Citizen science and digital data for trend analysis and impact assessment of *Prodiplosis* as an emerging pest in foliage crops

Ciencia ciudadana y datos digitales para el análisis de tendencias y la evaluación del impacto de *Prodiplosis* como plaga emergente en cultivos de follaje

Laura Alejandra Valbuena-Gaona[1], Hector Julio Villamil-Martha[2], Luz Mary Pardo-Ramirez[2], and Joaquín Guillermo Ramírez Gil[1,*]

ABSTRACT

Prodiplosis longifila is a pest of significant economic relevance, severely impacting crops like tomatoes and asparagus. Its effect on crops of ornamental foliage remains poorly documented, despite its growing importance in the agricultural sector. This study addresses our knowledge gap by implementing and validating digital tools of epidemiology (DE) and citizen science (CS) to enable a dynamic and participatory approach to pest monitoring. A trend analysis of scientific publications was conducted using web searches and social media interactions to identify topics concerning *Prodiplosis* over time, our knowledge gaps, and emerging areas of public interest. We assessed the impact of *Prodiplosis* on foliage crops, focusing on indirect effects and farmer-led management strategies shared through digital communication. Results show that digital tools such as trend monitoring on social media, web data analysis, WhatsApp group discussions, and farmer-managed digital platforms were effective for identifying the pest's distribution, significance, and control practices. DE and CS approaches revealed critical knowledge gaps concerning the biology, ecology, and management of *Prodiplosis*, particularly in ornamental crops. Field data confirmed the pest's negative impact on foliage yield and quality, with a strong dependence on chemical control methods, often applied without technical guidance. This study introduces an innovative methodology for assessing pest impacts through digital data analysis, offering practical insights for agricultural and policy decision-making. Moreover, the study highlights the potential of natural language processing as a powerful tool for synthesizing and detecting patterns in textual data and enhances the efficiency of pest surveillance and management systems.

Key words: bibliometrics, digital platforms, Google trends, social networks, spatial analysis.

RESUMEN

Prodiplosis longifila es una plaga de alta relevancia económica, que afecta gravemente cultivos como tomate y espárrago. Sin embargo, su impacto sobre cultivos de follaje ha sido poco estudiado, a pesar de su creciente importancia. Este estudio aborda dicha brecha mediante la implementación y validación de herramientas de epidemiología digital (ED) y ciencia ciudadana (CC), que permiten un enfoque participativo y dinámico para el monitoreo de esta plaga. Se realizó un análisis de tendencias basado en publicaciones científicas, búsquedas en internet e interacciones en redes sociales, con el objetivo de identificar los temas tratados, los vacíos de conocimiento y las áreas emergentes de interés. Adicionalmente, se evaluó el impacto de *Prodiplosis* en cultivos de follaje, describiendo sus efectos indirectos y las estrategias de manejo adoptadas por los agricultores a través de canales digitales. Los resultados muestran que herramientas digitales como el análisis de tendencias en redes sociales, la exploración de datos web, los grupos de WhatsApp y las plataformas digitales gestionadas por productores son eficaces para identificar la distribución, importancia y estrategias de control de *Prodiplosis*. Las metodologías de ED y CC también revelaron vacíos críticos en el conocimiento sobre la biología, ecología y manejo de esta plaga en cultivos ornamentales. El análisis de campo confirmó su impacto negativo en el rendimiento y la calidad del follaje, con una alta dependencia del control químico, usualmente sin asesoría técnica. Este estudio propone una metodología innovadora basada en datos digitales, destacando el potencial del procesamiento de lenguaje natural para fortalecer la vigilancia y gestión fitosanitaria.

Palabras clave: bibliometría, plataformas digitales, Google trends, redes sociales, análisis espacial.

Received for publication: February 17, 2025. Accepted for publication: April 25, 2025. Doi: 10.15446/agron.colomb.v43n1.118919

[1] Universidad Nacional de Colombia, Facultad de Ciencias Agrarias, Departamento de Agronomía, Laboratorio de Agrocomputación y Análisis epidemiológico, Center of Excellence in Scientific Computing, Bogotá (Colombia).
[2] Economía Agraria – Agropecuaria Villapard, Cachipay, Cundinamarca (Colombia).
[*] Corresponding author: jgramireg@unal.edu.co

Introduction

Digital data (DD) associated with digital epidemiology (DE) has become a crucial tool in the management of public health problems at the regional level, allowing the collection and analysis of real-time data for informed decision-making (Fulk *et al*., 2022; Salathé, 2018). This approach takes advantage of information and communication technologies (ICT), such as social networks and web platforms, in order to monitor and respond to disease outbreaks with more speed and precision (Cervellin *et al*., 2017; Fulk *et al*., 2022). Through the use of data from online platforms, DE has indirectly enabled the monitoring and analysis of the spread and distribution of diseases as an alternative to traditional epidemiological surveillance tools (Ekman & Litton, 2007; Fulk *et al*., 2022). The advantages include rapid data collection and early identification of epidemiological outbreaks. But these tools have some limitations, such as poor control over data quality, access to data, use of analytic tools, and access to and quality of the internet (Lippi & Cervellin, 2019; Park *et al*., 2018).

As a complement to the DE, citizen science (CS) is a research approach that actively involves the public in scientific research, contributing data collection, analysis, and problem-solving (Fraisl *et al*., 2022; Sherbinin *et al*., 2021). This participatory model democratizes science, leveraging the collective intelligence of non-experts to address complex challenges (Sherbinin *et al*., 2021). In the realm of public health, CS has gained significant importance, particularly in the field of DE (Carney *et al*., 2022). DE uses multiple computational and digital tools as methods to monitor and analyze disease patterns in real time (Fulk *et al*., 2022; Salathé, 2018). However, its success depends on engaging users who generate and share data through mobile apps, wearables, social media, and other digital platforms (Johnston *et al*., 2023; Rosas *et al*., 2022). This is where CS acts as a critical bridging tool, facilitating the interaction between DE and the general population.

In agricultural sciences, to our knowledge, the concept of DE and CS have not been widely explored. DE has been mainly used to monitor epidemics associated with public health in humans. One of the first reports was the study of the relationship between Google searches and the occurrence of respiratory diseases such as influenza (Ginsberg *et al*., 2009). Meanwhile, CS had its beginnings in the social and natural sciences, especially focused on ecology, and public health (Kullenberg & Kasperowski, 2016). Its fundamentals can be adapted to pest management that allows farmers access and real-time sharing of information on the presence and population dynamics of pests in their crops. This approach could not only improve the capacity to respond to phytosanitary problems, but also it promotes collaboration and knowledge sharing and fosters the democratization of knowledge (Ryan *et al*., 2018). In this way, the application of digital epidemiology to agriculture can lead to more efficient and sustainable pest management, reducing economic losses and promoting informed decisions for optimizing management strategies. The practical application of this concept is still in its early stages, but it can also help farmers improve their responses to phytosanitary problems (Ryan *et al*., 2018).

Empowering individuals to contribute plant health-related data can enhance CS surveillance systems, and this improves early outbreak detection, and supports data-driven decision-making. But it is necessary to recognize and identify biases and errors in the data (Tran *et al*., 2021). Additionally, CS surveillance systems foster public trust and awareness and encourage active participation in plant pest prevention and control. Initiatives such as participatory disease tracking apps exemplify how citizen engagement strengthens DE, transforming passive data collection into an interactive and collaborative process (Katapally *et al*., 2021). Ultimately, the synergy between CS and DE represents a paradigm shift, where communities become integral players in the global monitoring of plant health and response.

In the floricultural industry, Colombia is the world's second-largest exporter of cut flowers, and this production is complemented by that of green foliage, which represent an essential component in ornamental horticulture (Adebayo *et al*., 2020). Foliage is widely used in flower arranging, decoration, and landscaping because of its ability to beautify and enhance both indoor and outdoor spaces (Wei *et al*., 2023). The constant demand for foliage is reflected in a global market of high economic value and continuous growth (Wei *et al*., 2023).

In Colombia, green foliage took on greater importance in 2010 when the country began exporting bouquets, in which foliage and ferns played a complementary role to traditional flowers (Rojas Burgos, 2022). The most important foliage plants include Ruscus (*Ruscus aculeatus* L.) and Cocculus (*Cocculus laurifolius* L.), recognized for their potential for long vase life (Faust & Dole, 2021). The main foliage producing areas in Colombia are in the department of Cundinamarca, where they are grown in greenhouses, open field cultivation, and under shade nets.

Currently, foliage production in Colombia faces different threats, especially of phytosanitation. One of the most concerning pests is *Prodiplosis longifila* Gagné (Diptera: Cecidomyiidae) (Gagné, 1986). This pest is significant for damage to foliage, not only due to the damage it causes, but also due to the lack or inexistence of available information on its biology, ecology, population dynamics, impact, and sustainable management strategies. In the Cundinamarca highlands and particularly in the municipalities of Cachipay, Anolaima, Zipacón, and La Mesa, a considerable regional problem associated with the emergence of this emerged in recent years. The situation has generated uncertainty in the flower growing industry, since green foliage has been seriously affected by the presence of *P. longifila,* compromising the competitiveness and sustainability of these crops. One of the hosts reported for the species is the carnation (*Tagetes* sp.) (EPPO, 2017). Based on preferences for this species, it is possible that it could spread to other species of cut flowers and threaten their production. The importance of this pest lies in the fact that it is considered an invasive species in regions importing flowers from Colombia (EPPO, 2025; Goldsmith *et al*., 2013; Hernandez *et al*., 2015).

Prodiplosis longifila (commonly called *Prodiplosis* or *negrita* or *caracha*) is one of the main phytosanitary problems in the tomato (*Solanum lycopersicum* L.) and asparagus (*Asparagus officinalis* L.) crops (Geraud-Pouey *et al*., 2022; Hernandez *et al*., 2015; Castillo Valiente *et al*., 2020). This pest is a neotropical and polyphagous fly, considered to be of economic importance due to the serious damage it causes. It severely affects buds of Tahiti lime (*Citrus x latifolia* Tanaka ex Q.), (Jimenez, 2019), asparagus, various young shoots (asparagus, avocado, *Persea americana* Mill.), grapes (*Vitis vinifera* L.), leaves (tomato), inflorescences or flowers (tomato, Tahiti lime, asparagus), and vegetables (tomato), deforming them and limiting production (EPPO, 2017; Gagné, 1986; Hernandez *et al*., 2015). Adults of *P. longifila* are considered to be difficult to see in the field during the day, since they are active at dusk; females lay their eggs inside protected structures, such as in the buds of leaves and flowers, as well as under the calyx in tomato crops, or under the bracts or in flowers in asparagus (EPPO, 2017; Gagné, 1986; Hernandez *et al*., 2015).

The presence of *P. longifila* has been confirmed in Colombia, Ecuador, Peru, and the USA where larvae damage buds and tender tissues of asparagus, tomato, sweet pepper (*Capsicum annuum* L.), cucumber (*Cucumis* sp.), citrus, and other crops (Geraud-Pouey *et al*., 2022; Hernandez *et al*., 2015). In tomato crops in Colombia and Ecuador, damage from this pest is reported close to 100% (Geraud-Pouey *et al*., 2022). In Peru, the economic impact on potatoes and asparagus is about 16% (Cedano & Cubas, 2012; Kroschel *et al*., 2012). In the USA, the impact of this insect reaches up to 25%, when this plague was reported in the 1990's (Pena *et al*., 1987). There is a lack of recent information on its status. Due to the severe damage caused by this pest, the use of chemical pesticides has increased significantly, resulting in the selection of populations with a loss of sensitivity to the different pesticides used (Mujica & Kroschel, 2019).

Despite the growing interest in the cultivation of foliage and the significant potential impact of *P. longifila*, limited knowledge of this pest indicates the need to investigate key aspects of its biology and ecology, and to develop management alternatives that are based on scientific evidence. This research, conceived as a baseline, aimed to employ a set of tools within the DE, CS and data analysis approaches to identify and understand relevant trends about this pest in other hosts such as tomato or asparagus, and in foliage in particular. In this sense, our work focused on three main areas: academic trends through bibliometric analysis, trends in social networks such as X (previously Twitter) and YouTube, and information available on the Web under the Google trends platform. The perceptions of foliage producers in the production area of Cachipay, Cundinamarca, Colombia were explored using specifically designed surveys and direct communications in group chats. The present research defines potential alternatives for how these tools can be a solid basis for future research and control strategies for the floriculture and foliage production industry.

Materials and methods

Localization data origin

To analyze the tendencies of the foliage producers for aspects such as damage, importance, impact, management strategies, and others of *Prodiplosis*, we worked under the supervision and participation of the Economía Agrícola-Agropecuaria Villapard, and other farms located in the municipality of Cachipay in the department of Cundinamarca, Colombia (4°44’17.13” N, 74°25’37.68” W). Environmental conditions were in greenhouses under cover as well as in open fields of Ruscus and Cocculus foliage. Environmental conditions were semi-humid temperate conditions, with air temperatures ranging from 10°C -25°C, annual accumulated rainfall between 1300 mm and 2500 mm, and a bimodal distribution with dry periods in December – February and June – August. This research includes data collected from producers between the years 2018 (the time of the first pest reports) and 2023 (when the first part of the field work for this project was completed).

Identification of insects was carried out according to a protocol developed by the company. Identification concluded after a morphological and life cycle characterization that confirmed *P. longifila* as the pest population. This process was confirmed by the Museo Entomologico-UNAB of the Universidad Nacional de Colombia, Facultad de Ciencias Agrarias, Bogota, Colombia. As a complement and under the concept of pests of economic importance, the Institutio Colombiano Agropecuario (ICA) carried out the inspection, collected samples, and supported the identification of the pest. It should be noted that part of the producers’ work team, headed by the retired entomologist Professor Dario Corredor, sent photographs to the world expert Raymond J. Gagné, who stated that it was possible that it was the species *P. longifila*, but that the genus was under revision and that he was not committed to giving an accurate opinion. For the purpose of this research, we decided to consider this pest as *P. longifila*, but it is necessary that this be confirmed by taxonomic experts in the future.

Trends and perceptions of *Prodiplosis* in foliage crops and other hosts

A) Bibliometric

We carried out a primary analysis of scientific information in different bibliographic databases: Scopus(https://www.scopus.com/search/form.uri?display=basic#basic), PubMed (https://pubmed.ncbi.nlm.nih.gov/), and Google Scholar (https://scholar.google.com/?oi=gsb00&lookup=0&hl=es), as a basis for recognizing the state-of-the-art in research concerning *P. longifila*. For this objective, the search was limited to technical papers, theses, reports, and manuscripts published in English, Spanish, and Portuguese, and whose search equation was *Prodiplosis* which is identified as a common marker for the three selected languages. Some synonyms, or common names such as *negrita*, *caracha*, *caregato* were also included. The documents were then exported to BibTeX files, where the metadata contained authors, year of publication, journal or publication medium, country, author keywords, and abstract. The databases of the three bibliographic managers were merged, eliminating repeated results, and analyzed using the library pyBibx (Pereira *et al*., 2025), modifying some forms of visualization, executed in the Google Colab user environment under the Python programming language.

B) Web search and social networks (X and YouTube)

A detailed analysis was performed using various platforms, regardless of country, host, or other factors that might affect the usage of this type of tool (internet access, age, social status, academic strata, etc.) in order to identify and understand trends from the Web and social networks related to *Prodiplosis*. The platforms used for this analysis included X (previously known as Twitter) for the period in which information could be freely accessed (2011-2022), Google trends (2004-2024) and YouTube (2010-2024). The analysis focused on collecting search and published data on these platforms, covering a specific date range that allowed free use of the available data.

Initially, the occurrences of searches over time were evaluated, and the proportions of queries associated with different search algorithms were analyzed. From this information, trend curves were generated that illustrate fluctuations in the particular topic and relative searches over time. Subsequently, the text associated with these trends (comments, information, descriptions, among others) was extracted and integrated into the analysis, including content from indirectly linked pages that allowed for a more complete understanding of the topic. For the processing and analysis of these data, deep learning analysis based on natural language processing (NLP) tools were used (described in detail in the following section). We used the free software Python as the programming language, under the Google Colab user environment, specifically utilizing the sntwitter, googleapiclient, and pytrends libraries, which facilitated the extraction and systematic analysis of the information.

For the validation of the data associated with the search or count on social networks (X and YouTube) and the Web (Google trends) related to *Prodiplosis,* a historical database was used. This database was associated with reported cases of damage and the presence of the insect *Prodiplosis* from the Colombian phytosanitary service, the Plant Health Technical Directorate of the Instituto

Colombiano Agropecuario-ICA and monitoring data of the foliage production systems in Cundinamarca (n=50). It also included the different farmers dedicated to tomato production in Colombia that carry out monitoring as part of their agronomic practices (n=35, Antioquia, Cundinamarca, Boyacá, Caldas, Risaralda, Valle, Huila, Tolima, and Santander). For each social network or in Google trends, the historical series of available data in terms of searches, views, or tweets was used and integrated with the cases reported by the ICA in terms of both damage and presence of the insects associated with different hosts such as foliage and tomatoes.

Given that digital data presents several biases, both spatial and temporal, methodological adjustments were necessary to address these issues. Spatial biases arise from factors such as internet access, the location of the farmer, the technical assistant, or the individual conducting the search. Temporal bias is also significant since farmers tend to use sources like Google, videos, fellow producers, or technical advisors as their primary sources of information. These searches are often triggered by the ¨*identification of a phytosanitary problem*¨, knowledge about it, and the perceived economic impact. To reduce spatial bias, only areas relevant to tomato and foliage production in Colombia were considered, excluding unrelated queries that, while potentially real, could not be reliably validated. Temporal bias posed a greater challenge; thus, an approach was adopted to identify the statistically closest *lag* between online search behavior and actual field data. Techniques such as time lags, cross-correlation, and Granger causality tests were employed to assess the predictive or reactive nature of search activity about field observations

For the period of analysis, a series of lags were identified, which were associated with periods between 4 and 12 months and were adjusted for the digital data. Subsequently, a Spearman correlation analysis was performed, given the discrete origin of the variables. The confidence interval was calculated using the bootstrapping resampling method, and the results were visualized. The entire process was carried out in the Colab environment, utilizing the freely available programming language Python, ensuring a robust analysis of the relationship between online activity and reported agricultural cases.

C) Producer perceptions of *Prodiplosis* in foliage crops from surveys and use of WhatsApp social chat

Two methodological tools were used for this section. First, a semi-structured survey was designed and distributed to foliage producers in the Cachipay region and surrounding municipalities, inviting them to participate voluntarily. The survey was designed to gather farmer perceptions and management practices related to *Prodiplosis* in their cropping systems. It was structured into five thematic sections, each addressing specific aspects of the issue: i) Economic Perception and Impact (R1–R7). This section focused on understanding 'farmers' perceptions of the economic importance and impact of *Prodiplosis* on their crops, including yield losses, production costs, and overall farm profitability; ii) Use of Physical Control Measures (R8–R12). These questions addressed the use and perceived effectiveness of physical barriers such as protective curtains and traps in the management of *Prodiplosis*; iii) Population Dynamics and Environmental Factors (R13–R28). This block explored 'farmers' observations regarding *Prodiplosis* population levels and their potential associations with surrounding weeds, neighboring crops, and various agronomic practices such as fertilization, irrigation, and the presence of other pests; iv) Use of Inputs and Control Strategies (R29–R48). This section collected detailed information on pest control practices, including the use of bio-inputs and synthetic chemical products. Questions covered product names, dosages, timing and frequency of application (including time of day), as well as farmer perceptions of effectiveness and control success; v) Sociocultural and Farm Context (R49–R53). The final section captured background information on farmers, such as cultural practices, land tenure trends, educational level, access to technical assistance, and farm size. For more information, review the supplementary information 1.

In parallel, a group chat was accessed on the WhatsApp social network, in which farmers of foliage from the same region participated. The purpose of this study was to complement the information obtained in the survey and observe in real-time interactions and discussions related to the production of flowers and the situation associated with *Prodiplosis* during a period of analysis between 2018 and 2022. This group provided an additional platform for the exchange of knowledge and practices, allowing for the capture of qualitative data of interest.

In both cases, it was guaranteed that the ethical regulations related to the use of personal data were complied with. Informed consent was obtained from all participants, ensuring the confidentiality of sensitive information and the anonymity of the data collected. Furthermore, they were informed that the purpose of the study was strictly academic and details of how the data would be used in the context of the research were provided. These measures ensured that participation was conducted under rigorous ethical standards aligned with best research practices.

In the first phase, the survey was implemented through the Google Forms platform. This tool contained a total of 53 questions structured to evaluate different aspects associated with foliage and *Prodiplosis*. Of these, 32 questions were closed-ended and formulated based on a Likert-type response scale (strongly disagree, disagree, neutral or neither agree nor disagree, agree, and strongly agree) (Supplementary information 1). This allowed the measurement of the intensity of respondent perceptions and attitudes regarding various specific statements, enabling the quantification of opinions and a rigorous statistical analysis of the responses (Jebb *et al*., 2021; Joshi *et al*., 2015). Additionally, nine open-ended questions were included, designed to collect data for textual analysis (described in detail in Supplementary information 1). Seven dichotomous questions were formulated, with simple yes or no response options, focusing on capturing specific decisions and behaviors related to the study (Supplementary Information 1). The remaining seven questions were aimed at collecting personal and demographic information from the respondents. Once the data were collected, they were processed and analyzed using the likert library within the R programming environment version 4.4.2 (R Core Team, 2004), as well as other descriptive statistical methods.

Natural language processing (NLP) techniques were employed to analyze textual data obtained from the analysis of social networks and producer chats (Maulud *et al*., 2021; Tripathy *et al*., 2021). For this particular case, a neural network known as Skip-gram, which is a model that has shown promising results in NLP (Xiong *et al*., 2017), was retrained and implemented. We used words obtained from information found on X and pages linked to the tweet (6438 words, 40 tweets), titles of chats between farmers (19451 words, 1300 chats) and surveys (9359 words, 25 surveys). We then proceeded to optimize the parameters of the Skip-gram network. The World 2 Vec package was used for this process and the Skipgram network was executed in Python within the Google Colab user environment.

Design of a digital platform for connecting with producers in knowledge exchange processes

As a fundamental component to unify the concepts of digital epidemiology and citizen science, this study developed two interconnected digital applications. These applications were designed to perform partially similar yet complementary functions, facilitating data acquisition, visualization, and predictive modeling in plant pest monitoring (*Prodiplosis* case).

The first tool was a mobile application developed using an open cross-platform language (such as Dart) through the Flutter framework. This app was provided to the end users (a group of farmers participating in the trial). Five main modules were included: i) user module authentication and user profile management; ii) sensor configuration module settings for data acquisition from connected sensors; iii) data processing module graphical representation of collected and processed data; iv) georeferenced interactive map module visualization of spatial disease distribution and; v) disease forecasting and recommendations module integration of predictive models and disease management recommendations. The second tool was a web application developed using Python and the Flask framework. This platform includes the same core modules as the mobile application. But three additional features were introduced: i) API (Application Programming Interface) integration module facilitating access to external APIs for enhanced data analysis; ii) interactive forecast map module providing spatially explicit disease prediction visualizations and; iii) administrative dashboard module enabling real-time monitoring of measured variables. Communication between the platforms was established via a cloud-based infrastructure, utilizing a relational database and a virtual server. This setup enabled data storage, file hosting, and seamless information exchange between the Web and mobile applications, ensuring real-time synchronization and accessibility for end users.

Results

Bibliassociated with the current state of knowledge of *Prodiplosis* worldwide

The search of bibliographic databases yielded 957 documents of which 932 belonged to Google Scholar, 22 to Scopus, and 3 to PubMed, with 16 duplicates. The first document associated with this search dates to 1812, and no further papers were found until 1908. From 1987 onward there was a continuous increase in the number of papers published, culminating in a peak, in 2019 with 82 papers (Fig. 1A). Prior to this, 16 papers were reported. Most of these articles are associated with *P. longifila* in flowers or fruits of crops such as tomato, asparagus, chili, bell pepper, blueberry, and potato and encompass topics associated

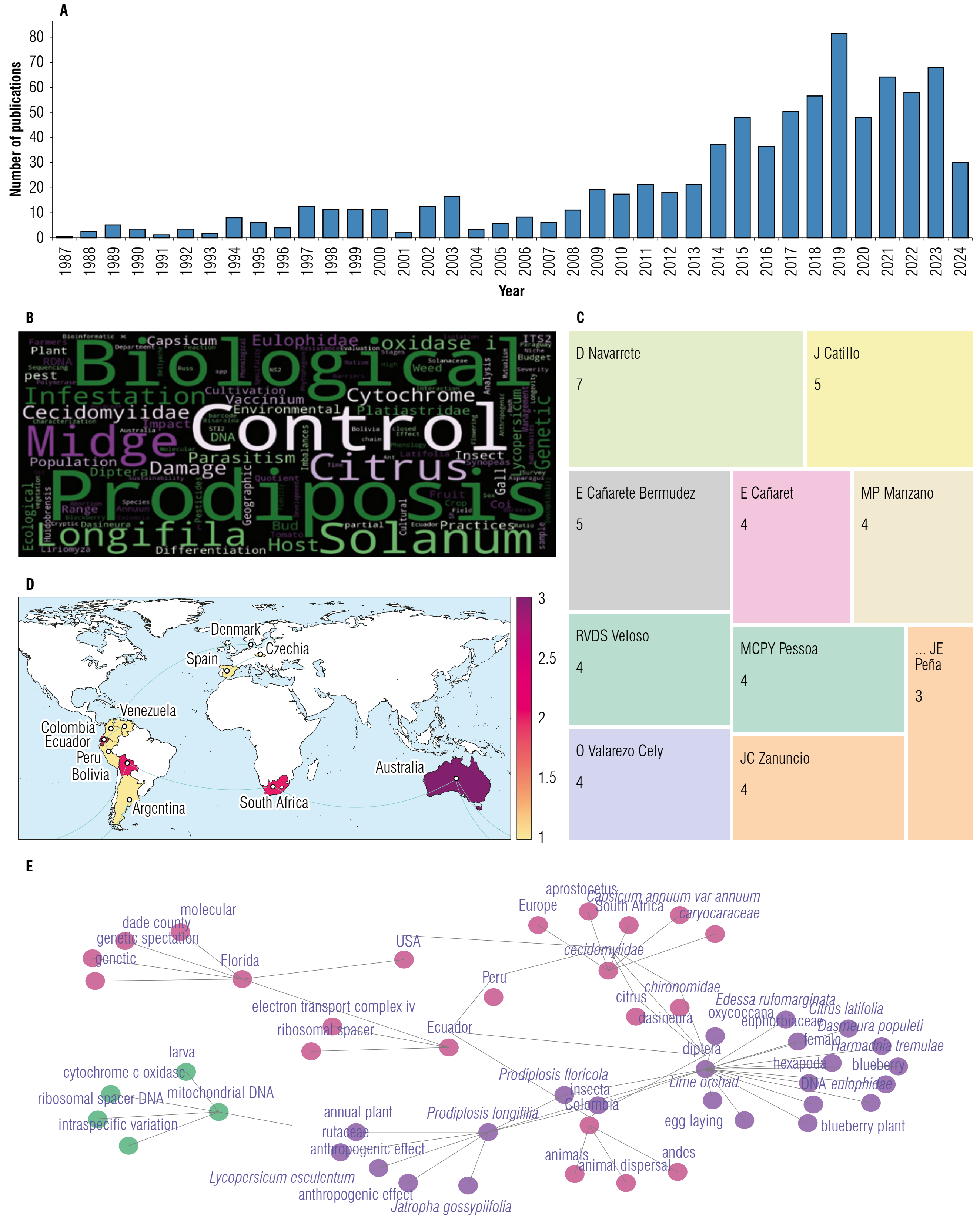

FIGURE 1. Meta-analysis of the scientific literature on *Prodiplosis*, based on a search query using the term "Prodiplosis" across the Scopus, PubMed, and Google Scholar databases. A-annual number of publications; B- most frequent terms extracted from abstracts; C-authors with the highest number of contributions; D-international collaboration network between countries; E-keyword co-occurrence clustering, illustrating thematic groupings within the literature.

with identification, distribution, damage, quality, yield, management, and genetics (Fig. 1B). Three main authors were highlighted (Fig. 1C). When analyzing the geographical relationships, 11 countries were identified, 6 of which are in Latin America (Fig. 1D), suggesting that this region is where this pest has the greatest impact.

When observing textual connectivity, 3 clusters were found. The first (green) is related to larval stages, neotropical regions, population structures, phylogenetics, quarantine measures, and molecular topics at the DNA and protein level (Fig. 1E). The second cluster (red) shows different subgups, specifically a relationship between Peru, Ecuador, and Florida, USA, with topics related to genetics, molecular biology, and the dipteran family Cecidomyiidae, associated with the host genus *Capsicum*. It also includes relationships between countries such as those involving South Africa, Europe, and the hymenopteran genus *Aprostocetus* and plants of the family Caryocaraceae. Slightly further away is the country of Colombia, which is closely tied to the Andes, animals, and to animal dispersal. Finally, cluster 3 (purple) shows three parts. The first of these is associated with *P. longifila,* which is linked to annual plants, Rutaceae, and anthropogenic effects, as well as plants such as *Jatropha gossypiifolia* L. and *S. esculentum*. The second is a bridge between the first and third groups and is associated with the class Insecta and *P. floricola*. Finally, the third subgroup is associated with Diptera, females, DNA, and some crops such as limes and blueberries (Fig. 1E).

Trends and perceptions of *Prodiplosis* using social media and web searches

The results obtained from the social network X revealed that the primary searches related to the term ¨Prodiplosis¨ were concentrated in Peru, Colombia, and, to a lesser extent, Ecuador (Fig. 2A). In both English and Spanish, the search trends showed peaks between 2016 and 2018. There has been a growing trend of Spanish-language tweets from 2016 onward, suggesting that this pest started to become a problem in Andean countries around this time (Fig. 2B). Furthermore, a word cloud analysis revealed that the most recurring terms were the species name, hosts (tomato, citrus, asparagus, among others), pest, control and life cycle, indicating that in recent years this insect has gained significant economic importance (Fig. 2C).

Similarly, web searches (Google Trends), which also include social media platforms such as YouTube, reflect findings like those of X. In this data source, *Prodiplosis* appears to be a pest of high importance in the previously mentioned Andean countries (Fig. 3A). In Colombia, the highest search volumes were associated with the departments of Cauca, Nariño, Cundinamarca, Huila, and the coffee growing region (Fig. 3A), aligning with the most productive areas for its well-known host, table tomatoes. Meanwhile, in Cundinamarca, the searches are linked to their potential impact on foliage and the threat to the floriculture sector. Temporal trends in Google searches reveal a significant peak in 2004, coinciding with reports of this pest in Peru, Ecuador, and Colombia. Furthermore, there has been a notable increase in searches since 2022, corresponding to a growing impact of this pest on host crops (Fig. 3B). Similarly, in the social network X, search terms in both English and Spanish focus on keywords such as name, control, populations, and hosts (Fig. 3C).

On YouTube, the highest number of views related to *Prodiplosis* occurred after 2019 in countries such as Peru, Colombia, and Ecuador (Fig. 4A and B). On this platform, videos related to integrated pest management, biological control, and *Prodiplosis* in tomatoes stand out, with most of the content in Spanish (Fig. 4C). Key terms frequently associated with this pest on YouTube include host plants, control methods, common and scientific names of the pest, presence in specific countries, and management practices (Fig. 4D).

Associated validation between social media searches, the Web and actual pest presence and reporting data

In the case of the validation of searches on social networks and the Web about the cases and damage to *Prodiplosis* reported by the ICA, a considerable potential for social networks and the Web is revealed to indirectly monitor phytosanitary problems in the agricultural sector (Fig. 5). According to the r metrics and the confidence interval, they indicate that the social network X, YouTube, and Google Trends presented values of 0.42, 0.51, and 0.73, respectively, based on the correlations observed within the lower and upper limits (Fig. 4A, B and C). These values suggested that, although all platforms have a degree of association with the *Prodiplosis* reports, Google Trends had the strongest relationship, indicating its greater effectiveness in reflecting actual infestation trends.

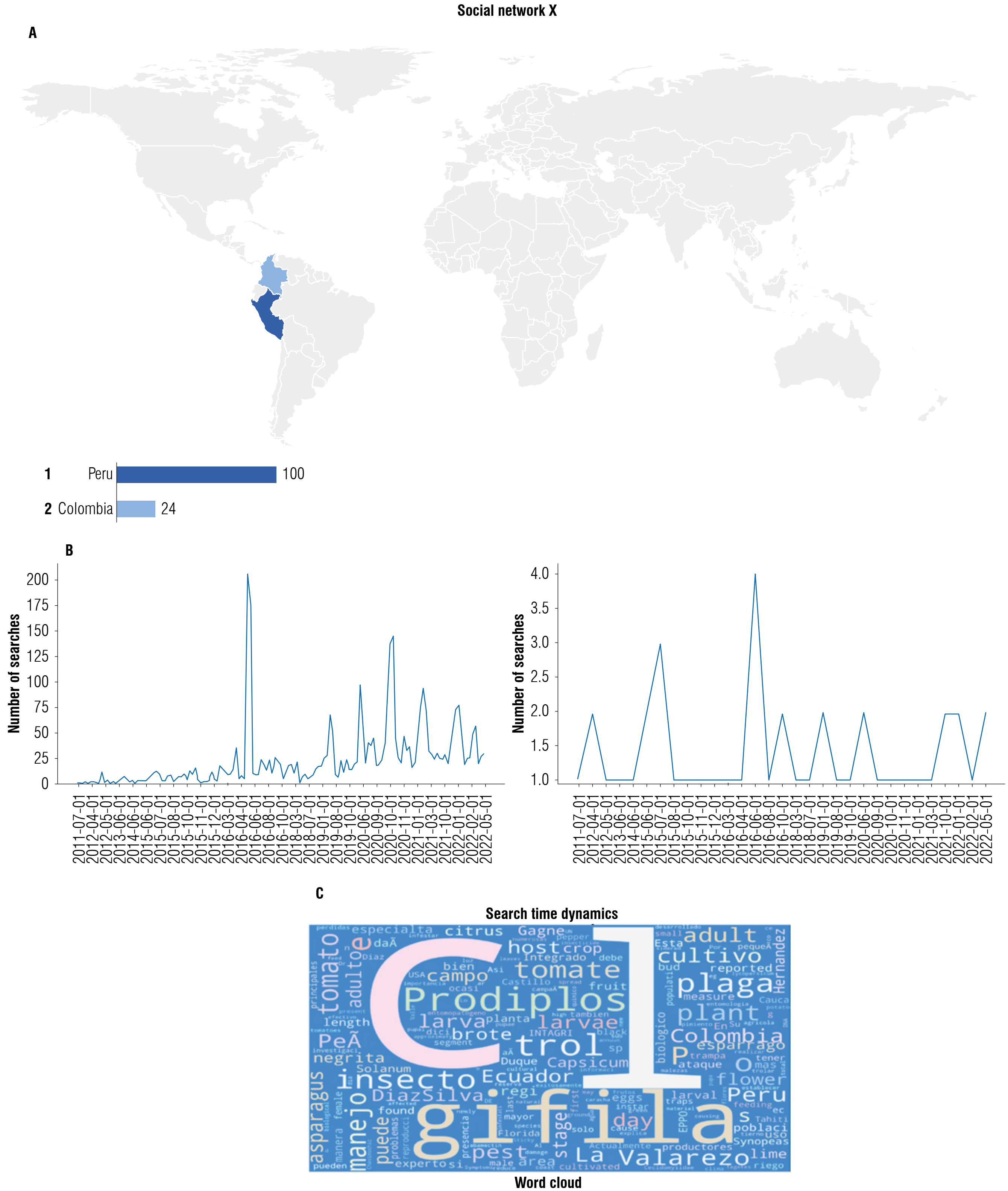

FIGURE 2. Aspects associated with trends in the social network X, the Web and YouTube related to aspects of *Prodiplosis* worldwide and in different hosts. A to C- analysis in social network X (previously Twitter). A-countries with most *Prodiplosis* publications; B-search time dynamic; C-world-cloud or related terms with *Prodiplosis*.

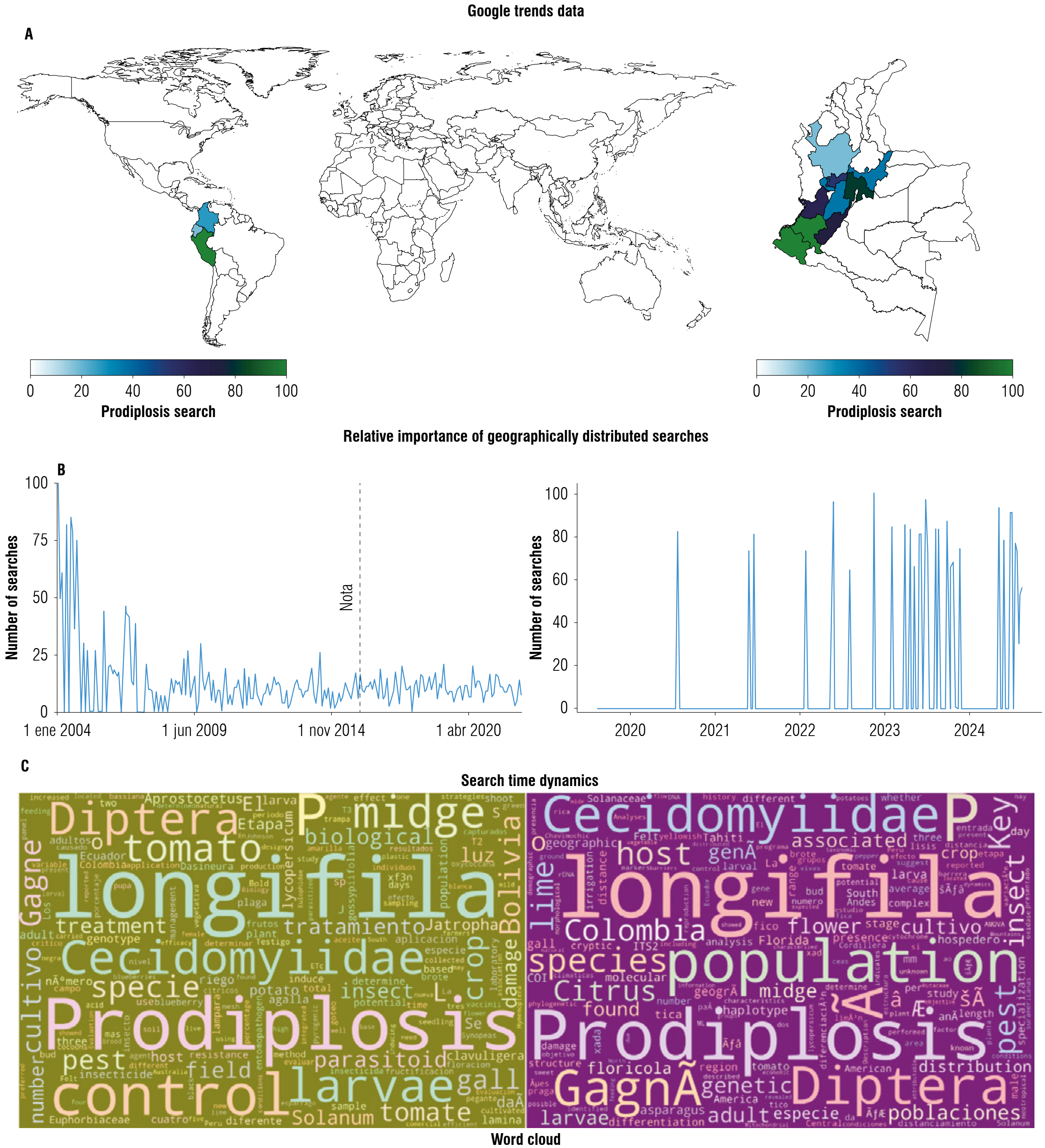

FIGURE 3. Aspects associated with trends in the social network X, the Web and YouTube related to aspects of *Prodiplosis* worldwide and in different hosts. A to C-analysis on the Web (Google trends); A-countries where people searched about *Prodiplosis* worldwide and in Colombia (right); B-search time dynamic worldwide and in Colombia (right); C-world cloud of related terms with *Prodiplosis* searches in worldwide.

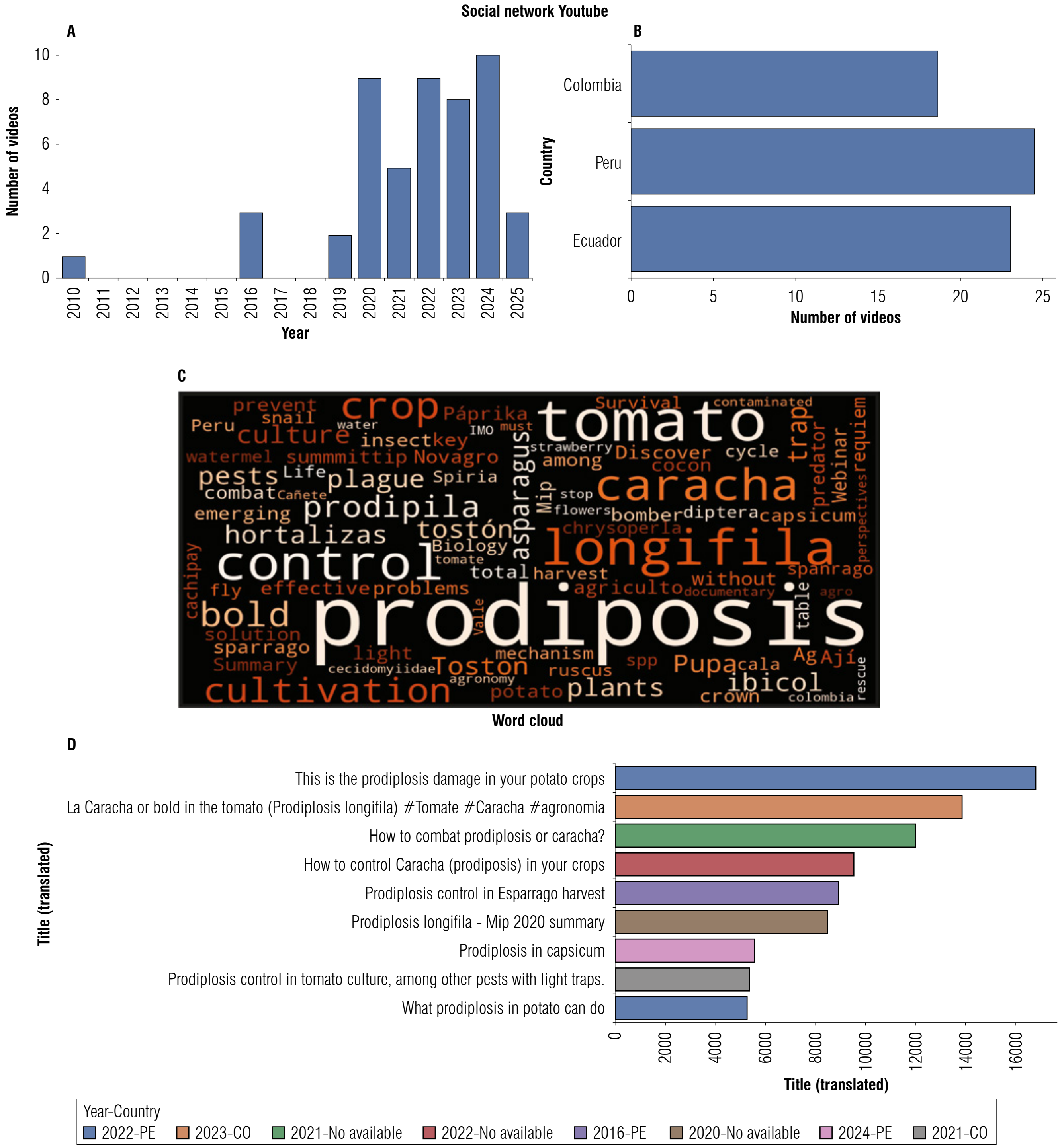

FIGURE 4. Aspects associated with trends in the social network X, the Web and YouTube related to aspects of *Prodiplosis* worldwide and in different hosts. A to D-YouTube social network analysis; A-number of videos published by year; B-countries where videos were published; C-world cloud of related terms of *Prodiplosis* videos; D-year and country of the most viewed videos, CO - Colombia, PE - Peru, EC - Ecuador.

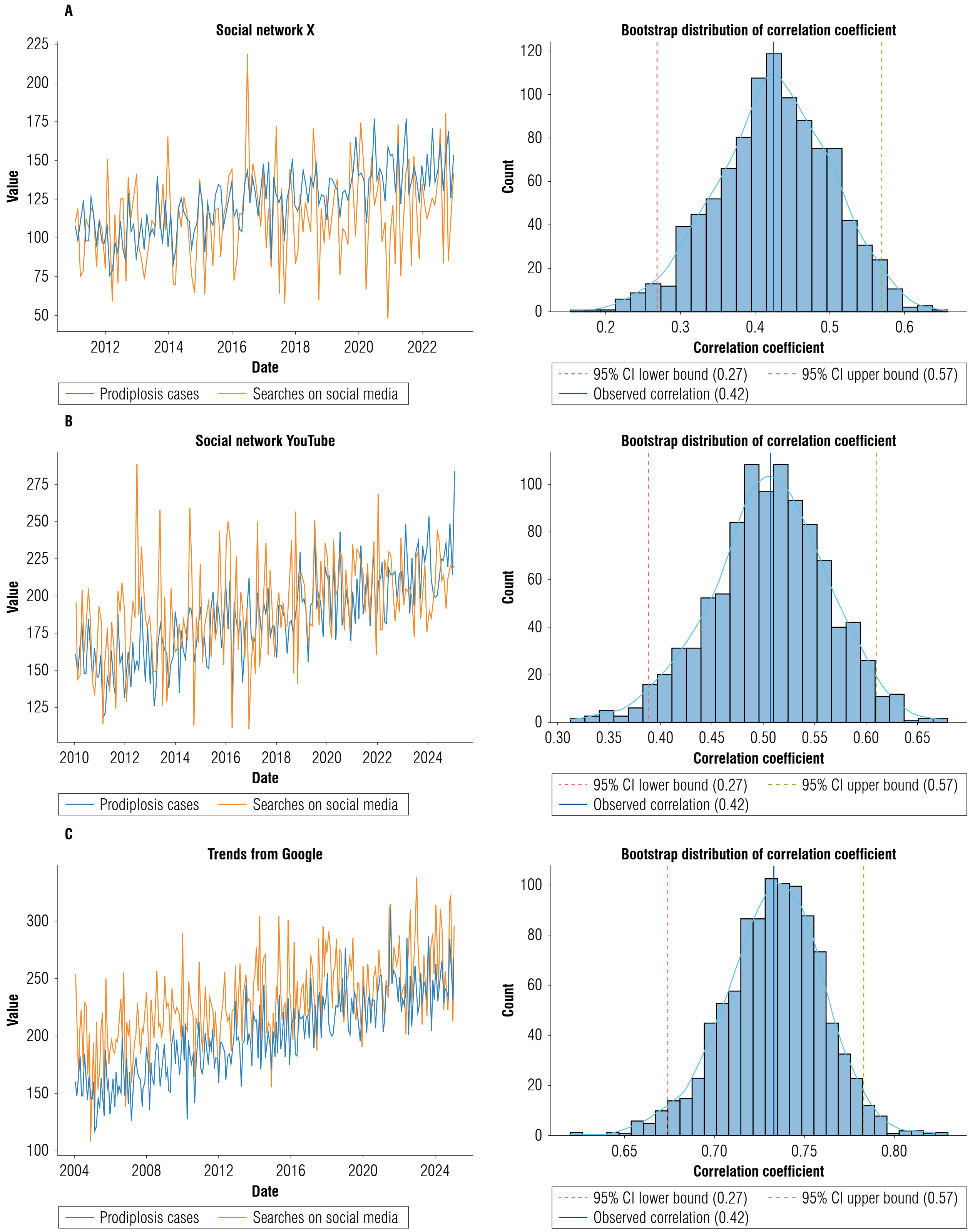

FIGURE 5. Validation of digital data sources related to *Prodiplosis* through comparison with actual reports from farmers and national phytosanitary systems in Colombia. A-analysis of posts and interactions on social network X (formerly Twitter), identifying geographic origin, frequency, and thematic content related to *Prodiplosis*; B-examination of YouTube content, including the number of views, user engagement, and the nature of videos referencing *Prodiplosis* or its damage on crops; C-trends from Google search queries, showing temporal dynamics and peaks of public interest at the national level.

Producer perceptions of the effect of *Prodiplosis* on foliage crops from surveys and the use of WhatsApp social chat

Analysis of trends observed in surveys (n = 25) revealed that more than 50% of *Ruscus* and *Cocculus* growers reported an increase in *Prodiplosis* populations since its emergence as a pest and increased production costs. However, growers still considered production to be profitable. Regarding spatial distributions, approximately 90% of the respondents confirmed that their neighbors also faced this pest and shared common weeds. This exacerbates the situation. However, only 47% perceived the management measures implemented as effective, although most agreed that the pest was related to technical shortcomings such as greenhouse design, changing climatic conditions, poor fertilization plans, and lack of monitoring tools (Fig. 6A and B). Most implemented curtains in greenhouses, with a perceived efficiency of 80%, while color (yellow and blue, which are the most abundant, not based on any evaluated criteria, but only on the recommendation of the ICA.) traps

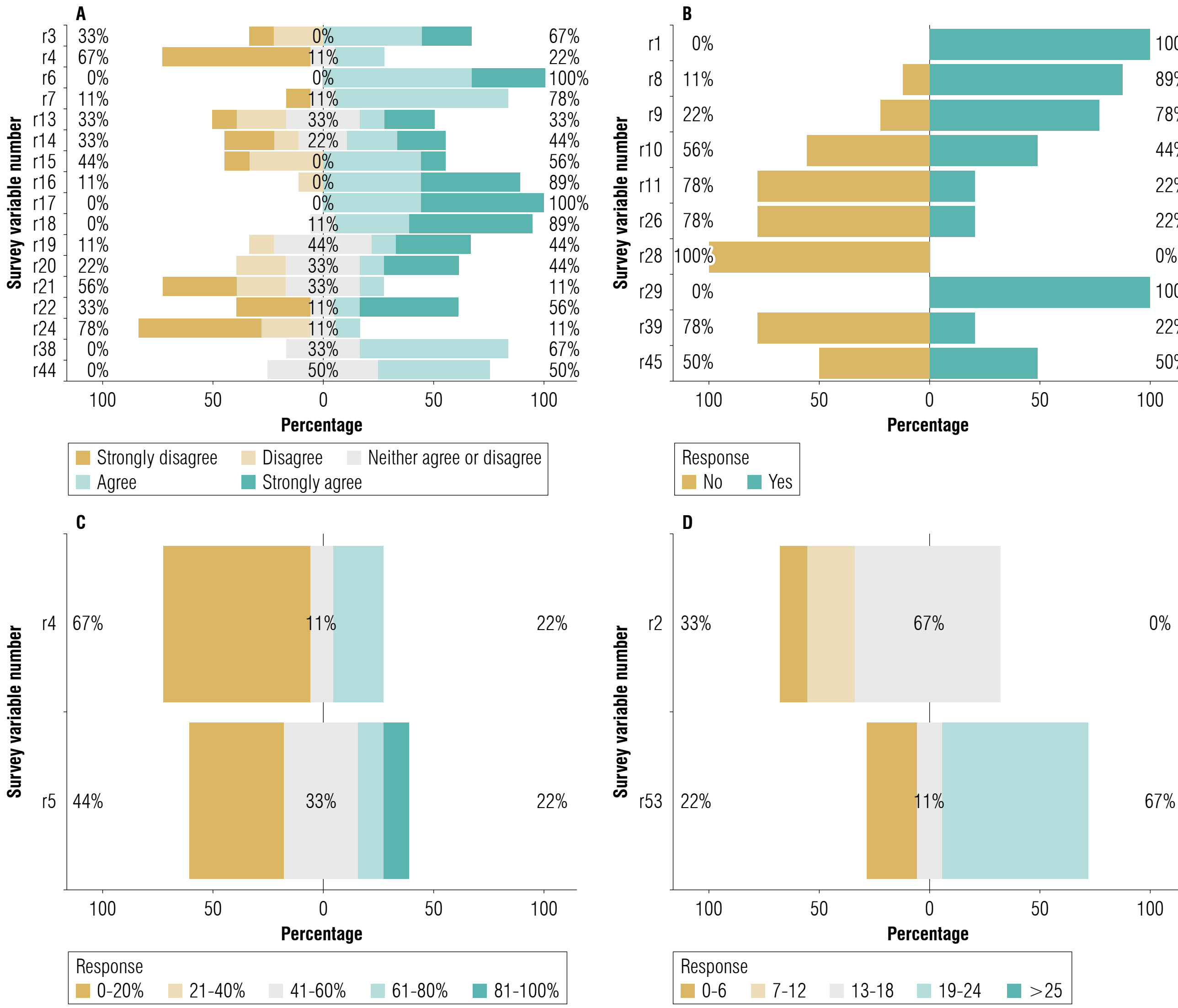

FIGURE 6. Summary of producer perceptions, management practices, and contextual variables related to *Prodiplosis*. in foliage crops from Cundinamarca, Colombia. A-general perception of *Prodiplosis*. And its relationship with *Ruscus* y *Cocculus* foliage based on the Likert scale; B-management carried out for the control of *Prodiplosis* based on dichotomous response questions; C-magnitude of the problem caused by *Prodiplosis* on the foliage crop; D-time since you consider the pest appeared in the crops (upper) based on the time you have been working (lower). *Prodiplosis* in Cachipay, Cundinamarca Colombia. r1–r7 includes economic impact of Prodiplosis, including yield losses and profitability. r8–r12 use and effectiveness of physical control methods. r13–r28 relationships between Prodiplosis populations and factors like weeds, neighboring crops, fertilization, irrigation, and other pests. r29–r48 pest management strategies, including bio-inputs and chemical controls covering product types, application practices, and perceived effectiveness. r49–r53 sociocultural and farm-related data, such as land tenure, education, technical assistance, and farm size. For more details, refer to Supplementary Information 1.

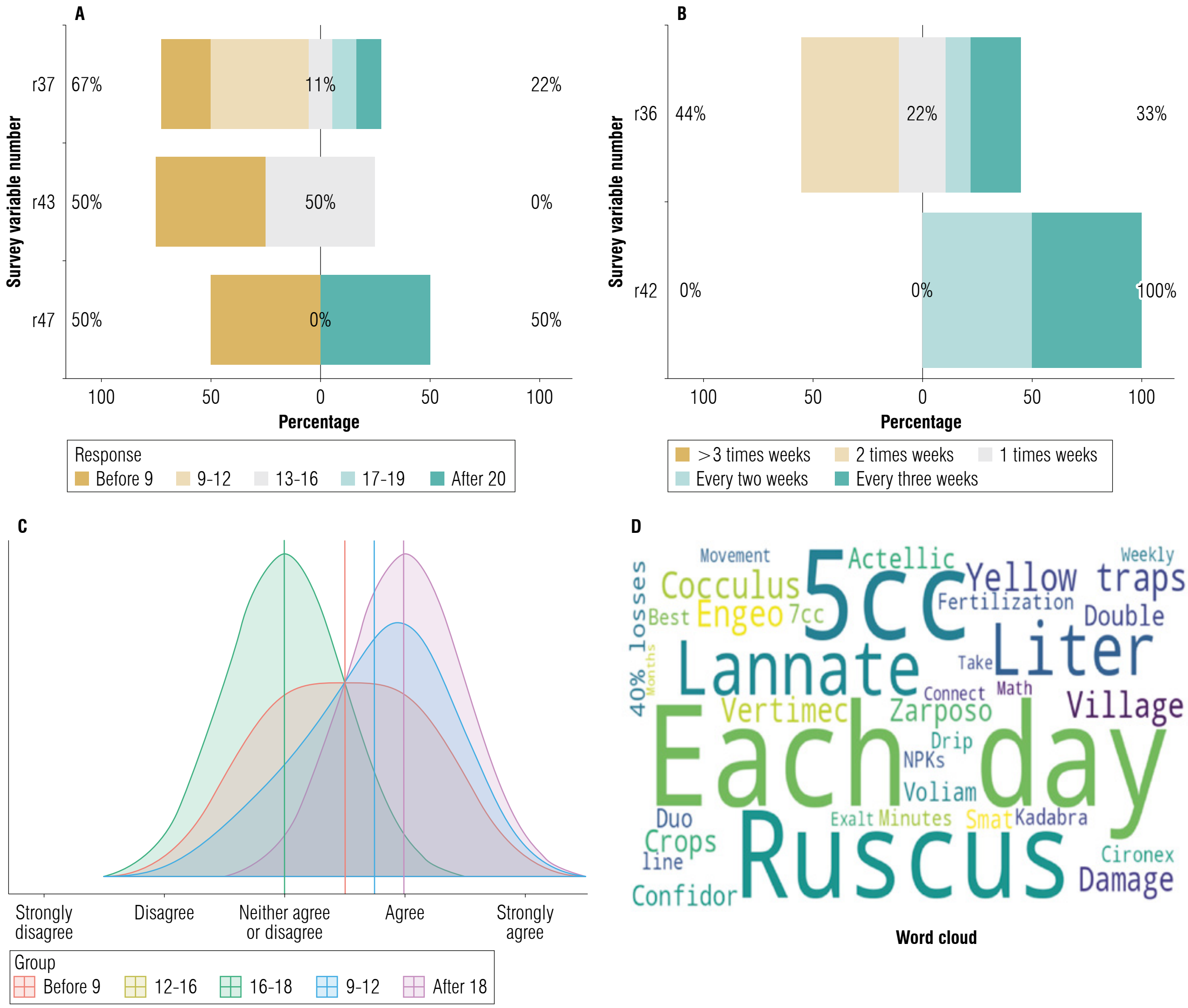

FIGURE 7. Summary of producer perceptions, management practices, and contextual variables related to *Prodiplosis*. in foliage crops from Cundinamarca, Colombia. A-time of application of chemical and biological treatments and bio inputs on the farms surveyed; B-frequency of application of chemical and biological products for the control of *Prodiplosis* in Cachipay, Cundinamarca, Colombia; C-perception of success in controlling *Prodiplosis* according to the time of application of chemicals; D-word cloud associated with the perception of producers associated with the use of the survey and WhatsApp chat in foliage producers in Cundinamarca. r29–r48 pest management strategies, including bio-inputs and chemical controls covering product types, application practices, and perceived effectiveness. For more details, refer to Supplementary Information 1.

for monitoring and biological control presented lower levels of acceptance and efficiency (Fig. 6A and B).

Regarding economic losses and affected areas, more than 50% of the respondents experienced losses between 20-40% (understood as the amount of foliage with the presence of larvae or damage that makes the foliage unsuitable for sale at market quality), although some farms showed up to 100% of the field was damaged (without loss, as they can make some process of damage removal or cosmetic appearance that improves the visual quality and can be sold, with a potential value request, depending on the degree of damage) (Fig. 6C and D). In terms of application frequency, chemical pesticides were applied more than three times a week in about 50% of farms. In contrast, biological applications were less frequent, once a month or every two weeks (Fig. 7B). Application times varied between producers, with chemical applications carried out mostly between 9 am and 12 m, and biological applications between 9 am and 4 pm (Fig.7A and C). The results associated with the use, impact and effect of the products used for the management of *Prodiplosis* suggested that the perception by farmers of success of chemical control was related to applications made after 6 pm each 2 or 3 weeks, coinciding with adult *Prodiplosis* activity (Fig. 7C and B). Most of the producers pointed out that the pest appeared since 2018 and colonized the foliage production systems according to the specific conditions of each farm, generating a high risk in the plantation, according to perceptions of damage, presence, impact, without being able to define a specific and forceful pattern (Fig. 6D). Meanwhile, the word cloud mainly was associated with the use and dosage of chemical products, to a lesser extent with aspects of traps related to color, and to a lesser extent yet with other management measures (Fig. 7D).

Design of a digital platform for connecting with producers in knowledge exchange processes

The development of two integrated digital platforms, a mobile application for field data collection and a web-based visualization and analysis interface, successfully fulfilled the objective of establishing a connection between foliage farmers and the academic community (Fig. 8). This integration served as a bridge, linking the scholarly perception of *Prodiplosis* related issues with the firsthand experiences

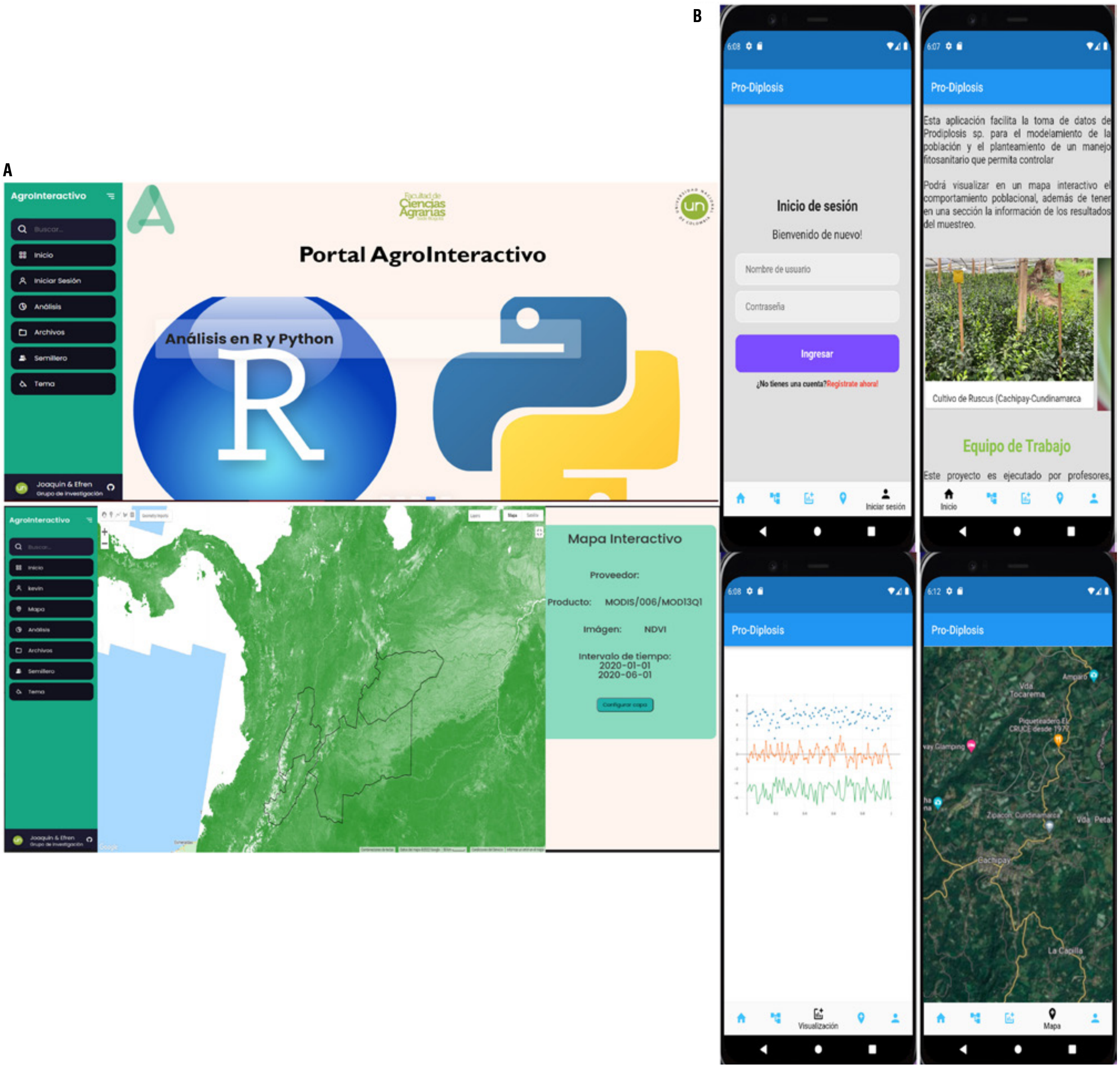

FIGURE 8. Visualization of digital Tools developed for connecting digital epidemiology and citizen science (in Spanish) in a case study on *Prodiplosis* population trends and dynamics. A- "Portal Agrointeractivo" – Web application - A data visualization and analysis platform that integrates interactive maps, predictive models, and external API connections, facilitating real-time monitoring and decision-making. B-¨Pro-Diplosis¨ – Mobile application- A field data collection tool that allows farmers and field technicians to record observations, configure sensor data acquisition, and visualize disease distribution through an interactive map.

of farmers in their fields. By facilitating this exchange, the platform enabled a more comprehensive understanding of the problem from both perspectives. Additionally, the platform provided a space for discussion and knowledge-sharing among farmers, serving as a foundation for the transmission and transfer of empirical knowledge. The system allowed real-time access to critical data related to the target problem through a bidirectional approach. Data were collected directly by farmers (citizen science), while digital data sources (digital epidemiology) extracted from various online platforms contributed to the analysis. This dual approach fostered interaction between the two data streams, enabling a validation process that enhanced the credibility and applicability of the collected information (Fig. 8).

Moreover, this study highlights the multiple challenges associated with integrating digital epidemiology and citizen science in agricultural settings. Among the most significant hurdles were data bases and the adoption of digital platforms, which impacted data quality and user engagement. These findings underscore the need for refined methodologies to mitigate biases and enhance user participation, ensuring more reliable and actionable insights in digital plant health monitoring.

Discussion

Our work demonstrated how the monitoring of web searches and social networks, bibliometric analysis, and further analysis of the information using data science tools (specifically visualization and advanced analytics elements such as the use of NLP as a deep learning algorithm) showed a high capacity to process and find patterns in textual data sources (Maulud *et al.*, 2021; Tripathy *et al.*, 2021). NLP is a valuable method for identifying empirical patterns in the behavior of pest populations, specifically at the agricultural level, under the focus of this study. These freely available tools allowed for a comprehensive view of the importance and geographical distribution of these pests. Focusing on the species *P. longifila*, we highlighted the geographic and thematic interest of the related queries, reflecting a particular interest in certain regions, possibly associated with its impact on crops of regional importance. In addition, we provided a detailed overview of search and discussion trends on digital platforms, proposing this methodology not only as an epidemiological monitoring strategy but also as a way to share information, convey messages, and develop more efficient management strategies.

Implementing a digital epidemiology and citizen science approach in agriculture could transform pest management by facilitating communication and connection between farmers and other actors such as researchers and agricultural technicians. Social networks, for example, can be used for rapid dissemination of alerts on pest outbreaks. At the same time web platforms can centralize geolocated data on infestations, facilitating the identification of patterns and trends of information (Charles-Smith *et al.*, 2015; Samaras *et al.*, 2020). Additionally, bibliometric analyses of the academic literature on specific pests can provide comprehensive overviews of existing research, guiding evidence-based management strategies and identifying knowledge gaps (Merigó & Yang, 2017).

In this sense, our work identified multiple gaps in the knowledge associated with *Prodiplosis* in foliage crops, particularly in aspects of biology, ecology, population dynamics, and management, at the level of knowledge accumulated in other hosts such as tomato, asparagus, and others. This work is also essential considering the scarcity of relevant information for green foliage crops. This lack of information aggravates the situation, given that farmers, as demonstrated in this work, are mostly focused on chemical control, very much associated with replicating what has been done in other crops with little or no success. This leads to the multiplication of errors and the abuse of chemical control.

The implementation of digital platforms successfully integrated citizen science and digital epidemiology but also revealed key challenges affecting data reliability and adoption. A significant issue was data quality control, as farmers lacked standardized protocols for measurement and monitoring. Despite the valuable practical experience possessed by farmers in the field of crop management, the study identified some common challenges associated with the morphological characterization of *Prodiplosis*. The presence of insects with similar morphology can make accurate identification difficult, even for experienced farmers, occasionally leading to misidentifications. Furthermore, the limited utilization of structured data management techniques, such as recording, storage, and retrieval of pest observations by growers, has led to the emergence of challenges in data management. Technical barriers such as poor internet access and low digital literacy further limited engagement with digital tools.

One of the primary challenges in the agricultural sector is the adoption and correct perception of technologies by farmers, who often prefer simple, low-cost, and accessible methods (Cáceres-Zambrano *et al.*, 2023; Tey & Brindal, 2012). In this context, we implemented an effective communication strategy using WhatsApp, a widely used platform in today's society with multiple advantages in communication, message learning exchange, and participatory monitoring, among others (Agnese *et al.*, 2024; Nain *et al.*, 2019). This tool facilitated communication, idea exchanges, and reporting on the impact, costs, and spread of the significant pest *Prodiplosis* in foliage crops, especially given the limited technical support available in the literature for designing integrated management strategies for this impactful pest. In addition, the provision of technical assistance from governmental or private entities is necessary to facilitate the reduction of existing gaps in identification and management.

Our approach achieved significant progress in enhancing information communication strategies and accessing real-time data, sharing both successful and unsuccessful experiences. This provided a regional approach to addressing a problem that directly affects farmers in the foliage production area of Cachipay municipality. The advantage of this chat group lies in its multidisciplinary nature, including farmers, technical assistants, export managers, and academia. Subsequently, to ensure the correct use and transmission of information, the data were analyzed using data science tools. This process enabled us to discard inconsistent or unverified information thereby facilitating the identification of patterns in the textual data through natural language processing analysis, which served as a potent instrument for the examination of ingestion of textual data (Choudhury *et al.*, 2018; Khurana *et al.*, 2023). This comprehensive approach improved the management of *Prodiplosis*, optimizing communication and decision-making in the agricultural sector.

The application of digital epidemiology tools in agriculture represents a significant advance in pest management. However, it also brings several challenges. One of the primary challenges is the automation of queries and data downloads, which requires the development of efficient algorithms to process large volumes of information in real-time. In addition, it is essential to have rapid modeling tools capable of generating early warnings of possible outbreaks, enabling a timely and effective response. In this regard, communication of the information obtained is another critical aspect, as the channels must be accessible and understandable to farmers, who often lack an adequate level of familiarity with new technologies.

To ensure the effectiveness of these digital tools, it is essential to have validation systems in place that use real field data, guaranteeing that predictions and recommendations are accurate and applicable in practice. Finally, access to the internet and the adoption of new technologies by farmers are crucial factors for the success of digital epidemiology. Without adequate connectivity and willingness to use these innovations, their implementation and effectiveness in the field could be severely limited. In addition, to enhance participation and improve data reliability, capacity-building initiatives are necessary, including training programs and user-friendly digital tools.

Addressing these limitations is crucial for ensuring the long-term success and scalability of digital epidemiology and citizen science in agriculture. Addressing spatial and temporal bias in digital data remains a critical challenge for improving the reliability of pest forecasting models. These biases can limit the accuracy of analyses, particularly when aiming to scale down to finer spatial resolutions necessary for localized decision-making. Spatial inconsistencies driven by uneven internet access and user location combined with time lags in information-seeking behavior reduce the precision of digital epidemiology tools. Overcoming these limitations requires robust frameworks that integrate citizen science strategies with structured data validation protocols. Encouraging participation from producers, technical advisors, and local institutions can enhance data quality and representativeness across key production zones. Citizen science, when properly implemented, offers a valuable avenue to complement digital trace data, providing contextual insight and ground-truthing capacity. Future efforts should prioritize the development of participatory systems that merge digital tools with on-the-ground validation, enabling more accurate and scalable models for agricultural risk management under dynamic environmental and socio-economic conditions.

Conclusions

Integration of digital epidemiological and citizen science tools within agricultural systems offers significant advances in understanding phytosanitary challenges. By enabling indirect monitoring and identifying searching trends at a regional scale, these tools provide valuable insights into the biology, ecology, and management strategies of pests. In the case of emerging pests such as *Prodiplosis*, the application of digital epidemiology permitted a more accurate and comprehensive approximation of real-time phytosanitary status, helping to delineate the pest's distribution, impact and emerging threats across diverse cropping systems.

In addition, we have successfully identified the extent of damage this emerging pest inflicts on foliage crops and its consequent impact on production systems, particularly in terms of economic losses. While this represents an initial approximation, it underscores the critical need for continued research into the life cycle and ecology of this pest as well as the influence of environmental factors on its population dynamics. Such efforts are crucial for development more effective management strategies to minimize both crop losses and production costs.

Acknowledgments

The authors would like to extend special thanks to all the foliage crop farmers involved in this study, whose knowledge, kindness, and willingness to help made it possible to successfully carry out this pilot test. We also express our gratitude to the students and professors of the course ¨Management Decisions: A Phytosanitary Approach¨ offered in the Graduate Program of the Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogotá. Their active participation in logistical organization and contributions of ideas were essential for validating fieldwork related to population trends. A very special thanks to Kevin Steven Quiroga and Camilo Borda, the developers who contributed their ideas and expertise in creating the digital prototypes implemented in this study as part of the pilot test. Lastly, we extend our sincere gratitude to Gabriel Alejandro Quinche, who, as a student at the time and now a professional in Computer Science, provided invaluable support in developing the code for Natural Language Processing analysis in this study. We also extend our gratitude to Professor Emeritus Darío Corredor for his efforts in establishing connections with the farmers in the study area, his in-depth discussions on this pest, and his motivation to develop this pilot test, which focuses on a pest of significant agricultural importance.

Conflict of interest statement

The authors declare that there is no conflict of interests regarding the publication of this article.

Author's contributions

LAVG: conceptualization, research, writing - original draft, visualization, writing, and editing. HJVM: conceptualization, writing - original draft, and editing. LMPR: conceptualization, writing - original draft, and editing JGRG: conceptualization, writing - original draft, and editing. All authors have read and approved the final version of the manuscript.

Supplementary information

Supplementary Information 1 can be consulted in the publicly available repository: https://github.com/agrocompuepidemlab/Citizen-science-and-digital-data-to-trend-analysis-and-impact-assessment-of-Prodiplosis

Literature cited

Adebayo, I. A., Pam, V. K., Arsad, H., & Samian, M. R. (2020). *The global floriculture industry: Status and future prospects*. In K. R. Hakeem (Ed.), *The global floriculture industry* (pp. 1–14). Apple Academic Press. https://api.pageplace.de/preview/DT0400.9781000751383_A40408500/preview-9781000751383_A40408500.pdf

Agnese, F., Othman, Z., Mitin, A., & Wan Yahaya, W. A. J. (2024). Participatory monitoring in farmer field school program through Whatsapp among indigenous farmers in rural Sarawak, Malaysia. *Interactive Learning Environments, 32*(9), 5699–5710. https://doi.org/10.1080/10494820.2023.2223241

Cáceres-Zambrano, J., Ramírez-Gil, J. G., & Barrios, D. (2023). Factors associated with the adoption of technologies for avocado production systems. *Agronomía Colombiana, 41*(3), Article 110579. https://doi.org/10.15446/agron.colomb.v41n3.110579

Carney, R. M., Mapes, C., Low, R. D., Long, A., Bowser, A., Durieux, D., Rivera, K., Dekramanjian, B., Bartumeus, F., Guerrero, D., Seltzer, C. E., Azam, F., Chellappan, S., & Palmer, J. R. B. (2022). Integrating global citizen science platforms to enable next-generation surveillance of invasive and vector mosquitoes. *Insects, 13*(8), Article 675. https://doi.org/10.3390/insects13080675

Castillo Valiente, J. R., Rodríguez Quispe, S. P., Apaza Tapia, W. E. Julca-Otiniano, A. M., Canto Saenz, M. A. , & Rosales Sanchez, T. (2020). *Prodiplosis longifila* Gagné (Diptera: Cecidomyiidae) in asparagus (*Asparagus officinalis*) crop in Chavimochic Irrigation Project. *Peruvian Journal of Agronomy, 4*(3), 75–81. https://doi.org/10.21704/pja.v4i3.1645

Cedano, C., & Cubas, P. (2012). *Baeuveria bassiana* (Bals.) Vuill. y *Metarhizium anisopliae* (Metsch.) Sorokin para el control de pupas de *Prodiplosis longifila* Gagné en el cultivo de espárrago. *Scientia Agropecuaria, 3*(1), 29–34. https://doi.org/10.17268/sci.agropecu.2012.01.04

Cervellin, G., Comelli, I., & Lippi, G. (2017). Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *Journal of Epidemiology and Global Health, 7*(3), 185–189. https://doi.org/10.1016/j.jegh.2017.06.001

Charles-Smith, L. E., Reynolds, T. L., Cameron, M. A., Conway, M., Lau, E. H. Y., Olsen, J. M., Pavlin, J. A., Shigematsu, M., Streichert, L. C., Suda, K. J., & Corley, C. D. (2015). Using social media for actionable disease surveillance and outbreak management: A systematic literature review. *PLoS ONE, 10*(10), Article e0139701. https://doi.org/10.1371/journal.pone.0139701

Choudhury, R. A., Mahaffee, W. F., McRoberts, N., & Gubler, W. D. (2018). Modeling uncertainty in grapevine powdery mildew epidemiology using fuzzy logic. *bioRxiv*, Article 264622. https://doi.org/10.1101/264622 de Sherbinin, A., Bowser, A., Chuang, T.-R., Cooper, C., Danielsen, F., Edmunds, R., Elias, P., Faustman, E., Hultquist, C., Mondardini, R., Popescu, I., Shonowo, A., & Sivakumar, K. (2021). The critical importance of citizen science data. *Frontiers in Climate, 3*, Article 650760. https://doi.org/10.3389/fclim.2021.650760

Ekman, A., & Litton, J.-E. (2007). New times, new needs; e-epidemiology. *European Journal of Epidemiology, 22*(5), 285–292. https://doi.org/10.1007/s10654-007-9119-0

EPPO. (2017). *Pest risk analysis for Prodiplosis longifila*. EPPO. http://www.eppo.int/QUARANTINE/Pest_Risk_Analysis/PRA_intro.htm

EPPO. (2025). *Prodiplosis longifila (PRDILO) [Categorization]*. EPPO Global Database. https://gd.eppo.int/taxon/PRDILO/categorization

Faust, J. E., & Dole, J. M. (2021). *The global cut flower and foliage marketplace*. https://doi.org/10.1079/9781789247602.0001

Fraisl, D., Hager, G., Bedessem, B., Gold, M., Hsing, P.-Y., Danielsen, F., Hitchcock, C. B., Hulbert, J. M., Piera, J., Spiers, H., Thiel, M., & Haklay, M. (2022). Citizen science in environmental and ecological sciences. *Nature Reviews Methods Primers, 2*(1), 1–20. https://doi.org/10.1038/s43586-022-00144-4

Fulk, A., Romero-Alvarez, D., Abu-Saymeh, Q., Onge, J. M. S., Peterson, A. T., & Agusto, F. B. (2022). Using Google Health Trends to investigate COVID-19 incidence in Africa. *PLOS ONE, 17*(6), Article e0269573. https://doi.org/10.1371/journal.pone.0269573

Gagné, R. J. (1986). Revision of *Prodiplosis* (Diptera: Cecidomyiidae) with descriptions of three new species. *Annals of the Entomological Society of America, 79*(1), 235–245. https://doi.org/10.1093/aesa/79.1.235

Geraud-Pouey, F., Garces, A., Contreras, N., & Geraud-Chirinos, J. E. (2022). *Prodiplosis longifila* (Diptera: Cecidomyiidae), evolución como plaga y un método para evaluar sus poblaciones en tomate. *Revista Colombiana de Entomología, 48*(1), Article e7807. https://doi.org/10.25100/socolen.v48i1.7807

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012–1014. https://doi.org/10.1038/nature07634

Goldsmith, J., Castillo, J., & Clarke-Harris, D. (2013). Gall midges (Cecidomyiidae) attacking horticultural crops in the Caribbean region and South America. In J. E. Peña (Ed.), *Potential invasive pests of agricultural crops* (pp. 240–250). CABI. https://doi.org/10.1079/9781845938291.0240

Hernandez, L. M., Guzman, Y. C., Martínez-Arias, A., Manzano, M. R., & Selvaraj, J. J. (2015). The bud midge *Prodiplosis longifila*: Damage characteristics, potential distribution and presence on a new crop host in Colombia. *Springer Plus, 4*(1), Article 205. https://doi.org/10.1186/s40064-015-0987-6

Jebb, A. T., Ng, V., & Tay, L. (2021). A review of key Likert scale development advances: 1995–2019. *Frontiers in Psychology, 12*, Article 637547. https://doi.org/10.3389/fpsyg.2021.637547

Johnston, A., Matechou, E., & Dennis, E. B. (2023). Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution, 14*(1), 103–116. https://doi.org/10.1111/2041-210X.13834

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *Current Journal of Applied Science and Technology*, 396–403. https://doi.org/10.9734/BJAST/2015/14975

Katapally, T. R., Hammami, N., & Chu, L. M. (2021). A randomized community trial to advance digital epidemiological and mHealth citizen scientist compliance: A smart platform study. *PLoS ONE, 16*(11), Article e0259486. https://doi.org/10.1371/journal.pone.0259486

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications, 82*(3), 3713–3744. https://doi.org/10.1007/s11042-022-13428-4

Kroschel, J., Mujica, N., Alcazar, J., Canedo, V., & Zegarra, O. (2012). Developing integrated pest management for potato: Experiences and lessons from two distinct potato production systems of Peru. In Z. He, R. Larkin, & W. Honeycutt (Eds.), *Sustainable potato production: Global case studies* (pp. 419–450). Springer. https://doi.org/10.1007/978-94-007-4104-1_25

Kullenberg, C., & Kasperowski, D. (2016). What is citizen science? – A scientometric meta-analysis. *PLoS ONE, 11*(1), Article e0147152. https://doi.org/10.1371/journal.pone.0147152

Lippi, G., & Cervellin, G. (2019). Is digital epidemiology reliable?—Insight from updated cancer statistics. *Annals of Translational Medicine, 7*(1), Article 15. https://doi.org/10.21037/atm.2018.11.55

Maulud, D. H., Zeebaree, S. R. M., Jacksi, K., Sadeeq, M. A. M., & Sharif, K. H. (2021). State of art for semantic analysis of natural language processing. *Qubahan Academic Journal, 1*(2), 21–28. https://doi.org/10.48161/qaj.v1n2a44

Merigó, J. M., & Yang, J.-B. (2017). A bibliometric analysis of operations research and management science. *Omega, 73*, 37–48. https://doi.org/10.1016/j.omega.2016.12.004

Mujica, N., & Kroschel, J. (2019). Ecological, economic, and environmental assessments of integrated pest management in potato: A case study from the Cañete Valley, Peru. *Food and Energy Security, 8*(1), Article e00153. https://doi.org/10.1002/fes3.153

Nain, M. S., Singh, R., & Mishra, J. R. (2019). Social networking of innovative farmers through WhatsApp messenger for learning exchange: A study of content sharing. *The Indian Journal of Agricultural Sciences, 89*(3), 556–558. https://doi.org/10.56093/ijas.v89i3.87605

Park, H.-A., Jung, H., On, J., Park, S. K., & Kang, H. (2018). Digital epidemiology: Use of digital data collected for non-epidemiological purposes in epidemiological studies. *Healthcare Informatics Research, 24*(4), 253–262. https://doi.org/10.4258/hir.2018.24.4.253

Peña, J. E., Baranowski, R. M., & McMillan, R. T. (1987). *Prodiplosis longifila* (Diptera: Cecidomyiidae) a new pest of citrus in Florida. *The Florida Entomologist, 70*(4), 527–529. https://doi.org/10.2307/3494798

Pereira, V., Basilio, M. P., & Santos, C. H. T. (2025). PyBibX – A Python library for bibliometric and scientometric analysis powered with artificial intelligence tools. *Data Technologies and Applications, 59*(2), 302–337. https://doi.org/10.1108/DTA-08-2023-0461

R Core Team. (2004). *R: A language and environment for statistical computing* [Software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rojas Burgos, E. E. (2022). Los ODS y el riesgo ambiental en la producción de follajes en la provincia del Tequendama (Colombia). *Human Review, 11*(2), 1–11. https://www.researchgate.net/publication/366420758_ODS_y_el_riesgo_ambiental_en_la_produccion_de_follajes_en_la_provincia_del_Tequendama_Colombia

Rosas, L. G., Espinosa, P. R., Jimenez, F. M., & King, A. C. (2022). The role of citizen science in promoting health equity. *Annual Review of Public Health, 43*, 215–234. https://doi.org/10.1146/annurev-publhealth-090419-102856

Ryan, S. F., Adamson, N. L., Aktipis, A., Andersen, L. K., Austin, R., Barnes, L., Beasley, M. R., Bedell, K. D., Briggs, S., Chapman, B., Cooper, C. B., Corn, J. O., Creamer, N. G., Delborne, J. A., Domenico, P., Driscoll, E., Goodwin, J., Hjarding, A., Hulbert, J. M., ..., Dunn, R. R. (2018). The role of citizen science in addressing grand challenges in food and agriculture research. *Proceedings of the Royal Society B: Biological Sciences, 285*(1891), Article 20181977. https://doi.org/10.1098/rspb.2018.1977

Salathé, M. (2018). Digital epidemiology: What is it, and where is it going? *Life Sciences, Society and Policy, 14*(1), Article 1. https://doi.org/10.1186/s40504-017-0065-7

Samaras, L., García-Barriocanal, E., & Sicilia, M.-A. (2020). Comparing social media and Google to detect and predict severe epidemics. *Scientific Reports, 10*(1), Article 4747. https://doi.org/10.1038/s41598-020-61686-9

Tey, Y. S., & Brindal, M. (2012). Factors influencing the adoption of precision agricultural technologies: A review for policy implications. *Precision Agriculture, 13*(6), 713–730. https://doi.org/10.1007/s11119-012-9273-6

Tran, T., Porter, W. T., Salkeld, D. J., Prusinski, M. A., Jensen, S. T., & Brisson, D. (2021). Estimating disease vector population size from citizen science data. *Journal of the Royal Society Interface, 18*(184), Article 20210610. https://doi.org/10.1098/rsif.2021.0610

Tripathy, J. K., Sethuraman, S. C., Cruz, M. V., Namburu, A., Mangalraj, Kumar, N., Ilango, S., & Vijayakumar, V. (2021). Comprehensive analysis of embeddings and pre-training in NLP. *Computer Science Review, 42*, Article 100433. https://doi.org/10.1016/j.cosrev.2021.100433

Wei, X., Khachatryan, H., Hodges, A., Hall, C., Palma, M., Torres, A., & Brumfield, R. (2023). Exploring market choices in the US ornamental horticulture industry. *Agribusiness, 39*(1), 65–109. https://doi.org/10.1002/agr.21769

Xiong, Z., Shen, Q., Xiong, Y., Wang, Y., & Li, W. (2019). New generation model of word vector representation based on CBOW or Skip-Gram. *Computers, Materials* & *Continua, 60*(1), 259–273. https://doi.org/10.32604/cmc.2019.05155