agronomy

MDPI

*Article*

# *Ensifer meliloti* L6-AK89, an Effective Inoculant of *Medicago lupulina* Varieties: Phenotypic and Deep-Genome Screening

**Marina L. Roumiantseva [1,*], Maria E. Vladimirova [1], Alla S. Saksaganskaia [1], Victoria S. Muntyan [1], Alexandra P. Kozlova [1], Alexey M. Afonin [2], Olga A. Baturina [3] and Boris V. Simarov [1]**

[1] Laboratory of Genetics and Selection of Microorganisms, Federal State Budget Scientific Institution, All-Russia Research Institute for Agricultural Microbiology (FSBSI ARRIAM), 196608 Saint Petersburg, Russia; mariiacherkasova@arriam.ru (M.E.V.); allasaksaganskaya@mail.ru (A.S.S.); vucovar@yandex.ru (V.S.M.); alexsandrak95@mail.ru (A.P.K.); genet@yandex.ru (B.V.S.)

[2] Laboratory of Genetics of Plant-Microbe Interactions, Federal State Budget Scientific Institution, All-Russia Research Institute for Agricultural Microbiology (FSBSI ARRIAM), 196608 Saint Petersburg, Russia; afoninalexeym@gmail.com

[3] Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences (ICBFM SB RAS), 630090 Novosibirsk, Russia; niboch@niboch.nsc.ru

* Correspondence: mroumiantseva@yandex.ru; Tel.: +7-(812)-476-28-02

**Abstract:** This paper presents a deep analysis of the accessory genome of an economically promising strain of *Ensifer* (*Sinorhizobium*) *meliloti*, L6-AK89, obtained as a result of next-generation high-throughput sequencing (MiSeq, MinIon). Strain L6-AK89 is a $Str^R$ mutant of the native strain CIAM1775, a symbiont of *Medicago lupulina* that adapted to a saline and arid habitat in NW Kazakhstan. CIAM1775 is an effective inoculant of *M. lupulina* cv. Mira (fodder type standard), cultivated on moderately acid soils in the NW agricultural region of Russia. Strain L6-AK89 makes it possible to obtain the expected high (>150%) increases in dry mass of the same plant variety in plant tests. The L6-AK89 genome has an increased proportion of sequences related to the accessory elements relative to reference strain Rm1021, 7.4% versus 4.8%. A set of 53 *nod/noe/nol/nif/fdx/fix* genes and 32 genes involved in stress tolerance together with 16S rRNA and *recA–atpD–glnII–gyrB–dnaJ* were evaluated. The high symbiotic efficiency of L6-AK89 with hop clover is most likely due to unique features of its genome, in combination with structural differences in its *nod* and stress-related genes, as well as unique clusters of quorum-sensing genes and osmoprotector synthesis.

**Keywords:** *Medicago lupulina*; *Ensifer* (*Sinorhizobium*) *meliloti*; symbiotic efficiency; accessor genome; cryptic plasmid; microarray; IS-elements; CRISPR/Cas; phage-related sequences; stress-related and *sym* genes



**Citation:** Roumiantseva, M.L.; Vladimirova, M.E.; Saksaganskaia, A.S.; Muntyan, V.S.; Kozlova, A.P.; Afonin, A.M.; Baturina, O.A.; Simarov, B.V. *Ensifer meliloti* L6-AK89, an Effective Inoculant of *Medicago lupulina* Varieties: Phenotypic and Deep-Genome Screening. *Agronomy* **2022**, *12*, 766. https://doi.org/10.3390/agronomy12040766

Academic Editor: Gavin Ash

Received: 3 November 2021
Accepted: 17 March 2022
Published: 22 March 2022



## 1. Introduction

Species of the genus *Medicago* are present as pioneer plants for neglected areas and soils undergoing desertification. *Medicago lupulina* L., or so-called black medic or hop clover, is a native species widespread in different geographical regions. Some varieties are able to grow on arid/semi-arid drought/salinated land, in mitigated barren areas, or along railway embankments or on lightly acidic desert soils, and are often used in phytoremediation [1,2]. Under suitable climatic conditions, this plant species can produce stems up to 2 m in high, and local varieties are known for their fast-developing stems after cutting, significant differences in green mass production, and adaptation to stress factors such as wide swings of temperature, drought, and salinity. *M. lupulina* combines diploid, tetraploid annual, and 2-to-3-year-old varieties with a well-developed root system and tender green mass, rich in proteins (up to 23%) and fats (up to 3%) [3–5], the accumulation of which occurs due to economically and environmentally friendly biological nitrogen fixation. Therefore,

obtaining a variety with a high biomass and seed yield that is highly reactive to inoculation with nodule bacteria is an important agricultural task.

*M. lupulina* interacts with mutualistic symbionts of both *Ensifer* (*Sinorhizobium, Rhizobium*) *meliloti* and *Ensifer medicae* species [6–8], forming a tight phylogenetic clade [9]. The potential for coevolution between *M. lupulina* plants and *Ensifer* microsymbionts was discovered, but no significant differences between *Ensifer* species have been reported [10]. Nevertheless, the authors have proposed that differences between *Ensifer* species are the only potential source of coevolutionary selection acting on *M. lupulina* [10]. *E. medicae* is known as a species showing host preference mostly for annual medics, while *E. meliloti* bacteria are predominantly symbionts of agriculturally valuable perennial alfalfa species [11,12]. The genomic characteristics of *M. lupulina* symbionts have been more frequently described for *E. medicae* than for *E. meliloti* [6,7,10,11,13].

Each prokaryote genome is represented by unique sets of core and accessory parts of the genome [14–16]. The accessory part of the genome is related to adaptation and strain diversification, while the core part comprises united genes common to all strains in a population encoding vitally essential cellular functions [17–20]. Core genes are usually localized near the origin of chromosome replication, *oriC* [21]. Internal recombination processes can significantly affect the positions of genes relative to their points at the origin and terminus of replication (*oriC* and *terC*). At the same time, the activity of mutation and recombination processes in the chromosome correlate with the positions of genes relative to *oriC* and *terC* [22,23]. The integration into chromosome of some elements, such as prophages and genomic islands (related to accessory genome), could provoke alterations in gene activity and, consequently, cell viability; that is, these changes can affect the phenotype in general [24,25]. A significant increase in the portion of foreign DNA that results from horizontal transfer is often detected in the chromosome region near *terC* [22,23,26]. Genes located in this region have higher mutation frequency, particularly associated with the transposition of IS elements [25,27,28]. Thus, conducting a comparative analysis using the alignment of a complete sequence of bacterial strain chromosomes is hampered by significant differences in their length and the lack of a single nomenclature for the replicons' orientation and starting points. In this work, we determined *terC* using bioinformatics methods, including GC-skew and localization of short non-coding repeats [29–34]. Altogether, it allowed us to align the chromosome sequences of compared strains and analyze the localization of genes, taking into account their localization corresponding to *oriC* and *terC*.

*Ensifer* species have a multipartite genome, united chromosome, and several plasmids of different sizes. In *E. meliloti*, there are two megaplasmids, which together make up two-thirds of the chromosome. pSymB is characterized as a chromid (chromosome-like replicon), which is structurally quite stable and can be experimentally mobilized for transfer under certain conditions [35,36]. The pSymA replicon can participate in intergenomic rearrangements or be mobilized for conjugative transfer in lab experiments with low frequency, and mobilization of pSymB is pSymA dependent [36–40]. Both megaplasmids should be considered rather stable characteristics of the *E. meliloti* genome, since according to the complete *Ensifer*/*Sinorhizobium* genomes in RefSeq (release 209, 5 November 2021), there are no data about *E. meliloti* native isolates lacking at least one megaplasmid, while both pSymA and pSymB are assigned to the mobilome. Non-symbiotic plasmids are revealed in nearly every strain (up to 86%) of *E. meliloti*, the size of which varies from 7 to 454 kb [41–45]. Modern approaches to rhizobia genomics research allow us to show that cryptic plasmids may contain sequences homologous to those at both megaplasmids and can carry genes functionally significant for bacteria fitness [36,46–48]. Some plasmids are self-transmissible or mobilizable and are an active part of the mobilome. The mobilome is a set of mobile genetic elements, such as plasmids, prophages, genomic islands, transposons, introns, and insertion sequences that are abundantly present in rhizobia genomes [49–53].

The accessory part of the genome is a diverse set of horizontally acquired strain-specific genes [16,39,54]. Its role is essential in specification of the prokaryotic genome, which in turn makes it possible for these organisms to occupy new ecological niches, including

interactions with their macrosymbionts, and gives them the ability to tolerate abiotic and biotic stresses [55]. Bacterial defense systems such as restriction–modification and CRISPR/Cas systems are frequently located on plasmids, and in those cases, they belong to the accessory part of the genome [56–59]. The above systems are related to bacterial immune systems and are essential to protect bacteria from foreign DNA invasion, including phages. In addition to the mentioned systems, some genes, such as *ropA*, encoding porin, may play a vital role in bacteria–phage interactions, since mutations in their sequences cause significant changes in resistance to phages [60]. Among the genes responsible for rhizobia tolerance to abiotic stress factors including salinity, acidity, desiccation, heat, etc., those related to the first two are the most studied. Thus, for *E. meliloti*, more than 1230 genes located on chromosome and plasmids have been shown to participate in responses to salt stress and acidic pH [61–63]. However, data supporting the functional role of genes responsible for the above types of stresses were reported for fewer than 40 genes [64–69].

An important group of accessory genes comprises those responsible for symbiotic rhizobia interactions with legume host plants. In the case of *E. meliloti*, these genes are mainly grouped into operons and localized on megaplasmid pSymA, but some are on megaplasmid pSymB and chromosome. The *nif*, *fix*, and *fdx* groups of genes are responsible for the nitrogen fixation process, and the *nod*, *noe*, and *nol* groups are responsible for the nodulation process [70,71]. Two of the three *nodD* regulatory genes, *nodD1* and *nodD2*, are activated by plant flavonoids, and *nodD3* is activated by SyrM, a bacterial transcriptional regulator [72]. The genes that determine the synthesis of the core part of the signaling Nod factor molecule are so-called common *nod* genes (*nodA*, *nodB*, *nodC*), which were identified in the vast majority of rhizobia [73]. *nodI* and *nodJ* genes encode the ABC transporters responsible for secretion of the Nod factor [74]. Other genes, including *nodE*, *nodF*, *nodG*, *nodH*, *nodL*, *nodM*, *nodP*, and *nodQ*, are characteristic of a particular rhizobia species or set of species [75]. Among the genes responsible for the interactions of particular rhizobia species with their host plants are *noe* genes, which are important for symbiosis of *E. meliloti* with *M. laciniata*, *M. littoralis*, and *M. lupulina* plants [76]. Some symbiotic genes have several copies located on different replicons of *E. meliloti*, such as *nodP* and *nodQ*, represented by two copies localized on pSymA and pSymB. Thus, *E. meliloti* species harbor about 50 symbiotic genes that are known for their functional importance in symbiosis with host plants from different genera (*Medicago* sp., *Melilotus* sp. and *Trigonella* sp.) [77].

All of the above symbiotic and stress-related genes (further *sym* and *str* genes) are well studied in reference strains of a wide range of species, and a large portion of them were not characterized when native isolates were studied; thus, our knowledge about allelic polymorphism of genes essential for symbiosis should not be considered comprehensive and objective. The evaluation of natively occurring polymorphisms of genes involved in symbiosis and stress tolerance is oriented toward searching for successful structural variations of genes, allowing rhizobia to form an efficient stress-tolerant symbiosis with legumes.

The objective of this research is to analyze the genome of *E. meliloti* L6-AK89, which is a streptomycin-resistant ($Str^R$) mutant of CIAM1775 (CIAM1775=AK89=A2=RCAM1775), an effective inoculant of *M. lupulina* cv. Mira (the fodder type standard) [4], cultivated on moderately acid soils in the NW agricultural region of Russia [78]. CIAM1775 is included in the GEBA-RNB [79]. This strain was isolated from pink nodule of wild-growing *M. lupulina* native to a secondary salinated area that is a part of the Mugodzhar center of alfalfa diversity in the NW of Kazakhstan, 500 km from the Aral Sea [80]. This region was explored for alfalfa diversity [81]. Both parent and mutant strains have a “salt-sensitive” phenotype (0.5 M NaCl), while a majority of strains from this region were shown to be tolerant to 0.6 M NaCl [80,82]. The phenotypic features of the L6-AK89 ($Str^R$) sparked our interest in conducting a deep screening analysis of its genome using NGS (MiSeq, Illumina) and MinIon (Oxford Nanopore Technologies) sequencing. Plant tests proving the symbiotic preference of *E. meliloti* L6-AK89 ($Str^R$) for cultivars, cultivar populations, and ecotypes of *M. lupulina* were performed in sterile microvegetative experiments. Our main interest was in evaluating the accessory part of the genome, represented by genes related to symbiosis

and tolerance to salinity, and exploring the mobilome of L6-AK89. Of primary interest was searching for and highlighting the unique genomic properties of *E. meliloti* L6-AK89 that provide this strain with the ability to form a highly effective symbiosis with *M. lupulina*, characterized by bud cleistogamy and adapted for vegetation in saline environments.

## 2. Materials and Methods

### *2.1. Bacterial Strains and Growth Conditions*

*S.* (*E.*) *meliloti* Rm1021 was used as a reference strain. Rm1021 is a streptomycin-resistant derivative of *E. meliloti* SU47 originally isolated from the alfalfa (*M. sativa*) nodule in New South Wales, Australia, in 1939 [83]. A native isolate of *E. meliloti* Md3/4 was trapped by *M. varia* from the soil sample collected in the NW of the Caucasus mountain region (laboratory collection) and was used as a contrast strain for phage typing. The *E. meliloti* strains were grown in liquid or on solid TY media at 28 °C [84].

### *2.2. Genome Sequences*

The 26 *Ensifer* (*Sinorhizobium*) spp. genome sequences from GenBank database (accessed in September 2021; Table 1) were used in this research.

**Table 1.** The genome sequences of *Ensifer* spp. strains used for analysis.

| Species | Strain | Assembly | Species | Strain | Assembly |
|---|---|---|---|---|---|
| *E. meliloti* | Rm1021 | GCA_000006965 | *E. meliloti* | KH35c | GCA_002197105 |
| | Rm2011 | GCA_000346065 | | KH46 | GCA_002197465 |
| | AK21 | GCA_009664245 | | RCAM1115 | GCA_014189595 |
| | AK555 | GCA_003044175 | | Rm41 | GCA_002197045 |
| | AK76 | GCA_016406285 | | RMO17 | GCA_000747295 |
| | AK83 | GCA_000147795 | | RU11/001 | GCA_001050915 |
| | B399 | GCA_002302375 | | S35m | GCA_015689095 |
| | B401 | GCA_002302355 | | SM11 | GCA_000218265 |
| | BL225C | GCA_000147775 | | T073 | GCA_002197145 |
| | CCMM B554 (FSM-MA) | GCA_002215195 | | USDA1021 | GCA_002197445 |
| | CXM1–105 | GCA_003044215 | | USDA1106 | GCA_002197065 |
| | GR4 | GCA_000320385 | | USDA1157 | GCA_002197025 |
| | HM006 | GCA_002197165 | *E. medicae* | WSM419 | GCA_000017145 |

### *2.3. Phage Typing*

Phage resistance of the target bacterial strains was assessed using the Fisher method according to [85]. For this process, 50 µL of overnight bacterial culture was plated on Petri dishes with TY medium containing 3% agar. The dish was divided into sectors, and 5 µL of different phage lysates was applied to each sector, with one used as the control. Incubation of bacterial strains with phage lysates lasted for 48 h at 28 °C. The 10 phages belonging to the *Siphoviridae*, *Myoviridae*, and *Podoviridae* families from the laboratory collection were used. These phages were isolated from soil samples collected from geographically distant regions (Dagestan and NW Russia).

### *2.4. Plant Tests*

Sterile plant tests were performed to evaluate the efficiency of symbiosis formed by *E. meliloti* L6-AK89 or Rm1021 strains with *M. lupulina* or *M. varia* plants [80]. We used *Medicago* seeds from the collections of the Federal Williams Research Center of Forage Production and Agroecology (FWRC FPA; Lobnya, Moscow region, Russia) and of the Federal Research Center N. I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR, Saint Petersburg, Russia). *M. varia* seeds were used in the experiment, since the parent strain of the reference strain *E. meliloti* Rm1021 was a native symbiont of *M. varia* [83]. Seeds from FWRC FPA were *M. lupulina* L. var. *perennans* Grossh. (cultivar Mira, cultivar populations VIK19 and VIK26), *M. lupulina* L. var. *vulgaris* Koch (cultivar population VIK32), and

*Medicago varia* Mart. (cultivar Taisiya). Seeds from VIR were *Medicago lupulina* L. var. *perennans* Grossh (ecotypes with catalogue numbers 52774 and 52741 from regions of Dagestan and Armenia, correspondingly).

*M. lupulina* and *M. varia* seeds were surface sterilized and scarified with concentrated sulfuric acid for 5 min and then washed with a large volume of sterile distilled water until pH 7.0. The seeds were germinated in Petri dishes for 3–4 days at 4 °C (stratification). Seedlings were placed by twos in 50 $cm^3$ test tubes with sterile vermiculite and 10 mL of Krasilnikov–Korenyako liquid medium without nitrogen and germinated for another 2 days at 20–22 °C [86]. After that, seedlings were inoculated with *E. meliloti* strains mixed with microelement solution. In all experiments, the inoculum was adjusted to approximately $5 \times 10^6$ cells $mL^{-1}$ [86]. Each plant × rhizobium combination in 8- to 20-fold replicas and control without inoculation (nci) for each combination in 10 replicas were grown in microvegetative sterile experiments with a 16 h photoperiod (16,000 lux, alternating 22 °C and 16 °C) for 56 days [80]. Variants of uninoculated plants of each tested variant of symbiosis were used as negative control of inoculation (nci). Three biological replicates were performed to obtain the results. The indicator of symbiotic efficiency of tested combinations was increased plant mass collected after 56 days of vegetation (yield parameter), length of shoots and roots (mm) and the ratio between plant morphological parameters calculated minus the corresponding nci (100%) according to [87–89]. Plant biomass was air-dried for 48 h at 60 °C in a drying chamber up to constant mass. Statistical data processing was performed by analysis of variance [90] using Microsoft Office Excel 2007 software.

### *2.5. Genome Sequencing and Assembly*

For DNA isolation, the strain was grown in liquid TY medium for 2 days at 28 °C on an orbital shaker at 180 rpm. DNA isolation was carried out as in [91,92]. The library for nanopore sequencing was prepared using the LSK108 kit and barcoded using the EXP-NBD103 according to the manufacturer's instructions.

Sequencing using the MinION sequencer was performed at the Core Centrum Genomic Technologies, Proteomics and Cell Biology FSBSI ARRIAM. The data from the run were demultiplexed, and the resulting number of reads for the strains was 30,284 (78 mln bp). Short-read sequencing was carried out on the Illumina MiSeq platform at the SB RAS Genomics Core Facility (ICBFM SB RAS), and 1.6 million paired reads with an average length of 296 bp were obtained.

Genome assembly and annotation of raw long and short reads were filtered prior to assembly. For Illumina reads, bbmap and bbduk utilities were used to remove adapter sequences, low-quality base pairs, and reads that did not belong to bacteria, as in [92]. Raw nanopore data were demultiplexed with Deepbinner (v 1.1; University of Melbourne, Parkville, Australia) [93] followed by Porechop (v 0.2.3; University of Melbourne, Parkville, Australia) [94], which also removed adapters and low-quality reads. Additionally, the reads were filtered with Filtlong (v 0.2.0; University of Melbourne, Parkville, Australia) [95], using Illumina reads for read improvement, and low-quality reads and reads shorter than 3000 nucleotides were filtered out. To assemble the genome, we used Flye assembler (v 2.8; University of California, San Diego, CA, USA) [96] with default parameters. The resulting assembly was polished 4 times with long reads using racon [97], followed by a run of medaka polisher (v. 1.2.1; Oxford Nanopore, Oxford, UK). We then followed the medaka step [98] with 5 rounds of iterative polishing with pilon [99] using the Illumina sequencing data, until the polishing stopped changing the assembly. This assembly was used for all subsequent analyses.

In order to make the coordinates of found genes consistent with previously published strains, we rotated the replicons according to those of Rm1021 using a custom script. The genome annotation was performed with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP v. 1.13; NCBI, Bethesda, MD, USA) algorithm [100–102].

### 2.6. Genome Analysis

The search for phage-related elements (prophages (Phs) and genomic islands (GIs)) in the genomes was carried out using the PHASTER web server [103,104] and Islander [105,106]. The analysis of codon usage was performed using the Sequence Manipulation Suite: Codon Usage with the standard genetic code [107]. CRISPR/Cas elements (CRISPR and Cas) were annotated using the CRISPRCasFinder program [108] using the default settings and other, different settings: minimal repeat length 20 bp, maximal repeat length 60 bp, detection of unordered *cas* gene using typing clustering model. The positions of minimum and maximum values of the cumulative GC-skew were determined using GenSkew (https://genskew.csb.univie.ac.at); the step size and window size were equal to 0.1% of the chromosome length. The search for tRNA sequences was carried out using tRNAscan-SE v. 2.0 (Todd Lowe Lab, Santa Cruz, CA, USA) [109,110]. The search for insertion sequence (IS) elements and their characterization down to the family was carried out using digIS [111] and ISfinder [112], correspondingly. The search for elements of conjugative transfer systems on the cryptic plasmid was performed using a web-based tool, oriTfinder [113]. Pairwise alignment of replicons of *E. meliloti* strains L6-AK89 and Rm1021 to construct the physical maps was performed using Mauve 2.4.0 (The Darling Lab, Sydney, Australia) [114]. Functional category analysis for gene products was performed using BlastKOALA [115]. Comparative analysis of gene nucleotide sequences and amino acid sequences of gene products was performed using MUSCLE [116], DnaSP v. 5.10.01 (University of Barcelona, Barcelona, Spain) [117] and MEGA v. 10.0.5 (Temple University, Philadelphia, PA, USA) [118]. Protein structure analysis was performed using ScanProsite (https://www.expasy.org/resources/scanprosite [119] and I-TASSER server [120]. Identification and analysis of the biosynthetic gene clusters (BGCs) in genomes of strains L6-AK89 and Rm1021 were performed with the antiSMASH pipeline (v. 5.1.1; Denmark, The Netherlands) using the "relaxed" setting set [121]. The genome of strain Rm1021 (GCA_000006965) was used as a reference.

### 2.7. Phylogenetic Analysis

The genome sequences of 26 *Ensifer spp.* strains were used in phylogenetic analysis (Table 1). Strain *E. medicae* WSM419 was used as an out-group. Multiple nucleotide sequence alignment was performed by MUSCLE [116]. Phylogenetic trees were constructed by IQ-TREE [122] with the maximum likelihood algorithm (1000 bootstrap replicates), and the trees were rendered with Dendroscope3 [123]. The relative rate of accumulation of substitutions (N) between sequences was calculated by using the formula: N = L/100 * n, where L is the sequence length and n is the number of substitutions per 100 nucleotides.

### 2.8. Microarray Analysis

The gene expression activity of bacterial cells of the two strains was assessed at 1 h and 12 h of salt stress using the technology of comparative genomic hybridization (DNA Microarray) on SM14kOli biochip (CEBITEC, Bielefeld, Germany) with oligonucleotides (70 bp) corresponding to fragments of 6205 genes of strain Rm1021 [62]. RNA was prepared from two independent cultures of strain L6-AK89 and reference strain Rm1021 grown in MOPS buffered GMS medium containing 0.5 M NaCl (induced cells). The reference RNA pool obtained from uninduced cells grown in MOPS buffered GMS medium was used in all hybridizations. These RNA samples were reverse transcribed into fluorescently labeled cDNAs using the combination control (Cy3)/induced(Cy5) and the dye swap combination control(Cy5)/induced(Cy3). In addition to these two biological replicates, one technical replicate was conducted by using RNA of the control and one of the induced cultures. Genes were considered differentially expressed if they were obtained for all four replicate spots and if the confidence indicator ($p$) was $\leq 0.05$, mean intensity (A) was $\geq 7$, and log2 of the expression ratio (M) was $\geq 1$ or $\leq -1$ (i.e., if there was at least a two-fold difference between the two experimental conditions) in all four experiments.

## 3. Results and Discussion

### *3.1. L6-AK89 in Symbiosis with Medicago spp.*

L6-AK89 ($Str^R$) was tested in plant tests with *M. lupulina* cultivar Mira, three cultivar populations, and two ecotypes, and also in symbiosis with *M. varia* cultivar Taisiya (see Section 2). Rm1021, whose parental strain was originally isolated from nodules of *M. varia* [83], was used to inoculate *M. lupulina* and *M. varia* plants as well. This approach was chosen in order to show the differences in the strain genotype × host plant genotype interaction, especially when each hop clover plant could be assumed to be a separate genotype because *M. lupulina* has bud cleistogamy. In total, the 14 different bacteria–host plant combinations with 8–20 replicates were evaluated by morphological parameters reflecting the intensity of symbiotrophic nutrition in tested strain × host plant combinations in agreement with [88,89,124,125].

When *M. lupulina* cv. Mira and cultivar populations VIK26 and VIK19 were inoculated with L6-AK89, the obtained average dry mass of plants corresponded to 22.29 ± 3.73, 22.7 ± 6.92, and 25.68 ± 0.92 mg/plant, respectively, after 56 days of vegetation. The increased dry mass amounted to an average of 144.5% (Figure 1A). This was significantly higher than the nci of *M. lupulina* plants ($\alpha = 0.001$) and the values obtained when *M. lupulina* plants were inoculated with Rm1021 (34.2%, $\alpha > 0.05$). For cultivar population VIK32 and ecotype no. 52741, the increases from L6-AK89 inoculation were 59.4 ± 26.8% and 71.2 ± 47.5% (19.66 ± 3.31 and 15.88 ± 4.4 mg/plant), respectively (Figure 1A). These differed significantly from nci, but not from Rm1021 (at $\alpha = 0.001$). Inoculation of ecotype no. 52774 with either strain L6-AK89 or Rm1021 did not lead to significantly increased plant biomass ($\alpha = 0.001$). The average plant dry mass of *M. varia* cv. Taisiya inoculated with L6-AK89 and Rm1021 amounted to 20.08 ± 1.6 and 20.94 ± 2.9 mg/plant, respectively. The increased dry mass amounted to 106.8 ± 16.5% and 115.65 ± 30.2%, respectively, compared to the nci of *M. varia* plants (Figure 1A). However, the symbiotic efficiency of both strains with *M. varia* did not differ significantly ($X^2 = 6.2 \cdot 10^{-2}$; $p = 0.80$).

According to the calculated coefficient of variation (V) of the dry mass of plants, it was found that the combinations formed by strain L6-AK89 were more aligned than those formed by strain Rm1021 with *M. lupulina* cultivar Mira and cultivar populations VIK26 and VIK19 (V = 21.0%, 37.2%, and 4.76% for L6-AK89 and 30.8%, 41.7%, and 13.0% for Rm1021, correspondingly).

The average value of dry mass for symbiotic combinations formed by L6-AK89 and hop clover (cv. Mira and VIK19, VIK26, VIK32; no. 52741) was 112.8 ± 31.6%, whereas it was about four times less for similar combinations formed with strain Rm1021 (29.9 ± 23.6%). Thus, yield responsiveness of hop clover toward inoculation with strain L6-AK89 of *E. meliloti* was observed.

It was found that inoculation with strain L6-AK89 of *M. lupulina* cv. Mira, as well as *M. varia* cv. Taisiya, contributed to a significant increase in the aboveground part of plants (9.5 ± 0.1 and 10.4 ± 0.6 cm, respectively), by more than 73% on average relative to nci ($\alpha = 0.001$). When inoculated with strain Rm1021 of *M. lupulina* cv. Mira, this value did not exceed 50.2%, whereas inoculation of the Taisiya cultivar of *M. varia* produced an 87.5% increase in the length of shoots compared with nci (8.2 ± 0.1 and 11.3 ± 1.3 cm, respectively) (Figure 1B). Inoculation with L6-AK89 of plants of cultivar populations VIK26, VIK19, and VIK32 contributed to an increase in the length of plant shoots by up to 60% (in the first two cases, the values are reliable: more than 8.7 ± 0.6 cm; $X^2 = 6.6$ and 6.4; $p = 0.01$) (Figure 1B). On the contrary, inoculation of plants of these cultivar populations with the Rm1021 strain led to a significant shortening of shoot length relative to nci ($X^2 = 13.9$; $p = 7.5 \cdot 10^{-3}$) (Figure 1B). Thus, it was found that the inoculation of plants of cultivar populations VIK26 and VIK19 with strains L6-AK89 and Rm1021 led to significant differences in the length of plant shoots.

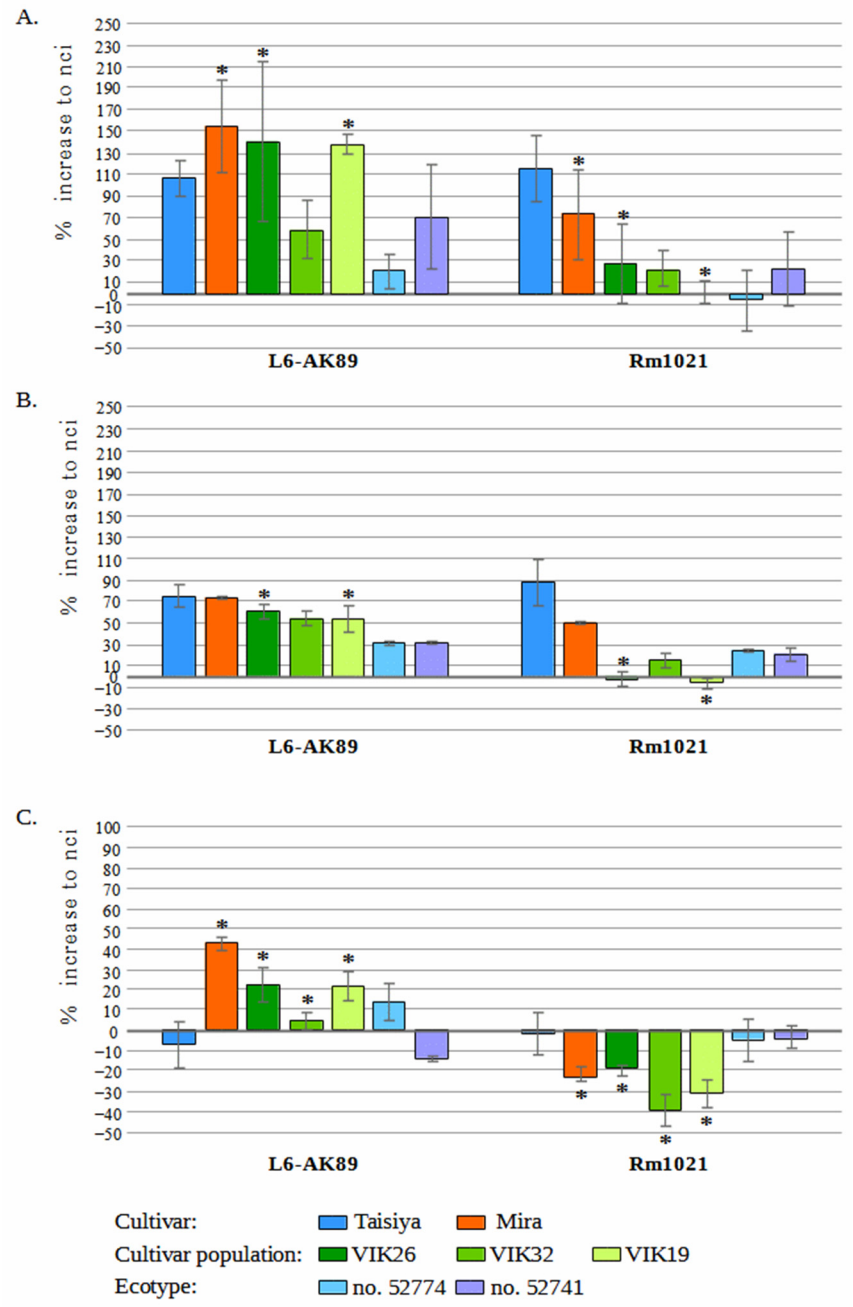


**Figure 1.** Symbiotic efficiency and morphological parameters of *Medicago* spp. in symbiosis with *Ensifer meliloti* L6-AK89 and Rm1021: (**A**) increase in dry mass of plants (% relative to control without inoculation); (**B**) relative change in length of inoculated alfalfa shoots (%); (**C**) relative change in length of inoculated alfalfa roots (%). Error bars represent ±SE. * Significant differences between analyzed strains ($p < 0.05$).

Similarly, an analysis of the length of the roots of inoculated plants was carried out. Inoculation with strain L6-AK89 of *M. lupulina* plants of cultivar populations VIK26 and VIK19 led to an increase in root length by more than 20% (12.3 ± 0.9 and 14.1 ± 0.8 cm), while inoculation of cv. Mira plants with the same strain contributed to an increase in more than 42.8% (12.3 ± 0.3 cm) compared to the corresponding nci. At the same time, inoculation with reference strain Rm1021 resulted in significant suppression of the root system growth of plants of these hop clover varieties (19.4%, 31.1%, and 21.8%, respectively; $X^2 = 4.39, 7.79, 9.08$; $p = 3.6 \cdot 10^{-2}, 5.3 \cdot 10^{-3}, 2.6 \cdot 10^{-3}$, respectively) (Figure 1C). Inoculation of *M. varia* cv. Taisiya with strain Rm1021 did not significantly affect root length, whereas

inoculation with L6-AK89 contributed to a decrease in root length. The differences between the combinations of strains and *M. varia* (host plant) were not significant ($X^2 = 0.09$; $p = 0.76$). This shows that inoculation with strains L6-AK89 and Rm1021 leads to a significant change in the root length of plants of cultivar Mira and cultivar populations VIK26 and VIK19.

Joint analysis of both morphological parameters showed that the increase in the ratio of shoot to root length in examined combinations of strain × host plant ranged from 20% to 90%. These data support significant differences between tested combinations. It was revealed that the ratios calculated for combinations between plants of cultivar population VIK26 (30.8 ± 9.9%, 21.8 ± 7.0%) and VIK19 (25.8 ± 8.9%, 36.8 ± 1.2%) and ecotype no. 52774 (15.1 ± 9.9%, 31.2 ± 22.7%) did not significantly differ when L6-AK89 or Rm1021 was used as inoculant, respectively. At the same time, the ratio calculated for combinations between plants of cultivar Mira (21.5 ± 11.8%, 91.9 ± 9.5%) and cultivar population VIK32 (46.7 ± 7.9%, 89.1 ± 4.5%) with L6-AK89 or Rm1021, respectively, showed significant differences between strains used as inoculant ($X^2 = 6.25$; $p = 0.012$). Development of plants with shorter roots was predominant in cases where the reference strain was tested as a microsymbiont. A similar situation was observed when strain L6-AK89 was used as inoculant for plants of ecotype no. 52741. Inoculation of *M. varia* cv. Taisiya plants with L6-AK89 or Rm1021 did not lead to significant changes in the corresponding ratios (87 ± 29.5% and 90 ± 24.4%, respectively). A comparison of the ratio of shoot to root length in all combinations (*M. varia*, *M. lupulina*) showed that the difference between strains was significant ($X^2 = 20.81$; $p = 0.002$).

Thus, it was shown that inoculation of tested cultivar/cultivar populations/ecotypes of *M. lupulina* with strain L6-AK89 leads to significantly more efficient accumulation of biomass yield (excluding ecotype no. 52774), increased shoot and root length of plants of cultivar populations, and possibly more balanced development of VIK26 and VIK19 plants, and the comparison was with the nci, as well as with Rm1021.

The presented data demonstrate that the genetic complementarity that develops in non-obligatory mutualistic symbiosis between *M. lupulina* and *E. meliloti* may differ significantly from symbiotic combinations between similar microsymbionts and *M. varia* plants. This is partly because, as *M. lupulina* exhibits bud cleistogamy (and each plant can be considered as a separate genotype), the genetics of symbiotic plant–microbe interactions are less effective, while symbiotic combinations between cultivars and strain L6-AK89 prove the obviousness of improving their further selection in future research. Therefore, the genomic characteristics of the L6-AK89 microsymbiont may be of particular importance.

### *3.2. L6-AK89 Genome Characteristics*

The size of the L6-AK89 genome was established as 6.79 Mb, which is 102 kb larger than the genome of reference strain Rm1021 (Table 2). The L6-AK89 genome contains chromosome, pSymA, pSymB, and accessory cryptic plasmid pL6-AK89 (0.35 Mb). The GC (%) composition of both megaplasmids is similar in both strains; however, the GC (%) of the chromosome of L6-AK89 is lower (on 0.8%) than that of Rm1021 (Table 2).

**Table 2.** Characteristics of replicons of L6-AK89 compared to reference strain Rm1021.

| Replicon | Size (bp) * | Number of | | | | GC Content (%) * |
|---|---|---|---|---|---|---|
| | | tRNA * | rRNA * | tmRNA * | ORFs * | |
| Chromosome | 3,580,141/3,654,135 | 53 **/52 | 9/9 | 1/1 | 3333/3315 | 62.9/63.7 |
| pSymA | 1,217,792/1,354,226 | 1/2 | 0/0 | 0/0 | 1156/1176 | 60.4/60.4 |
| pSymB | 1,645,902/1,683,333 | 2/1 | 0/0 | 0/0 | 1486/1463 | 62.5/62.4 |
| pL6-AK89 | 349,886 | 1 | 0 | 0 | 353 | 59.1 |
| Total: | 6,793,721/6,691,694 | 57/55 | 9/9 | 1/1 | 6328/5954 | 62.1/62.2 |

* Data are presented for L6-AK89/Rm1021. ** There is an additional truncated tRNA-Lys$^{(TTT)}$ (1797964–1798049, according to tRNAscan-SE v. 2.0) as part of the GI.

In order to make a comparative analysis between chromosomes and megaplasmids of L6-AK89 and Rm1021, the same origin points were designed and used to construct physical replicon maps (Figure 2). For chromosomes of L6-AK89 and Rm1021, the first nucleotide of the *oriC* sequence was designated as the origin point. The *oriC* region of L6-AK89 (477 bp) was identical to the *oriC* of Rm1021 (SMc04880).

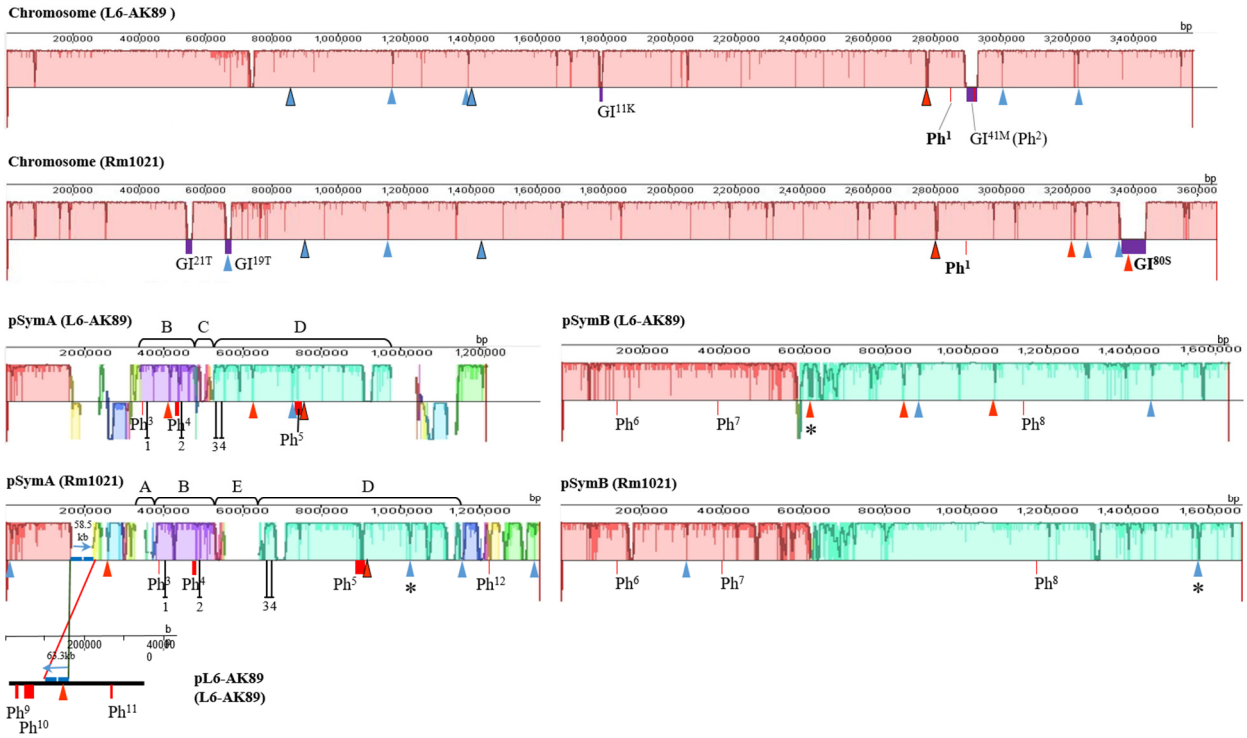


**Figure 2.** Full-genome analysis of *E. meliloti* L6-AK89 compared with reference Rm1021. Designations: IS-elements: I-XI, ISRm17; I-XIX, ISRm2011–2; /. Figures framed in black have identical localization of ISRm17 and ISRm2011–2, respectively, on replicons of strains L6-AK89 and Rm1021. * Disrupted IS element sequences (2 ORFs identified). Genomic islands (GIs): Width of rectangle corresponds to GI coordinates in chromosome: $GI^{11K}$, $GI^{41M}$, $GI^{19T}$, $GI^{21T}$, $GI^{80S}$, where superscript indicates length of GI in kb and capital letter corresponds to tRNA gene of a certain amino acid (site of integration). Incomplete prophages (Phs): Width of rectangle corresponds to Ph coordinates in replicon: $Ph^{1-12}$, where superscript indicates number of phages in list: [1] *Pseudomonas* phage nickie(NC_042091); [2] *Sulfitobacter* phage NYA-2014a(NC_027299); [3] *Salmonella* phage vB_SosS_Oslo (NC_018279); [4] *Xanthomonas* phage Carpasina (NC_047962); [5,10] Stx2-converting phage Stx2a_F451(NC_049924); [6] *Escherichia* phage ArgO145(NC_049918); [7] *Synechococcus* phage S-CBWM1(NC_048106); [8] *Streptococcus* phage Dp-1(NC_015274); [9] *Sinorhizobium* phage phiLM21(NC_029046); [11] *Escherichia* phage SH2026Stx1(NC_049919); [12] Stx2-converting phage 1717(NC_011357). On pSymA only: B and D collinear homologous regions; A, E, and C unique regions for Rm1021 or L6-AK89; ⊥ *sym* genes: 1 (*nodD2*), 2 (*nodM*), 3 (*fixS1*), 4 (*fixL*).

The *terC* region of the L6-AK89 chromosome was identified using two distinct approaches. The asymmetry in the distribution of a non-coding eight-base motif (NCMs), GGGCAGGG, on positive and negative DNA strands and the nucleotide composition (GC-skew) were analyzed according to [34,126]. The results obtained by the two methods were completely consistent with each other (Figure 3). The region in which the NCM transition from the positive to the negative DNA strand took place was defined as *terC*. In total, 278 NCMs were detected on the L6-AK89 chromosome, 157 on the positive strand and 121 on the negative strand (Figure 3). The length of the *terC* region of the L6-AK89 chromosome was found to be 9932 bp, and it had coordinates 1,697,784–1,707,715. In the same region, the maximum point of the cumulative graph of GC content on the L6-AK89 chromosome (1,704,080) was revealed (Figure 3). The *terC* region was found to be 1671 bp

longer than the similar region of strain Rm1021 (8261 bp), where the similarity of corresponding parts of *terC* was more than 99.9%. The difference in length between *terC* regions of analyzed strains was due to the identified sequence ISRel26 (of the ISNCY family) in the *terC* region of L6-AK89.

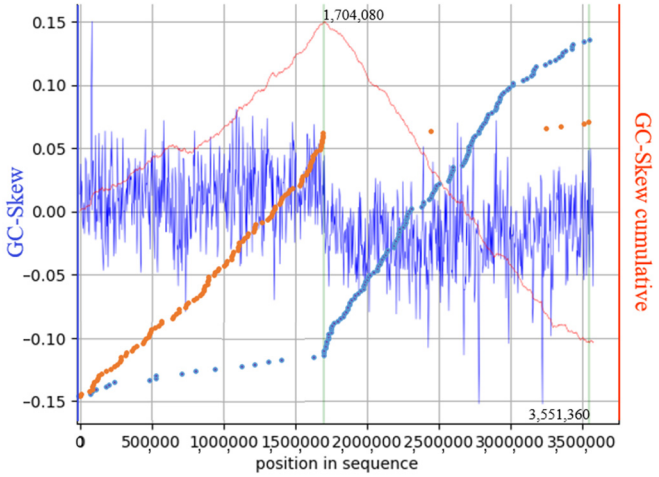


**Figure 3.** Distribution of non-coding 8-base motif and GC-skew on L6-AK89 chromosome. Dark blue line: GC-skew normal graph (step and window size, 0.1% of chromosomal length); red line: GC-skew cumulative graph; vertical green line: maximum and minimum of cumulative GC-skew graph; orange points: non-coding 8-base motifs (NCMs) located on positive DNA strand; purple points: NCMs located on negative DNA strand.

The alignment of chromosomal and megaplasmid sequences relative to their origins made it possible to correctly assess their sequences in genetically unrelated strains. The chromosome and pSymB lengths were shorter by 74 and 37.4 kb, respectively, and pSymA was 136.4 kb shorter in L6-AK89 compared with corresponding replicons of Rm1021. In total, three basic replicons were 248 kb shorter in L6-AK89, while it had an accessory cryptic plasmid pL6-AK89 with a fragment homologous to the 58.5 kb region of pSymA of Rm1021.

### *3.3. Analysis of Coding Sequences*

In total, 6328 proteins encoding ORFs were identified in the L6-AK89 genome, which is 374 (or 5.9%) more ORFs than in the Rm1021 genome (Table 2).

A phylogenetic analysis of the consensus sequence of the 16S rRNA gene sequence of the three *rrn* operons of *E. meliloti* and five core gene sequences of strain L6-AK89 was conducted in comparison with 25 strains of *E. meliloti* and *E. medicae* WSM419, taken as an outgroup (Table 1). All three sequences of the 16S rRNA gene were identical to each other in both strains. Since the 16S rRNA sequences in the genomes of some *E. meliloti* strains were different, the consensus sequences of the 16S rRNA gene conservative for *E. meliloti* species (Pi = 0.00025) were determined and a phylogenetic tree was constructed (Section 2, Figure S1A). The majority of sequences of *E. meliloti* strains (21 out of 26 tested strains, including L6-AK89 and Rm1021) were in the B2 cluster. The 16S rRNA gene sequences of this cluster were identical both within each genome and between strains. Cluster C contained consensus sequences of 16S rRNA of the three strains (CXM1–105, S35m, and KH35c) that were identical to each other, but different from those of L6-AK89 (evolutionary distance (N) was 0.01 bp per 100 bp). Cluster B1 contained only the sequence of strain AK76. The evolutionary distance (N) between L6-AK89 and AK76 was 0.02 bp per 100 bp.

All 16S rRNA sequences of studied *E. meliloti* strains differed from similar sequences of *E. medicae* WSM419 (Figure S1A). The relative rate of accumulation of substitutions (N) between the consensus sequences of L6-AK89 (also Rm1021) and WSM419 was 0.04 bp per 100 bp. Thus, the 16S rRNA gene sequences were identical for 80.8% of the tested *E.*

*meliloti* genomes, including Rm1021 and L6-AK89; thus, the latest strain is a typical strain of this species.

A phylogenetic analysis was performed using sequences of core genes used in similar studies [127–131]. The concatenate of the corresponding sequences, with a length of 7227 bp, was *recA-atpD-glnII-gyrB-dnaJ* (Pi = 7.4 × $10^{-3}$). Clusters A and B included all 26 concatenates of *E. meliloti* strains (Pi = 2.3 × $10^{-3}$) and the corresponding sequences of WSM419 were determined. Top-down clustering analysis showed that concatenates of L6-AK89 and Rm1021 were part of different groups, AI and AII, with these combining sequences of 12 and 8 strains, respectively, of *E. meliloti* (bootstrap 100%, Pi = 6.7 × $10^{-4}$ and 0.00, respectively) (Figure S1(B1,B2)). The relative rate of accumulation of substitutions (N) between the concatenate sequences of L6-AK89 and Rm1021 was 0.021 bp for every 100 bp of the sequence. Between the concatenate sequences of L6-AK89 and WSM419, the value of N was greater than between Rm1021 and WSM419 (9.987 and 9.007 substitutions per 100 bp, respectively). According to the data obtained, strain L6-AK89 is phylogenetically more distant from *Ensifer medicae* than Rm1021.

The analysis of the most ancient and conservative genes encoding RNA in the L6-AK89 genome resulted in 67 sequences encoding tRNA, rRNA, and tmRNA (Table 2). Nine genes encoding rRNA were related to three *rrn* operons; one gene encoding transfer-messenger RNA (tmRNA) was identical to the corresponding gene of Rm1021 (SMc04478). Out of 53 genes encoding tRNA on the L6-AK89 chromosome, 51 were identical to similar genes of Rm1021. The other two tRNAs were tRNA-Arg$^{(ACG)}$ and tRNA-Lys$^{(TTT)}$; the former had an SNP and was 99% identical to the corresponding gene of Rm1021, and the latter had similarity with the sequence on a genomic island (GI$^{41M}$) integrated into tRNA-Lys$^{(TTT)}$ (Figure 2).

On the pSymA of L6-AK89, tRNA-Sec (TCA) was detected, while a similar megaplasmid of Rm1021 contained that tRNA as well as tRNA-Met (CAT). The pSymB of L6-AK89 contained tRNA-Val$^{(GAC)}$ and tRNA-Arg$^{(CCG)}$; the same megaplasmid of Rm1021 had only the latter. On the accessory plasmid pL6-AK89, the identified tRNA-Val$^{(CAC)}$ was identical to tRNA on accessory plasmid of USDA1021, on pSymA of RU11/001 and SM11 strains, and on the GI located on chromosome of strain AK83 (Table 1).

Thus, some differences in the locations and sequences of tRNA genes in the L6-AK89 genome, as compared to Rm1021, were revealed. The fact that some tRNAs can be localized on GIs or accessory plasmids indicates that such tRNAs can be additional sites for the integration of foreign DNA, which is consistent with [131,132]. In addition, the fact that tRNAs are located on small plasmids indicates that these ancient conserved sequences are involved in horizontal gene transfer.

A search for gene clusters that determine the biosynthesis of secondary metabolites that potentially possess signaling or antimicrobial effects was carried out in L6-AK89 and Rm1021 using antiSMASH 5.0 and BLASTn. As a result, eight secondary metabolite clusters, including five homologous clusters on the chromosomes, were revealed in the genomes of both strains (cover = 100%, identity > 99%; Table 3). In addition, one cluster determining the biosynthesis of N-acetylglutaminylglutamine amide (NAGGN) was localized on the pSymB of both strains (cover = 99%, identity = 96.93%), which demonstrates its key role in cell osmoprotection. L6-AK89 has an additional osmotic shock response gene cluster on the pL6-AK89 plasmid that determines the biosynthesis of phosphate and ectoine, suggesting a strong response in order to maintain osmotic homeostasis. The pSymA of L6-AK89 contains an additional homoserine lactone cluster that may be responsible for the regulation of quorum sensing. Conversely, the reference strain had a cluster of antibiotic biosynthesis genes and a cluster of siderophore biosynthesis genes on megaplasmids, which are not present in the genome of the L6-AK89.

Thus, we show that the L6-AK89 genome has additional clusters of genes for osmotic shock response and quorum sensing, which can improve the viability of the symbiotically active L6-AK89 strain that possesses a salt-sensitive phenotype but was originally isolated from a region that suffered from salinity.

**Table 3.** Biosynthetic gene clusters in L6-AK89 and Rm1021 genomes.

| Replicon | Cluster Coordinates | | Cover/Identity (%) *** | Cluster Name | Putative Cluster Function |
|---|---|---|---|---|---|
| | L6-AK89 | Rm1021 | | | |
| Chromosome | 593,047–640,555 * | 620,572–668,086 * [620,572–658,665] ** | 100/99.54 | T1PKS | polyketide synthesis [133] |
| | 1,210,650–1,253,307 | 1,244,088–1,286,745 [1,244,088–1,286,734] | 100/99.94 | NRPS-like | nonribosomal peptide synthesis [134] |
| | 1,619,597–1,640,439 | 1,653,352–1,674,194 [1,653,352–1,674,194] | 100/99.98 | terpene | terpene biosynthesis |
| | 1,963,942–1,984,586 | 1,986,863–2,007,507 [1,986,863–2,007,507] | 100/99.97 | hserlactone | quorum sensing |
| | 2,224,345–2,246,364 | 2,245,746–2,267,765 [2,245,742–2,267,771] | 100/99.91 | TfuA-related | trifolitoxin production |
| pSymA | 493,358–513,978 | - | - | hserlactone | quorum sensing |
| | - | 1,304,062–1,319,558 | - | siderophore | rhizobactin synthesis [135] |
| pSymB | 459,736–481,971 | - [487,573–503,480] | 99/96.93 | NAGGN | osmotic shock response [136] |
| | - | 548,445–559,374 | - | bacteriocin | antibiotic |
| pL6-AK89 | 201,714–279,196 | - | - | phosphonate, ectoine | osmotic stress protection [137] |

* According to antiSMASH; ** identified by homology (BLASTn) with nucleotide sequences of clusters of strain L6-AK89 (identified by antiSMASH); *** according to BLASTn.

On the L6-AK89 chromosome, sequences LJD24_04610 and LJD24_08900, homologous to, respectively, the *ccrM* and *nifS* genes of Rm1021, were revealed. According to KEGG analysis conducted by BlastKOALA [115], the first one belongs to RM-II, and the second one encodes cysteine desulfurase and probably participates in the DNA phosphothiolation system [138]. These genes belong to different types of cell defense systems against foreign DNA [115].

The analysis of pSymA and pSymB of L6-AK89 was carried out in the direction of identifying sequences important for the maintenance of these replicons in the cell. On pSymA, the replication module *repABC* and genes related to conjugative transfer (*traA*, *traC*, *traD*, *traG*) were identified, as well the type IV secretion system genes (*virB1-virB11*) and transcriptional regulator *rctA*, which regulates conjugative transfer. On pSymB, the replication genes, represented by modules *repABC* and *rep*AB, were identified, as well the *traA* gene encoding relaxase. The localization and orientation of these genes on pSymA and pSymB of L6-AK89 were similar to corresponding replicons in Rm1021. The identity between *repABC* located on pSymA and pSymB of L6-AK89 and Rm1021 was 91.11 and 98.34%, respectively.

The accessory plasmid pL6-AK89 contained the replication module *repABC*, and the *repA* gene was detected at a distance of 59 kb from it; the product of the latter had an extremely low level of homology with RepA of the above module (cover = 84%, identity = 41.78%, E value = $9 \cdot 10^{-99}$). The identity between *repABC* on pL6-AK89 and the similar module on pSymA and pSymB was low (cover = 39 and 21%, identity = 66.45% and 68.11%, respectively). Elements of the conjugative transfer systems *oriT* (57 bp) and *traA*, the T4SS gene cluster, and type IV coupling protein were identified. In total, 353 ORFs were detected on pL6-AK89. The *groESL* operon on pL6-AK89 was similar to the *groESL2* on pSymA and to the *groESL1* operon on the chromosome (cover = 98% and 100%, identity = 99.94 for *groEL* in both cases; identity 97.64% and 97.98% for *groES* genes, respectively), which in Rm1021 was shown to be essential for symbiosis [139]. A similar *groESL* operon was detected on several accessory plasmids of strains AK83, SM11, RU11/001, GR4, USDA1021, and M162 (Table 1).

Module *acdS-lrpL* (LJD24_RS17290-LJD24_RS17285) encoding 1-aminocyclopropane-1-carboxylate (ACC) deaminase (AcdS)) and its regulatory protein leucine-responsive regulatory protein (LrpL), detected on pL6-AK89, was homologous to a similar module described on symbiotic megaplasmid pSmeSM11c of SM11 (cover = 100%, identity = 99.96%) [140], but it was not detected on Rm1021 (Figure S2). The length of this particular sequence was 16,507 bp and it had 17 ORFs and a fragment of IS6 (ISEc59). A part of the above region, specifically a fragment of 2083 bp (cover = 100%, identity = 99.23%) with the *acdS-lrpL* module flanked on the right side by 26 bp homologous to the 3′ end of IS5 (ISRm33), was homologous to a corresponding fragment on accessory plasmid pSmeSM11a of SM11 [46,141]. The sequences of *acdS* located on accessory plasmids of L6-AK89 and SM11 were identical, whereas *acdS*, located on megaplasmid pSmeSM11c with SNPs, did not provoke significant changes in protein conformation (cover = 99%, identity = 100%). Homologous *acdS-lrpL* modules identified in accessory plasmids of rhizobia isolated in geographically distant regions are voted for their abundance by horizontal gene transfer.

### *3.4. Mobilome of L6-AK89*

#### 3.4.1. Phage-Related Sequences Analysis

In total, 12 phage-related sequences (Ph and GI) were detected on replicons of pL6-AK89 using PHASTER and the Islander algorithm, respectively, and cross BLASTn alignment between phage-related elements annotated in genomes of these and other strains (Table 4).

Two out of the three chromosomal sequences (10.6 and 40.8 kb) were GIs site-specifically integrated into sequences encoding tRNA-Lys$^{(TTT)}$ and tRNA-Met$^{(CAT)}$ located at the third and fourth quarter of the chromosome, respectively (Table 4). In one of the above GIs, 16.8 out of 40.8 kb were related to incomplete prophage (according to PHASTER: score = 40, gene hit count = 8) and similar to the sequences of *Sulfitobacter* phages NYA-2014a$^{(NC_027299)}$, pCB2047_A$^{(NC_020858)}$, and pCB2047_C$^{(NC_020856)}$ (Table 4). Different amino acid codon usage by each GI was revealed, as well as between GIs and tested core genes, and these differences were significant ($X^2 > 118$, $p < 2.9 \cdot 10^{-5}$). These data confirm that GIs were acquired in the process of horizontal gene transfer, which is an agreement with [142–144]. The third 12.2 kb sequence on the L6-AK89 chromosome (in the fourth quarter) was similar to the sequence of the inc-Ph$^1$ of *Pseudomonas* sp. (according to PHASTER: score = 40, gene hit count = 2; Table 4), *Salmonella* phage SSU5$^{(NC_018843)}$ and *Rhizobium* phage RHEph10$^{(NC_034248)}$. Remarkably, a similar sequence was also detected on the Rm1021 chromosome (cover = 100%, identity = 99.84%).

Nine out of twelve prophages related to phages from the *Siphoviridae*, *Podoviridae, Myoviridae, and Herelleviridae* families were on megaplasmids pSymA and pSymB and cryptic plasmid pL6-AK89 (Table 4). All prophage sequences on pSymB and pL6-AK89 were identified using PHASTER, whereas prophage sequences on pSymA were detected by cross-alignment, with prophages detected on similar replicons of Rm1021 (using PHASTER). The length of seven Phs ranged from 5.8 to 9.5 kb, while Ph$^{10}$, located on pL6-AK89, was 23.5 kb (Figure 2). The three incomplete Phs localized on pSymB (Figure 2, Table 4) were similar to corresponding sequences on the same replicon of Rm1021 (cover = 100%, identity = 97.50–99.50%). Ph$^5$ had the same localization on pSymA of both strains (region D, Figure 2). The identity between these Ph sequences was 98.6%, since Ph$^5$ (31.8 kb) had a 7 kb deletion in L6-AK89 (Figure 2, Table 4). The incomplete Ph$^3$ and Ph$^4$ had the same localization in region B on pSymA in both strains, and inc-Ph$^4$ was between *nodF* and *nodC* genes (Figure 2). Both Phs showed homology to corresponding sequences in both strains (cover = 100%, identity = 99.18–99.63%).

Thus, *E. meliloti* may carry similar phage-related sequences not only on the chromosome, but on both megaplasmids as well. The revealed phage sequences are the trash from phage infection that probably occurred in the ancestral genome. The accessory plasmid inherited the three unique sequences related to phages from the order *Caudovirales* (*Siphoviridae* and *Podoviridae* families), two of which are associated with Stx phages from

the same order and are usually found in *Escherichia coli*. Our finding demonstrates the activity of phages from the same order in horizontal gene transfer between bacteria of distinct classes.

**Table 4.** Phage-related elements in *E. meliloti* strain L6-AK89 genome.

| L6-AK89 Replicon | Type of Insertion Sequence (GI[I]/Ph[II]) | Integration Site of GI/attP of Ph | Ph | GI/Ph Coordinates | GI/Ph Length (bp) | GI/Ph GC% |
|---|---|---|---|---|---|---|
| Chromosome | GI | tRNA-Lys$^{(TTT)}$ | - | 1,787,380–1,797,997 | 10618 | 60.4 |
| | inc-Ph[1] | TGCAGCGTTCATA | *Pseudomonas* phage nickie (NC_042091) III−1 | 2,847,359–2,859,574 | 12216 | 62.5 |
| | GI (inc-Ph[2]) | tRNA-Met$^{(CAT)}$ | *Sulfitobacter* phage III−2 | GI: 2,889,780–2,930,558 (Ph: 2,903,909–2,920,668) | GI: 40,779 Ph: 16,760 | GI: 59.3 Ph: 61.4 |
| pSymA | inc-Ph[3] * | - | *Salmonella* phage vB_SosS_Oslo (NC_018279) III−3 | 347,835–353,703 | 5869 | 58.8 |
| | inc-Ph[4] * | - | *Xanthomonas* phage Carpasina (NC_047962) III−4 | 423,074–432,147 | 9074 | 59.6 |
| | inc-Ph[5] *, ** | TGGCGGGCGCTC | Stx2-converting phage Stx2a_F451 (NC_049924) III−5 | 728,573–746,657, 746,650–753,387 | 18,085 6738 | 58.9 58.2 |
| pSymB | inc-Ph[6] | - | *Escherichia* phage ArgO145 (NC_049918) | 133,663–143,079 | 9459.6 | 62.1 |
| | inc-Ph[7] | - | *Synechococcus* phage S-CBWM1 (NC_048106) | 387,843–396,892 | 9050 | 62.3 |
| | inc-Ph[8] | - | *Streptococcus* phage Dp-1 (NC_015274) III−6 | 1,142,006–1,148,376 | 6371 | 62.5 |
| pL6-AK89 | inc-Ph[9] | - | *Sinorhizobium* phage phiLM21 (NC_029046) III−7 | 17,254–27,049 | 9796 | 58.7 |
| | inc-Ph[10] | CCGTCGCCTTCC | Stx2-converting phage Stx2a_F451 (NC_049924) | 43,164–66,618 | 23,455 | 58.1 |
| | inc-Ph[11] | - | *Escherichia* phages SH2026Stx1(NC_049919) III −8 | 264,406–272,442 | 8037 | 57.8 |

[I] Genomic island (GI) determined according to Islander algorithm. * inc-Ph indicated in Figure 2. [II] inc-Ph, incomplete prophage according to PHASTER. * inc-Ph according to cross BLASTn alignment. ** Phage fragments according to cross BLASTn alignment. [III] NCBI reference sequence according to PHASTER. III−1 Similarity was also found to *Salmonella* phage SSU5(NC_018843) and *Rhizobium* phage RHEph10(NC_034248). III−2 Similarity was also found to *Sulfitobacter* phages NYA-2014a (NC_027299), pCB2047_A(NC_020858), and pCB2047_C(NC_020856). III−3 Similarity was also found to *Pectobacterium* phage ZF40 (NC_019522), *Prochlorococcus* phage *p*-TIM68 (NC_028955), *Pseudomonas* phage phiPSA1 (NC_024365), YMC11/02/R656 (NC_028657), JBD44 (NC_030929), *Escherichia* phage Cajan (NC_028776), slur01 (NC_028831), Stx2-converting phage 1717(NC_011357), Seurat (NC_027378), *Cronobacter* phage ENT47670 (NC_019927), *Alteromonas* phage vB_AcoS-R7M (NC_048878), *Pantoea* phage vB_PagS_Vid5 (NC_042120), and *Lactococcus* phage bIL312 (NC_002671). III−4 Similarity was also found to *Xanthomonas* virus XcP1 (NC_048147), *Planktothrix* phage PaV-LD (NC_016564), *Ralstonia* phage phiRSL1 (NC_010811), *Escherichia* phage SH2026Stx1 (NC_049919), Stx2-converting phage 1717(NC_011357), Stx2a_F451 (NC_049924), 500465–2 (NC_049343), *Bacillus* phage Eldridge (NC_030920), BCU4 (NC_047735), Spbeta (NC_001884), G (NC_023719), and *Prochlorococcus* phage *p*-TIM68 (NC_028955). III−5 *Escherichia* phage SH2026Stx1 (NC_049919), Stx2-converting phage 1717(NC_011357), *Ralstonia* phage RsoM1USA (NC_049432), *Pseudomonas* phage MD8 (NC_031091). III−6 Similarity was also found to *Arthrobacter* phage Mufasa8(NC_049478), Sonali(NC_048152), *Mycobacterium* phage CRB2(NC_051585), and *Bacillus* phage BCD7(NC_019515). III−7 Similarity was also found to *Paenibacillus* phage Tripp(NC_028930). III−8 SH2026Stx1(NC_049919), Stx2-converting phage 1717(NC_011357), Stx2a_F451 (NC_049924).

### 3.4.2. CRISPR/Cas Analysis

The L6-AK89 genome was studied for immune response systems. Seven sequences in intergenic regions corresponded to characteristics of clustered regularly interspaced short palindromic repeats (CRISPR) cassettes, and three *cas* genes were revealed on the chromosome. One CRISPR sequence and one *cas* gene were on pSymB, but none of

the above-mentioned sequences were on pSymA or accessory plasmid pL6-AK89. The identified CRISPR cassettes and *cas* genes encoding type I proteins were not clustered together or associated with phage-related elements in the L6-AK89 genome. All detected CRISPR cassettes had one spacer, except one that had two spacers. The repeat size varied from 20 to 57, and the spacer size in terms of repeat size was 0.65 to 2.8. CRISPR cassette spacers shared similarities with phages from the *Siphoviridae*, *Podoviridae*, *Myoviridae*, and *Autographiviridae* families along with several eukaryotic viruses.

Chromosomes of tested strains contained identical sequences of *cas* genes. The sequence of *cas* gene on pSymB of L6-AK89 had six SNPs (one was nonsynonymous) and 3 bp insertion (CGG in positions 1428–1430) compared with the corresponding gene of Rm1021 (99% identity). These nucleotide substitutions led to amino acid replacement (Pro401 to Ser) and the appearance of Gly477, which did not occur in the active center of Cas; however, protein folding was altered (according to I-TASSER protein structure prediction) [120].

The described localization and organization of elements of CRISPR/Cas systems were common not only for L6-AK89 and Rm1021, but also for 24 *E. meliloti* strains for which ge-nomes were assembled (Table 1). Most CRISPR cassettes had SNPs in spacers and/or palindromic repeat sequences in all of the above indicated strains. Corresponding chromosomal sequences of *cas* genes showed 99.2–100% homology (cover = 100%); more than 55% of sequences of two *cas* genes and more than 74% of sequences of one *cas* gene were identical to L6-AK89 *cas* genes. The similarity between sequences of *cas* genes on pSymB in all of the above indicated strains was 97.8 to 99.9% (cover = 100%).

Summarizing, these unique data demonstrate that unclustered elements of the type 1 CRISPR/Cas system were revealed mostly on chromosome and pSymB in all strains of root nodule bacteria of *E. meliloti* species studied thus far. While the localization and structure of CRISPR cassettes and *cas* genes are quite conservative for that particular rhizobia species, their spacers showed identity to mostly abundant phages from the *Caudovirales* order. The role of these elements of the CRISPR/Cas system in the emergence of resistance of rhizobia to phages, or any other aspect of their fitness requires additional research.

#### 3.4.3. Abundance of IS Elements in Genome of L6-AK89

IS elements belong to the essential part of the mobilome and can affect gene activity. The L6-AK89 genome was analyzed for an abundance of IS elements using digIS [111] and ISfinder [112] programs (see Section 2). In total, 80 sequences related to IS elements were identified. Among them, 40 were full-sized IS elements, 24 were fragments, and 16 had no homology with IS elements in the ISfinder database (Table 5). Almost twice as many IS elements were found on the chromosome than on pSymB and accessory plasmid, and three times as many on pSymA of L6-AK89. At the same time, IS fragments, as well as n.s. sequences that could not be identified as specific to *E. meliloti* or as foreign sequences, were more frequently detected on pL6-AK89 than on the chromosome or both megaplasmids (Table 5).

Out of 40 full-sized IS elements of L6-AK89, 35 had 98–100% identity with IS elements of *Ensifer meliloti* (Table 5). ISSme2 and two copies of it were similar to IS elements of closely related *E. medicae* species (95–97% sequence identity). The other two IS elements were similar to ISRel16 of *Rhizobium etli* and ISRle5 of *R. leguminosarum* (80% and 89% sequence identity, respectively). The analysis of fragments showed that 14 out of 24 tested sequences were related to 7 IS elements from 7 IS families (Table 5). The other 10 fragments showed similarity to eight IS elements from other bacterial species: *R. etli*, *Nitrobacter* sp., *R. leguminosarum*, *P. versutus*, and *E. coli* (similarity did not exceed 91% with coverage of 44–95%; Table 5).

Some IS elements and their fragments were introduced into the genome of L6-AK89 by prophages. Thus, *Sinorhizobium* inc-Ph$^{9}$ contained a fragment of ISRel13 of *R. etli*; inc-Ph$^{5}$ harbored ISSme3 of *S. medicae*, ISRm4–1 of *S. meliloti*, and a fragment of ISEc59 of *E. coli*; and inc-Ph$^{11}$ had a full-sized ISRm14 of *S. meliloti* (Tables 4 and 5, Figure 2).

**Table 5.** Full-sized IS elements and their fragments in L6-AK89 genome[VI].

| IS Element | | | Number of IS Elements (Full-Sized/Fragments, n.c.) per Replicon [II/III] | | | | Origin [IV] | Identity [V] (%) | Total Number: (Full-Sized/ Fragments, n.c.) |
|---|---|---|---|---|---|---|---|---|---|
| IS Family [I] | Group [I/II] | IS Element [II] | Chromosome | pSymA | pSymB | pL6-AK89 | | | |
| IS256 | - | ISRm5 | 2/0 | - | 1/2 (≤79%, ≤14%) | 0/1 (≤45%) | *E. mel.* | 98–99 | 3/3, 0 |
| | - | ISRm3 | - | 1/0 | 2/0 | 2/0 | *E. mel.* | 99 | 5/0, 0 |
| IS4 | IS4Sa | ISRm16 (ISRm22) [VI] | 5/0 | - | - | - | *E. mel.* | 99–100 | 5/0, 0 |
| IS630 | - | ISRm2011–2 | 6/0 | 0/1 (≤91%) | 2/0 | - | *E. mel.* | 99–100 | 8/1, 0 |
| | - | n.c. [III] | - | - | $1_{714}$, $1_{696}$ | $1_{669}$ | - | - | 0/0, 3 |
| ISAs1 | - | ISRm21 | 1/0 | - | - | - | *E. mel.* | 98 | 1/0, 0 |
| IS1595 | ISNwi1 | ISRm32 | - | 0/1 (≤66%) | - | 0/2 (≤61%, ≤18%) | *E. mel.* | 92–96 | 0/3, 0 |
| IS481 | - | ISRm20 | 2/2 (≤84%, ≤17%) | - | - | - | *E. mel.* | 99–100 | 2/2, 0 |
| | - | n.c. | - | - | $1_{1143}$ | - | - | - | 0/0, 1 |
| IS3 | IS407 | ISRm7 | - | 1/0 | - | - | *E. mel.* | 99 | 1/0, 0 |
| | IS2 | ISRm1 | - | - | - | 1/2 (≤97%, ≤3%) | *E. mel.* | 95–99 | 1/2, 0 |
| | IS407 | ISRle5 | - | - | - | 1/0 | *R. leg.* | 89 | 1/0, 0 |
| ISNCY | ISDol1 | ISRm17 | 1/0 | 3/0 | 2/2 (≤96%, ≤4%) | 1/0 | *E. mel.* | 99–100 | 7/2, 0 |
| | IS1202 | ISRel26 | 0/1 (≤87%) | 0/1 (≤88%) | - | 0/1 (≤87%) | *R. etli* | 83–85 | 0/3, 0 |
| | IS1202 | ISRel10 | - | 0/1 (≤26%) | - | - | *R. etli* | 83 | 0/1, 0 |
| IS5 | IS903 | ISRm33 | - | - | - | 0/1 (≤72%) | *E. mel.* | 98 | 0/1, 0 |
| | IS1031 | ISRm4–1 | - | - | - | *1/0 | *E. mel.* | 98 | 1/0, 0 |
| | None model1 | ISNisp7 | - | 0/1 (≤70%) | - | - | *Nitr. sp.* | 79 | 0/1, 0 |
| | IS427 | ISRel13 | - | - | - | *0/1 (≤93%) | *R. etli* | 91 | 0/1, 0 |
| IS66 | - | ISRm14 | - | - | - | *1/0 | *E. mel.* | 97 | 1/0, 0 |
| | - | ISSme3 | - | 1/0 | - | *1/0 | *E. med.* | 95 | 2/0, 0 |
| | - | ISRle3 | - | - | - | 0/1 (≤84%) | *R. leg.* | 82 | 0/1, 0 |
| | - | n.c. | - | - | - | $1_{732}$ | - | - | 0/0, 1 |
| | - | n.c. | - | $1_{1533}$ | $1_{1533}$ | - | - | - | 0/0, 2 |
| | - | n.c. | $1_{1596}$ | - | - | - | - | - | 0/0, 1 |
| IS21 | - | ISPve1 | | | | 0/1 (≤44%) | *p. ver.* | 81 | 0/1, 0 |
| | - | ISRel16 | - | - | 0/1 (≤95%) | 1/0 | *R. etli* | 79–80 | 1/1, 0 |
| | - | ISSme2 | - | - | - | 1/0 | *E. med.* | 97 | 1/0, 0 |
| | - | n.c. | - | - | - | $1_{999}$ | - | - | 0/0, 1 |
| IS6 | - | ISEc59 | - | - | - | * 0/1 (≤85%) | *E. coli* | 82 | 0/1, 0 |
| IS30 | - | n.c. | - | $1_{905}$ | - | - | - | - | 0/0, 1 |
| IS110 | - | n.c. | $1_{453}$ | - | $1_{990}$ | - | - | - | 0/0, 2 |
| | - | n.c. | - | - | - | $1_{1068}$ | - | - | 0/0, 1 |
| | - | n.c. | - | - | - | $1_{657}$ | - | - | 0/0, 1 |
| | - | n.c. | - | - | - | $1_{864}$ | - | - | 0/0, 1 |
| ISL3 | - | n.c. | - | - | - | $1_{1740}$ | - | - | 0/0, 1 |
| In total: 15 IS-families | Total number: full-sized/fragments, n.c. IS | | 17/3, 2 | 6/5, 2 | 7/5, 5 | 10/11, 7 | - | - | 40/24, 16 |

[I] Determined with digIS. [II] IS elements determined according to ISfinder (yellow squares), sequences characterized as mobile element by digIS but did not have significant alignments in ISfinder database (unclear IS elements, further n.c.) (gray squares). [III] Value of coverage (%) for fragments in brackets, n.c. sequence length in lowercase; full-sized IS elements had 100% coverage. [IV] Original microorganism hosted IS elements. Origins: *E. mel.*, *Ensifer* (*Sinorhizobium* or *Rhizobium*) *meliloti*; *E. med.*, *Ensifer medicae*; *R. leg.*, *Rhizobium leguminosarum*; *R. etli*, *Rhizobium etli*; *Nitr.* sp., *Nitrobacter* sp.; *p. ver.*, *Paracoccus versutus*; *E. coli*, *Escherichia coli*. [V] Percent of nucleotide identity of IS element with reference sequence according to ISfinder. [VI] ISRm16 according to ISfinder, ISRm22 according to NCBI Annotation Tool. * IS elements identified in prophages.

A detailed analysis of IS localization in the L6-AK89 genome showed that some elements can be targets for the insertion of other IS elements. Thus, on the chromosome, the IS element (1596 bp; IS66 family) was flanked on both sides by ISRm20 (IS481), or ISRm3 was located within two fragments of ISRm17 on pSymB, or the two linked IS elements, ISSme2 and IS110, were flanked by ISRm17 fragments on pL6-AK89 (Table 4). Thus, ISRm20, ISRm17, and ISRm1 may be targets for insertion of IS66, ISRm3, and IS110, respectively, as revealed in the L6-AK89 genome.

Thus, it was demonstrated that the portion of *E. meliloti* IS elements was 1.4 times larger than the portion of mobile elements of bacteria from other species (foreign sequences). The ratio of IS elements of *E. meliloti* to "foreign" IS elements on chromosome compared with pL6-AK89 ($X^2 = 7.96$, $p = 0.005$), or on chromosome compared with pSymA ($X^2 = 5.16$, $p = 0.02$), or on pSymB compared with pL6-AK89 ($X^2 = 4.3$, $p = 0.04$) significantly differed. Thus, an overwhelming number of foreign IS elements were present in pL6-AK89 and in incomplete prophages.

All 80 sequences related to IS elements belonged to 15 out of 29 known families [111]. IS elements or their fragments belonging to four families (IS630, IS256, IS66, and ISNCY) were identified on all replicons of L6-AK89. Thirty different IS elements were represented by one to three copies, and only six different IS elements were represented by four to nine copies in the genome (Table 5). Thus, ISRm16 (ISRM22) of the IS4 family is represented by five full-sized copies on the L6-AK89 chromosome, while in the genome of Rm1021, there are eight full-sized copies on chromosome and one more copy on pSymB.

ISRm2011–2 (IS630 family) was assumed to be a species-specific IS element. The DNA hybridization pattern of this IS element suggested applying as a strain barcode for *E.* (*S.* or *R.*) *meliloti* species [145–147]. Eight full-sized copies (1053 bp) of ISRm2011–2 were revealed in the L6-AK89 genome, six of which were located on the chromosome (Table 5, Figure 2). Only two copies of ISRm2011–2 on the chromosome had localization similar to the corresponding IS elements of Rm1021 (Figure 2). However, in the L6-AK89 genome, despite the presence of ISRm2011–2, no introns were detected, unlike Rm1021, which was shown to have introns [148].

Another element of interest is ISRm17, which participates in the regulation of gene expression in *E. meliloti* strains [149]. Seven full-sized copies (1664 bp) and one copy of ISRm17 disrupted by ISRm3 (two fragments of 1600 and 60 bp) were detected in the L6-AK89 genome (Figure 2, Table 5). Only one of these ISRm17 copies had similar localization on the chromosomes of both strains, but in the case of L6-AK89 it was inverted.

Summarizing, the L6-AK89 genome abounds in full-sized IS elements, the majority of which show identity to IS elements of *E. meliloti* and other rhizobia species (*R. leguminosarum, R. etli)*, and even taxonomically distant bacterial species (*Nitrobacter* sp. *p. versutus, E. coli*). Some IS elements were identified in prophages of potentially mobile accessory plasmids, which allows us to suggest that the L6-AK89 genome is intensively involved in the process of horizontal gene transfer.

#### 3.4.4. Potential Effect of IS Elements on Adjacent Genes in "Head-to-Tail" Position

Since IS elements can significantly affect the metabolic activity of host cells [150], the location and orientation of the 40 full-sized IS elements and their fragments (80 sequences in total), including their potential influence on the activity of adjacent genes, were studied in the L6-AK89 and Rm1021 genomes using transcriptome profiling by microarray-based comparative genomic hybridization (CGH). It was determined that out of the 80 sequences related to IS elements, only 13 (16%) coincided in sequence and localization with IS elements of Rm1021. These IS elements were grouped into the following variants: (i) the location and orientation of IS elements and adjacent genes are similar in both strains; (ii) the IS elements are lacking in the genome of one of the strains and adjacent genes differ or are absent in the genome of the second strain; (iii) genes are disrupted by the insertion of IS elements; (iv) the IS elements are lacking in the genome of one of the strains, but adjacent genes are similar (Table S1). Variants iii and iv were identified in 5 and 23 cases, respectively, in L6-AK89, and both were of interest for assessment.

In the L6-AK89 genome, SMc01357, SMb21043, SMa0308, and SMa1787 (genes indicated the Rm1021 genome according to IANT 1021 DB) are disrupted by ISRm2011–2, ISRm5, ISRm17, and a fragment of ISNisp7, respectively (variant iii). All of these genes in the Rm1021 genome did not show changes in activity under salt stress (M value was in the range of $-1$ to 1, and A value was not statistically significant; see Section 2). Thus, it was concluded that all four genes are not involved in the response to salt stress, nor apparently do they play a significant role in the response to salt stress in L6-AK89. An alternative situation was when ISRm1 disrupted the sequence of gene encoding autoinducer binding domain-containing protein, which is annotated as two ORFs, SMc03896 and SMc03899, in the Rm1021 genome (Table S1). A similar gene in L6-AK89 is LJD24_RS16750, which was upregulated under 12 h of salt stress (M value = 1.33; Table S1).

The localization of IS elements in head-to-tail orientation was detected for 23 genes of L6-AK89 and 28 genes of Rm1021. Of these, activity was examined for 8 L6-AK89 and 15 Rm1021 genes that were homologous in both strains (Table S1). Significant upregulation of activity under conditions of 12 h salt stress was revealed for the SMa2377 gene of Rm1021 (M value = 1.05). This gene encodes MFS permease, and ISRm2011–2 is inserted in front of it, while the activity of the similar gene (LJD24_24730) in L6-AK89 did not change under stress conditions (M value = $-0.24$).

Summarizing, the analysis of the localization of IS elements in L6-AK89 and Rm1021 (variants iii-iv) showed that under 1 and 12 h salt stress, as analyzed by CGH, changes

in the activity of upstream genes occurred with a frequency not higher than 0.048. Thus, the head-to-tail construction relative to IS elements and adjacent upstream genes do not unambiguously lead to a change in the activity of these particular genes.

### *3.5. Fitness Genes*

#### 3.5.1. Stress-Related (str) Genes

The ability of bacteria to tolerate various abiotic stresses and phage infections is of vital importance for saprophyte rhizobia in soil. Taxonomically different groups of microorganisms are recruited for processing proteins/enzymes of similar function, frequently determined by homologous genes. A group of genes are involved in the formation of responses to some types of stress, such as salt or acidity stress, and phage infection, which are common in agricultural regions where parent strain CIAM1775 was used as inoculant [78]. This particular group included 28 genes related to stress tolerance (*str* genes), which have been widely studied [65–67,69,151–157]. Among them are 11 genes involved in response to salt shock (the first step of salt stress response, characterized by the accumulation of water and potassium ions by passive or active transport across the cell membrane), 10 genes involved in response to prolonged exposure to salt stress (a second step of salt stress tolerance results in the accumulation of osmolytes to maintain the turgor of cells) (Table S2), and seven genes related to ability of bacteria to resist moderately acidic pH conditions ($pH_{H2O}$ 5.5). Among genes of interest was a group of four *rop* genes encoding putative outer membrane proteins, porins, which have been shown to have amyloidogenic properties, and are themselves involved in cell–phage interactions [60,158].

The *ropA1*, *ropA2*, *ropB1*, and *ropB2* genes were identified on the L6-AK89 chromosome. The *ropB2* sequence (639 bp) of strain L6-AK89 has a single-nucleotide synonymous substitution relative to the similar sequence of Rm1021 (identity = 99.84%). The sequence lengths of *ropA1* and *ropA2* were 1038 and 1050 bp in the two strains, respectively, but the amino acid sequence identity of these proteins was 78.4%. Two nucleotide substitutions were detected in the *ropA1* sequence in L6-AK89, one of which led to an amino acid substitution in the extracellular domain of RopA1 (E81D) and non-synonymous substitution in the *ropA2* sequence. The revealed amino acid substitution is a product of the *ropA1* gene, which is a receptor for some phages from the *Myoviridae* family [60], leading to differences between L6-AK89 and Rm1021 in tolerance to 10 lytic phages. Lytic bacteriophage no. 2, isolated from soils of SW Dagestan and belonging to the *Myoviridae* family, was specific to strain L6-AK89 (Figure S3, Table S3).

Both L6-AK89 and Rm1021 have RCI site-specific recombinase between *ropA1* and *ropA2* genes. ISRm2011–2 (947 bp) was determined between *rci* (*xerC*) and *ropA2* genes in the Rm1021 genome, while this was not a case for the L6-AK89 genome. A downregulating effect of this IS element on the expression of *ropA1* and *ropA2* genes under 1 h salt stress was revealed for Rm1021 (M value = −1.28/−1.94; 12 h M value = −0.00/0.70 for *ropA1*/*ropA2*, respectively), while for L6-AK89 no change in the expression level of both *ropA* genes was detected (M value = −0.46/−0.92 for 1 h and −0.27/−0.66 for 12 h salt stress for *ropA1*/*ropA2*, respectively). Thus, a significant difference in the expression of both *ropA* genes was revealed between L6-AK89 and Rm1021 as well as in the phage resistance spectrum, particularly phages from the *Myoviridae* family, which may be of biotechnological interest in the future.

In all, 28 genes related to stress tolerance were studied for their copy number and localization in the L6-AK89 genome compared with Rm1021. Some differences were identified in L6-AK89, among them the presence of *trkH2* and *acdS* genes but the absence of *aqpZ2*. When nucleotide sequences of all indicated genes were analyzed, identity was shown for only 10 genes localized on the chromosome of both studied strains (*betA*, *betC*, *betI*, *trkA*, *trkH*, *actR*, *actS*, *actJ*, *actK*, *chvI*) (Figure 4).

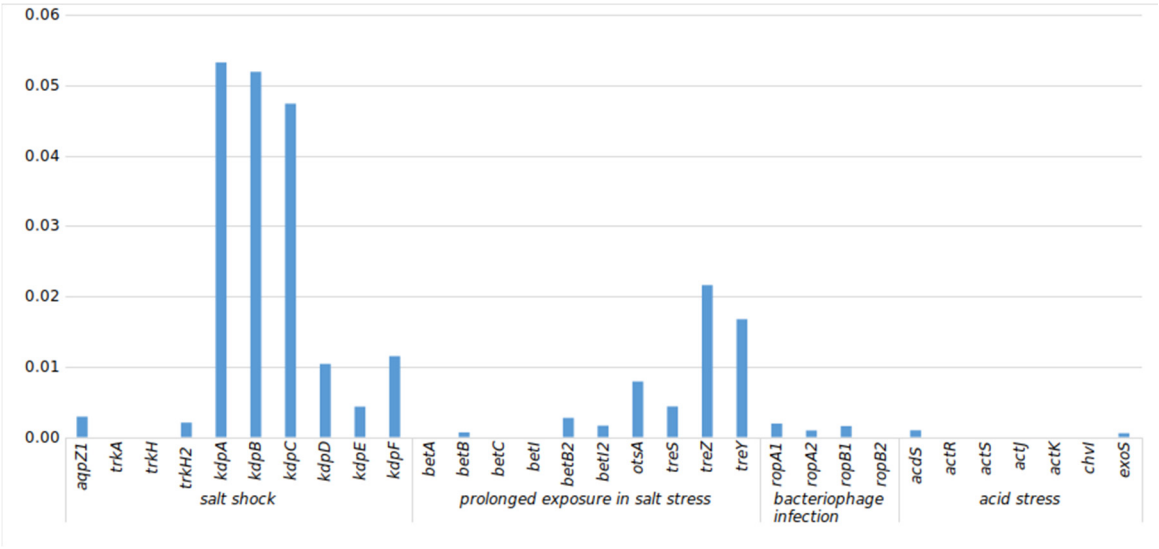


**Figure 4.** *p*-distance values for stress genes of *E. meliloti* strains L6-AK89 and Rm1021.

The average value of evolutionary distance (*p* distance) between the 28 studied nucleotide sequences was $7.9 \times 10^{-3}$; for genes located on the chromosome, it was $5.1 \times 10^{-4}$ and on pSymA and pSymB, it was $2.06 \times 10^{-2}$ and $1.27 \times 10^{-2}$, respectively, and was only $9.8 \times 10^{-4}$ for sequences on accessory plasmid (Figure 4). The largest nucleotide distance was identified for sequences of the *kdpFABCDE* operon, encoding proteins involved in active transport of potassium cations (first step of salt stress response) and for *treY*, *treS*, *treZ*, and *otsA* gene sequences (second step of salt stress tolerance).

A phylogenetic analysis of nucleotide sequences of a group of 10 *str* genes was performed for L6-AK89 in comparison with 25 strains of *E. meliloti* and *E. medicae* WSM419 (Table 1). The sequences of these strains were concatenated, the order of the genes was *betABCI–kdpABCDE–otsA*, and the size of the consensus sequence was 14,447 bp (Pi = 0.02571). These particular genes were selected, as they are involved in the response of bacterial cells to stress and were predicted to be part of the ancestral gene set of the ancestor of alpha-proteobacteria [159]. As a result, two clusters, A and B, were identified (Figure 5). Cluster B included only one concatenated sequence of WSM419, and cluster A included all 26 concatenated sequences of *E. meliloti* strains (Pi = 2.3·10–3). Top-down clustering analysis showed that concatenated sequences of L6-AK89 and Rm1021 were part of different groups, A2.1 and A2.2, respectively (bootstrap 67%). Group A2.1 contained sequences of 11 strains of *E. meliloti*, including sequences of strains AK83, AK555, AK21, and AK76 isolated from the same region as L6-AK89 (Pi = 0.0031). Group A2.2 contained sequences of 10 strains of *E. meliloti* and was three times more diverse than group A2.1 (Pi = 0.00927).

The relative rate of substitution accumulation (N) between concatenated sequences *betABCI–kdpABCDE–otsA* of *E. meliloti* L6-AK89 and Rm1021 was 4.87 per 100 bp, while the rate between concatenated sequences of putative ancestral *E. meliloti* and *E. medicae* sequences (EmelA and EmedA, respectively) was more than 15 times higher; however, the rate between *E. meliloti* L6-AK89 and *E. medicae* WSM419 was less than that between Rm1021 and WSM419 (74.8 and 77.8 substitutions per 100 bp, respectively).

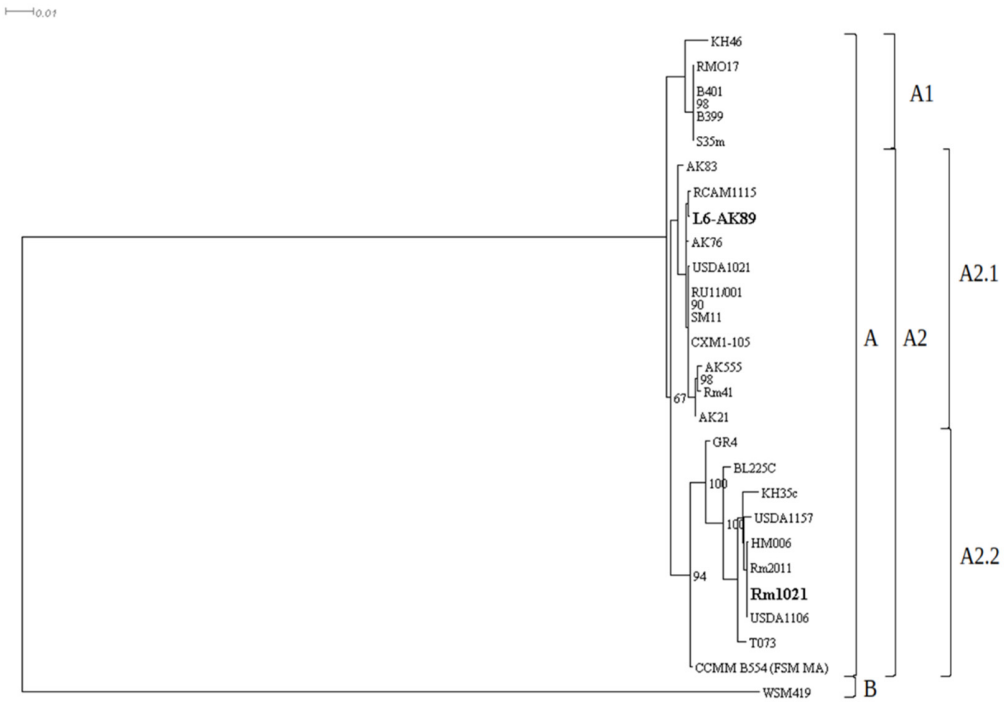


**Figure 5.** Phylogenetic analysis of concatenated nucleotide sequences of *bet-kdp-ots*A genes of *Ensifer* spp. Nucleotide substitution models selected for analysis were GTR + F + R5. Scale bar = 0.01 for concatenated *bet*/*kdp*/*ots*A nucleotide substitutions per site. Rooted tree was created by using maximum likelihood algorithm. Statistical support for internal nodes was determined by 1000 bootstrap replicates.

Consequently, the concatenated sequence of *E. meliloti* L6-AK89 is phylogenetically closer to the similar concatenated sequence of *E. medicae* WSM419 and closer than the corresponding sequence of reference strain *E. meliloti* Rm1021. Thus, phylogenetically, *betABCI–kdpABCDE–otsA* genes related to the stress metabolism of L6-AK89 are evolutionarily closer to corresponding genes of *E. medicae* WSM419 and are more conservative than similar genes of reference *E. meliloti* Rm1021.

According to the microarray data, the expression of the above genes under 1 h and 12 h of salt stress (0.5 M NaCl) did not significantly differ from that of genes with a constitutive level of expression in both strains, excluding *kdpD* in Rm1021, which showed downregulated expression under 1 h stress (M value = −1.71; Table S2). Thus, the 28 genes known to be related to stress were not significantly upregulated L6-AK89 under both durations of salt stress (1 and 12 h).

Additionally, we studied chaperones, since they play a significant role in the viability and stress tolerance of rhizobia and are involved in symbiosis formation [139,160]. It is known that Rm1021 has four *groESL* operons and one *groEL*, according to [139]. However, five *groESL* operons and two *groEL* genes were identified in the L6-AK89 genome. The additional *groEL* is located on pSymA, and the *groESL* operon is on accessory plasmid pL6-AK89. The latter showed high homology to *groESL1* and *groESL2* operons on the chromosome and pSymA of L6-AK89, respectively (cover = 98%, identity = >99.5%). Both *groESL1* and *groESL2* operons of L6-AK89 were downregulated under 1 and 12 h of salt stress (0.5 M NaCl; M value = −1.15 and −1.27, −1.43 and −1.48 for 1 and 12 h, respectively), while the expression of *groEL2* on pSymA was constitutive in that strain as well as the reference strain. The expression of *groEL1*, located on chromosome and involved in symbiosis, was downregulated in Rm1021 after 1 h (M value = −1.29), whereas the expression of *groEL1* was constitutive under 1 and 12 h of salt stress in L6-AK89. Thus, the GroEL chaperone located on the chromosome may be involved in the response to salt shock

(1 h). The obtained data expand our understanding of the functional role of chaperones in *E. meliloti* stress tolerance.

Thus, the activity of the *ropA1*, *ropA2*, *kdpD*, and *groEL* genes was downregulated in Rm1021, whereas similar genes in L6-AK89 were expressed constitutively under salt shock and during prolonged salt exposure, which can support the vitality of the salt-sensitive phenotype of this strain in a salinized environment.

#### 3.5.2. Sym Region Analysis

The pSymA of L6-AK89 and Rm1021 contained extended collinear homologous regions, both designated as region B, with lengths of 141.2 and 141.9 kb, respectively (Figure 2). Each region B is flanked by *dgoK2* and LJD24_21270 (SMa0937) genes. The structural differences between the two regions were mainly due to the indels and/or IS elements in intergenic regions. Inside each region, there is a Sym region more than 80 kb in size (81,834 and 80,770 bp), which in both strains is flanked by *nodD2* and *nodM* genes (Figure 2). The Sym region contains 40 *sym* genes in both strains: 17 *nod* genes, 2 *noe*, *nol*, and *fdx* genes, 8 *nif* genes, and 11 *fix* genes. Extended collinear homologous regions D (448.6 and 503 kb, respectively, in L6-AK89 and Rm1021) contains 13 *fix* genes (*fixS1*, *fixI1*, *fixH*, *fixG*, *fixP1*, *fixQ1*, *fixO1*, *fixN1*, *fixM*, *fixK1*, *fixT1*, *fixJ*, *fixL*) in both strains. Regions B and D are bordered by extended nonhomologous regions A and E or C. Regions C and E are adjacent to the 3′ end of region B on pSymA of L6-AK89 and Rm1021, respectively. The length of region E is 120.1 kb, while region C is more than 2.5 times shorter (46.2 kb) (Figure 2). Region A, with a length of 53.5 kb, is flanked by *fixN3* and *SMa0702* and adjacent to the 5′ end of region B in Rm1021 but was not found in L6-AK89. In total, 52 ORFs were identified in region A, including several genes related to amino acid metabolism and encoding transposases. It also contained operon $fixNOQP_3$, similar to that in the "common ancestor" Rhizobiaceae [161], located together with *fixI2* and *fixS2* genes. Two other operons *fixNOQP* involved in the symbiotic respiration process were present on pSymA in Rm1021 and L6-AK89 as well. $fixNOQP_1$ is localized in region D and $fixNOQP_2$ in region B.

The analysis of the structure of pSymA in L6-AK89 showed that it contains six fewer *fix* genes and the whole region of *nod*/*noe*/*nol*/*nif* /*fdx*/*fix* genes has a more compact structure than in the reference strain.

Nucleotide sequence analysis revealed that only seven *sym* genes on pSymA (*fixA*, *fixB*, *fixO1*, *fixU*, *fdxB*, *nodD2*, *nodF*) and *nifS* the L6-AK89 chromosome were identical to similar genes of Rm1021. Nucleotide diversity of 20 other *fix* genes varied greatly (*p* distance values from 0.001 to 0.135), and the maximum level of diversity was detected in *fixQ2* and *fixH* (0.118 and 0.135; Figure 6). A high level of nucleotide variability was also revealed for the *nifA* gene, encoding global transcriptional regulator NifA (*p* distance = 0.022) in L6-AK89 in contrast to Rm1021 (Figure 6). For other *nif* genes, *p* distance values did not exceed 0.005.

Synonymous nucleotide substitutions were identified in eight genes (*fixC*, *fixQ1*, *fixT2*, *nifD*, *nifH*, *nodA*, *nodB, nodG*) of L6-AK89 compared to Rm1021. For five genes (*nodN*, *nodQ1, nodQ2, nodP2*, and *nodP1*), deletions and insertions resulted in a frameshift, and in the first cases, stop codons were revealed and corresponding genes were annotated as pseudogenes. Only in the *nodP1* gene did a disruption of transcription not occur. *nolR*, a negative regulator of nodulation, is inactivated in Rm1021 due to single nucleotide insertion [162], but no insertion or other substitutions were found in *nolR,* and this gene could be functional in L6-AK89.

Phylogenetic analysis was carried out for aligned concatenated sequences of 26 strains of *E. meliloti* and *E. medicae* WSM419 (Table 1, Figure 7). Concatenates consisted of sequences of 21 *nod*/*noe*/*nol* genes combined into one consensus sequence 24,823 bp in size, according to the order of their localization on pSymA: *nodD2–nodL–noeA–noeB–nodD3–nodH–nodF–nodE–nodG–nodP1–nodQ1–nodJ–nodI–nodC–nodB–nodA–nodD1–nodN–nolG–nolF–nodM*. As a result, it was found that the relative rate of accumulation of substitutions (N) between clades containing concatenated sequences of *nod–nol–noe* genes of two rhizobia species

was 17.5 bp per 100 bp. Subcluster B1.1a2b1, obtained as a result of descending clustering (Figure 7) and containing the concatenate of L6-AK89 strain and three promising highly efficient strains used in practice, combined 42% of the concatenates and was 152.7 times more diverse than cluster B1.1a2b2 containing the Rm1021 concatenate, according to the calculated values of Pi (0.00458 and 0.00003, respectively). Moreover, the concatenate of L6-AK89 was phylogenetically farther from the ancestral sequence than those of Rm1021 (N = 5.26 and 4.12).

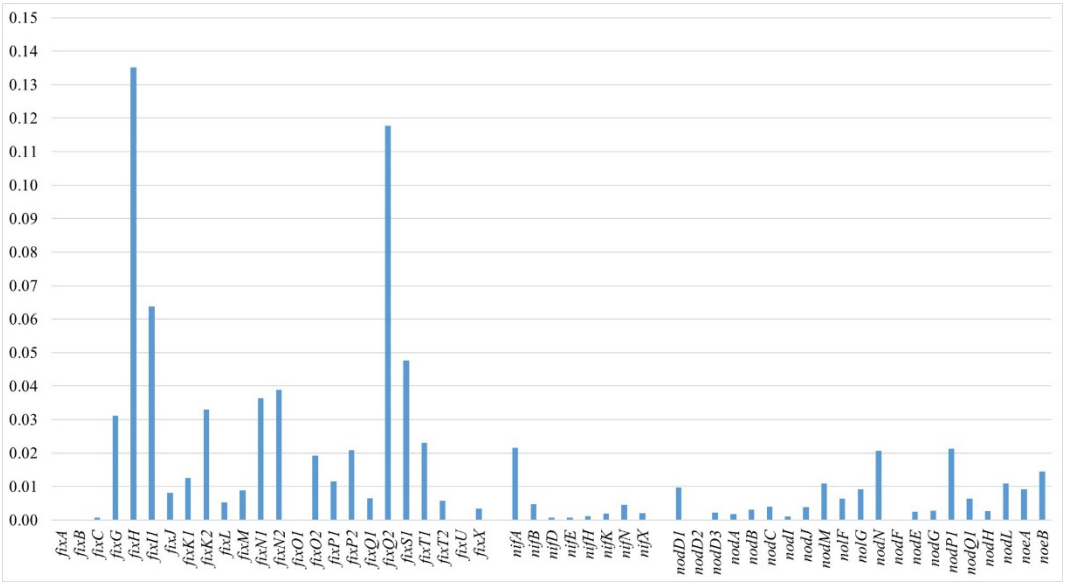


**Figure 6.** *p*-distance values for *sym* genes of *E. meliloti* L6-AK89 compared with Rm1021.

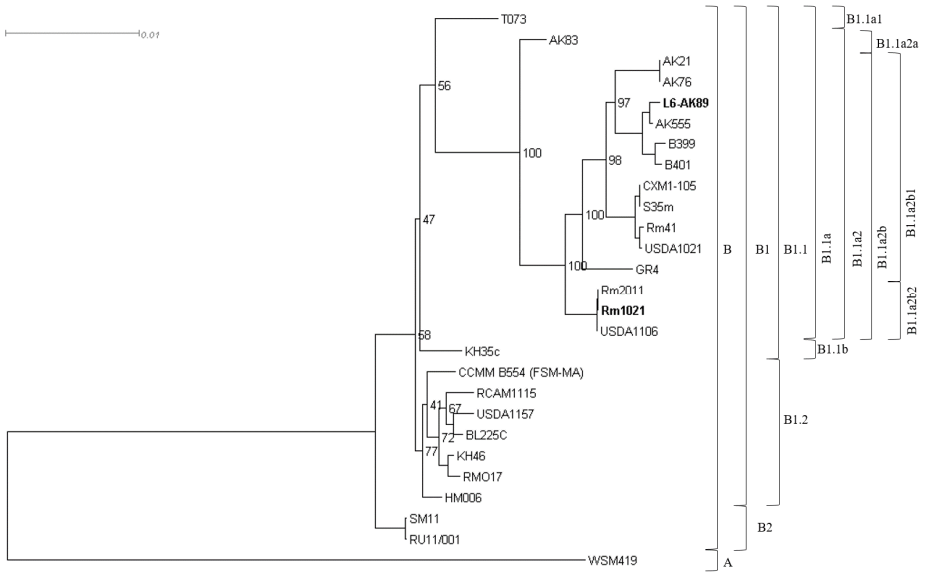


**Figure 7.** Phylogenetic analysis of 21 concatenated *nod/noe/nol* nucleotide sequences in *Ensifer* spp. Nucleotide substitution model selected for analysis was TN + F + I + G4. Scale bar = 0.01 for *nod/noe/nol* substitutions per site. Rooted tree was created by using maximum likelihood algorithm. Statistical support for internal nodes was determined by 1000 bootstrap replicates.

Differences in amino acid sequences were revealed in 13 out of 21 *nod/noe/nol* genes localized on pSymA (*nodD1*, *nodD3*, *nodC*, *nodI*, *nodJ*, *nodM*, *nolF*, *nolG*, *nodE*, *nodH*, *nodL*, *noeA*, *noeB*) and in chromosomal *nodN2* of L6-AK89. Amino acid substitutions also resulted in replacement of one to another polar amino acid in NodD1 (L90I) and NodD3 (S256N).

Both substitutions did not affect corresponding HTH motifs within the domains responsible for DNA binding. Two substitutions were detected in NodC, one in the N-terminal domain (A39S) and another in the C-terminal domain (M330I), responsible for the synthesis of Nod factor chain from glucosamine monomers. One substitution of nonpolar on polar amino acid (A81T) was in the ATP-binding cassette of in NodI, involved in the secretion of Nod factor.

Four amino acid substitutions were revealed in NodM, three of which were in the catalytic domain (according to in silico analysis). Two substitutions replaced one polar acid by another one, which were hydrophilic amino acids in both cases (N10H and R23P). The third replacement (T132A) was the appearance of another nonpolar amino acid in the protein structure. The isoelectric point of NodM in L6-AK89 and Rm1021 did not change significantly (5.66 and 5.64, respectively). All of the substitutions described above could affect the structure of the NodM catalytic domain in L6-AK89, while the constant of protein did not change significantly. Two amino acid substitutions, G54R and P170A, were revealed in NodL (O-acetyltransferase) of L6-AK89. The former was a replacement of nonpolar on polar hydrophilic amino acid. This enzyme is necessary for the synthesis of the O-acetylated species-specific Nod factor, and it has a key function in rhizobia interaction with roots of hop clover plants (*M. lupulina*) [76].

Three amino acid substitutions (D6E, D61V, N458R) were revealed in NoeA of L6-AK89, and the isoelectric point of the protein differed from those in Rm1021 (7.81 and 7.00, respectively). Nine substitutions were revealed in NoeB; two of them were the replacement of one nonpolar by another one (F54V, I522M) and two were the replacement of nonpolar by polar amino acids (H225R, H497), three were the replacement of polar by nonpolar amino acids (E95G, S239L, T287I), and two were the replacement of nonpolar by polar amino acids (V237E, F261Y). The isoelectric point of NoeB of L6-AK89 is 7.27, while in Rm1021, it is 6.81. Thus, significant amino acid substitutions were revealed in products of *nodL*, *noeA*, and *noeB* genes in L6-AK89, a native symbiont of *M. lupulina*, compared with the Rm1021 symbiont of *M. varia*.

The analysis of the Sym region of L6-AK89 revealed a high level of variability in the nucleotide sequences of *fix* genes, while *nif* genes encoding the nitrogenase complex are highly conserved, a finding consistent with the literature. Significant differences in the nucleotide and amino acid sequences of regulatory gene *nifA* and its product (*p* distance = 0.022 and 0.028, respectively) of strain L6-AK89 with respect to the reference Rm1021 were revealed. *E. meliloti* L6-AK89 showed significant differences in nucleotide and amino acid sequences of both regulatory and structural (common and species-specific) nodulation genes (*nodD3*, *nodD1*, *nodC*, *nodI*, *nodJ*, *nodM*, *nolF*, *nolG*, *nodE*, *nodH*, *nodL*, *noeA*, *noeB*) and their products, that together might contribute to its symbiotic complementarity with *M. lupulina* host plant.

## 4. Conclusions

This paper presents an in-depth analysis of the accessory genome of an economically promising strain of *Ensifer* (*Sinorhizobium*) *meliloti*, L6-AK89, obtained as a result of next-generation high-throughput sequencing (MiSeq, MinIon). High-yield responsiveness of hop clover varieties, including the world standard fodder type *M. lupulina* cv. Mira, three cultivar populations, and ecotypes, to inoculation with the L6-AK89 strain of *E. meliloti* was shown in microvegetative experiments.

The genome of L6-AK89 is represented by a chromosome and two megaplasmids, and the length of each replicon was shorter comparing with corresponding replicons in the reference strain Rm1021. Additionally, the strain L6-AK89 has an accessory plasmid pL6-AK89 (349.9 kb) due to which its genome size was more than on 100 kb larger than the genome of Rm1021. Since L6-AK89 is a typical representative of the genus *Ensifer* according to phylogenetic analysis of the consensus 16S rRNA gene sequences and concatenated core genes, the accessory part of the genome was our main focus. The portion of sequences related to the accessory part of the genome, which were detected on all four replicons, was 1.5 times higher compared to the genome of reference strain Rm1021 (7.4 and 4.8%,

respectively). All of these elements are related to the mobilome and are involved in horizontal gene transfer according to earlier reports [49–51,163–166]. In total, 10 inc-Phs homologous to phages from the *Siphoviridae, Podoviridae*, and *Myoviridae* families, two genomic islands, and 64 full-sized IS elements and their fragments belonging to more than half of the known IS families were detected in the L6-AK89 genome. Our findings show that the portion of IS elements related to *E. meliloti* was 1.4 times greater than the portion belonging to other rhizobia species or to taxonomically distant bacteria species, such as *Nitrobacter* sp., *p. versutus,* and *E. coli*. The evaluation of potential activity of 23 homologous genes adjacent to IS elements (head-to-tail orientation) under two types of salt stress clearly showed that a change in gene activity is a quite rare event in genomes of both tested strains.

Evaluation of CRISPR/Cas systems for L6-AK89 strongly supports our earlier reports [57,167,168], while some nucleotide substitutions that affect Cas folding were revealed in this strain. The localization and structure of CRISPR cassette and *cas* genes are quite conservative for *S. meliloti*, and their spacers showed identity to mostly abundant phages from the *Caudovirales* order. The attempt to analyze *rop* genes encoding porins, which are used by phages for cell adsorption [60], showed that strain L6-AK89 has significant structural changes in RopA1 and RopA2, and the cells of this strain are lysed by phages from the *Mioviridae* family, in contrast to Rm1021, and this may be due to a different density of interaction with this phage, according to [169]. There is additional role in tolerance to phages for tmRNA, together with the ClpXP/ClpAX complex, involved in regulating gene expression [170–172], both of which were detected in L6-AK89.

An additional replicon, pL6-AK89, accumulated in three inc-prophages related to phages of the *Siphoviridae* and *Podovirida* families, but also to Stx phages from the same order *Caudovirales*, while no sequences comprising the Shiga toxin operon were detected. This potentially conjugative replicon contained *traA*, the T4SS gene cluster, and type IV coupling protein, as well as a foreign IS element, *groESL* operon homologous to similar chromosomal *groESL*1 operon essential for symbiosis [139], *acdS-lrpL* module, and additional osmotic shock response gene clusters. The three additional tRNA sequences, which may be additional sites for site-specific integration of phage DNA, as was demonstrated for *Ensifer* (*Sinorhizobium*) spp. [131,132], were revealed on pL6-AK89 and both megaplasmids. pSymA has an additional homoserine lactone cluster that may be responsible for the regulation of quorum sensing genes [173–175]. Thus, both accessory plasmid pL6-AK89 and pSymA carry additional clusters of genes for osmotic shock response and quorum sensing, which can improve the viability of L6-AK89 with a salt-sensitive phenotype but was originally isolated from a region that suffered from salinity.

Besides mobilome elements, 53 *nod/noe/nol/nif/fdx/fix* genes and 28 well-known genes responsible for rhizobia fitness and stress tolerance were also part of our focus. About 70% of *sym* and *str* genes had significant non-synonymous nucleotide substitutions relative to the reference strain. All 28 genes in the L6-AK89 strain did not show significant upregulation of activity under 0.5 M NaCl stress of 1 and 12 h, while some tested genes had a constitutive level of expression, which could support the fitness of the salt-sensitive phenotype of L6-AK89 in a salinized environment. A high level of variability was found in the nucleotide sequences of *fix* genes, while *nif* genes encoding the nitrogenase complex are highly conserved, a finding consistent with the literature. Significant amino acid substitutions were revealed in products of regulatory and structural (common and species-specific) genes that together contribute to the specificity of L6-AK89 toward different varieties of diploid *M. lupulina*.

Finally, phylogenetic analysis using the consensus sequences of the 16S rRNA gene and the aligned concatenated sequences of genes involved in the response to stress factors showed that these sequences in the L6-AK89 strain are phylogenetically closer to the corresponding putative *E. meliloti* ancestral sequences (EmelA) and *E. medicae* ancestral sequences (EmedA) than similar sequences of the Rm1021 strain according to the calculated values of the relative rate of substitution accumulation (N = $3.45 \cdot 10^{-5}$/0.04053 and 1.1464/71.79 per 100 bp for L6-AK89 to EmelA and EmedA, respectively; N = $6.90 \cdot 10^{-5}$/0.04057 and

4.09017/74.73 per 100 bp for Rm1021 to EmelA and EmedA, respectively). The analysis of aligned concatenated sequences of *nod–nol–noe* genes and of the core genome showed that these sequences in L6-AK89 are phylogenetically more distant from the corresponding putative *E. meliloti* ancestral sequences than similar sequences of Rm1021, according to the calculated values of the relative accumulation rate of substitutions (N = 5.26/22.76 and 0.411/9.01 per 100 bp for L6-AK89 and N = 4.12/21.62 and 0.391/8.99 per 100 bp for Rm1021 to EmelA and EmedA, respectively).

This makes it possible to suggest that accessory parts of the genome related to bacterial mechanisms of withstanding abiotic stress appeared phylogenetically earlier in the genome of the L6-AK89 strain compared to the Rm1021 strain, regardless of the formation of the gene pool associated with the interaction of rhizobia and host plants, and the compact arrangement of blocks with symbiotic genes, as in the L6-AK89 genome, is a later phylogenetic event, which is probably related to its symbiotic preference for *Medicago lupulina* host plant.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy12040766/s1, Figure S1: Phylogenetic analysis of (A) consensus 16S rRNA gene nucleotide sequences and (B) concatenated *recA–atpD–glnII–gyrB–dnaJ* gene nucleotide sequences in *Ensifer* spp.; Figure S2: Localization of *acdS–lrpL* module on the graphic linear part of cryptic plasmid of *E. meliloti* L6-AK89; Figure S3: Lysis of L6-AK89 strain by bacteriophage no. 2; Table S1: IS elements and genes in head-to-tail orientation in genomes of strains L6-AK89 and Rm10211; Table S2: Presence and localization of genes involved in response to stress in genomes of strains L6-AK89 and Rm1021; Table S3: Phagotyping of *E. meliloti* strains.

**Author Contributions:** Conceptualization, M.L.R.; methodology, M.L.R., M.E.V., and V.S.M.; software, A.M.A.; formal analysis, M.E.V., A.S.S., V.S.M., A.P.K., and A.M.A.; investigation, M.L.R., V.S.M., and M.E.V.; data curation, M.L.R., M.E.V., A.S.S., V.S.M., A.P.K., A.M.A., O.A.B., and B.V.S.; writing—original draft preparation, M.L.R., M.E.V., A.S.S., V.S.M., and A.P.K.; writing—review and editing, M.L.R., M.E.V., A.S.S., V.S.M., and B.V.S.; project administration, M.L.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This paper was produced with support from the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement no. 075–15-2020–920, dated 16 November 2020, on providing grants in the form of subsidies from the federal budget of the Russian Federation. The grant was provided for state support for the creation and development of a World-class Scientific Center, "Agrotechnologies for the Future".

**Data Availability Statement:** The assemblies and sequence data have been uploaded to the NCBI.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Amer, N.; Al Chami, Z.; Al Bitar, L.; Mondelli, D.; Dumontet, S. Evaluation of *Atriplex halimus*, *Medicago lupulina* and *Portulaca oleracea* for phytoremediation of Ni, Pb, and Zn. *Int. J. Phytoremediation* **2013**, *15*, 498–512. [CrossRef] [PubMed]
2. Jakupović, L.; Kalvarešin, M.; Bukovina, K.; Poljak, V.; Vujić, L.; Zovko Končić, M. Optimization of two eco-friendly extractions of black medick (*Medicago lupulina* L.) phenols and their antioxidant, cosmeceutical, α-glucosidase and α-amylase inhibitory properties. *Molecules* **2021**, *26*, 1610. [CrossRef] [PubMed]
3. Butkutė, B.; Padarauskas, A.; Cesevičienė, J.; Pavilonis, A.; Taujenis, L.; Lemežienė, N. Perennial legumes as a source of ingredients for healthy food: Proximate, mineral and phytoestrogen composition and antibacterial activity. *J. Food Sci. Technol.* **2017**, *54*, 2661–2669. [CrossRef] [PubMed]
4. Vorsheva, A.V.; Stepanova, G.V. Estimation of the productivity and nutritionality of forage mass of new varieties of *Medicago lupulina*.. In Proceedings of the Plant Growing and Meadow Growing, Moscow, Russia, 18–19 November 2020; RSAU—MTAA: Moscow, Russia, 2020; pp. 628–631. (In Russian) [CrossRef]
5. Stepanova, G.V.; Vorsheva, A.V. Formation of bicarp populations of hop medic. In *Multifunctional Adaptive Fodder Production*, 25th ed.; Cherniavskih, V.I., Kosolapov, V.M., Shpakov, A.S., Kutuzova, A.A., Pobednov, Y.A., Kostenko, S.I., Georgiadi, N.I., Svechnikova, G.N., Eds.; FSBOU DPO RAKO APK: Moscow, Russia, 2021; pp. 9–20. (In Russian)

6. Bailly, X.; Giuntini, E.; Sexton, M.C.; Lower, R.P.; Harrison, P.W.; Kumar, N.; Young, J.P. Population genomics of *Sinorhizobium medicae* based on low-coverage sequencing of sympatric isolates. *ISME J.* **2011**, *5*, 1722–1734. [CrossRef]
7. Harrison, T.L.; Wood, C.W.; Heath, K.D.; Stinchcombe, J.R. Geographically structured genetic variation in the *Medicago lupulina-Ensifer* mutualism. *Evolution* **2017**, *71*, 1787–1801. [CrossRef]
8. Boyle, J.A.; Simonsen, A.K.; Frederickson, M.E.; Stinchcombe, J.R. Priority effects alter interaction outcomes in a legume-rhizobium mutualism. *Proc. Biol. Sci.* **2021**, *288*, 20202753. [CrossRef]
9. Martens, M.; Delaere, M.; Coopman, R.; De Vos, P.; Gillis, M.; Willems, A. Multilocus sequence analysis of *Ensifer* and related taxa. *Int. J. Syst. Evol. Microbiol.* **2007**, *57*, 489–503. [CrossRef]
10. Harrison, T.L.; Wood, C.W.; Borges, I.L.; Stinchcombe, J.R. No evidence for adaptation to local rhizobial mutualists in the legume *Medicago lupulina*. *Ecol. Evol.* **2017**, *7*, 4367–4376. [CrossRef]
11. Biondi, E.G.; Pilli, E.; Giuntini, E.; Roumiantseva, M.L.; Andronov, E.E.; Onichtchouk, O.P.; Kurchak, O.N.; Simarov, B.V.; Dzyubenko, N.I.; Mengoni, A.; et al. Genetic relationship of *Sinorhizobium meliloti* and *Sinorhizobium medicae* strains isolated from Caucasian region. *FEMS Microbiol. Lett.* **2003**, *220*, 207–213. [CrossRef]
12. Singh, R.; Interrante, S.M.; Butler, T.J.; Young, C.A. Characterization and effectiveness of co-inoculation of *Sinorhizobium* strains on annual medics. *Crop Sci.* **2012**, *52*, 932–942. [CrossRef]
13. Roberts, R.; Jackson, R.W.; Mauchline, T.H.; Hirsch, P.R.; Shaw, L.J.; Döring, T.F.; Jones, H.E. Is there sufficient *Ensifer* and *Rhizobium* species diversity in UK farmland soils to support red clover (*Trifolium pratense*), white clover (*T. repens*), lucerne (*Medicago sativa*) and black medic (*M. lupulina*)? *Appl. Soil Ecol.* **2017**, *120*, 35–43. [CrossRef]
14. Young, J.P.W.; Crossman, L.C.; Johnston, A.W.B.; Thomson, N.R.; Ghazoui, Z.F.; Hull, K.H.; Wexler, M.; Curson, A.R.J.; Todd, J.D.; Poole, P.S.; et al. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.* **2006**, *7*, R34. [CrossRef]
15. Tettelin, H.; Riley, D.; Cattuto, C.; Medini, D. Comparative genomics: The bacterial pan-genome. *Curr. Opin. Microbiol.* **2008**, *11*, 472–477. [CrossRef]
16. Segerman, B. The genetic integrity of bacterial species: The core genome and the accessory genome, two different stories. *Front. Cell. Infect. Microbiol.* **2012**, *2*, 116. [CrossRef]
17. Zhaxybayeva, O.; Lapierre, P.; Gogarten, J.P. Genome mosaicism and organismal lineages. *Trends Genet.* **2004**, *20*, 254–260. [CrossRef]
18. Sim, S.H.; Yu, Y.; Lin, C.H.; Karuturi, R.K.; Wuthiekanun, V.; Tuanyok, A.; Chua, H.H.; Ong, C.; Paramalingam, S.S.; Tan, G.; et al. The core and accessory genomes of *Burkholderia pseudomallei*: Implications for human melioidosis. *PLoS Pathog.* **2008**, *4*, e1000178. [CrossRef]
19. Mauchline, T.H.; Hayat, R.; Roberts, R.; Powers, S.J.; Hirsch, P.R. Assessment of core and accessory genetic variation in *Rhizobium leguminosarum* symbiovar trifolii strains from diverse locations and host plants using PCR-based methods. *Lett. Appl. Microbiol.* **2014**, *59*, 238–246. [CrossRef]
20. Jiao, J.; Ni, M.; Zhang, B.; Young, J.P.W.; Chan, T.-F. Coordinated regulation of core and accessory genes in the multipartite genome of *Sinorhizobium fredii*. *PLoS Genet.* **2018**, *14*, e1007428. [CrossRef]
21. Prozorov, A.A. Regularities of the location of genes having different functions and of some other nucleotide sequences in the bacterial chromosome. *Microbiology* **2007**, *76*, 383–392. (In Russian) [CrossRef]
22. Oliveira, P.H.; Touchon, M.; Cury, J.; Rocha, E.P.C. The chromosomal organization of horizontal gene transfer in bacteria. *Nat. Commun.* **2017**, *8*, 841. [CrossRef]
23. Kivisaar, M. Mutation and recombination rates vary across bacterial chromosome. *Microorganisms* **2019**, *8*, 25. [CrossRef] [PubMed]
24. Sousa, C.; de Lorenzo, V.; Cebolla, A. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology* **1997**, *143*, 2071–2078. [CrossRef] [PubMed]
25. Block, D.H.; Hussein, R.; Liang, L.W.; Lim, H.N. Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res.* **2012**, *40*, 8979–8992. [CrossRef] [PubMed]
26. Kopejtka, K.; Lin, Y.; Jakubovicova, M.; Koblizek, M.; Tomasch, J. Clustered core- and pan-genome content on *Rhodobacteraceae* chromosomes. *Genome Biol. Evol.* **2019**, *11*, 2208–2217. [CrossRef]
27. Mira, A.; Ochman, H. Gene location and bacterial sequence divergence. *Mol. Biol. Evol.* **2002**, *19*, 1350–1358. [CrossRef]
28. Juurik, T.; Ilves, H.; Teras, R.; Ilmjarv, T.; Tavita, K.; Ukkivi, K.; Kivisaar, M. Mutation frequency and spectrum of mutations vary at different chromosomal positions of *Pseudomonas putida*. *PLoS ONE* **2012**, *7*, e48511. [CrossRef]
29. Lobry, J.R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **1996**, *13*, 660–665. [CrossRef]
30. Salzberg, S.L.; Salzberg, A.J.; Kerlavage, A.R.; Tomb, J.F. Skewed oligomers and origins of replication. *Gene* **1998**, *217*, 57–67. [CrossRef]
31. Lobry, J.R.; Louarn, J.M. Polarisation of prokaryotic chromosomes. *Curr. Opin. Microbiol.* **2003**, *6*, 101–108. [CrossRef]
32. Song, J.; Ware, A.; Liu, S.L. Wavelet to predict bacterial *ori* and *ter*: A tendency towards a physical balance. *BMC Genom.* **2003**, *4*, 17. [CrossRef]
33. Lesterlin, C.; Pages, C.; Dubarry, N.; Dasgupta, S.; Cornet, F. Asymmetry of chromosome replichores renders the DNA translocase activity of FtsK essential for cell division and cell shape maintenance in *Escherichia coli*. *PLoS Genet.* **2008**, *4*, e1000288. [CrossRef]

34. Kono, N.; Arakawa, K.; Sato, M.; Yoshikawa, H.; Tomita, M.; Itaya, M. Undesigned selection for replication termination of bacterial chromosomes. *J. Mol. Biol.* **2014**, *426*, 2918–2927. [CrossRef]
35. Galardini, M.; Pini, F.; Bazzicalupo, M.; Biondi, E.G.; Mengoni, A. Replicon-dependent bacterial genome evolution: The case of *Sinorhizobium meliloti*. *Genome Biol. Evol.* **2013**, *5*, 542–558. [CrossRef]
36. Lagares, A.; Sanjuán, J.; Pistorio, M. The plasmid mobilome of the model plant-symbiont *Sinorhizobium meliloti*: Coming up with new questions and answers. In *Plasmids*; Tolmasky, M.E., Alonso, J.C., Eds.; ASM Press: Washington, DC, USA, 2015; pp. 277–293. [CrossRef]
37. Oresnik, I.J.; Liu, S.L.; Yost, C.K.; Hynes, M.F. Megaplasmid pRme2011a of *Sinorhizobium meliloti* is not required for viability. *J. Bacteriol.* **2000**, *182*, 3582–3586. [CrossRef]
38. Guo, X.; Flores, M.; Mavingui, P.; Fuentes, S.I.; Hernández, G.; Dávila, G.; Palacios, R. Natural genomic design in *Sinorhizobium meliloti*: Novel genomic architectures. *Genome Res.* **2003**, *13*, 1810–1817. [CrossRef]
39. Pistorio, M.; Giusti, M.A.; Del Papa, M.F.; Draghi, W.O.; Lozano, M.J.; Tejerizo, G.T.; Lagares, A. Conjugal properties of the *Sinorhizobium meliloti* plasmid mobilome. *FEMS Microbiol. Ecol.* **2008**, *65*, 372–382. [CrossRef]
40. Nogales, J.; Blanca-Ordóñez, H.; Olivares, J.; Sanjuán, J. Conjugal transfer of the *Sinorhizobium meliloti* 1021 symbiotic plasmid is governed through the concerted action of one- and two-component signal transduction regulators. *Environ. Microbiol.* **2013**, *15*, 811–821. [CrossRef]
41. Velázquez, E.; Mateos, P.F.; Pedrero, P.; Dazzo, F.B.; Martinez-Molina, E. Attenuation of symbiotic effectiveness by *Rhizobium meliloti* SAF22 related to the presence of a cryptic plasmid. *Appl. Environ. Microbiol.* **1995**, *61*, 2033–2036. [CrossRef]
42. Barran, L.R.; Ritchot, N.; Bromfield, E.S. *Sinorhizobium meliloti* plasmid pRm1132f replicates by a rolling-circle mechanism. *J. Bacteriol.* **2001**, *183*, 2704–2708. [CrossRef]
43. Roumiantseva, M.L.; Onishchuk, O.P.; Belova, V.S.; Kurchak, O.N.; Simarov, B.V. Polymorphism of *Sinorhizobium meliloti* strains isolated from diversity centers of alfalfa in various soil and climatic conditions, *Russ. J. Genet. Appl. Res.* **2011**, *1*, 97–102. [CrossRef]
44. Saksaganskaia, A.; Vladimirova, M.; Afonin, A.; Simarov, B.; Roumiantseva, M. Cryptic plasmids essential for *Sinorhizobium meliloti* fitness. In *Proceedings of the 20th International Multidisciplinary Scientific GeoConference SGEM 2020, Varna, Bulgaria, 18–24 August 2020*; STEF92 Technology Ltd.: Sofia, Bulgaria, 2020; pp. 223–230. [CrossRef]
45. Saksaganskaia, A.; Muntyan, V.; Muntyan, A.; Afonin, A.; Simarov, B.; Roumiantseva, M. Are cryptic plasmids of *Sinorhizobium meliloti* attractive within phage infection? *FEBS Open Bio* **2021**, *11*, 133. [CrossRef]
46. Stiens, M.; Schneiker, S.; Keller, M.; Kuhn, S.; Pühler, A.; Schlüter, A. Sequence analysis of the 144-kilobase accessory plasmid pSmeSM11a, isolated from a dominant *Sinorhizobium meliloti* strain identified during a long-term field release experiment. *Appl. Environ. Microbiol.* **2006**, *72*, 3662–3672. [CrossRef] [PubMed]
47. Stiens, M.; Schneiker, S.; Puhler, A.; Schlüter, A. Sequence analysis of the 181-kb accessory plasmid pSmeSM11b, isolated from a dominant *Sinorhizobium meliloti* strain identified during a long-term field release experiment. *FEMS Microbiol. Lett.* **2007**, *271*, 297–309. [CrossRef] [PubMed]
48. López, J.L.; Lozano, M.J.; Lagares, A., Jr.; Fabre, M.L.; Draghi, W.O.; Del Papa, M.F.; Pistorio, M.; Becker, A.; Wibberg, D.; Schlüter, A.; et al. Codon usage heterogeneity in the multipartite prokaryote genome: Selection-based coding bias associated with gene location, expression level, and ancestry. *MBio* **2019**, *10*, e00505-19. [CrossRef]
49. Dobrindt, U.; Hochhut, B.; Hentschel, U.; Hacker, J. Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* **2004**, *2*, 414–424. [CrossRef]
50. Frost, L.S.; Leplae, R.; Summers, A.; Toussaint, A. Mobile genetic elements: The agents of open source evolution. *Nat. Rev. Microbiol.* **2005**, *3*, 722–732. [CrossRef]
51. Stokes, H.W.; Gillings, M.R. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiol. Rev.* **2011**, *35*, 790–819. [CrossRef]
52. Jørgensen, T.S.; Kiil, A.S.; Hansen, M.A.; Sørensen, S.J.; Hansen, L.H. Current strategies for mobilome research. *Front. Microbiol.* **2015**, *5*, 750. [CrossRef]
53. Rodríguez-Beltrán, J.; Sørum, V.; Toll-Riera, M.; de la Vega, C.; Peña-Miller, R.; San Millán, Á. Genetic dominance governs the evolution and spread of mobile genetic elements in bacteria. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 15755–15762. [CrossRef]
54. Di Cenzo, G.C.; Checcucci, A.; Bazzicalupo, M.; Mengoni, A.; Viti, C.; Dziewit, L.; Finan, T.M.; Galardini, M.; Fondi, M. Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat. Commun.* **2016**, *7*, 12219. [CrossRef]
55. Poole, P.; Ramachandran, V.; Terpolilli, J. Rhizobia: From saprophytes to endosymbionts. *Nat. Rev. Microbiol.* **2018**, *16*, 291–303. [CrossRef]
56. Bernheim, A.; Sorek, R. The pan-immune system of bacteria: Antiviral defence as a community resource. *Nat. Rev. Microbiol.* **2020**, *18*, 113–119. [CrossRef]
57. Vladimirova, M.; Muntyan, V.; Kozlova, A.; Grudinin, M.; Roumiantseva, M. CRISPR/Cas in genomes of *Sinorhizobium meliloti*. In *Proceedings of the 20th International Multidisciplinary Scientific GeoConference SGEM 2020, Varna, Bulgaria, 18–24 August 2020*; STEF92 Technology Ltd.: Sofia, Bulgaria, 2020; pp. 215–222. [CrossRef]
58. Zheng, Y.; Li, J.; Wang, B.; Han, J.; Hao, Y.; Wang, S.; Ma, X.; Yang, S.; Ma, L.; Yi, L.; et al. Endogenous type I CRISPR-Cas: From foreign DNA defense to prokaryotic engineering. *Front. Bioeng. Biotechnol.* **2020**, *8*, 62. [CrossRef]

59. Pinilla-Redondo, R.; Russel, J.; Mayo-Muñoz, D.; Shah, S.A.; Garrett, R.A.; Nesme, J.; Madsen, J.S.; Fineran, P.C.; Sørensen, S.J. CRISPR-Cas systems are widespread accessory elements across bacterial and archaeal plasmids. *Nucleic Acids Res.* **2021**, *1*, gkab859. [CrossRef]
60. Crook, M.B.; Draper, A.L.; Guillory, R.J.; Griffitts, J.S. The *Sinorhizobium meliloti* essential porin RopA1 is a target for numerous bacteriophages. *J. Bacteriol.* **2013**, *95*, 3663–3671. [CrossRef]
61. Rüberg, S.; Tian, Z.X.; Krol, E.; Linke, B.; Meyer, F.; Wang, Y. Construction and validation of a *Sinorhizobium meliloti* whole genome DNA microarray: Genome-wide profiling of osmoadaptive gene expression. *J. Biotechnol.* **2003**, *106*, 255–268. [CrossRef]
62. Domínguez-Ferreras, A.; Pérez-Arnedo, R.; Becker, A.; Olivares, J.; Soto, M.J.; Sanjuán, J. Transcriptome profiling reveals the importance of plasmid pSymB for osmoadaptation of *Sinorhizobium meliloti*. *J. Bacteriol.* **2006**, *188*, 7617–7625. [CrossRef]
63. Hellweg, C.; Pühler, A.; Weidner, S. The time course of the transcriptomic response of *Sinorhizobium meliloti* 1021 following a shift to acidic pH. *BMC Microbiol.* **2009**, *15*, 37. [CrossRef]
64. Tiwari, R.P.; Reeve, W.G.; Dilworth, M.J.; Glenn, A.R. Acid tolerance in *Rhizobium meliloti* strain WSM419 involves a two-component sensor-regulator system. *Microbiology* **1996**, *142*, 1693–1704. [CrossRef]
65. Domínguez-Ferreras, A.; Muñoz, S.; Olivares, J.; Soto, M.J.; Sanjuán, J. Role of potassium uptake systems in *Sinorhizobium meliloti* osmoadaptation and symbiotic performance. *J. Bacteriol.* **2009**, *191*, 2133–2143. [CrossRef]
66. Domínguez-Ferreras, A.; Soto, M.J.; Pérez-Arnedo, R.; Olivares, J.; Sanjuán, J. Importance of trehalose biosynthesis for *Sinorhizobium meliloti* osmotolerance and nodulation of alfalfa roots. *J. Bacteriol.* **2009**, *191*, 7490–7499. [CrossRef] [PubMed]
67. Wang, C.; Kemp, J.; Da Fonseca, I.O.; Equi, R.C.; Sheng, X.; Charles, T.C.; Sobral, B.W.S. *Sinorhizobium meliloti* 1021 loss-of-function deletion mutation in *chvI* and its phenotypic characteristics. *Mol. Plant Microbe Interact.* **2010**, *23*, 153–160. [CrossRef] [PubMed]
68. Da-Silva, J.R.; Alexandre, A.; Brígido, C.; Oliveira, S. Can stress response genes be used to improve the symbiotic performance of rhizobia? *AIMS Microbiol.* **2017**, *3*, 365–382. [CrossRef] [PubMed]
69. Albicoro, F.J.; Draghi, W.; Martini, M.C.; Salas, M.E.; Torres Tejerizo, G.T.; Lozano, M.; Lopez, J.L.; Vacca, C.; Cafiero, J.H.; Pistorio, M.; et al. The two-component system ActJK is involved in acid stress tolerance and symbiosis in *Sinorhizobium meliloti*. *J. Biotechnol.* **2021**, *329*, 80–91. [CrossRef] [PubMed]
70. Rutten, P.J.; Poole, P.S. Oxygen regulatory mechanisms of nitrogen fixation in rhizobia. *Adv. Microb. Physiol.* **2019**, *75*, 325–389. [CrossRef] [PubMed]
71. Lindström, K.; Mousavi, S.A. Effectiveness of nitrogen fixation in rhizobia. *Microb. Biotechnol.* **2020**, *13*, 1314–1335. [CrossRef]
72. Barnett, M.J.; Long, S.R. The *Sinorhizobium meliloti* SyrM regulon: Effects on global gene expression are mediated by *syrA* and *nodD3*. *J. Bacteriol.* **2015**, *197*, 1792–1806. [CrossRef]
73. Debellé, F.; Moulin, L.; Mangin, B.; Dénarié, J.; Boivin, C. Nod genes and Nod signals and the evolution of the *Rhizobium* legume symbiosis. *Acta Biochim. Pol.* **2001**, *48*, 359–365. [CrossRef]
74. Aoki, S.; Ito, M.; Iwasaki, W. From β- to α-proteobacteria: The origin and evolution of rhizobial nodulation genes *nodIJ*. *Mol. Biol. Evol.* **2013**, *30*, 2494–2508. [CrossRef]
75. Geddes, B.A.; Kearsley, J.; Morton, R.; diCenzo, G.C.; Finan, T.M. The Genomes of Rhizobia. In *Advances in Botanical Research*; Elsevier: Amsterdam, The Netherlands, 2020; Volume 94, pp. 213–249. ISBN 978-0-08-102798-1.
76. Ardourel, M.; Lortet, G.; Maillet, F.; Roche, P.; Truchet, G.; Promé, J.C.; Rosenberg, C. In *Rhizobium meliloti*, the operon associated with the *nod* box n5 comprises *nodL*, *noeA* and *noeB*, three host-range genes specifically required for the nodulation of particular *Medicago* species. *Mol. Microbiol.* **1995**, *17*, 687–699. [CrossRef]
77. Bena, G. Molecular phylogeny supports the morphologically based taxonomic transfer of the "medicagoid" *Trigonella* species to the genus *Medicago* L. *Plant Syst. Evol.* **2001**, *229*, 217–236. [CrossRef]
78. Roumiantseva, M.L.; Vladimirova, M.E.; Muntyan, V.S.; Stepanova, G.V.; Saksaganskaya, A.S.; Kozhemyakov, A.P.; Orlova, A.G.; Becker, A.; Simarov, B.V. Highly effective root nodule inoculants of alfalfa (*Medicago varia* L.): Molecular-genetic analysis and practical usage in cultivar creation. *Agric. Biol.* **2019**, *54*, 1306–1323. [CrossRef]
79. Reeve, W.; Ardley, J.; Tian, R.; Eshragi, L.; Yoon, J.W.; Ngamwisetkun, P.; Seshadri, R.; Ivanova, N.N.; Kyrpides, N.C. A Genomic Encyclopedia of the Root Nodule Bacteria: Assessing genetic diversity through a systematic biogeographic survey. *Stand. Genom. Sci.* **2015**, *10*, 14. [CrossRef] [PubMed]
80. Ibragimova, M.V.; Rumyantseva, M.L.; Onishchuk, O.P.; Belova, V.S.; Kurchak, O.N.; Andronov, E.E.; Dziubenko, N.I.; Simarov, B.V. Symbiosis between the root-nodule bacterium *Sinorhizobium meliloti* and alfalfa (*Medicago sativa*) under salinization conditions. *Microbiology* **2006**, *75*, 77–81. [CrossRef]
81. Greene, S.L.; Kisha, T.J.; Dzyubenko, N.I. Conserving alfalfa wild relatives: Is past introgression with Russian varieties evident today? *Crop Sci.* **2008**, *48*, 1853–1864. [CrossRef]
82. Roumiantseva, M.L.; Muntyan, V.S. Root nodule bacteria *Sinorhizobium meliloti*: Tolerance to salinity and bacterial genetic determinants. *Microbiology* **2015**, *84*, 303–318. [CrossRef]
83. Vincent, J.M. Serological studies of the root-nodule bacteria. I. Strains of *Rhizobium meliloti*. *Proc. Linn. Soc. N.S.W.* **1941**, *66*, 145–154.
84. Beringer, J.E. R factor transfer in *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **1974**, *84*, 188–198. [CrossRef]
85. Clokie, M.R.J.; Kropinski, A.M. (Eds.) *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions*, 1st ed.; Humana Press: New York, NY, USA, 2009; pp. 69–77. [CrossRef]

86. Rumyantseva, M.L.; Simarov, B.V.; Onishchuk, O.P.; Andronov, E.E.; Chizhevskaya, E.P.; Belova, V.S.; Kurchak, O.N.; Muntyan, A.N.; Rumyantseva, T.B.; Zatovskaya, T.V. *Biodiversity of Rhizobia in Ecosystems and Agrocenoses: Theoretical Bases and Methods*, 1st ed.; ICZR: St. Petersburg, Russia, 2011; pp. 21–85. (In Russian)
87. Smurygin, M.A.; Novoselova, A.S.; Konstantinova, A.M.; Rubcov, M.I.; Filimonov, M.A.; Bekhtin, N.S.; Voshchinin, P.A.; Ezhakova, O.F.; Katkov, V.A.; Kuleshov, G.F.; et al. *Guidelines for the Selection of Perennial Grasses*; FWRC FPA: Moscow, Russia, 1985; pp. 98–99. (In Russian)
88. Roumiantseva, M.L.; Stepanova, G.V.; Kurchak, O.N.; Onishchuk, O.P.; Muntyan, V.S.; Dzyubenko, E.A.; Dzyubenko, N.I.; Simarov, B.V. Selection of salt tolerant alfalfa (*Medicago* L.) plants from different varieties and their morfo biological and symbiotic properties analysis. *Agric. Biol.* **2015**, *50*, 673–684. [CrossRef]
89. Provorov, N.A.; Onishchuk, O.P.; Kurchak, O.N. Impacts of inoculation with *Sinorhizobium meliloti* strains differing in salt tolerance on the productivity and habitus of alfalfa (*Medicago sativa* L.). *Agric. Biol.* **2016**, *51*, 343–350. [CrossRef]
90. Lakin, G.F. *Biometrics*, 4th ed.; Vysshaya Shkola: Moscow, Russia, 1990; pp. 159–179. (In Russian)
91. Green, M.R.; Sambrook, J. Isolation of high-molecular-weight DNA using organic solvents. *Cold Spring Harb. Protoc.* **2017**, *3*, pdb-prot093450. [CrossRef]
92. Afonin, A.M.; Gribchenko, E.S.; Sulima, A.S.; Zhukov, V.A. Complete genome sequence of an efficient *Rhizobium leguminosarum* bv. viciae strain, A1. *Microbiol. Resour. Announc.* **2020**, *9*, e00249-20. [CrossRef]
93. Wick, R.R.; Judd, L.M.; Holt, K.E. Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLoS Comput. Biol.* **2018**, *14*, e1006583. [CrossRef]
94. Porechop. Available online: https://github.com/rrwick/Porechop (accessed on 20 October 2021).
95. Filtlong. Available online: https://github.com/rrwick/Filtlong (accessed on 20 October 2021).
96. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [CrossRef]
97. Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **2017**, *27*, 737–746. [CrossRef]
98. Oxford Nanopore Technologies. Medaka. Available online: https://github.com/nanoporetech/medaka (accessed on 20 October 2021).
99. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef]
100. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **2016**, *44*, 6614–6624. [CrossRef]
101. Haft, D.H.; DiCuccio, M.; Badretdin, A.; Brover, V.; Chetvernin, V.; O'Neill, K.; Li, W.; Chitsaz, F.; Derbyshire, M.K.; Gonzales, N.R.; et al. RefSeq: An update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* **2018**, *46*, D851–D860. [CrossRef]
102. Li, W.; O'Neill, K.R.; Haft, D.H.; DiCuccio, M.; Chetvernin, V.; Badretdin, A.; Coulouris, G.; Chitsaz, F.; Derbyshire, M.K.; Durkin, A.S.; et al. RefSeq: Expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Res.* **2021**, *49*, D1020–D1028. [CrossRef]
103. Zhou, Y.; Liang, Y.; Lynch, K.H.; Dennis, J.J.; Wishart, D.S. PHAST: A fast phage search tool. *Nucleic Acids Res.* **2011**, *39*, W347–W352. [CrossRef]
104. Arndt, D.; Grant, J.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.; Wishart, D.S. PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21. [CrossRef] [PubMed]
105. Hudson, C.M.; Lau, B.Y.; Williams, K.P. Islander: A database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Res.* **2015**, *43*, D48–D53. [CrossRef] [PubMed]
106. Bertelli, C.; Laird, M.R.; Williams, K.P.; Lau, B.Y.; Hoad, G.; Winsor, G.L.; Brinkman, F.S.L. IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* **2017**, *45*, W30–W35. [CrossRef] [PubMed]
107. Stothard, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **2000**, *28*, 1102–1104. [CrossRef]
108. Couvin, D.; Bernheim, A.; Toffano-Nioche, C.; Touchon, M.; Michalik, J.; Néron, B.; Rocha, E.P.C.; Vergnaud, G.; Gautheret, D.; Pourcel, C. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **2018**, *46*, W246–W251. [CrossRef]
109. Lowe, T.M.; Chan, P.P. tRNAscan-SE on-line: Search and contextual analysis of transfer RNA genes. *Nucleic Acids Res.* **2016**, *44*, W54–W57. [CrossRef]
110. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* **2019**, *1962*, 1–14. [CrossRef]
111. Puterová, J.; Martínek, T. digIS: Towards detecting distant and putative novel insertion sequence elements in prokaryotic genomes. *BMC Bioinform.* **2021**, *22*, 258. [CrossRef]
112. Siguier, P.; Perochon, J.; Lestrade, L.; Mahillon, J.; Chandler, M. ISfinder: The reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **2006**, *34*, D32–D36. [CrossRef]
113. Li, X.; Xie, Y.; Liu, M.; Tai, C.; Sun, J.; Deng, Z.; Ou, H.Y. oriTfinder: A web-based tool for the identification of origin of transfers in DNA sequences of bacterial mobile genetic elements. *Nucleic Acids Res.* **2018**, *46*, W229–W234. [CrossRef]

114. Darling, A.C.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [CrossRef] [PubMed]
115. Kanehisa, M.; Sato, Y.; Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **2016**, *428*, 726–731. [CrossRef] [PubMed]
116. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef]
117. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [CrossRef] [PubMed]
118. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef] [PubMed]
119. De Castro, E.; Sigrist, C.J.A.; Gattiker, A.; Bulliard, V.; Langendijk-Genevaux, P.S.; Gasteiger, E.; Bairoch, A.; Hulo, N. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **2006**, *34*, W362–W365. [CrossRef] [PubMed]
120. Yang, J.; Zhang, Y. I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Res.* **2015**, *43*, W174–W181. [CrossRef] [PubMed]
121. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S.Y.; Medema, M.H.; Weber, T. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019**, *47*, W81–W87. [CrossRef]
122. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef]
123. Huson, D.H.; Scornavacca, C. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **2012**, *61*, 1061–1067. [CrossRef]
124. Bláha, L. Importance of root—shoot ratio for crops production. *J. Agron. Agri. Sci.* **2019**, *2*, 12. [CrossRef]
125. Allito, B.B.; Ewusi-Mensah, N.; Logah, V.; Hunegnaw, D.K. Legume-rhizobium specificity effect on nodulation, biomass production and partitioning of faba bean (*Vicia faba* L.). *Sci. Rep.* **2021**, *11*, 3678. [CrossRef] [PubMed]
126. Lawrence, J.G.; Hendrickson, H. Lateral gene transfer: When will adolescence end? *Mol. Microbiol.* **2003**, *50*, 739–749. [CrossRef] [PubMed]
127. Young, J.P.W.; Wexler, M. Sym plasmid and chromosomal genotypes are correlated in field populations of *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **1988**, *134*, 2731–2739. [CrossRef]
128. Zhang, Y.M.; Tian, C.F.; Sui, X.H. Robust markers reflecting phylogeny and taxonomy of rhizobia. *PLoS ONE* **2012**, *7*, e44936. [CrossRef]
129. Guo, H.J.; Wang, E.T.; Zhang, X.X.; Li, Q.Q.; Zhang, Y.M.; Tian, C.F.; Chen, W.X. Replicon-dependent differentiation of symbiosis-related genes in *Sinorhizobium* strains nodulating *Glycine max*. *Appl. Environ. Microbiol.* **2014**, *80*, 1245–1255. [CrossRef]
130. Stefan, A.; Van Cauwenberghe, J.; Rosu, C.M.; Stedel, C.; Labrou, N.E.; Flemetakis, E.; Efrose, R.C. Genetic diversity and structure of *Rhizobium leguminosarum* populations associated with clover plants are influenced by local environmental variables. *Syst. Appl. Microbiol.* **2018**, *41*, 251–259. [CrossRef]
131. Cherkasova, M.E.; Muntyan, V.S.; Saksaganskaia, A.S.; Simarov, B.V.; Roumiantseva, M.L. *Sinorhizobium meliloti*: Chromosomal types and genomic islands. *Ecol. Genet.* **2019**, *17*, 23–38. [CrossRef]
132. Roumiantseva, M.L.; Muntyan, V.S.; Cherkasova, M.E.; Andronov, E.E.; Saksaganskaya, A.S.; Dzyubenko, E.A.; Dzyubenko, N.I.; Simarov, B.V. A comparative analysis of genomic characters of reference *Sinorhizobium meliloti* strains, the alfalfa symbionts (review). *Agric. Biol.* **2017**, *52*, 928–939. [CrossRef]
133. Wang, B.; Guo, F.; Huang, C.; Zhao, H. Unraveling the iterative type I polyketide synthases hidden in *Streptomyces*. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 8449–8454. [CrossRef]
134. Strieker, M.; Tanović, A.; Marahiel, M.A. Nonribosomal peptide synthetases: Structures and dynamics. *Curr. Opin. Struct. Biol.* **2010**, *20*, 234–240. [CrossRef] [PubMed]
135. Lynch, D.; O'Brien, J.; Welch, T.; Clarke, P.; Cuív, P.O.; Crosa, J.H.; O'Connell, M. Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. *J. Bacteriol.* **2001**, *8*, 2576–2585. [CrossRef] [PubMed]
136. Sagot, B.; Gaysinski, M.; Mehiri, M.; Guigonis, J.M.; Le Rudulier, D.; Alloing, G. Osmotically induced synthesis of the dipeptide N-acetylglutaminylglutamine amide is mediated by a new pathway conserved among bacteria. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 12652–12657. [CrossRef] [PubMed]
137. Rivera-Araya, J.; Pollender, A.; Huynh, D.; Schlömann, M.; Chávez, R.; Levicán, G. Osmotic imbalance, cytoplasm acidification and oxidative stress induction support the high toxicity of chloride in acidophilic bacteria. *Front. Microbiol* **2019**, *10*, 2455. [CrossRef] [PubMed]
138. Wang, L.; Jiang, S.; Deng, Z.; Dedon, P.C.; Chen, S. DNA phosphorothioate modification-a new multi-functional epigenetic system in bacteria. *FEMS Microbiol. Rev.* **2019**, *43*, 109–122. [CrossRef]
139. Bittner, A.N.; Foltz, A.; Oke, V. Only one of five *groEL* genes is required for viability and successful symbiosis in *Sinorhizobium meliloti*. *J. Bacteriol.* **2007**, *189*, 1884–1889. [CrossRef]
140. Schneiker-Bekel, S. The complete genome sequence of the dominant *Sinorhizobium meliloti* field isolate SM11 extends the *S. meliloti* pan-genome. *J. Biotechnol.* **2011**, *155*, 20–33. [CrossRef]

141. Kuhn, S.; Stiens, M.; Pühler, A.; Schlüter, A. Prevalence of pSmeSM11a-like plasmids in indigenous *Sinorhizobium meliloti* strains isolated in the course of a field release experiment with genetically modified *S. meliloti* strains. *FEMS Microbiol. Ecol.* **2008**, *63*, 118–131. [CrossRef]
142. Garcia-Vallve, S.; Guzman, E.; Montero, M.A.; Romeu, A. HGT-DB: A database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic Acids Res.* **2003**, *31*, 187–189. [CrossRef]
143. Tuller, T. Codon bias, tRNA pools, and horizontal gene transfer. *Mob. Genet. Elements* **2011**, *1*, 75–77. [CrossRef]
144. Bedhomme, S.; Amorós-Moya, D.; Valero, L.M.; Bonifaci, N.; Pujana, M.À.; Bravo, I.G. Evolutionary changes after translational challenges imposed by horizontal gene transfer. *Genome Biol. Evol.* **2019**, *11*, 814–831. [CrossRef]
145. Selbitschka, W.; Arnold, W.; Jording, D.; Kosier, B.; Toro, N.; Pühler, A. The insertion sequence element ISRm2011-2 belongs to the IS630-Tc1 family of transposable elements and is abundant in *Rhizobium meliloti*. *Gene* **1995**, *163*, 59–64. [CrossRef]
146. Roumiantseva, M.L.; Andronov, E.E.; Sharypova, L.A.; Dammann-Kalinowski, T.; Keller, M.; Young, J.P.; Simarov, B.V. Diversity of *Sinorhizobium meliloti* from the Central Asian alfalfa gene center. *Appl. Environ. Microbiol.* **2002**, *68*, 4694–4697. [CrossRef]
147. Biondi, E.G.; Toro, N.; Bazzicalupo, M.; Martínez-Abarca, F. Spread of the group II intron RmInt1 and its insertion sequence target sites in the plant endosymbiont *Sinorhizobium meliloti*. *Mob. Genet. Elements* **2011**, *1*, 2–7. [CrossRef]
148. Toro, N.; Martínez-Abarca, F.; Molina-Sánchez, M.D.; García-Rodríguez, F.M.; Nisa-Martínez, R. Contribution of mobile group II introns to *Sinorhizobium meliloti* genome evolution. *Front. Microbiol.* **2018**, *9*, 627. [CrossRef]
149. Schlüter, J.P.; Reinkensmeier, J.; Daschkey, S.; Evguenieva-Hackenberg, E.; Janssen, S.; Jänicke, S.; Becker, J.D.; Giegerich, R.; Becker, A. A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium *Sinorhizobium meliloti*. *BMC Genom.* **2010**, *11*, 245. [CrossRef]
150. Hernandez-Lucas, I.; Ramirez-Trujillo, J.A.; Gaitan, M.A.; Guo, X.; Flores, M.; Martinez-Romero, E.; Perez-Rueda, E.; Mavingui, P. Isolation and characterization of functional insertion sequences of rhizobia. *FEMS Microbiol. Lett.* **2006**, *261*, 25–31. [CrossRef]
151. Mandon, K.; Osterås, M.; Boncompagni, E.; Trinchant, J.C.; Spennato, G.; Poggi, M.C.; Le Rudulier, D. The *Sinorhizobium meliloti* glycine betaine biosynthetic genes (*betlCBA*) are induced by choline and highly expressed in bacteroids. *Mol. Plant Microbe Interact.* **2003**, *16*, 709–719. [CrossRef]
152. Wang, D.; Xue, H.; Wang, Y.; Yin, R.; Xie, F.; Luo, L. The *Sinorhizobium meliloti ntrX* gene is involved in succinoglycan production, motility, and symbiotic nodulation on alfalfa. *Appl. Environ. Microbiol.* **2013**, *79*, 7150–7159. [CrossRef]
153. Siarot, L.; Toyazaki, H.; Hidaka, M.; Kurumisawa, K.; Hirakawa, T.; Morohashi, K.; Aono, T. A Novel regulatory pathway for $K^+$ uptake in the legume symbiont *Azorhizobium caulinodans* in which TrkJ represses the *kdpFABC* operon at high extracellular $K^+$ concentrations. *Appl. Environ. Microbiol.* **2017**, *83*, e01197-17. [CrossRef]
154. van Loo, B.; Schober, M.; Valkov, E.; Heberlein, M.; Bornberg-Bauer, E.; Faber, K.; Hyvönen, M.; Hollfelder, F. Structural and mechanistic analysis of the choline sulfatase from *Sinorhizobium melliloti*: A class I sulfatase specific for an alkyl sulfate ester. *J. Mol. Biol.* **2018**, *430*, 1004–1023. [CrossRef]
155. Sharma, M.P.; Grover, M.; Chourasiya, D.; Bharti, A.; Agnihotri, R.; Maheshwari, H.S.; Pareek, A.; Buyer, J.S.; Sharma, S.K.; Schütz, L.; et al. Deciphering the role of trehalose in tripartite symbiosis among rhizobia, arbuscular mycorrhizal fungi, and legumes for enhancing abiotic stress tolerance in crop plants. *Front. Microbiol.* **2020**, *11*, 509919. [CrossRef] [PubMed]
156. Vanaporn, M.; Titball, R.W. Trehalose and bacterial virulence. *Virulence* **2020**, *11*, 1192–1202. [CrossRef] [PubMed]
157. Mom, R.; Muries, B.; Benoit, P.; Robert-Paganin, J.; Réty, S.; Venisse, J.S.; Padua, A.; Label, P.; Auguin, D. Voltage-gating of aquaporins, a putative conserved safety mechanism during ionic stresses. *FEBS Lett.* **2021**, *595*, 41–57. [CrossRef] [PubMed]
158. Kosolapova, A.O.; Belousov, M.V.; Sulatskaya, A.I.; Belousova, M.E.; Sulatsky, M.I.; Antonets, K.S.; Volkov, K.V.; Lykholay, A.N.; Shtark, O.Y.; Vasilieva, E.N.; et al. Two novel amyloid proteins, RopA and RopB, from the root nodule bacterium *Rhizobium leguminosarum*. *Biomolecules* **2019**, *9*, 694. [CrossRef]
159. Boussau, B.; Karlberg, E.O.; Frank, A.C.; Legault, B.A.; Andersson, S.G. Computational inference of scenarios for alpha-proteobacterial genome evolution. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9722–9727. [CrossRef]
160. Paço, A.; Brígido, C.; Alexandre, A.; Mateos, P.F.; Oliveira, S. The symbiotic performance of chickpea rhizobia can be improved by additional copies of the *clpB* chaperone gene. *PLoS ONE* **2016**, *11*, e0148221. [CrossRef]
161. Kopat, V.V.; Chirak, E.R.; Kimeklis, A.K.; Safronova, V.I.; Belimov, A.A.; Kabilov, M.R.; Andronov, E.E.; Provorov, N.A. Evolution of *fixNOQP* genes encoding cytochrome oxidase with high affinity to oxygen in rhizobia and related bacteria. *Russ. J. Genet.* **2017**, *53*, 766–774. [CrossRef]
162. Cren, M.; Kondorosi, A.; Kondorosi, E. NolR controls expression of the *Rhizobium meliloti* nodulation genes involved in the core Nod factor synthesis. *Mol. Microbiol.* **1995**, *15*, 733–747. [CrossRef]
163. Burrus, V.; Waldor, M.K. Formation of SXT tandem arrays and SXT-R391 hybrids. *J. Bacteriol.* **2004**, *186*, 2636–2645. [CrossRef]
164. Kung, V.L.; Ozer, E.A.; Hauser, A.R. The accessory genome of *Pseudomonas aeruginosa*. *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 621–641. [CrossRef]
165. Guerillot, R.; Da Cunha, V.; Sauvage, E.; Bouchier, C.; Glaser, P. Modular evolution of TnGBSs, a new family of ICEs associating IS transposition, plasmid replication and conjugation for their spreading. *J. Bacteriol.* **2013**, *195*, 1979–1990. [CrossRef]
166. Berger, C.; Rückert, C.; Blom, J.; Rabaey, K.; Kalinowski, J.; Rosenbaum, M.A. Estimation of pathogenic potential of an environmental *Pseudomonas aeruginosa* isolate using comparative genomics. *Sci. Rep.* **2021**, *11*, 1370. [CrossRef]
167. Bakaev, M.; Vladimirova, M.; Kozlova, A.; Muntyan, V.; Afonin, A.; Grudinin, M.; Roumiantseva, M. Abundant of CRISPR/Cas elements in genomes of nitrogen-fixing bacteria from the genus *Sinorhizobium*. *FEBS Open Bio* **2021**, *11*, 134. [CrossRef]

168. Muntyan, V.S.; Afonin, A.M.; Vladimirova, M.E.; Saksaganskaya, A.S.; Gribchenko, E.S.; Baturina, O.; Roumiantseva, M.L. Complete genome sequence of *Sinorhizobium meliloti* S35m, a salt-tolerant isolate from alfalfa rhizosphere in soil native to the Caucasus region. *Microbiol. Resour. Announc.* **2021**, *10*, e01417-20. [CrossRef]
169. Charbonneau, D.M.; Aubé, A.; Rachel, N.M.; Guerrero, V.; Delorme, K.; Breault-Turcot, J.; Masson, J.F.; Pelletier, J.N. Development of *Escherichia coli* asparaginase II for immunosensing: A trade-off between receptor density and sensing efficiency. *ACS Omega* **2017**, *2*, 2114–2125. [CrossRef]
170. Zhouravleva, G.A.; Rovinski, N.S. Adaptive role of tmRNA. *Ecol. Genet.* **2005**, *3*, 12–18. [CrossRef]
171. Gillet, R.; Felden, B. Emerging views on tmRNA mediated protein tagging and ribosome rescue. *Mol. Microbiol.* **2001**, *42*, 879–885. [CrossRef]
172. Withey, J.H.; Friedman, D.I. The biological roles of trans-translation. *Curr. Opin. Microbiol.* **2002**, *5*, 154–159. [CrossRef]
173. Marketon, M.M.; González, J.E. Identification of two quorum-sensing systems in *Sinorhizobium meliloti*. *J. Bacteriol.* **2002**, *184*, 3466–3475. [CrossRef]
174. Gurich, N.; González, J.E. Role of quorum sensing in *Sinorhizobium meliloti*-alfalfa symbiosis. *J. Bacteriol.* **2009**, *191*, 4372–4382. [CrossRef] [PubMed]
175. Gao, M.; Coggin, A.; Yagnik, K.; Teplitski, M. Role of specific quorum-sensing signals in the regulation of exopolysaccharide II production within *Sinorhizobium meliloti* spreading colonies. *PLoS ONE* **2012**, *7*, e42611. [CrossRef] [PubMed]