Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 345 (2005) 73–80
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Comparative analysis of two genomic regions among four strains of *Buchnera aphidicola*, primary endosymbiont of aphids

Vicente Pérez-Brocal, Amparo Latorre, Rosario Gil, Andrés Moya*

Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado de correos 22085, 46071 València, Spain

Received 13 September 2004; received in revised form 25 October 2004; accepted 15 November 2004
Available online 29 December 2004
Received by U. Bastolla

Abstract

Preliminary analysis of two selected genomic regions of *Buchnera aphidicola* BCc, the primary endosymbiont of the cedar aphid *Cinara cedri*, has revealed a number of interesting features when compared with the corresponding homologous regions of the three *B. aphidicola* genomes previously sequenced, that are associated with different aphid species. Both regions exhibit a significant reduction in length and gene number in *B. aphidicola* BCc, as it could be expected since it possess the smallest bacterial genome. However, the observed genome reduction is not even in both regions, as it appears to be dependent on the nature of their gene content. The region *fpr–trxA*, that contains mainly metabolic genes, has lost almost half of its genes (45.6%) and has reduced 52.9% its length. The reductive process in the region *rrl–aroK*, that contains mainly ribosomal protein genes, is less dramatic, since it has lost 9.3% of genes and has reduced 15.5% of its length. Length reduction is mainly due to the loss of protein-coding genes, not to the shortening of ORFs or intergenic regions. In both regions, G+C content is about 4% lower in BCc than in the other *B. aphidicola* strains. However, when only conserved genes and intergenic regions of the four *B. aphidicola* strains are compared, the G+C reduction is higher in the *fpr–trxA* region.

Keywords: Evolutionary genomics; Genome reduction; Gene disintegration; G+C content

1. Introduction

Bacterial symbioses are widespread among insects (Baumann et al., 2000; van Ham et al., 2004), probably being one of the key factors of their evolutionary success, since they may have allowed access to novel ecological niches and to new unbalanced food resources. A well studied case is the nutritional symbiotic association between aphids (Hemiptera: Aphididae) and *Buchnera aphidicola*, a γ-proteobacterium (Unterman et al., 1989). The bacteria provide essential amino acids to their hosts, which are deficient in their phloematic diet. On the other hand, the hosts harbour the bacteria in a nutrient-rich and protected environment, within specialised cells called bacteriocytes that constitute bacteriomes, organs located in the hemocoel, surrounding the gut of the aphids (Houk and Griffiths, 1980; Douglas, 1989). The aphids–*B. aphidicola* association was established about 80–150 Myr (von Dohlen and Moran, 2000) by a single infection followed by strict vertical transmission from mother to eggs or embryos. The symbiosis is obligatory for both partners, as it is revealed by the impossibility of culturing *B. aphidicola* out of the aphids, while treatments with antibiotics causes abnormal growth, lack of reproduction, and premature death of the insects (Sasaki et al., 1991; Douglas, 1996).

Three *B. aphidicola* genomes have been sequenced to date, two of them from bacteria harboured by aphids from the subfamily Aphidinae, *Acyrtosiphon pisum* (*B. aphidicola* BAp; Shigenobu et al., 2000) and *Schizaphis graminum* (*B. aphidicola* BSg; Tamas et al., 2002), which

Abbreviations: G+C, molar fraction of guanine and cytosine in DNA; ORF, open reading frame; Myr, million years; GC content; kb, kilobase(s); PFGE, pulse field gel electrophoresis; BLAST, basic local alignment search tool; tRNA, transfer RNA; bp, base pair(s); rRNA, ribosomal RNA; COG, cluster of orthologous genes; *df*, degrees of freedom; IR, intergenic region(s).

* Corresponding author. Tel.: +34 96 354 3480; fax: +34 96 354 3670.
E-mail address: andres.moya@uv.es (A. Moya).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.11.021

diverged 50 Myr, and a third one from the subfamily Pemphiginae, *Baizongia pistaciae* (*B. aphidicola* BBp; van Ham et al., 2003), which diverged from the aforementioned ones about 150 Myr. *B. aphidicola* shows an extreme reduction of its genome size, compared to their free-living relatives such as *Escherichia coli*, with chromosome sizes of 640, 641, and 616 kb, and 560, 545, and 504 protein-coding genes in BAp, BSg, and BBp genomes, respectively. Such genomic reduction is characteristic of endosymbiotic and parasitic bacteria, as a consequence of their adaptation to intracellular life. In these conditions, many gene products became unnecessary, since they can be provided by the environment. In absence of selective pressure and horizontal gene transfer, only essential genes for each particular life conditions tend to be preserved, causing a dramatic genome size reduction (Wernegreen, 2002).

By PFGE, Gil et al. (2002) determined that the genome of several *B. aphidicola* strains from different aphid subfamilies has suffered an even more dramatic reductive process. The smallest genomes analyzed correspond to three *B. aphidicola* strains from aphids of subfamily Lachninae: BCcu (endosymbiont of *Cinara cupresi*) and BCtu (endosymbiont of *Cinara tujafilina*), with an estimated chromosome size of 475 kb, and BCc (endosymbiont of *Cinara cedri*), with an estimated size of 450 kb, the smallest bacterial genome reported so far. The sequencing and characterization of the gene content of this extremely reduced genome, and its comparison with the three previously sequenced *B. aphidicola* genomes, will deepen our knowledge about which genes are essential for endosymbiotic life, and will allow a better understanding of the factors involved in the process of genome shrinkage, as well as the evolutionary process involved in the adaptation of each *B. aphidicola* strain to its particular aphid host.

Besides *B. aphidicola*, a certain number of aphid lineages harbour other types of bacteria designated as secondary symbionts (Fukatsu and Ishikawa, 1998; Sandström et al., 2001; Russell et al., 2003). The aphid *C. cedri* also harbours a high amount of secondary symbionts (Gómez-Valero et al., 2004b), which cannot be completely eliminated in bacteriocyte extractions, making it difficult to obtain sufficiently pure *B. aphidicola* BCc genomic DNA for a shotgun sequencing approach. Therefore, we decided to perform the sequencing of this genome by genomic walking techniques.

The comparative genomic analysis of the three previously sequenced *B. aphidicola* strains revealed a nearly perfect gene order conservation (Tamas et al., 2002; Silva et al., 2003; van Ham et al., 2003). Our preliminary sequencing results also show that most of the *B. aphidicola* BCc genome has maintained the same gene order. In the present work we report the analysis of two selected genome regions of BCc that were chosen on the nature of their gene content, and its comparison with the corresponding homologous regions in the three already sequenced *B. aphidicola* genomes. The first region (*rrl–aroK* region) is located between the genes that encode ribosomal 23S RNA and shikimate kinase I, and it contains mostly essential genes coding for ribosomal proteins. The second region (*fpr–trxA* region) is located between the coding genes for ferredoxine-NADP reductase and thioredoxin, and it contains several putative nonessential metabolic genes.

2. Materials and methods

2.1. B. aphidicola genomic DNA purification from C. cedri

Aphids were collected in May and June 2002 from cedar trees in Llíria (Valencia, Spain). Extraction and purification of bacteriocytes containing *B. aphidicola* were carried out as reported (Gil et al., 2002). Part of the bacteriocytes was used to obtain total DNA, following the CTAB/Na method (Ausubel et al., 1995). The remaining bacteriocytes were included in low melting point agarose plugs (Riethman et al., 1997; Charles and Ishikawa, 1999). Contaminants (cell membranes, cut, or damaged DNA from both *B. aphidicola* and secondary symbionts, etc.) were eliminated by PFGE as described (Gil et al., 2002). Subsequently, the plugs were removed from the wells and the included genomic DNA was linearized with 40 units of *Rsr*II. A second PFGE with a gradient of 4.5 V/cm and switch times of 10–90 s for 24 h was run to separate the 450 kb band that corresponds to the *B. aphidicola* linear chromosome. *Saccharomyces cerevisiae* chromosomes (New England Biolabs) were used as molecular weight markers. The band was cut out from the gel and purified by electroelution in a dialysis cellulose membrane (SIGMA) for 1 h at 100 V/cm. Finally, DNA was cleaned by phenol extraction and concentrated by precipitation, using standard protocols.

2.2. Construction and sequencing of partial genomic libraries

Purified genomic linear DNA was digested with four-cutter restriction enzymes generating blunt (*Alu*I and *Rsa*II) or cohesive ends (*Sau*3AI and *Taq*I). Blunt ended fragments were cloned into pUC18/*Sma*I (Amersham Pharmacia Biotech) and *Sau*3AI and *Taq*I fragments were cloned into pUC18/*Bam*HI (Amersham Pharmacia Biotech) and pUC18/*Acc*I (MBI Fermentas) plasmids, respectively. Plasmid DNA purification was carried out on 96-well plates with a MultiPROBE II Ex engine, using Qiagen and Millipore kits. Sequencing was performed in 96-well PCR plates with BigDye3 kit (Applied Biosystems) by using universal pUC primers.

2.3. Long Expand PCR and minilibraries

Sequences scattered along the genome, obtained from the partial genomic libraries previously described, were