NYMPHSTAR: an accurate high-throughput quantitative method for whitefly (Aleurotrachelus socialis Bondar) resistance phenotyping in cassava

Adriana Bohórquez-Chaux
International Center for Tropical Agriculture: Alliance of Bioversity International and International Center for Tropical Agriculture
María Isabel Gómez-Jiménez
International Center for Tropical Agriculture: Alliance of Bioversity International and International Center for Tropical Agriculture
Luisa Fernanda Leiva-Sandoval
Swedish University of Agricultural Sciences: Sveriges lantbruksuniversitet
Luis Augusto Becerra Lopez-Lavalle (labecerra66@outlook.com)
International Center for Tropical Agriculture: CIAT https://orcid.org/0000-0003-3520-2270

Research Article

Keywords: whitefly nymphs, counts, leaf, image analysis

Posted Date: January 27th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1247625/v1

Abstract

Background

Whitefly (*Aleurotrachelus socialis* Bondar) is an important pest causing high economic losses in cassava production systems in the north of South America. It reduces the plant's photosynthesis by colonizing the cassava leaves and either directly feeding on their phloem sap or excreting substances that allow the growth of sooty mold and the consequent reduction of the photosynthetic area. The deployment of the crop's natural resistance to this pest is the most effective approach to its management. Phenotypic evaluation to identify germplasm with superior whitefly-resistance (WFR) levels from that showing a whitefly susceptible (WFS) response will benefit from the availability of an accurate high-throughput, quantitative phenotyping method.

Results

We developed an accurate and efficient image-based phenotyping method (Nymphstar) to quantify the total number of third and fourth instar nymphs through red, green, and blue color space (RGB) image analysis as a plugin for ImageJ. Nymphstar estimates both the number of third and fourth instar nymphs and the percentage of leaf area they occupy. We tested 19 cassava genotypes and classified them after data analysis as resistant or susceptible to *A. socialis* attack. Benchmarking Nymphstar against manual nymph counts performed by a specialist revealed a highly significant correlation between direct nymph counts and those estimated using Nymphstar, which analyzed images and counted nymphs 150 times faster than by manual counts. Full-bench caging for a free-choice assay facilitated our assessment of WFR in the cassava germplasm and early replicated trials in the experimental population and enhanced the efficiency of whitefly (*A. socialis*) colonization on each cassava genotype to accurately depict the intensity of the resistance/susceptible response under a choice regime, simultaneously reducing human operator bias.

Conclusions

Nymphstar is a fast, accurate image analysis screening tool for the automated counting of nymphs and quantification of leaf area they occupy, allowing for the assessment of cassava resistance to whitefly on a large number of cassava plants in a glasshouse-based assay while avoiding the potential bias normally associated with field assessment and manual counting.

Background

For more than 800 million people in the tropics, cassava (*Manihot esculenta* Crantz) is a basic staple food [1]. This crop—adapted to adverse climatic and harsh soil conditions—produces good yields where

most other crops fail [2, 3]. Cassava thus provides food security and has a great potential for climate change adaptation in the tropics [4].

Cassava production is susceptible to large yield losses caused by several pests and diseases—cassava whiteflies (Hemiptera: Aleyrodidae) being one of the most important in the Neotropics, Africa, and parts of Asia, because they inflict not just indirect damage as virus vectors but also direct damage by feeding on the phloem, causing chlorosis and early leave drop [5, 6]. Additionally, because the whitefly diet is rich in sugars, they cover the foliage in sticky and sugary excreta that support the proliferation of sooty mold fungi [7].

The Americas support the largest cassava whitefly diversity, with 11 reported species [8, 9]. In the north of South America, *Aleurotrachelus socialis* Bondar, one of the most important whitefly species feeding and rearing on cassava, causes significant yield loss when prolonged attack occurs [10]. Fortunately, resistance to the attack of this species has been characterized in hundreds of cassava genotypes [11, 12].

Resistance to insect attack *in planta*—a defense mechanism resulting from plant–insect co-evolutionary forces—is essential for the plant's survival [13]. However, the introduction of novel un-adapted, exotic germplasm with a narrow genetic base is likely to lack key genetic features for traits such as pest and disease resistance. Moreover, the rapid adoption of these materials by farmers, together with the selection and propagation of fewer varieties, would lead to genetic uniformity, loss of genetic diversity, and vulnerability to new pests and diseases, as is the case in Africa, where cassava is an exotic crop. Over the course of 500 years from its introduction to the African continent [14], endemic organisms in Africa (for instance, *Bemisia tabaci* in sub-Saharan Africa–East) and Central Africa (SSA–ECA) or the African cassava mosaic virus (ACMV, a Begomovirus) have overcome cassava defense mechanisms, establishing through adaptation or mutation a new set of pests and diseases to the crop [15, 16]. After years of research, scientists are concluding that increases in *B. tabaci* population abundance play a key role not only in the yield losses caused by physical damage but also in the incidence and spread of both Begomoviruses (cassava mosaic disease (CMD) and Ipomoviruses (cassava brown streak disease (CBSD)) [17].

Reducing whitefly populations is thus an important step toward controlling viral diseases and the yield losses they cause. Significant reduction of whitefly populations can thus be achieved by both monitoring cassava growing regions to identify infestation patterns, so that appropriate IPM measures can be deployed, and screening novel germplasm to identify new sources of WFR for crop improvement. The essential insect counts for both processes are almost exclusively performed either manually under the stereomicroscope or by direct counting in the field [11, 18]. Manual counting requires access to a highly trained professional and is usually tedious, time-consuming and prone to errors. Furthermore, where thousands of samples need to be processed, manual counting is an unsuitable method for monitoring whiteflies in the field or for breeding purposes.

Recent advances in the quality of digital and electronic images have promoted the use of image analysis and processing methods for pest monitoring and phenotyping of resistance to insects [19, 20]. Early

detection systems of pest attacks in the glasshouse have been developed using computer image analysis and based on monitoring and surveillance of the adults using images or video. These methods have proved robust in not only counting insects [21] but also detecting and discriminating between groups such as whiteflies, aphids, and thrips in the glasshouse [19, 22, 23], and in estimating the damage caused by different species [24]. Specifically for whiteflies, current image-based identification methods offer a fast and accurate way to count the number of adults using machine learning routines; some of these methods have been used for phenotyping plant resistance to *B. tabaci* and *Trialeurodes vaporariorum* but none currently exist for *A. socialis* [24, 25, 26].

Although counting of adult whiteflies is used for assessing the relative levels of varietal resistance in several crops, including cassava [18], the plant has three mechanisms to counterattack the pests—antixenosis, antibiosis, and tolerance [27, 28].

Given that adults can choose where to feed and oviposit, varietal resistance by antixenosis is assessed by the number of adults that visit and remain on their host in several crops, including cassava [18]. However, factors such as time of day, weather, and the high mobility of adults affect adult counts [18]. Evaluation of an immobile population at the last immature stages avoids the effect of these external factors, simultaneously identifying the survival population for the next generation of whiteflies after antixenotic (oviposition preference) or antibiotic (mortality of eggs and first and second instar nymphs) plant strategies have been deployed. Defense properties of resistant genotypes cause a higher mortality of *A. socialis* first and second instar nymphs than that of third and fourth instar nymphs [11].

To the best of our knowledge, in 2014 [29] the most advanced image analyses-based method for estimating immature whitefly numbers was developed using an algorithm to estimate the number of third and fourth instar nymphs of *B. tabaci* on soybean leaves using a digital image processing routine. This method is based on color transformation from the original images by splitting the image channels and mathematical morphology operations to highlight and delimit the nymphs, exoskeletons, and lesions on leaf sections, resulting in a faster estimation with high correlation between manual counting and the automated method, but discarding the nymphs near the veins and leaving spurious objects, causing estimation errors [29]. The features extracted by implementing this image analysis routine using machine learning techniques can be exploited to increase the precision of nymph identification. No novel methods exploring this improvement for whitefly nymph identification and counts have been reported.

Importantly, all currently available image-based whitefly identification and counting methods using nymph stages provide a user-friendly, easily accessible final application for further assessment of their practical use in field or glasshouse studies of whitefly crop infestation.

Hence, we aimed here to: 1) propose a phenotyping glasshouse-based assay for a large-scale evaluation of cassava genotypes to assess whitefly (*A.socialis*) varietal resistance, and 2) develop a user-friendly, accurate, high-throughput, image-based software application for estimating the number of third and fourth instar *A. socialis* nymphs developed on young cassava leaves.

Materials And Methods

Mass rearing of *A. socialis* colony

The process for mass rearing of *A. socialis* comprises three phases: 1) field production of standard cassava plant material (genotype COL1468) to obtain a regular supply of seed-stakes for host plantlet propagation; 2) screen house host plantlet production (COL1468) for controlled glasshouse infestations; and 3) massive adult whitefly production through rearing in the glasshouse. We developed the method described here, which facilitates the production of an average of 7600 whitefly adults per plant, by adapting previous studies [11].

1.1 Field production of seed-stakes: Every three months, we vertically planted 300 stakes of the genotype COL1468, which tolerates a high whitefly population density, in the field, separating both plants and rows by 1 meter. The optimal length of each seed-stake was approximately 20 cm or 5 axillary buds. We fertilized these plants one month after planting with N:P:K (15:15:15) and during the rest of the growing cycle supplemented them with micronutrients when the plants showed deficiencies (such as iron and zinc). We monitored the plants each week and controlled pests and diseases with pesticides as needed. At eight months, we began harvesting of the planting materials (seed-stakes) for growing plants in the screen house. We immediately interrupted pesticide use to avoid any effect on whitefly colony development.

1.2 Growing of host plantlets in screenhouse: On a weekly basis, we collected one hundred seed-stakes (aged 8 to 10 months) from the field. Thereafter, we planted seed-stakes in 2 L pots containing sterile substrate (1:3 sand to black soil; no clay topsoil) and maintained them in a whitefly-free screenhouse for six weeks (Figure 1A). We applied fertilization 15 days after planting with N:P:K (15:15:15) and watered when needed. Manual control of pests required their continuous monitoring; we avoided the use of agrochemicals for mites, thrips, and other organisms at this stage, as traces of pesticides could significantly affect whitefly development and population reproduction fitness.

1.3 Whitefly colony: We maintained the *A. socialis* colony permanently in a glasshouse with daily average temperature of 27.5 °C ± 0.1 °C and relative humidity of 66% ± 0.3% (mean ± SEM). We separated the glasshouse into two spaces: i) the infestation chamber and ii) the development chamber. In the infestation chamber (i), we permanently kept two groups of plants: a) infested COL1468 plants with fourth instar nymphs, and b) clean COL1468 plants. Twice a week, we moved 30 six-week-old COL1468 host plantlets from the screen house into the infestation chamber, where we allowed whiteflies to oviposit for 72 to 96 hours (Figure 1B). Immediately thereafter, we introduced another group of 30 six-week-old COL1468 host plantlets to the infestation chamber and shook the previously infested group of plants to remove the adults. We transferred the group of adult-free plants infested with oviposited eggs to the development chamber (ii) (Figure 1C & D).

When *A. socialis* nymphs reach the fourth instar stage (approximately 30 days after infestation), we water-sprayed each leaf of the plants to remove exuviae, honeydew, sooty mold, and most of the white

wax this species produces in their immature stages. This procedure did not disturb the nymph development cycle and allowed leaf by leaf inspection for opportunistic undesirable pests, permitting their manual removal prior to the leaves being placed into the infestation chamber where whitefly adults would emerge. Because adult whiteflies prefer the youngest leaves, we cut the shoot apices of the introduced plants to improve the infestation of the new un-infested plants (Figure 1E & F).

Phenotypic assay to measure cassava defense responses against whitefly infestation

We developed a robust and easy-to-use whitefly infestation assay to assess the relative levels of cassava defense responses (resistance vs susceptible) against whitefly infestation. We designed a full-bench caging free-choice experiment to be performed under practical glasshouse-base condition (hereafter referred to as a glasshouse WFR assay). We evaluated two full-sib cassava families (240 and 198 individuals for each family), segregating for resistance to whitefly in eight infestation trials with 176 replications across four years (2013, 2016, 2017, and 2018). We included 19 cassava genotypes in these experiments as checks, of which 10 genotypes had a known resistance response to *A. socialis* infestation based on previous studies [12].

We therefore chose high levels of whitefly infestation to standardize the WFR glasshouse assay. Nine genotypes used as checks carried traits of economic importance to the cassava program's stakeholders, although their resistant response to high whitefly infestation levels were unknown (Table 1). Thus, for the purpose of validating the effectiveness of glasshouse-based assay, we used only the data gathered from the 19 *M. esculenta* hereafter (Additional file 1).

Table 1. Cassava genotype checks used in bioassays of resistance to whitefly *A. socialis*, carried out during 2013–2018. WFR: resistant to whitefly; WFS: susceptible to whitefly. CBSD: cassava brown streak disease; CMD: cassava mosaic disease.

Genotype	Whitefly response and other biotic stress response and author reference	Type of variety, origin
COL1468	WFS, is host for *A. socialis* mass rearing [11]	Landrace, Colombia
COL2182	WF response unknown, CBSD resistant [30]	Landrace, Colombia
COL2246	WFS, is parental of segregant family [12]	Landrace, Colombia
ECU19	WF response unknown, CBSD resistant [30]	Landrace, Ecuador
ECU41	WF response unknown, CBSD resistant [30]	Landrace, Ecuador
ECU72	WFR, is parental of segregant family [11]	Landrace, Ecuador
ECU183	WFS [12]	Landrace, Ecuador
PAR41	WF response unknown, CBSD resistant [30]	Landrace, Paraguay
PER183	WFS [12]	Landrace, Peru
PER226	WF response unknown, CBSD resistant [30]	Landrace, Peru
PER317	WFR [12]	Landrace, Peru
PER335	WFR [12]	Landrace, Peru
PER368	WFR [12]	Landrace, Peru
PER415	WFR [12]	Landrace, Peru
PER556	WF response unknown, CBSD resistant [30]	Landrace, Peru
PER597	WF response unknown, CBSD resistant [30]	Landrace, Peru
PER608	WFR [12]	Landrace, Peru
TMS60444	WFS, is parental of segregant family [31]	African improved variety
TME3	WF response unknown, CMD resistant [32]	African improved variety

2.1 Production of clean cassava planting material: We re-propagated eight-week-old *in vitro* plantlets of the genotypes listed in Table 1 at CIAT´s cassava program tissue culture lab. Once these materials showed four expanded leaves, we transferred them to a screen house for tissue hardening in soil, where they were transplanted into black plastic bags (10 cm W x 15 cm H) filled with sterile soil substrate (1:2 sand: black soil).

2.2 Infestation with *A. socialis*: Approximately two months after soil transfer, we moved the new cassava plantlets displaying at least five fully expanded leaves to the infestation glasshouse to conduct the WFR

assay. Here we placed the plants on a table (18 m L x 3 m W), each separated by 20 cm, with total capacity of 100 plants (Figure 2A). On the experimental table, we placed one COL1468 plant and one ECU72 plant as a susceptible and resistant checks, respectively, and one host plantlet of COL1468 as an infestation control in each replicate. We covered each table with a large white mesh tent (18 m L x 3 m W x 3 m H) to confine the whitefly adults after infestation (Figure 2B). Prior to whitefly infestation, we identified and marked those leaves preferred by *A. socialis* adults as Leaf-1 and Leaf-2, where Leaf-1 corresponded to the youngest expanding leaf and Leaf-2 to a young fully expanded leaf (Figure 2C & D). We then marked the stem with a permanent ink marker below Leaf-2 to monitor the position of this leaf when whiteflies reached the fourth instar stage. Infestation occurred with the adults perched on six COL1468 plantlets that remained 72 to 96 hours in the infestation chamber of the whitefly colony glasshouse (Figure 2E). We transferred these plants to the infestation glasshouse and shook them above the experimental plants, releasing approximately 22,000 adults.

Seven days after infestation, we transferred the plants to another screen house to facilitate the development of immature whiteflies, simultaneously avoiding unwanted infestations by other undesirable pests. Forty days after infestation, when most nymphs had reached the fourth instar, we marked Leaf-2 on the upper side of the petiole with a permanent ink marker for easy recognition during image capturing. We water-sprayed Leaf-1 and Leaf-2 as described in the whitefly colony methodology (Figure 2F). We collected clean infested leaves, placed them outspread between two reusable paper towels, and stored them at 4 ℃ prior to image acquisition. With this method, we were able to store the leaves for several weeks until the image could be captured.

Image acquisition to develop the Nymphstar plugin

Here we proposed an image-based nymph count scoring system (Nymphstar) to reduce labor and accelerate the data acquisition necessary to assess whitefly infestation levels *in planta*. Previously published data prioritized high-quality images as a prerequisite for building an image-based whitefly nymph identification tool. To minimize potential shortcomings arising while retrieving meaningful data from each cassava whitefly-infested leaf image, we pre-treated the leaves with 50% ethanol to eliminate undesirable residues such as white wax and honeydew, which may introduce noise into the analysis. This process revealed a high contrast between the characteristic black color of the third and fourth instar nymphs of *A. socialis*, and the green color of the leaves (Figure 3-1).

To capture the images, we immediately placed each leaf into the ORTech Photo-e-Box Bio using a black fabric that absorbed less light as a background to increase the contrast and favor even lighting of the leaf. We fixed the camera—a 12.3-megapixel regular camera Nikon D300s, with an AF-S DX Micro-NIKKOR 40 mm f/2.8G lens—onto a Copy-Stand (Kaiser Reproduction Stand RS1/RA1 5510) at 70 cm from the black background (Figure 3-3). In cases where the leaves were bigger than the available visual field, we divided them into two or three pieces (Figure 3-2). To standardize the image-capture settings, we used the Nikon Control Pro 2 software. We captured the images through the RGB color model (red, green, blue color space) with a resolution of 4228 x 2848 pixels and stored them in JPG format.

Nymphstar: Image analysis for nymph counting and nymph density estimation

We developed the image analysis application Nymphstar in Java language, as a plugin for ImageJ software (National Institute of Health, USA). To phenotype plants with different leaves sizes and morphologies, we analyzed captured leaf images using the application to obtain the total number of nymphs and the leaf area they covered. The Nymphstar plugin operates on Linux, Mac OS X and Windows, in both 32-bit and 64-bit modes.

Based on preliminary work [33], we designed the Nymphstar application for image analysis to follow three main steps; we illustrate an overview of the Image processing flow of Nymphstar in Figure 4.

(1) Pre-processing comprised performing operations on the images to suppress undesired objects that distort nymph detection; the first step was noise reduction of the original image with a Gaussian blur filter [34,35]. We accomplished this specific step with a machine learning technique called Bayesian learning, using color as the training feature.

(2) Processing comprised application of different methods to extract the desired information from the image, according to [29]. We used the image sharpening filter “Unsharp Mask” to generate a smooth effect and loss of edges [36]. To remove undesirable objects on the produced image, we used the ImageJ plugin “Analyze Particles” [37]. It may be difficult to count nymphs crowded on leaves with high infestation without losing some data during the process. In order to account for each individual nymph within the cluster, we used the “Watershed Segmentation” plugin of ImageJ as described before [38].

(3) Post-processing comprised the analysis and interpretation of the extracted nymph data. We carried out nymph quantification by applying the Euler number implemented in the MorphoLibJ package of imageJ [39].

Accuracy and efficiency test of Nymphstar

We tested the data accuracy and image processing speed of Nymphstar against those of manual counting of ground-truth images. We randomly selected 2% of the total images obtained from the 19 *M. esculenta* checks evaluated and classified them into one of three infestation levels adapted from the population scale of six levels proposed by previous studies [11].

An expert entomologist, a person with an intermediate level of experience in nymph counting, and a beginner performed manual counting using the selected images taken with the protocol of image acquisition and quantified the nymphs by visual inspections on the computer screen using a digital counter to mark and count the number of nymphs. For manual counting, the time was registered with a digital stopwatch recording the time per picture, while for digital counting, running Nymphstar on a computer with an Intel Core I7-7500U processor with a speed of 2.7 GHz and 16 GB of RAM, we estimated the time from the creation time (hh:mm:ss) of each image recorded in file properties.

Finally, we contrasted the original images with the output images produced by Nymphstar to verify the segmentation between background, leaf, and nymphs.

Statistical analysis

We performed statistical analysis using SAS software 9.3 for Linux with the PROC GLM procedure. We estimated the effect of whitefly (*A. socialis*) infestation on 19 cassava clones (Table 1) by averaging the number of nymphs found in leaves 1 and 2 per plant, obtained from Nymphstar, across the experiments performed in 2013, 2016, 2017, and 2018 for mean nymph numbers and in 2017 and 2018 for percentage of area occupied by nymphs. Our preliminary descriptive analysis of the data showed that the distribution of the nymph variable corresponded to a negative binomial distribution; we therefore used a generalized linear model for this type of distribution, for nymph numbers per leaf, before establishing differences between means of genotypes using independent-samples t-tests (LSD). We regarded $P < 0.0001$ as significant in detecting statistical differences. We used the same model and test of comparison of means to evaluate the percentage of areas occupied by the insects but adjusting the model to a binomial distribution for this type of data. For the accuracy test of Nymphstar, concordance correlation method was used to evaluate the agreement between manual counting (Expert, intermediate and beginner evaluators) versus Nymphstar plugin counting. The epiR R package was employed to calculate the concordance correlation coefficient (CCC) and the respective confidence interval at 95%. Bias was determinate for each pair of comparison by computing the average difference of both measurements. Correlation and Bland-Altman plots were produced using ggplot2 R package.

Results

Nymphstar: Image Analysis for nymph counting and nymph density estimation

Pre-processing, background removing

To extract the desired features (nymph count data) from the leaf images, we decoupled the leaf information from the image background by extracting the pixels corresponding to the leaf from those corresponding to the background. We accomplished this specific step using a machine learning technique called Bayesian learning, using color as the training feature. First, we filtered the original image (see Figure 5A) with a Gaussian blur filter, reducing details and noise, including that of the nymphs (Figure 5B). We used the trained Bayesian learning method to extract the leaf color information, resulting in a segmented image with two categories: (i) background (black = zero value pixels) and (ii) leaf (white = 255 value pixels). We removed all particles on the background (Figure 5C and D). With this new image providing necessary information, we calculated the total leaf area to estimate nymph occupancy area. We then used the resulting image (Figure 5D) as a mask for the original image (Figure 5A) to produce a new RGB image for further processing (Figure 5E).

Processing: Image segmentation and detection of desired objects

The new RGB image without background, alone, did not provide enough information for accurate image segmentation. We thus decomposed the image into its three-color channels—red, green, and blue (RGB)—to assess which of the three contained most of the information for capturing nymphs on the leaves, finding the green channel to contain more information on the leaf and the blue channel, more information on the nymphs. Subtracting the blue channel from the green channel, the black regions, in our case corresponding to the nymphs, become very dark, and the regions of the leaf veins and borders become grayish, facilitating segmentation—a key feature to allow its counting in subsequent steps (Figure 5D). We corrected the effect of smoothness and loss of edges using the image sharpening filter "Unsharp Mask" (Figure 5F). We next filtered the pixels by color. Since the leaves of the plants were uniformly illuminated above a black background, and after applying the subtraction of channels to isolate the nymphs´ pixels, the nymph's color took a pixel value of zero; we therefore used a low-pass filter set at that zero value of nymph´s color. With this filter we assessed the intensity values of each image pixel and set those pixels with intensity levels higher than zero value (grays and whites) to zero value and eliminated them, while setting those that were at the same zero value as the nymphs at 255 (white) and considered them as information. This process rendered as a result a binary image with white-colored nymphs on a black background (Figure 5G).

However, in some instances and after this thorough process, some undesirable objects remained in the produced image; to remove them, we used the ImageJ plugin "Analyze Particles" set to a minimum and maximum pixel size of 20 and 700, respectively. We also set up a circularity range between 0 and 1. Combining the original image with this new binary one with an "AND" logical operator, we observed that some nymphs crowded together were considered as one, with the concomitant loss of some data. In order to account for each individual nymph within the cluster, we used the ImageJ "Watershed Segmentation" plugin. We once again used the "Analyze Particles" plugin to filter, based on shape and size, the remaining undesired objects. After the separation of the nymphs, we set the new range to 30 for the minimum and infinite for the maximum.

Post-processing and data analysis

After each image had been segmented and all informative objects detected, we quantified the nymphs by applying the Euler number from the MorphoLibJ package of imageJ that quantifies the number of objects. This algorithm is the result of the number of white particles (N) minus the number of holes in those objects (H), , and by using the 8-connectivity among pixels, it computes the Euler number measurement. As the nymphs do not have holes inside them, the result was the number of nymphs on the leaf. Additionally, to estimate the percentage of area occupied by nymphs, we used the image histogram to extract the number of white pixels. Finally, for a better visual appreciation of the results, we combined the processed image (Figure 5G) with the segmented image of the leaf generated in the pre-processing step with colors inverted (Figure 5H). Hence, the full data acquisition package (Nymphstar) gives a JPG processed grayscale image including (i) the total nymph number estimation, and (ii) leaf area and percentage area with nymphs (Figure 5I).

Furthermore, our newly developed ImageJ plugin, Nymphstar, can be used for the acquisition of single image data, as well as for a group of images through batch analysis. For a group of images, traits are analyzed and exported to a CSV file, and processed images are stored in a target folder chosen by the user (Additional file 2, Batch-processing section). For a single image, data processing results are immediately displayed in a log window; in both cases, the resulting image (s) is saved in a JPG format.

Nymphstar's accuracy and efficiency test

To test the accuracy of the Nymphstar application, we contrasted the manual nymph counts from ground-truth images with the results given by Nymphstar (Table 2). (Additional file 3 contains the results of the 57 images analyzed for the accuracy test). We randomly selected a total of 57 images with a resolution of 12 MPX (4228 by 2848 pixels) from the set of 3871 images obtained from the 1464 plants evaluated (two leaves from each plant and, depending on size, one or several pictures from each leaf), with a range of infestation between 57 and 4107 nymphs per leaf (Figure 6A and Additional file 4).

Table 2. Comparison of the average number of nymphs and time spent obtained with manual counting and using the Nymphstar plugin; for all variables we show the range in parenthesis. We adapted the infestation levels based on the population scale described before [11]. This scale has values from 1 to 6; therefore, the low level would be equivalent to 1 (no whitefly stages present) and 2 (1–200 individuals per leaf). The medium level would be equivalent to 3 (201–500 individuals per leaf) and 4 (501–2000 individuals per leaf). The high level would be equivalent to 5 (2001–4000 individuals per leaf) and 6 (>4000 individuals per leaf).

Infestation level	**No. images**	**Manual**		**Nymphstar plugin**	
		Number of Nymphs (Mean)	Time Mean (sec)	Number of Nymphs (Mean)	Time Mean (sec)
Low (0–200)	16	120.6 (22–199)	64.3 (23.8–93.3)	136.1 (57–206)	19 (18–22)
Medium (201–2000)	20	955.5 (201–2021)	339.76 (97.5–902)	959.7 (237–1670)	19,5 (18–24)
High (>2001)	21	3044.6 (1702–4941)	651.87 (196–1929.7)	2648,04 (1954–4107)	19,33 (18–20)

For the accuracy test, we performed a Lin's concordance index (see the statistical results in Additional file 5), obtaining the result r = 0.98 between the number of nymphs counted by the Nymphstar plugin and the manual count of nymphs. The Bland and Altman plot is used to compare two measurement techniques

based on the differences between their measurements, the differences are plotted to be able to observe the dispersion. In an ideal situation, the difference would be zero and all points should lie on the horizontal line y = 0 (dotted line). The solid black line represents the average of the differences obtained, ideally it should be a horizontal line at y = 0, but when the differences are not zero, this value would represent the bias.

The difference is calculated based on Nymphstar, so a positive bias would indicate that the Nymphstar count yielded, on average, higher values than the counts made by the person. A negative bias would indicate that the person counted more than Nymphstar. (Figure 7).

Analysis efficiency: In terms of nymph quantification speed, using the ImageJ plugin Nymphstar we analyzed the 57 images in 1534 seconds (25,56 minutes) with an average of 19,175 seconds per image (Table 2). The same images were analyzed manually in 25513,7 sec (5,97 hours) with an average of 6,29 minutes per image.

Glasshouse-based whitefly-resistance (WFR) bioassay

Here, we used the WFR bioassay for characterizing cassava relative resistance levels to whitefly infestation under glasshouse conditions to evaluate 19 cassava genotypes with known (9 genotypes) and unknown (10 genotypes) resistant and susceptible responses to *A. socialis* infestation (Table 1).

The cassava genotypes tested for WFR differed significantly in their responses to the whitefly *A. socialis* based on the number of nymphs found per leaf ($F_{(18, 3451)} = 22.32$, $P < 0.0001$) (Table 4, Additional file 6). We based the establishment of resistance/susceptibility categories on these results. The most susceptible genotype, PER556—with a previously unknown response to *A. socialis*—had significantly more nymphs than all other genotypes (1634.8 nymphs per leaf), followed by the susceptible TMS60444, PAR41, ECU183, PER226 and PER415 genotypes with more than 1150 but less of 1300 nymphs per leaf. The third group were genotypes with a mean number of nymphs per leaf between 840 and 1100: COL2246, ECU19, COL1468, PER597, TME3, COL2182, PER335, ECU41, and PER183, which we categorized as “Intermediate” because they fell between two statistically different genotypes—PER451 (group “BC”) and PER608 (group “DEF”), with a number of nymphs per leaf of 1154.9 and 841.4, respectively. The resistant genotypes PER608, PER317 and PER368 did not differ among them (841.4, 800.6 and 635.1, respectively). ECU72—a genotype previously studied and categorized as resistant to *A. socialis*—was conspicuously more resistant to whitefly attack (521,3 nymphs per leaf) over all the genotypes apart from PER368 (635,1 nymphs per leaf).

Table 4. Mean ± standard error of the total number of nymphs per leaf in checks analyzed for 4 years (2013, 2016, 2017, and 2018) in eight different experiments (number of plants per experiment shown in Additional file 1). We made these measurements using WFR bioassays.

clone	Mean ± S. E. (Nymphs per leaf) *	N Obs (# plants)	Whitefly response previous studies	Whitefly response using Nymphstar
PER556	1634.8 ± 213.5 **A**	4	Unknown	High WFS
TMS60444	1277.6 ± 38 **B**	134	WFS	WFS
PAR41	1163.7 ± 130.7 **BC**	20	Unknown	WFS
ECU183	1122.8 ± 110.4 **BC**	30	WFS	WFS
PER226	1156.8 ± 84.3 **BC**	20	Unknown	WFS
PER415	1154.9 ± 80.5 **BC**	36	WFR	WFS
COL2246	1102.4 ± 41.6 **BCD**	183	WFS	Intermediate
ECU19	1060.7 ± 127.2 **BCDE**	20	Unknown	Intermediate
COL1468	1041.1 ± 113 **BCDE**	185	Unknown	Intermediate
PER597	1046.1 ± 38.5 **BCDE**	18	Unknown	Intermediate
TME3	1037.9 ± 58.7 **BCDE**	81	WFS	Intermediate
COL2182	991.1 ± 105.2 **CDE**	19	Unknown	Intermediate
PER335	946 ± 75.7 **CDE**	61	WFR	Intermediate
ECU41	931.5 ± 104.2 **CDE**	20	Unknown	Intermediate
PER183	913.2 ± 65 **CDEF**	44	WFS	Intermediate
PER608	841.4 ± 46.1 **DEF**	156	WFR	WFR
PER317	800.6 ± 51.5 **EF**	89	WFR	WFR
PER368	635.1 ± 62.4 **FG**	60	WFR	WFR
ECU72	521.3 ± 18.3 **G**	284	WFR	High WFR

*For each clone, the means within a column followed by the same letter are not significantly different (independent-samples t-tests (LSD), DF = 3433, $P < 0.0001$)

Discussion

Automated identification and counts of pests in agriculture

In the last two decades, with the development of image digitization and data processing automation, a new body of research has emerged with a view to achieving precise and accurate early detection of important pests in agriculture [40,41]. Most of these investigations exploit digital imaging and are focused on the development of software or algorithms for identification of different pests [26,29,40, 42,

43,44]. To this end, for the whitefly, thrips and aphids, most of the studies undertaken today target the digital identification of adults, using in most cases sticky traps to capture them [25,41,44]. Very few cases in the literature have proposed insect identification at early stages of development (that is, eggs or nymphs) [24,29]. To the best of our knowledge, only two studies—one [45] developing a methodology using convolutional neural networks to identify and count nymphs of aphids directly on the leaves and one [29] developing an algorithm to count whitefly (*Bemisia tabaci*) third and fourth instar nymphs—are the best examples of attempts to identify insect pest in early stages of their life cycle. Unfortunately, neither of these two methods have been sufficiently automated to be deployed in nymph monitoring studies. Likewise, an automated nymph count methodology for whitefly species *Aleurotrachelus socialis*, which is an important pest of cassava in South America, is lacking.

The need to develop a precise, quick phenotyping method based on digitized image analysis for *A. socialis* whiteflies is urgent. The model for *A. socialis* identification must rapidly provide appropriate early detection of the insect *in planta* or evaluate the plant's potential resistance factors. The rapid and early identification of plants carrying resistance to whitefly attack is an important aspect of pest management because the early deployment of resistance in the field will prevent these species from becoming superabundant and widespread, and increasing the risk of yield loss or the spread of other viral diseases. To help solve this problem, we developed a tool as a plugin for the open access software ImageJ—Nymphstar—for the identification and quantification of *A. socialis* third and fourth instar nymphs and estimation of the leaf area occupied by nymphs. Differences in sizes and color of the leaves and levels of infestation do not affect identification and count accuracy. Moreover, Nymphstar vastly improves data acquisition time. Across eight independent glasshouse-based whitefly-resistance (WFR) phenotyping trials, we analyzed 19 *M. esculenta* checks with their replicas, for a total of 1937 plants, which corresponded to 3874 leaves bearing 775050 nymphs. While counting the nymphs manually on these ground-truth images could take an average of 39471 minutes (82 standard 8-hour labor days), analysis of all images using the automatic batch image analysis of Nymphstar would be completed in 20,63 hours, turning a multi-day task into one accomplished in just hours; Nymphstar promises to reduce the burden of routine *A. socialis* monitoring presently faced by agricultural scientists and extensionists. A practical application of this tool for field breeding would potentially reduce the WFR selection time in the absence of advanced molecular markers.

Currently, the traditional method for assessing the abundance of whitefly populations and their incidence on crop plants is manual, both in the field and in the glasshouse. Hence, the capacity to conduct large epidemiology surveys or assess potential plant resistance is doubly hampered by the need for access to both adequate laboratory facilities and well-trained entomologists. Nymphstar not only offers a solution to these limitations but also presents the opportunity to undertake large epidemiological surveys or to innovate in the phenotypic characterization of WFR. Although this tool was tested on cassava plants, it could be used for other plant types and other whitefly species, with some modifications in the processing of the images.

Image analysis

In the development of the plugin for counting of A. *socialis* nymphs we used some techniques to improve its efficiency and accuracy. To identify the pixels corresponding to the leaf and background, we used the naive Bayes approach. This technique is a probabilistic classifier based on the Bayes theorem, which computes the class of each observation using a likelihood-trained model [46]. Here, we trained the program with different color samples of leaves and backgrounds to obtain a binary image output where the image was labeled as (i) background (black = zero value pixels) and (ii) leaf (white = 255 value pixels). In this way, we were able to separate the leaf from the background and measure the area corresponding to the leaf.

Image reproducibility is key to the accuracy and efficiency of the Nymphstar data aquisition performance [43]. The PhotoBox allows for the use of the same light levels for each image to avoid changes in the characteristics of the image in subsequent analysis. The control of the light conditions and camera height improve image acquisition, avoiding the presence of shadows that could generate unhelpful data and maintaining the same height to compare leaf areas between samples. Nymph counts and measuring of the leaf area are usually performed manually or semi-automatically. The high-throughput method described here is an automated image-based phenotyping system, which can be easily adapted to other whiteflies and plants.

Measuring resistance to whitefly in cassava

Phenotyping assays

Central to the development of a whitefly-resistant variety is the identification of effective sources of resistance. Natural infestation (choice bioassay) of insect pests in cassava has been an effective but time-consuming process developed over nearly a quarter of a century [11,12]. Previous studies [11] suggested that a marked size reduction in the second and third instar nymphs and pupae in genotypes showing resistance was evident, suggesting a probable antibiosis mechanism. However, in choice bioassays, both antibiosis and antixenosis are plant and pest response strategies deployed during host infestation. To best account for both, we proposed a glasshouse-based WFR bioassay (Figure 2).

Here, we present a phenotyping methodology that allows screening for WFR (choice bioassay) of a large quantity of cassava genotypes and plants per trial. Our method has been applied to plants propagated in pots or bags and on different substrates, from *in vitro*, micro, or regular size stakes (data not published). Using this glasshouse-based methodology, it is possible to have results three months after planting, leading to the classification of plants in various resistant/susceptible categories without infestation from other organisms or the effects of changing weather in field experiments.

With the phenotyping methodology presented here we were able to differentiate and identify with statistically significant power the WFR (WFR vs WFS) of each of the cassava genotypes evaluated (Table 4). Our results are consistent with those based on measurements using damage and population scales in previous studies [11,12], in which ECU72 showed high levels of resistance to *A. socialis*. ECU72, a landrace from Ecuador, is most frequently used as a resistance check in all *A. socialis* resistance studies

and has been used as a female parent of two segregating mapping populations for WFR. In contrast, with our methodology we observed that genotypes PER368, PER317 and PER608—previously characterized as highly resistant [11]—did not display the high levels of resistance observed in ECU72. We classified other genotypes of great importance in cassava breeding since they have other characteristics such as resistance to CBSD in Africa [30], as is the case for ECU19, ECU41, and COL2182, as well as TME3, an African genotype resistant to the virus produced by CMD [47]; all these genotypes showed intermediate resistance to whitefly. Nymphstar allowed us to classify the most susceptible genotypes used across different years, such as PER556, TMS60444, PAR41, PER226, PER415, and ECU183. The susceptible category is represented by a wide variety of important cassava genotypes used for different purposes in breeding, including TMS60444, used as a model plant for cassava genetic transformation, and a male susceptible parent in the establishment of a WFR mapping population (Becerra Lopez-Lavalle, personal communication). Our glasshouse-based whitefly bioassay revealed that genotypes previously considered susceptible—such as COL1468 (used for the mass rearing of *A. socialis*)—are intermediate in the scale. Similarly, COL2246, used as a male susceptible parent in a full-sib segregating family, is now categorized as intermediate based on its response displayed to whitefly infestation with our new methodology.

For cassava breeding programs, precise and reliable phenotyping for WFR is extremely useful; additionally, evaluation of large collections of plants in a short time, makes the analysis more reliable, delivering quantitative measures such as nymph counts and leaf area occupied by nymphs. Both measures presented a high correlation and could be used separately or together in the analyses, giving equivalent results in this case study, in which we use different landraces of *Manihot esculenta.* In other cases, such as the phenotyping resistance in wild relatives of *Manihot* (data not shown), the differences in size and morphology of the leaves between these different *Manihot* spp made the measurement of percentage of leaf area occupied by nymphs (nymph density) more useful than the number of nymphs alone.

These highly accurate quantitative measurements of WFR are ideally suited to genomic and genetic studies searching for resistance genes, using QTLs or GWAS analysis [48, 49]. The automated methodology excludes errors and bias caused by manual counting. Previous QTL analyses for resistance to whiteflies conducted field testing using scales of damage and population of whiteflies as phenotypic data. Previous studies developed and used a scale of 1 to 6 to describe the damage caused by the whitefly population to cassava. This scale allowed for the characterization of cassava germplasm into resistant and susceptible categories, which were validated by later studies [12]. Although this scale is discriminated for the whitefly-resistant and -susceptible genotypes in cassava, it is not the best measure for quantitative genetic analysis. Preliminary analysis of these discrete data prevented the identification of QTLs in a whitefly-resistant segregating population (personal communication, Becerra Lopez-Lavalle). The Nymphstar data—quantitative and continuous in nature—may offer a better resolution in accounting for the variability of WFR across this segregating population.

Here we present the advantages of a quicker and less laborious application, where the outcome can be visually controlled and it is possible to acquire extra parameters, such as the leaf area occupied by

nymphs. We also highlight the use of the plugin outside working hours. It can be used whenever it is required, because it allows use in batch mode without human supervision.

Conclusions

We developed a high-throughput image analysis-based tool called Nymphstar to quantify the number and the area occupied by *A. socialis* nymphs on cassava. This application, along with a glasshouse-based WFR bioassay to measure relative levels of whitefly-resistance response across different *M. esculenta* landraces, together with experimentally segregating populations, has proven highly efficient in obtaining quantitative measurements such as the number of third and fourth instar nymphs and the leaf area they occupy. This tool offers an over eighty-fold reduction in the time taken for WFR evaluation. The Nymphstar application enables fast and accurate screening of multiple breeding populations to select superior genotypes for decreasing the relative population size of whitefly in cassava cropping systems, thus reducing the potential for insect-transmitted viruses (CMD and/or CBSD) to mutate into more virulent forms. Furthermore, this tool can assist plant epidemiologists in tracking the distribution of this pest.

Abbreviations

β: threshold that is applied on an image to separate the object of interest from background pixels, resulting in a binary image

CSV: data file of comma separated values

I: binary image. Pixels with values are those representing the objects of interest and those with zeros are those representing non-desired objects.

GLM: Generalized linear model used for statistical analysis.

N: number of nymphs on a leaf.

R, G, and B: red, green, and blue channels of the RGB image

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare they have no competing interests.

Funding

The present work has been funded by Natural Resources Institute, University of Greenwich, from a grant provided by the Bill and Melinda Gates Foundation (Grant OPP1058938).

Authors' contributions

ABC and LABLL were responsible of the experimental design and drafted the manuscript with sections written by ABC, MIGJ, LFLS, and LABLL developed and optimized the NYMPHSTAR protocol. LFLS developed the machine leaning code for ImageJ. ABC and MIGF conducted all the statistical analyses. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Carmen Adriana Bolanos, Carlos Ordonez, Vianey Barrera, Pablo Herrera , Teresa Cuasialpud, Daniel Encarnacion, Gerardino Perez, Fernando Mondragon, Harrison Moran, German Patiño and Adriana Vasquez for technical assistance and Prof. Linda Walling, Prof. Paul Fraser and Laura Perez-Fons for helpful discussion of data.

References

1. Lebot V. Tropical root and tuber crops: cassava, sweet potato, yams, and aroids. Wallingford, UK: CABI Publishing; 2009.
2. El-Sharkawy MA. Cassava biology and physiology. Plant Mol Biol. 2004;56:481-501.
3. Raphael IO. Technical efficiency of cassava farmers in Southeastern Nigeria: stochastic frontier approach. J Agric Sci. 2008; 3:152-6.
4. Burns A, Gleadow R, Cliff J, Zacarias A, Cavagnaro T. Cassava: the drought, war and famine crop in a changing world. Sustainability. 2010; 2:3572-607.
5. Bellotti AC, Braun AR. Arias B. Castillo JA, Guerrero JM: Origin and management of Neotropical cassava arthropod pests. Afr Crop Sci J. 1994; 2:407-17.
6. Herrera-Campo BV, Hyman G, and Bellotti AC: Threats to cassava production: known and potential geographic distribution of four key biotic constraints. 2011. Food Secur. 3, 329e345.
7. Nelson S. Sooty mold. 2008. Plant disease PD-52. Mänoa, Honolulu, Hawaii, Cooperative extension Service. College of Tropical Agriculture and Human Resources, University of Hawaii.

8. Ovalle TM, Parsa S, Hernández MP, Becerra Lopez-lavalle LA. Reliable molecular identification of nine tropical whitefly species. Ecol Evol. 2014; 4:3778-87.
9. Vásquez-Ordoñez AA, Hazzi NA, Escobar-Prieto D, Paz-Jojoa D, Parsa S. A geographic distribution database of the Neotropical cassava whitefly complex (Hemiptera, Aleyrodidae) and their associated parasitoids and hyperparasitoid (Hymenoptera). ZooKeys. 2015; 545:75-87.
10. Bellotti AC, Herrera-Campo BV, Hyman G. Cassava production and pest management: present and potential threats in a changing environment. Trop Plant Biol. 2012; 5:39-72.
11. Bellotti AC, Arias B. Host plant resistance to whiteflies with emphasis on cassava as a case study. Crop Protection. 2001; 20:813-23.
12. Parsa S, Medina C, Rodríguez V. Sources of pest resistance in cassava. Crop Prot. 2015; 68:79-84.
13. Berlinger MJ. Plant resistance to insects. In: CAPINERA, JL, editors. Encyclopedia of entomology. Dordrecht: Springer Netherlands; 2008.
14. Jones WO. Manioc: an example of innovation in African economies. Econ Dev Cult Change. 1957; 5:97-117.
15. Herren HR. Cassava pest and disease management: an overview. Afr Crop Sci J. 1994; 2:345-53.
16. Omongo CA, Kawuki R, Bellotti AC, Alicai T, Baguma Y, Maruthi MN, Bua A, Colvin J. African cassava whitefly, *Bemisia tabaci*, resistance in African and South American cassava genotypes. J Integr Agric. 2012; 11:327-36.
17. Milenovic M, Wosula EN, Rapisarda C, Legg JP. Impact of host plant species and whitefly species on feeding behavior of *Bemisia tabaci*. Front Plant Sci. 2019; 10:1.
18. Sseruwagi P, Sserubombwe WS, Legg JP, Ndunguru J, Thresh JM. Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: a review. Virus Res. 2004; 100:129-42.
19. Mundada RG, Gohokar VV. Detection and classification of pests in greenhouse using image processing. J Electron Commun Eng. March–April 2013; 5:57-63.
20. Berger B, de Regt B, Tester M. High-throughput phenotyping of plant shoots. Methods Mol Biol. 2012; 918:9-20.
21. Bechar I, Moisan S. On-line counting of pests in a greenhouse using computer vision; 2010. Available from: https://hal.inria.fr/inria-00515624.
22. Jongman C, Junghyeon C, Mu Q, Chang-woo J, Hwang-young K, Ki-baik U, Tae-soo C. Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis. Int J Math Comput Simul. 2007; 1:46-52.
23. Zayas I, Pomeranz Y, Lai F. Discrimination of wheat and nonwheat components in grain samples by image analysis. Cereal Chem. 1989; 66:233-7.
24. Bhadane G, Sharma S, Nerkar VB. Early pest identification in agricultural crops using image processing techniques. International Journal of Electrical. p. 2277-626 and Computer Engineering 2: 77-82; 2013. Electronics ISSN no [online].

25. Sun Y, Cheng H, Cheng Q, Zhou H, Li M, Fan Y, Shan G, Damerow L, Lammers PS, Jones SB. A smart-vision algorithm for counting whiteflies and thrips on sticky traps using two-dimensional Fourier. Transform spectrum. Biosyst. Eng. 2017, 153, 82–88.26.
26. Xia C, Chon TS, Ren Z, Lee JM. Automatic identification and counting of small size pests in greenhouse conditions with low computational cost. Ecol Inform. 2015; 29:139-46.
27. Painter RH. Insect resistance in crop plants. New York: Macmillan. 520 p.
28. Kogan M, Ortman E. Antixenosis-A new term proposed to define Painter´s" Nonpreference" modality of resistance. Vol. 24; 1978 [ESA bulletin]. p. 175-6.
29. Barbedo JGA. Using digital image processing for counting whiteflies on soybean leaves. J Asia Pac Entomol. 2014; 17:685-94.
30. Sheat S, Fuerholzner B, Stein B, Winter S. Resistance against cassava brown streak viruses from Africa in cassava germplasm from south America. Front Plant Sci. 2019; 10:567.
31. Irigoyen ML, Garceau DC, Bohorquez-Chaux A, Lopez-Lavalle LAB, Perez-Fons L, Fraser PD, Walling LL. Genome-wide analyses of cassava Pathogenesis-related (PR) gene families reveal core transcriptome responses to whitefly infestation, salicylic acid and jasmonic acid. BMC Genomics. 2020; 21:93.
32. Akano O, Dixon O, Mba C, Barrera E, Fregene M. Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. Theor Appl Genet. 2002; 105:521-5.
33. Müller-Linow M, Wilhelm J, Briese C, Wojciechowski T, Schurr U, Fiorani F. Plant Screen Mobile: an open-source mobile device app for plant trait analysis. Plant Methods. 2019; 15:Article number: 2.
34. Davies E. Machine vision: theory, algorithms and practicalities. Academic Press; 1990. p. 42-4.
35. Gonzalez R, Woods R. Digital image processing. Addison-Wesley Publishing Company; 1992. p. 191.
36. Polesel A, Ramponi G, Mathews VJ. Image enhancement via adaptive unsharp masking. IEEE Trans Image Process. 2000 ;9:505-10.
37. Igathinathane C, Pordesimo LO, Columbus EP, Batchelor WD, Methuku SR. Shape identification and particles size distribution from basic shape parameters using ImageJ. Comput Electron Agric. 2008;63(2, October):168-82.
38. Soille P, Vincent LM. Determining watersheds in digital pictures via flooding simulations. Proc SPIE. 1360, Visual Communications and Image Processing '90: Fifth in a Series, (1 September 1990).
39. Chen M, Yan PF. A fast algorithm to calculate the Euler number for binary images. Pattern Recognit Lett. 1988; 8:295-7.
40. Maharlooei M, Sivarajan S, Bajwa SG, Harmon JP, Nowatzki J. Detection of soybean aphids in a greenhouse using an image processing technique. Comput. Electron. Agric. 2017, 132, 63–70.
41. Wang ZB, Wang KY, Liu ZQ, Wang X, Pan S. A cognitive vision method for insect pest image segmentation. IFAC (international Federation of Automatic Control). Papers. 2018 85-89:51-17.
42. Solis-Sánchez L, García-Escalante J, Castañeda-Miranda R, Torres-Pacheco I, Guevara-González R. Machine vision algorithm for whiteflies (*Bemisia tabaci* Genn.) scouting under greenhouse

environment.
43. Espinoza K, Valera DL, Torres JA, López A, Molina-Aiz FD. Combination of image processing and artificial neural networks as a novel approach for the identification of *Bemisia tabaci* and *Frankliniella occidentalis* on sticky traps in greenhouse agriculture. Comput Electron Agric. 2016; 127:495-505.
44. Deng L, Wang Y, Han Z, Yu R. Research on insect image detection and recognition based on bio-inspired methods. Biosyst Eng. 2018; 169:139-48.
45. Chen J, Fan Y, Wang T, Zhang C, Qiu Z, He Y. Automatic segmentation and counting of aphid nymphs on leaves using convolutional neural networks. Agronomy. 2018; 8:129.
46. Hsu SC, Chen IC, Huang CL. Image classification using naive Bayes classifier with pairwise local observations. J Inf Sci Eng. 2017;32.
47. Colvin J, Omongo CA, Maruthi MN, Otim-Nape GW, Thresh JM. Dual Begomovirus infections and high Bemisia tabaci populations: two factors driving the spread of a cassava mosaic disease pandemic. Plant Pathol. 2004; 53:577-84.
48. Nzuki I, Katari MS, Bredeson JV, Masumba E, Kapinga F, Salum K, et al. QTL Mapping for Pest and Disease Resistance in Cassava and Coincidence of Some QTL with Introgression Regions Derived from Manihot glaziovii. Front Plant Sci. 2017; 8:1168.
49. Kayondo SI, Pino Del Carpio D, Lozano R, Ozimati A, Wolfe M, Baguma Y, Gracen V, Offei S, Ferguson M, Kawuki R, Jannink JL. Genome-wide association mapping and genomic prediction for CBSD resistance in *Manihot esculenta*. Sci Rep. 2018;8.

Figures

Figure 1

Graphical representation of *A. socialis* mass rearing. (**A**) Genotype COL1468 stakes planted in 2 L pots containing sterile substrate. (**B**) COL1468 plants in infestation chamber with whitefly adults. (**C**) Infested COL1468 in development chamber. (**D**) *A. socialis* life cycle on COL1468. (**E & F**) Shoot tips cut to force adult whiteflies to colonize new un-infested plants to start a new infestation cycle.

Figure 2

Graphical representation of phenotyping glasshouse-based assay. (**A & B**): Experimental cassava plants infested with adult whitefly were covered with a large white mesh tent. (**C & D**): Stem of each plant was marked with a permanent ink marker under the oldest leaf (Leaf-2) next to the youngest one with at least one lobe completely opened (Leaf-1). (**E**) From the *A. socialis* colony, COL1468 plants infested with adult

whiteflies were transferred to the infestation glasshouse and shaken above the experimental plants. (**F**): Harvesting at 38–40 days post-infestation, from Leaf-1 and Leaf-2, marked on the day of infestation, of all the genotypes under study.

Figure 3

(**1**) Cassava leaves before the acquisition of images: a) Leaf before washing, b) leaf after collection and storing at 4°C, and c) leaf after moistening with 50% ethanol. (**2**) Leaves fit to the visual field: complete leaf that fits into the visual range (left) and big leaf that was cut into two parts (right). (**3**) System implemented for image acquisition.

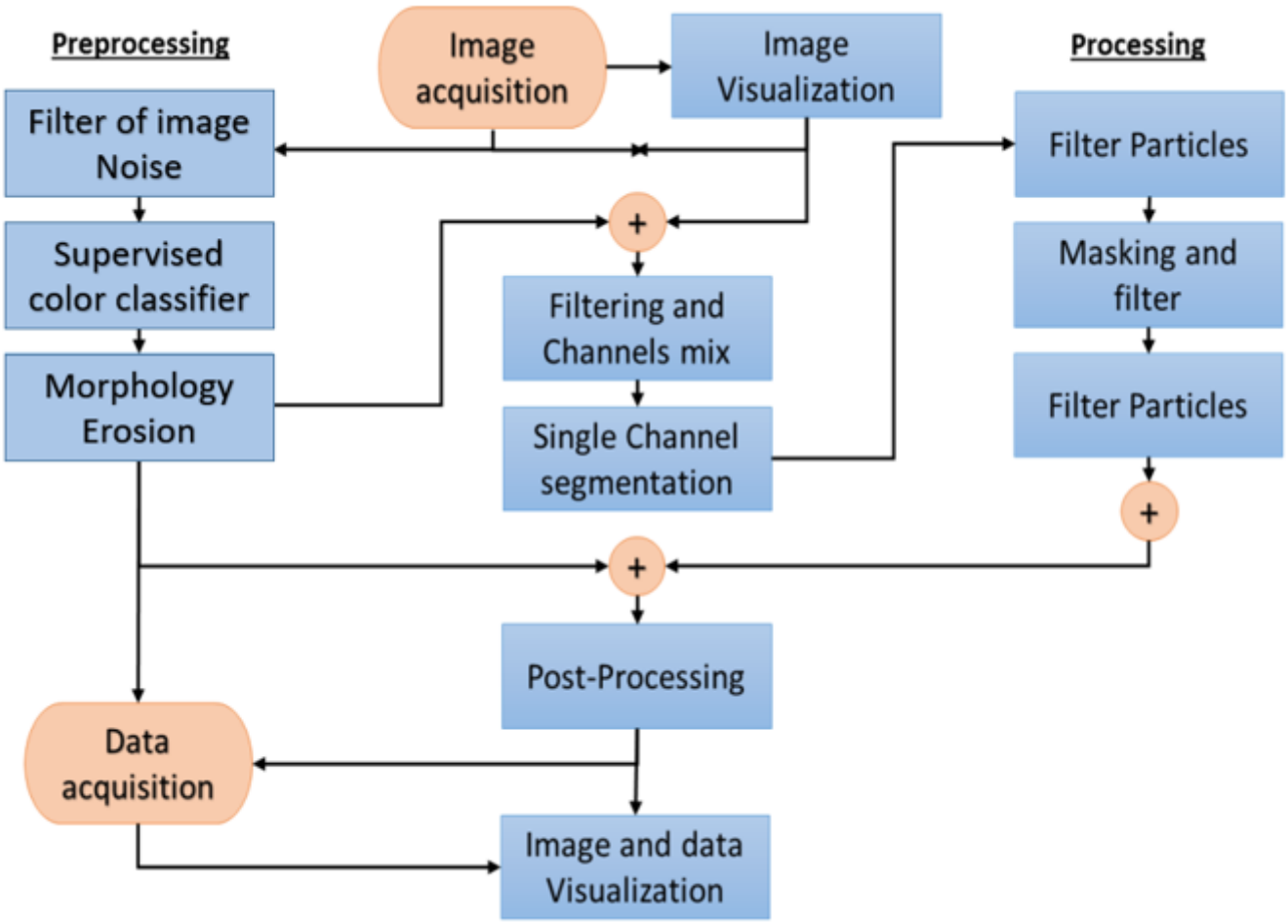

Figure 4

Flow chart showing the Nymphstar image processing steps from image acquisition to data acquisition.

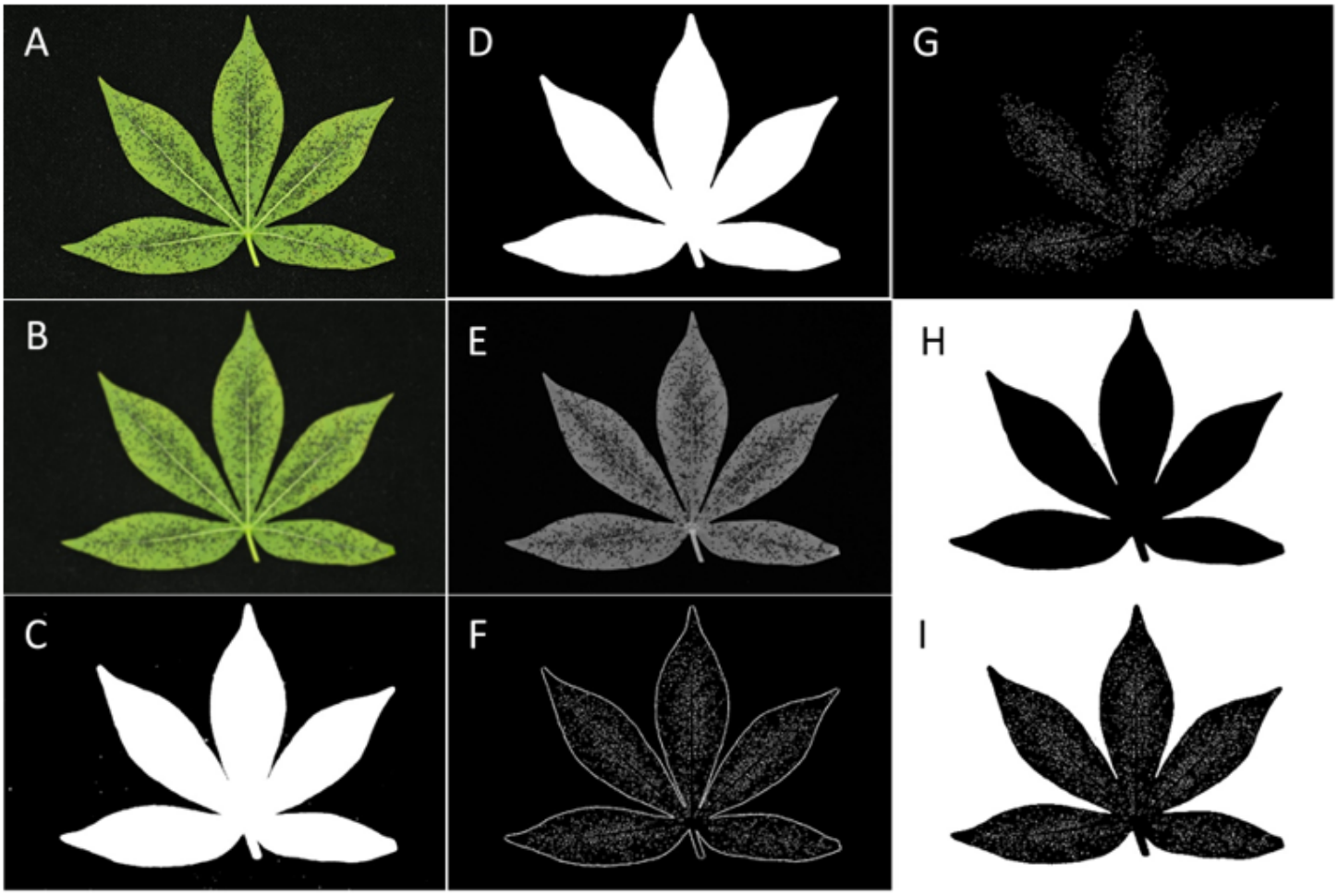

Figure 5

Steps of the image analysis process on a cassava leaf. A) Original image; B) Gaussian blur filter; C) image segmented by the supervised classifier; D) mask for subtracting background from figure (A); E) RGB image for decomposition on channels (channel green–channel blue image); F) filtered image with “Unsharp Mask” for edge sharpening; G) filtered pixels by color, detection of objects touching each other, filtered particles by shape and size, to obtain a binary image with nymphs colored in black; H) Figure 5G is combined with Figure 5C inverted; I) final JPG image.

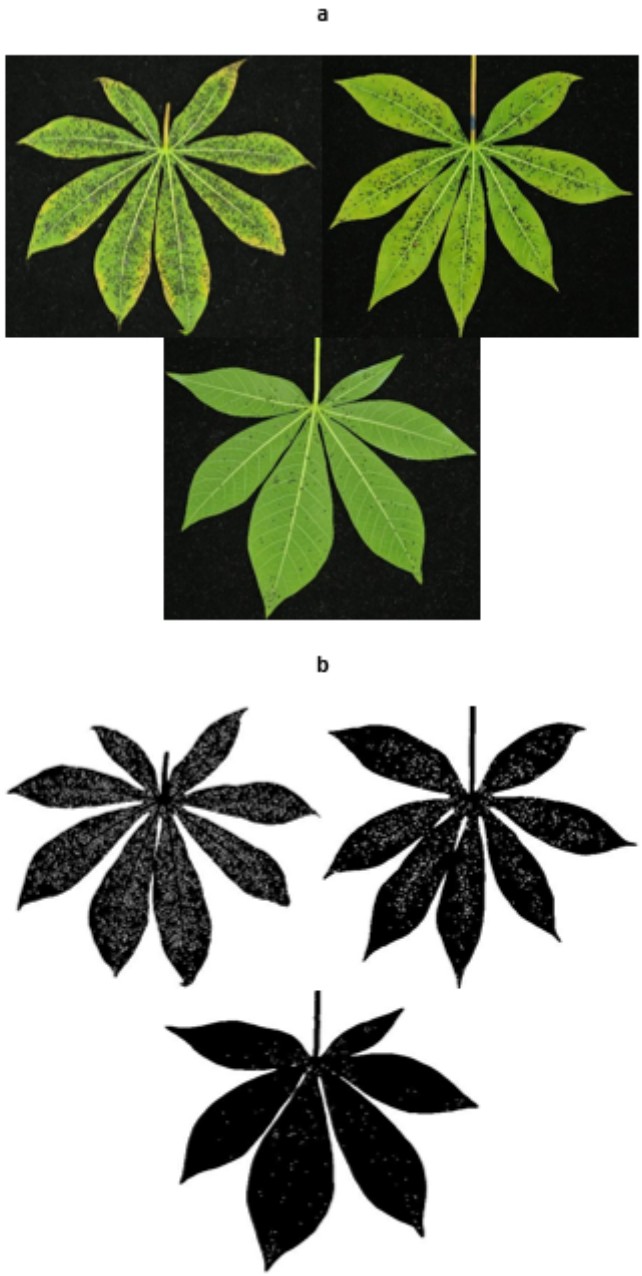

Figure 6

A, Cassava leaves with different levels of infestation of *A. socialis* nymphs—high (left), medium (center), and low (right), with populations between n = 22 and n = 4941 nymphs per leaf. **B**, Images of the same leaves obtained after the processing by the plugin Nymphstar.

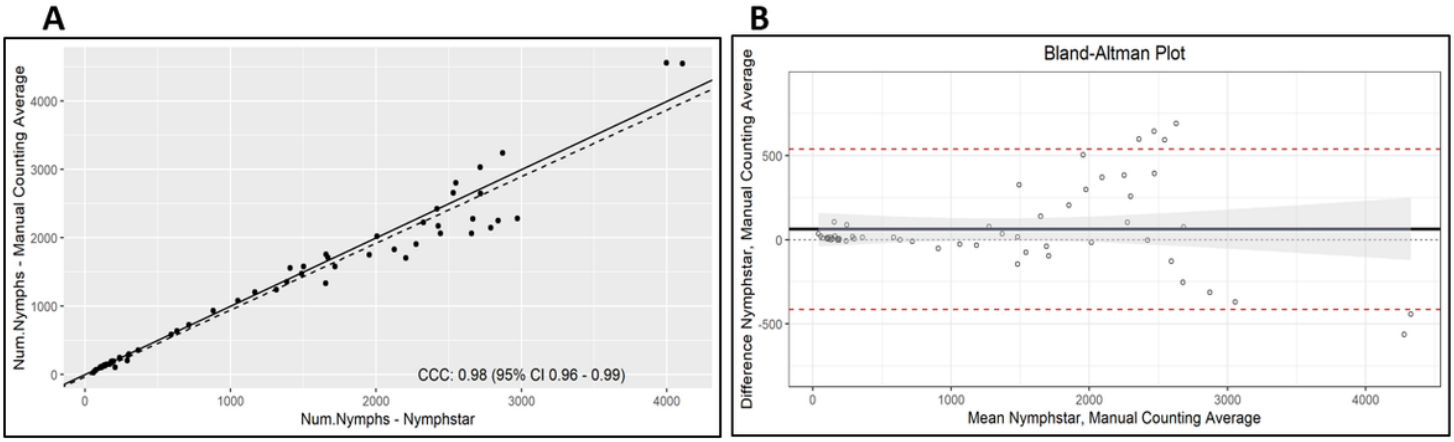

Figure 7

Lin's concordance index between the results of nymph counts (A) obtained by manual counts by observation and Nymphstar counts and (B), Bland-Altman plot where the dispersion of the data is observed when comparing manual counting vs counting using Nymphstar, based on 57 images.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1Eightinfestationtrials19checks.xlsx
- Additionalfile2Nymphstarmanual.pdf
- Additionalfile3LeaveinfestedbyWFCounts.xlsx
- Additionalfile4LeaveinfestedbyWF.jpg
- Additionalfile5Linsconcordanceindexanalysis.docx
- Additionalfile6Statisticalanalysis19checks.docx