Reconstructing the tomato root-associated metatranscriptome using host-targeted RNAseq data

The abstract must include the **Motivation, Methods, and Results sections**. It cannot be longer than **5,000 characters (including spaces) or longer than 4,000 characters (including spaces) when a figure is also included** (see Figure field below)

Title:

Reconstructing the tomato root-associated metatranscriptome using host-targeted RNAseq data

Chialva M(1), Salvioli A(1), Ghignone S(2), Novero M(1), Bonfante P(1)

Affiliation
(1) Department of Life Sciences and Systems Biology, University of Torino, Viale P.A. Mattioli 25, Torino, Italy
(2) Institute for Sustainable Plant Protection (IPSP), Italian National Research Council (CNR), Viale P.A. Mattioli 25, I-10125 Torino, Italy

Motivation
Plants, like humans, have their own microbiota, which may exert a powerful effect on their health. In the last years, Next Generation Sequencing technologies, applied to develop 16S/18S metabarcoding data sets and metagenomes have become a standard to routinely describe microbiota diversity. As a step forward, metatranscriptomic approaches applied to host-associated niches are providing details on the functions activated by microbial communities and their potential interaction with their host. However, such approaches are limited by several factors, among which the high-coverage sequencing needed which results into major economic efforts as well as technical limitations, such as RNA extraction from environmental samples, rRNA depletion procedures, and computationally intensive bioinformatics. However, understanding how soil microbiota (one of the most diverse ecosystem on Earth) influences plant responses is becoming a priority in plant sciences and will pave the way for next-generation agriculture.
In this work, we analysed the tomato root-associated fungal metatranscriptome using host-targeted short-reads RNA-seq data from Chialva *et al*. (2018) and set up a new bioinformatic pipeline taking advantage of published tools and R programming environment. Furthermore, we tried to link the reconstructed metatranscriptome at both diversity and functional level with the host transcriptional profile, testing the hypothesis that molecular functions in the plant-microbiota complex (plant holobiont) are connected.

Methods
Total RNA was extracted from tomato roots grown on native soils as well as on an artificial substrate, and including two tomato genotypes as detailed in Chialva *et al*. (2018). Poly-A enriched RNA was converted in cDNA and sequenced using Illumina HiScanSQ (50 bp reads).
Reads were trimmed using TRIMMOMATIC v0.35 and host sequences removed by aligning on tomato genome (SL.2.50) using STAR c2.2.5a. Residual host (transcript variants) and contaminant reads were detected mapping on a NCBI-NT subset containing all tomato and human sequences using Bowtie2 v2.2.9 and removed. rRNA sequences were removed with SortMeRNA v2.1. The remaining reads were considered as metatranscriptome and annotated. Due to the short read length, we chose to annotate reads rather than generating an assembly to avoid the assembly of chimeric transcripts. Sequences were mapped on the full NCBI Nucleotide database using Taxoner64 v1.7 software and Bowtie2. For taxonomical assignment, the LCA (last common ancestor) algorithm at 97% sequence identity on the reference database using NCBI taxonomy was adopted. Residual contaminants (Viridiplantae and Metazoa) as well as low-abundance *taxa* (<5 raw reads in at least 3 libraries) were discarded.
Reads were functionally annotated by mapping them to the eggNOG database v4.5. Since sequencing length did not allow functional assignment, we generated a set of ‘pseudo-reads’ extending taxa-annotated reads 50 bp upstream and downstream on their mapping reference generating 150 bp fragments using 'getfasta' function in BEDTools v2.26.0-19. Pseudo-reads from

different samples were than pooled and mapped on eggNOG using DIAMOND-BLASTX v0.8.24.
Plant transcriptome was analysed mapping reads on tomato reference genome with TopHat2 and normalizing reads count with DESeq2 as described in Chialva *et al*. (2018).
Analysis and statistical tests were performed in R statistical programming environment (R Core Team, 2017).

Results

Using a custom pipeline we were able to reconstruct tomato root metatranscriptome using host-targeted RNA-seq libraries. Even though the sequencing coverage for microbial transcripts was low, we successfully reconstructed fungal functional diversity associated with plant host. As reported in literature, only a small part of non-host reads can be successfully assigned and, in our case, this was due to the short reads length. The fungal microbiota, irrespectively of the soil type, was dominated by Ascomycota and other *taxa* commonly associated to rhizosphere such as Glomeromycotina. However, differential abundance analysis revealed that these taxa were more abundant in both native soils when compared to an artificial substrate. Even if the pipeline did not allow to assign functions to *taxa*, we successfully reconstruct overall expressed microbial functions (1,568-56,515 reads per library mapped). Among functional categories, in native soils, we found as upregulated different COGs corresponding to ribosomal proteins, mirroring a higher microbial metabolic activity. Finally, to validate our hypothesis, we performed a variance partitioning analysis: it revealed that microbial functional categories, more than diversity, hold a statistically supported correlation with the tomato transcriptome explaining more than 40% of its variability.