# Complete genome sequence of Courdo11 virus, a member of the family *Mimiviridae*

**Niyaz Yoosuf · Isabelle Pagnier · Ghislain Fournous · Catherine Robert · Bernard La Scola · Didier Raoult · Philippe Colson**

**Abstract** Giant viruses of amoebae were discovered 10 years ago and led to the description of two new viral families: *Mimiviridae* and *Marseilleviridae*. These viruses exhibit remarkable features, including large capsids and genomes that are similar in size to those of small bacteria and their large genetic repertoires include genes that are unique among viruses. The family *Mimiviridae* has grown during the past decade since the discovery of its initial member, Mimivirus, and continues to expand. Here, we describe the genome of a new giant virus that infects *Acanthamoeba* spp., Courdo11 virus, isolated in 2010 by inoculating *Acanthamoeba* spp. with freshwater collected from a river in southeastern France. The Courdo11 virus genome is a double stranded DNA molecule composed of 1,245,674 nucleotides. The comparative analyses of Courdo11 virus with the genomes of other giant viruses showed that it belongs to lineage C of mimiviruses of amoebae, being most closely related to *Megavirus chilensis* and LBA 111, the first mimivirus isolated from a human. Major characteristics of the *M. chilensis* genome were identified in the Courdo11 virus genome, found to encode three more tRNAs. Genomic architecture comparisons mirrored previous findings that showed conservation of collinear regions in the middle part of the genome and diversity towards the extremities. Finally, fourteen ORFans were identified in the Courdo11 virus genome, suggesting that the pan-genome of mimiviruses of amoeba might reach a plateau.

**Keywords** Courdo11 virus · Mimivirus · *Mimiviridae* · Giant virus · Megavirales · Nucleocytoplasmic large DNA viruses · Amoeba

N. Yoosuf · I. Pagnier · G. Fournous · C. Robert · B. La Scola · D. Raoult · P. Colson (✉)
Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes (URMITE), UM63, CNRS 7278, IRD 198, INSERM U1095, Facultés de Médecine et de Pharmacie, Aix-Marseille Univ., 27 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France
e-mail: philippe.colson@univ-amu.fr

I. Pagnier · B. La Scola · D. Raoult · P. Colson
Fondation Institut Hospitalo-Universitaire (IHU) Méditerranée Infection, Pôle des Maladies Infectieuses et Tropicales Clinique et Biologique, Fédération de Bactériologie-Hygiène-Virologie, Centre Hospitalo-Universitaire Timone, Assistance Publique – Hôpitaux de Marseille, 264 rue Saint-Pierre, 13385 Marseille Cedex 05, France
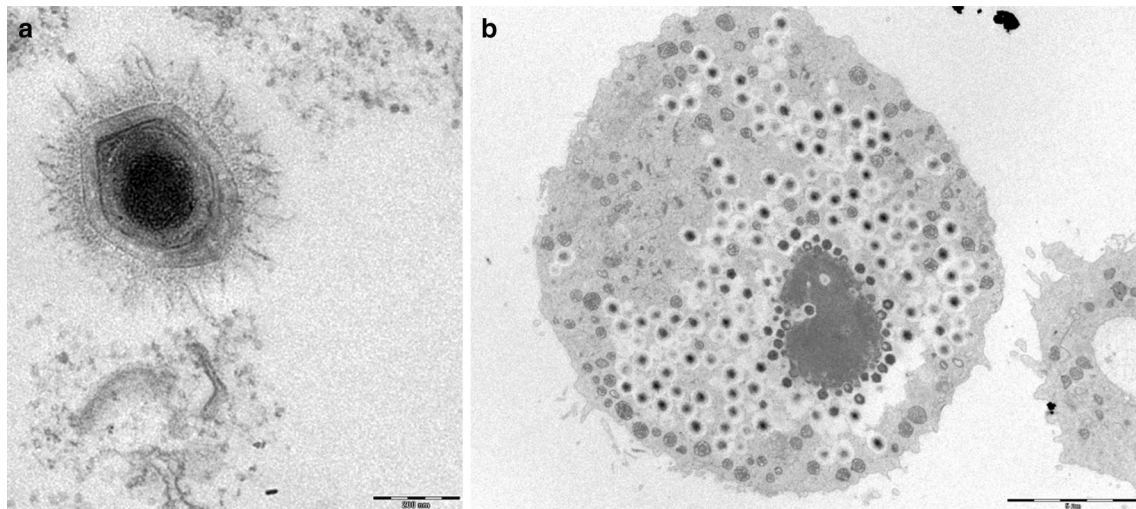
## Introduction

Giant viruses of amoebae were discovered 10 years ago and led to the description of two new viral families: *Mimiviridae* and *Marseilleviridae* [1–7]. These viruses exhibit remarkable features, including large capsids and genomes that are similar in size to those of small bacteria [7–10]. Furthermore, their large genetic repertoires include genes that encode components of the translation apparatus that are unique among viruses. Even larger amoebae viruses were recently described; the genomes of these viruses are 1.9 and 2.5 kbp in size [11]. Altogether, these findings have altered the definition of a virus [1, 11–13]. The family *Mimiviridae* has grown during the past decade since the discovery of its initial member, Mimivirus, and continues to expand [6, 9, 14]. The first group of this family is composed of mimiviruses that infect *Acanthamoeba* spp., among which three lineages, A, B and C, have been delineated with Mimivirus [1], Moumouvirus [5] and *Megavirus chilensis* [4] as the prototype members, respectively [7]. *Cafeteria roenbergensis* virus is a smaller,

**Fig. 1** Electron microscopy of Courdo11 virus particles (**a**, *left*) and *A. polyphaga* infected with Courdo11 virus 16 h post-infection (**b**, *right*). *Scale bars* represent 200 nm (**a**) and 5 μm (**b**)

distantly related mimivirus that infects a marine dinoflagellate [15]. Mimiviruses of amoebae can exchange DNA with other microorganisms that live sympatrically with the mimiviruses inside *Acanthamoeba* spp. [15], and they harbour a specific mobilome [16, 17]. These giant viruses are common in environmental samples, as shown by high rates of isolation from water and by metagenomic analyses [14, 18, 19]. In addition, LBA111 and Shan, two members of the lineage C of amoeba-associated mimiviruses were recently isolated from the bronchoalveolar fluid and stools, respectively, of Tunisian patients presenting pneumonia [20, 21]. These findings indicate that mimiviruses might cause pneumonia [22, 23]. Here, we describe the genome of a new mimivirus that infects *Acanthamoeba* spp.

## Materials and methods

Courdo11 virus was isolated in 2010 by inoculating *A. polyphaga*, as previously described [24], with freshwater drawn by one of us (IP) slightly below the surface of the Le Peyron creek in Saint-Raphael city, southeastern France. Its capsid size is approximately 450 nm (Fig. 1). The Courdo11 virus genome was sequenced on a 454-Roche GS20 device as described previously [1, 5], and the obtained reads were assembled de novo with Newbler Assembly software [25]. A second set of reads was obtained using an AB SOLiD instrument and was mapped on the previously assembled genome using CLC Bio software (http://www.clcbio.com/index.php?id=28). Open reading frames were predicted using an in-house pipeline and GeneMarkS software [26]. The genomic architectures were compared using Owen [27], Mauve [28] and MUMmer [29] softwares. The strategy of reciprocal best

BLASTp hits [30] was used to identify orthologous sets of genes using the Proteinortho tool [31]. An e-value below 1e-3, an amino acid identity above 30 % and sequence coverage above 70 % were used to consider hits as significant. The tRNAScanSE tool was used to search for transfer RNA genes [32]. BLASTp was performed against the NCBI GenBank non-redundant protein sequence database (nr) with an e-value lower than 1e-3 and alignment length greaterthan 80 amino acids (for alignment lengths <80 amino acids, we used an e-value of 1e-5) to identify ORFans. The same e-value cutoff has been used in previous studies to define ORFans [33]. The sequence alignments were built using the muscle program [34], and the alignments were trimmed using the trimAl tool [35]. The phylogenetic trees were constructed using PhyML [36] and were visualized using FigTree software (http://www. umiacs.umd.edu/∼morariu/figtree/).

## Results

The Courdo11 virus genome (GenBank Accession No. JX975216) is a double stranded DNA molecule composed of 1,245,674 nucleotides. It is 13,523 nt shorter than the *M. chilensis* genome [4] and is currently the second largest mimivirus genome that has been described. A total of 1,166 predicted proteins were tentatively identified in the Courdo11 virus genome, and 1,085 of these were greater than 100 amino acids in size. The sizes of the Courdo11 virus predicted proteins range from 48 to 2,605 amino acids in length with a mean size of 312 amino acids. The genome has a high coding density of 84 % with a mean distance of 176 nucleotides separating the coding sequences. The predicted genes are evenly distributed on both strands; 597
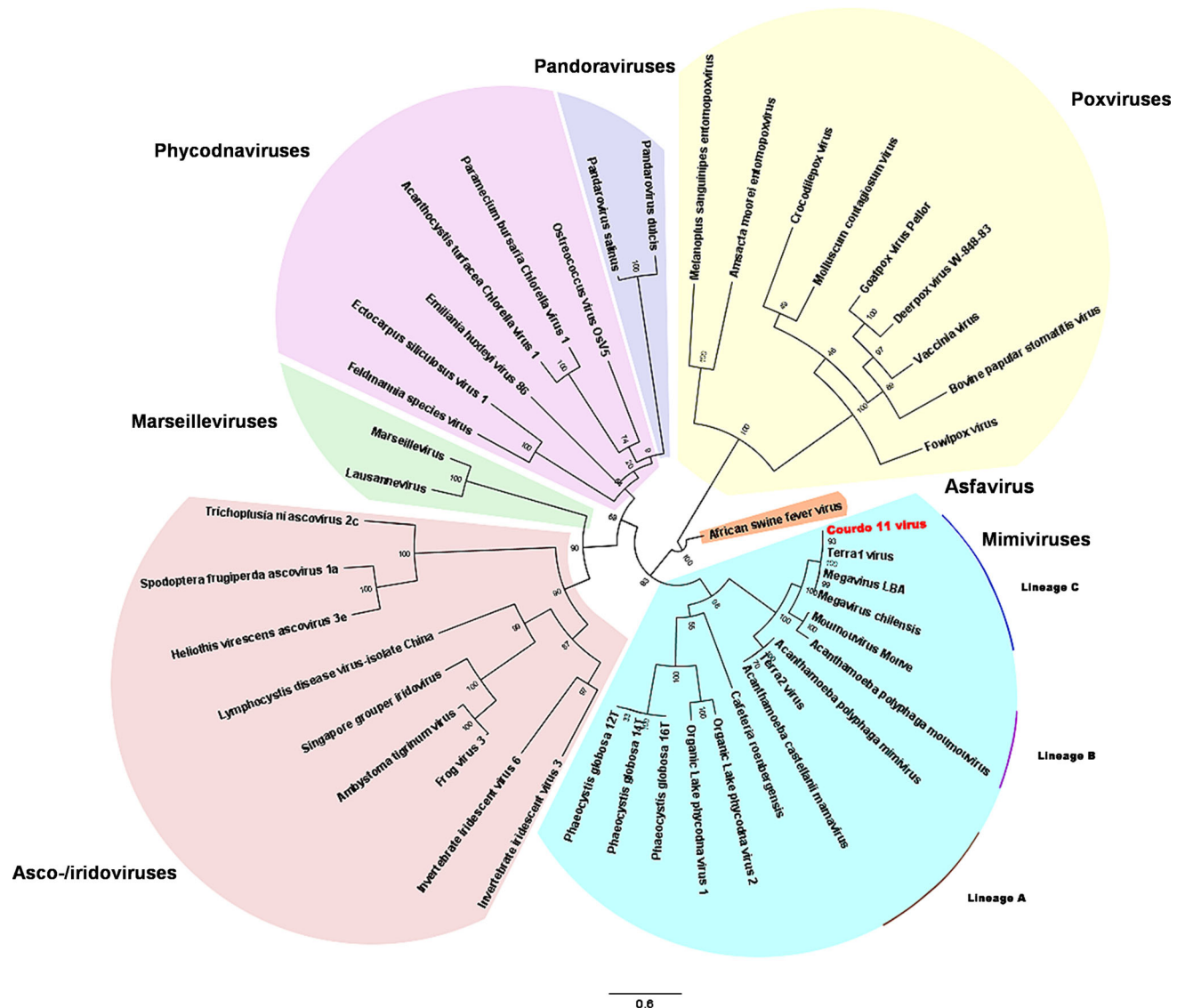
**Fig. 2** The phylogenetic tree was constructed based on the family B DNA polymerases from selected members of the order "Megavirales" [7–9] and from *Pandoravirus dulcis* and *P. salinus* [11] using the maximum likelihood method. The sequence alignment was generated using the muscle program [34], and the trimAl tool [35] was used for automated alignment trimming. The phylogenetic tree was constructed using PhyML [36] and visualized with the FigTree software (http://www.umiacs.umd.edu/∼morariu/figtree/). The numbers of tree nodes indicate bootstrap replicates of 100

and 569 predicted genes are located on the positive and negative strands, respectively. Phylogenetic reconstructions of the family B DNA polymerase of the proposed order "Megavirales" (that encompasses nucleocytoplasmic large DNA viruses) [7–9] indicates that Courdo11 virus belongs to the lineage C of amoeba-associated mimiviruses [7] and is closely related to *M. chilensis* recovered from marine coastal water in Chile (Fig. 2) [4].

Comparative analyses of Courdo11 virus and other mimiviruses showed that the genomes of Courdo11 virus and *M. chilensis* are highly similar and collinear, and the highest levels of divergence were located near the ends of the Courdo11 virus genome. Overall, the alignment of

these two viral genomes using Owen revealed 70 regions larger than 5,000 nt with a mean nucleotide identity of 98.1 % (range, 93.8–99.6) (Fig. 3). A total of 1,137 of the 1,166 predicted proteins of Courdo11 virus (97.5 % of its gene repertoire) are homologous to *M. chilensis* with a mean amino acid identity of 95.5 %. Moreover, Courdo11 virus shares 1,130 (96.9 %) homologous proteins with LBA111. The reciprocal best hits strategy using Proteinortho to compare the Courdo11 virus predicted proteins with those of the members of family *Mimiviridae* showed that the Courdo11 virus shares a high number of orthologous genes, 860 (73.7 %) and 857 (73.4 %) of its gene repertoire, with *M. chilensis* and LBA 111, respectively,
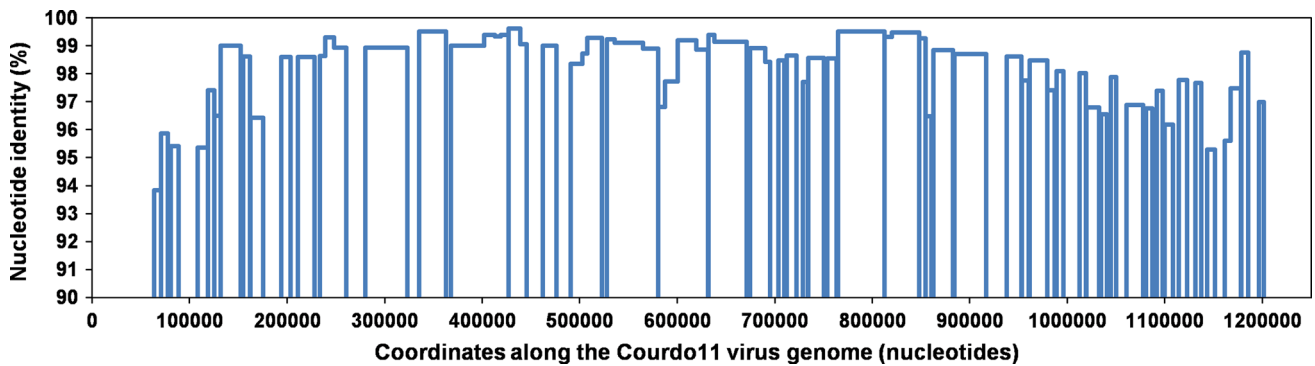
**Fig. 3** Nucleotide identity along the Courdo11 virus and *Megavirus chilensis* [4] genomes for the largest (>5,000 nt) matching regions
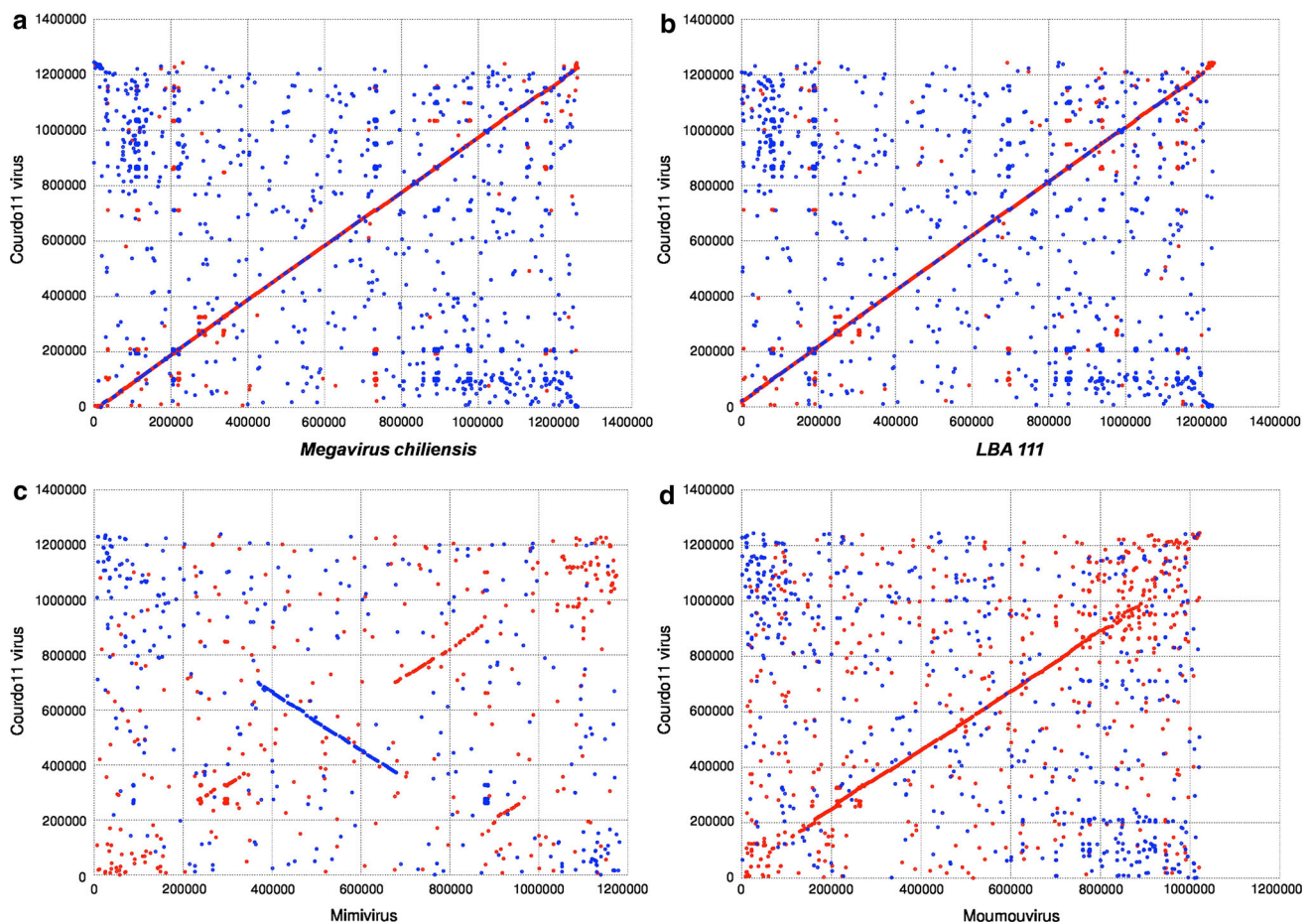


**Fig. 4** Genomic dot plots for the Courdo11 virus against amoebae-associated mimiviruses belonging to lineages A, B and C. The Courdo11 virus genome was compared to the **a** *Megavirus chilensis* [4], **b** LBA 111, **c** Mimivirus [1] and **d** Moumouvirus [5] genomes. *Dot plots* were constructed using MUMmer 3.22 software [29], and nucleotide-based alignments were performed with MUMmer. *Dot plots* were generated by the MUMmerplot script and the program gnuplot (www.gnuplot.info/docs_4.0/gnuplot.html). The aligned segments are represented by *dots* or *lines*. Forward matches are shown in *red*, and reverse complement matches are shown in *blue* (Color figure online)

which are members of lineage C of mimiviruses of amoebae. In addition, Courdo11 virus shares 393 (33.7 %) and 450 (38.5 %) orthologous genes with Mimivirus, the founding member of lineage A and with Moumouvirus, the founding member of lineage B, respectively. The dot plots

of gene repertoires for the Courdo11 virus genome with the genomes of members of lineage C, *M. chilensis* and LBA111, shows high levels of synteny and a near-perfect collinearity (Fig. 4a, b). The genomic dot plot of Courdo11 virus against Mimivirus [1] shows shorter interrupted

collinear regions with a large inversion in the central part of the genome (Fig. 4c), whereas the dot plot against Moumouvirus [5] shows a near-perfect collinearity in the middle part of the genome and rearrangements towards the extremities (Fig. 4d).

Overall, BLASTp searches against the NCBI GenBank non-redundant protein sequence database identified *bona fide* homologues for 1,152 Courdo11 virus predicted proteins (99 % of the gene repertoire), and 14 (1.2 %) ORFans were identified. The main components of the *M. chilensis* gene content are also found in the Courdo11 virus genome, including three amino acyl-tRNA synthetases (mg743, mg844, mg358) that are absent in Mimivirus, a putative DNA photolyase (mg779) and a uridine monophosphate kinase (mg431). In contrast, 6 tRNAs were detected in the Courdo11 virus genome, including three tRNA-Leu, one tRNA-Cys, one tRNA-His and one tRNA-Trp, whereas only 3 tRNAs were identified in *M. chilensis* (1 Trp, and 2 Leu). We extended our analysis of the Courdo11 virus by comparing it with the newly identified viruses *Pandoravirus dulcis* and *Pandarovirus salinus* [11]. A BLASTp search of the Courdo11 virus gene content against those of pandoraviruses using 1e-3 as e-value cut off yielded 150 (12.8 %) and 132 (11.3 %) significant hits with *P. salinus* and *P. dulcis,* respectively.

## Discussion

The comparative analyses of Courdo11 virus with the genomes of other members of family *Mimiviridae* showed that this giant virus is a *bona fide* new member of the family *Mimiviridae* and belongs to lineage C of mimiviruses of amoebae. Genomic architecture comparisons mirrored previous findings that showed conservation of collinear regions in the middle part of the genome and diversity towards the extremities; this feature was indeed described in other mimiviruses and in poxviruses [4, 5, 37–39]. Further analyses showed that the Courdo11 virus genome is most closely related to *M. chilensis* [4] and LBA 111 [20], the first mimivirus isolated from a human. The evolutionary relationship between Courdo11 virus and other mimiviruses isolated from freshwater (Mimivirus and Moumouvirus), marine coastal water (*M. chilensis*) and soil (Terra1 virus and Terra2 virus) indicates that these giant viruses have a broad host range and can survive in different habitats. Major characteristics of the *M. chilensis* genome were identified in the Courdo11 virus genome, which was found to encode three more tRNAs. Fourteen ORFans were identified in the Courdo11 virus genome, suggesting that the pan-genome of mimiviruses of amoeba might reach a plateau.

## References

1. D. Raoult, S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan, J.M. Claverie, Science **306**, 1344–1350 (2004)
2. M. Boyer, N. Yutin, I. Pagnier, L. Barrassi, G. Fournous, L. Espinosa, C. Robert, S. Azza, S. Sun, M.G. Rossmann, M. Suzan-Monti, B. La Scola, E.V. Koonin, D. Raoult, Proc. Natl. Acad. Sci. USA **106**, 21848–21853 (2009)
3. V. Thomas, C. Bertelli, F. Collyn, N. Casson, A. Telenti, A. Goesmann, A. Croxatto, G. Greub, Environ. Microbiol. **13**, 1454–1466 (2011)
4. D. Arslan, M. Legendre, V. Seltzer, C. Abergel, J.M. Claverie, Proc. Natl. Acad. Sci. USA **108**, 17486–17491 (2011)
5. N. Yoosuf, N. Yutin, P. Colson, S.A. Shabalina, I. Pagnier, C. Robert, S. Azza, T. Klose, J. Wong, M.G. Rossmann, B. La Scola, D. Raoult, E.V. Koonin, Genome Biol. Evol. **4**, 1324–1330 (2012)
6. B. La Scola, A. Campocasso, R. N'Dong, G. Fournous, L. Barrassi, C. Flaudrops, D. Raoult, Intervirology **53**, 344–353 (2010)
7. P. Colson, X. de Lamballerie, G. Fournous, D. Raoult, Intervirology **55**, 321–332 (2012)
8. P. Colson, X. de Lamballerie, N. Yutin, S. Asgari, Y. Bigot, D.K. Bideshi, X.W. Cheng, B.A. Federici, J.L. Van Etten, E.V. Koonin, B. La Scola, D. Raoult, Arch. Virol. (2013). doi:10.1186/1743-422X-10-158
9. N. Yutin, P. Colson, D. Raoult, E.V. Koonin, Virol. J. **10**, 106 (2013)
10. N. Yutin, E.V. Koonin, Virol. J. **9**, 161 (2012)
11. N. Philippe, M. Legendre, G. Doutre, Y. Coute, O. Poirot, M. Lescot, D. Arslan, V. Seltzer, L. Bertaux, C. Bruley, J. Garin, J.M. Claverie, C. Abergel, Science **341**, 281–286 (2013)
12. D. Raoult, P. Forterre, Nat. Rev. Microbiol. **6**, 315–319 (2008)
13. E.V. Koonin, N. Yutin, Intervirology **53**, 284–292 (2010)
14. I. Pagnier, D.G. Reteno, H. Saadi, M. Boughalmi, M. Gaia, M. Slimani, T. Ngounga, M. Bekliz, P. Colson, D. Raoult, B. La Scola, Intervirology **56**, 354–363 (2013)
15. M.G. Fischer, M.J. Allen, W.H. Wilson, C.A. Suttle, Proc. Natl. Acad. Sci. USA **107**, 19508–19513 (2010)
16. C. Desnues, B. La Scola, N. Yutin, G. Fournous, C. Robert, S. Azza, P. Jardot, S. Monteil, A. Campocasso, E.V. Koonin, D. Raoult, Proc. Natl. Acad. Sci. USA **109**, 18078–18083 (2012)
17. N. Yutin, D. Raoult, E.V. Koonin, Virol. J. **10**, 158 (2013)
18. P. Colson, B. La Scola, D. Raoult, Intervirology **56**, 376–385 (2013)
19. P. Colson, D. Raoult, Megavirales composing a fourth domain of life: Mimiviridae and Marseilleviridae, *Viruses: essential agents of life* (Springer, Dordrecht, 2012)
20. H. Saadi, I. Pagnier, P. Colson, J. Kanoun Cherif, M. Beji, M. Boughalmi, S. Azza, N. Armstrong, C. Robert, G. Fournous, B. La Scola, D. Raoult, Clin. Infect. Dis. **57**, e127–e134 (2013)
21. H. Saadi, D.G. Ikanga Reteno, P. Colson, S. Aherfi, P. Minodier, I. Pagnier, F. Fenollar, C. Robert, D. Raoult, B. La Scola, Intervirology **56**, 424–429 (2013)
22. A. Vincent, B. La Scola, L. Papazian, Intervirology **53**, 304–309 (2010)
23. P. Colson, L. Fancello, G. Gimenez, F. Armougom, C. Desnues, G. Fournous, N. Yoosuf, M. Million, B. La Scola, D. Raoult, J. Clin. Virol. **57**(3), 191–200 (2013)
24. B. La Scola, C. Desnues, I. Pagnier, C. Robert, L. Barrassi, G. Fournous, M. Merchat, M. Suzan-Monti, P. Forterre, E. Koonin, D. Raoult, Nature **455**, 100–104 (2008)
25. M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen,

S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, J.M. Rothberg, Nature **437**, 376–380 (2005)

26. J. Besemer, M. Borodovsky, Nucleic Acids Res. **33**, W451–W454 (2005)
27. A.Y. Ogurtsov, M.A. Roytberg, S.A. Shabalina, A.S. Kondrashov, Bioinformatics **18**, 1703–1704 (2002)
28. A.C. Darling, B. Mau, F.R. Blattner, N.T. Perna, Genome Res. **14**, 1394–1403 (2004)
29. A.L. Delcher, S.L. Salzberg, A.M. Phillippy, Curr. Protoc. Bioinformatics (2003). doi:10.1002/0471250953.bi1003s00. chapter 10:Unit 10.3
30. I.K. Jordan, I.B. Rogozin, Y.I. Wolf, E.V. Koonin, Genome Res. **12**, 962–968 (2002)
31. M. Lechner, S. Findeiss, L. Steiner, M. Marz, P.F. Stadler, S.J. Prohaska, BMC Bioinformatics **12**, 124 (2011)
32. P. Schattner, A.N. Brooks, T.M. Lowe, Nucleic Acids Res. **33**, W686–W689 (2005)
33. Y. Yin, D. Fischer, BMC Genomics **9**, 24 (2008)
34. R.C. Edgar, Nucleic Acids Res. **32**, 1792–1797 (2004)
35. S. Capella-Gutierrez, J.M. Silla-Martinez, T. Gabaldon, Bioinformatics **25**, 1972–1973 (2009)
36. S. Guindon, F. Lethiec, P. Duroux, O. Gascuel, Nucleic Acids Res. **33**, W557–W559 (2005)
37. T.G. Senkevich, E.V. Koonin, J.J. Bugert, G. Darai, B. Moss, Virology **233**, 19–42 (1997)
38. A. McLysaght, P.F. Baldi, B.S. Gaut, Proc. Natl. Acad. Sci. USA **100**, 15655–15660 (2003)
39. P. Colson, G. Gimenez, M. Boyer, G. Fournous, D. Raoult, PLoS One **6**, e18935 (2011)