The PEG-responding desiccome of the alder microsymbiont Frankia alni

Kais Ghedira1, Emna Harigua-Souiai2, Cherif Ben Hamda1,3, Pascale Fournier4, Petar Pujic4, Sihem Guesmi5,6, Ikram Guizani2, Guylaine Miotello7, Jean Armengaud7, Philippe Normand4 & Hâïtham Sghaier5,8

Actinorhizal plants are ecologically and economically important. Symbiosis with nitrogen-fixing bacteria allows these woody dicotyledons to colonise soils under nitrogen deficiency, water-stress or other extreme conditions. However, proteins involved in xerotolerance of symbiotic microorganisms have yet to be identified. Here we characterise the polyethylene glycol (PEG)-responding desiccome from the most geographically widespread Gram-positive nitrogen-fixing plant symbiont, Frankia alni, by next-generation proteomics, taking advantage of a Q-Exactive HF tandem mass spectrometer equipped with an ultra-high-field Orbitrap analyser. A total of 2,052 proteins were detected and quantified. Under osmotic stress, PEG-grown F. alni cells increased the abundance of envelope-associated proteins like ABC transporters, mechano-sensitive ion channels and Clustered Regularly Interspaced Short Palindromic Repeats CRISPR-associated (cas) components. Conjointly, dispensable pathways, like nitrogen fixation, aerobic respiration and homologous recombination, were markedly down-regulated. Molecular modelling and docking simulations suggested that the PEG is acting on Frankia partly by filling the inner part of an up-regulated osmotic-stress large conductance mechanosensitive channel.

Actinobacteria belonging to the genus Frankia do establish nitrogen-fixing nodular symbiosis with the roots of 23 angiosperm genera that are collectively called “actinorhizals”5. These plants form root nodules in which Frankia fixes nitrogen, thus permitting them to thrive in pioneer soils poor in nitrogen and organic matter, such as glacial moraines, lava fields, forest burnouts or anthropogenic sites such as mine spoils or hydrodam dykes6. Frankia establishes a symbiotic association with the roots of several dicotyledonous plants. The different Frankia lineages form a coherent cluster at the root of the aerobic actinobacteria phylum7, and F. alni in particular establishes symbiosis with alder (Alnus) and bayberry (Morella) species8.

The interaction has evolved over several million years with a sophisticated dialogue that does not imply acylated N-acetyl-glucosamine oligomeric Nod factors9. The Frankia determinants of symbiosis are still poorly known, which is for the most part due to the lack of a genetic transformation system. Transcriptomics has shown genes coding for nitrogenase (nif), hydrogenase uptake (hup), hopanoids (shc, hpn), iron-sulfur (suf) clusters as among the most up-regulated10. Proteomics has also been used to analyze the symbiosis11 to identify up-regulated peptides and it has shown the presence of Nif, Hup, Suf, Hop proteins as expected but also several transporters, regulators and various proteins involved in stress responses.

1Laboratory of Bioinformatics, Biomathematics and Biostatistics - LR16IPT09, Institut Pasteur de Tunis, Université de Tunis el Manar, Tunis, 1002, Tunisia. 2Laboratory of Molecular Epidemiology and Experimental Pathology - LR11IPT04, Institut Pasteur de Tunis, Université de Tunis el Manar, Tunis, 1002, Tunisia. 3Université de Carthage, Faculté des Sciences de Bizerte, Tunis, 7021, Tunisia. 4Université de Lyon, Université Lyon 1, Lyon; CNRS, UMR 5557, Ecologie Microbienne, UMR1418, INRA, 69622 Cedex, Villeurbanne, France. 4Laboratory "Energy and Matter for Development of Nuclear Sciences" (LR16CNSTN02), National Center for Nuclear Sciences and Technology (CNSTN), Sidi Thabet Technopark, 2020, Tunisia. 4National Agronomy Institute (INAT), Avenue Charles Nicolle, 1082, Tunis, Mahrajène, Tunisia. 5Laboratoire Innovations Technologiques pour la Détection et le Diagnostic (Li2D), Service de Pharmacologie et Immunoanalyse (SPI), CEA, INRA, F-30207, Bagnols sur Cèze, France. 6Associated with Laboratory "Biotechnology and Nuclear Technology" (LR16CNSTN01) & Laboratory "Biotechnology and Bio-Geo Resources Valorization" (LR11ES31), Sidi Thabet Technopark, 2020, Tunisia. Correspondence and requests for materials should be addressed to P.N. (email: philippe.normand@univ-lyon1.fr)
Osmotic stress is a constant challenge for bacteria living in a range of soils\textsuperscript{4,9}, from those having low salinity in rainy cold latitudes to highly saline ones in warm and dry parts of the world. It also affects symbionts\textsuperscript{10} that must alternate between two main biotopes, the soil and the root tissues. Soils, especially poor soils, have low osmotic potential while plant tissues have a much higher osmotic potential. Most studies on the effect of salinity on actinorhizal symbionts have focused on the plant, especially \textit{Casuarina} to illustrate various adaptations such as sodium partitioning or proline accumulation\textsuperscript{11}. Only a few recent studies on \textit{Frankia} have shown the effects of salinity on ammonium (NH\textsubscript{4}\textsuperscript{+}) assimilation\textsuperscript{12}, on cell wall/membrane biogenesis functions and on some transport proteins\textsuperscript{13}. \textit{Almus-}, \textit{Casuarina-} and \textit{Elaegnus-} infective isolates grow best at 50 mM but are nevertheless able to grow well in medium containing up to 200 mM NaCl, but not at 500 mM\textsuperscript{14}. The effect of NaCl on one strain, \textit{Cc16}, was apparently less severe than that of another osmolyte, sucrose\textsuperscript{15}. Nevertheless, little is known about the molecular adaptations and the present study was undertaken to better understand how \textit{Frankia} cope with osmolytes using proteogenomics, an approach recently used to better decipher how \textit{Geodermatophilaceae} cope with desiccation\textsuperscript{16}.

NaCl causes various impairments to cells, some related to osmotic stress and some related to ionic stress and these two effects are clearly distinguishable\textsuperscript{17}. PEG is a non-ionic polymer of various molecular weights (MW) that has been used often for osmotic studies when it is sought to have only the osmotic effect.

Mechanisms involved in desiccation tolerance have been well studied in Cyanobacteria\textsuperscript{18}, Actinobacteria\textsuperscript{19,20} as well as xerotolerant members of the \textit{Bosea}, \textit{Chelatococcus}, \textit{Deinococcus} and \textit{Methylobacterium} genera\textsuperscript{21}. Briefly, the different studies have shown that desiccation damages cell membranes, proteins and DNA due to oxidative stress and production of reactive oxygen species (ROS)\textsuperscript{22,23}. Strategies adopted by these microorganisms for desiccation tolerance include prevention of ROS damage, osmoprotection through the accumulation of sucrose and/or trehalose\textsuperscript{24,25}, the production of extracellular polysaccharides (EPS) that reduce water loss\textsuperscript{26}, or the ability to limit protein oxidation during dehydration\textsuperscript{27}. Previous studies\textsuperscript{20,21,25–27} highlighted changes in expression of several genes following water stress including genes for synthesis of trehalose/sucrose, sugar transporters, chaperone genes (groES/EL, \textit{dnaK}/\textit{J}), oxidative stress protection genes (\textit{dps}, thioredoxin), ABC transporters, dehydrogenases, esterases, proteases, hydrolases and lyases. A \textit{Casuarina-}infective \textit{Frankia} strain was recently shown to have few proteins upregulated under salt stress, among which cell wall/membrane biogenesis functions and some transport proteins\textsuperscript{13}.

The desiccome, a term coined by Potts and colleagues\textsuperscript{28}, can be defined as the set of genes, proteins and metabolites that are necessary for desiccation tolerance\textsuperscript{29}. Here we report the first PEG-responding desiccome of a nitrogen-fixing symbiotic bacterium, \textit{F. alni}. We highlight the role of ABC transporters, mechanosensitive ion channels and Cas components in response to this stress.

Results

Growth in polyethylene glycol (PEG). Upon inoculation in BAP- medium at a 0.1 OD\textsubscript{600} growth of \textit{F. alni} strain ACN14\textsuperscript{a} was inhibited at all concentrations of PEG tested. Nitrogen fixation was active without PEG but even the lowest concentration of PEG completely inhibited it (data not shown). We chose an inoculation density of 0.1 to have sufficient biomass and after 7 days, growth without PEG had reached an OD\textsubscript{600} of 0.677, i.e. a 0.74 fold in density increase. In the presence of PEG 0.9\%, OD\textsubscript{600} reached 0.42 or a 0.47 fold increase. Growth in the presence of PEG was thus 63\% that without PEG. Hyphae, vesicles and sporangia looked similar under both conditions.

Comprehensive proteome coverage of \textit{F. alni} cells. Thanks to the high speed and high resolution of this analyser, a total of 295,788 high quality MS/MS spectra were recorded. Amongst these, a set of 208,018 MS/MS spectra (70.3\%) was assigned to peptide sequences. This dataset allowed detecting a total of 20,825 peptide sequences, which pointed at 2,454 polypeptides, with 2,052 proteins certified with at least two distinct peptide sequences. These proteins and their characteristics are listed in Supplementary Data, Table S1. On the average, a total of 20,801 spectral counts were measured per sample with a remarkably low standard deviation (5.3\%). With an average of 10 peptides and 100 spectral counts per polypeptide, an extensive coverage of the proteome was reached that compares favourably to most proteomic studies of bacteria\textsuperscript{18,30,31}.

Differentially expressed proteins. When comparing both conditions and compiling the five biological replicates of each condition, a total of 1,951 proteins were detected whatever the conditions. A set of 29 proteins was specifically detected only in the reference while 85 were detected only in the PEG-treated cells. Differentially detected protein abundances were identified using a Fold Change (FC) threshold of 1.5 and \textit{p}-value below 0.05. Four classes were delineated as follows: i) Blue class proteins for which identifications satisfied both, the fold change (>1.5) and statistical criteria (\textit{p}-value < 0.05); ii) Orange class proteins for which identifications did not meet the fold criterion but have low \textit{p}-values; iii) Green class proteins for which identifications satisfied the fold criterion but not the statistical criterion; iv) and finally Red class for which identifications did not meet the fold and \textit{p}-value criteria. The Blue class comprised 294 proteins. A total of 211 proteins were more abundant in PEG-treated cells compared to the reference while 83 were found less abundant. Figure 1 shows the distribution of the proteins affected by the PEG treatment compared to the control within the \textit{F. alni} genome and their expression levels. Genes coding for more abundant proteins upon PEG-induction are highlighted in red whereas those more abundant in the control were highlighted in cyan in Fig. 1. Tables 1 and 2 list the most up- and down-regulated PEG-modulated proteins of \textit{F. alni}, respectively. Twelve ribosome-associated translation proteins (FRAAL1099–50S ribosomal subunit protein L30 (2.10), etc.) were also up-regulated by PEG but are not shown in Table 1.

Re-annotation of differentially expressed proteins following polyethylene glycol (PEG) treatment. Among the blue class differentially expressed proteins (294 proteins) identified under PEG treatment,
53 were characterised as hypothetical proteins and/or conserved hypothetical proteins. We therefore conducted a new analysis using PSI-BLAST tool to reannotate them. Seventeen ORFs remained as proteins with unknown function because of their low similarity to characterised proteins and the function of 36 ORFs could be deduced from closely-related known proteins (Supplementary Data, Table S2. Re-annotation of differentially expressed proteins following polyethylene glycol (PEG) treatment based on a Psi-Blast approach.). Among re-annotated ORFs, several envelope-associated proteins were identified including the most up-regulated protein (FRAAL2567 (5.67)), a putative auto-transporter adhesin. These results consolidate previous data (see Table 1) that have evidenced the extreme importance of envelope-associated proteins in response to water stress.

**Functional analysis.** The functional enrichment analysis applied on the lists of differentially expressed proteins following PEG treatment, covering Gene Ontology and pathway annotations, was performed through the STRING database. Among the enriched KEGG pathways within the blue class differentially expressed proteins, we denote the general metabolic pathways (map01100, FDR = 0.000147), ribosome (map03010, FDR = 0.000147) and sesquiterpenoid and triterpenoid biosynthesis (map00909, FDR = 0.0032). Among the most significantly enriched biological processes associated to the blue class of differentially expressed proteins, we found: nitrogen compound metabolic process (GO.0006807, FDR = 7.79e-05), metabolic process (GO.0008152, FDR = 0.00022), protein metabolic process (GO.0019538, FDR = 0.00022), cellular nitrogen compound metabolic process (GO.0034641, FDR = 0.00022), cellular metabolic process (GO.0044237, FDR = 0.00022) and translation...
(GO:0006412, FDR = 0.000706). Outputs of biological processes enrichment data were analysed with GOPlot to generate insightful plots. Figure 2 and Supplementary Figures 1 and 2 highlight the most enriched biological processes associated to the blue differentially expressed, orange and red class proteins, respectively. There are no significantly enriched biological processes in the green class proteins.

Among the top blue class enriched biological processes (Fig. 2), we denote the nitrogen compound metabolic process, which includes proteins belonging to Nif family (e.g., FRAAL6811/NifK), Rpm family (e.g., FRAAL1926/RpmA), Rpl family (e.g., FRAAL5214/RplT) and other proteins like FRAAL6814/NifV, FRAAL6803/NifB, FRAAL6810/NifD, FRAAL6809/NifE, FRAAL2491/Shc2, FRAAL6797/Ferredoxin, FRAAL6799/2-oxoglutarate ferredoxinoxidoreductase beta-subunit, FRAAL1432/Scg1, FRAAL4563/Transport protein associated with Fe-S cluster assembly, FRAAL4147/Cytochrome c oxidase polypeptide I (AA3 subunit 1), FRAAL5625/Poly(3-hydroxybutyrate) depolymerase, FRAAL6005/Biotin carboxyl carrier protein, FRAAL6082/Succinate-semialdehyde dehydrogenase I, NADP-dependent, FRAAL5481/Malate synthase, FRAAL0315/Aldolase, FRAAL1829/Aldehyde dehydrogenase, FRAAL1039/ADH-quinone oxidoreductase chain, FRAAL6528/ATP-dependent protease, DNA repair protein, and FRAAL5811/ATP-dependent DNA helicase.

Table 1. Summary of the most up-regulated polyethylene glycol (PEG)-induced proteins of *F. alni*.

| Response mechanism | Protein | Product | Tfold |
|--------------------|---------|---------|-------|
| Membrane transport proteins | FRAAL2567 | Putative autotransporter adhesion | 5.67 |
| | FRAAL2549 | Secondary metabolite secretion ATP-binding transport | 4.80 |
| | FRAAL1420 | Nitrate/sulfonate ATP-binding (ABC) transporter | 2.89 |
| | FRAAL2085 | Trk system potassium uptake protein TrkA | 1.65 |
| Intracellular osmolytes concentration | FRAAL0995 | Osmotic-stress mechanosensitive channel MscL | 4.30 |
| | FRAAL1888 | Glutamine amidotransferase | 2.20 |
| | FRAAL1197 | 2-deoxyribose-5-phosphate aldolase, NAD(P)-linked | 2.15 |
| | FRAAL5155 | Peptidase S51, dipeptidase E | 1.91 |
| | FRAAL2257 | Non-ribosomal peptide synthase | 2.20 |
| | FRAAL2542 | Non-ribosomal peptide synthase | 1.80 |
| | FRAAL2545 | Non-ribosomal peptide synthase | 1.60 |
| | FRAAL2500 | Polyketide decyclase | 1.79 |
| Molecular chaperones | FRAAL431 | Clp-family ATP-binding protease | 4.00 |
| | FRAAL4781 | Acyl-coenzyme A thioesterase | 3.00 |
| | FRAAL6258 | Acyl carrier protein (ACP) | 2.33 |
| | FRAAL1744 | Phospholipase D | 1.82 |
| | FRAAL3585 | Esterase | 1.75 |
| | FRAAL5087 | 1-acylglycerol-3-phosphate O-acyltransferase | 1.50 |
| | FRAAL2171 | Acyl-CoA N-acyltransferase | 1.50 |
| Membrane lipid biosynthesis | FRAAL0095 | Osmotic-stress mechanosensitive channel MscL | 4.30 |
| | FRAAL1888 | Glutamine amidotransferase | 2.20 |
| | FRAAL1197 | 2-deoxyribose-5-phosphate aldolase, NAD(P)-linked | 2.15 |
| | FRAAL5155 | Peptidase S51, dipeptidase E | 1.91 |
| | FRAAL2257 | Non-ribosomal peptide synthase | 2.20 |
| | FRAAL2542 | Non-ribosomal peptide synthase | 1.80 |
| | FRAAL2545 | Non-ribosomal peptide synthase | 1.60 |
| | FRAAL2500 | Polyketide decyclase | 1.79 |
| CO₂ sensing and metabolism | FRAAL1222 | Carbonic anhydrase | 2.80 |
| | FRAAL6022 | Gamma-aminobutyrilaldehyde dehydrogenase | 2.17 |
| | FRAAL1417 | Methylmalonate-semialdehyde dehydrogenase (acylating) | 1.50 |
| | FRAAL5703 | Global regulator (repressor) for SOS regulon LexA | 1.56 |

Table 2. Summary of the most down-regulated polyethylene glycol (PEG)-modulated proteins of *F. alni*.

| Response mechanism | Protein | Product | Tfold |
|--------------------|---------|---------|-------|
| Nitrogen fixation | FRAAL6814 | Nitrogenase-associated homocitrate synthase NifV | −6.50 |
| | FRAAL6803 | FeMo cofactor biosynthesis protein NifB | −5.24 |
| | FRAAL6810 | Nitrogenase iron-molybdenum cofactor biosynthesis NifE | −3.80 |
| | FRAAL6809 | Nitrogenase iron-molybdenum cofactor biosynthesis NifN | −3.50 |
| | FRAAL2491 | squalene-hopencyclase Shc2 | −2.17 |
| | FRAAL6797 | Ferredoxin | −2.00 |
| | FRAAL6799 | 2-oxoglutarate ferredoxinoxidoreductase beta-subunit | −1.74 |
| | FRAAL1432 | Squalene-hopencyclase Shc1 | −1.62 |
| | FRAAL4563 | Transport protein associated with Fe-S cluster assembly | −1.60 |
| Respiration and energy production and conversion | FRAAL4147 | Cytochrome c oxidase polypeptide I (AA3 subunit 1) CtaD | −3.70 |
| | FRAAL5625 | Poly(3-hydroxybutyrate) depolymerase Pha | −2.83 |
| | FRAAL6005 | Biotin carboxyl carrier protein | −2.00 |
| | FRAAL6082 | Succinate-semialdehyde dehydrogenase I, NADP-dependent | −2.00 |
| | FRAAL5481 | Malate synthase G AceB | −1.75 |
| | FRAAL0315 | Aldehyde dehydrogenase | −1.71 |
| | FRAAL1829 | [NiFe] uptake hydrogenase, large subunit | −1.70 |
| | FRAAL1039 | NADH-quinone oxidoreductase chain H NuoH | −1.50 |
| DNA repair system | FRAAL6528 | ATP-dependent protease, DNA repair protein | −2.36 |
| | FRAAL5811 | ATP-dependent DNA helicase RecG | −1.50 |
Scientific REPORTS | (2018) 8:759 | DOI:10.1038/s41598-017-18839-0

Figure 2. Chord diagram showing the most enriched biological processes (GO terms) with their differentially expressed blue class proteins following PEG treatment. In each chord, enriched GO biological processes are shown on the right, and *F. alni* blue class differentially expressed proteins contributing to this enrichment are shown on the left. On the left side of the circle, each differentially expressed protein (locus tag (FRAAL)) is represented by a rectangle which colour is correlated to the value of the Tfold (logFC). PEG up-regulated proteins are displayed in red whereas down-regulated proteins are displayed in blue. Chords connect protein names with biological process GO term groups. Each GO term is represented by one coloured line.

**F. alni** protein-protein interactions. Supplementary Figures 3, 4, 5 and 6 represent *F. alni* protein-protein interactions within the blue, orange, green and red class proteins, respectively. Green nodes colour in Supplementary Figure 3 denotes up-regulated proteins whereas red nodes colour denotes down-regulated proteins. For instance, the network corresponding to the blue class proteins (Supplementary Figure 3) highlights an interaction between an up-regulated putative Clp-family ATP-binding protease (FRAAL4431) and a down-regulated putative signal peptide (FRAAL4430).

Cluster of Orthologous Genes (COGs). Figure 3 and Supplementary Figures 7, 8 and 9 show the distribution of the blue, orange, green and red classes proteins into the COG functional categories, respectively. Green colour within the Fig. 3 denotes up-regulated proteins whereas red colour denotes down-regulated proteins. We were not able to assign several hypothetical proteins to any COG category. For the blue class (Fig. 3), five COG categories/classes contain only up-regulated proteins—D (cell cycle control, cell division, chromosome partitioning), F (nucleotide transport and metabolism), S (function unknown), U (intracellular trafficking, secretion and vesicular transport) and W (extracellular structures). These results provide parallel pieces of evidence (see also Table 1) in support of the importance of protein clusters related to cellular processes and signalling and metabolism as well as some poorly characterised proteins during growth of *F. alni* under water stress.
Comparison between differentially expressed proteins following polyethylene glycol (PEG)-induced water stress and symbiosis. We compared up- and down-regulated proteins related to tolerance to PEG-induced water stress and symbiosis in *F. alni*. Common up- (128) and down-regulated (37) proteins were identified and highlighted in Fig. 4, yielding evidence of common pathways between xerotolerance and symbiosis in *F. alni*.

FRAAL0095-polyethylene glycol (PEG) interaction investigation. Among the most up-regulated proteins, we chose the FRAAL0095 (Q0RUG3) as a case study. The amino acid sequence of this putative mechanosensitive channel of large conductance (MscL) was submitted to the Swiss Model server \(^\text{34}\). Templates search identified the structure of the MscL of *Mycobacterium tuberculosis* (PDBid: 2OAR) as the template with the highest sequence identity \((47.6\%)\) with the target. It is crystallised as a homo-pentamer with three gold ions. This template was selected for model building with respect to its oligo-state. The resulting structure model shown in Supplementary Figure 10 exhibited satisfactory quality scores with a Global Model Quality Estimation (GMQE) score of 0.72. This metric, comprised between zero and one, reflects the expected accuracy of the model with respect to the target-template alignment \(^\text{34}\). The higher the GMQE the more accurate is the model. Another quality assessment was performed through the Ramachandran diagrams (see Supplementary Figure 11) of the homo-pentamer model of FRAAL0095. It presented 11 residues \((1.7\%)\) within the outliers region, distributed on all five chains. They mainly occur at the top of the channel (extra-cellular part), or in the trans-membrane region. The global quality of the model was thus considered as satisfactory.

Docking of the PEGbasic and related probes (Fig. 5) was performed on this structure model. The resulting docking poses were merged and used by the SOM-BSfinder tuned version to define a 3D map of the preferential interaction spots of the PEG and its related probes with the protein surface. The resulting map and hot spots are shown in Fig. 6(A–F). Three populated regions, called Consensual Clusters (CCs), could be identified at equivalent positions between equivalent helices from pairs of adjacent chains of the homo-pentamer (Fig. 6(C)). A fourth CC, ranked as least important and coloured in red on Fig. 6(C), was detected in a higher position than the other CCs.

Moderate and long chain PEG molecules are polymers with the repeating unit -\(\text{CH}_2\text{CH}_2\text{O}\)-. Their length determines their molecular weights which are computed as follows: \(\text{MW} = 18.02 + 44.05\times n\), where \(n\) is the number of the repeating units. For instance, PEG400 is composed of 8–9 repeating units and PEG8000 is composed of 181–182 repeating units. PEG400 was herein used as a basic model for moderate to long chain PEGs to conduct molecular docking simulations. In fact, the moderate length of PEG400 can be handled by docking programs for accurate calculations while larger sizes \((\text{e.g.}, \text{PEG8000})\) would be time consuming with poor quality \((\text{data not shown})\). We hypothesised that data collected with PEG400 along with the different probes will provide insightful results on the PEG-protein interactions for our case study. PEG400 (see Fig. 5) was thus docked on the structure model of the FRAAL0095. The SOM-BSfinder analysis returned a 3D map with a different shape as compared to the one obtained with the PEGbasic and related probes. It presented higher global density, which can be explained by the length and the flexibility of the polyethylene chain of PEG400. Still, a CC was detected between each two equivalent helices from adjacent chains (Fig. 6(F)), which is consistent with the results obtained with the PEG-basic and the probes. Moreover, an additional CC \((\text{coloured in green on Fig. 6(F)})\) was identified in a higher position than the other CCs. The docking results suggested a folding of the PEG400 inside the channel formed by the five chains. This is consistent with the literature\(^\text{35,36}\). In fact, it was proposed that moderate \((\text{PEG8000})\) to long chain \((\text{PEG4700} \text{and higher})\) PEGs would fold themselves and form a corona that mainly establishes hydrophobic...
interactions with proteins. We further identified the residues that may interact with the PEG400 as those located at a maximal distance of 4.0 Å from any of its atoms; and obtained residues alanine 25 (A25), alanine 28 (A28), valine 29 (V29), threonine 31 (T31), valine 34 (V34), lysine 35 (K35), arginine 65 (R65), glutamate 67 (E67), asparagine 70 (N70), serine 71 (S71), alanine 74 (A74) and phenylalanine 75 (F75). These interacting residues were mostly hydrophobic (A, V, F) or polar with neutral side chains (T, S, N). On the protein surface, they constituted a series of symmetric hydrophobic pockets (see Supplementary Figure 10) that define what have previously been described as the hydrophobic core of MscL. Hydrophobic and polar residues appeared to be central to the PEG-FRAAL0095 interactions.

Discussion

Little is known about the means by which Frankia cells tolerate desiccation in the soil. In view of the fact that increasing concentrations of greenhouse gases in the atmosphere and climate change are expected to modify the global water cycle and to increase the odds of worsening drought in the next decades, an understanding of the responses of F. alni to desiccation and thus to osmotic stress becomes an important issue in the attempt to adapt symbiosis to desiccated niches. In the present work we have analysed the expression of proteins in the alder symbiont F. alni in response to PEG-induced water stress to characterise the impact of desiccation on nitrogen-fixing symbiotic bacteria. We used liquid medium since it easier to have reproducible conditions with healthy cells devoid of agar.

The results of the experiments summarised in Tables 1 and 2 indicated six most up-regulated and three most down-regulated PEG-modulated response mechanisms, respectively. The most remarkable differences in the analysed proteomes of F. alni cells grown with and without PEG could be assigned to envelope-associated proteins (e.g., FRAAL2567 (5.67) and FRAAL2549 (4.8); cf. Table 1). FRAAL2567 is a putative autotransporter adhesin, which is concordant with previous studies indicating the importance of proteins associated with the modulation of the structure and function of the three-dimensional extracellular matrix under stress conditions—desiccation, UV irradiation and oxidation. In addition, cell envelope remodelling, transcriptional and translational regulation and salt-stress responsive hypothetical proteins were shown to be important for salt-tolerant Frankia strains. A previous study that concords with present results (Table 1), showed that Rhodococcus jostii RHA1 desiccation-specific transcriptome had several genes associated with cellular processes including lipid metabolism and cell envelope modification. ABC transporters, like FRAAL2549, are integral membrane proteins that typically use cellular energy to translocate solutes across cellular membranes in all phyla. Through the transport of molecules, ABC transporters are involved in diverse cellular processes including osmotic homeostasis and resistance to xenotoxins. Moreover, up-regulation of a potassium transporter (FRAAL2085 (1.65)) was observed. Orthologs of this protein were reported to scavenge K+ ions from the environment and maintain cell hydrostatic pressure in response to desiccation. In addition, MscL, one of the best studied highly conserved mechanosensitive channels throughout the bacterial kingdom, was found to be highly up-regulated in F. alni.
following exposure to PEG (FRAAL0095 (4.30)). This transmembrane channel acts through switching from closed to open conformation under severe osmotic challenges by opening a water-filled pore44,45. Molecular modelling revealed potential interacting sites of the PEG molecule inside the channel formed by five monomers of MscL of *F. alni*. These sites were located between each pair of adjacent chains in a symmetric frame. Interactions with the PEG mainly included hydrophobic and polar residues that define the hydrophobic core that is sine qua non to the channel activity46. Recently, the co-crystal structure of the C-terminal SH3 domain of myosin IB from *Entamoeba histolytica* with PEG was reported47. Gautam and colleagues47 brought insightful elements for a better understanding of the interactions of PEG with protein surfaces and reported in particular the corona-like binding mode previously described48–51. Our findings are consistent with the literature, which adds confidence to our computational approach. In fact, long-chain PEGs are flexible polymers with multiple rotatable bonds and high degrees of freedom, which makes them a challenging case for molecular docking software. In the present work, we established a protocol to circumvent this issue by using the repeating unit of the PEG and its chemical features.

---

**Figure 5.** Chemical structures of the PEG400, PEGbasic and derived probes. Probes (0–4) are chemical substructures of PEG.

---

**Figure 6.** Surface mapping of the FRAAL0095 and identified hot-spots. Panels in the first line present results obtained with the docking of PEG basic and derived probes. Panels in the second line present results obtained with the docking of PEG 400. (A) A face-view of the protein structure along with the u-Matrix obtained with SOM-BSfinder analysis of the PEGbasic docking results, with a zoom on the u-Matrix. (B) A top-view of the protein structure along the u-Matrix of the PEGbasic docking results. (C) A top-view of the protein structure along with the CCs (hot-spots) identified using the SOM-BSfinder analysis of the PEGbasic docking results. (D) A face-view of the protein structure along with the u-Matrix obtained with SOM-BSfinder analysis of the PEG400 docking results, with a zoom on the u-Matrix. (E) A top-view of the protein structure along the u-Matrix of the PEG400 docking results. (F) A top-view of the protein structure along with the CCs (hot-spots) identified using the SOM-BSfinder analysis of the PEG400 docking results.
as probes in order to infer a simplistic model for docking. Then, the SOM-based method permitted to generate an accurate map of the binding hot spots on the MscL surface, along with a prediction of the overall binding mode. Our results suggest that long-chain PEGs may act as an inner coat of the channel66,67. This mode of binding may either induce an increase in water molecules conductance through the channel or induce a mechanistic blocking of its entrance. Our findings constitute a response element to the role of MscL (FRAAL0095) in the F. alni desiccome, and a starting point to conduct further investigations on the PEG-FRAAL0095 interactions.

As expected, envelope stress would induce a cascade of other defence pathways. Most notable is the induction of one CRISPR-associated proteins (FRAAL0457 (1.67) being part of the Green class) supporting published observations that expression of CRISPR-Cas components in several bacterial species can be induced following envelope stress68. Supplementary Table S3 lists the CRISPR-associated PEG-induced proteins of F. alni. Also, the present analysis revealed the up-regulation of proteins associated with cell osmoprotection. For instance, FRAAL1888 (2.20) and FRAAL2171 (1.5) are involved in the synthesis of the dipeptide N-acetylglutaminylglutamine amide (NAGGN) that plays a key role in osmoprotection34. In addition to its role in NAGGN biosynthesis, FRAAL1888 is involved in the conversion of glutamine into glutamate, one of the most abundant and up-regulated amino acids in Alnus and Casuarina nodules, for which a role as osmolyte was postulated by Brooks and Benson35. Also, the FRAAL5155 peptidase (1.91), by the cleavage of dipeptides into single peptides, would provide needed intracellular osmolytes to offset the high extracellular osmotic pressure.

In addition, consistent with previous studies, we observed a significant up-regulation of FRAAL4431 (4) encoding a Clp protease. Indeed, ClpP proteases play an essential role in removing damaged proteins from low GC Gram-positive bacteria under stress conditions such as osmotic shock69. Furthermore, proteolysis of such proteins would release amino acids that would in turn provide osmolytes to help compensate the osmotic shock. Similarly, a Clp-protease was induced during desiccation and dehydration in F. alni (2.8) in PEG-grown F. alni cells. It is expected that this enzyme, by hydrating CO₂ to carbonate (H₂CO₃), would increase the concentration of protons within the cell and decrease the pH. A previous study70 showed an increase in the expression of an orphan carbonic anhydrase in the cyanobacterium Microcoleus vaginatus using a whole genome transcriptional time course assay in response to hydration. The observed increase in several ROS-scavenging proteins and translation-associated ribosome proteins provided evidence in support of a higher demand of detoxification and protein synthesis during growth under PEG-induced water stress conditions. For example, FRAAL6022 (2.17) encodes a gamma-aminobutyraldehyde dehydrogenase (GABA synthesis). GABA can act as a hydroxyl radical scavenger28.

Deinococcus radiodurans’ xerotolerance was explained by the phenomenon of gene sharing79: during land colonization by Terrabacteria, gene products (e.g., DNA repair genes, etc.) contributing to this bacterium’s ionizing radiation were recruited to serve an additional function that is desiccation tolerance60. In F. alni, we observed an upregulation of the SOS response genes transcriptional repressor LexA (1.56) with an expected downregulation of some DNA repair proteins (Table 2). Furthermore, the investigation of the relationship between desiccation tolerance and nitrogen metabolism in Anaabaena sp. PCC 7120 suggested that terrestrial cyanobacteria may acclimate to desiccation stress via nitrogen (N₂) fixation by using desiccation inducible genes81. In our study, TFold analysis pinpointed three classes of differentially down-regulated proteins (e.g., FRAAL6814 (−6.5), FRAAL4147 (−3.7) and FRAAL6528 (−2.36); cf. Table 2). The remarkable decrease in protein expression related to nitrogen fixation (more than sixfold), aerobic respiration (more than threefold) and homologous recombination (more than twofold) quantitatively reflects the magnitude of the impact of nitrogen-, oxygen- and homologous-recombination-associated energy demands in PEG-grown F. alni cells.

Identified common pathways between symbiosis and desiccation tolerance (Fig. 4) are in accordance with the fact that plant-growth-promoting (PGP) rhizobacteria augment plant tolerance to drought82. In line with this, it was recently demonstrated that bacterial endophytic communities promote date palm (Phoenix dactylifera L.) growth under drought conditions83. Also, it was shown that plant growth promotion ability exerted by bacteria is a drought-induced effect84. In addition, a previous work85 showed an induction of the threonine dehydratase (DR_0567)—converting threonine into NH₃/NH₄⁺—and the nodulation efficiency protein D (NfeD, DR_2142 (WP_010888773.1)) following desiccation of D. radiodurans R₁.

Penetration into host tissues implies many stresses, one of which is osmotic since Alnus glutinosa, as most alder species, thrives on river banks where water has low salt levels, while root tissues have a higher, isotonic osmotic potential. Furthermore, alder synthesizes many peptides upon entry of Frankia, many of which are defensive blends that bind to the cell membrane and modify its porosity86 and thus the ability to cope with osmolytes. Rhizobia have been shown to have many determinants associated with desiccation tolerance, many of which have similarities to those seen upregulated in Frankia such as transporters87, the mechanosensitive channel88 or disaccharide accumulation89. However the link between osmotic response and symbiosis establishment was not very strong, implying the responses seen here would have more usefulness for saprotrophic soil existence.

Conclusions

The present paper is the first high-throughput proteomic study of F. alni subjected to water-related stress. This research, analysing the most representative proteome of a Frankia strain, contributes to a better comprehension of environmental stress adaptation particularly in desiccated soils. In the future, our data might be used in further comparative proteogenomic studies of nitrogen-fixing plant symbionts.

Materials and Methods

Bacterial growth and treatment with polyethylene glycol (PEG). F. alni strain ACN14a90 was grown in liquid BAP medium without ammonium as described earlier90. The cells were then syringed thoroughly with needles of decreasing width (21G–27G), the cells OD₆₀₀ measured and the suspension diluted prior to inoculation.
An OD$_{600}$ inoculation of 0.1 was used to make a growth dynamic and to monitor ARA activity at 0.3, 0.9 and 2.8% PEG8000 (Sigma). Inoculations were then made at the same OD$_{600}$ in fresh BAP medium without or with PEG8000 (Sigma) added to a final concentration of 0.9% w/v. The cells were grown in 125 mL Erlenmeyers with 40 mL medium at 28 °C on an orbital shaker at 200 rpm for 7 days. The cells were harvested, their OD$_{600}$ measured, an aliquot inoculated onto LB agar plates to detect contaminations and the cells observed under the microscope. The cells were then sedimented (1,500 g for 5 minutes) and the pellets frozen until proteomics analysis.

**Proteomic sample preparation and nanoscale liquid chromatography coupled to tandem mass spectrometry (nano LC-MS/MS) analysis of tryptic peptides.** The protein content of cells grown without PEG and in the presence of PEG was established by a shotgun procedure. Following trypsin proteolysis, peptides were analysed by nanoLC-MS/MS with a Q-Exactive HF tandem mass spectrometer incorporating an ultra-high field Orbitrap analyser. Five independent biological replicates were analysed for both conditions: PEG-treated bacteria and untreated bacteria. A volume of 5 μL of LDS (1X) was added per mg of bacterial pellet prior a 5 min heat treatment at 99 °C followed by a treatment for 5 min in an ultrasonic bath. Cells and debris were transferred into a 2 mL Precellys (Bertin Technologies, F-78180 Montigny le Bretonneux) tube containing 200 μg of glass beads and the samples were subjected to 3 cycles of grinding at 6,500 rpm for 20 sec by means of a Precellys grinder (Bertin technologies). Samples were then centrifuged for 40 sec at 16,000 g. The resulting supernatants were incubated for 10 min at 99 °C and were subjected to a short SDS-PAGE migration as previously described.$^{31}$ The polyacrylamide bands containing the whole solubilised protein content of each sample were processed as previously described.$^{32}$ Briefly, they were subjected to DTT reduction, treated with iodoacetamide and then, proteolysed with Sequencing Grade Trypsin (Roche, F-38240 Meylan) using 0.01% of proteaseMAX detergent (Promega, F-69260 Charbonnières-les-Bains). The resulting peptides (10 μL) were analysed in data-dependent mode with an ESI-Q Exactive HF mass spectrometer (ThermoFisher Scientific, F-91963, Courtaboeuf) equipped with an ultra-high field Orbitrap analyser and coupled to an Ultimate 3000 176 RSL Nano LC System (ThermoFisher). Peptides were injected onto a reverse phase Acclaim PepMap 100 C18 column (3 μm, 100 Å, 75 μm i.d. × 500 mm) and resolved at a flow rate of 0.2 μL/min with a 60 min gradient of CH3CN in presence of 0.1% formic acid. A dataset of 10 nanoLC-MS/MS runs were recorded. The Q-Exactive HF instrument was operated with Top20 standard parameters and a dynamic exclusion of 10 sec as previously described$^{33}$. MS/MS spectra were searched using MASCOT 2.2.04 software (Matrix Science, London, W1U 7GB, UK) against the *F. alni* database (5,804 protein sequences) with the following parameters: full-trypsin specificity, maximum of two missed cleavages, mass tolerances of 5 ppm on the parent ion and 0.02 Da on the higher energy collisional dissociation-induced peptideic fragments, fixed modification of carboxymidomethylated cysteine (+57.0215), and oxidised methionine (+15.9949) as dynamic modifications. Peptide-to-MS/MS spectrum matching with a MASCOT score below a p-value of 0.05 were selected and assigned to unique peptide sequence following the parsimony principle. A protein was considered valid when at least two different peptides were detected. The false-positive rate for protein identification was estimated by a search with a reverse decoy database to be below 0.1% using the same parameters. Proteins were quantified based on their spectral counts. The normalised spectral abundance factor (NSAF) was calculated by dividing the spectral count for each observed protein by the polypeptide theoretical mass, as described previously$^{30}$ and is presented as percentage of the NSAF sum considering all proteins. Proteome comparison between both conditions was done taking into account the five biological replicates with the TFold module from the PatternLab software and standard normalisation, as previously described$^{31}$.

**Homology-based functional analysis and Clusters of Orthologous Groups of proteins (COGs) prediction.** Protein sequences of the PEG-responding differentially expressed proteins were retrieved from the National Center for Biotechnology Information (NCBI) database$^{34}$ and uploaded into the STRING database$^{35}$ analysis tool for Gene Ontologies and metabolic pathways functional enrichments. Biological processes and metabolic pathways with False Discovery Rate (FDR) ≤ 0.005 were considered as significant. *F. alni* Clusters of Orthologous Groups of proteins (COGs) were inferred for an important set of proteins based on BLASTP$^{36}$ best reciprocal hits with *Frankia* sp. EAN1pec for which these clusters are known and available in the database of COGs$^{37}$.

**Protein-protein interaction network.** The sequences of the 294 proteins of *F. alni* belonging to the PatternLab blue class (for which identifications satisfied both the fold ($>$1.5) and statistical criteria (p-value $<$ 0.05)) were downloaded and saved in a multi-FASTA file. This file was then uploaded in the STRING database$^{35}$ and *F. alni* was chosen as a query microorganism. A total of 294 hits with 100% identity were detected and their protein-protein interactions were predicted. In order to generate protein-protein interactions integrating the expression data information of each protein, data were integrated into the Cytoscape tool.$^{27}$ Expression values were imported as node attributes. The same steps were followed to analyse protein-protein interaction networks of the other detected proteins (belonging to the PatternLab orange, green and red classes (see Results section for details)). All single nodes were not represented with Cytoscape which considers only interacting proteins.

**Re-annotation of differentially expressed hypothetical proteins.** Differentially expressed proteins that were annotated as hypothetical (or conserved hypothetical) were re-annotated. For this, we conducted new analyses using the Position-Specific Iterated BLAST (PSI-BLAST) program$^{38}$ at http://blast.ncbi.nlm.nih.gov/ with default parameters against a database of non-redundant protein sequences (nr).

**Molecular modelling and docking.** *Comparative modelling*. Protein sequence of the FRAAL0095, one of the most up-regulated *Frankia* genes (4.30), identified as coding for an osmotic-stress large conductance mechanosensitive channel, was submitted to the Swiss Model Server$^{34}$. Search for templates of experimentally
determined protein structures within the Swiss-Model Template Library (SMTL) was performed. Identified templates, ranked according to their sequence identity rate (IR) and their coverage rate (CR) with the target sequence, were examined. The template(s) with the highest IR and CR were selected for model building. The model with the most satisfactory quality assessment scores was considered for the subsequent analysis. Ramachandran diagrams of the selected model were generated using the RAMPAGE server [46], and used to assess the 3D model quality.

**Molecular docking and surface mapping.** Molecular docking of the PEG and related molecules (herein called ligands) on the structure model of the protein target (herein called receptor) was performed using AutoDock Vina 1.1.2 [3]. The latter program requires input files of the ligands and the receptor in PDBQT (Protein Data Bank) [3], Partial Charge (‘Q’), and Atom Type (‘T’) format. Default parameters were used and a maximum of 20 lowest-energy poses were kept for each ligand. Two forms of the PEG were considered: (i) a short form having the formula \( \text{C}_2\text{H}_10\text{O}_3 \) and is referred to as PEGbasic and (ii) a longer form having the formula \( \text{C}_{18}\text{H}_{38}\text{O}_{10} \) to which we refer as PEG400. Additionally, five chemical substructures of the PEG were generated and used as probes to map hot spots on the protein surface. Simplified Molecular Input Line Entry System (SMILES) conformations of PEGbasic, PEG400 and the probes were generated manually, then converted to PDBQT format using the OpenBabel package (http://openbabel.org) [43]. The PDB file of the target was used to generate the corresponding PDBQT format by adding hydrogen atoms and atomic partial charges using the OpenBabel package [3].

**Proteomics data repository.** The mass spectrometry proteomic data were deposited at the ProteomeXchange Consortium (http://proteomecentral.proteomeexchange.org) via the PRIDE partner repository with the data set identifiers PXD007226 and DOI 10.6019/PXD007226.

**References**

1. Normand, P., Benson, D. R., Berry, A. M. & Tisa, L. S. In _The Prokaryotes: Actinobacteria_ (eds Eugene Rosenberg et al.) 339–356 (Springer Berlin Heidelberg, 2014).

2. Benson, D. R. & Silvester, W. B. Biology of _Frankia_ strains, actinomycete symbionts of actinorrhizal plants. _Microbiol Rev_ 57, 293–319 (1993).

3. Sen, A. _et al._ Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders ‘Frankiales’ and _Micrococcales_ should be split into coherent entities: proposal of Frankiales ord. nov., _Geodermatophilales_ ord. nov., _Acidothermales_ ord. nov. and _Nakamurellales_ ord. nov. _Int J Syst Evol Microbiol_ 64, 3821–3822, https://doi.org/10.1099/ijs.0.063966-0 (2014).

4. Oshone, R. _et al._ Proposal of a type strain for _Frankia abii_ (Woronin 1866) V. Tuleuba 1985, emended description of _Frankia abii_, and recognition of _Frankia cassuarinae_ sp. nov. and _Frankia elaeagni_ sp. nov. _Int J Syst Evol Microbiol_ 66, 5201–5210, https://doi.org/10.1099/ijsem.0.014916 (2016).

5. Normand, P. _et al._ Genome characteristics of facultatively symbiotic _Frankia_ sp. strains reflect host range and host plant biogeography. _Genome Res_ 17, 7–15, https://doi.org/10.1101/gr.7598407 (2007).

6. Alloisio, N., Kucho, K.-i., Pujic, P. & Normand, P. In _Biological Nitrogen Fixation_ 757–768 (John Wiley & Sons, Inc, 2015).

7. Oshone, R. _et al._ Intraspecies variation in sodium partitioning, potassium and proline accumulation under salt stress in _Cassuarina equisetifolia_ Forst. _Symbiosis_ 70, 117–127, https://doi.org/10.1007/s13199-016-0424-9 (2016).

8. Srivastava, A., Singh, A., Singh, S. S. & Mishra, A. K. Salt stress-induced changes in antioxidative defense system and proteome profiles of salt-tolerant and sensitive _Frankia_ strains. _J Environ Sci Health. Part A, Toxic/hazardous substances & environmental engineering_ 52, 420–428, https://doi.org/10.1080/10994529.2016.1270672 (2017).

9. Lebre, P. _et al._ Xerotolerant bacteria: surviving through a dry spell. _Nat Rev Micro_ 15, 285–296, https://doi.org/10.1038/nrmicro.2017.16 (2017).

10. Kakumanu, M. L., Cantrell, C. L. _et al._ Microbial community response to varying magnitudes of desiccation in soil: a test of the osmolyte accumulation hypothesis. _Soil Biol Biochem_ 57, 644–653 (2013).

11. Burkeleigh, S. H. & Dawson, J. O. Desiccation tolerance and trehalose production in _Frankia_ hyphae. _Soil Biol Biochem_ 26, 593–598 (1994).

12. Selvakumar, R. R. _et al._ Intraspecies variation in sodium partitioning, potassium and proline accumulation under salt stress in _Cassuarina equisetifolia_ Forst. _Symbiosis_ 70, 117–127, https://doi.org/10.1007/s13199-016-0424-9 (2016).

13. Srivastava, A., Singh, A., Singh, S. S. & Mishra, A. K. Salt stress-induced changes in antioxidative defense system and proteome profiles of salt-tolerant and sensitive _Frankia_ strains. _J Environ Sci Health. Part A, Toxic/hazardous substances & environmental engineering_ 52, 420–428, https://doi.org/10.1080/10994529.2016.1270672 (2017).

14. Oshone, R. _et al._ Genomic, transcriptomic, and proteomic approaches towards understanding the molecular mechanisms of salt tolerance in _Frankia_ strains isolated from _Cassuarina_ trees, _RMC Genome_ 18, 633, https://doi.org/10.1186/s12864-017-4056-0 (2017).

15. Dawson, J. O. _et al._ A. H. Sensitivity of selected _Frankia_ isolates from _Cassuarina, Allocasuarina_ and North American host plants to sodium chloride. _Physiol Plant_ 70, 272–278, https://doi.org/10.1111/j.1399-3054.1987.tb06144.x (1987).

16. Oshone, R., Mansour, S. R. _et al._ S. Effect of salt stress on the physiology of _Frankia_ sp strain CcI6. _Cellulose_ 38, 699–702 (2013).

17. Sghaier, H. _et al._ Stone-dwelling actinobacteria _Blastococcus saxobsidens, Modestobacter marinus_ and _Geodermatophilus obscurus_ proteogenomes. _The ISME J_ 10, 21–29, https://doi.org/10.1038/ismej.2015.108 (2016).

18. Umezawa, T., Mizuno, K. _et al._ Discrimination of genes expressed in response to the ionic or osmotic effect of salt stress in soybean with cDNA-AFLP. _Plant Cell Environ_ 1617–1625 (2002).

19. Raeyev, R. _et al._ Dynamic cyanobacterial response to hydration and dehydration in a desert biological soil crust. _ISME J_ 7, 2178–2191, https://doi.org/10.1038/ismej.2013.83 (2013).

20. Mohammadpanah, F. _et al._ _Frankia_ from arid and desert habitats: Diversity and biological activity. _Front Microbiol_ 6, 1541, https://doi.org/10.3389/fmicb.2015.01541 (2015).

21. Fornellos, J. C., Gonçalves, E. R. _et al._ Global response to desiccation stress in the soil actinomycete _Rhodococcus jostii_ RHA1. _Appl Environ Microbiol_ 74, 2627–2636, https://doi.org/10.1128/aem.02711-07 (2008).

22. Mclntyre, H. _et al._ Trehalose biosynthesis in _Rhizobium leguminosarum bv. trifoli and its role in desiccation tolerance. _Appl Environ Microbiol_ 73, 3984–3992, https://doi.org/10.1128/AEM.00412-07 (2007).

23. Boncompagni, E., Oteras, M., Poggi, M. C. _et al._ R. Occurrence of choline and glycine betaine uptake and metabolism in the family _rhizobiales_ and their roles in osmoprotection. _Appl Environ Microbiol_ 65, 2072–2077 (1999).

24. De Philippis, R. _et al._ _Frankia_ strains reflect host range and host plant biogeography. _FEMS Microbiol Rev_ 22, 151–175, https://doi.org/10.1111/j.1574-6976.1998.tb00365.x (1998).
1. Carro, L. 
2. Tyynismaa, M. 
3. Dörflinger, M. 
4. Jones, P. M. & George, A. M. The ABC transporter structure and mechanism: perspectives on recent research. 
5. Gautam, G., Rehman, S., Pandey, P. & Gourinath, S. Crystal structure of the PEG-bound SH3 domain of myosin IB from 
6. Cherif, H. 
7. Yang, J., Kloepper, J. W. & Ryu, C. M. Rhizosphere bacteria help plants tolerate abiotic stress. 
8. Haswell, E., Phillips, R. & Rees, D. Mechanosensitive channels: what can they do and how do they do it? 
9. Sghaier, H., Mitomo, H. & Narumi, I. Genomic confession of 
10. Contreras-Porcia, L., Thomas, D., Flores, V. & Correa, J. A. Tolerance to oxidative stress induced by desiccation in 
11. Brooks, J. M. & Benson, D. R. Comparative metabolomics of root nodules infected with 
12. Knowles, D. B., Lim, K. & Herron, J. N. In 
13. Wu, J. 
14. Hašek, J. Poly(ethylene glycol) interactions with proteins. Zeitschrift für Kristall Suppl 23, 613–618 (2006). 
15. Wu, J. et al. Binding characteristics between polyethylene glycol (PEG) and proteins in aqueous solution. J Mat Chem B 2, 2983–2992, https://doi.org/10.1039/C4TB0025B (2014). 
16. Knowles, D. B. et al. Chemical interactions of polyethylene glycols (PEGs) and glycerol with protein functional groups: applications to effects of PEG and glycerol on protein processes. Biochemistry 54, 3528–3542, https://doi.org/10.1021/acs.biochem.5b00246 (2015). 
17. Lim, K. & Herron, J. N. In Poly(ethylene glycol) chemistry: biotechnical and biomedical applications (ed. Milton Harris, J.) 29–56 (Springer US, 1992). 
18. Perozo, E., Cortes, D., Sompornsiput, P., Khoda, A. & Martinac, B. Open channel structure of Mscl and the gating mechanism of mechanosensitive channels. Nature 418, 942–948 (2002). 
19. Prudhomme, C. et al. Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment. Proc Nat Acad Sci USA 111, 3262–3267, https://doi.org/10.1073/pnas.1207313111 (2014). 
20. Jones, P. M. & George, A. M. The ABC transporter structure and mechanism: perspectives on recent research. Cell Mol Life Sci 61, 682–699, https://doi.org/10.1007/s00018-003-3336-9 (2004). 
21. Martinoia, E. et al. Multifunctionality of plant ABC transporters–more than just detoxifiers. Planta 214, 345–355 (2002). 
22. Zhong, D. & Blount, P. Phosphatidylinositol is crucial for the mechanosensitivity of Mycobacterium tuberculosis Mscl. Biochemistry 52, 5415–5420, https://doi.org/10.1021/bi080790o (2013). 
23. Sukhave, S., Blount, P., Martinac, B. & Kung, C. Mechanosensitive channels of Escherichia coli: the Mscl gene, protein, and activities. Annu Rev Physiol. 59, 633–657 (1997). 
24. Hamill, O. & Martina, B. Molecular basis of mechanotransduction in living cells. Physiol Rev. 81, 685–740 (2001). 
25. Häse, C., Le Dain, A. & Martinac, B. Molecular dissection of the large mechanosensitive ion channel (Mscl) of E. coli: mutants with altered channel gating and pressure sensitivity. J Membr Biol. 157, 17–25 (1997). 
26. Gautam, G., Rehman, S., Pandey, P. & Gournath, S. Crystal structure of the PEG-bound SH3 domain of myosin IB from Entamoeba histolytica reveals its role of ligand recognition. Acta Crystallogr D Struct Biol. 73, 672–682, https://doi.org/10.1107/S2059798317006339 (2017). 
27. Lim, S. T., Jane, J. L., Rajagopalan, S. & Seib, P. A. Effect of Starch Granule Size on Physical Properties of Starch-Filled Polyethylene Film. Biotechnol Prog 8, 51–57 (1992). 
28. Häse, J. Poly (ethylene glycol) interactions with proteins. Z Kristallogr Suppl. 123, 613–618 (2006). 
29. Wu, J. et al. Binding characteristics between polyethylene glycol (PEG) and proteins in aqueous solution. J Mat Chem B 2, 2983–2992 (2014). 
30. Knowles, D. B. et al. Chemical Interactions of Polyethylene Glycols (PEGs) and Glycerol with Protein Functional Groups: Applications to Effects of PEG and Glycerol on Protein Processes. Biochem J 384, 3528–3542 (2015). 
31. Haswell, E., Phillips, R. & Rees, D. Mechanosensitive channels: what can they do and how do they do it? Structure 19, 1356–1369, https://doi.org/10.1016/j.str.2011.09.005. (2011). 
32. Sampson, T. R. & Weiss, D. S. CRISPR-Cas systems: new players in gene regulation and bacterial physiology. Front Cell Inf Microbiol 4, 37, https://doi.org/10.3389/fcimb.2014.00037 (2014). 
33. Sagot, B. et al. Osmotically induced synthesis of the dipeptide N-acetylaminoarginine amide is mediated by a new pathway conserved among bacteria. Proc Nat Acad Sci USA 107, 12652–12657, https://doi.org/10.1073/pnas.1003631107 (2010). 
34. Brooks, J. M. & Benson, D. R. Comparative metabolomics of root nodules infected with Frankia sp. strains and uninfected roots from Abies alba and Vicia faba. J Exp Bot 63, 5357–5366 (2012). 
35. Knowles, D. B. et al. Poly(Ethylene Glycol) chemistry: biotechnical and biomedical applications. Acta Crystallogr D Struct Biol. 73, 672–682, https://doi.org/10.1107/S2059798317006339 (2017). 
36. Knowles, D. B. et al. Chemical Interactions of Polyethylene Glycols (PEGs) and Glycerol with Protein Functional Groups: Applications to Effects of PEG and Glycerol on Protein Processes. Biochem J 384, 3528–3542 (2015). 
37. Haswell, E., Phillips, R. & Rees, D. Mechanosensitive channels: what can they do and how do they do it? Structure 19, 1356–1369, https://doi.org/10.1016/j.str.2011.09.005. (2011). 
38. Sampson, T. R. & Weiss, D. S. CRISPR-Cas systems: new players in gene regulation and bacterial physiology. Front Cell Inf Microbiol 4, 37, https://doi.org/10.3389/fcimb.2014.00037 (2014). 
39. Sagot, B. et al. Osmotically induced synthesis of the dipeptide N-acetylaminoarginine amide is mediated by a new pathway conserved among bacteria. Proc Nat Acad Sci USA 107, 12652–12657, https://doi.org/10.1073/pnas.1003631107 (2010). 
40. Brooks, J. M. & Benson, D. R. Comparative metabolomics of root nodules infected with Frankia sp. strains and uninfected roots from Abies alba and Vicia faba. J Exp Bot 63, 5357–5366 (2012).
66. Vanderlinde, E. M. et al. Identification of a novel ABC transporter required for desiccation tolerance, and biofilm formation in *Rhizobium leguminosarum* bv. *viciae* 3841. *FEMS Microbiol Ecol* 71, 327–340, https://doi.org/10.1111/j.1574-6949.2009.00824.x (2010).

67. Balleza, D., Gomez-Lagunas, F. & Quinto, C. Cloning and functional expression of an Mscl ortholog from *Rhizobium etli*: characterization of a mechanosensitive channel. *J Membrane Biol* 234, 13–27, https://doi.org/10.1007/s00232-010-9235-8 (2010).

68. Goiffi, K., Pica, N., Pichereau, V. & Blanco, C. Disaccharides as a new class of nonaccumulated osmoprotectants for *Sinorhizobium meliloti*. *Appl Environ Microbiol* 65, 1491–1500 (1999).

69. Normand, P. & Lalonde, M. Evaluation of *Frankia* strains isolated from provenances of two *Alnus* species. *Can J Microbiol* 28, 1133–1142, https://doi.org/10.1139/m82-168 (1982).

70. Carro, L. et al. Physiological effects of major up-regulated *Alnus glutinosa* peptides on *Frankia* sp. ACN14a. *Microbiology* 162, 1173–1184, https://doi.org/10.1099/mic.0.000291 (2016).

71. Hartmann, E. M., Allain, E., Guillard, J. C., Pible, O. & Armengaud, J. Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria. *Meth Molec Biol* 1197, 275–285, https://doi.org/10.1007/978-1-4939-1261-2_16 (2014).

72. Clair, G., Armengaud, J. & Dupont, C. Restricting fermentative potential by proteome remodeling: an adaptive strategy evidenced in *Bacillus cereus*. *Molec Cell Proteome: MCP* 11, M111 013102, https://doi.org/10.1074/mcp.M111.013102 (2012).

73. Klein, G. et al. RNA-binding proteins are a major target of silica nanoparticles in cell extracts. *Nanotoxicology* 10, 1555–1564, https://doi.org/10.1080/17435390.2016.1244299 (2016).

74. Database Resources of the National Center for Biotechnology Information. *Nucl Acids Res* 45, D12–D17, https://doi.org/10.1093/nar/gkw1071 (2017).

75. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* 215, 403–410, https://doi.org/10.1016/0022-2836(90)90320-E (1990).

76. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucl Acids Res* 28, 33–36 (2000).

77. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498–2504, https://doi.org/10.1101/gr.1239303 (2003).

78. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* 25, 3389–3402 (1997).

79. Lovell, S. C. et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* 50, 437–450, https://doi.org/10.1002/prot.10286 (2003).

80. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31, 455–461, https://doi.org/10.1002/jcc.21334 (2010).

81. Rose, P. W. et al. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucl Acids Res* 45, D271–D281, https://doi.org/10.1093/nar/gkw1000 (2017).

82. O’Boyle, N. M. et al. Open Babel: an open chemical toolbox. *J Cheminform* 3, 33, https://doi.org/10.1039/c1jm10363j (2011).

83. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8.2.1 (2015).

Acknowledgements

This work was supported by the Tunisian National Center for Nuclear Sciences and Technology (CNSTN), the Institut Pasteur de Tunis (Tunis) and the Ministry of Higher Education and Scientific Research in Tunisia.

Author Contributions

K.G. carried out bioinformatics analysis of data and drafted the manuscript. E.H.-S. performed the structural bioinformatics study and took part in writing the manuscript. C.B.H. helped to analyse the results and generated some of the figures. P.F. did the PEG tests. P.P. planned the PEG tests and analysed the results. S.G. collected some of the figures. P.N. planned the PEG tests and analysed the results. I.G. helped in the structural bioinformatics study and took part in writing the manuscript. K.G. carried out bioinformatics analysis of data and drafted the manuscript. E.H.-S. performed the structural bioinformatics study and took part in writing the manuscript. H.S. participated in the design and coordination of the study and drafted the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-18839-0.

Competing Interests: The authors declare that they have no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.