

# Whole-genome sequence of a non-pathogenic *Fusarium* strain isolated from alfalfa in southern Tunisian Oases

Ben Alaya Amani\*, Chaouachi Manel, Khiari Bilel, Ben Slimene Imen, Djebali Naceur\*\*

Laboratory of Bioactive Substances, Center of Biotechnology of Borj Cedria, BP 901, Hammam-Lif 2050, Hammam-Lif, Tunisia

\* Corresponding author: amani.ben.alaya@gmail.com ; \*\*Co-corresponding author: dnaceur@yahoo.fr

## Abstract

The use of non-pathogenic *Fusarium* strains as biofertilizers and biocontrol agents represents a promising and sustainable alternative to conventional agricultural practices. In this context, we present a draft genome of *Fusarium* sp. TZ1MST22, a non-pathogenic endophytic strain isolated from *Medicago sativa* plants collected in an oasis in southern Tunisia. Whole-genome sequencing of the strain yielded a 39.76 Mb assembly comprising 160 contigs longer than 500 bp. Based on comprehensive metrics from FASTQC and QUAST, the resulting genome assembly was determined to be of high quality.

## Introduction

*Fusarium* is a large, cosmopolitan genus of filamentous Ascomycete fungi belonging to the order Hypocreales (Nectriaceae), ranked among the most species-rich and economically destructive fungal groups worldwide (Crous et al. 2021). It encompasses a vast array of ubiquitous and phylogenetically distinct species, a diversity that has long attracted attention due to its intricate and, at times, controversial taxonomy, as well as its profound impact on agriculture (Summerell et al. 2019).

Taxonomically, *Fusarium* species are classified within the Ascomycota phylum and are characterized by distinctive morphological features such as canoe-shaped macroconidia, the production of chlamydospores, and the formation of both sexual and asexual reproductive structures. Historically, the genus was circumscribed by Link in 1809 based on multiseptate macroconidia, but later studies revealed that this character evolved convergently in different Ascomycete lineages and was even lost within the *Fusarium solani* species complex (Gräfenhan

**et al., 2011; O'Donnell et al., 2013).** For a long time, there has been confusion in *Fusarium* taxonomy due to **Snyder and Hansen's (1940)** nine-species system, misleading overlaps caused by convergent evolution and character loss, the phenomenon of cultural degeneration, and the strong opinions of taxonomists and plant pathologists who have worked on them. The genus has always been a taxonomically controversial group, whereupon the number of *Fusarium* species has varied from nine to more than a thousand depending on the taxonomists (**Summerell and Leslie 2011**). Yet, the species were later redistributed into species complexes after the introduction of modern molecular tools (**O'Donnell et al. 2000 a & b; Geiser et al. 2013; O'Donnell et al. 2013; Aoki et al. 2014**). Advances in molecular phylogenetics subsequently redefined *Fusarium* as a monophyletic genus comprising at least 20 species complexes and several independent lineages, a framework now widely accepted (**Geiser et al., 2013; O'Donnell et al., 2013**). The most recent *Fusarium* phylogeny arising from a 19 orthologous protein-coding gene study was proposed by Geiser et al. (**2021**), who identified the largest species complexes as *F. sambucinum* (FSAMSC), *F. incarnatum-equiseti* (FIESC), *F. fujikuroi* (FFC), *F. oxysporum* (FOSC), and one of the more basal lineages, *F. solani* (FSSC). Currently, a total of 23 species complexes are recognized within the genus *Fusarium*, englobing 431 *Fusarium* species (**Geiser et al., 2021**). However, thousands more species await formal taxonomic classification.

The economic and agricultural importance of *Fusarium* lies primarily in its pathogenic members, which are responsible for devastating diseases such as *Fusarium* wilt, root rot, and head blight in staple crops including wheat, maize, and legumes. These infections cause yield losses and quality degradation, threatening food security at a global scale. Interestingly, there is increasing evidence that some *Fusarium* species also live as endophytes and benefit plant growth and stress tolerance (**Maciá-Vicente et al. 2009; Skiada et al. 2019**), suggesting that our current knowledge of *Fusarium* biology and their strain-specific interactions with host plants is still incomplete. In fact, *Fusarium* is not restricted to pathogenic lifestyles: some species occur as soil saprophytes, endophytes (**Walsh et al. 2010; Cavalcanti et al. 2020**), latent pathogens, or even fungicolous fungi (**Torbati et al. 2021**), while others are associated with animals, including humans (**Santos et al. 2020 ; Meza-Menchaca et al. 2020**).

In one study conducted by **Ben Alaya et al. (2024)** to identify the causal agent of wilt disease observed on *Medicago sativa* plants in southern Tunisian oases, 79 fungal strains were isolated

from roots, collars, and stems. The molecular identification showed that 73.42% of the isolated fungal strains belong to the genus *Fusarium*, including the species complexes of *F. oxysporum*, *F. solani*, *F. fujikuroi*, *F. incarnatum equiseti*, etc. The inoculation of the fungal strains on *Medicago sativa* and *Medicago truncatula* showed that some of the strains enhanced plant growth (6–16%), in particular, the strain TZ1MST22 of *Fusarium* sp. exhibited plant growth-promoting (PGP) properties on both *Medicago* species and on date palm. Due to its high potential as PGP agent, the whole genome of TZ1MST22 strain was sequenced in order to characterize the genetic determinants underlying its beneficial traits and to assess its safety for agricultural use.

## **Material and methods**

### **Fungal strain isolation**

The fungal strain TZ1MST22 was isolated from the stems of *Medicago sativa* plants showing wilting symptoms collected from an oasis in Tozeur, southern Tunisia (GPS coordinates: longitude: 8.163181, latitude: 33.935854). The isolation procedure was performed following the methodology outlined by **Ben Alaya et al. (2024)**.

### **DNA extraction**

DNA extraction was carried out by MicrobesNG (Birmingham, UK) following their standard protocols. The fungal strain was shipped in compliance with the instructions that MicrobesNG supplied, available on their website (<https://microbesng.com>). The strain was cultivated as single-spore culture on PDA (Potato Dextrose Agar; Pronadisa, Madrid, Spain) medium at  $25 \pm 2$  °C in darkness.

### **Illumina NovaSeq 6000 sequencing**

Genomic DNA libraries were prepared using the Nextera XT Library Prep Kit (Illumina, San Diego, USA), following the manufacturer's protocol with two modifications: the input DNA quantity was doubled, and the PCR elongation time was extended to 45 seconds. Sequencing was performed on an Illumina NovaSeq 6000 platform using a 250 bp paired-end strategy. Adapter trimming of raw reads was carried out using Trimmomatic v 0.30 (**Bolger et al., 2014**), applying a sliding window quality cutoff of Q15. De novo assembly of the trimmed reads was conducted using SPAdes v3.7 (**Bankevich et al., 2012**).

## Sequence processing

Read quality scores were assessed with FastQC version 0.11.2 (Andrews, 2010) implemented in the Galaxy platform (<https://usegalaxy.org/>). FastQC was executed with default parameters. The quality of the assembled genome was assessed using QUAST version 5.2.0 (Gurevich et al., 2013). The pipeline was executed with the following key parameters: a minimum contig length of 500 bp, minimum alignment identity of 95%, and minimum alignment length of 65 bp. Gene prediction, ribosomal RNA detection, and conserved single-copy ortholog assessment were enabled through integrated modules, including Barrnap and BUSCO (using AUGUSTUS).

## Data availability

The assembled genome has been deposited in the NCBI GenBank under BioProject ID PRJNA930339, and Biosample ID SAMN33005115 with the associated accession numbers JBRJOB000000000 (assembly) and SRS16690020 (raw reads). This version represents the first draft genome of the non-pathogenic *Fusarium* strain TZ1MST22.

## Results & discussion

In total, 13,996,888 high-quality clean readings were obtained (Table 1). The genome assembly of the *Fusarium* strain TZ1MST22 was evaluated using the Quast tool v 5.2.0, which provided critical metrics for the quality of the de novo assembly achieved. Among the 301 contigs obtained, 140 exceeded 1000 bp in length, and 141 were less than 500 bp in size. Finally, 160 contigs longer than 500 bp were retained for additional analysis. The longest contig measures 3072088 bp, and the estimated total genome size is 39756103 bp. The GC content is 47.93%. The assembly exhibits a high degree of continuity with a N50 of 1209769 bp. The N90 value of 219101 bp indicates that contigs of at least this size make up 90% of the genome. The L50 and L90 values are documented as 12 and 42, respectively, indicating that the assembly is relatively compact, requiring just a small number of contigs to cover 50% and 90% of the genome. Furthermore, the average coverage depth is high (167x), with almost the entire genome covered at least 10x (99.95 %), indicating good sequencing depth and genome coverage (Table 1).

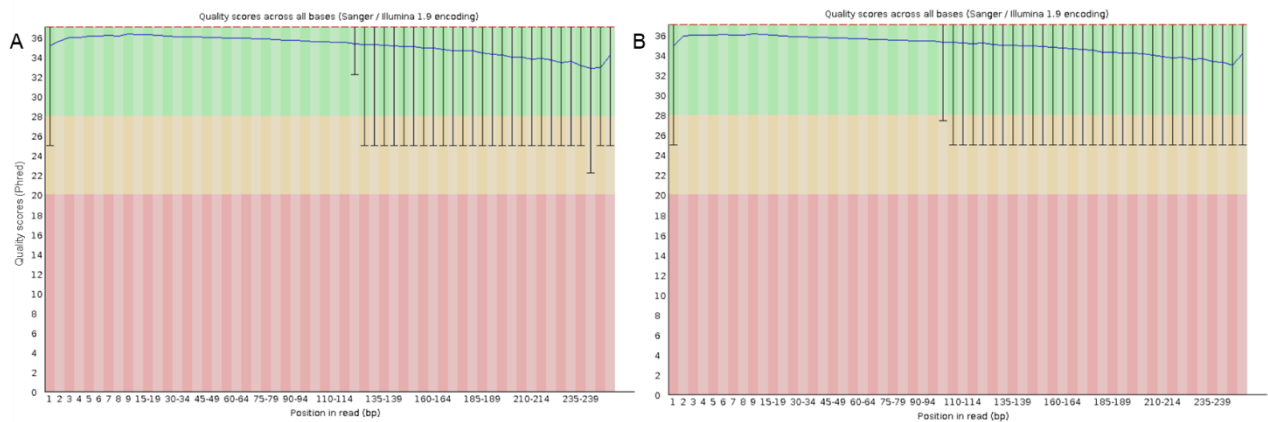
**Table 1.** Assembly statistics of the whole genome sequence of TZ1MST22 *Fusarium* sp. strain.

| Genome features                | Value    |
|--------------------------------|----------|
| No. contigs                    | 160      |
| No. contigs ( $\geq 1000$ bp)  | 140      |
| Total length ( $\geq 0$ bp)    | 39813733 |
| Total length ( $\geq 1000$ bp) | 39742206 |
| Largest contig                 | 3072088  |
| Genome size (bp)               | 39756103 |
| GC (%)                         | 47,93    |
| N50                            | 1209769  |
| N75                            | 497463   |
| N90                            | 219101   |
| L50                            | 12       |
| L75                            | 25       |
| L90                            | 42       |
| Average coverage depth         | 167      |
| No. tRNA                       | 337      |
| No rRNA                        | 76       |
| Coverage $\geq 5x$             | 99.97 %  |
| Coverage $\geq 10x$            | 99.95 %  |

Read quality scores were assessed with FastQC version 0.11.2 (Andrews, 2010) implemented in the Galaxy platform (<https://usegalaxy.org/>). The results of this analysis are detailed in the section below.

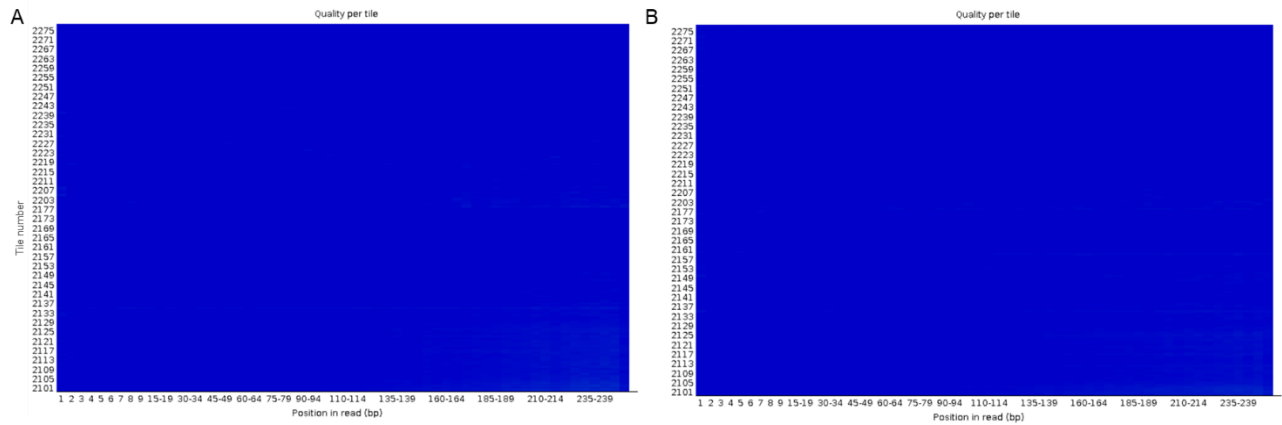
The Figure 1 depicts the quality (Phred score, on the y-axis) of each base (on the x-axis) across all reads in the dataset. At each read position, a boxplot depicts the quality of all reads. The median appears in red, while the mean is in blue. The quality profiles showed consistently high Phred scores at all base positions. Specifically, the median quality ratings for all base positions exceeded 37, indicating base call accuracy greater than 99.98%. The color codes indicate the scores for extremely good quality (green), good quality (orange), and low quality (red). Our values are well

within the "green" quality zone of the FastQC plots, indicating high base-calling accuracy throughout the read length. Quality scores varied minimally at each position. The absence of interquartile ranges indicates that the great majority of reads remained of uniformly good quality throughout the sequencing run. Furthermore, there were no noticeable quality decreases at the start or end of the reads, which is frequent in lower-quality sequencing runs (Figure 1).



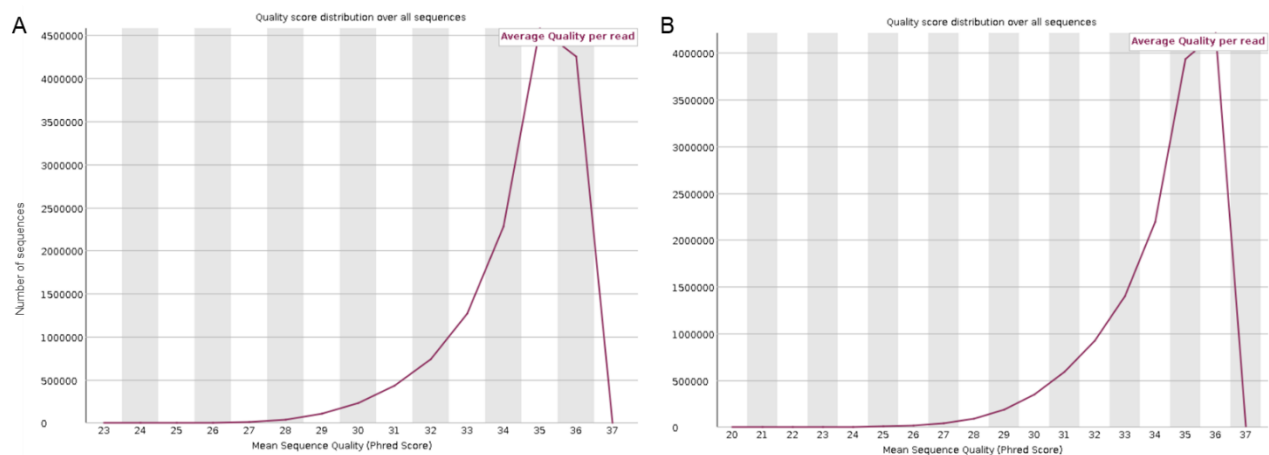
**Figure 1.** Per base sequence quality profiles for (A) forward reads and (B) reverse reads.

Per-tile sequence quality analysis, which is particularly relevant to Illumina-based sequencing data, evaluates the consistency of base-calling accuracy across different regions of the flow cell. Each tile corresponds to a distinct physical section of the flow cell, and the quality distribution across these tiles can reveal systematic issues in specific areas during the sequencing process. In our dataset (Figure 2), the per-tile quality heatmaps for both forward and reverse reads displayed a uniform blue coloration, indicating minimal deviation from the average per-base quality. The predominance of cool colors across all tiles reflects a high level of consistency and uniform quality across the entire flow cell, with no evidence of localized quality degradation. This uniformity suggests optimal performance of the sequencing instrument during the run and supports the overall high quality of the dataset.



**Figure 2.** Per tile sequence quality profiles for (A) forward reads and (B) reverse reads.

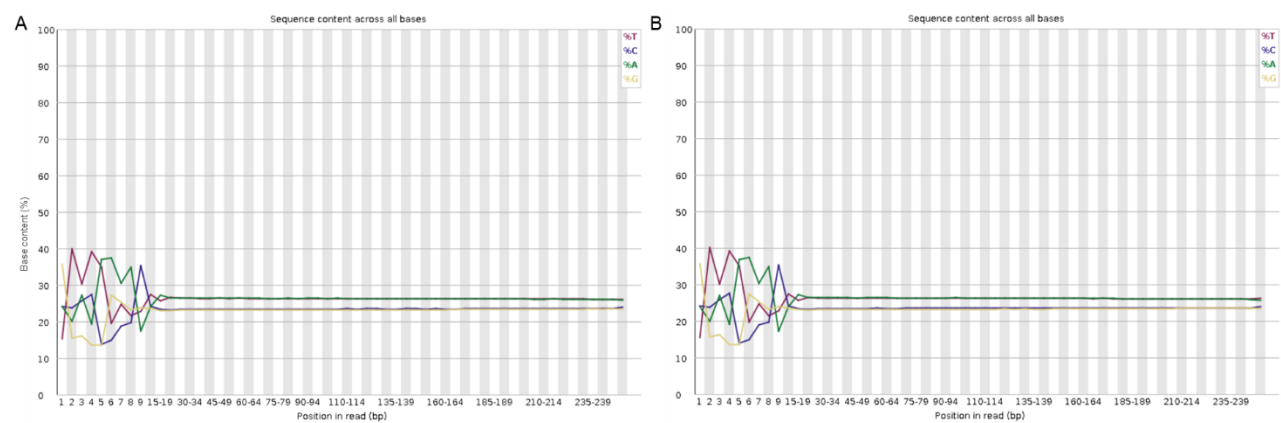
The distribution of average quality scores per read was evaluated to assess the overall accuracy of the sequencing data. In this analysis, the x-axis represents the average Phred quality score for individual reads, while the y-axis indicates the number of reads corresponding to each score. Our results (Figure 3) showed a sharp, unimodal distribution with a pronounced peak between Phred scores of 35 and 37, indicating that the majority of reads exhibit high base-calling accuracy. The steep curve observed at the higher end of the quality spectrum is characteristic of high-quality sequencing runs and reflects the overall consistency and reliability of the dataset.



**Figure 3.** Per sequence quality scores for (A) forward reads and (B) reverse reads.

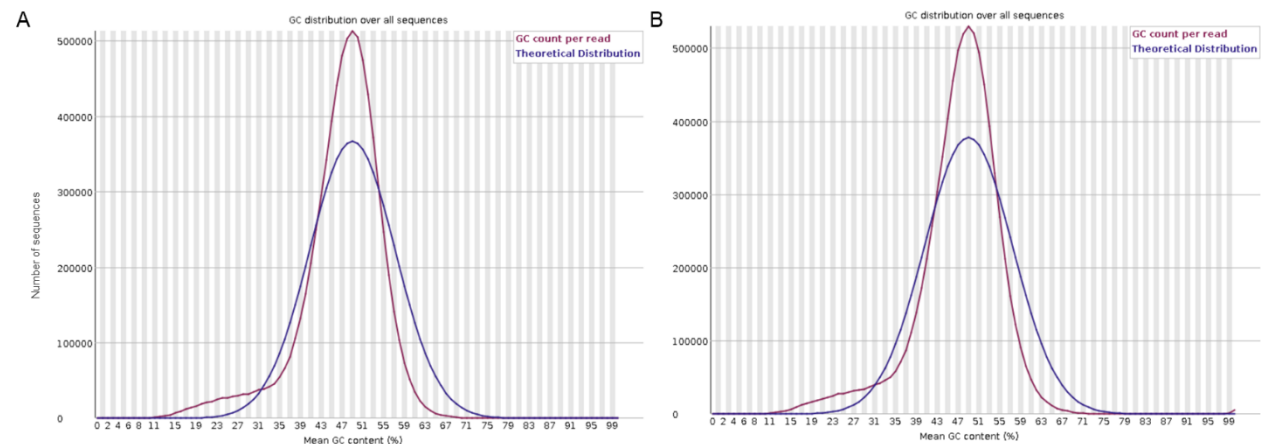
The per base sequence content analysis displays the proportion of each nucleotide (A, T, G, and C) at every position across all reads. In a random library, especially one without sequence-specific bias (e.g., from random priming or fragmentation), the relative abundance of each base is expected

to be relatively uniform across read positions, with the base composition lines running roughly parallel. In our dataset, the sequence content plot for both forward and reverse reads showed near-identical proportions of adenine and thymine, as well as guanine and cytosine, across all positions. This symmetry suggests an absence of strong base composition bias and supports the conclusion that the library preparation and sequencing processes maintained a balanced representation of nucleotides (Figure 4).



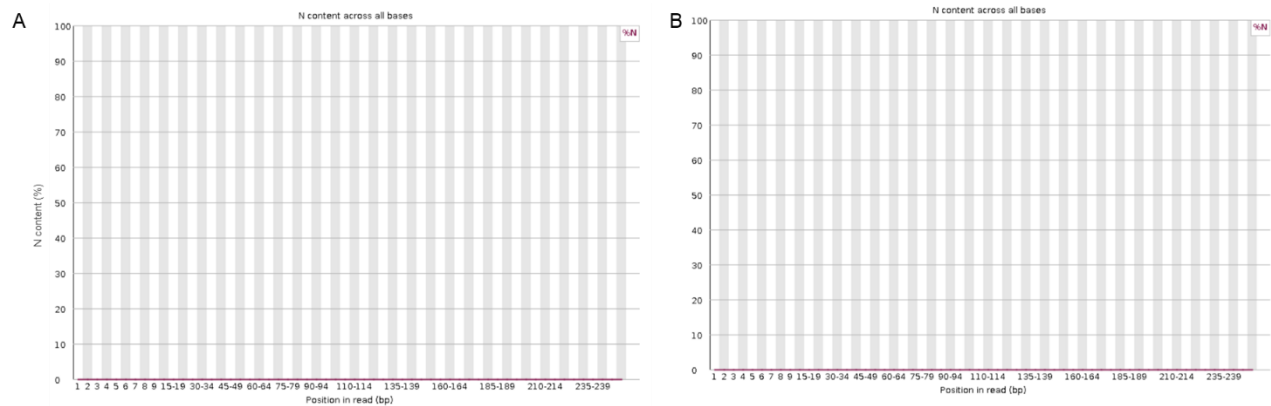
**Figure 4.** Per base sequence content for (A) forward reads and (B) reverse reads.

The per sequence GC content showed a unimodal distribution centered at 47%, consistent with the expected composition of the target genome. The absence of significant deviation or abnormal patterns indicates a high-quality library with no evident sequence bias (Figure 5).



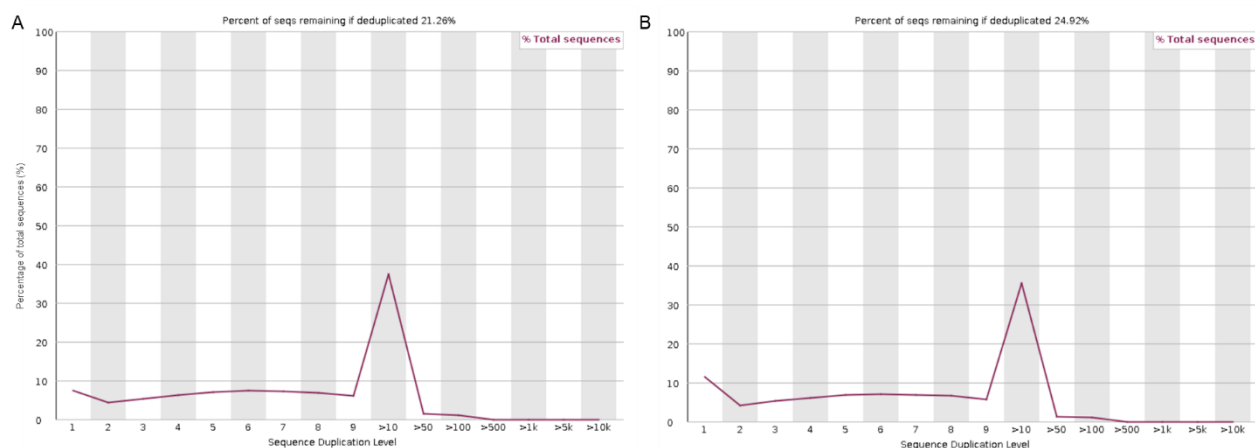
**Figure 5.** Per sequence GC content for (A) forward reads and (B) reverse reads.

The per base N content analysis indicates the proportion of undetermined bases (N) at each position in the reads. In high-quality sequencing runs, the occurrence of N bases is expected to be minimal or absent. In our dataset (Figure 6), the N content was consistently 0% across all positions, confirming the absence of ambiguous base calls and indicating excellent sequencing accuracy.



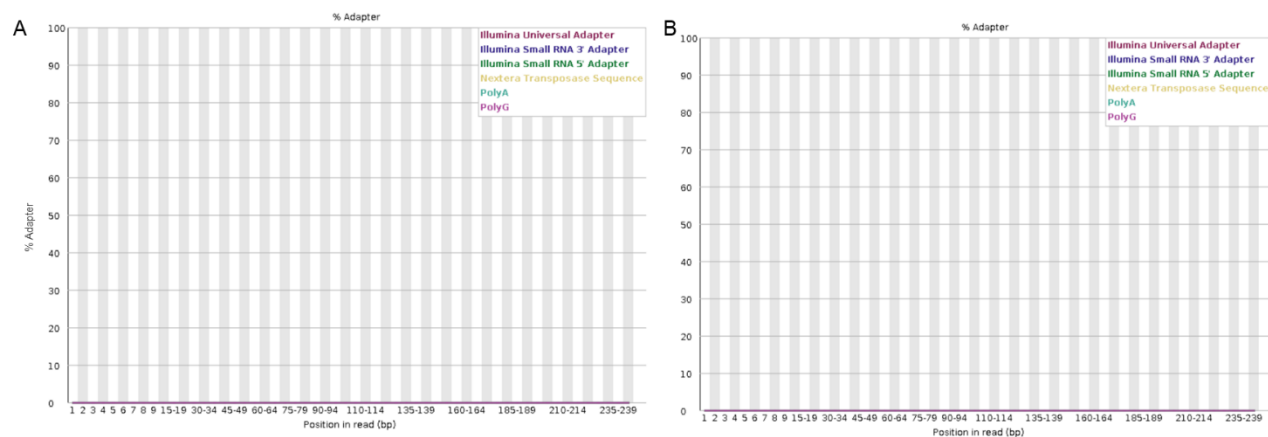
**Figure 6.** Per base N content for (A) forward reads and (B) reverse reads.

Sequence duplication analysis revealed that 21.26% of forward reads and 24.92% of reverse reads would remain unique after deduplication, indicating a moderate level of sequence redundancy. A slight peak was observed beyond the 10-duplicate level, with around 40% of sequences in this range. However, no overrepresented sequences were detected in the dataset, suggesting that the duplication is not driven by specific contaminant sequences or dominant fragments. Instead, it likely reflects normal duplication associated with sequencing depth and library preparation. Overall, these results do not indicate any major quality concerns regarding library complexity (Figure 7).



**Figure 7.** Sequence duplication levels for (A) forward reads and (B) reverse reads.

Adapter content analysis showed no detectable presence of adapter sequences in either forward or reverse reads. This indicates effective adapter trimming during preprocessing or high-quality library preparation with minimal adapter contamination, supporting the overall integrity of the sequencing data (Figure 8).



**Figure 8.** Adapter content for (A) forward reads and (B) reverse reads.

## Conclusion

The genomic resources generated in this project provide new tools to study non-pathogenic *Fusarium* strains in order to identify genes and pathways associated with plant-growth promotion properties, to predict secondary-metabolite biosynthetic gene clusters involved in pathogen antagonism and plant-growth promotion effects, and to screen for the absence of known virulence determinants. Comparative genomic analyses with reference *Fusarium* genomes will also enable

us to place TZ1MST22 in a phylogenomic context, confirm its relatedness to non-pathogenic strains, and evaluate its suitability for development as a safe agricultural biofertilizer.

### **Acknowledgement**

The sequencing of the fungal strain was supported by GetGenome and The Sainsbury Laboratory, Norwich, UK, with support from the Gatsby Charitable Foundation and the Biotechnology and Biological Sciences Research Council (BBSRC). We extend our heartfelt thanks to Pr. Sophien Kamoun for his insightful guidance and continuous support. Our sincere appreciation also goes to Dr. James Canham and Dr. Joe Win for their generous technical assistance.

### **Statement of continuing work**

The datasets are being shared prior to their formal publication and should be regarded as preliminary. We value and invite feedback from the community. Please contact us at (amani.ben.alaya@gmail.com or dnaceur@yahoo.fr) for additional information.

### **References**

- Andrews, S. (2010). FastQC: a Quality Control Tool for High Throughput Sequence Data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Aoki, T., O'Donnell, K., & Geiser, D. M. (2014). Systematics of key phytopathogenic *Fusarium* species: current status and future challenges. *Journal of General Plant Pathology*, 80(3), 189-201.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... & Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, 19(5), 455-477.
- Ben Alaya, A., Khiari, B., Ben Slimene, I. et al. Diversity and spectrum of pathogenicity of fungi associated with wilt disease of alfalfa in southern Tunisian oases. *Eur J Plant Pathol* 169, 799–823 (2024). <https://doi.org/10.1007/s10658-024-02871-8>.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.
- Cavalcanti, A.D., Santos, A.C.S., de Oliveira Ferro, L., Bezerra, J.D., Souza-Motta, C.M., Magalhães, O.M.C. (2020). *Fusarium massalimae* sp. nov. (*F. lateritium* species complex) occurs

endophytically in leaves of *Handroanthus chrysotrichus*. *Mycol Prog.* 19(10):1133–1142.  
<https://doi.org/10.1007/s11557-020-01622-3>.

Crous, P.W., Hernández-Restrepo, M., van Iperen, A.L., Starink-Willemse, M., Sandoval-Denis, M., Groenewald, J.Z. (2021). Citizen science project reveals novel fusarioid fungi (Nectriaceae, Sordariomycetes) from urban soils. *Fung Syst Evol* 8:101–127.  
<https://doi.org/10.3114/fuse.2021.08.09>.

Geiser, D. M., Aoki, T., Bacon, C. W., Baker, S. E., Bhattacharyya, M. K., Brandt, M. E., et al. (2013). LETTER TO THE EDITOR: One fungus, one name: Defining the genus *Fusarium* in a scientifically robust way that preserves longstanding use. *Phytopathology*, 103, 400–408.

Geiser, D. M., Al-Hatmi, A. M., Aoki, T., Arie, T., Balmas, V., Barnes, I., ... & Viljoen, A. (2021). Phylogenomic analysis of a 55.1-kb 19-gene dataset resolves a monophyletic *Fusarium* that includes the *Fusarium solani* species complex. *Phytopathology*®, 111(7), 1064-1079.

Gräfenhan, T., Schroers, H.-J., Nirenberg, H. I., & Seifert, K. A. (2011). An overview of the taxonomy, phylogeny and typification of nectriaceous fungi in *Cosmospora*, *Acremonium*, *Fusarium*, *Stilbella* and *Volutella*. *Studies in Mycology*, 68, 79–113.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072-1075.

Maciá-Vicente, J. G., Rosso, L. C., Ciancio, A., Jansson, H. B., & Lopez-Llorca, L. V. (2009). Colonisation of barley roots by endophytic *Fusarium equiseti* and *Pochonia chlamydosporia*: effects on plant growth and disease. *Annals of Applied Biology*, 155(3), 391-401.

Meza-Menchaca, T., Singh, R.K., Quiroz-Chávez, J., García-Pérez, L.M., Rodríguez-Mora, N., Soto-Luna, M., Quiroz-Figueroa, F.R. (2020). First demonstration of clinical *Fusarium* strains causing cross-kingdom infections from humans to plants. *Microorganisms* 8(6):947.  
<https://doi.org/10.3390/microorganisms8060947>.

O'Donnell, K. (2000 a). Molecular phylogeny of the *Nectria haematococca-Fusarium solani* species complex. *Mycologia*, 92(5), 919-938.

O'Donnell, K., Nirenberg, H. I., Aoki, T., & Cigelnik, E. (2000 b). A multigene phylogeny of the *Gibberella fujikuroi* species complex: detection of additional phylogenetically distinct species. *Mycoscience*, 41(1), 61-78.

O'Donnell, K., Rooney, A. P., Proctor, R. H., Brown, D. W., McCormick, S. P., Ward, T. J., et al. (2013). Phylogenetic analyses of RPB1 and RPB2 support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria. *Fungal Genet. Biol.* 52, 20–31. doi: 10.1016/j.fgb.2012.12.004.

Santos, A.C.S., Diniz, A.G., Tiago, P.V., Oliveira, N.T. (2020). Entomopathogenic *Fusarium* species: a review of their potential for the biological control of insects, implications and prospects. *Fungal Biol Rev* 34(1):41–57. <https://doi.org/10.1016/j.fbr.2019.12.002>

Skiada, V., Faccio, A., Kavroulakis, N., Genre, A., Bonfante, P., & Papadopoulou, K. K. (2019). Colonization of legumes by an endophytic *Fusarium solani* strain FsK reveals common features to symbionts or pathogens. *Fungal Genetics and Biology*, 127, 60-74.

Snyder, W. C., & Hansen, H. N. (1940). The species concept in *Fusarium*. *American Journal of Botany*, 64-67.

Summerell B.A. (2019). Resolving *Fusarium*: current status of the genus. *Annu Rev Phytopathol* 57:323–339. <https://doi.org/10.1146/annurev-phyto-082718-100204>.

Summerell, B. A., & Leslie, J. F. (2011). Fifty years of *Fusarium*: how could nine species have ever been enough?. *Fungal Diversity*, 50(1), 135-144.

Torbati, M., Arzanlou, M., Santos, A.C.S. (2021). Fungicolous *Fusarium* species: Ecology, diversity, isolation, and identification. *Curr Microbiol* 78:2850–2859. <https://doi.org/10.1007/s00284-021-02584-9>.

Walsh, J.L., Laurence, M.H., Liew, E.C.Y., Sangalang, A.E., Burgess, L.W., Summerell, B.A., Petrovic, T. (2010). *Fusarium*: two endophytic novel species from tropical grasses of northern Australia. *Fungal Divers* 44:149–159. <https://doi.org/10.1007/s13225-010-0035-3>.