelecer sistemática com base em informações recuperadas de genomas completos (THOMPSON et al., 2013). Por exemplo, alguns métodos *in silico*, como hibridização digital DNA-DNA (dDDH) e identidade média de nucleotídeos (ANI), têm sido amplamente adotados por microbiologistas para delineação de espécies (AUCH et al., 2010). A aplicação generalizada de métodos genômicos na taxonomia e sistemática procariótica revelou vários erros de identificação, com propostas de novos taxa e reclassificações (COLSTON et al., 2014). Além da taxonomia e da sistemática, a análise em dados genômicos proporcionou melhor compreensão das características

genotípicas/fenotípicas inerentes aos organismos estudados individualmente (ZHANG et al., 2016). Vários trabalhos revelaram que múltiplas cepas de uma espécie apresentam variações genômicas em termos de conteúdo gênico (KIM et al., 2017).

Na última década, a exploração de genomas foi estabelecida como uma tecnologia complementar às abordagens "clássicas" para identificar novos compostos que são rotineiramente usados por equipes de pesquisa acadêmica e industrial (LIU et al., 2015). Alguns softwares de bioinformática examinam dados de genomas para identificar e analisar os clusters de genes biossintéticos para uma ampla gama de produtos naturais (WEBER et al., 2015). Esses *softwares* podem identificar aglomerados de genes responsáveis pela biossíntese de diferentes classes de metabólitos secundários, incluindo tiopeptídeos, lantipeptídeos, lassopeptídeos, peptídeos obtidos a partir da sintetase não ribossômal (NRPS) e policetídeo sintases (PKS), entre outros (WEBER et al., 2015).

Os arranjos CRISPr/Cas (clustered regularly interspaced short palindromic repeat) fazem parte do sistema imunológico das bactérias, protegendo-as contra invasores como bacteriófagos e plasmídeos (RATH et al., 2015). São amplamente distribuídos em bactérias e archaea (CAIN e BOINETT 2013). Estudos destacam o uso de lócus CRISPr para fornecer informações filogenéticas entre bactérias intimamente relacionadas (SUPPLY et al., 2013). Como os lócus CRISPr mostram conservação específica de linhagem no nível de nucleotídeos, eles provaram ser marcadores valiosos para tais estudos e, em sequenciamento do genoma completo, pode fornecer "insights" sobre as relações filogenéticas entre diferentes bactérias (CAIN e BOINETT, 2013). Nos últimos anos esses sistemas têm sido utilizados também como ferramentas importantes para a edição de genomas (ALTENBUCHNER, 2016).

Neste trabalho, aplicaram-se várias abordagens de genômica comparativa para estudar a espécie *Bacillus velezensis*; uma bactéria aeróbia, Gram-positivo, formadora de endósporos, que promove o crescimento de plantas. Indivíduos desta espécie foram frequentemente isolados de vários nichos ecológicos, incluindo solo, rizosfera, associados à planta, fezes de animais, alimentos, entre outros (EARL et al., 2008). Inúmeras cepas de *B. velezensis* são amplamente usadas em formulações comerciais para promover o controle de patógenos (IDRISS et al., 2002). Análises genômicas revelaram que *B. velezensis* possui grupos específicos de genes relacionados à biossíntese de metabólitos secundários, que

desempenham papéis significativos tanto na supressão de patógenos quanto na promoção do crescimento de plantas (PALAZZINI et al., 2016).

Este é o primeiro estudo a incluir todos os genomas completos de *B. velezensis*, e das espécies intimamente relacionadas como *Bacillus amyloliquefaciens* e *Bacillus siamensis.* Objetivou-se então apresentar a sequência completa do genoma da cepa UFLA258, revisar o posicionamento taxonômico dos genomas completos dessas espécies, que estão depositados no banco de dados e fornecer seus limites de espécie. Além disso, decidiu-se apresentar e investigar agrupamentos gênicos de metabólitos secundários e a presença de matrizes CRISPr/Cas e fagos nos genomas estudados.

REFERÊNCIAS

Altenbuchner, J. Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 system.
Applied and Environmental Microbiology, v. 82, p. 5421-5427, 2016.

AUCH, A. F.; VON, J. M.; KLENK, H. P.; GOKER, M. Digital DNA-DNA hybridization for
microbial species delineation by means of genome-to-genome sequence comparison.
Standards in genomic sciences, v. 2, p. 117-134, 2010.

CAI, X. C.; LIU, C. H.; WANG, B. T.; XUE, Y. R. Genomic and metabolic traits endow
Bacillus velezensis CC09 with a potential biocontrol agent in control of wheat powdery
mildew disease. **Microbiological Research**, v. 196, p. 89-94, 2017.

CAIN, A. K.; BOINETT, C. J. A CRISPR view of genome sequences. **Nature Reviews
Microbiology**, v. 11, p. 226, 2013.

COLSTON, S. M.; FULLMER, M.; BEKA, L.; LAMY, B.; GOGARTEN, J. P.; GRAF, J.
Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using
Aeromonas as a test case. mBio 5(6):e02136-14. 2014.

EARL, A. M.; LOSICK, R.; KOLTER, R. Ecology and genomics of *Bacillus subtilis*. **Trends
in Microbiology**, v. 16, p. 269-275, 2008.

FIRA, D.; DIMKIĆ, I.; BERIĆ, T.; LOZO, J.; STANKOVIĆ, S. Biological controlo f plant
pathogens by *Bacillus* species. **Journal of Biotechnology**, v. 285, p. 44-55, 2018.

IDRISS, E. E.; MAKAREWICZ, O.; FAROUK, A.; ROSNER, K.; GREINER, R.;
BOCHOW, H.; et al. Extracellular phytase activity of *Bacillus amyloliquefaciens* FZB45
contributes to its plant-growth-promoting effect. **Microbiology**, v. 148, p. 2097-2109, 2002.

KIM, Y.; KOH, I.; LIM, M. Y.; CHUNG, G. H.; RHO, M. Pan-genome analysis of *Bacillus*
for microbiome profiling. **Scientific Reports**, v. 7, 10984, 2017.

MUKHERJEE, A. K.; DAS, K. Correlation between diverse cyclic lipopeptides production and regulation of growth and substrate utilization by *Bacillus subtilis* strains in a particular habitat. **FEMS Microbiology Ecology**, v. 54, p. 479-489, 2005.

PALAZZINI JM, DUNLAP CA, BOWMAN MJ, CHULZE SN. 2016. *Bacillus velezensis* RC 218 as a biocontrol agent to reduce *Fusarium* head blight and deoxynivalenol accumulation: Genome sequencing and secondary metabolite cluster profiles. *Microbiological Res.* **192**:30-36.

RATH, D.; AMLINGER, L.; RATH, A.; LUNDGRENB, M. The CRISPR-Cas immune system: Biology, mechanisms and applications. **Biochimie**, v. 117, p. 119-128, 2015.

RAVEL, J.; FRASER, C. M. Genomics at the genus scale. **Trends in Microbiology**, v. 13 , p. 95-97, 2005.
SHAFI, J.; TIAN, H.; MINGSHAN, J. *Bacillus* species as versatile weapons for plant pathogens: a review. **Biotechnology & Biotechnological Equipment**, v. 31, p. 446-459, 2017.

SUPPLY, P.; MARCEAU, M.; MANGENOT, S.; et al. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. **Nature Genetics**, v. 45, p. 172-179, 2013.

THOMPSON, C. C.; CHIMETTO, L.; EDWARDS, R. A.; SWINGS, J.; STACKEBRANDT, E.; THOMPSON, F. L. Microbial genomic taxonomy. **BMC Genomics**, v. 14, 913, 2013.

VERNIKOS, G.; MEDINI, D.; RILEY, D. R.; TETTELIN, H. Tem years of pan-genome analyses. **Current Opinion in Microbiology**, v. 23, p. 148-154, 2015.

WEBER T, et al. 2015. AntiSMASH 3.0 - a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res*. **43**: 237-243.

ZHANG, N.; YANG, D.; KENDAL, J. R. A.; BORRISS, R.; DRUZHININA, I. S.; KUBICEK, C. P.; SHEN, Q.; ZHANG, R. Comparative genomic analysis of *Bacillus*

amyloliquefaciens and *Bacillus subtilis* reveals evolutional traits for adaptation to plant-associated habitats. **Frontiers in Microbiology**, v. 7, 2039, 2016.

ARTIGO 2 - Complete genome sequence of the biocontrol agent *Bacillus velezensis* UFLA258 and its comparison with related species: diversity within the commons

Authors

Fabíola de Jesus Silva[1], Larissa Carvalho Ferreira[2], Vicente Paulo Campos[1], Valter Cruz-Magalhães[1], Aline Ferreira Barros[1], Jackeline Pereira Andrade[3], Daniel P. Roberts[4], Jorge Teodoro de Souza[1*]

Institutional Affiliation

[1]Plant Pathology Department, Federal University of Lavras, 37200-000, Lavras, MG, Brazil.

[2]Plant Pathology Department, Aberystwyth University, Aberystwyth, United Kingdom.

[3]Biological Sciences Department, Feira de Santana State University, Feira de Santana, BA, Brazil.

[4]USDA-Agricultural Research Service, Sustainable Agricultural Systems Laboratory, Beltsville, MD, USA 20705.

***Author for Correspondence**: Jorge Teodoro de Souza, Plant Pathology Department, Federal University of Lavras, Brazil.

Abstract

In this study, the full genome sequence of *Bacillus velezensis* strain UFLA258, a biological control agent of plant pathogens was obtained, assembled and annotated. With a comparative genomics approach, *in silico* analyses of all complete genomes of *B. velezensis* and closely related species available in the database were performed. The genome of *B. velezensis* UFLA258 consisted of a single chromosome of 3.95 Mbp in length, with a mean GC content of 46.69%. It contained 3,949 genes encoding proteins, and 27 RNA genes. Analyses based on ANI and dDDH and a phylogeny with complete sequences of the *rpo*B gene confirmed that 19 strains deposited in the database as *B. amyloliquefaciens* were in fact *B. velezensis*. In total, 115 genomes were analyzed and taxonomically classified as follows: 105 were *B. velezensis*, 9 were *B. amyloliquefaciens* and 1 was *B. siamensis*. Although these species are phylogenetically close, the combined analyses of several genomic characteristics, such as the presence of biosynthetic genes encoding secondary metabolites, CRISPr/Cas arrays, ANI and dDDH, and other information about the strains, including isolation source, allowed their unequivocal classification. This genomic analysis expands our knowledge about the closely related species, *B. velezensis*, *B. amyloliquefaciens* and *B. siamensis*, with emphasis on their taxonomical status.

Keywords: Biological control, *Bacillus* spp., Comparative genomics, CRISPr/Cas, Secondary metabolites.

1 Introduction

Numerous microorganisms have been successfully developed as biopesticides at the commercial level (Shafi et al. 2017). Members of the *Bacillus* genus have been used for this purpose due to their ability to produce a large number of biologically active molecules with growth-promoting activity and inhibitory effects against plant pathogens (Fira et al. 2018; Jiang et al. 2018; Olishevska et al. 2019). The potential of *Bacillus* isolates for commercial development is enhanced by their fast growth rate and resistance to adverse environmental conditions (Shafi et al. 2017).

Bacillus velezensis was originally described in 2005 (Ruiz-Garcia et al. 2005), and since then its biopesticide potential has been unequivocally shown (Jiang et al. 2018; Gao et al. 2017; Cai et al. 2017). This species synthesizes several types of lipopeptides as products of secondary metabolism. Some of these compounds are active against plant pathogens and/or induce systemic resistance in plants, conferring an adaptive advantage in specific ecological niches (Mukherjee & Das 2005; Ruiz-Garcia et al. 2005; Lim et al. 2017; Yamamoto et al. 2015).

Initially, *B. velezensis* was shown to be closely related to *B. subtilis* and *B. amyloliquefaciens* (Ruiz-Garcia et al. 2005). Subsequently, *B. velezensis* was shown to be a heterotypic synonym of *B. amyloliquefaciens* subsp. *plantarum*, *B. methylotrophicus*, and *B. oryzicola* (Wang et al. 2008; Dunlap et al. 2016). Although all these species were reclassified as *B. velezensis*, this information still needs to be integrated into a well-organized resource.

Strain UFLA258 of *B. velezensis* was isolated from soil around the roots of healthy cotton plants and shown to have potential to control plant pathogens (Medeiros et al. 2011; 2015; Martins et al. 2013; 2018; 2019). In this study, we sequenced the genome of strain UFLA258 and compared it with all genomes of closely related species. Additionally, a taxonomic re-evaluation of the clade *B. velezensis*-*B. amyloliquefaciens* was performed.

Although closely related, they are distinct species with many commonalities and minor differences.

2 Material and Methods

2.1 Isolation and DNA extraction

Bacillus velezensis UFLA258 was isolated from a soil sample collected in a cotton (*Gossypium hirsutum* L.) field in Mato Grosso state, Brazil. DNA extraction was done according to the method described by Lee et al. (2003).

2.2 Genome sequencing and assembly

The sequence data was generated with an Illumina NextSeq-500 using the run kit Illumina NextSeq® 500/550 High Output Kit v2. Sequencing resulted in 22,196,922 reads, with length varying from 32 to 151 bases, which comprised a total of 3,351,735,222 bases and represented 849-fold genome coverage. The quality was checked with the program FastQC v0.11.5 (Andrews 2010). The genome was assembled employing the assembly service "auto" available in PATRIC (Pathosystems Resource Integration Center; Wattam et al. 2014). This strategy implements BayesHammer (Nikolenko et al. 2013) in short reads, followed by three assembly strategies that include Velvet (Zerbino & Birney 2008), IDBA 1.1.1 (Peng et al. 2010) and SPAdes 3.10.0 (Bankevich et al. 2012). Based on each assembly score provided by the QUAST (Quality Assessment Tool for Genome Assemblies) algorithm (Gurevich et al. 2013), the SPAdes assembly was chosen to move on to the subsequent steps. The 1,304

contigs generated were united into 12 scaffolds using the CONTIGuator web server (Galardini et al. 2011) with *B. velezensis* strain UCMB5113 (accession number NC_022081.1) as the reference genome. The gene *dnaA* was determined as the beginning of the chromosome using an in-house script. Finally, gaps were closed by a de novo strategy with FGAP (Piro et al. 2014), and by reference using NCBI's BLASTn (Altschul et al. 1999) and read mapping in CLC Genomics Workbench 11 (Qiagen Inc.).

2.3 Genome annotation and manual curation

The UFLA258 genome was annotated using the RASTtk (Rapid Annotation Using Subsystem Technology; Brettin et al. 2015) annotation service in PATRIC. Manual curation was conducted through Artemis 16.0.0 software (Rutherford et al. 2000) and insertion/deletion (indels) errors were checked in CLC Genomics Workbench 11, and only adjusted when there was depth coverage. Genes with potential frameshifts were compared to other complete genes with BLASTn against the NR database at NCBI. Translated protein sequences were determined with BLASTp against the UniProt database (Wasmuth et al. 2017). Ribosomal RNA genes were verified using the web-tool RNAmmer 1.2 (Lagesen et al. 2007) and tRNA genes were verified with tRNAscan-SE 2.0 (Lowe et al. 2016). Clusters of orthologous groups (COGs) were defined with the eggNOG v. 4.5.1 database (Huerta-Cepas et al. 2016).

2.4 Comparative genomics

All complete genome sequences of *B. velezensis*, *B. amyloliquefaciens* and *Bacillus siamensis* strains available in the GenBank database (https://www.ncbi.nlm.nih.gov) as of June 24, 2019 were used in this study. Digital DNA-DNA Hybridization (dDDH) and

Average Nucleotide Identity (ANI) comparisons were calculated using JspeciesWS (Richter et al. 2015) and Kostas Lab (Rodriguez-R & Konstantinidis 2014), respectively. The genome sequence of the type strain *B. velezensis* FZB42 was used as a reference (accession number NC_009725.1). CRISPr (Clustered Regularly Interspaced Short Palindromic Repeat) matrices and phages were identified using the web-tool CRISPRfinderCAS (Couvin et al. 2018) and PHASTER (Arndt et al. 2016), respectively. Clusters of biosynthetic genes from secondary metabolites were predicted using antiSMASH 4.0.2 (Weber et al. 2015). Principal component analysis (PCA) was performed with the comparative genomics data with packages cluster and factoextra implemented in the R software (R Core Team, 2019).

2.5 Phylogenetic analyses

Complete *rpoB* gene (β subunit of the RNA polymerase) sequences were retrieved from the genomes under study and used for the phylogenetic analysis. Alignments were performed with MAFFT v7.0 (Katoh et al. 2017). A maximum likelihood tree with the T92+G+I model was constructed using MEGA v10.1 (Kumar et al. 2018) with 1,000 bootstrap replicates.

3 Results and Discussion

3.1 Properties of the genome of *B. velezensis* UFLA258

The genome of *B. velezensis* UFLA258 is comprised of an unique chromosome of 3.95 Mb (Figure 1), which falls between 3.71 and 4.39 Mb, the size range reported for this species (Table S1). The chromosome is predicted to include 3,949 protein-encoding genes,

from which 3,747 genes were functionally assigned while the remaining genes were annotated as hypothetical proteins. Pseudogenes accounted for 1.7% of the total number of genes. There were 84 tRNA genes and 9 copies of the ribosomal RNA operon distributed throughout the genome, which represented 27 rRNA genes (Table S1). From the predicted genes, 3,439 (87.08%) were classified into 20 functional COG categories, while the remaining 510 (12.92%) were not classified into COGs (Figure 1). The most numerous COGs contained genes with unknown function (806 genes), genes involved in the transport and metabolism of amino acids (281 genes) and genes involved in transcription (270 genes). COG categories with the lowest number of genes were genes related to chromatin dynamics and structure and a gene for RNA processing (Figure 1).

3.2 Phylogeny and species boundaries in the clade *B. amyloliquefaciens - B. velezensis*

The complete genome of strain UFLA258, and 86 *B. velezensis,* 28 *B. amyloliquefaciens* and 1 *B. siamensis* genomes available in the NCBI database were used in the analyses. According to the dDDH and ANI values (Supplementary Tables S2 and S3) and the phylogenetic analyses with the *rpoB* gene (Supplementary Figure S1), strain UFLA258 and 104 additional strains belonged in the species *B. velezensis*, whereas 9 other strains were *B. amyloliquefaciens* and 1 was *B. siamensis*. ANI and dDDH values were above the cutoff for the delimitation of each species (ANI > 95%, Auch et al. 2010; dDDH > 70%, Richter et al. 2015).
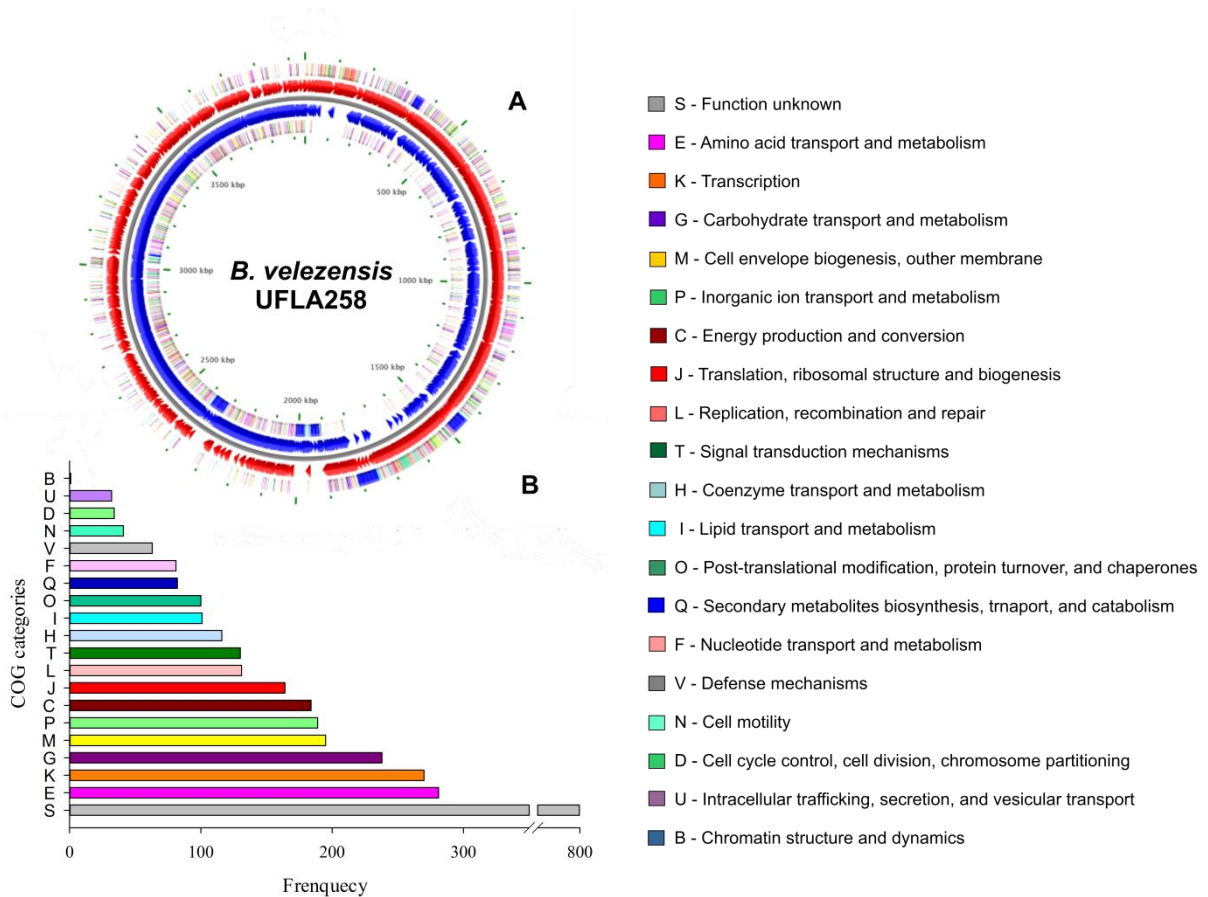
Figure 1. A. Graphical circular map of *Bacillus velezensis* strain UFLA258 chromosome. From outer circle to the center: CDS on forward strand (colored according to COG categories), all CDS and RNA genes on forward strand, all CDS and RNA genes on reverse strand, CDS on reverse strand (colored according to COG categories). The map was generated using Bacterial Annotation System, BASys (Van Domselaar et al. 2005). B. COG functional classification of the 3,439 proteins.

The phylogenetic analysis corroborated the ANI and dDDH results, showing that most strains identified as *B. amyloliquefaciens* were indeed *B. velezensis* (Figure S1). The resolution power of the *rpoB* gene in phylogenetic analyses has been shown by several authors (Sharma and Patil 2011; Fan et al. 2017). Recently, the designation "operational group *B. amyloliquefaciens*" has been proposed to name the closely related species *B.*

amyloliquefaciens, *B. siamensis* and *B. velezensis* (Fan et al. 2017). However, we propose the identification by species names as it is easy enough to perform by the methods described above, including *rpoB* phylogeny and genomic indices (ANI and dDDH). The use of individual species names will facilitate scientific communication. Additionally, the species *B. amyloliquefaciens* is less frequently encountered than *B. velezensis*.

3.3 Comparison of UFLA258 with genomes of related species

All deposited genomes of *B. velezensis* were used in this part of the analysis, including genomes re-identified as above, totaling genomes from 105 strains. Among the complete genomes of *B. velezensis*, the number of genes ranged from 3,683 to 4,744 and the number for *B. velezensis* UFLA258 fits within this range (Supplementary Table S2). Similarly, the GC content (46.69%) and mean size of the genome (4.03 Mb) of *B. velezensis* UFLA258 were comparable to deposited genomes of other *B. velezensis* strains (Supplementary Table S2).

The genomes of *B. velezensis* encoded twelve groups of genes involved in the production of antimicrobial compounds (Supplementary Table S4). Among these, five groups of non-ribosomal peptide synthase (NRPS) genes, for bacilysin, bacillibactin, fengyncin, bacillaene and surfactin biosynthesis were absolutely conserved in all 105 *B. velezensis* genomes used in this study, whereas the polyketide synthase genes (PKS) for difficidin and macrolactin biosynthesis were not present in one strain, AGVL-005 (Supplementary Table S4). Genes for the compounds plantathizolicin, mersacidin, subtilin, bacillomycin and locilomycin showed a more variable pattern and occurred in 18, 6, 5, 1 and 1 genomes, respectively. Genomes of *B. amyloliquefaciens* generally did not harbor PKS genes, except for two strains that encoded macrolactin. Another difference was the absence of the compounds described above as having a variable pattern of occurrence in genomes of *B.*

velezensis (Supplementary Table S4). However, these differences may not be consistent among *B. amyloliquefaciens* due to the low number of *B. amyloliquefaciens* genomes available. Most of the above compounds have surfactant and antibiotic activities and were shown to be active against plant pathogens (Scholz et al. 2011, 2014; Sumi et al. 2015).

The principal component analysis (PCA) performed with strains from different continents, isolation sources, ANI, dDDH, secondary metabolite profile, among others, revealed that 103 out of the 105 strains of *B. velezensis* grouped closer, while the two remaining strains were separated from this main group (Figure 2A). Strain OSY-S3 of *B. velezensis* separated from the others due to the presence of a greater number of CRISPr arrays in its genome and the presence of genes encoding the compounds bacilomycin and plantathiazolincin; whereas strain AGVL-005 lacked difficidin and macrolactin genes, which also happens in genomes of *B. amyloliquefaciens*. Strains of *B. amyloliquefaciens* and the only *B. siamensis* strain available clustered apart from each other and from *B. velezensis* (Figure 2A).

The secondary metabolites difficidin and macrolactin, along with ANI and dDDH, were the variables that contributed the most in the PCA analysis (Figure 2B). Despite the fact that *B. velezensis* represents a globally distributed species, 75% of the isolates deposited in the database came from the Asian continent, mostly from China. The others were from the American (14.3%) and European (9.5%) continents, with only one isolate from Africa. Most *B. velezensis* strains were obtained from rhizosphere/plant (40%), soil (30%) and food (22%). Within *B. amyloliquefaciens*, most isolates were also Asian, but with the predominance of strains isolated from food (Supplementary Table S4).
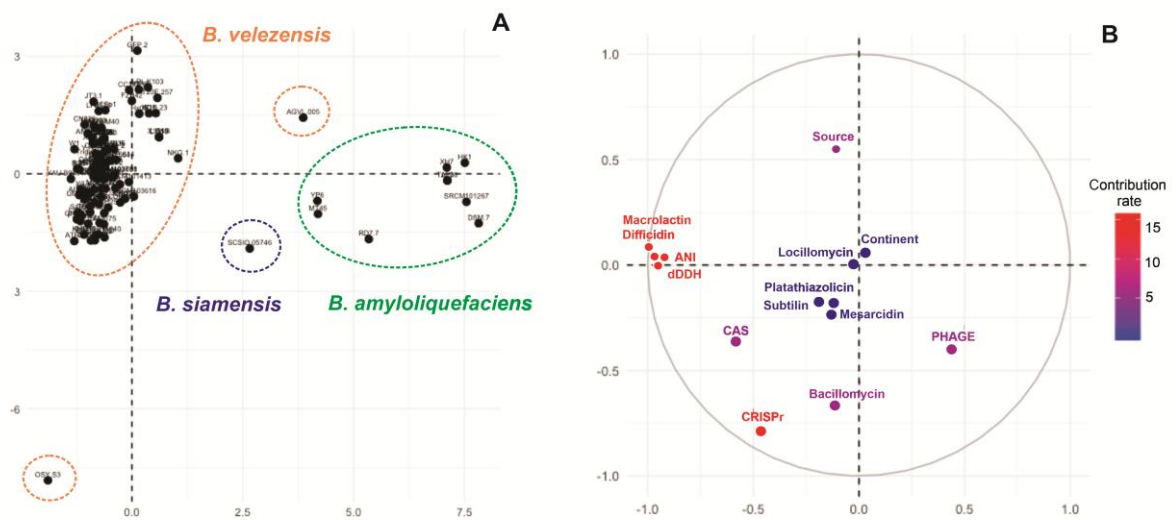
Figure 2. Principal components analysis (PCA) of 104 strains of *Bacillus velezensis* performed based on ANI, dDDH, secondary metabolite profiles, origin, source, presence of CRISPr/Cas arrays and phages. A) Clustering of the 104 strains of *B. velezensis* and 9 strains of *B. amyloliquefaciens*, including the strain type DSM 7. B) Variables genomic, colored according to the contribution rates in the analysis.

CRISPr/Cas arrays (clustered regularly interspaced short palindromic repeats) are part of the immune system of bacteria and archaea (Cain & Boinett 2013), protecting them against invaders such as bacteriophages and plasmids (Rath et al. 2015). These arrays were found in more than 85% of the genomes of *B. velezensis* studied. When present, the number of Cas copies was always higher than the number of CRISPr copies, with the exception of strain OSY-S3, with 5 copies of each. On the other hand, only 33% of the genomes of *B. amyloliquefaciens* possessed CRISPr/Cas arrays. However, due to the low number of genomes of this species, little can be inferred at this time. CRISPr loci may be used to provide phylogenetic relationships among bacterial lineages and more recently have been used as tools for transformation (Supply et al. 2013; Cain & Boinett 2013; Altenbuchner, 2016).

This comparative analysis provided a comprehensive understanding of the genomes of *B. velezensis* and closely related species. Special emphasis was given to the taxonomic classification of this group, where genomes of 115 strains were evaluated. From these, a total of 19 strains deposited as *B. amyloliquefaciens* were reclassified as *B. velezensis*. In summary, 105 strains were shown to be *B. velezensis*, 9 were *B. amyloliquefaciens* and 1 was *B. siamensis* (Supplementary Table S2).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1999. Basic local alignment search tool. J Mol Biol. 215:403-410.

Altenbuchner J. 2016. Editing of the *Bacillus subtilis* genome by the CRISPR-Cas9 system. Appl Environ Microbiol. 82:5421-5427.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc, 2010.

Arndt D et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44:16-21.

Bankevich A et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comp Biol. 19:455-477.

Brettin T et al. 2015. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 5:8365.

Cai XC, Liu CH, Wang BT, Xue YR. 2017. Genomic and metabolic traits endow *Bacillus velezensis* CC09 with a potential biocontrol agent in control of wheat powdery mildew disease. Microbiol Res. 196:89-94.

Cain AK, Boinett CJ. 2013. A CRISPR view of genome sequences. Nat Rev Microbiol. 11(4):226.

Couvin D et al. 2018. CRISPRCasFinder: an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* 46:246-251.

Dunlap CA, Kim SJ, Kwon SW, Rooney AP. 2016. *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens*; *Bacillus methylotrophicus*, *Bacillus*

amyloliquefaciens subsp. *plantarum* and 'Bacillus oryzicola' are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. Int J Syst Evol Microbiol. 66:1212–1217.

Fan B, Blom J, Klenk HP, Borriss R. *Bacillus amyloliquefaciens*, *Bacillus velezensis* and *Bacillus siamensis* form an "operational group B. amyloliquefaciens" within the *B. subtilis* species complex. Front Microbiol. 8:22.

Fira D, Dimkić I, Berić T, Lozo J, Stanković S. 2018. Biological control of plant pathogens by *Bacillus* species. J Biotechnol. 285:44-55.

Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. Source Code Biol Med. 6:11.

Gao Z, Zhang B, Liu H, Han J, Zhang Y. 2017. Identification of endophytic *Bacillus velezensis* ZSY-1 strain and antifungal activity of its volatile compounds against *Alternaria solani* and *Botrytis cinerea.* Biol Control 105:27-39.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.

Huerta-Cepas J et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res. 44:286-293.*

Jiang CH et al. 2018. *Bacillus velezensis*, a potential and efficient biocontrol agent in control of pepper gray mold caused by *Botrytis cinerea*. Biol Control 126:147-157.

Katoh K, Rosewicki J, Yamada KD. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinf. in press: bbx108.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 35(6):1547-1549.

Lagesen K et al. 2007. DW. RNammer: consistent annotation of rRNA genes in genomic sequences. Nucl Acids Res. 35:3100-3108.

Lee YK, Kim HW, Liu CL, Lee HK. 2003. A simple method for DNA extraction from marine bacteria that produce extracellular materials. J Microbiol Method. 52:245-250.

Lim SM et al. 2017. Diffusible and volatile antifungal compounds produced by an antagonistic *Bacillus velezensis* G341 against various phytopathogenic fungi. Plant Pathol J. 33:488-498.

Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucl Acids Res. 44:54-57.

Martins SA et al. 2018. Common bean (*Phaseolus vulgaris* L.) growth promotion and biocontrol by rhizobacteria under *Rhizoctonia solani* suppressive and conducive soils. Appl Soil Ecol. 127:129-135.

Martins SJ, Medeiros FHV, Souza RM, Resende MLV, Ribeiro-Jr. PM. 2013. Biological control of bacterial wilt of common bean by plnt growth-promoting rhizobacteria. Biol Control 66:65-71.

Martins SJ et al. 2019. Microbial volatiles organic compounds control anthracnose (*Colletotrichum lindemuthianum*) in common bean (*Phaseolus vulgaris* L.). Biol Control 131:36-42.

Medeiros FHV et al. 2011. Transcriptional profiling in cotton associated with *Bacillus subtilis* (UFLA285) induced biotic-stress tolerance. Plant Soil 347:327-337.

Medeiros FHV et al. 2015. Screening of endospore-forming bacteria for cotton seed treatment against bacterial blight and damping-off. Adv Plants Agric Res. 2:00056.

Mukherjee A, Das K. 2005. Correlation between diverse cyclic lipopeptides production and regulation of growth and substrate utilization by *Bacillus subtilis* strains in a particular habitat**.** FEMS Microbiol Ecol. 54:479-489.

Nikolenko SI, Korobeynikov AI, Alekseyev MA. 2013. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC genomics 14:S7.

Olishevska S, Nickzad A, Déziel E. 2019. *Bacillus* and *Paenibacillus* secreted polyketides and peptides involved in controlling human and plant pathogens. Appl Microbiol Biotechnol. 103:1189-1215.

Peng Y, Leung HCM, Yiu SM, Chin FYL. 2010. IDBA–a practical iterative de Bruijn graph de novo assembler. Res Comp Mol Biol. 6044:426-440.

Piro VC et al. 2014. FGAP: an automated gap closing tool. BMC Res Notes **7**:371.

Rath D, Amlinger L, Rath A, Lundgrenb M. 2015. The CRISPR-Cas immune system: Biology, mechanisms and applications. Biochimie 117:119-128.

Richter M, Rosselló-Móra R, Glöckner FO, Peplies J. 2015. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32:929-931.

Rodriguez RLM, Konstantinidis KT. 2014. Bypassing cultivation to identify bacterial species. Microbe 9:111-118.

Ruiz-García C, Béjar V, Martínez-Checa F, Llamas I, Quesada E. 2005. *Bacillus velezensis* sp. nov., a surfactant-producing bacterium isolated from the river Vélez in Málaga, southern Spain. Int J Syst Evol Microbiol. 55:191-195.

Rutherford K et al. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16(10): 944-945.

Scholz R et al. 2011. Plantazolicina novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42. J. Bacteriol. 193:215-224.

Scholz R et al. 2014. Amylocyclicin, a novel circular bacteriocin produced by *Bacillus amyloliquefaciens* FZB42. J. Bacteriol. 196:1842-1852.

Sharma V, Patil PB. 2011. Resolving the phylogenetic and taxonomic relationship of *Xanthomonas* and *Stenotrophomonas* strains using complete *rpoB* gene sequence. PLoS Curr. 3:RRN1239.

Shafi J, Tian H, Mingshan J. 2017. *Bacillus* species as versatile weapons for plant pathogens: a review. Biotech Biotechnol Equip. 31:446-459.

Sumi CD, Yang BW, Yeo IC, Hahm YT. 2015. Antimicrobial peptides of the genus *Bacillus*: a new era for antibiotics. Can J Microbiol. 61:93-103.

Supply P et al. 2013. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. Nat Genet. 45:172-179.

Van Domselaar GH et al. 2005. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res. 33:455-459.

Wasmuth EV, Lima CD. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45:158-169.

Wattam AR et al. 2014. PATRIC, the bacterial bioinfor- matics database and analysis resource. Nucleic Acids Res. 42:581-591.

Wang LT, Lee FL, Tai CJ, Kuo HP. 2008. *Bacillus velezensis* is a later heterotypic synonym of *Bacillus amyloliquefaciens.* Int J Syst Evol Microbiol. 58(3):671-675.

Weber T et al. 2015. AntiSMASH 3.0 - a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43:237-243.

Yamamoto S, Shiraishi S, Suzuki S. 2015. Are cyclic lipopeptides produced by *Bacillus amyloliquefaciens* S13-3 responsible for the plant defence response in strawberry against *Colletotrichum gloeosporioides*. Lett Appl Microbiol. 60:379-386.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821-829.

Table S1. Assembly statistics and genome features of *Bacillus velezensis* strain UFLA258.

a) Genome statistics

Attribute	Value
Contigs	55
Largest contig (bp)	999,521
Total length (bp)	3,947,620
N50	623,714
L50	3
L75	4
GC (%)	46.5

b) Genome features

Attribute	Value
Chromosome size (bp)	3,947,620
Chromosomal genes (number)	3,949
Protein coding genes	3,747
RNA genes	116
Pseudogenes	66
Genes with function prediction	3,439
CRISPR arrays	1

Table S2. Species boundaries for *Bacillus velezensis*, *B. amyloliquefaciens* and *B. siamensis* based on genomic properties and indexes.

Strain	Size (Mb)	GC (%)	Protein	ANI (%)	dDDH (%)
B. velezensis					
UFLA258	3.95	46.69	3,147	98.87	92.00
104 strains	3.68-4.39	43.23-47.00	2,251-4,185	97.57-98.89	80.10-92.00
B. amyloliquefaciens					
9 strains	3.68-4.08	45.69-46.30	3,501-4,028	97.88-99.98	81.30-100.0
B. siamensis					
1 strain	4.28	45.97	4,150	100.00	100.00
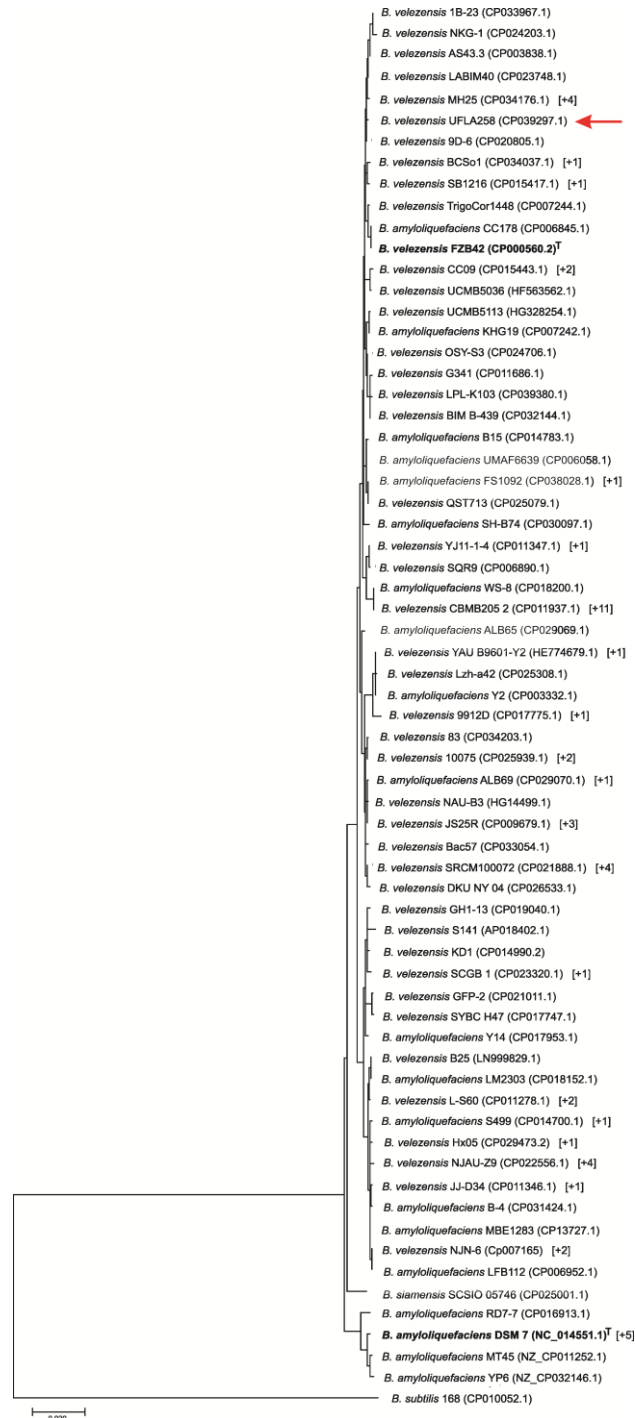
Figure S1. Phylogenetic tree with the taxonomic placement of strain UFLA258. The phylogenetic tree was constructed in MEGA X v10.1 (Kumar et al. 2018) based on complete nucleotide sequences of the *rpoB* gene (3,582 bp) aligned in MAFFT (Katoh et al. 2017). The tree was constructed with the Maximum Likeliood method and the Kimura 2-parameter model. Bootstrap values were calculated with 1,000 resamplings. The scale indicates the number of substitutions per site.

Table S3. Genomes used in this study with the taxonomical re-identification of some *B. amyloliquefaciens* based on ANI and dDDH values.

Strains	Accession	Current classification	FZB42^T		DMS 7^T		Re-identification
			ANI (%)	dDDH (%)	ANI (%)	dDDH (%)	
83	NZ_CP034203.1	*B. velezensis*	98.15	85.80	-	-	=
157	NZ_CP022341.1	*B. velezensis*	98.21	85.80	-	-	=
10075	NZ_CP025939.1	*B. velezensis*	97.91	85.30	-	-	=
8_2	NZ_CP028439.1	*B. velezensis*	97.64	80.60	-	-	=
131-4	NZ_CP028441.1	*B. velezensis*	97.62	80.60	-	-	=
1B-23	NZ_CP033967.1	*B. velezensis*	98.55	91.10	-	-	=
9912D	NZ_CP017775.1	*B. velezensis*	97.77	83.20	-	-	=
9D-6	NZ_CP020805.1	*B. velezensis*	98.74	91.20	-	-	=
AGVL-005	CP024922.1	*B. velezensis*	98.30	89.70	-	-	=
ANSB01E	NZ_CP036518.1	*B. velezensis*	98.11	85.30	-	-	=
AS43.3	NC_019842.1	*B. velezensis*	98.74	91.40	-	-	=
ATR2	NZ_CP018133.1	*B. velezensis*	98.01	85.60	-	-	=
B25	NZ_LN999829.1	*B. velezensis*	97.60	80.40	-	-	=
Bac57	NZ_CP033054.1	*B. velezensis*	97.79	84.40	-	-	=
BCSo1	NZ_CP034037.1	*B. velezensis*	98.89	90.70	-	-	=

BIM B-439D	NZ_CP032144.1	*B. velezensis*	98.74	90.40	-	-	=
BS-37	NZ_CP023414.1	*B. velezensis*	98.68	90.40	-	-	=
CAU B946	NC_016784.1	*B. velezensis*	97.51	80.20	-	-	=
CBMB205	NZ_CP011937.1	*B. velezensis*	98.12	85.40	-	-	=
CBMB205	NZ_CP014838.1	*B. velezensis*	98.12	85.40	-	-	=
CC09	NZ_CP015443.1	*B. velezensis*	98.38	88.90	-	-	=
CGMCC11640	NZ_CP026610.1	*B. velezensis*	98.26	88.90	-	-	=
CMT-6	NZ_CP025341.1	*B. velezensis*	97.54	81.00	-	-	=
CN026	NZ_CP024897.1	*B. velezensis*	98.15	85.60	-	-	=
DKU_NT_04	NZ_CP026533.1	*B. velezensis*	97.71	82.90	-	-	=
DR-08	NZ_CP028437.1	*B. velezensis*	98.13	85.40	-	-	=
DSYZ	NZ_CP030150.1	*B. velezensis*	98.25	88.70	-	-	=
FZB42	NC_009725.1	*B. velezensis*	100.00	100.00	-	-	=
G341	NZ_CP011686.1	*B. velezensis*	98.73	90.40	-	-	=
GFP-2	NZ_CP021011.1	*B. velezensis*	97.71	80.80	-	-	=
GH1-13	NZ_CP019040.1	*B. velezensis*	97.43	80.80	-	-	=
GQJK49	NZ_CP021495.1	*B. velezensis*	98.10	85.40	-	-	=
GYL4	NZ_CP020874.1	*B. velezensis*	98.12	86.80	-	-	=

Hx05	NZ_CP029473.1	*B. velezensis*	97.53	80.10	-	-	=
J7-1	NZ_CP028440.1	*B. velezensis*	97.59	80.60	-	-	=
JJ-D34	NZ_CP011346.1	*B. velezensis*	97.50	80.30	-	-	=
JS25R	NZ_CP009679.1	*B. velezensis*	98.19	85.80	-	-	=
JT3-1	NZ_CP032506.1	*B. velezensis*	98.09	85.40	-	-	=
JTYP2	NZ_CP020375.1	*B. velezensis*	98.09	85.40	-	-	=
K26	NZ_CP023075.1	*B. velezensis*	97.87	84.40	-	-	=
KD1	NZ_CP014990.2	*B. velezensis*	97.71	81.30	-	-	=
L-1	NZ_CP023859.1	*B. velezensis*	98.71	91.10	-	-	=
LAMBIM40	NZ_CP023748.1	*B. velezensis*	98.72	91.00	-	-	=
LB002	NZ_CP037417.1	*B. velezensis*	97.36	80.40	-	-	=
LDO2	NZ_CP029034.1	*B. velezensis*	98.10	85.40	-	-	=
L-H15	NZ_CP010556.1	*B. velezensis*	97.62	80.20	-	-	=
LPL-K103	NZ_CP039380.1	*B. velezensis*	98.45	89.90	-	-	=
LS69	NZ_CP0159111	*B. velezensis*	98.34	85.30	-	-	=
L-S60	NZ_CP011278.1	*B. velezensis*	97.62	80.30	-	-	=
Lzh-a42	NZ_CP025308.1	*B. velezensis*	98.10	85.60	-	-	=
M75	NZ_CP016395.1	*B. velezensis*	97.58	80.60	-	-	=

MH25	NZ_CP034176.1	*B. velezensis*	98.71	91.10	-	-	=
NAU-B3	NC_022530.1	*B. velezensis*	98.09	85.80	-	-	=
NJAU-Z9	NZ_CP022556.1	*B. velezensis*	97.62	80.70	-	-	=
NJN-6	NZ_CP007165.1	*B. velezensis*	97.39	80.50	-	-	=
NKG-1	NZ_CP024203.1	*B. velezensis*	98.71	92.00	-	-	=
NY12-2	NZ_CP033576.1	*B. velezensis*	97.79	84.40	-	-	=
OSY-S3	CP024706.1	*B. velezensis*	98.65	89.60	-	-	=
QST713	NZ_CP025079.1	*B. velezensis*	98.45	88.30	-	-	=
S141	NZ_AP018402.1	*B. velezensis*	98.66	89.80	-	-	=
S3-1	NZ_CP016371.1	*B. velezensis*	98.12	85.40	-	-	=
SB1216	CP015417.1	*B. velezensis*	98.76	90.20	-	-	=
SCDB 291	NZ_CP022654.2	*B. velezensis*	97.57	81.30	-	-	=
SCGB 1	NZ_CP023320.1	*B. velezensis*	97.58	81.20	-	-	=
SCGB 574	NZ_CP023431.1	*B. velezensis*	98.19	85.90	-	-	=
SQR9	NZ_CP006890.1	*B. velezensis*	98.06	86.80	-	-	=
SRCM100072	NZ_CP021888.1	*B. velezensis*	97.75	84.50	-	-	=
SRCM101413	NZ_CP021890.1	*B. velezensis*	97.82	84.20	-	-	=
SRCM103616	NZ_CP035410.1	*B. velezensis*	97.83	84.20	-	-	=

SRCM103691	NZ_CP035393.1	*B. velezensis*	97.89	84.20	-	-	=
SRCM103788	NZ_CP035399.1	*B. velezensis*	97.85	84.20	-	-	=
sx01604	NZ_CP018007.1	*B. velezensis*	98.09	85.40	-	-	=
SYBC H47	NZ_CP017747.1	*B. velezensis*	97.73	82.00	-	-	=
T20E-257	NZ_CP021976.1	*B. velezensis*	97.60	80.60	-	-	=
TB1501	NZ_CP022531.1	*B. velezensis*	98.78	91.10	-	-	=
TJ02	NZ_CP024797.1	*B. velezensis*	98.74	91.30	-	-	=
TrigoCor1448	NZ_CP007244.1	*B. velezensis*	98.78	91.60	-	-	=
UCMB-5033	NC_022075.1	*B. velezensis*	98.65	90.50	-	-	=
UCMB5036	NC_020410.1	*B. velezensis*	98.55	89.00	-	-	=
UCMB5113	NC_022081.1	*B. velezensis*	98.81	90.50	-	-	=
UFLA258	NZ_CP039297.1	*B. velezensis*	98.87	92.00	-	-	=
W1	NZ_CP028375.1	*B. velezensis*	98.15	85.50	-	-	=
YAUB9601Y2	NC_017061.1	*B. velezensis*	98.06	85.70	-	-	=
YJ11-1-4	NZ_CP011347.1	*B. velezensis*	98.10	86.80	-	-	=
ZF2	NZ_CP032154.1	*B. velezensis*	98.13	85.40	-	-	=
ZL918	NZ_CP021338.1	*B. velezensis*	97.59	80.20	-	-	=
ALB65	NZ_CP029069.1	*B. amyloliquefaciens*	98.02	85.30	**93.39**	**56.10**	*B. velezensis*

ALB69	NZ_CP029070.1	*B. amyloliquefaciens*	98.17	86.40	**93.35**	**55.80**	*B. velezensis*
ALB79	NZ_CP029071.1	*B. amyloliquefaciens*	98.23	88.50	**93.29**	**56.00**	*B. velezensis*
B15	NZ_CP014783.1	*B. amyloliquefaciens*	97.56	80.20	**93.37**	**55.50**	*B. velezensis*
B-4	NZ_CP031424.1	*B. amyloliquefaciens*	98.43	88.60	**93.4**	**55.60**	*B. velezensis*
CC178	NC_022653.1	*B. amyloliquefaciens*	99.99	100.00	**93.41**	**56.10**	*B. velezensis*
DSM 7	NC_014551.1	*B. amyloliquefaciens*	**93.84**	**56.20**	100.00	100.00	=
FS1092	NZ_CP038028.1	*B. amyloliquefaciens*	98.14	88.30	**93.23**	**55.80**	*B. velezensis*
HK1	NZ_CP018902.1	*B. amyloliquefaciens*	**93.82**	**56.10**	99.98	100.00	=
IT-45	NC_020272.1	*B. amyloliquefaciens*	97.6	80.50	**93.44**	**55.50**	*B. velezensis*
KHG19	NZ_CP007242.1	*B. amyloliquefaciens*	98.74	90.50	**93.44**	**56.10**	*B. velezensis*
LFB112	NC_023073.1	*B. amyloliquefaciens*	97.45	80.30	**93.25**	**55.60**	*B. velezensis*
LL3	NC_017190.1	*B. amyloliquefaciens*	**93.74**	**55.70**	99.47	96.40	=
LM2303	NZ_CP018152.1	*B. amyloliquefaciens*	97.53	80.30	**93.28**	**55.20**	*B. velezensis*
MBE1283	NZ_CP013727.1	*B. amyloliquefaciens*	97.52	79.90	**93.33**	**55.20**	*B. velezensis*
MT45	NZ_CP011252.1	*B. amyloliquefaciens*	**93.92**	**56.10**	98.09	85.70	=
RD7-7	NZ_CP016913.1	*B. amyloliquefaciens*	**93.99**	**56.30**	97.57	81.30	=
S499	NZ_CP014700.1	*B. amyloliquefaciens*	97.65	80.50	**93.46**	**55.50**	*B. velezensis*
SH-B74	NZ_CP030097.1	*B. amyloliquefaciens*	98.53	89.90	**93.31**	**56.10**	*B. velezensis*
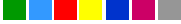
SRCM101267	NZ_CP021505.1	*B. amyloliquefaciens*	**93.78**	**56.10**	99.69	98.90	=
TA208	NC_017188.1	*B. amyloliquefaciens*	**93.69**	**55.70**	99.28	95.20	=
UMAF6614	NZ_CP006960.1	*B. amyloliquefaciens*	98.22	86.40	**93.39**	**56.00**	*B. velezensis*
UMAF6639	NZ_CP006058.1	*B. amyloliquefaciens*	98.21	88.60	**93.35**	**55.70**	*B. velezensis*
WS-8	NZ_CP018200.1	*B. amyloliquefaciens*	98.13	85.40	**93.27**	**55.40**	*B. velezensis*
XH7	NC_017191.1	*B. amyloliquefaciens*	**93.76**	**55.70**	99.31	95.40	=
Y14	NZ_CP017953.1	*B. amyloliquefaciens*	98.15	85.60	**93.25**	**55.90**	*B. velezensis*
Y2	NC_017912.1	*B. amyloliquefaciens*	97.59	80.20	**93.24**	**55.60**	*B. velezensis*
YP6	NZ_CP032146.1	*B. amyloliquefaciens*	**93.90**	**56.20**	97.88	85.90	=
SCSIO 05746	NZ_CP025001.1	*B. siamensis*	**94.15**	**58.00**	**92.99**	**54.60**	=

Digital DNA-DNA Hybridization (dDDH) and Average Nucleotide Acid Identity (ANI) comparisons with the FZB42 (*B. velezensis* type strain) and DMS 7 (*B. amyloliquefaciens* type strain) genomes. Values out of the range for species delineation are shown in bold (ANI> 95%, Auch et al. 2010; dDDH> 70%, Richter et al. 2015).

Table S4. Genomic characteristics, source and origin of the *B. velezensis*, *B. amyloliquefaciens* and *B. siamensis* strains used in this study. These data, together with ANI and dDDH were used in the PCA analysis.

Strain	Accession	Continent	Source	CRISPR*	Cas*	Phage*	Metabolites**
B. velezensis							
83	NZ_CP034203.1	American	Plant	1	8	0	▮▮▮▮▮▮▮
157	NZ_CP022341.1	Asian	Plant	1	7	1	▮▮▮▮▮▮▮
10075	NZ_CP025939.1	Asian	Food	2	7	3	▮▮▮▮▮▮▮▮
8_2	NZ_CP028439.1	Asian	Soil	1	6	2	▮▮▮▮▮▮▮
131-4	NZ_CP028441.1	Asian	Soil	0	0	2	▮▮▮▮▮▮▮
1B-23	NZ_CP033967.1	American	Rhizosphere	0	0	2	▮▮▮▮▮▮▮
9912D	NZ_CP017775.1	Asian	Soil	1	7	3	▮▮▮▮▮▮▮
9D-6	NZ_CP020805.1	American	Rhizosphere	1	7	0	▮▮▮▮▮▮▮
AGVL-005	CP024922.1	American	Plant	0	0	2	▮▮ ▮▮ ▮
ANSB01E	NZ_CP036518.1	Asian	Other	2	7	1	▮▮▮▮▮▮▮
AS43.3	NC_019842.1	American	Rhizosphere	1	7	1	▮▮▮▮▮▮▮▮
ATR2	NZ_CP018133.1	Asian	Soil	3	7	1	▮▮▮▮▮▮▮▮
B25	NZ_LN999829.1	Asian	Rhizosphere	2	8	0	▮▮▮▮▮▮▮
Bac57	NZ_CP033054.1	Asian	Soil	2	8	1	▮▮▮▮▮▮▮

BCSo1	NZ_CP034037.1	European	Other	1	7	1	🟩🟦🟥🟨🟦🟪⬜
BIM B-439D	NZ_CP032144.1	European	Soil	3	7	2	🟩🟦🟥🟨🟦🟪⬜
BS-37	NZ_CP023414.1	Asian	Food	1	7	2	🟩🟦🟥🟨🟦🟪🟧
CAU B946	NC_016784.1	Asian	Rhizosphere	1	7	1	🟩🟦🟥🟨🟦🟪⬜🟧
CBMB205	NZ_CP011937.1	Asian	Rhizosphere	2	7	1	🟩🟦🟥🟨🟦🟪⬜
CBMB205	NZ_CP014838.1	Asian	Rhizosphere	1	7	1	🟩🟦🟥🟨🟦🟪⬜
CC09	NZ_CP015443.1	Asian	Plant	2	6	1	🟩🟦🟥🟨🟦🟪⬜
CGMCC11640	NZ_CP026610.1	Asian	Soil	3	6	2	🟩🟦🟥🟨🟦🟪⬜
CMT-6	NZ_CP025341.1	Asian	Food	1	7	1	🟩🟦🟥🟨🟦🟪⬜
CN026	NZ_CP024897.1	European	Other	2	7	0	🟩🟦🟥🟨🟦🟪⬜
DKU_NT_04	NZ_CP026533.1	Asian	Food	1	6	1	🟩🟦🟥🟨🟦🟪⬜
DR-08	NZ_CP028437.1	Asian	Soil	1	7	1	🟩🟦🟥🟨🟦🟪⬜
DSYZ	NZ_CP030150.1	Asian	Rhizosphere	3	6	1	🟩🟦🟥🟨🟦🟪⬜
FZB42	NC_009725.1	European	Rhizosphere	0	0	0	🟩🟦🟥🟨🟦🟪⬜🟧
G341	NZ_CP011686.1	Asian	Rhizosphere	2	7	1	🟩🟦🟥🟨🟦🟪⬜
GFP-2	NZ_CP021011.1	Asian	Other	0	0	0	🟩🟦🟥🟨🟦🟪⬜
GH1-13	NZ_CP019040.1	Asian	Soil	2	8	1	🟩🟦🟥🟨🟦🟪⬜ 🟦
GQJK49	NZ_CP021495.1	Asian	Rhizosphere	1	7	1	🟩🟦🟥🟨🟦🟪⬜

Continue...

GYL4	NZ_CP020874.1	Asian	Plant	1	7	0	
Hx05	NZ_CP029473.1	Asian	Rhizosphere	0	0	1	
J7-1	NZ_CP028440.1	Asian	Soil	1	7	2	
JJ-D34	NZ_CP011346.1	Asian	Food	1	7	1	
JS25R	NZ_CP009679.1	Asian	Plant	1	7	1	
JT3-1	NZ_CP032506.1	Asian	Other	1	7	0	
JTYP2	NZ_CP020375.1	Asian	Plant	1	7	1	
K26	NZ_CP023075.1	Asian	Food	2	7	1	
KD1	NZ_CP014990.2	Asian	Food	0	0	1	
L-1	NZ_CP023859.1	Asian	Soil	2	6	2	
LAMBIM40	NZ_CP023748.1	American	Faeces	1	7	1	
LB002	NZ_CP037417.1	Asian	Soil	1	7	3	
LDO2	NZ_CP029034.1	Asian	Soil	2	7	0	
L-H15	NZ_CP010556.1	Asian	Soil	0	0	2	
LPL-K103	NZ_CP039380.1	Asian	Plant	0	0	1	
LS69	NZ_CP0159111	Asian	Plant	1	7	1	
L-S60	NZ_CP011278.1	Asian	Soil	0	0	2	
Lzh-a42	NZ_CP025308.1	Asian	Soil	2	7	1	

M75	NZ_CP016395.1	Asian	Soil	2	7	2	
MH25	NZ_CP034176.1	Asian	Rhizosphere	2	6	3	
NAU-B3	NC_022530.1	Asian	Plant	1	7	1	
NJAU-Z9	NZ_CP022556.1	Asian	Soil	1	7	1	
NJN-6	NZ_CP007165.1	Asian	Plant	1	7	2	
NKG-1	NZ_CP024203.1	Asian	Soil	0	0	4	
NY12-2	NZ_CP033576.1	Asian	Food	2	8	2	
OSY-S3	CP024706.1	American	Food	5	5	2	
QST713	NZ_CP025079.1	European	Soil	2	7	2	
S141	NZ_AP018402.1	Asian	Rhizosphere	1	8	1	
S3-1	NZ_CP016371.1	Asian	Rhizosphere	1	7	1	
SB1216	CP015417.1	American	Soil	3	7	2	
SCDB 291	NZ_CP022654.2	Asian	Soil	1	8	3	
SCGB 1	NZ_CP023320.1	Asian	Food	2	8	1	
SCGB 574	NZ_CP023431.1	Asian	Food	1	7	1	
SQR9	NZ_CP006890.1	Asian	Rhizosphere	0	0	0	
SRCM100072	NZ_CP021888.1	Asian	Food	1	8	1	
SRCM101413	NZ_CP021890.1	Asian	Food	1	6	3	

SRCM103616	NZ_CP035410.1	Asian	Food	1	7	4	
SRCM103691	NZ_CP035393.1	Asian	Food	1	7	2	
SRCM103788	NZ_CP035399.1	Asian	Food	1	7	2	
sx01604	NZ_CP018007.1	Asian	Soil	2	7	1	
SYBC H47	NZ_CP017747.1	Asian	Food	1	8	1	
T20E-257	NZ_CP021976.1	Asian	Plant	0	0	2	
TB1501	NZ_CP022531.1	Asian	Soil	2	6	1	
TJ02	NZ_CP024797.1	Asian	Soil	2	7	3	
TrigoCor1448	NZ_CP007244.1	American	Rhizosphere	1	8	1	
UCMB-5033	NC_022075.1	European	Soil	1	7	1	
UCMB5036	NC_020410.1	European	Rhizosphere	1	7	1	
UCMB5113	NC_022081.1	European	Soil	2	6	1	
UFLA258	NZ_CP039297.1	American	Soil	1	7	1	
W1	NZ_CP028375.1	Asian	Faeces	2	6	0	
YAUB9601Y2	NC_017061.1	Asian	Rhizosphere	2	7	0	
YJ11-1-4	NZ_CP011347.1	Asian	Food	1	7	0	
ZF2	NZ_CP032154.1	Asian	Plant	1	7	1	
ZL918	NZ_CP021338.1	Asian	Plant	1	7	2	

ALB65	NZ_CP029069.1	American	Food	1	7	1	
ALB69	NZ_CP029070.1	American	Food	1	7	2	
ALB79	NZ_CP029071.1	American	Plant	2	7	1	
B15	NZ_CP014783.1	Asian	Plant	1	8	1	
B-4	NZ_CP031424.1	Asian	Plant	1	8	1	
CC178	NC_022653.1	Asian	Plant	0	0	0	
FS1092	NZ_CP038028.1	American	Food	2	7	1	
IT-45	NC_020272.1	American	Rhizosphere	2	6	1	
KHG19	NZ_CP007242.1	Asian	Food	2	7	1	
LFB112	NC_023073.1	Asian	Other	2	6	1	
LM2303	NZ_CP007242.1	Asian	Faeces	1	7	1	
MBE1283	NC_023073.1	Asian	Food	1	7	2	
S499	NZ_CP018152.1	Africa	Soil	3	7	0	
SH-B74	NZ_CP013727.1	Asian	Soil	3	6	1	
UMAF6614	NZ_CP014700.1	European	Rhizosphere	1	7	2	
UMAF6639	NZ_CP030097.1	European	Rhizosphere	3	7	1	
WS-8	NZ_CP006960.1	Asian	Soil	2	7	1	
Y14	NZ_CP006058.1	Asian	Rhizosphere	1	7	1	

							Metabolites**
Y2	NZ_CP018200.1	Asian	Rhizosphere	2	7	1	▪▪▪▪▪▪▪ ▪
B. amyloliquefaciens							
DSM 7	NC_014551.1	European	Soil	0	0	5	▪▪ ▪▪ ▪
HK1	NZ_CP018902.1	Asian	Other	0	0	4	▪▪ ▪▪ ▪
LL3	NC_017190.1	Asian	Food	0	0	2	▪▪ ▪▪ ▪
MT45	NZ_CP011252.1	Asian	Food	1	7	1	▪▪ ▪▪▪▪
RD7-7	NZ_CP016913.1	Asian	Food	2	8	0	▪▪ ▪▪ ▪
SRCM101267	NZ_CP021505.1	Asian	Food	0	0	4	▪▪ ▪▪ ▪
TA208	NC_017188.1	Asian	Food	0	0	2	▪▪ ▪▪ ▪
XH7	NC_017191.1	Asian	Rhizosphere	0	0	2	▪▪ ▪▪ ▪
YP6	NZ_CP032146.1	Asian	Rhizosphere	1	7	1	▪▪ ▪▪▪▪
B. siamensis							
SCSIO 05746	NZ_CP025001.1	Asian	Soil	2	6	2	▪▪▪▪▪▪▪

* Number of copies of each element in the genome.

** Metabolites: ▪ Bacilysin, ▪ bacillibactin, ▪ difficidin, ▪ fengyncin, ▪ bacillaen, ▪ macrolactin, ▪ surfactin, ▪ plantathizolicin, ▪ subtilin, ▪ mersacidin, ▪ bacillomycin, ▪ locilomycin.

CURRÍCULO DA AUTORA

Fabíola de Jesus Silva é natural de Janaúba - MG, onde nasceu em 1990. Em 2008, ingressou no curso de Agronomia na Universidade Estadual de Montes Claros. Durante a graduação fez estágio voluntário (2010) no Laboratório de Fitopatologia e Microbiologia do Solo. Foi bolsista PIBIC/FAPEMIG (2011 – 2013) no mesmo laboratório. Concluída a graduação em 2013, ingressou no curso de Mestrado em Produção Vegetal no Semiarido na mesma instituição, sob a orientação da professora Dra. Regina Cássia Ferreira Ribeiro. Em 2016 iniciou o curso de Doutorado em Fitopatologia na Universidade Federal de Lavras, sob orientação do professor Dr. Vicente Paulo Campos.