Systematic and Applied Microbiology 38 (2015) 176–183

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Genetic divergence and gene flow among *Mesorhizobium* strains nodulating the shrub legume *Caragana*

Zhaojun Ji[a,b], Hui Yan[a,b], Qingguo Cui[c], Entao Wang[a,d], Wenxin Chen[a,b], Wenfeng Chen[a,b,*]

[a] State Key Laboratory of Agrobiotechnology, Beijing 100193, China
[b] College of Biological Sciences and Rhizobia Research Center, China Agricultural University, Beijing 100193, China
[c] State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[d] Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, México D. F. 11340, Mexico

ARTICLE INFO

Article history:
Received 16 December 2014
Received in revised form 27 February 2015
Accepted 28 February 2015

Keywords:
Mesorhizobium
Caragana
Microevolution
Mutation
Recombination
Phylogeny

ABSTRACT

Although the biogeography of rhizobia has been investigated extensively, little is known about the adaptive molecular evolution of rhizobia influenced by soil environments and selected by legumes. In this study, microevolution of *Mesorhizobium* strains nodulating *Caragana* in a semi-fixing desert belt in northern China was investigated. Five core genes—*atpD*, *glnII*, *gyrB*, *recA*, and *rpoB*, six heat-shock factor genes—*clpA*, *clpB*, *dnaK*, *dnaJ*, *grpE*, and *hlsU*, and five nodulation genes—*nodA*, *nodC*, *nodD*, *nodG*, and *nodP*, of 72 representative mesorhizobia were studied in order to determine their genetic variations. A total of 21 genospecies were defined based on the average nucleotide identity (ANI) of concatenated core genes using a threshold of 96% similarity, and by the phylogenetic analyses of the core/heat-shock factor genes. Significant genetic divergence was observed among the genospecies in the semi-fixing desert belt (areas A–E) and Yunnan province (area F), which was closely related to the environmental conditions and geographic distance. Gene flow occurred more frequently among the genospecies in areas A–E, and three sites in area B, than between area F and the other five areas. Recombination occurred among strains more frequently for heat-shock factor genes than the other genes. The results conclusively showed that the *Caragana*-associated mesorhizobia had divergently evolved according to their geographic distribution, and have been selected not only by the environmental conditions but also by the host plants.

Introduction

In the northern region of China, deserts, Gobi, and sand dunes, (Supplementary Fig. S1) have a great influence on the agriculture and livelihood of local people and animals. Sandstorms are the main natural calamities in this region, especially during the spring season. As excellent windproof and sand-fixing plants, legumes such as *Caragana* spp., *Hedysarum scoparium* [17], *Hedysarum mongolicum* [39], *Alhagi sparsifolia* [18], and *Sophora alopecuroides* [41], are widely distributed in the arid and semi-arid deserts. Nitrogen fixation of these legumes by establishing symbiosis with rhizobia plays an important role in their healthy growth, especially in barren sand soil with low nutrient content [19].

Of the above psammophytic vegetation, *Caragana* species have been planted widely in the semi-fixed deserts. The predominant microsymbionts nodulating these legumes, especially the *Caragana* species, are *Mesorhizobium* species, and they have a good fitness in alkaline sands due to their acid production capability, as well as high- or low-temperature and drought resistance [4,17]. It has been documented that 10 species, including *Mesorhizobium septentrionale*, *M. amorphae*, *M. gobiense*, *M. mediterraneum*, *M. temperatum*, *M. caraganae*, *M. huakuii*, *M. tianshanense*, *M. metallidurans*, and *M. shangrilense*, could nodulate *Caragana* in extreme arid and semi-arid deserts, or on mountains with perennially low temperature [26,40].

To date, the existence of rhizobial biogeographic patterns has been well demonstrated [3,19,25,40], but little is known about their underlying processes. The central goals of rhizobial biogeography are to understand their diversification mechanisms, as well as the effects of natural selection, genetic drift, dispersal, and gene mutation, on the microevolution of the rhizobia concerned [10]. Currently, a series of interventions on soil stability and biodiversity

* Corresponding author at: College of Biological Sciences and Rhizobia Research Center, China Agricultural University, Beijing 100193, China. Tel.: +86 1062734009.
E-mail address: chenwf@cau.edu.cn (W. Chen).

http://dx.doi.org/10.1016/j.syapm.2015.02.007
0723-2020/

preservation have provided tools and strategies to address this problem. The *Caragana*-nodulating rhizobia may be used as an additional model to answer these questions because they can survive in the stressful environments of nutrient-poor deserts or inhabit the root nodules of *Caragana* [1,22]. In these rhizobia, the nodulation and heat-shock related genes play important roles in the symbiosis with legumes and adaptation to the stressful environments. Analyses of the genetic divergence and gene flow (the movement and successful establishment of genotypes from one to another) may reflect their histories of adaptive molecular evolution.

In order to help characterize the rhizobial microevolutionary mechanism, three types of genes, including six heat-shock factor genes, five nodulation genes, and five core genes, were chosen in the present study to investigate the genetic diversity, gene recombination/mutation events, and gene flow, among the *Caragana*-nodulating *Mesorhizobium*.

Materials and methods

Rhizobial strains

A total of 724 *Caragana*-nodulating rhizobial strains were selected according to the results of previous studies [25,40]. These strains originated from six areas including the southeast Tengger Desert (area A, 11 strains), south Mu Us Desert (area B, 447 strains), east Kubuqi Desert (area C, 65 strains), southeast Hunshandake Desert (area D, 72 strains), south Horqin Desert (area E, 113 strains), and the mountains of northwest Yunnan (area F, 16 strains) (Supplementary Fig. S1). The areas A–E were distributed along the northern semi-fixing desert belt in northern China that has a typical temperate desert climate characterized by extremely dry, dramatic day/night (approximately 50 °C), and seasonal (−30 to 50 °C) variation in temperature with frequent strong winds; whereas area F was on a subtropical plateau with a damp climate and moderate temperature variation (approximately 25 °C for day/night and between 0 and 32 °C for different seasons) [20]. Strains obtained from three neighboring sites (B-1, B-2, and B-3) in area B were analyzed as a subset in order to study the influence of geographic distance on the gene exchanges between the mesorhizobia. Eight type strains, originating from other hosts, for the species *Mesorhizobium amorphae*, *M. gobiense*, *M. tianshanense*, *M. mediterraneum*, *M. temperatum*, *M. metallidurans*, *M. huakuii*, and *M. septentrionale*, were used as references to determine the phylogenetic positions of the *Caragana* mesorhizobia. All these strains were cultured in TY medium at 28 °C and preserved in 20% glycerol at −80 °C [32].

Gene amplification and sequencing

Template DNA was extracted as described previously [36]. The PCR amplification protocols and functions of the six heat-shock factor genes *clpA*, *clpB*, *dnaK*, *dnaJ*, *grpE*, *hlsU* [2,4], five nodulation genes *nodA*, *nodC*, *nodD*, *nodG*, *nodP* [23,37], and five core genes *atpD*, *glnII*, *gyrB*, *recA*, *rpoB* [37], are shown in detail in Supplementary Table S2 [7,29,30]. PCR products were purified and commercially sequenced by an ABI 3730xl sequencer using the corresponding primers. All sequences obtained were checked using Chromas Pro (Ver. 1.7.6), and edited manually using DNAMAN (Ver. 7.212). A total of 1137 nucleotide sequences were obtained in this study and they were deposited in the GenBank database under accession numbers KP250993 to KP252129 (see Supplementary Table S8 for detail).

Phylogenetic analyses

The nucleotide sequences were aligned using the ClustalW program [33]. Neighbor-joining (NJ) phylogenetic trees were constructed using the MEGA5 program [34] with the Kimura 2-parameter model. Maximum likelihood (ML) trees were constructed using the phyML program [8] based on models selected by the Akaike information criterion (AIC) test in the MODELTEST 3.7 package [28]. SPLITSTREE 4.0 was used to assess the degree of a tree-like structure for alleles of each locus, and to reveal the potentially incompatible signals in the evolutionary history with split phylogenetic networks (1000 bootstraps) [13]. The Shimodaira–Hasegawa (SH) test [31] was performed in order to evaluate the topological consistency between the phylogenetic trees using the PAUP* 4.0b1 program [33].

Nucleotide polymorphism and population genetics analyses

The software package DnaSP (DNA sequence polymorphism) [21] was used to estimate the nucleotide polymorphisms, including the number of haplotypes (h), haplotype diversity (Hd), nucleotide diversity (π), and π_N/π_S ratios (π_S is the number of synonymous substitutions per synonymous site; π_N is the number of non-synonymous substitutions per non-synonymous site) [27], as well as average nucleotide divergence (Dxy), and gene flow (the number of migrants, Nm). CLONALFRAME was used to deduce the effect of recombination on the evolution with three independent runs (100,000 burn-in iterations, plus 1,000,000 sampling iterations for each run); satisfactory judgment of which was based on the method described previously [6]. Two recombination rates: r/m (the relative impact of recombination compared with that of the point mutation in the genetic diversification of the lineage), and ρ/θ (the relative frequency of the occurrence of recombination compared with that of the point mutation in the history of the lineage) were also calculated [6]. Minimal recombination events (Rm) within the populations were estimated using DnaSP [12]. Admixture levels of the *Mesorhizobium* genospecies were investigated using STRUCTURE with the LOCPRIOR model [11,15], and a K value of 4 was chosen for each of the three types of genes with 100,000 burn-in and 1,000,000 sampling iterations.

Results

Phylogenies based on concatenated gene sequences

The phylogenetic tree of all 724 rhizobial strains was firstly grouped based on *recA* sequences (not shown), and 72 strains (see Supplementary Table S1) were selected to represent each of the genotypes. These representatives fully covered the diverse *Caragana* rhizobial populations due to their high haplotype diversity values ($Hd = 0.8 \sim 1.0$) (see Supplementary Table S3). Topological structures of the phylogenetic trees based on each core gene showed no significant differences using the neighbor-joining (NJ), and maximum likelihood (ML) methods (not shown). Based on the concatenated core genes (Fig. 1), 21 genospecies were differentiated according to the ANI threshold value of 96%. Nine genospecies were identified as defined species that corresponded to *M. septentrionale*, *M. amorphae*, *M. gobiense*, *M. mediterraneum*, *M. temperatum*, *M. caraganae*, *M. shangrilense*, *M. loti*, and *M. huakuii*, respectively. The other 12 genospecies showed ANI values lower than 96% with the defined species and they were designated as novel genospecies named *Mesorhizobium* spp. I–XII, respectively.

All the strains were grouped according to the ANI-based genospecies, and the topological structures were similar between the core (Fig. 1) and heat-shock factor (Supplementary Fig. S2) gene trees, although some variations could be observed in several cases, including CCBAU 01583, CCBAU 01790, CCBAU 01718, and CCBAU 01764. However, the phylogenetic relationships of the concatenated sequences of the nodulation genes (Fig. 2) were significantly