Letter

Discovery of protein-coding palindromic repeats in *Wolbachia*

Hiroyuki Ogata, Karsten Suhre and Jean-Michel Claverie

Information Génomique et Structurale, CNRS, UPR2589, IBSM, 31 chemin Joseph Aiguier, 13402, Marseille, Cedex 20, France

Recurrent use of existing genes through gene duplication or lateral transfer is the most common evolutionary mechanism to generate new protein-coding genes in bacteria [1,2]. However, this leaves the question as to the origin of the primordial gene pool unanswered, and does not provide a comprehensive model for *de novo* creation of new protein sequences during the course of evolution. *Rickettsia* are rather unique in this context, uniquely exhibiting mobile palindromic repeats (RPEs) that are capable of creating new peptides (35–50 amino acids) by inserting themselves within protein-coding genes [3]. The recurrent application of such a process, coupled with the accelerated evolution of the inserted peptides, might in fact account for *bona fide* genomic innovations [4,5]. Given its potentially enormous evolutionary significance, it is paradoxical that this phenomenon is not encountered in many other bacteria. However, the analysis of recently obtained *Wolbachia* genomes led to the identification of (i) RPEs and (ii) a new family of protein-coding palindromic repeats in these genomes, suggesting that this puzzling evolutionary process is also at work in *Wolbachia*.

Rickettsia and *Wolbachia* are obligate intracellular parasites of the same family *Rickettsiaceae* of the α-proteobacteria, associated with a variety of invertebrate species. However, the two bacterial genera are quite distant (Figure 1a), their divergence going back 400 to 800 million years (My) [6]. The first evidence was obtained upon analysis of the genome of *Wolbachia pipientis* *w*Mel, an arthropod-associated parasite [7]. In this genome, two copies of the *Rickettsia* palindromic element-1 (RPE-1) [4] were identified that exhibit significant sequence similarities (E-value $<1.0\times10^{-4}$) [8] to previously identified RPE-1 nucleotide sequences. The two *w*Mel RPE-1 nucleotide sequences were predicted to fold into stable hairpin-like structures. One of them (119bp) was found within a non-coding region, the other (147bp) was found within the *carB* gene. The protein encoded by *carB* is likely to be functional, despite the presence of the inserted repeat, for the following reasons. *carB* is predicted to encode the large subunit of carbamoyl-phosphate synthase, a key enzyme of the pyrimidine biosynthesis pathway. The RPE-1 is inserted in-frame within the gene and has a 49 amino acid peptide sequence similar to the coding RPE-1 of *Rickettsia*, which has been described previously (Figure 1b). The *w*Mel-genome contains a single copy of *carB* and a single copy of *carA* (encoding the small subunit of the enzyme). All genes for the carbamoyl-phosphate to UMP pathway have been identified in the genome [7]. Furthermore, the analysis performed on the basis of the three-dimensional structure of *Escherichia coli* CarBA (PDB: 1C3O) indicates that the RPE insertion site corresponds to a surface loop (Figure 1c), and is compatible with the overall structure and functions of the enzyme [9].

Evidence of a new family of palindromic repeats was obtained through the examination of homologous long insertions in different open-reading frames (ORFs) of the *w*Mel genome. Twenty-seven copies of repeated sequences were identified (83 to 321bp long) and found to be distributed throughout the genome. These repeats resemble RPEs in that they have a conserved 108bp-long palindromic structure, and that they frequently occur within predicted protein-coding genes (24 out of 27 cases). However, this new repeat family exhibits no relationship with the previously defined RPE families at the sequence level. These repeats have been named WPEs for *Wolbachia* palindromic elements. Among the ORFs that exhibit WPE inserts, five ORFs have precise functional annotations: cell-wall-associated enzymes MurF and MurE, cytochrome c biogenesis protein CcmF, biotin protein ligase BirA, and geranyltranstransferase IspA. The peptide sequences derived from these WPEs are shown in Figure 1b. When three-dimensional structures of homologues were available, we were able to verify that the insertions sites appeared structurally compatible with the normal folding and function of the host proteins.

Recently, the genome of another *Wolbachia* species (http://tools.neb.com/wolbachia/), a symbiont of the filariasis-causing nematode *Brugia malayi* (*w*Bm), has been determined. No clear homologues of RPEs could be found, but WPE homologues were identified both in coding and non-coding regions of the genome. The recent finding that *W. pipientis* and *Rickettsia felis* can infect identical flea species [10] suggests one route by which RPE-1 could have been acquired by *W. pipientis* from a close encounter with *Rickettsia*. Such a lateral transfer might not have occurred in *w*Bm.

The creation of new protein segments by palindromic repeat insertion has therefore occurred in two related but quite distant genera of α-proteobacteria. In general, the genomes of obligate intracellular bacteria tend to accumulate neutral or slightly deleterious mutations [3]. However, similar phenomena have not yet been described in

Corresponding author: Ogata, H. (Hiroyuki.Ogata@igs.cnrs-mrs.fr).
Available online 2 April 2005

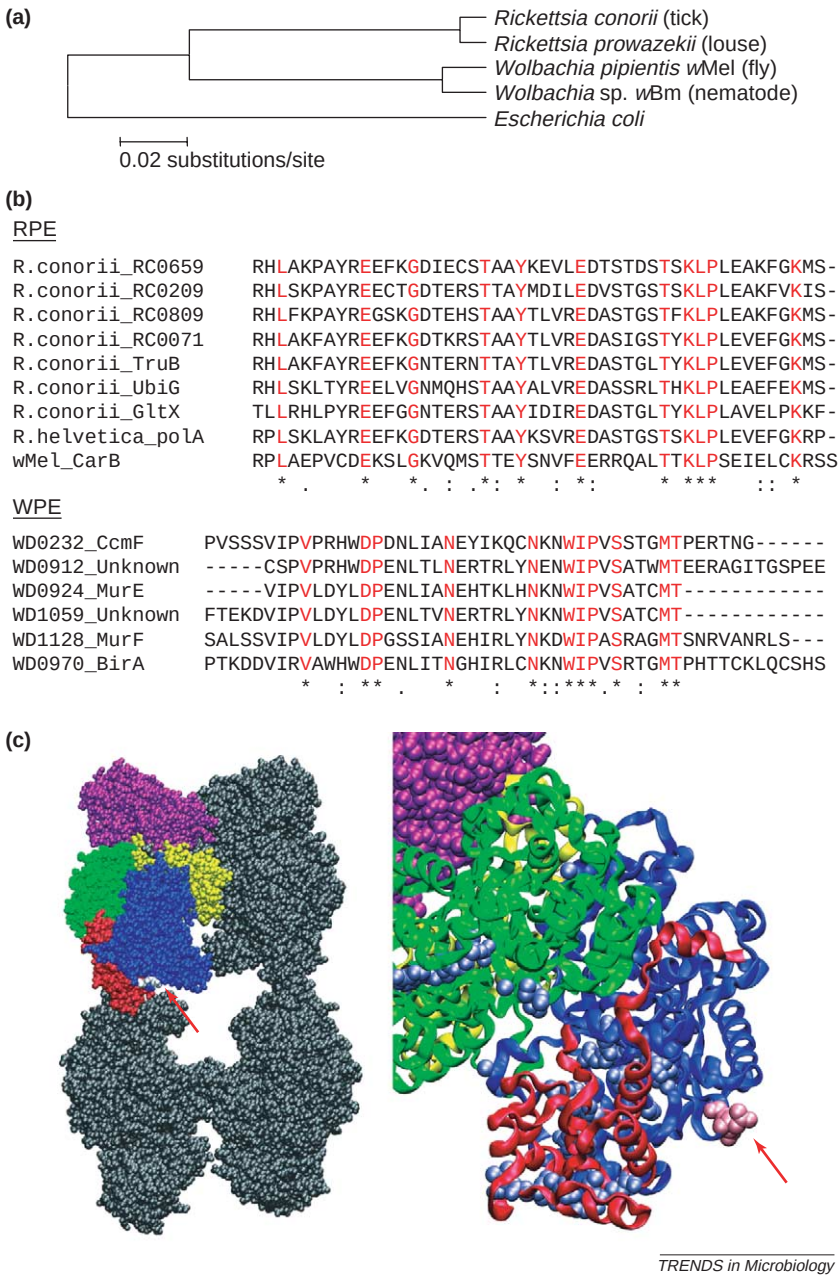

Figure 1. Protein-coding palindromic repeats in *Wolbachia* and *Rickettsia*. **(a)** A 16S rRNA phylogenetic tree showing the relationships between *Rickettsia* and *Wolbachia*. Invertebrate hosts are indicated in parentheses on the right of the species names. The tree was constructed with the unweighted pair-group method with arithmetic mean (UPGMA) method using MEGA [11]. **(b)** Amino acid sequence alignments of the RPE-1 repeats in *Wolbachia pipientis w*Mel and *Rickettsia,* and WPE repeats identified in *W. pipientis w*Mel. For the WPE, only the conserved central part of the alignment is shown. The WPE in lspA exhibited a high level of sequence difference compared with other WPEs and is not included here. The alignment was constructed using T-Coffee [12]. (c) Predicted insertion site of the *w*Mel RPE-1 in the large subunit of carbamoyl-phosphate synthase indicated by red arrows in the reference structure of the *Escherichia coli* enzyme (PDB: 1C3O). The *E. coli* enzyme is a tetramer of heterodimers, $(\alpha, \beta)_4$. CarA is colored in magenta. Four different domains of CarB are colored in green, yellow, blue and red. Blue spheres on the right panel represent residues involved in catalytic activities of CarB. Similar analyses have been done for four WPE-containing ORFs, for which the three-dimensional strucure of homologues were available [PDBs: 1E8C (MurE); 1BIB (BirA); 1GG4 (MurF); 1RQJ (lspA)].

other obligate intracellular bacteria. The presence of unrelated families of coding palindromic repeats in *Rickettsia* and *Wolbachia* now sheds a new light on these peculiar genetic elements, prompting us to speculate that they might be more centrally linked to the emergence or adaptive mechanisms of these parasitic α-proteobacteria than previously thought. It appears now more crucial than ever to experimentally elucidate the consequences of the repeat insertion on the functions of the host proteins.

References

1 Ohno, S. (1970) *Evolution by gene duplication*, Springer-Verlag
2 Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* 284, 2124–2129
3 Ogata, H. *et al*. (2000) Selfish DNA in protein-coding genes of *Rickettsia*. *Science* 290, 347–350
4 Ogata, H. *et al*. (2002) Protein coding palindromes are a unique but recurrent feature in *Rickettsia*. *Genome Res.* 12, 808–816
5 Claverie, J.M. and Ogata, H. (2003) The insertion of palindromic repeats in the evolution of proteins. *Trends Biochem. Sci.* 28, 75–80

6 Moran, N.A. *et al*. (1993) A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lond. B. Biol. Sci.* 253, 167–171
7 Wu, M. *et al*. (2004) Phylogenomics of the Reproductive Parasite *Wolbachia pipientis w*Mel: A Streamlined Genome Overrun by Mobile Genetic Elements. *PLoS Biol.* 2, 327–341
8 Eddy, S.R. (1996) Hidden Markov models. *Curr. Opin. Struct. Biol.* 6, 361–365
9 Holden, H.M. *et al*. (1999) Carbamoyl phosphate synthetase: an amazing biochemical odyssey from substrate to product. *Cell. Mol. Life Sci.* 56, 507–522
10 Rolain, J.M. *et al*. (2003) Molecular detection of *Bartonella quintana*, *B. koehlerae*, *B. henselae*, *B. clarridgeiae*, *Rickettsia felis*, and *Wolbachia pipientis* in cat fleas, France. *Emerg. Infect. Dis.* 9, 338–342
11 Kumar, S. *et al*. (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17, 1244–1245
12 Notredame, C. *et al*. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302, 205–217

0966-842X/$ - see front matter
doi:10.1016/j.tim.2005.03.013