INSTITUT PASTEUR

Research in Microbiology 161 (2010) 830–837

www.elsevier.com/locate/resmic

Guidelines for nomenclature assignment of Ger receptors

Christian A. Ross [b,*], Ernesto Abel-Santos [a,1]

[a] Department of Chemistry, University of Nevada, Las Vegas, Las Vegas, NV 89154, USA
[b] School of Life Sciences, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154, USA

Received 5 May 2010; accepted 2 August 2010
Available online 24 September 2010

Abstract

Some members of the orders *Bacillales* and *Clostridiales* form dormant spores when subjected to environmental stress. These resistant spores return to normal vegetative growth upon encountering nutrients. A family of membrane-bound proteins called germination (Ger) receptors is tasked with detecting metabolites that serve as germination signals. During the characterization of tripartite Ger receptor proteins from the genera *Bacillus* and *Clostridium*, we found numerous nomenclature inconsistencies. In this work, we issued guidelines to remediate this problem. We generated a sample of sequenced *Bacillus*, *Clostridium* and related endospore-forming genera. Ger receptor proteins were recovered from these genomes by PSI-BLAST. The resulting Ger receptor protein sequences were then clustered by neighbor-joining based on sequence alignment. Inconsistencies in Ger receptor labeling were noted and new names proposed. Finally, a systematic approach for naming of new Ger receptors was designed.

Keywords: Ger receptors; *Bacillus*; *Clostridium*; Spore germination

1. Introduction

In the Gram-positive spore-formers of the orders *Bacillales* and *Clostridiales,* endospores are formed when survival conditions are suboptimal for vegetative growth. These environmentally hardened spores remain viable for long periods until conditions improve. It is the function of germination (Ger) receptors to detect small molecules signals that trigger the emergence of spores from dormancy. Each Ger receptor is generally expressed from a tricistronic *ger* receptor operon. The tripartite receptors form a protein family, some members of which have established roles in recognizing specific germinants (Atluri et al., 2006; Fisher and Hanna, 2005; Hornstra et al., 2006). The Ger receptor complex is believed to be composed of two putative transmembrane proteins termed subunit A and subunit B (e.g., GerAA and GerAB) and a third, potentially membrane anchored protein termed subunit C (e.g., GerAC) (Ross and Abel-Santos, 2010).

Ger receptors were first named in *Bacillus subtilis* after a concerted effort by geneticists to map and phenotype mutants in the spore germination pathway (Moir et al., 1979). The first studied Ger receptor responded to a single germinant, L-alanine, and was designated GerA (Yasuda and Tochikubo, 1984). Two other Ger receptors were shown to act cooperatively to trigger germination with either L-asparagine supplemented with glucose, fructose, and potassium ions (GFK) or L-alanine supplemented with GFK (Irie et al., 1996). These two receptors were designated GerB and GerK, respectively.

The first designations for genes encoding Ger receptors in *B. subtilis*, *gerA* and *gerB*, were regarded as temporary (Moir et al., 1979), but have remained in use without regards for phylogenetic relationships or functional identity. Furthermore, *B. subtilis* also possess two hypothetical *ger* receptor operons, *yfkQRT* and *yndDEF* (Paidhungat and Setlow, 2000) that encode proteins with no assigned function.

Among the *Bacillus cereus* species group, six *ger* receptor operons (*gerA*, *gerH*, *gerK*, *gerL*, *gerS*, and *gerY*) are found in

* Corresponding author. Tel.: +1 702 895 5834.
E-mail addresses: rossc6@unlv.nevada.edu (C.A. Ross), ernesto.abelsantos@unlv.edu (E. Abel-Santos).
[1] Tel.: +1 702 895 2608.

0923-2508/$ - see front matter © 2010 Institut Pasteur. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.resmic.2010.08.002

the sequenced *Bacillus anthracis* Ames genome (Fisher and Hanna, 2005; Read et al., 2003) and a seventh (*gerX*) is located on the virulence plasmid pX01 (Chantal et al., 1999). The closely related *B. cereus* strain E33L has seven chromosomally located *ger* receptor operons. Because of inconsistent nomenclature, seven labels (*gerA*, *gerB*, *gerH*, *gerI*, *gerK*, *gerL* and *gerQ*) have been entered into GenBank (Han et al., 2006). An additional *ger* receptor partial operon located on plasmid pE33L466 has also been labeled *gerK*. Another strain of the same species, *B. cereus* ATCC 14579, was found to have seven *ger* receptor operons (Ivanova et al., 2003) and they were labeled *gerG*, *gerI*, *gerK*, *gerL*, *gerQ, gerR* and *gerS* (Hornstra et al., 2006). The difference in nomenclature in evolutionarily related *ger* receptor operons creates confusion in the identification of individual genes.

With the expansion of the number of Ger receptors characterized and the availability of a substantial number of completely sequenced bacterial genomes, the limitations of the existing designations have become apparent. In many cases, inconsistent or confusing notations have been applied. For example, directly orthologous Ger receptors in closely related *Bacillus* species have been given different names (e.g. the proteins encoded by the *gerH* operon in *B. anthracis* and the *gerI* operon in *B. cereus* are almost identical). Furthermore, designations have become largely independent of their evolutionary and functional roles. For example, *gerK* has been used to designate *Clostridium* and *Bacillus ger* receptor operons that have little sequence homology.

In the present study, we introduce a systematic approach to clarify the nomenclature of Ger receptors. Our goal is to make a minimum number of nomenclature changes, to avoid changing protein names that already reflect functional roles. Furthermore, we attempt to reflect consistent evolutionary relationships as best as they can be inferred at present. In order to systematize the Ger receptor nomenclature, we selected a representative sample of sequenced *Bacillus*, *Clostridium* and related endospore formers. We then extracted proteins encoded by Ger receptor operons through PSI-BLAST searches, and clustered them via protein sequence alignment and neighbor-joining. After reviewing the Ger receptor names included in the associated Genbank records, we proposed a series of nomenclature changes, provided a list of Entrez accession numbers for these receptors and their orthologues, and suggested guidelines for future Ger receptor annotations.

2. Materials and methods

2.1. Construction of germination protein database

Protein PSI-BLAST was performed on October 28, 2008, using GerAA (NP_391185), GerAB (NP_391186), GerAC (NP_391187) as queries. The results were used as seeds to develop a position-specific weight matrix (PSWM) model to search the Genbank sequenced genome database. Data from PSWM were used to develop a protein family (Altschul et al., 1997). The result of each PSI-BLAST iteration was filtered against the list of organisms in Table S1. PSI-BLAST results with E-values of more than 10^{-15} were discarded. PSI-BLAST was continued until convergence was achieved for all three Ger receptor subunits (no new sequences discovered with subsequent searches).

2.2. Protein alignment

Each family of related proteins from the PSI-BLAST search was imported into MEGA 4.0 software (Tamura et al., 2007). Proteins were aligned with ClustalW using the Gonnet matrix. The aligned proteins were then trimmed to remove leading and trailing short peptides not common to all members of the family. Sequences that were incomplete or aligned poorly were not included in phylogenetic reconstruction. The resultant trimmed alignments were then used to construct neighbor-joining trees in MEGA software using homogenous Poisson correction for estimating distances (Nei and Kumar, 2000). Complete deletion was used for missing sites. Bootstrap values were calculated from 1000 replicate trees.

2.3. Procedure for nomenclature assignment in B. cereus ATCC 14579 and Bacillus megaterium QM B1551

Identification of *ger* receptor operons encoding the GerG, GerR and GerL receptors was accomplished by nucleotide BLAST against the NCBI genome sequence of *B. cereus* ATCC 14579 (NC_004722) using the specificity regions of primers provided (Hornstra et al., 2006) as *in silico* probes. The remaining Ger receptors were retrieved from *B. cereus* ATCC 14579 by BLAST using an amino acid sequence of *B. subtilis* GerAA. Identification of the *B. megaterium* QM B1551 GerU receptor, GerVB, and GerWB proteins was accomplished by BLAST using the amino acid sequence of *B. subtilis* GerAA.

Presumptive identification of Ger receptor protein orthologues from the list of organisms in Table S1 was accomplished by protein BLAST. To confirm membership in each orthologous grouping, protein sequences were imported into the existing alignment, realigned using ClustalW, and a new neighbor-joining tree constructed as described above.

3. Results and discussion

3.1. Number of receptors

BLAST searches conducted with GerAA from *B. subtilis* and limited to one sample strain from each sporulating bacteria species with a finished sequenced genome (listed in Supplementary Table S1) resulted in 139 hits in the NCBI RefSeq database, 20 of which were identified as SpoVAF-like proteins. Similarly, 125 and 112 sequences were discovered using GerAB, and GerAC as seeds. Use of alternate Ger receptor proteins (GerXA, GerXB, GerXC from pXO1) as seed sequences did not affect the results. In all, the number of encoded Ger receptors varies between species from one to seven distinct operons. For example, the *B. anthracis* 'Ames strain' appears to contain four intact chromosomally encoded