



OPEN ACCESS

EDITED BY

George V. Popescu,
Mississippi State University, United States

REVIEWED BY

Beatriz Xoconostle-Cázares,
National Polytechnic Institute of Mexico
(CINVESTAV), Mexico
Alessandro Cestaro,
National Research Council (CNR), Italy

*CORRESPONDENCE

Nunzio D'Agostino

✉ nunzio.dagostino@unina.it

[†]These authors have contributed
equally to this work and share
first authorship

RECEIVED 20 August 2024

ACCEPTED 21 October 2024

PUBLISHED 08 November 2024

CITATION

Fruggiero C, Aufiero G, D'Angelo D, Pasolli E
and D'Agostino N (2024) Refining dual RNA-
seq mapping: sequential and combined
approaches in host-parasitic plant dynamics.
Front. Plant Sci. 15:1483717.
doi: 10.3389/fpls.2024.1483717

COPYRIGHT

© 2024 Fruggiero, Aufiero, D'Angelo, Pasolli
and D'Agostino. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Refining dual RNA-seq mapping: sequential and combined approaches in host-parasitic plant dynamics

Carmine Fruggiero^{1,2†}, Gaetano Aufiero^{2†}, Davide D'Angelo^{1,2},
Edoardo Pasolli² and Nunzio D'Agostino^{2*}

¹Department of Electrical Engineering and Information Technology, University of Naples Federico II,
Naples, Italy, ²Department of Agricultural Sciences, University of Naples Federico II, Portici,
Naples, Italy

Transcriptional profiling in host plant-parasitic plant interactions is challenging due to the tight interface between host and parasitic plants and the percentage of homologous sequences shared. Dual RNA-seq offers a solution by enabling *in silico* separation of mixed transcripts from the interface region. However, it has to deal with issues related to multiple mapping and cross-mapping of reads in host and parasite genomes, particularly as evolutionary divergence decreases. In this paper, we evaluated the feasibility of this technique by simulating interactions between parasitic and host plants and refining the mapping process. More specifically, we merged host plant with parasitic plant transcriptomes and compared two alignment approaches: sequential mapping of reads to the two separate reference genomes and combined mapping of reads to a single concatenated genome. We considered *Cuscuta campestris* as parasitic plant and two host plants of interest such as *Arabidopsis thaliana* and *Solanum lycopersicum*. Both tested approaches achieved a mapping rate of ~90%, with only about 1% of cross-mapping reads. This suggests the effectiveness of the method in accurately separating mixed transcripts *in silico*. The combined approach proved slightly more accurate and less time consuming than the sequential approach. The evolutionary distance between parasitic and host plants did not significantly impact the accuracy of read assignment to their respective genomes since enough polymorphisms were present to ensure reliable differentiation. This study demonstrates the reliability of dual RNA-seq for studying host-parasite interactions within the same taxonomic kingdom, paving the way for further research into the key genes involved in plant parasitism.

KEYWORDS

plant-parasitic plant interaction, transcriptomics, dual RNA-sequencing, read mapping, sequential approach, combined approach

1 Introduction

Parasitism in angiosperms involves a parasitic plant deriving nutrients from host plants. This complex ecological strategy has evolved independently approximately a dozen times, resulting in more than 290 genera and 4,700 species of parasitic plants (Westwood et al., 2010; Nickrent, 2020). Notably, certain parasitic species exhibit generalist behaviors, enabling them to parasitize multiple host species. At least 25 genera are recognized crop pathogens, including *Striga* (witchweeds), *Orobanchaceae* (broomrapes), and *Cuscuta* (dodder), posing significant threats to agriculture (Nickrent and Musselman, 2004). Quantifying yield losses can be challenging, yet the impact of parasitic weeds on international agriculture is undeniably on the rise (Hegenauer et al., 2017).

The evolutionary transition from non-parasitic ancestors to parasitic plants marked a shift from autotrophy (self-sustained nutrition through photosynthesis) to varying degrees of heterotrophy (reliance on external sources for sustenance) (Westwood et al., 2010). One way for classifying parasitic plants is based on their photosynthetic capacity. Hemiparasites can photosynthesize but primarily rely on hosts for water and mineral nutrients, while holoparasites lack photosynthesis and depend entirely on hosts for nutrition. Another classification is based on their dependency from hosts to complete their lifecycle: obligate parasites require a host, whereas facultative parasites can reproduce independently. While holoparasites are necessarily obligate due to their lack of photosynthesis, also some hemiparasites still need hosts for greatly enhanced the reproductive success thanks to the increased intake of mineral elements (Klaren and Janssen, 1978; Lambers and Oliveira, 2019). These interactions between host and parasite can have a significant impact on growth, reproduction, physiology, and ecosystem dynamics of the host (Hegenauer et al., 2017).

The most renowned and widespread stem holoparasitic genus is *Cuscuta*, comprising of about 170-200 species (Park et al., 2019). These plants have degenerated roots and leaves and stems spiral counterclockwise around their host plants.

The trophic connection between *Cuscuta* and its host relies on the development of a specialized structure called haustorium. The haustorium develops in stages, with haustorial cells forming hyphae that penetrate the vascular tissues of the host, mimicking xylem or phloem conduits. This intimate connection allows the parasitic plant to acquire water and nutrients and facilitates the horizontal transfer of macromolecules, including messenger RNA, small and long non-coding RNA, proteins, and even pathogens like viruses, phytoplasmas and viroids (Hosford, 1967; van Dorst and Peters, 1974; Kamińska and Korbin, 1999; Haupt et al., 2001; Birschwilks et al., 2006; Kim et al., 2014; Johnson and Axtell, 2019; Subhankar et al., 2021; Jhu and Sinha, 2022; Wu et al., 2022).

The exchange of RNA between hosts and *Cuscuta* can occur in both directions. For example, a study on *Cuscuta pentagona* parasitizing *Arabidopsis thaliana* found that 45% of the genes expressed in *Arabidopsis* were detected in *Cuscuta*, and conversely 24% of the genes expressed in *Cuscuta* were found in the *Arabidopsis* stem (Kim et al., 2014). The exact role of these

mobile RNAs is not fully understood, although different studies have suggested that some of them may be translated into proteins that affect plants' physiology or act as modulators of gene expression in response to abiotic and biotic stress (Westwood and Kim, 2017; Shahid et al., 2018; Hudzik et al., 2020; Maizel et al., 2020; Park et al., 2022).

The interface region acts as a battleground, where gene expression changes involve both parasite and host to disrupt various physiological processes of the other. These include recognition via pattern recognition receptors (PRRs), production of cytotoxic compounds, establishment of physical barriers, release of reactive oxygen species (ROS), and initiation of plant cell death (Albert et al., 2020). Understanding this complex conflict requires a deep understanding of the gene expression changes occurring in both host and parasite plant as they engage in interaction.

Since its introduction, RNA-sequencing (RNA-seq; known as bulk RNA-seq) has emerged as the favored technology for this purpose. Traditionally, RNA-seq reveals mRNA and/or non-coding RNA to provide a snapshot of gene expression in a sample (Mortazavi et al., 2008). Estimation of the average gene expression levels across a population of sampled cells provides insights into tissue-specific molecular mechanisms.

RNA-seq analysis comprises various steps: experimental design and sample acquisition; quality control and pre-processing of raw data; read mapping; gene/transcript level quantification; identification of differently expressed genes; and functional analysis (Conesa et al., 2016; Galise et al., 2021).

In experiments designed to capture transcriptional profiles of interacting organisms at the interface region, precise tissue sampling and RNA isolation are crucial steps.

Traditionally, the isolation of total transcripts has relied on preparing plant tissue sections for laser capture microdissection (LCM) followed by RNA isolation and high-throughput sequencing (Honaas et al., 2013; Jhu and Sinha, 2022). LCM combines microscopy with laser beams to isolate tissue types from host-parasite combined samples. Although this method has undergone refinement to yield efficient outcomes within a feasible timeframe, the intricate nature of the interface tissue (comprising host plant, haustorial, and hyphae tissue) poses challenges that require trained personnel and specialized equipment (Park et al., 2022). An alternative to address biases in separation techniques has been represented by the dual RNA-seq approach. This technique relies on sampling the entire host plant-parasitic plants interface, and expression profiles are discerned computationally by mapping reads to the respective reference sequences (genome/transcriptome) (Westermann et al., 2012; Naidoo et al., 2018; Wolf et al., 2018). The sampling from multiple tissues simultaneously introduces computational challenges in managing reads coming from different organisms. As example, mapping becomes non-trivial as it involves managing multiple mapping events within a single reference sequence (i.e., when reads map equally well on multiple loci within single organism), in addition to cross-mapping events occurring between the two reference sequences.

In this study, we categorized cross-mapping events into two types: (1) one-side cross-mapping, where reads from one organism are exclusively assigned to the other organism (often due to missing

mapping to the first organism genome, presumably reflecting genome incompleteness); (2) two-side cross-mapping, where reads from one organism are assigned to both organisms.

Within-genome multiple mapping results from gene duplication events, while cross-mapping is mainly due to insufficient divergence in the gene sequences between interacting organisms. For this reason, several reads can ambiguously map within and between organisms, an issue that is further emphasized when using short-read technologies (Westermann et al., 2017; Deschamps-Francoeur et al., 2020).

The percentage of cross-mapped reads is influenced by various factors, particularly the evolutionary divergence between interacting organisms. Dual RNA-seq approaches have been broadly employed to study various host plant-parasitic non-plant interactions (Liao et al., 2019; Du et al., 2022; Walker et al., 2024). In these cases, the interacting organisms are phylogenetically distant, resulting in substantial sequence divergence that reduces the probability of cross-mapping. Nevertheless, the entity of cross-mapping involving phylogenetically close organisms as in host plant-parasitic interaction has not been well explored. A naïve approach to handle ambiguously assigned reads is to discard them outright. However, this could potentially underestimate gene/transcript abundance levels.

Therefore, it is crucial to develop and implement more advanced techniques capable of accurately aligning reads to the reference genome in dual RNA-seq applications. The accuracy of this procedure depends on various factors, including the choice of alignment algorithm, the quality of the reference sequence, and the configuration of the algorithm parameters (Srivastava et al., 2020).

In this study, we assessed the feasibility of using dual RNA-seq to investigate interactions between phylogenetically close parasite and host species, both belonging to the Plantae kingdom, with focus on challenges related to multiple mapping and cross-mapping.

More specifically, we simulated two *in silico* interactions involving the parasitic plant *Cuscuta campestris* with two different hosts: *A. thaliana* (as a model organism) and *Solanum lycopersicum* (as a crop phylogenetically closer to *C. campestris*). The mapping was performed using the available reference genomes of *C. campestris*, *A. thaliana* and *S. lycopersicum*. We did not consider *de novo* transcriptome assemblies due to the uncertainty about horizontal transfer between host and parasitic plants, which could result in misattributed transcripts.

We compared two approaches to differentiate mixed reads, as summarized in Figure 1: mapping reads sequentially to the genomes of both species (the sequential approach), and mapping reads to a single combined genome (the combined approach).

2 Materials and methods

To manage the data, scripts in R (version 4.3.3) and GNU bash (version 5.0.17(1)) were developed in house (GNU, 2007; R Core Team, 2024). The Sankey plot was generated using SankeyMATIC (<https://sankeymatic.com/>) and further modified with Inkscape version 1.3 (<https://inkscape.org/>).

2.1 Phylogenetic analysis of *Cuscuta* spp. and their host range

We considered the large subunit of the ribulose-bisphosphate carboxylase (*rbcL*) gene (Manhart, 1994) to perform a phylogenetic analysis with the aim of comparing *Cuscuta* species and their host plants. The *rbcL* protein sequences were used to build a multiple alignment using MAFFT v7.520 with the iterative refinement method (Katoh, 2002). IQ-TREE version 2.2.2.6 was used to

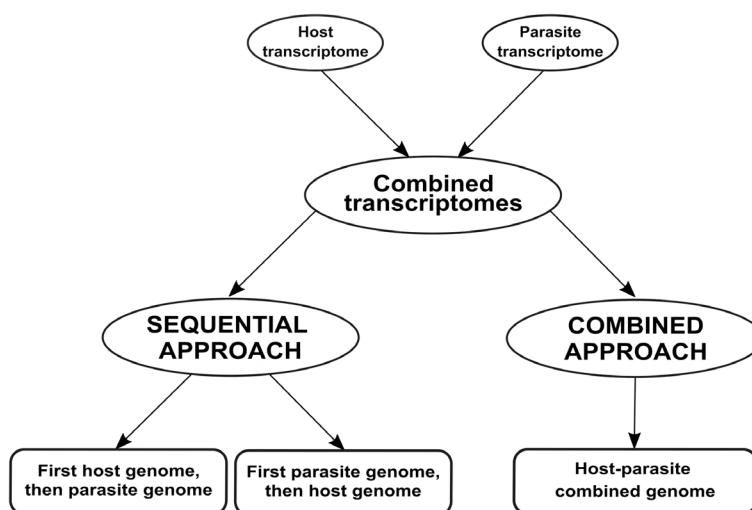


FIGURE 1

Workflow employed for analyzing dual RNA-seq data in this case study. We simulated a dual RNA-seq experiment by merging transcriptomes from two species involved in a parasitic relationship. Subsequently, we analyzed the merged transcriptomes using two approaches: sequential and combined. In the sequential approach, the merged transcriptomes are aligned first to the host plant and then to the parasitic plant (or vice versa). In the combined approach, the merged transcriptomes are aligned to a single genome that combines both host and parasite genomes.

construct a maximum-likelihood phylogenetic tree with 1,000 bootstrap replicates (Minh et al., 2020). The LG+I+G4 substitution model was identified as the best-fitting model for the analysis. Finally, the resulting tree was visualized via the R packages Treeio v1.26.0 and ggtree v3.10.1 (Wang et al., 2020; Yu, 2020). Detailed identifiers of the sequences used in tree construction can be found in Supplementary Table S1.

2.2 Retrieval of input data and their pre-processing

We considered *C. campestris* as parasitic plant and *A. thaliana* and *S. lycopersicum* as host plants.

The reference genomes of these species were retrieved from the European Nucleotide Archive (ENA) repository: *A. thaliana* (GCF_000001735.4), *S. lycopersicum* (GCF_000188115.5) and *C. campestris* (GCA_900332095.2).

For each species, we downloaded RNA-seq data from independent studies ensuring that three biological replicates were included for each (Table 1): *A. thaliana* data from the stem tissue (ENA acc. no.: SRR22559142, SRR22559143, SRR22559144) of the Columbia ecotype (Col-0) sampled during the vegetative stage (~20.1 M of reads on average); *S. lycopersicum* data from the stem tissue (ENA acc. no.: SRR25558913, SRR25558914, SRR25558915) of the Heinz 1706 cultivar (~43.2 M of reads on average); *C. campestris* data from developing haustoria (ENA acc. no.: SRR12763776, SRR12763787, SRR12763788), without host contact (~14.7 M of reads on average) (Bawin et al., 2022). All replicates were selected based on the use of NovaSeq 6000 sequencing technology and to ensure similar average read lengths (150 bases for the hosts and 100 bases for the parasite). This approach was chosen to minimize mapping performance biases related to sequencing technology platform. These RNA-Seq paired-end libraries were filtered using Trimmomatic version 0.39 with parameters: LEADING=20; TRAILING=20; SLIDING WINDOW=4 (Bolger et al., 2014). Only reads ≥ 75 nucleotides were retained.

2.3 Merging of the transcriptome data

We simulated a dual RNA-seq experiment in which the acquired reads included both host and parasitic plant sequences. Specifically, the first replicate of one species was merged with the

first replicate of the other one, and similarly for the other replicates. Consequently, the *C. campestris* transcriptome was merged with that of *A. thaliana* to create three merged transcriptome replicates (acc. no.: SRR22559142 + SRR12763776; SRR22559143 + SRR12763787; SRR22559144 + SRR12763788), which were used for downstream analysis. The same procedure was followed for *C. campestris* and *S. lycopersicum* (acc. no.: SRR25558913 + SRR12763776; SRR25558914 + SRR12763787; SRR25558915 + SRR12763788).

2.4 Dual RNA-Seq simulation via sequential and combined approach

In the sequential approach, host plant and parasitic plant genomes were individually indexed using STAR version 2.5.2b (Dobin et al., 2013). The merged pre-processed reads were mapped using STAR with parameters `-outFilterMultimapNmax 10` and `-outFilterMismatchNmax 5`, to minimize multiple mapping and cross-mapping issues. The reads were aligned to the host genome, and the resulting unmapped reads were then aligned to the parasite genome. The same procedure was repeated by swapping the order of mapping.

In the combined approach, host plant and parasitic plant genomes were first combined and then a single index was created. At this point, the merged pre-processed reads were mapped with STAR as previously described.

2.5 Evaluation metrics

Various performance metrics were computed for aligned reads. If one mate of a pair of reads mapped entirely while the other did not map at all, both mates were discarded and labelled as unmapped. As for the sequential, when considering reads initially mapped to the genome of the host plant, we defined correct assignment of host plant reads to the host plant genome as S-TP_{h1} and S-TN_{p1} (true positives for the host plant and true negatives for the parasitic plant). Conversely, incorrect assignment to the genome of the parasitic plant was labelled as S-FN_{h2} and S-FP_{p2} (false negatives for the host plant and false positives for the parasitic plant). Similarly, we defined correct assignment of parasitic plant reads to the parasitic plant genome as S-TN_{h2} and S-TP_{p2} (true negatives for the host plant and true

TABLE 1 The input RNA-seq data of the three selected species.

Plant species	ENA study accession number	ENA run accession number	Number of raw reads	Number of pre-processed reads
<i>A. thaliana</i>	PRJNA899009	SRR22559142, SRR22559143, SRR22559144	20248052, 20784415, 22169416	19336553, 19864492, 21188340
<i>S. lycopersicum</i>	PRJNA1003223	SRR25558913, SRR25558914, SRR25558915	43307357, 41684242, 49831874	41602968, 40278306, 47801733
<i>C. campestris</i>	PRJNA666991	SRR12763776, SRR12763787, SRR12763788	15454773, 15497676, 15864650	14613644, 14578626, 14978212

Specifically, the ENA bioproject and run accession numbers are reported along with the number of reads before and after the pre-processing step.

positives for the parasitic plant), while incorrect assignment to the plant host genome was designated as S-FP_{h1} and S-FN_{p1} (false positives for the host plant and false negative for the parasitic plant) (Figure 2).

When considering reads initially mapped to the genome of the parasitic plant, we defined correct assignment of parasitic plant reads to the parasitic plant genome as S-TP_{p1} and S-TN_{h1} (true positive for the parasitic plant and true negatives for the host plant). Conversely, incorrect assignment to the genome of the host plant was labelled as S-FN_{p2} and S-FP_{h2} (false negatives for the parasitic plant and false positives for the host plant). Correspondingly, we defined correct assignment of host plant reads to the genome of the host plant as S-TN_{p2} and S-TP_{h2} (true negatives for the parasitic plant and true positives for the host plant), while incorrect assignment to the genome of the parasitic plant was designated as S-FP_{p1} and S-FN_{h1} (false positives for the parasitic plant and false negative for the host plant). As for the combined approach, we defined correct assignment of plant host reads to the genome of host plant as C-TP_h and C-TN_p (true positives for the host plant and true negatives for the parasitic plant). Conversely, incorrect assignment to the genome of the parasitic plant was labelled as C-FN_h and C-FP_p (false negatives for the host plant and false positives for the parasitic plant). Correspondingly, we defined correct assignment of parasitic plant reads to the genome as C-TN_h and C-TP_p (true negatives for the host plant and true positives for the parasitic plant), while incorrect assignment to the genome of the host plant was designated as C-FP_h and C-FN_p (false positives for the host plant and false negatives for the parasitic plant). BAM files were processed using Samtools version 1.14 (Danecek et al., 2021).

Precision, sensitivity, specificity and accuracy metrics were calculated as follows:

$$\text{precision} = \frac{\text{truepositive}}{\text{truepositive} + \text{falsepositive}}$$

$$\text{sensitivity} = \frac{\text{truepositive}}{\text{truepositive} + \text{falsenegative}}$$

$$\text{specificity} = \frac{\text{truenegative}}{\text{truenegative} + \text{falsepositive}}$$

$$\text{accuracy} = \frac{\text{truepositive} + \text{truenegative}}{\text{truepositive} + \text{truenegative} + \text{falsepositive} + \text{falsenegative}}$$

2.6 Investigating one-side cross-mapped reads in the sequential and combined approaches

Exploration about one-side cross-mapped reads was performed considering reads labelled as S-FN_{h2} for host and S-FN_{p2} for parasite in sequential approach. Regarding combined approach, reads extractable through the complement C-FN_h (total cross-mapped reads) respect to C-TP_h (host reads correct assigned) were considered for host, while reads extractable through the complement C-FN_p (total cross-mapped reads) respect to C-TP_p (host reads correct assigned) were considered for parasite. Resulting reads were remapped to their respective genomes using STAR with less stringent parameters. Specifically, the parameters `-outFilterMultimapNmax 10` and `-outFilterMismatchNmax 10` were set. The statistical results were averaged between replicates. Summarization of these reads was achieved using the `htseq-count` function embedded in STAR version 2.5.2b and relative genome GFF3 annotation files.

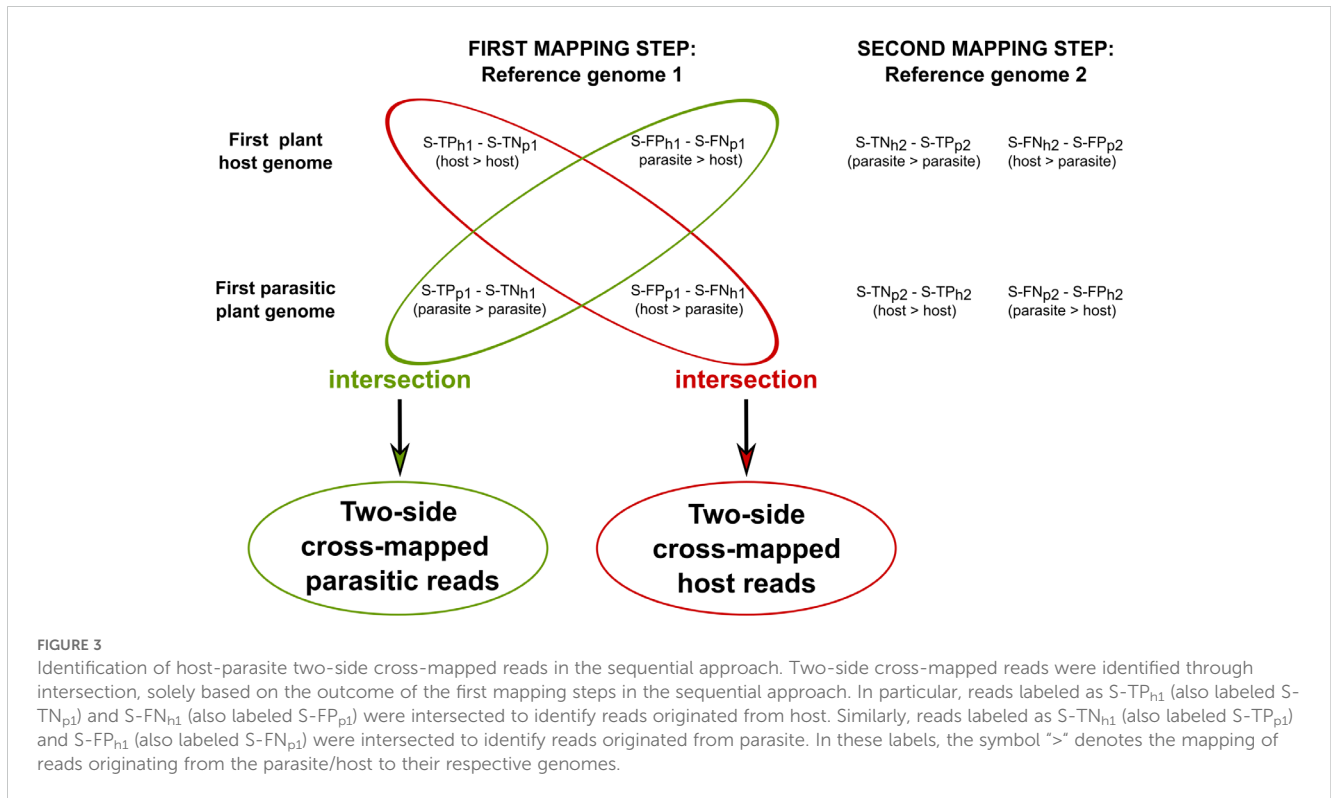
2.7 Investigating two-side cross-mapped reads in the sequential and combined approaches

To investigate host loci where two-side cross-mapped reads align, S-TP_{h1} and S-FN_{h1} were intersected in sequential approach. Similarly, S-TP_{p1} and S-FN_{p1} were intersected to identify two-side cross-mapped reads for parasitic loci (Figure 3).

		First mapping step		Second mapping step	
SEQ	First Host	H > H S-TP _{h1} [S-TN _{p1}]	P > H S-FP _{h1} [S-FN _{p1}]	P > P S-TN _{h2} [S-TP _{p2}]	H > P S-FN _{h2} [S-FP _{p2}]
	First Parasite	P > P S-TN _{h1} [S-TP _{p1}]	H > P S-FN _{h1} [S-FP _{p1}]	H > H S-TP _{h2} [S-TN _{p2}]	P > H S-FP _{h2} [S-FN _{p2}]
		Single Mapping step			
COM		H > H C-TP _h [C-TN _p]	P > H C-FP _h [C-FN _p]	P > P C-TN _h [C-TP _p]	H > P C-FN _h [C-FP _p]

FIGURE 2

Categorization of reads based on the employed mapping procedure. The "SEQ" row illustrates the sequential approach, detailing mappings for both host and parasitic plants, separately. In contrast, the "COM" row represents the combined approach, where a single mapping encompasses both host and parasitic plants. The mapping of reads originated from the parasite/host to their respective genomes is indicated by the symbol ">", with the assigned labels resulting from the mapping displayed below. Labels without brackets refer to the host plant, while those within brackets refer to the parasitic plant. For example, to indicate reads belonging to parasitic plant but mapping to the genome of the host plant in the first mapping step "P > H", the label "S-FP_{h1}" (Sequential - False Positive host plant first mapping step) is used for the host plant reference, and "S-FN_{p1}" (Sequential - False Negative parasitic plant first mapping step) is used for the parasitic plant reference.



These reads were summarized using the htseq-count function embedded in STAR version 2.5.2b and relative genome GFF3 annotation files. From these, the relative functional description attributes (the “product” tag) were extracted.

In the combined approach, two-side cross-mapped reads were identified by intersecting C-TP_h with C-FN_h for host loci, and C-TP_p and C-FN_p for parasite loci. Next, the functional description attributes (the “product” tag) of each identified transcript were then extracted from the corresponding GFF3 annotation files.

3 Results

3.1 Phylogenetic analysis of *Cuscuta* spp. and their host range

We performed a phylogenetic analysis based on the rbcL protein sequences from six *Cuscuta* species and 28 host species. In the resulting tree (Figure 4), we identified a large clade supported by a high bootstrap value (0.93) that comprises the parasitic plant *C. campestris*, other *Cuscuta* species and six host species. Five of these species belonged to the Solanales order, including the crop *S. lycopersicum*, along with one species from the Convolvulus order. Other host species of interest, such as *A. thaliana* from the Brassicales order, were clearly phylogenetically distant from the clade that includes *Cuscuta* species. Based on this, we considered *S. lycopersicum* as host species phylogenetically close to *C. campestris*, whereas *A. thaliana* represents a host with a more divergent evolutionary relationship.

3.2 Dual RNA-seq simulation via sequential and combined approaches

After quality filtering and library merging, the three resulting merged libraries involving *A. thaliana* and *C. campestris* included an average of ~34.85 M reads, whereas the three resulting merged libraries involving *S. lycopersicum* and *C. campestris* included an average of ~57.95 M reads.

3.2.1 Sequential approach

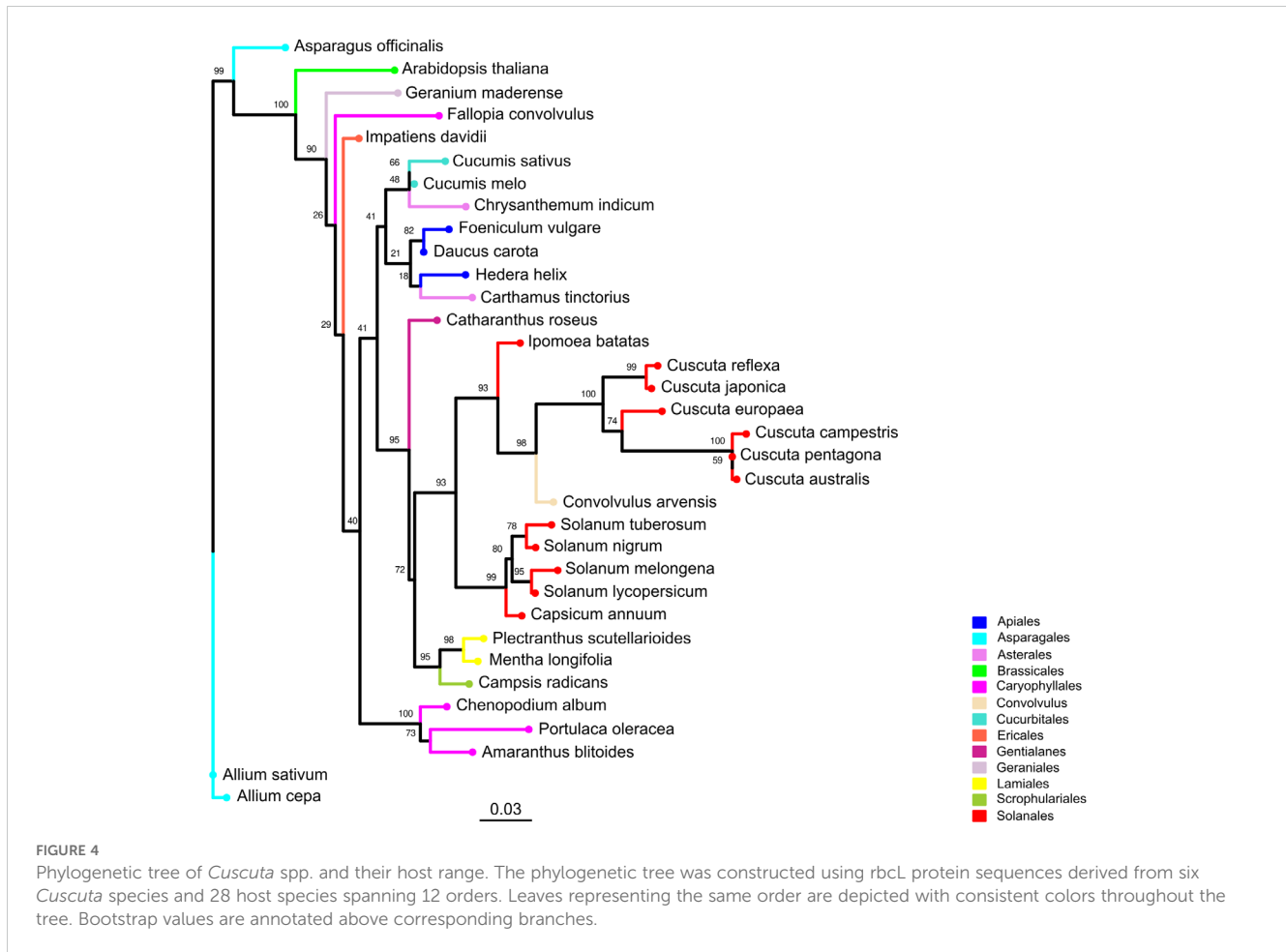
3.2.1.1 Mapping of *A. thaliana* and *C. campestris* reads

Mapping first to host and then to parasite genome, approximately 19.81 M reads were assigned to *A. thaliana* and 14.22 M reads to *C. campestris*. On average, uniquely mapped reads assigned to host were ~18.99 M (~54.49%) while ~13.06 M (~37.46%) were assigned to parasite.

Multiple mapped reads assigned to host were ~0.82 M (~2.35%) and ~1.17 M (~3.35%) were assigned to parasite (Table 2.1; Figure 5).

Among the mapped reads, the total cross-mapped reads (both one-side and two-side), originating from *A. thaliana* but assigned to *C. campestris* were ~0.004 M (~0.002%). Conversely, reads originating from *C. campestris* and assigned to *A. thaliana* were ~0.13 M (~0.64%) (Supplementary Tables S2, S3). Notably, the evaluation metrics (i.e., precision, sensitivity, accuracy, and specificity) were all close to one (Table 2.2).

We repeated the analysis by swapping the order of mapping; in this scenario, reads were initially mapped to the *C. campestris* genome and then to the *A. thaliana* genome.



Approximately 19.24 M reads were assigned to *A. thaliana* and 14.79 M reads to *C. campestris*. On average, the uniquely mapped reads assigned to the host totalled ~18.78 M (~53.87%), while ~13.32 M (~38.22%) were assigned to the parasite. The multiple mapped reads assigned to the host accounted for ~0.47 M (~1.34%), whereas those assigned to parasite were ~1.47 M (~4.22%) (Table 2.1; Figure 5). Among mapped reads, the total cross-mapped (both one-side and two-side), originating from *A. thaliana* and assigned to *C. campestris* amounted to ~0.49 M (~3.33%). Conversely, reads originating from *C. campestris* and assigned to *A. thaliana* were ~0.05 M (~0.28%) (Supplementary Tables S2, S3). Also in this case, the evaluation metrics were close to one (Table 2.2).

3.2.1.2 Mapping of *S. lycopersicum* and *C. campestris* reads

Mapping first to host and then to parasite genome, approximately 37.52 M of reads were assigned to *S. lycopersicum* and 14.38 M reads to *C. campestris*. In the scenario where reads were initially mapped to the *S. lycopersicum* genome, on average, the uniquely mapped reads assigned to the host were ~36.97 M (~63.79%), while ~13.1 M (~89.0%) were assigned to the parasite. The multiple mapped reads assigned to the host accounted for ~0.6 M (~1.3%), while those assigned to the parasite were ~1.3 M (~8.7%) (Table 3.1; Figure 5). Among mapped reads, the total

cross-mapped (both one-side and two-side), originating from *S. lycopersicum* and assigned to *C. campestris*, were ~0.13 M reads (~0.92%). Conversely, reads originating from *C. campestris* and assigned to *S. lycopersicum* were ~0.09 M reads (~0.25%) (Supplementary Tables S2, S4). Ultimately, all statistical metrics approached unity (Table 3.2).

Reversing the order of use of the reference genomes, approximately 37.33 M reads were assigned to *S. lycopersicum*, while 14.58 M reads to *C. campestris*. Among these, the uniquely mapped reads assigned to the host totaled ~36.8 M (~85.2%), whereas those assigned to the parasite were ~13.2 M (~89.4%). The multiple mapped reads assigned to the host accounted for ~0.5 M (~1.1%), while those assigned to the parasite were ~1.4 M (~9.6%) (Table 3.1; Figure 5).

The total cross-mapped reads (both one-side and two-side), originating from *S. lycopersicum* and assigned to *C. campestris* amounted to ~0.28 M (~1.93%). Conversely, reads originating from *C. campestris* and assigned to *S. lycopersicum* were ~0.05 M (~0.13%) (Supplementary Tables S2, S4). Consistent with previous mapping results, all statistical metrics approached unity (Table 3.2).

3.2.2 Combined approach

3.2.2.1 Mapping of *A. thaliana* and *C. campestris* reads

When aligning reads to the combined genome of *A. thaliana* and *C. campestris*, approximately 19.68 M reads were assigned to *A.*

TABLE 2.1 Statistics on the number of mapped reads by considering a sequential approach.

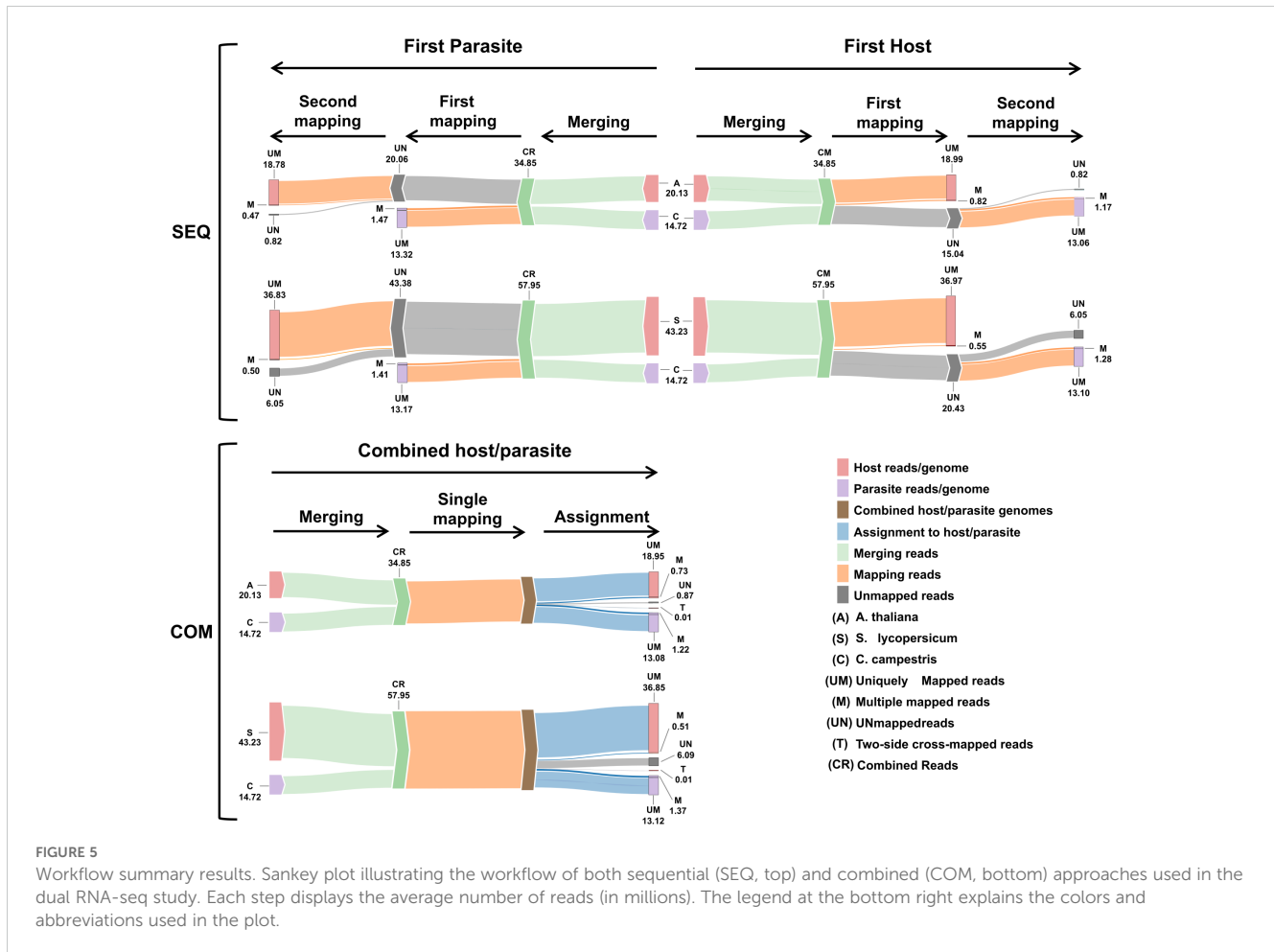
Libraries (<i>A. thaliana</i> + <i>C. campestris</i>)	Replicate	Order mapping	Processed reads	Uniquely mapped to host genome	Multiple mapped to host genome	Uniquely mapped to parasite genome	Multiple mapped to parasite genome	Unmapped
SRR22559142 + SRR12763776	Replicate 1	Host; parasite	33950197	18603604 (55%)	410240 (1%)	12948080 (38%)	1165669 (3%)	822604 (2%)
SRR22559143 + SRR12763787	Replicate 2	Host; parasite	34443118	18102724 (53%)	1618200 (5%)	12953498 (38%)	1126992 (3%)	641704 (2%)
SRR22559144 + SRR12763788	Replicate 3	Host; parasite	36166552	20268051 (56%)	426377 (1%)	13270916 (37%)	1208650 (3%)	992558 (3%)
SRR22559142 + SRR12763776	Replicate 1	Parasite; host	33950197	18564026 (55%)	327509 (1%)	12992108 (38%)	1243959 (4%)	822595 (2%)
SRR22559143 + SRR12763787	Replicate 2	Parasite; host	34443118	17553088 (51%)	752222 (2%)	13639241 (40%)	1857011 (5%)	641556 (2%)
SRR22559144 + SRR12763788	Replicate 3	Parasite; host	36166552	20210535 (56%)	324911 (1%)	13331684 (37%)	1306968 (4%)	992454 (3%)

Reads were first mapped to the *A. thaliana* (host) genome and then to the *C. campestris* (parasite) genome (Order mapping: "Host; parasite") or vice versa.

TABLE 2.2 Evaluation metrics including precision, sensitivity, accuracy, and specificity were used to assess read mapping based on the sequential approach applied to *A. thaliana* (host) and *C. campestris* (parasite) interaction.

Libraries (<i>A. thaliana</i> + <i>C. campestris</i>)	Replicate	Order mapping	Processed reads	Mapped	Precision (host)	Sensitivity (host)	Accuracy (host)	Specificity (host)	Precision (parasite)	Sensitivity (parasite)	Accuracy (parasite)	Specificity (parasite)
SRR22559142 + SRR12763776	Replicate 1	Host; parasite	33950197	33127593 (97.58%)	0.9940452	0.9999979	0.9965810	0.9920416	0.9999972	0.9920416	0.9965810	0.9999979
SRR22559143 + SRR12763787	Replicate 2	Host; parasite	34443118	33801414 (98.14%)	0.9929953	0.9999623	0.9958913	0.9902841	0.9999475	0.9902841	0.9958913	0.9999623
SRR22559144 + SRR12763788	Replicate 3	Host; parasite	36166552	35173994 (97.26%)	0.9936527	0.9999860	0.9962574	0.9910098	0.9999801	0.9910098	0.9962574	0.9999860
SRR22559142 + SRR12763776	Replicate 1	Parasite; host	33950197	33127602 (97.58%)	0.9974339	0.9969504	0.9967967	0.9965926	0.9959512	0.9965926	0.9967967	0.9969504
SRR22559143 + SRR12763787	Replicate 2	Parasite; host	34443118	33801562 (98.14%)	0.9968245	0.9317532	0.9587400	0.9959116	0.9137517	0.9959116	0.9587400	0.9317532
SRR22559144 + SRR12763788	Replicate 3	Parasite; host	36166552	35174098 (97.26%)	0.9972441	0.9958850	0.9959853	0.9961265	0.9942195	0.9961265	0.9959853	0.9958850

Each metric was calculated and reported separately for both the host and the parasite.



thaliana, while 14.30 M reads to *C. campestris*. Among these, the uniquely mapped reads assigned to host totaled ~18.95 M (~54.37%) while for the parasite they were ~13.08 M (~37.53%). The multiple mapped reads assigned to host accounted for ~0.73 M (~2.10%), whereas those assigned to the parasite were ~1.22 M (~3.49%) (Table 4.1; Figure 5). The total cross-mapped reads (both one-side and two-side), originating from *A. thaliana* and assigned to *C. campestris* amounted to ~0.01 M (~0.03%). Conversely, reads originating from *C. campestris* and assigned to *A. thaliana* were ~0.01 M (~0.04%) (Supplementary Tables S5, S6). Ultimately, all evaluation metrics were close to unity (Table 4.2).

3.2.2.2 Mapping of *S. lycopersicum* and *C. campestris* reads

When aligning reads to the combined genome of *S. lycopersicum* and *C. campestris*, approximately 37.35 M of reads were assigned to *S. lycopersicum*, while 14.50 M reads to *C. campestris*. Among these, the uniquely mapped reads assigned to host accounted for ~36.85 M (~63.58%) while for the parasite they were ~13.12 M (~22.65%). The multiple mapped reads assigned to the host totaled ~0.51 M (~0.87%), whereas those assigned to the parasite were ~1.37 M (~2.37%) (Table 4.1; Figure 5). The total cross-mapped (both one-side and two-side), originating from *S. lycopersicum* and assigned to *C. campestris* amounted to ~0.21 M (~0.56%). Conversely, reads

originating from *C. campestris* and assigned to *S. lycopersicum* were ~0.01 M (~0.05%) (Supplementary Tables S5, S6). Finally, all evaluation metrics were close to unity (Table 4.2).

3.3 Investigating one-side cross-mapped reads

On average, in the sequential approach involving *A. thaliana* and *C. campestris*, the number of reads labelled as S-FN_{h2} were not significant, while S-FN_{p2} accounted for ~0.05 M reads; of these only ~1.83% successfully remapped to the *C. campestris* genome. Similarly, when considering *S. lycopersicum* and *C. campestris*, the number of reads labelled as S-FN_{h2} were ~0.13 M, while S-FN_{p2} accounted for ~0.05 M; of these only ~59.82% and ~1.92% successfully remapped to the *S. lycopersicum* and *C. campestris* genomes, respectively (Supplementary Table S7).

In the combined approach involving *A. thaliana* and *C. campestris*, the number of reads labelled as C-FN_h and C-FN_p were both not significant. For *S. lycopersicum* and *C. campestris*, the number of reads labelled as C-FN_h were ~0.20 M, while the number of reads labelled as C-FN_p were not significant. The remapping rate of C-FN_h reads to *S. lycopersicum* were ~76.53%. Moreover, reads did not map to annotated loci (Supplementary Table S8).

TABLE 3.1 Statistics on the number of mapped reads by considering a sequential approach.

Libraries (<i>S. lycopersicum</i> + <i>C. campestris</i>)	Replicate	Order mapping	Processed reads	Uniquely mapped to the host genome	Multiple mapped to the host genome	Uniquely mapped to the parasite genome	Multiple mapped to the parasite genome	Unmapped
SRR25558913 + SRR12763776	Replicate 1	Host; parasite	56216612	35405921 (63%)	495390 (1%)	13001831 (23%)	1292197 (2%)	6021273 (11%)
SRR25558914 + SRR12763787	Replicate 2	Host; parasite	54856932	35128009 (64%)	538023 (1%)	12993036 (24%)	1226568 (2%)	4971296 (9%)
SRR25558915 + SRR12763788	Replicate 3	Host; parasite	62779945	40372884 (64%)	625278 (1%)	13314939 (21%)	1321988 (2%)	7144856 (11%)
SRR25558913 + SRR12763776	Replicate 1	Parasite; host	56216612	35253494 (63%)	441975 (1%)	13066350 (23%)	1433890 (3%)	6020903 (11%)
SRR25558914 + SRR12763787	Replicate 2	Parasite; host	54856932	35008504 (64%)	485523 (1%)	13052054 (24%)	1339693 (2%)	4971158 (9%)
SRR25558915 + SRR12763788	Replicate 3	Parasite; host	62779945	40238892 (64%)	562252 (1%)	13378826 (21%)	1454850 (2%)	7145125 (11%)

Reads were first mapped to the *S. lycopersicum* (host) genome and then to the *C. campestris* genome (Order mapping: "Host; parasite) or vice versa.

TABLE 3.2 Evaluation metrics including precision, sensitivity, accuracy, specificity were used to assess read mapping based on the sequential approach applied to *S. lycopersicum* (host) and *C. campestris* (parasite) interaction.

Libraries (<i>S. lycopersicum</i> + <i>C. campestris</i>)	Replicate	Order mapping	Processed reads	Mapped	Precision (host)	Sensitivity (host)	Accuracy (host)	Specificity (host)	Precision (parasite)	Sensitivity (parasite)	Accuracy (parasite)	Specificity (parasite)
SRR25558913 + SRR12763776	Replicate 1	Host; parasite	56216612	50195339 (89.29%)	0.9976894	0.9956543	0.9952329	0.9941667	0.9976894	0.9956543	0.9952329	0.9941667
SRR25558914 + SRR12763787	Replicate 2	Host; parasite	54856932	49885636 (90.94%)	0.9971581	0.9969043	0.9957543	0.9928673	0.9971581	0.9969043	0.9957543	0.9928673
SRR25558915 + SRR12763788	Replicate 3	Host; parasite	62779945	55635089 (88.62%)	0.9976331	0.9968478	0.9959311	0.9933556	0.9976331	0.9968478	0.9959311	0.9933556
SRR25558913 + SRR12763776	Replicate 1	Parasite; host	56216612	50195709 (89.29%)	0.9988087	0.9910511	0.9927392	0.9970097	0.9988087	0.9910511	0.9927392	0.9970097
SRR25558914 + SRR12763787	Replicate 2	Parasite; host	54856932	49885774 (90.94%)	0.9985610	0.9934882	0.9943193	0.9964058	0.9985610	0.9934882	0.9943193	0.9964058
SRR25558915 + SRR12763788	Replicate 3	Parasite; host	62779945	55634820 (88.62%)	0.9987852	0.9931971	0.9940920	0.9966060	0.9987852	0.9931971	0.9940920	0.9966060

Each metric was calculated and reported separately for both the host and the parasite.

TABLE 4.1 Statistics on the number of mapped reads by considering combined approach.

Interaction	Libraries (host + parasite)	Replicate	Processed reads	Uniquely mapped to the host genome	Multiple mapped to the host genome	Uniquely mapped to the parasite genome	Multiple mapped to the parasite genome	Two-side cross-mapped	Unmapped
<i>Arabidopsis-Cuscuta</i>	SRR22559142 + SRR12763776	Replicate 1	33950197	18572122 (54.7%)	330452 (0.97%)	12965830 (38.19%)	1211131 (3.57%)	1709 (0.01%)	868953 (2.56%)
<i>Arabidopsis-Cuscuta</i>	SRR22559143 + SRR12763787	Replicate 2	34443118	18046908 (52.4%)	1526290 (4.43%)	12979800 (37.68%)	1178915 (3.42%)	13583 (0.04%)	697622 (2.03%)
<i>Arabidopsis-Cuscuta</i>	SRR22559144 + SRR12763788	Replicate 3	36166552	20227851 (55.93%)	337908 (0.93%)	13292023 (36.75%)	1260436 (3.49%)	2831 (0.01%)	1045503 (2.89%)
<i>Solanum-Cuscuta</i>	SRR25558913 + SRR12763776	Replicate 1	56216612	35261137 (62.72%)	452535 (0.8%)	13019757 (23.16%)	1396199 (2.48%)	8192 (0.01%)	6078792 (10.81%)
<i>Solanum-Cuscuta</i>	SRR25558914 + SRR12763787	Replicate 2	54856932	35013923 (63.83%)	490718 (0.89%)	13016735 (23.73%)	1307742 (2.38%)	7000 (0.01%)	5020814 (9.15%)
<i>Solanum-Cuscuta</i>	SRR25558915 + SRR12763788	Replicate 3	62779945	40262060 (64.13%)	575318 (0.92%)	13335341 (21.24%)	1415630 (2.25%)	8307 (0.01%)	7183289 (11.44%)

Reads were mapped to the host plant-parasitic plant combined genome. The reported two-side cross-mapped reads represent the cumulative count between the two organisms.

TABLE 4.2 Evaluation metrics including precision, sensitivity, accuracy, specificity were used to assess read mapping based on the combined approach.

Interaction	Libraries (host + parasite)	Replicate	Processed reads	Mapped	Precision (host)	Sensitivity (host)	Accuracy (host)	Specificity (host)	Precision (parasite)	Sensitivity (parasite)	Accuracy (parasite)	Specificity (parasite)
<i>Arabidopsis-Cuscuta</i>	SRR22559142 + SRR12763776	Replicate 1	33950197	33079535 (97.44%)	0.9998379	0.9999910	0.9999932	0.9997838	0.9999035	0.9999035	0.9997838	0.9999932
<i>Arabidopsis-Cuscuta</i>	SRR22559143 + SRR12763787	Replicate 2	34443118	33731913 (97.94%)	0.9997350	0.9998929	0.9999225	0.9996337	0.9998013	0.9998013	0.9996337	0.9999225
<i>Arabidopsis-Cuscuta</i>	SRR22559144 + SRR12763788	Replicate 3	36166552	35118218 (97.1%)	0.9997497	0.9999470	0.9999625	0.9996464	0.9998315	0.9998315	0.9996464	0.9999625
<i>Solanum-Cuscuta</i>	SRR25558913 + SRR12763776	Replicate 1	56216612	50129628 (89.17%)	0.9999063	0.9832704	0.9932917	0.9997641	0.9951223	0.9951223	0.9997641	0.9932917
<i>Solanum-Cuscuta</i>	SRR25558914 + SRR12763787	Replicate 2	54856932	49829118 (90.83%)	0.9998413	0.9881445	0.9952389	0.9996021	0.9964788	0.9964788	0.9996021	0.9952389
<i>Solanum-Cuscuta</i>	SRR25558915 + SRR12763788	Replicate 3	62779945	55588349 (88.54%)	0.9998616	0.9863561	0.9950951	0.9996118	0.9962778	0.9962778	0.9996118	0.9950951

Each metric was calculated and reported separately for both the host and the parasite.

3.4 Investigating two-side cross-mapped reads

3.4.1 Sequential approach

3.4.1.1 *A. thaliana* and *C. campestris*

The intersection between *A. thaliana* S-TP_{h1} reads and *C. campestris* S-FN_{h1} reads yielded an average of 0.49 M reads. Similarly, the intersection between *C. campestris* S-TP_{p1} reads and *A. thaliana* S-FN_{p1} reads resulted in an average of 0.07 M reads.

The resulting loci span across all five *A. thaliana* chromosomes (including organelle genomes) were reported in [Supplementary Table S9](#). None of the loci were annotated in *C. campestris*.

3.4.1.2 *S. lycopersicum* and *C. campestris*

The intersection between *S. lycopersicum* S-TP_{h1} reads and *C. campestris* S-FN_{h1} reads resulted in an average of 0.17 M reads. Similarly, the intersection between *C. campestris* S-TP_{p1} reads and *S. lycopersicum* S-FN_{p1} reads yielded an average of 0.04 M reads. The resulting loci span across all twelve *S. lycopersicum* chromosomes (including organelle genomes) were reported in [Supplementary Table S10](#). None of the loci were annotated in *C. campestris*.

3.4.2 Combined approach

3.4.2.1 *A. thaliana* and *C. campestris*

On average of 0.005 M reads from *A. thaliana* and 0.001 M reads from *C. campestris* were tagged as two-side cross-mapped reads, by intersecting C-TP_h with C-FN_h for the host and C-TP_p with C-FN_p for the parasite. Annotated loci resulting from this process were identified on chromosome 2, chromosome 3, and the mitochondrial and plastidial genomes of *A. thaliana* ([Supplementary Table S11](#)). None of the loci were annotated in *C. campestris*.

3.4.2.2 *S. lycopersicum* and *C. campestris*

On average 0.01 M reads from *S. lycopersicum* and 0.002 M reads from *C. campestris* were tagged as two-side cross-mapped reads, by intersecting C-TP_h with C-FN_h for the host and C-TP_p with C-FN_p for the parasite. Annotated loci resulting from this process were identified on chromosomes 3 and 11, and the mitochondrial genome of *S. lycopersicum* ([Supplementary Table S12](#)). None of the loci were annotated in *C. campestris*.

4 Discussions

Transcriptome investigations on host plant-parasitic plant interaction have typically involved separate analysis of the two organisms. For this purpose, the use of LCM to isolate the host from the parasite has represented the standard in the field ([Jhu and Sinha, 2022](#)). While valuable, LCM is costly, time-consuming, and requires specialized expertise limiting its accessibility. For this reason, *in silico* separation of transcripts via dual-RNA-seq offers a promising alternative.

Dual RNA-seq allows for the identification of core genes involved in the interaction by sampling and analyzing infected

tissues. Examining gene expression changes in the parasite alongside the host response, uncovers critical mechanisms in the host-parasite interplay. This method is economical and practical since the tissues do not need to be physically separated. It requires that reads must be assigned through read mapping onto their respective reference genomes. If one species lacks an assembled genome, its assembled transcriptome can be used. Anyway, the choice of genome as reference, instead of assembled transcriptome, is required in order to avoid the potential presence of transferred transcript. Two approaches are used: the sequential method, where reads are mapped sequentially to both species reference genomes, and the combined method, where reads are mapped to a single concatenated genome. The sequential method can be used in the presence of almost one reference genome (parasite or host), while the combined approach is exclusively employed when both host and parasitic reference genomes are accessible. Dual RNA-seq complexity arises from handling sequences from two organisms simultaneously, with a key challenge being cross-mapping reads; these latter could represent misleading or loss information. Here, we classified cross-mapping events into two main categories, namely “one-side cross-mapping” and “two-side cross-mapping”. The first indicates reads from one organism but exclusively assigned to the other, while the second refers to reads ambiguously assigned to both organisms.

Although distinguishing reads originating from two eukaryotes poses additional challenges, dual RNA-seq has been successfully applied to study plant-pathogen interactions across various plant species and eukaryotic pathogens and parasites, including fungi, oomycetes, and nematodes ([Petitot et al., 2016](#); [Naidoo et al., 2018](#)). Despite demonstrated utility, performing this analysis between two plants is still relatively uncharted territory, which may seem daunting at first glance. In fact, the key to separating reads from bacterial and eukaryotic cells lies in their divergence and distinct content of their RNA molecules ([Marsh et al., 2018](#); [Westermann et al., 2017](#)). A previous study by Ikeue et al. ([Ikeue et al., 2015](#)) attempted to separate *in silico* reads from two distinct plant species, *Impatiens balsamina* and *Cuscuta japonica*. They exclusively used sequential approach without prior knowledge of sequence origin.

Our study aims to evaluate dual RNA-seq feasibility between host and parasite within the same taxonomic kingdom. Using assembled genomes as references, we employed both sequential and combined approaches. We generated artificial datasets to replicate interaction of two host-parasite systems: *A. thaliana*-*C. campestris* and *S. lycopersicum*-*C. campestris*.

A. thaliana, a *Brassicaceae* family member, and *S. lycopersicum*, a *Solanaceae* family member, were chosen to investigate whether a host plant, phylogenetically further from *C. campestris* (member of the *Convolvulaceae* family), could improve read assignment accuracy due to sequence divergence. A phylogenetic analysis using the sequences of the rbcL protein ([Manhart, 1994](#)), commonly used in plant phylogenetics, determined the evolutionary distance between host and parasite. This analysis serves as preliminary step in comparing the extent of cross-mapping in two host plants with varying evolutionary distances from the parasitic plant. As anticipated, *S. lycopersicum* is phylogenetically closer to *C. campestris* than *A. thaliana*. These findings align with previous studies supporting the

monophyletic nature of *Convolvulaceae* family as sister group of *Solanaceae* family, placing *Convolvulaceae* and *Brassicaceae* into distinct clades ('Classification and System in Solanales', 2008; Soltis et al., 2011).

Accurately assigning reads to their respective genomes could be challenging, especially when interacting organisms belong to the same taxonomic kingdom. This challenge stems from the presence of homologous sequences, which are often highly similar. To verify this hypothesis, the RNA-seq libraries were chosen to utilize the maximum read length available in ENA repository, in order to enhance the accuracy of sequence assignment to their respective reference genomes. The libraries come from tissues that could be attacked during infection process, namely the stem for the two host plants and the dodder haustoria, closely mimicking genuine parasitic conditions. Furthermore, to prevent read contaminations between host into the parasitic and vice versa, due to RNA transfer phenomenon during the infection, we selected samples collected when they did not interact, ensuring confident attribution of reads to their sources. We combined the host and parasitic plant transcriptomes to simulate two distinct interactions, allowing accurate assessment of mapped reads, multiple mapped reads, cross-mapped reads, and computation of evaluation metrics: precision, sensitivity, specificity and accuracy. Both interactions were thoroughly analyzed through sequential and combined approaches using reference genome of interested species. Typically, RNA-seq analysis yields mapping percentages ranging 70%-90% (Conesa et al., 2016). In this study, alignment of simulated mixed RNA-seq resulted in high mapping rates (around 90% in both approaches), demonstrating the method's effectiveness. Our findings suggest that, in sequential approach, when merged reads were firstly mapped to the host genome, this latter tends to retain a little percentage of reads (cross-mapping near 1%) belonging to the parasite and vice versa; however, in the combined approach, this trend was less pronounced (cross-mapping less than 0.2%). This variation can be attributed to the alignment step being a single operation in the combined approach. Thus, the mapping tool selects the genome to which each read aligns best, leveraging homologous sequence polymorphisms between the two species (Wolf et al., 2018; Espindula et al., 2019; O'Keeffe and Jones, 2019). The single operation of the combined approach is less time-consuming. Specifically, we observed a halving of the processing time in the combined approach compared to the sequential approach.

Moreover, combined approach offers a further advantage. It is possible to isolate the two-side cross-mapped reads after a single mapping step. Differently, sequential approach needs a bit intricates strategy, previously requiring almost three mapping operations and swapping the order of reference genomes. The sequential and combined approach exhibit minimal differences in the number of multiple mapped reads. Despite these differences, all evaluation metrics indicate near parity, affirming the equal reliability of both methods. No significant differences were observed in cross-mapping percentages between reads of *S. lycopersicum* (phylogenetically closer to *C. campestris* than *A. thaliana*) and *A. thaliana* when combined with those of *C.*

campestris thanks to the stringent parameters adopted. By utilizing these parameters, we effectively addressed the struggle highlighted by O'Keeffe and Jones (2019), namely the direct relationship between read discrimination among host-pathogen species and their taxonomic divergence. Another intriguing observation, from both the sequential and combined approach pertains to the failure of a few reads labelled S-FN_{h2}, S-FN_{p2}, C-FN_h and C-FN_p (one-side cross-mapped reads) to align with their respective genomes. This could stem from either incomplete genome assemblies or excessively stringent mapping criteria. Our results indicate that relaxing these criteria would result in reallocating over 50% of previously unaligned reads to the respective genome. Notably, during the remapping process for *C. campestris* a significantly lower percentage of reads were reallocated, suggesting that genome incompleteness may be the primary factor in this instance.

We examined loci and the relative annotation, to evaluate the extent of misleading or loss information, when similar reads map equally well in both genomes (two-side cross-mapping). In the sequential approach, we identified a few hundred thousand two-side cross-mapped reads, whereas in the combined approach, we observed a reduction of over 90%.

These loci were annotated only for the hosts due to potential incompleteness in *C. campestris* annotation. They span all chromosomes, including the organelle genomes, and are involved in various cellular processes, such as protein synthesis, respiration, and some are annotated as non-coding RNA. All genes are part of the plant basal metabolism, suggesting they are unlikely to be crucial in the interaction. Moreover, given their low abundance, this loss of information may not be significant.

Our study underscores the reliability of dual RNA-seq as an effective choice, particularly when the host and parasite belong to the same kingdom. This insight paves the way to new experiments expanding our understanding of the key genes involved in plant parasitism.

5 Conclusions

Dual RNA-seq can be applied to host plant – parasitic plant interaction, providing a reliable *in silico* separation of mixed reads. Cross-mapped reads, due to homologous sequences between the two genomes, did not represent a significant misleading information. This study establishes that sequential and combined approach are both equally trustworthy. However, combined approach results to be less time-consuming, and slightly mitigates the cross-mapping phenomenon.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

CF: Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft. GA: Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft. DD: Formal analysis, Writing – review & editing. EP: Investigation, Writing – review & editing. ND: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was carried out in the frame of Programme STAR Plus, financially supported by UniNA and Compagnia di San Paolo. Carmine Fruggiero's PhD was funded by the Italian Ministry of Education, University, and Research through the *PON Ricerca e Innovazione 2014-2020* initiative (D.M. n. 1061, 10-08-2021).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1483717/full#supplementary-material>

SUPPLEMENTARY TABLE 1

Protein sequences accession numbers involved in the phylogenetic analysis. List of NCBI accession numbers for the *rbcL* proteins of *Cuscuta* spp. and their host plants. Six *Cuscuta* species were selected, along with 28 hosts spanning twelve different orders, categorized as crops, ornamentals, weeds, and model organisms.

SUPPLEMENTARY TABLE 2

Sequential approach cross-mapping. Summary of both one-side and two-side cross-mapped reads in the sequential approach.

SUPPLEMENTARY TABLE 3

A. thaliana and *C. campestris* cross-mapping in sequential approach. Statistics on the number of correctly assigned reads and cross-mapped reads (both one-side and two-side) in the sequential approach applied to *A. thaliana* and *C. campestris*.

SUPPLEMENTARY TABLE 4

S. lycopersicum and *C. campestris* cross-mapping in sequential approach. Statistics on the number of correctly assigned reads and cross-mapped reads (both one-side and two-side) in the sequential approach applied to *S. lycopersicum* and *C. campestris*.

SUPPLEMENTARY TABLE 5

Cross-mapping in sequential approach. Summary of both one-side and two-side cross-mapped reads in the combined approach.

SUPPLEMENTARY TABLE 6

Cross-mapping in combined approach. Statistics on the number of correctly assigned reads and cross-mapped reads (both one-side and two-side) in the combined approach.

SUPPLEMENTARY TABLE 7

One-side cross-mapping in sequential approach. Remapping result of one-side cross-mapped read to their respective genomes for the sequential approach involving *A. thaliana*-*C. campestris* and *S. lycopersicum*-*C. campestris* interactions.

SUPPLEMENTARY TABLE 8

One-side cross-mapping in combined approach. Remapping result of one-side cross-mapped read to their respective genomes for combined approach involving *A. thaliana*-*C. campestris* and *S. lycopersicum*-*C. campestris* interactions.

SUPPLEMENTARY TABLE 9

A. thaliana and *C. campestris* two-side cross-mapping in sequential approach. List of *A. thaliana* loci identified investigating on two-side cross-mapped reads from the *A. thaliana*-*C. campestris* interaction via the sequential approach. Chromosome (Chr), locus identifier (Parent) and description of the gene function (Product) were extracted from *A. thaliana* genome annotation (gff3 file) were reported.

SUPPLEMENTARY TABLE 10

S. lycopersicum and *C. campestris* two-side cross-mapping in sequential approach. List of *S. lycopersicum* loci identified investigating on two-side cross-mapped reads from the *S. lycopersicum*-*C. campestris* interaction via the sequential approach. Chromosome (Chr), locus identifier (Parent) and description of the gene function (Product) were extracted from *S. lycopersicum* genome annotation (gff3 file) and reported.

SUPPLEMENTARY TABLE 11

A. thaliana and *C. campestris* two-side cross-mapping in sequential approach. List of *A. thaliana* loci identified investigating on two-side cross-mapped reads from *A. thaliana*-*C. campestris* interaction via the combined approach. Chromosome (Chr), locus identifier (Parent) and description of the gene function (Product) were extracted from *A. thaliana* genome annotation (gff3 file) and reported.

SUPPLEMENTARY TABLE 12

S. lycopersicum and *C. campestris* two-side cross-mapping in combined approach. List of *S. lycopersicum* loci identified investigating on two-side cross-mapped reads from *S. lycopersicum*-*C. campestris* interaction via combined approach. Chromosome (Chr), locus identifier (Parent) and description of the gene function (Product) were extracted from *S. lycopersicum* genome annotation (gff3 file) and reported.

References

- 'Classification and System in Solanales' (2008). *Solanaceae and Convolvulaceae: Secondary Metabolites: Biosynthesis, Chemotaxonomy, Biological and Economic Significance (A Handbook)* (Berlin, Heidelberg: Springer Berlin Heidelberg), 11–31.
- Albert, M., Axtell, M. J., and Timko, M. P. (2020). Mechanisms of resistance and virulence in parasitic plant–host interactions'. *Plant Physiol.* 185, 1282–1291. doi: 10.1093/plphys/kiab064
- Bawin, T., Bruckmüller, J., Olsen, S., and Krause, K. (2022). A host-free transcriptome for haustoriogenesis in *Cuscuta campestris*: Signature gene expression identifies markers of successive development stages. *Physiol. Plant.* 174, e13628. doi: 10.1111/ppl.v174.2
- Birschwilks, M., Haupt, S., Hofius, D., and Neumann, S. (2006). Transfer of phloem-mobile substances from the host plants to the holoparasite *Cuscuta* sp. *J. Exp. Bot.* 57, 911–921. doi: 10.1093/jxb/erj076
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17, 13. doi: 10.1186/s13059-016-0881-8
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. doi: 10.1093/gigascience/giab008
- Deschamps-Francoeur, G., Simoneau, J., and Scott, M. S. (2020). Handling multi-mapped reads in RNA-seq. *Comput. Struct. Biotechnol. J.* 18, 1569–1576. doi: 10.1016/j.csbj.2020.06.014
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Du, H., Yang, J., Chen, B., Zhang, X., Xu, X., Wen, C., et al. (2022). Dual RNA-seq Reveals the Global Transcriptome Dynamics of *Ralstonia solanacearum* and Pepper (*Capsicum annuum*) Hypocotyls During Bacterial Wilt Pathogenesis. *Phytopathology*® 112, 630–642. doi: 10.1094/PHYTO-01-21-0032-R
- Espindula, E., Sperb, E. R., Bach, E., and Passaglia, L. M. P. (2019). The combined analysis as the best strategy for Dual RNA-Seq mapping. *Genet. Mol. Biol.* 42, e20190215. doi: 10.1590/1678-4685-gmb-2019-0215
- Galise, T. R., Esposito, S., and D'Agostino, N. (2021). "Guidelines for Setting Up a mRNA Sequencing Experiment and Best Practices for Bioinformatic Data Analysis," in *Crop Breeding*. Ed. P. Tripodi (Springer US, New York, NY), 137–162.
- GNU, P. (2007). *Free Software Foundation*, Bash (3.2.48)[Unix shell program].
- Haupt, S., Oparka, K. J., Sauer, N., and Neumann, S. (2001). Macromolecular trafficking between *Nicotiana tabacum* and the holoparasite *Cuscuta reflexa*. *J. Exp. Bot.* 52, 173–177. doi: 10.1093/jxb/52.354.173
- Hegenauer, V., Körner, M., and Albert, M. (2017). Plants under stress by parasitic plants. *Curr. Opin. Plant Biol.* 38, 34–41. doi: 10.1016/j.pbi.2017.04.006
- Honaas, L. A., Wafula, E. K., Yang, Z., Der, J. P., Wickett, N. J., Altman, N. S., et al. (2013). Functional genomics of a generalist parasitic plant: Laser microdissection of host-parasite interface reveals host-specific patterns of parasite gene expression. *BMC Plant Biol.* 13, 9. doi: 10.1186/1471-2229-13-9
- Hosford, R. M. (1967). Transmission of plant viruses by dodder. *Bot. Rev.* 33, 387–406. doi: 10.1007/BF02858742
- Hudzik, C., Hou, Y., Ma, W., and Axtell, M. J. (2020). Exchange of small regulatory RNAs between plants and their pests I. *Plant Physiol.* 182, 51–62. doi: 10.1104/pp.19.00931
- Ikeue, D., Schudoma, C., Zhang, W., Ogata, Y., Sakamoto, T., Kurata, T., et al. (2015). A bioinformatics approach to distinguish plant parasite and host transcriptomes in interface tissue by classifying RNA-Seq reads. *Plant Methods* 11, 34. doi: 10.1186/s13007-015-0066-6
- Jhu, M.-Y., and Sinha, N. R. (2022). *Cuscuta* species: Model organisms for haustorium development in stem holoparasitic plants. *Front. Plant Sci.* 13, 1086384. doi: 10.3389/fpls.2022.1086384
- Johnson, N. R., and Axtell, M. J. (2019). Small RNA warfare: exploring origins and function of trans-species microRNAs from the parasitic plant *Cuscuta*. *Curr. Opin. Plant Biol.* 50, 76–81. doi: 10.1016/j.pbi.2019.03.014
- Kamińska, M., and Korbin, M. (1999). Graft and dodder transmission of phytoplasma affecting lily to experimental hosts. *Acta Physiol. Plant.* 21, 21–26. doi: 10.1007/s11738-999-0023-y
- Katoh, K. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kim, G., LeBlanc, M. L., Wafula, E. K., dePamphilis, C. W., and Westwood, J. H. (2014). Genomic-scale exchange of mRNA between a parasitic plant and its hosts. *Science* 345, 808–811. doi: 10.1126/science.1253122
- Klaren, C. H., and Janssen, G. (1978). Physiological Changes in the Hemiparasite *Rhinanthus serotinus* before and after Attachment. *Physiol. Plant.* 42, 151–155. doi: 10.1111/j.1399-3054.1978.tb01556.x
- Lambers, H., and Oliveira, R. S. (2019). "Biotic Influences: Parasitic Associations," in *Plant Physiological Ecology*. Eds. H. Lambers and R. S. Oliveira (Springer International Publishing, Cham), 597–613.
- Liao, Z.-X., Ni, Z., Wei, X.-L., Chen, L., Li, J.-Y., Yu, Y.-H., et al. (2019). Dual RNA-seq of *Xanthomonas oryzae* pv. *oryzicola* infecting rice reveals novel insights into bacterial-plant interaction. *PLoS One* 14, e0215039. doi: 10.1371/journal.pone.0215039
- Maizel, A., Markmann, K., Timmermans, M., and Wachter, A. (2020). To move or not to move: roles and specificity of plant RNA mobility. *Curr. Opin. Plant Biol.* 57, 52–60. doi: 10.1016/j.pbi.2020.05.005
- Manhart, J. R. (1994). Phylogenetic analysis of green plant rbcL sequences. *Mol. Phylogenet. Evol.* 3, 114–127. doi: 10.1006/mpev.1994.1014
- Marsh, J. W., Hayward, R. J., Shetty, A. C., Mahurkar, A., Humphrys, M. S., and Myers, G. S. (2018). Bioinformatic analysis of bacteria and host cell dual RNA-sequencing experiments. *Briefings Bioinf.* 19, 1115–1129. doi: 10.1093/bib/bbx043
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Naidoo, S., Visser, E. A., Zwart, L., Toit, Y. D., Bhaduria, V., and Shuey, L. S. (2018). Dual RNA-sequencing to elucidate the plant-pathogen duel. *Curr. Issues Mol. Biol.* 27, 127–142. doi: 10.21775/cimb.027.127
- Nickrent, D. L. (2020). Parasitic angiosperms: How often and how many? *TAXON* 69, 5–27. doi: 10.1002/tax.12195
- Nickrent, D., and Musselman, L. (2004). Introduction to parasitic flowering plants. *Plant Health Instruct.* 13. doi: 10.1094/PHI-I-2004-0330-01
- O'Keefe, K. R., and Jones, C. D. (2019). Challenges and solutions for analysing dual RNA-seq data for non-model host–pathogen systems. *Methods Ecol. Evol.* 10, 401–414. doi: 10.1111/2041-210X.13135
- Park, S.-Y., Shimizu, K., Brown, J., Aoki, K., and Westwood, J. H. (2022). Mobile Host mRNAs Are Translated to Protein in the Associated Parasitic Plant *Cuscuta campestris*. *Plants* 11, 93. doi: 10.3390/plants11010093
- Park, I., Song, J.-H., Yang, S., Kim, W. J., Choi, G., and Moon, B. C. (2019). *Cuscuta* species identification based on the morphology of reproductive organs and complete chloroplast genome sequences. *Int. J. Mol. Sci.* 20, 2726. doi: 10.3390/ijms20112726
- Petitot, A., Dereeper, A., Agbessi, M., Da Silva, C., Guy, J., Ardisson, M., et al. (2016). Dual RNA-seq reveals *Meloidogyne graminicola* transcriptome and candidate effectors during the interaction with rice plants. *Mol. Plant Pathol.* 17, 860–874. doi: 10.1111/mpp.2016.17.issue-6
- R Core Team (2024). R: A Language and Environment for Statistical Computing. Vienna (Austria: R Foundation for Statistical Computing).
- Shahid, S., Kim, G., Johnson, N. R., Wafula, E., Wang, F., Coruh, C., et al. (2018). MicroRNAs from the parasitic plant *Cuscuta campestris* target host messenger RNAs. *Nature* 553, 82–85. doi: 10.1038/nature25027
- Soltis, D. E., Smith, S. A., Cellinese, N., Wurdack, K. J., Tank, D. C., Brockington, S. F., et al. (2011). Angiosperm phylogeny: 17 genes, 640 taxa. *Am. J. Bot.* 98, 704–730. doi: 10.3732/ajb.1000404
- Srivastava, A., Malik, L., Sarkar, H., Zakeri, M., Almodaresi, F., Sonesson, C., et al. (2020). Alignment and mapping methodology influence transcript abundance estimation. *Genome Biol.* 21, 239. doi: 10.1186/s13059-020-02151-8
- Subhankar, B., Yamaguchi, K., Shigenobu, S., and Aoki, K. (2021). Trans-species small RNAs move long distances in a parasitic plant complex. *Plant Biotechnol.* 38, 187–196. doi: 10.5511/plantbiotechnology.21.0121a
- van Dorst, H. J. M., and Peters, D. (1974). Some biological observations on pale fruit, a viroid-incited disease of cucumber. *Netherlands J. Plant Pathol.* 80, 85–96. doi: 10.1007/BF01980613
- Walker, P. L., Belmonte, M. F., McCallum, B. D., McCartney, C. A., Randhawa, H. S., and Henriquez, M. A. (2024). Dual RNA-sequencing of *Fusarium* head blight resistance in winter wheat. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1299461
- Wang, L.-G., Lam, T. T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., et al. (2020). Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Mol. Biol. Evol.* 37, 599–603. doi: 10.1093/molbev/msz240
- Westermann, A. J., Barquist, L., and Vogel, J. (2017). Resolving host–pathogen interactions by dual RNA-seq. *PLoS Pathog.* 13, e1006033. doi: 10.1371/journal.ppat.1006033
- Westermann, A. J., Gorski, S. A., and Vogel, J. (2012). Dual RNA-seq of pathogen and host. *Nat. Rev. Microbiol.* 10, 618–630. doi: 10.1038/nrmicro2852
- Westwood, J. H., and Kim, G. (2017). RNA mobility in parasitic plant – host interactions. *RNA Biol.* 14, 450–455. doi: 10.1080/15476286.2017.1291482

Westwood, J. H., Yoder, J. L., Timko, M. P., and dePamphilis, C. W. (2010). The evolution of parasitism in plants. *Trends Plant Sci.* 15, 227–235. doi: 10.1016/j.tplants.2010.01.004

Wolf, T., Kämmer, P., Brunke, S., and Linde, J. (2018). Two's company: studying interspecies relationships with dual RNA-seq. *Curr. Opin. Microbiol.* 42, 7–12. doi: 10.1016/j.mib.2017.09.001

Wu, Y., Luo, D., Fang, L., Zhou, Q., Liu, W., and Liu, Z. (2022). Bidirectional lncRNA transfer between *Cuscuta* parasites and their host plant. *Int. J. Mol. Sci.* 23, 561. doi: 10.3390/ijms23010561

Yu, G. (2020). Using ggtree to visualize data on tree-like structures. *Curr. Protoc. Bioinf.* 69, e96. doi: 10.1002/cpbi.v69.1