Research for Practice LAIMBURG JOURNAL

Short Paper

From Chaos to Clarity. The Apple TCP Gene Naming Makeover

Vom Chaos zur Klarheit. Die Neuordnung der Apfel TCP Gen-Benennung

Dal Caos all'ordine. Il restauro dei geni TCP di melo

Mattia Tabarelli[1]

[1]Laimburg Research Centre, 39040 Auer/Ora, BZ, Italy

ABSTRACT

Apple Proliferation (AP) is a widespread disease affecting apple orchards throughout Europe, including South Tyrol.The disease is associated with a phytoplasma, '*Candidatus* Phytoplasma mali' ('*Ca.* P. mali'), which manipulates the plant's physiological processes through the secretion of small peptides called effectors. One well-studied phytoplasma effector is SAP11, which targets Teosinte Branched 1/Cycloidea/Proliferating Cell Factor 1 (TCP) genes in multiple plant species, including apple. TCP genes encode plant-specific transcription factors involved in various biological processes, such as growth, development, and responses to stimuli. The identification and naming of TCP genes have been inconsistent, leading to confusion and redundancy in sequence databases. In apple, 52 TCP genes (MdTCP) were identified in 2014 on the 2010 assembly of the apple genome, which is now considered outdated. To address this issue, a comprehensive investigation was conducted to identify and name TCP genes in apple using the high-quality genome assembly, GDDH13v1.1. Existing TCP sequences were aligned with the genome excluding redundant, fragmented, and non-TCP sequences. The revised set comprised 40 unique MdTCPs, including three novel genes. To establish a standardised nomenclature, each MdTCP was BLASTedagainst the Arabidopsis thaliana gene database, in order to identify the best hits for each MdTCP. The MdTCP genes were then renamed, incorporating the letter "a" or "b" to differentiate between MdTCPs showing homology to the same AtTCP, and also between the existing and proposed nomenclatures. The present study highlights the need for clarity and organisation in sequence databases, especially in respect of TCP genes. The presence of redundant and fragmented sequences in databases complicates accurate identification and annotation of genes. By providing a comprehensive and standardised nomenclature system, we aim to enhance the coherence and interoperability of future TCP gene research in plant and crop science. In our view, this work will serve as a valuable resource for researchers studying TCP genes in apple and will provide insights into the evolutionary dynamics and functional roles of TCP genes in plants.

KEYWORDS

Malus domestica, MdTCP, nomenclature system, genomic data reorganisation

CITE ARTICLE AS

Tabarelli Mattia (2024). From Chaos to Clarity. The Apple TCP Gene Naming Makeover. Laimburg Journal 06/2024.008 DOI: 10.23796/LJ/2024.008.

CORRESPONDING AUTHOR

Tabarelli Mattia, Laimburg Research Centre, Laimburg 6 – Pfatten/Vadena, 39040 Auer/Ora, BZ, Italy
mattia.tabarelli@laimburg.it, +390471969736

INTRODUCTION

APPLE PROLIFERATION

Apple Proliferation (AP) is a widespread disease found in apple orchards across Europe, including South Tyrol. Its symptoms include reddening or yellowing of the leaves, shoot proliferation, abnormal leaf and root growth, and the production of small, tasteless, and colourless fruits. Effective strategies for managing AP involve removing infected trees, controlling the insect vectors responsible for transmission, and utilising disease-free propagation material. The disease is associated with a phytoplasma, '*Candidatus* Phytoplasma mali' ('*Ca*. P. mali') [1]. Phytoplasmas are wall-less bacteria that colonise the plant's phloem. One main feature of phytoplasmas is their extremely reduced genome size, the result of their adaptation to the obligatory intracellular parasites' lifestyle. This evolutionary phenomenon, widely observed in nature, is the result of the loss of many genes and the conservation of those essential for survival and replication [2]. The phytoplasma genome lacks, for instance, genes for amino acid and fatty acid biosynthesis, oxidative phosphorylation and even ATP synthase, but they do possess genes that encode for small peptides, known as effectors, that are secreted into host cells. These peptides can modify the plant's physiological processes, providing a competitive advantage for the bacteria [3]. The first identified effectors were named Secreted AY-WB Proteins (SAP) after Aster Yellow Phytoplasma (AY-WB), where they were initially described [4]. SAP11 stands out among the well-studied effectors, as it has been demonstrated to target and inactivate various members of the Teosinte Branched 1, Cycloidea, Proliferating Cell Factors (TCP) gene family in multiple plant species [5]. In the case of '*Ca*. P. mali', a SAP11 homologue called ATP00189 or SAP11CaPM has been identified, which also targets TCP genes in apple [6] [7].

THE TCP TRANSCRIPTION FACTOR GENE FAMILY

TCP consists of a large group of genes encoding for plant-specific transcription factors (TF) found in all land plants, from mosses to eudicots. TFs bind on specific regions on the DNA initiating and regulating the expression of genes, that is the rate at which the DNA is converted into RNA that will then be translated into proteins. TCPs regulate a large number of biological processes, mainly relative to plant growth, development, and responses to various stimuli, such as branching regulation, and floral and leaf development [8] [9]. These proteins are also believed to play a role in plant defence mechanisms as it has been observed that some pathogens, among which *Pseudomonas syringae*, *Hyaloperonospora arabidopsidis* [10] and the previously mentioned phytoplasmas, tend to target TCP genes.

The number of TCP genes varies significantly among species, particularly between basal and higher plants, due to numerous gene duplication events that influenced the evolution of this gene family. In fact, a relatively small number of TCP genes is found in the basal groups of land plants, whereas higher plants usually possess a larger repertoire: for example, the moss *Physcomitrella patens* and the lycophyte *Se*laginella *moellendorffii* have five and seven TCP genes, respectively, whereas 24 are present in *Arabidopsis thaliana* (*A. thaliana*) and 46 in *Zea mays* [11]. All TCP proteins share a conserved basic-helix-loop-helix (bHLH) motif, called TCP domain, which mediates nuclear localisation, DNA binding and protein-protein interactions. Nearly half of the TCP proteins display a conserved deletion of four amino acids in the TCP domain, allowing their classification into Class I and Class II, with and without deletion, respectively. The division into two classes has significant implications for the functional characteristics of the proteins: the DNA binding sequences of the two TCP classes differs slightly, indicating that members of different classes have similar yet distinct target sequences on the DNA [12]. Furthermore, TCPs can form both homo- and heterodimers, which also exhibit varying degrees of specificity for different sequences. Lastly, functional evidence indicates that to some extent, Class I and Class II may regulate the same biological process in an antagonistic manner [13].

The first (and only) description of the TCP gene family in Malus domestica (M. domestica) (TCPs from *M. domestica* will be named MdTCPs in this article) can be traced back to 2014, when Xu and collaborators [14] identified 52 TCPs on the first apple genome assembly,released in 2010 by Velasco and collaborators [15]. The 52 genes were numbered based on their relative order on the chromosomes, that is corresponding to the order in which they were found on different chromosomes: the first gene on chromosome 1 was designated MdTCP1, the second MdTCP2, and so on for all the 17 pseudochromosomes. The basic and essential work carried out by Xu *et al*. required an update for two main reasons: firstly, in 2017 a new assembly of the *M. domestica* genome, GDDH13v1.1, was released [16] and is currently considered to be the reference genome for apple. Secondly, a manual analysis of the TCP gene sequences revealed several anomalies, such as reciprocal identical sequences and premature stop codons.

THE CHALLENGES OF GENOMIC DATA ORGANISATION

The drastic decrease in sequencing costs has led to a massive increase in the availability of complete and partial nucleotide and protein sequences, which are published continuously every year. While the accessibility and abundance of genomic data facilitate proliferation and expansion of knowledge, it is essential to establish a system that brings order to this vast amount of data. At the time of writing, the GenBank database, one of the largest collections of gene,

protein, and genomic data, contains more than 2.4 billion records [17]. Among these records, some are annotated, meaning that specific features of the sequences, such as coding regions, untranslated regions (UTRs), introns, and exons, have been described. However, many entries consist of fragments, uncharacterised sequences, or whole genome raw sequences. It is common to find multiple instances of the same gene sequence in different entries, arising from different studies or through the process of "automatic annotation", where internal algorithms automatically determine the characteristics of the sequences. Consequently, the same gene may be present multiple times with different names. This chaotic situation is further amplified by works of gene identification in genomes, where entire gene families are identified and named. For example, in the case of TCP genes identification in apple in 2014, genes were named based on their relative order on the chromosome. In some instances, genes are named using the principle of homology, which involves comparing their sequences to those of a reference organism. In the field of plant science, the commonly used reference organism for this purpose is *A. thaliana*. Homology-based naming relies on identifying similarities in gene sequences with the reference organism to assign appropriate names to the genes. The choice of using one naming system over the other is at the discretion of the authors or, sometimes, the journal. A clear example of this is the case of the identification of TCP genes in grapevine, where two independent groups published the same study only a few weeks apart, each using a different naming system [18] [19]. As a result, the same sequences were assigned different names.

In light of this, we conducted a comprehensive investigation aimed at identifying MdTCP genes within the high-quality genome assembly GDDH13v1.1. Additionally, we propose a standardised nomenclature system for MdTCP genes, drawing upon their homology with *A. thaliana* genes. This initiative seeks to establish a systematic framework that brings clarity and organisation to the TCP gene landscape. By aligning the nomenclature with a well-established model organism, such as *A. thaliana*, we aim to enhance the coherence and interoperability of future studies in the field of TCP gene research in plant and crop science.

RESULTS

Initially, all putative TCP-containing sequences of M. domestica available in public databases were collected and aligned with the previously published set of 52 MdTCP sequences. The alignment revealed that the sequences retrieved from the databases either corresponded to the published MdTCP gene set, represented fragments of TCPs, or did not contain a TCP domain at all. Notably, no additional TCP-containing sequence was found beyond those already mentioned in the MdTCP set, which were subsequently utilised for further analyses. A manual evaluation of the 52 MdTCP sequences confirmed that four pairs (MdTCP5/MdTCP6, MdTCP13/MdTCP14, MdTCP19/-MdTCP20, MdTCP43/MdTCP44) and one trio (MdTCP8/MdTCP9/-MdTCP10) exhibited 100% reciprocal identity. The alignment of these 11 sequences to the *M. domestica* double haploid genome GDDH13v1.1. highlighted that each pair or trio mapped to a single position on the genome. Consequently, six redundant sequences, one from each pair, and two from the trio, were excluded from the gene set. Next, the remaining 46 MdTCP sequences were aligned to the GDDH13v1.1 assembly. The alignment revealed that 33 sequences matched with predicted protein-coding genes, while 13 sequences mapped onto intergenic regions (stretches of DNA sequences between genes) or partially overlapped with predicted gene regions. The nucleotide sequences corresponding to the 13 non-TCP mapped genes were retrieved and assessed for the presence of the TCP domain. Among them, seven deduced amino acid sequences (MdTCP3, MdTCP4, MdTCP19, MdTCP36, MdTCP37, MdTCP41, MdTCP50) displayed premature stop codons, and two sequences (MdTCP7, MdTCP42) were fragments of non-TCP gene sequences. Consequently, these nine sequences were removed from the set. Conversely, the remaining four non-TCP mapping sequences (MdTCP2, MdTCP17, MdTCP23, MdTCP52) contained a full open reading frame (a potential protein-coding region) with a TCP domain, and were therefore retained in the set.

During the revision process, a total of 15 out of the initial 52 sequences were excluded from the set. The reasons for exclusion included redundancy (six sequences), presence of premature stop codons (seven sequences), or constituting fragments of non-TCP genes (two sequences). Out of the remaining 37 sequences, 30 were listed as TCP-containing in the predicted mRNA gene list derived from GDDH13v1.1, which is an automatic annotation performed by the authors on the whole genome. Additionally, three novel putative TCP genes that were not reported in the original set by Xu *et al.* were identified in this predicted mRNA set. Interestingly, three sequences (MdTCP33, MdTCP40, MdTCP45) previously observed to match with protein-coding genes on the GDDH13v1.1 assembly, were automatically annotated as uncharacterised proteins. The three novel sequences, which contained a start and a stop codon as well as a TCP domain, were provisionally named "nc" (for “not classified”) 1, 2, and 3 (MdTCPnc1, MdTCPnc2, and MdTCPnc3) and included in the set. Alignment of the deduced amino acid sequences of these three novel MdTCPs revealed their similarity. Specifically, MdTCPnc3 and MdTCPnc2 respectively aligned with the 3’ and 5’ ends of MdTCPnc1, which is the longest of the three.

To determine the *A. thaliana* homo-

logue of each MdTCP, a BLAST query against the *A. thaliana* gene database was performed (TCPs from *A. thaliana* will be named AtTCPs in this article). This made it possible to identify the best hits for each gene, which were further manually investigated and confirmed. MdTCP genes were thus re-named, appending the letter “a” and, eventually, “b” after the number to differentiate between the current and proposed nomenclatures. In two cases the same *A. thaliana* sequence exhibited similarity to three MdTCP sequences (AtTCP9 with MdTCP12, MdTCP26, and MdTCP32; AtTCP1 with MdTCPnc1, MdTCPnc2, and MdTCPnc3), and a cluster of four genes (MdTCP2, MdTCP17, MdTCP23, and MdTCP52) showed similarity to a single *A. thaliana* gene (AtTCP17). In total, 19 AtTCP sequences were identified as homologues for the complete set of 40 MdTCPs. However, a corresponding M. domestica orthologue could not be found for the remaining five AtTCPs, namely AtTCP11, AtTCP16, AtTCP22, AtTCP23, and AtTCP24.

The final revised set of MdTCP genes consists of 40 unique sequences, categorised as follows: 33 deduced from both the 2010 and GDDH13v1.1 genome assemblies, four deduced from the 2010 assembly, but not annotated in GDDH13v1.1, and three deduced from GDDH13v1.1 but absent in the 2010 assembly, and thus not predicted previously. Figure 1 displays the MdTCP phylogenetic tree, in which the separation of MdTCPs into the two classes and subclasses thereof is visible: 17 genes belong to TCP Class I (22 in the previous classification by Xu and coauthors), 16 belong to the subclass CIN of Class II (26 in the previous classification) and seven belong to the Class II subclass CYC/TB1 (four in the previous classification). For comparison, in *A. thaliana*, Class I contains 13 genes, Class II-CIN contains eight, and Class II-TB1/CYC contains three.

DISCUSSION

In the present study we reidentified the MdTCP gene family from the high-quality M. domestica genome assembly, GDDH13v1.1. Compared to the previous classification, the number of genes decreased from 52 to 40, due to the addition of three novel sequences and the exclusion of 15. The decision to exclude these sequences was motivated by the fact that nine did not possess a TCP-domain-containing open reading frame, and six showed redundancies with other sequences. Interestingly, the redundant sequences had sequential names reflecting their physical proximity on the chromosomes, as the previous nomenclature was based on the relative position of each gene. Consequently, the duplications responsible for the redundant MdTCP sequences in the 2010 assembly were absent in GDDH13v1.1 and were likely artifacts generated during the previous *in silico* genome assembly process.

Additionally, we identified 19 AtTCP genes as homologues of the 40 MdTCP genes. Notably, in the classification based on the previous draft genome assembly, only 15 AtTCP genes were identified

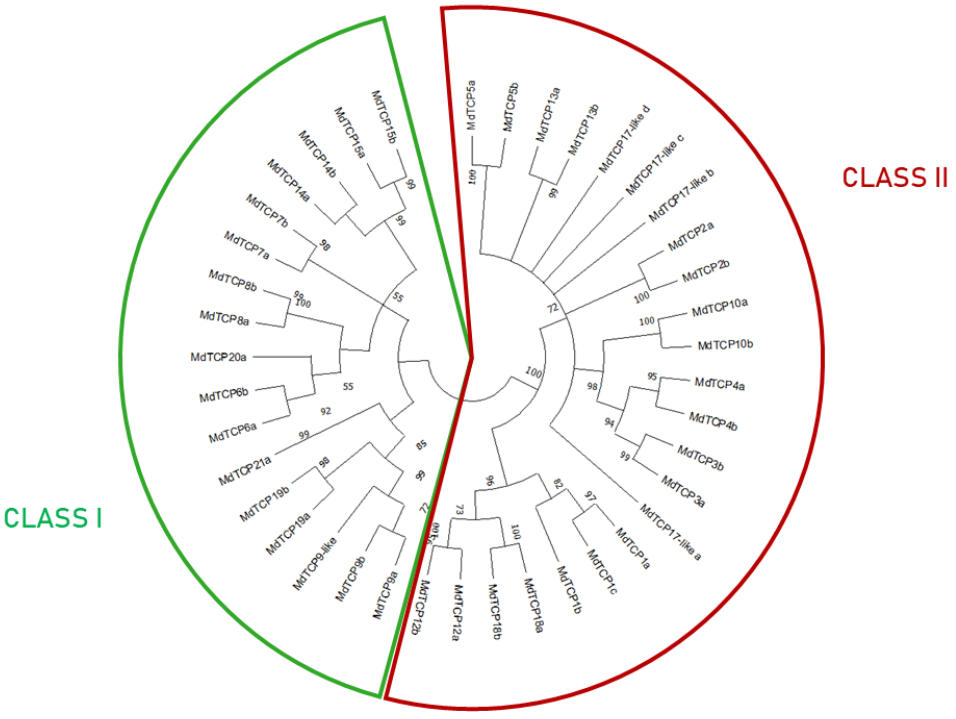

Fig. 1: Neighbour-joining phylogenetic tree calculated on the amino acid sequences of the 40 MdTCP genes identified in the present work. The two classes of TCP are indicated in green (Class I) and red (Class II). Node support is indicated as bootstrap values, resulting from 10 000 reiterations. Nodes with a support lower than 50 are collapsed.

as homologues for the complete set of MdTCP genes. Despite the higher number of AtTCP homologues found in our study, none of the five remaining AtTCP genes had a corresponding MdTCP gene. It is worth mentioning that each Class II AtTCP gene had two orthologues, except for AtTCP24, while several members of Class I seemed to have only one or no orthologues in M. domestica. As previously mentioned, the evolutionary history of TCPs has been significantly influenced by genome duplication events. The presence of approximately two orthologues of apple for each AtTCP is consistent with the whole genome duplication event that occurred approximately 50 million years ago in the Malinae group [15]. This observation also suggests that, in most cases, newly derived TCP genes resulting from gene or genomic duplications are retained and remain active genes, being transcribed and possessing open reading frames, and do not become pseudogenes, i.e., genes that no longer encode for functional proteins. This indicates a positive evolutionary pressure to maintain the functional redundancy of TCP genes. In regard to the absence of apple orthologues for four of the AtTCP genes, two possibilities exist: either these genes exist but have not yet been identified, or they have been secondarily lost .

The nomenclature of TCP genes, as well as other gene families, is currently based on a numbering system that can be either linked to the relative order of the sequences on the chromosomes or on the homology of the sequences with their corresponding counterparts in *A. thaliana.* While the situation is more regulated in the context of medical data, such as those related to humans, rats and mice, there is currently no obligation regarding plant science. This dual adoption of different standards for classifying newly discovered gene sequences can result in ambiguities, where identical sequences are assigned two or even three different names. Such a situation presents a significant obstacle to research, as more genes are being characterised in terms of their functional roles in physiological pathways, interaction partners, and expression patterns. To overcome this issue and avoid misinterpretation or underutilisation of data, we propose the establishment of a common standard, starting with the TCP gene family nomenclature in apple. In our study, we systematically renamed the identified TCP genes based on their homology with their counterparts in *A. thaliana*. We believe that this standardised classification will facilitate future research by reducing ambiguities associated with the random numbering of TCP genes across different species.

ACKNOWLEDGMENTS

The present work [20] was performed as part of a PhD project with the University of Udine, co-funded by the Edmund Mach Foundation and Laimburg Research Centre [21]. In Laimburg the work was part of the APPLIII and APPLIV projects within the framework agreement in the field of invasive species in fruit growing and major pathologies, which is co-funded by the South Tyrolean Apple Consortium and the Autonomous Province of Bozen/Bolzano. I would like to thank my former supervisor, Dr. Mickael Malnoy, leader of the Plant Biotechnology Unit of the E. Mach Foundation, and my former co-supervisor and current group leader, Dr. Katrin Janik.

ZUSAMMENFASSUNG

Apfeltriebsucht (AP) ist eine weit verbreitete Krankheit, die Apfelplantagen in Europa, einschließlich Südtirol, betrifft. Ein effizientes Management von AP umfasst das Entfernen infizierter Bäume, die Kontrolle von Insektenvektoren und die Verwendung von krankheitsfreiem Vermehrungsmaterial. Die Krankheit wird mit einem Phytoplasma namens '*Candidatus* Phytoplasma mali' ('*Ca.* P. mali') in Verbindung gebracht, das die physiologischen Prozesse der Pflanze durch die Sekretion kleiner Peptide, sogenannten Effektoren, manipuliert. Ein gut untersuchter Phytoplasma-Effektor ist SAP11, der auf TCP-Gene in mehreren Pflanzenarten, einschließlich Apfel, abzielt. TCP-Gene kodieren pflanzenspezifische Transkriptionsfaktoren, die an verschiedenen biologischen Prozessen wie Wachstum, Entwicklung und Reaktionen auf Reize beteiligt sind. Die Identifizierung und Benennung von TCP-Genen war bisher uneinheitlich, was zu Verwirrung und Redundanz in Sequenzdatenbanken geführt hat. 2014 wurden anhand der 2010 erstellten und mittlerweile als veraltet geltenden Assemblierung des Apfelgenoms im Apfel 52 TCP-Gene identifiziert. Zur Lösung dieses Problems wurde unter Verwendung der hochwertigen Genomassemblierung GDDH13v1.1 eine umfassende Untersuchung zur Identifizierung und Benennung von TCP-Genen im Apfel durchgeführt. Vorhandene TCP-Sequenzen wurden mit dem Genom abgeglichen, wodurch redundante, fragmentierte und nicht-TCP-Sequenzen ausgeschlossen wurden. Der überarbeitete Satz umfasste 40 eindeutige TCP-Gene, darunter befinden sich auch drei neuartige Gene. Um eine standardisierte Nomenklatur zu erstellen, wurden BLAST-Abfragen in der Gendatenbank von Arabidopsis thaliana durchgeführt, um die besten Übereinstimmungen für jedes MdTCP-Gen zu ermitteln. Die MdTCP-Gene wurden dann umbenannt und mit den Buchstaben "a" oder "b" versehen, um zwischen MdTCPs, die eine Homologie mit demselben AtTCP aufweisen, sowie zwischen der bestehenden und der vorgeschlagenen Nomenklatur zu unterscheiden. Die vorliegende Studie unterstreicht den Bedarf an Klarheit und Organisation in Sequenzdatenbanken, insbesondere im Zusammenhang mit TCP-Genen. Das Vorhandensein redundanter und fragmentierter Sequenzen in Datenbanken erschwert die genaue Identifizierung und Annotation von Genen. Durch die Bereitstellung eines umfassenden und standardisierten Nomenklatursystems wollen wir die Kohärenz und Interoperabilität der zukünftigen TCP-Genforschung in den Pflanzen- und Nutzpflanzenwissenschaften verbessern. Aus unserer Sicht stellt diese Arbeit eine wertvolle Ressource für Forscher dar, die TCP-Gene im Apfel untersuchen, und liefert Einblicke in die evolutionäre Dynamik und die funktionelle Rolle von TCP-Genen in Pflanzen.

RIASSUNTO

Lo scopazzo del melo (AP) è una malattia molto diffusa che colpisce i meleti in Europa, Alto Adige compreso. Una gestione efficace di AP prevede la rimozione degli alberi infetti, il controllo degli insetti vettori e l'utilizzo di materiale di propagazione sano. La malattia è associata a un fitoplasma, 'Candidatus Phytoplasma mali' ('Ca. P. mali'), che manipola i processi fisiologici della pianta attraverso la secrezione di piccoli peptidi, chiamati effettori. Un effettore di fitoplasma ben studiato è SAP11, che ha come bersaglio i geni TCP in diverse specie vegetali, tra cui il melo. I geni TCP codificano fattori di trascrizione specifici di pianta coinvolti in vari processi biologici, tra i quali crescita, sviluppo e risposta agli stimoli. L'identificazione e la denominazione dei geni TCP sono state incoerenti, determinando confusione e ridondanza nei database di sequenze. Nel 2014 sono stati identificati 52 geni TCP nel melo basandosi sulla versione del genoma pubblicata nel 2010, ora considerata obsoleta. Nel presente studio è stata condotta un'indagine completa per identificare i geni TCP in Malus domestica sul genoma di alta qualità GDDH13v1.1, pubblicato nel 2014. Le sequenze TCP preesistenti sono state allineate con il genoma, escludendo così quelle ridondanti, frammentate o non TCP. Il set rivisto comprende 40 geni TCP unici, inclusi tre nuovi geni mai identificati in precedenza. Per stabilire una nomenclatura standardizzata, sono state eseguite delle query BLAST sul database dei geni di Arabidopsis thaliana per identificare i migliori risultati per ciascun gene MdTCP. I geni MdTCP sono stati poi rinominati di conseguenza, incorporando la lettera "a" o "b" per distinguere MdTCP simili allo stesso AtTCP e, allo stesso tempo, tra la nomenclatura esistente e quella proposta. Il presente studio evidenzia la necessità di chiarezza e organizzazione nei database di sequenze, soprattutto per quanto riguarda i geni TCP. La presenza di sequenze ridondanti e frammentate nei database complica l'identificazione e l'annotazione accurata dei geni. Fornendo un sistema di nomenclatura completo e standardizzato, miriamo a migliorare la coerenza e l'interoperabilità della futura ricerca sui geni TCP. Riteniamo che questo lavoro costituisca una risorsa preziosa per i ricercatori che studiano i geni TCP in Malus domestica e fornisca approfondimenti sulle dinamiche evolutive e sui ruoli funzionali dei geni TCP nelle piante.

REFERENCES

[1] Seemüller E., Schneider B. (2004). 'Candidatus Phytoplasma mali', 'Candidatus Phytoplasma pyri' and 'Candidatus Phytoplasma prunorum', the causal agents of apple proliferation, pear decline and European stone fruit yellows, respectively. International journal of systematic and evolutionary microbiology 54 (4), 1217-1226, DOI:10.1099/ijs.0.02823-0.

[2] Oshima K., Maejima K., Namba S. (2013). Genomic and evolutionary aspects of phytoplasmas. Frontiers in microbiology 4:230, DOI:10.3389/fmicb.2013.00230.

[3] Sugio A., Kingdom H.N., MacLean A.M. et al. (2011). Phytoplasma protein effector SAP11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. Proceedings of the National Academy of Sciences 108 (48), E1254-1263, DOI:10.1073/pnas.1105664108.

[4] Bai X., Correa V.R., Toruño T.Y. et al. (2009). AY-WB phytoplasma secretes a protein that targets plant cell nuclei. Molecular plant-microbe interactions 22 (1), 18-30, DOI:10.1094/MPMI-22-1-0018.

[5] Sugio A., MacLean A.M., Hogenhout S.A. (2014). The small phytoplasma virulence effector SAP11 contains distinct domains required for nuclear targeting and CIN-TCP binding and destabilization. New Phytologist 202 (3), 838-848, DOI:10.1111/nph.12721.

[6] Mittelberger C., Hause B., Janik K. (2022). The 'Candidatus Phytoplasma mali' effector protein SAP11CaPm interacts with MdTCP16, a class II CYC/TB1 transcription factor that is highly expressed during phytoplasma infection. PLoS ONE 17 (12), e0272467, DOI:10.1371/journal.pone.0272467.

[7] Janik K., Mithöfer A., Raffeiner M. et al. (2017). An effector of apple proliferation phytoplasma targets TCP transcription factors-a generalized virulence strategy of phytoplasma. Molecular Plant Pathology 18 (3), 435-442, DOI:10.1111/mpp.12409.

[8] Martín-Trillo M., Cubas P. (2010). TCP genes. A family snapshot ten years later. Trends in Plant Science 15 (1), 31-39, DOI:10.1016/j.tplants.2009.11.003.

[9] Nicolas M., Cubas P. (2016). TCP factors. New kids on the signaling block. Current Opinion in Plant Biology 33 (October), 33-41, DOI:10.1016/j.pbi.2016.05.006.

[10] Mukhtar M.S., Carvunis A.-R., Dreze M. et al. (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science 333 (6042), 596-601, DOI:10.1126/science.1203659.

[11] Zhou H., Hwarari D., Ma H. et al. (2022). Genomic survey of TCP transcription factors in plants. Phylogenomics, evolution and their biology. Frontiers in genetics 13:1060546, DOI:10.3389/fgene.2022.1060546.

[12] Kosugi S., Ohashi Y. (2002). DNA binding and dimerization specificity and potential targets for the TCP protein family. The Plant journal 30 (3), 337-348, DOI:10.1046/j.1365-313x.2002.01294.x.

[13] Li S. (2015). The Arabidopsis thaliana TCP transcription factors: A broadening horizon beyond development Plant Signaling & Behavior 10 (7). DOI:10.1080/15592324.2015.1044192.

[14] Xu R., Sun P., Jia F. et al. (2014). Genome-wide analysis of TCP transcription factor gene family in Malus domestica. Journal of Genetics 93 (3), 733-746, DOI:10.1007/s12041-014-0446-0.

[15] Velasco R., Zharkikh A., Affourtit J. et al. (2010). The genome of the domesticated apple (Malus × domestica Borkh.). Nature Genetics 42 (10), 833-839, DOI:10.1038/ng.654.

[16] Daccord N., Celton J.-M., Linsmith G. et al. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. Nature Genetics 49 (7), 1099-1106, DOI:10.1038/ng.3886.

[17] Clark K., Karsch-Mizrachi I., Lipman D.J. et al. (2016). GenBank. Nucleic Acids Research 44 (D1), D67-72, DOI:10.1093/nar/gkv1276.

[18] Leng X., Wei H., Xu X. et al. (2019). Genome-wide identification and transcript analysis of TCP transcription factors in grapevine. BMC Genomics 20:786, DOI:10.1186/s12864-019-6159-2.

[19] Jiu S., Xu Y., Wang J. et al. (2019). Genome-Wide Identification, Characterization, and Transcript Analysis of the TCP Transcription Factors in Vitis vinifera. Frontiers in genetics 10:1276, DOI:10.3389/fgene.2019.01276.

[20] Tabarelli M., Malnoy M., Janik K. (2022). Chasing Consistency. An Update of the TCP Gene Family of Malus × domestica. Genes 13 (10), 1696, DOI:10.3390/genes13101696.

[21] Tabarelli, M. (2021). Investigating the molecular targets of Apple Proliferation Phytoplasma effector SAP11CaPM in apple. (Dissertation). University of Udine/Edmund Mach Foundation. S. Michele a. Adige, Italy. Retrieved May 12, 2023, from https://hdl.handle.net/11390/1208644